Methods in Molecular Biology 1837

Springer Protocols

Remus T. Dame *Editor*

Bacterial Chromatin

Methods and Protocols

EXTRAS ONLINE

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Bacterial Chromatin

Methods and Protocols

Edited by

Remus T. Dame

Leiden Institute of Chemistry and Centre for Microbial Cell Biology, Leiden University, Leiden, The Netherlands

Humana Press

Editor
Remus T. Dame
Leiden Institute of Chemistry
and Centre for Microbial Cell Biology
Leiden University
Leiden, The Netherlands

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-8674-3 ISBN 978-1-4939-8675-0 (eBook)
https://doi.org/10.1007/978-1-4939-8675-0

Library of Congress Control Number: 2018950447

This Humana Press imprint is published by the registered company Springer Science+Business Media, LLC part of Springer Nature.
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Bacterial chromatin is the nucleoprotein complex consisting of genomic DNA and its associated proteins in bacteria. In recent years, it has been increasingly appreciated that bacterial chromatin is organized at various length scales and that organization affects processes occurring within this structure and *vice versa*. Nevertheless, detailed information of this structure–function relationship is lacking. The methods brought together in this book can be used to explore this relationship ranging from the molecular to the cellular scale. Next to detailed experimental protocols of *in vivo* and *in vitro* approaches, theoretical contributions describing approaches to genome structure modeling and data analysis are included. The book will be a useful up-to-date reference work for scholars in the bacterial chromatin field, as well as others entering the field from adjacent research fields (transcription, replication, and repair in bacteria), as well as the eukaryotic chromatin field.

Leiden, The Netherlands ***Remus T. Dame***

Contents

Contributors

Julius Battjes • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

Nicole A. Becker • *Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine and Science, Rochester, MN, USA*

Beth A. Boudreau • *Department of Biochemistry, University of Wisconsin – Madison, Madison, WI, USA*

Thomas B. Brouwer • *Huygens-Kamerlingh Onnes Laboratory, Leiden Institute of Physics, Leiden University, Leiden, The Netherlands*

Cedric Cagliero • *Transcription Control Section, RNA Biology Laboratory, National Cancer Institute, National Institutes of Health, Frederick, MD, USA; Jecho Laboratories Inc., Frederick, MD, USA*

Debasish Chaudhuri • *Institute of Physics, Sachivalaya Marg, Bhubaneswar, India; Homi Bhaba National Institute, Mumbai, India*

De Chen • *Ras Initiative, Frederick National Laboratory for Cancer Research (FNLCR), Leidos Biomedical Research, Inc., Frederick, MD, USA*

Onuma Chumsakul • *Graduate School of Biological Sciences, Nara Institute of Science and Technology, Nara, Japan; Basic Research and Development Division, Rohto Pharmaceutical Co., Ltd., Kyoto, Japan*

Ivan Corbeski • *Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*

Frédéric G. Crémazy • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

Remus T. Dame • *Leiden Institute of Chemistry and Centre for Microbial Cell Biology, Leiden University, Leiden, The Netherlands*

David C. Grainger • *School of Biosciences, Birmingham University, Birmingham, UK*

James R. Haycocks • *School of Biosciences, Birmingham University, Birmingham, UK*

Dieter W. Heermann • *Institute for Theoretical Physics, Heidelberg University, Heidelberg, Germany*

Joost Heinsman • *Leiden Institute of Physics, Leiden University, Leiden, The Netherlands*

Bram Henneman • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

Andreas Hofmann • *Institute for Theoretical Physics, Heidelberg University, Heidelberg, Germany*

Velten Horn • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

Akira Ishihama • *Research Center for Micro-Nano Technology, Hosei University, Tokyo, Japan*

Shu Ishikawa • *Graduate School of Biological Sciences, Nara Institute of Science and Technology, Nara, Japan; Graduate School of Science, Technology and Innovation, Kobe University, Kobe, Hyogo, Japan*

Ding Jun Jin • *Transcription Control Section, RNA Biology Laboratory, National Cancer Institute, National Institutes of Health, Frederick, MD, USA*

Artur Kaczmarczyk • *Huygens-Kamerlingh Onnes Laboratory, Leiden Institute of Physics, Leiden University, Leiden, The Netherlands*

Hyeongjun Kim • *Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA*
Matthew V. Kotlajich • *Department of Biochemistry, University of Wisconsin – Madison, Madison, WI, USA*
Lisa E. Lamberte • *School of Biosciences, Birmingham University, Birmingham, UK*
Robert Landick • *Department of Biochemistry, University of Wisconsin – Madison, Madison, WI, USA; Department of Bacteriology, University of Wisconsin – Madison, Madison, WI, USA*
Szu-Ning Lin • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands; Department of Physics and Astronomy, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
Joseph J. Loparo • *Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA*
L. James Maher III • *Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine and Science, Rochester, MN, USA*
Carmen Mata Martin • *Transcription Control Section, RNA Biology Laboratory, National Cancer Institute, National Institutes of Health, Frederick, MD, USA*
Hugo Maruyama • *Department of Bacteriology, Osaka Dental University, Osaka, Japan*
Geri F. Moolenaar • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*
Kazuya Morikawa • *Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan*
Bela M. Mulder • *Institute AMOLF, Amsterdam, The Netherlands*
Kensuke Nakamura • *Department of Life Science and Informatics, Maebashi Institute of Technology, Maebashi, Gunma, Japan*
Hiroshi Ogasawara • *Division of Gene Research, Research Center for Supports to Advanced Science, Shinshu University, Ueda, Nagano, Japan*
Ryosuke L. Ohniwa • *Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan; Center for Biotechnology, National Taiwan University, Taipei, Taiwan*
Taku Oshima • *Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, Japan; Department of Biotechnology, Toyama Prefectural University, Imizu, Toyama, Japan*
Ulric B. le Paige • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands; Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
Justin P. Peters • *Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine and Science, Rochester, MN, USA*
Chi Pham • *Huygens-Kamerlingh Onnes Laboratory, Leiden Institute of Physics, Leiden University, Leiden, The Netherlands*
Liang Qin • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*
Fatema-Zahra M. Rashid • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*
Daniel Schindler • *Chromosome Biology Group, LOEWE Center for Synthetic Microbiology—SYNMIKRO, Philipps-Universität Marburg, Marburg, Germany; School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester, UK*
Karin Schnetz • *Institute for Genetics, University of Cologne, Cologne, Germany*
Tomohiro Shimada • *School of Agriculture, Meiji University, Kawasaki, Kanagawa, Japan*

Zhe Sun • *Transcription Control Section, RNA Biology Laboratory, National Cancer Institute, National Institutes of Health, Frederick, MD, USA*
Kunio Takeyasu • *Center for Biotechnology, National Taiwan University, Taipei, Taiwan; Graduate School of Biostudies, Kyoto University, Kyoto, Japan*
Ramon A. van der Valk • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*
Hugo van Ingen • *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands; Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
John van Noort • *Huygens-Kamerlingh Onnes Laboratory, Leiden Institute of Physics, Leiden University, Leiden, The Netherlands*
Torsten Waldminghaus • *Chromosome Biology Group, LOEWE Center for Synthetic Microbiology—SYNMIKRO, Philipps-Universität Marburg, Marburg, Germany*
Yanwei Wang • *Department of Physics, Wenzhou University, Wenzhou, China*
Gijs J. L. Wuite • *Department of Physics and Astronomy, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
Guangcan Yang • *Department of Physics, Wenzhou University, Wenzhou, China*
Celine Zumkeller • *Chromosome Biology Group, LOEWE Center for Synthetic Microbiology—SYNMIKRO, Philipps-Universität Marburg, Marburg, Germany*

Part I

In Vivo Approaches

Chapter 1

Determination of the 3D Genome Organization of Bacteria Using Hi-C

Frédéric G. Crémazy, Fatema-Zahra M. Rashid, James R. Haycocks, Lisa E. Lamberte, David C. Grainger, and Remus T. Dame

Abstract

The spatial organization of genomes is based on their hierarchical compartmentalization in topological domains. There is growing evidence that bacterial genomes are organized into insulated domains similar to the Topologically Associating Domains (TADs) detected in eukaryotic cells. Chromosome conformation capture (3C) technologies are used to analyze in vivo DNA proximity based on ligation of distal DNA segments crossed-linked by bridging proteins. By combining 3C and high-throughput sequencing, the Hi-C method reveals genome-wide interactions within topological domains and global genome structure as a whole. This chapter provides detailed guidelines for the preparation of Hi-C sequencing libraries for bacteria.

Key words Hi-C, Chromosome, Bacterial chromatin

1 Introduction

Over the last decade, Chromosome Conformation Capture (3C)-based methods have been used to resolve the 3D conformation of genomes in eukaryotes [1]. The techniques have revealed how genomes are organized and how this organization can influence gene expression through long-range physical interactions between regulatory elements [2, 3]. Hi-C is possibly the most powerful of the 3C-based methods as it allows the interrogation of all possible interactions across the genome in an unbiased manner at a resolution of up to 5 kb [4].

Only recently, Hi-C has been used to probe chromosome conformation in prokaryotes [5–7]. The first genome-wide interaction frequency maps obtained from *Bacillus subtilis* and *Caulobacter crescentus* demonstrated that, as in eukaryotes, bacterial

Frédéric G. Crémazy and Fatema-Zahra M. Rashid contributed equally to this work

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_1,

genomes are organized into large-scale insulated chromosomal structures called Chromosome Interactions Domains (CID). The formation and maintenance of these domains is correlated with DNA transactions such as transcription [6]. Hi-C has also revealed how the 3D structure of the *ori* macrodomain influences DNA replication initiation and chromosome segregation [7]. These results underline the relevance of using Hi-C to correlate 3D structure of genomes with genome function.

Like other 3C-based methods, Hi-C relies on the proximity ligation of DNA fragments connected together by architectural proteins to probe the 3D organization of the chromosome (*see* Fig. 1). First, a snapshot of the chromosome is obtained by treating

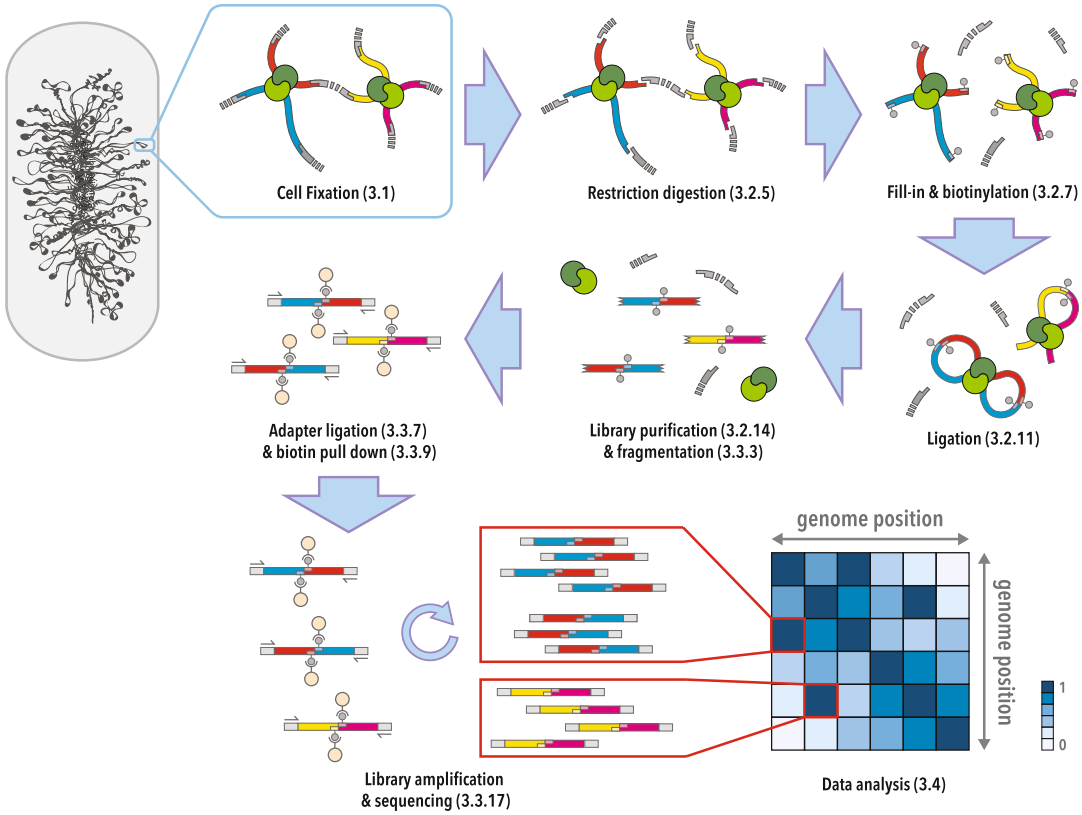


Fig. 1 Schematic overview of the Hi-C method. Bacteria are cultured and fixed using formaldehyde to crosslink proteins to DNA (proteins are shown in green, DNA fragments of interest in various colors and other DNA fragments in gray dotted lines). After cell lysis, bacterial chromatin is digested with a restriction enzyme and the cohesive extremities are filled using T4 DNA polymerase in buffer containing a biotinylated nucleotide (light colors and gray circles). The digested DNA fragments are then diluted tenfold and ligated overnight. After cross-link reversal and protein digestion, the purified DNA is fragmented enzymatically. For next-generation sequencing, barcoded adapters are ligated to the DNA and the library is enriched for biotinylated junctions using streptavidin beads (light brown circles). The library is further amplified by PCR and analyzed by paired-end sequencing. After mapping and filtering of all the junctions, genome-wide interactions are displayed as a colored matrix

bacterial cells with a fixative, usually formaldehyde. After lysis of the cells, the chromatin is digested using a restriction enzyme, either a 6- or 4-base pair cutter, fragmenting the genome into a heterogeneous population of DNA-protein complexes. Interactions between distal DNA segments are retained after fragmentation. Hence, overhanging DNA ends generated by restriction can be filled using biotin-labeled nucleotides and blunt end ligated. Ligation is carried out in dilute conditions to favor ligation of DNA fragments in the same cross-linked chromatin complex. Hence, the chance of ligation between unlinked DNA fragments is minimized. After reverse crosslinking and DNA purification, biotin labels from unligated restriction fragment ends are removed using the exonuclease activity of T4 DNA polymerase. The library produced is then fragmented and the biotin-labelled junctions are enriched using streptavidin-coated magnetic beads. The ligation-junctions are identified by paired-end sequencing. After filtering and mapping of the reads to a reference genome, a genome-wide proximity map is generated in the form of a matrix of pairwise interaction frequencies across the genome.

In this chapter, we present a detailed protocol for the preparation of Hi-C libraries from the bacterium *Escherichia coli* and subsequent high-throughput sequencing. Data analysis and methods for generating Hi-C interaction maps are explained in more detail in Chapter 2 of this book.

2 Materials

2.1 Cell Fixation

1. 1× PBS solution: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.4. Place the solution on ice.
2. 37% w/v formaldehyde solution.
3. 3% v/v formaldehyde solution: Add 1.6 mL of 37% w/v formaldehyde solution to 18.4 mL of cold 1× PBS solution (*see* **Note 1**). Freshly prepare the 3% formaldehyde and place it on ice. Keep the solution away from direct sunlight.
4. 2.5 M Glycine solution filter-sterilized using a sterile, non-pyrogenic 0.2 μm filter.
5. 1× TE Buffer: 10 mM Tris–HCl, 1 mM EDTA·Na_2, pH 8.0.
6. Dry ice.
7. Refrigerated benchtop centrifuge (e.g., Eppendorf centrifuge 5810R or 5430R).
8. Roller bench.
9. −80 °C freezer.
10. 50 mL Falcon tubes.

11. Serological pipettes.
12. Pipette pump.
13. Elongated Pasteur pipette (made of soda lime silica glass): Heat a 1 cm space of the Pasteur pipette above the tip and stretch the heated region to reduce the diameter of the pipette bore. Use a pair of tweezers to carefully break the pipette at the elongated region to produce a Pasteur pipette with a narrow opening.
14. 1000 µL micropipette.
15. Sterile 1.5 mL microfuge tubes.

2.2 Hi-C Library Preparation

1. 1× TE Buffer: 10 mM Tris–HCl, 1 mM EDTA·Na_2, pH 8.0.
2. 40,000 U/µL Ready-Lyse Lysozyme (Epicentre).
3. 1.0 M NaCl.
4. Nuclease-free water.
5. 10% w/v sodium dodecyl sulfate (SDS) solution: Dissolve 1.0 g of SDS (ACS reagent, ≥99.0%) in 10 mL of nuclease-free water.
6. 10% v/v Triton X-100 solution: Dilute 1.0 mL of Triton X-100 (molecular biology grade) with 9.0 mL of nuclease-free water. Store the solution at room temperature (20 °C–25 °C) in the dark.
7. 10× restriction enzyme buffer (as supplied with the restriction enzyme).
8. Selected restriction enzyme (*see* **Note 2**).
9. 100 mM deoxynucleotides (ultrapure dNTPs as separate solutions). Repeated freeze-thaw cycles may result in the degradation of dNTPs. To minimize degradation of the 100 mM stock, dilute 20.0 µL of 100 mM dNTP with 180.0 µL of nuclease-free water. Store the 10 mM dNTP solution at −20 °C as 50.0 µL aliquots.
10. Separate solutions of 1 mM dNTP: Dilute one 10 mM dNTP aliquot tenfold with nuclease-free water. Store the dNTP solution at −20 °C and thaw on ice when necessary.
11. 0.4 mM biotin-14-dATP (*see* **Note 3**).
12. 5 U/µL DNA Polymerase I, Large (Klenow) fragment.
13. 20 mg/mL bovine serum albumin (BSA).
14. 10× T4 DNA Ligase buffer: 0.5 M Tris–HCl, 0.1 M $MgCl_2$, 10 mM ATP, 0.1 M DTT, pH 7.5.
15. 2000 U/µL T4 DNA Ligase.
16. 0.5 M EDTA, pH 8.0.
17. 10 mg/mL RNase A solution.
18. 20 mg/mL Proteinase K solution.

19. 5.0 M NaCl solution.
20. 25:24:1 Phenol:chloroform:isoamyl alcohol solution.
21. Chloroform.
22. 1.0 M sodium acetate (NaOAc), pH 8.0.
23. 5.0 mg/mL glycogen.
24. Cold 100% and 70% ethanol solutions.
25. 10 mM Tris, pH 8.0.
26. dsDNA quantification kit (e.g., Qubit® dsDNA HS Assay Kit from ThermoFisher Scientific).
27. 10× T4 DNA Polymerase buffer: 0.5 M NaCl, 0.1 M Tris–HCl, 0.1 M $MgCl_2$, 1 mg/mL BSA, pH 7.9 (available as NEBuffer 2.1).
28. 3 U/μL T4 DNA Polymerase.
29. Thermomixer (e.g., Eppendorf Thermomixer® C).
30. Benchtop centrifuge (e.g., Eppendorf MiniSpin® plus).
31. Vacuum concentrator.
32. Vortex mixer.
33. −20 °C freezer.
34. 2.5 μL, 20 μL, 200 μL, and 1000 μL micropipettes.
35. Sterile 1.5 mL microfuge tubes.
36. Agarose gel electrophoresis setup.

2.3 NGS Library Preparation

1. Solid Phase Reverse Immobilization (SPRI) beads (e.g., Agencourt AMPure XP from Beckman). Store the beads as 1.0 mL aliquots at 4 °C. Ensure that the beads are equilibrated to room temperature and vortexed before use.
2. Elution buffer: 10 mM Tris, pH 8.0.
3. Elution buffer + 0.1% Tween 20.
4. Illumina® library preparation kit (e.g., KAPA HyperPlus kit from KAPA Biosystems).
5. Illumina®-compatible paired-end sequencing adapters at a concentration of 15 μM (e.g., KAPA Single-Indexed Adapter Sets A and B from KAPA Biosystems). The sequencing adapters should be barcoded to allow multiplexing during NGS.
6. Magnetic rack (e.g., DynaMag™-PCR Magnet from ThermoFisher Scientific).
7. Streptavidin-coupled magnetic beads optimized for the enrichment of biotin-labeled nucleic acids (e.g., Dynabeads® MyOne™ Streptavidin T1 from ThermoFisher Scientific). Store the beads at 4 °C and vortex well before use.

8. 2× Beads Wash Buffer (2× BWB): 10 mM Tris pH 7.5, 1 mM EDTA, 2 mM NaCl, 0.1% Tween 20.
9. 1× Beads Wash Buffer (1× BWB): 5 mM Tris pH 7.5, 0.5 mM EDTA, 1 mM NaCl, 0.1% Tween 20.
10. dsDNA quantification kit (e.g., Qubit® dsDNA HS Assay Kit from ThermoFisher Scientific).
11. Tapestation (Agilent Technologies) or Bioanalyzer (Agilent Technologies).
12. Thermal cycler with a heated lid.
13. Vortex mixer.
14. PCR tubes.

3 Methods

This protocol is optimized for HiC in *Escherichia coli*. To process other bacteria using this protocol, optimize the cell lysis and solubilization steps (*see* **steps 2** and **3** of Subheading 3.2) by adjusting the treatment duration and/or the concentrations of Ready-Lyse Lysozyme and SDS.

3.1 Cell Fixation

1. Transfer 6.0 mL of the *Escherichia coli* culture at the mid-exponential phase (OD_{600} of 1.0) into a prechilled 50 mL Falcon tube (*see* **Note 4**).
2. Collect the cells by centrifugation at 5000 rpm (3400 × *g*) for 5 min at 4 °C. Decant the supernatant and resuspend the cell pellet in 18.0 mL of cold 1× PBS to wash the bacterial cells (*see* **Note 1**).
3. Centrifuge the cells at 5000 rpm (3400 × *g*) for 5 min at 4 °C and discard the supernatant.
4. Resuspend the cell pellet in 18.0 mL of 3% formaldehyde solution. Incubate the cell suspension at 4 °C for 1 h on a roller bench at 30 rpm.
5. Add 3.2 mL of 2.5 M glycine to the fixation reaction (to a final concentration of 0.375 M) to quench the excess formaldehyde. Incubate the cells for 15 min at 4 °C on a roller bench at 30 rpm.
6. Collect the fixed cells by centrifugation at 5000 rpm (3400 × *g*) for 5 min at 4 °C. Pour out the supernatant and carefully remove the rest with a micropipette. Resuspend the cell pellet in 5.0 mL of 1× TE.
7. Divide the suspension of fixed cells into five aliquots of 1.0 mL in sterile 1.5 mL microfuge tubes (*see* **Note 5**).

8. Pellet the fixed cells by centrifugation at 10,000 rpm (10,600 × *g*) for 2 min at 4 °C. Remove the supernatant with an elongated Pasteur pipette.
9. Flash freeze the cell pellets in dry ice and store the fixed cells at −80 °C for up to 2 months.

3.2 Hi-C Library Preparation

1. Resuspend a pellet of fixed cells in 50.0 μL of 1× TE and place the cell suspension on ice.
2. Prepare a 1:10 dilution of 40,000 U/μL Ready-Lyse Lysozyme (Epicentre) with 4.0 μL of 1× TE, 0.5 μL of 1.0 M NaCl, and 0.5 μL of 40,000 U/μL Ready-Lyse Lysozyme. Add 0.5 μL of the diluted enzyme to the cell suspension. Incubate the tube for 15 min at 37 °C in a Thermomixer at 1000 rpm to lyse the cells.
3. Add 2.5 μL of 10% SDS to the lysis reaction and mix by pipetting. Incubate the tube for 15 min at 37 °C in a Thermomixer at 1000 rpm.
4. 5.0 μL of the cell lysate diluted with 45.0 μL of 1× TE may be kept aside at −20 °C as "chromatin." This control serves as a test to determine whether the extracted chromatin undergoes degradation during cell lysis and solubilization (*see* Fig. 2) (*see* **steps 2** and **3** of Subheading 3.2). Continue with processing this control at **step 13** of Subheading 3.2 and adjust the volume of the cell lysate to 53.0 μL with 5.0 μL of 1× TE to continue with Hi-C library preparation.
5. Transfer 53.0 μL of the cell lysate into 186.0 μL of the restriction digestion mix (136.0 μL of nuclease-free water, 25.0 μL of 10% Triton X-100, and 25.0 μL of 10× restriction enzyme buffer). Incubate the sample at 37 °C for 10 min in a Thermomixer at 1000 rpm to sequester SDS in the cell lysate with Triton X-100 in the digestion mix.
6. Add 100 U of Restriction enzyme to the digestion mix and, if necessary, make up the volume of the reaction to 250 μL with nuclease-free water. Incubate the reaction at 37 °C for 3 h with shaking at 450 rpm.
7. Fill in the restriction ends and tag them with biotin by adding the following to 250 μL of the digestion reaction: 3.0 μL of 1.0 mM dTTP, 3.0 μL of 1.0 mM dGTP, 3.0 μL of 1.0 mM dCTP, 7.5 μL of 0.4 mM biotin-14-dATP, 1.5 μL of 20 mg/mL BSA, 26.0 μL of nuclease-free water, and 6.0 μL of 5 U/μL DNA Polymerase I, Large (Klenow) fragment (*see* **Note 3**). Mix well by gently inverting the tube and incubate the reaction for 45 min at 25 °C in a Thermomixer at 450 rpm.
8. Inactivate the restriction enzyme and Klenow fragment by adding 15.8 μL of 10% SDS to the fill-in reaction to a final

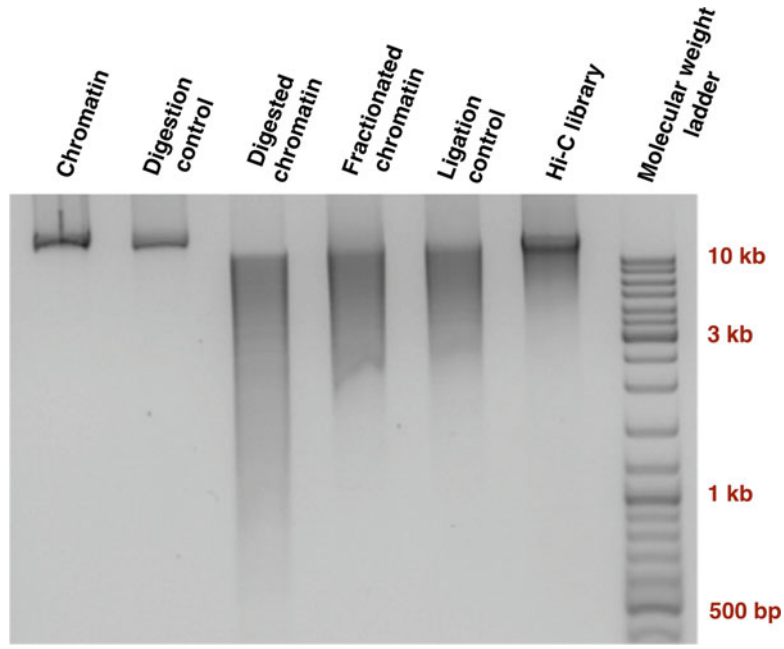


Fig. 2 Typical results obtained for different steps of the Hi-C library preparation. Chromatin: Chromatin extracted from fixed cells during lysis and solubilization runs as a high molecular weight band >10 kb. Digestion control: Chromatin incubated in the restriction digestion mix for three hours in the absence of a restriction enzyme should appear as a high molecular weight band similar to the chromatin lane. A smear in this control indicates that the chromatin undergoes degradation during restriction digestion. Digested chromatin: Chromatin digested with a 6-cutter should run as a smear with a bulk of the fragments having a size >2 kb. Fractionated chromatin: Fractionation pellets protein-cross linked DNA. "Free floating" DNA fragments that may contribute to random ligation and self-circularized products are eliminated. This step is associated with a loss of shorter fragments. Ligation control: Fractionated chromatin in the ligation mix in the absence of T4 DNA Ligase should run on the agarose gel with a profile comparable to fractionated chromatin. Degradation will be observed as a shift of the profile towards shorter DNA lengths. Hi-C library: A well-ligated Hi-C library shows a shift of the DNA profile towards high molecular weights

concentration of 0.5%. Mix well by gentle inversion and incubate the sample for 20 min at 25 °C in a Thermomixer at 1000 rpm. Avoid inactivation at higher temperatures to minimize premature reverse cross-linking of the chromatin fragments.

9. Fractionate the chromatin by centrifuging the sample at 16,000 rpm (25,200 × *g*) for 1 h at 4 °C and discard the supernatant. Resuspend the gel-like pellet in 200.0 μL of Nuclease-Free Water. Use 2.0 μL of the sample to determine the DNA concentration using the Qubit® dsDNA HS Assay Kit.

10. Keep a 1.0 μg aliquot aside as the R+ control. This control is used to visualize the efficiency of restriction digestion (*see* Fig. 2). Make up the volume of the control to 50.0 μL using 1× TE and store it at −20 °C. Continue with processing the R+ control at **step 13** of this subsection.
11. Adjust the volumes of 1.0–2.0 μg aliquots of resuspended chromatin to 820 μL with Nuclease-Free Water. Add 100.0 μL of 10× T4 DNA Ligase Buffer, 75.0 μL of 10% Triton X-100, and 5.0 μL of 20 mg/mL BSA. Mix well by inversion and incubate the samples at 16 °C for 20 min in a Thermomixer at 1000 rpm to sequester SDS.
12. Add 3.0 μL of 2000 U/μL T4 DNA Ligase and incubate the reaction for 16 h at 16 °C. Terminate the ligation reaction by adding 20.5 μL of 0.5 M EDTA.
13. Add 16.6 μL of 1:10 diluted 100 mg/mL RNase A to the ligation reaction. Use 0.8 μL for the Chromatin and R+ control (*see* **steps 4** and **10** of Subheading 3.2). Mix by pipetting and incubate the samples for 30 min at 37 °C in a Thermomixer at 450 rpm.
14. Add 12.5 μL of Qiagen Proteinase K solution and 120.8 μL of 5.0 M NaCl to the ligation reaction. To the Chromatin and R+ control (*see* **steps 4** and **10** of Subheading 3.2), add 0.64 μL of Qiagen Proteinase K solution and 6.2 μL of 5.0 M NaCl. Incubate the samples for 6 h at 65 °C in a Thermomixer at 450 rpm to reverse formaldehyde-mediated cross-links and digest proteins.
15. Divide the ligation reaction over two tubes. Add an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol to ligation samples and controls and vortex the tubes for 10 s. Centrifuge the samples for 10 min at 14,000 rpm (19,300 × *g*) on a benchtop centrifuge and transfer the top aqueous layer into a sterile 1.5 mL microfuge tube. Perform this step twice.
16. Place the ligation solutions in a Vacuum Concentrator for 1 h at 60 °C to reduce the sample volumes from 600 to 200 μL. Pool the solutions.
17. Add 0.1 volumes of 1.0 M Sodium Acetate pH 8.0 and 0.025 volumes of 5.0 mg/mL Glycogen to each sample and vortex well. Add 2.5 volumes of cold 100% ethanol. Mix the solutions well by inversion or vortexing and incubate the samples overnight at −20 °C.
18. Centrifuge the precipitation solutions for 20 min at 16,000 rpm (25,200 × *g*) at 4 °C. Remove the supernatant with an elongated Pasteur pipette. A white pellet should be visible at the base of the microfuge tube.

19. Add 500 µL of cold 70% ethanol to the microfuge tube to wash the pellet. Centrifuge the tube at 16,000 rpm (25,200 × *g*) for 5 min at 4 °C and remove the supernatant with an elongated Pasteur pipette. Perform this step twice.
20. Dry the pellets in a 37 °C incubator for 15 min or on a benchtop for 1 h. Dissolve each pellet in 17.0 µL of 10 mM Tris.
21. Use 2.0 µL of the purified libraries and controls to determine DNA concentration using the Qubit® dsDNA HS Assay Kit.
22. Visualize between 100 and 500 ng of each sample on a 1.5% Agarose gel. Ensure that between 500 ng to 1.0 µg of the purified, ligated Hi-C library is available for the next steps of the protocol (*see* Fig. 2).
23. To 15.0 µL of the purified Hi-C library, add 23.75 µL of nuclease-free water, 5.0 µL of 10× NEBuffer 2.1, 5.0 µL of 1 mM dGTP, 0.25 µL of 20 mg/mL BSA, and 1.0 µL of 3 U/µL T4 DNA Polymerase. Incubate the reaction at 16 °C for 3 h in a Thermomixer at 450 rpm (*see* **Note 6**).
24. Add an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol to the sample and vigorously vortex the tube for 10 s. Centrifuge the samples for 10 min at 14,000 rpm (19,300 × *g*) on a benchtop centrifuge and transfer the top aqueous layer into a sterile 1.5 mL microfuge tube.
25. Add 0.1 volumes of 3.0 M Sodium Acetate pH 5.6 and 0.025 volumes of 5.0 mg/mL Glycogen and vortex the sample well. Add 2.5 volumes of cold 100% ethanol and mix well by inversion. Incubate the sample at −20 °C for at least 1 h.
26. Centrifuge the precipitation solution for 20 min at 16,000 rpm (25,200 × *g*) at 4 °C and remove the supernatant with an elongated Pasteur pipette. A white pellet should be visible at the base of the microfuge tube.
27. Add 500 µL of cold 70% ethanol to the microfuge tube to wash the pellet. Centrifuge the tube at 16,000 rpm (25,200 × *g*) for 5 min at 4 °C and remove the supernatant with an elongated Pasteur pipette.
28. Dry the pellet in a 37 °C incubator for 15 min or on a benchtop for 1 h. Dissolve the pellet in 17.0 µL of 10 mM Tris. Use 2.0 µL of the library to determine DNA concentration using the Qubit® dsDNA HS Assay Kit. Ensure that at least 250 ng of the Hi-C library is available for NGS Library preparation.

3.3 NGS Library Preparation

1. Remove glycogen and excess salt from the Hi-C library by mixing 45.0 µL of Agencourt AMPure XP beads with 15.0 µL of the library in a PCR tube. Incubate the sample at room temperature for 5 min and follow the wash steps

according to the manufacturer's instructions. Elute the DNA in 37.0 μL of elution buffer (*see* **Note 7**).

2. Use 2.0 μL of the library to determine DNA concentration using the Qubit® dsDNA HS Assay Kit. Proceed with library preparation if more than 200 ng of the Hi-C library is available.
3. Add 5.0 μL of 10× KAPA Frag buffer and 10.0 μL of the KAPA Frag enzyme to the Hi-C library on ice.
4. Transfer the reaction to a thermocycler at 4 °C and run the following program:

Step	Temperature	Time
Pre-cooled block	4 °C	N/A
Fragmentation	37 °C	10 min
Hold	4 °C	∞

Maintain the temperature of the heated lid at 45 °C during fragmentation (*see* **Note 8**).

5. Mix 7.0 μL of the End Repair and A-Tailing Buffer and 3.0 μL of the End Repair and A-Tailing Enzyme Mix on ice (supplied with the KAPA HyperPlus kit).
6. Immediately transfer the mix (10.0 μL) to the fragmentation reaction at 4 °C and run the program below on a thermocycler:

Step	Temperature	Time
Pre-cooled block	4 °C	N/A
End repair and A-tailing	65 °C	30 min
Hold	20 °C	∞

Maintain the temperature of the heated lid at 85 °C during end repair and A-tailing (*see* **Note 8**).

7. Immediately add the following to 60.0 μL of end repaired and A-tailed DNA: 5.0 μL of 15 μM Illumina® paired-end sequencing adapter (barcoded), 5.0 μL of elution buffer + 0.1% Tween 20, 30.0 μL of ligation buffer (KAPA HyperPlus), and 10.0 μL of DNA Ligase (KAPA HyperPlus). To sequence several Hi-C libraries in a single Illumina® lane, ligate each library to a different sequencing adapter.
8. Incubate the reaction at 20 °C for 2 h.
9. Transfer 50.0 μL of Dynabeads® MyOne™ Streptavidin T1 into a PCR tube. Place the tube on a magnetic rack until the solution clears and carefully remove the supernatant with a pipette.

10. Resuspend the beads in 50.0 μL of 1× BWB by vortexing. Place the beads back on the magnetic rack until the solution clears and remove the supernatant. Perform this step twice.
11. Resuspend the beads in 50.0 μL of 2× BWB by vortexing.
12. Clean up the ligation reaction (*see* **step 8**) with 110.0 μL of Agencourt AMPure XP beads and elute the adapter-ligated Hi-C library in 50.0 μL of elution buffer.
13. Mix the library with 50.0 μL of Dynabeads® MyOne™ Streptavidin T1 in 2× BWB (*see* **step 11**). Incubate the sample for 45 min at 37 °C in a ThermoMixer at 500 rpm.
14. Place the sample on a magnetic rack to collect the beads and carefully discard the supernatant.
15. Resuspend the beads in 50.0 μL of elution buffer + 0.1% Tween 20 by vortexing. Collect the beads with the magnetic rack and discard the supernatant. Repeat this step twice.
16. Resuspend the beads in 20.0 μL of elution buffer + 0.1% Tween 20.
17. To the adapter-ligated Hi-C library on beads (**step 16**), add: 25.0 μL of 2× KAPA HiFi HotStart ReadyMix, and 5.0 μL of 10× Library Amplification Primer Mix as supplied with the KAPA HyperPlus kit. Run the following program in a thermocycler to amplify the library:

Step	Temperature	Time	Cycles
Initial denaturation	98 °C	45 s	1
Denaturation Annealing Extension	98 °C 60 °C 72 °C	15 s 30 s 30 s	6–8 (*see* **Note 9**)
Final extension	72 °C	1 min	1
Hold	4 °C	∞	1

18. Collect the beads with the magnetic rack and transfer the supernatant (referred to hereafter as the sequencing library) into a new PCR tube. Place the tube on ice until needed. Proceed with processing the sequencing library at **step 21**.
19. Resuspend the beads in 50.0 μL of elution buffer + 0.1% Tween 20 by vortexing and collect the beads using the magnetic rack. Remove the supernatant by pipetting. Repeat this step twice.
20. Resuspend the beads in 20.0 μL of elution buffer + 0.1% Tween 20 and store the sample at 4 °C. This adapter-ligated Hi-C library on beads can be reused on a later date to prepare the sequencing library as outlined in **steps 17** to **24** of this section.

21. Purify the sequencing library (*see* **step 18**) with 35.0 μL of Agencourt AMPure XP beads (*see* **Note 7**).
22. Elute the sequencing library in 11.0 μL of elution buffer.
23. Use 1.0 μL of the sequencing library to determine DNA concentration using the Qubit® dsDNA HS Assay Kit. Use the elution buffer to adjust library concentration to 5.0 ng/μL.
24. Measure the concentration of the sequencing library to ensure that the dilution was performed accurately.
25. Multiple libraries, each marked with distinct Illumina® paired-end sequencing adapters, can now be pooled for sequencing. The ratio at which the libraries are pooled determines the fraction of total reads available for each library.
26. Assess the quality of the sequencing library by determining the size distribution of the DNA fragments using a Tapestation (Agilent Technologies) or Bioanalyzer (Agilent Technologies). If small amounts of adapters or adapter dimers remain, perform a second 0.7× SPRI bead purification.

3.4 Data Analysis and Contact Maps Generation

After deep sequencing, reads are mapped using an iterative method, then assigned to each restriction fragment. Self-circularized, unligated fragments and PCR artifacts are discarded using filtering and valid junctions are assigned to bins by their midpoint coordinate to generate a first raw contact map. If the binning process can reduce the resolution of the map, it also increases signal-to-noise ratio. The final normalized contact map is obtained after matrix balancing based on the nonuniform distribution of the length of restriction fragments and the nucleotide composition of the genome. In our experience, 15 million valid junction reads are enough to produce a normalized 5 kb resolution map. Chapter 2 provides a more detailed description of the analysis workflow for processing raw data into contact maps. Several software packages and online tools are also available to support researchers in analyzing Hi-C data (*see* **Note 10**).

4 Notes

1. Buffers used for cell fixation should lack compounds that contain nucleophilic groups such as thiols, amines, and amides. The carbonyl moiety of formaldehyde reacts with these groups reducing the efficiency of cell fixation. It is therefore also necessary to wash the bacterial cells with 1× PBS prior to fixation (*see* **step 2**; Subheading 3.1) to remove components of the growth medium that react with formaldehyde.

2. Restriction enzymes define the resolution of the Hi-C library. The enzyme is selected based on the criterion that it produces a non-degenerate 5′ overhang upon cleavage of its restriction site. This is necessary as 3′ overhangs cannot be filled in, and degenerate sticky ends may interfere with biotin tagging. It is necessary that the enzyme functions optimally at 37 °C (or at an even lower temperature; for instance, the optimal temperature for CviAII activity is 25 °C) as restriction digestion at higher temperatures leads to premature reverse cross-linking of DNA-protein complexes. To ensure a good digestion efficiency during library preparation, it is important that the activity of the selected enzyme is not blocked by DNA methylation. The size distribution of the predicted restriction fragments generated by the enzyme should also be evaluated to ensure that the enzyme generates limited, if any, restriction fragments with sizes greater than the chosen Hi-C map resolution; long restriction fragments introduce biases in the contact matrix.
3. During the preparation of Hi-C libraries, restriction fragment overhangs are filled in with biotin-labeled nucleotides to tag ligation junctions generated during the subsequent proximity ligation step. Ensure that the fill-in of the 5′ overhang generated by restriction digestion allows the incorporation of the selected biotin-labelled nucleotide. For instance, do not use biotin-14-dATP to fill-in overhangs generated by HpaII (C^CGG).
4. Cells can be harvested for fixation at any stage of growth. However, it is necessary to maintain cell concentration during fixation. For this, adjust the volume of the harvested culture when fixation is performed on *E. coli* cells at a different OD_{600}
5. Each 1 mL aliquot contains approximately 10^9 fixed cells. This corresponds to ~10 μg of DNA.
6. Some chromatin fragments may have been filled in with biotin-labeled nucleotides, but may not have undergone ligation to give a chimeric product. Unligated "dangling ends" that are bound by streptavidin-coated magnetic beads and thereafter sequenced reduce the number of useful reads per sequencing run. Biotin removal from such ends is performed with T4 DNA Polymerase under conditions that favor the 3′ exonuclease activity of the enzyme in removing nucleotides past blunted ends. This condition is provided by maintaining the concentration of selected free nucleotides in the reaction mix below 100 μM [7, 8]. For instance, for the removal of biotin-14-dATP from the dangling ends of PsuI- or BglII-digested Hi-C libraries (RGA*TC), 100 μM of dGTP is added to the reaction mix, whereas dATP, dCTP, and dTTP are not (*see* **step 23** of Subheading 3.2). This way, nucleotides up to but

excluding G will be removed as a result of the activation of the 5′ polymerase function of the enzyme at that site. The 3′ exonuclease activity of T4 DNA Polymerase is also favored by increasing enzyme concentration, extending reaction times, and incubation at higher temperatures, for instance, 16 °C—as opposed to 12 °C—to favor the breathing of DNA ends.

7. Solid Phase Reverse Immobilization (SPRI) beads reversibly bind DNA fragments in solution in the presence of Polyethylene glycol (PEG). The immobilization is size selective and depends on the concentration of PEG in the DNA-bead solution. Since PEG is present in the SPRI beads storage buffer (as supplied by the manufacturer), the size selective immobilization of DNA onto the SPRI beads is determined by the volumetric ratio rather than the concentration ratio of beads to DNA. Larger DNA fragments bind to the beads at low PEG concentrations. Thus, a 0.7× volumetric ratio of beads to DNA solution is used to remove the short-length adapters and adapter dimers from the sequencing library during NGS library preparation (*see* **step 21**; Subheading 3.3). Shorter DNA fragments can only be immobilized at higher PEG concentrations. Therefore, to remove glycogen and excess salts from the HiC library for NGS library preparation, a 3.0× volumetric ratio of beads to DNA solution is used (*see* **step 1**; Subheading 3.3).
8. Using a heated lid at a temperature > 50 °C during library fragmentation (*see* **step 4**; Subheading 3.3) or > 85 °C during end repair and A-tailing (*see* **step 6**; Subheading 3.3) overheats the reactions. This may result in a partial denaturation of the enzymes and a lower efficiency of the reactions.
9. These are the recommended number of cycles for 500–200 ng of input DNA as measured in **step 2** of Subheading 3.3.
10. Software packages and online tools available to analyze and display Hi-C interaction data:
    - Juicer: https://github.com/theaidenlab/juicer
    - HiCUP: http://www.bioinformatics.babraham.ac.uk/projects/hicup/
    - Hi-C inspector: https://github.com/HiC-inspector/HiC-inspector
    - Hi-C pro: https://github.com/nservant/HiC-Pro
    - HiTC: http://www.bioconductor.org/packages//2.10/bioc/html/HiTC.html

Acknowledgments

This work was supported by grants from the Netherlands Organization for Scientific Research [VICI 016.160.613] and the Human Frontier Science Program (HFSP) [RGP0014/2014].

References

1. Denker A, de Laat W (2016) The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev 30 (12):1357–1382. https://doi.org/10.1101/gad.281964.116
2. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485 (7398):376–380. https://doi.org/10.1038/nature11082
3. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021
4. Lieberman-Aiden E, Lieberman-Aiden E, van Berkum NL, van Berkum NL, Williams L, Williams L, Imakaev M, Imakaev M, Ragoczy T, Ragoczy T, Telling A, Telling A, Amit I, Amit I, Lajoie BR, Lajoie BR, Sabo PJ, Sabo PJ, Dorschner MO, Dorschner MO, Sandstrom R, Sandstrom R, Bernstein B, Bernstein B, Bender MA, Groudine M, Groudine M, Gnirke A, Gnirke A, Stamatoyannopoulos J, Stamatoyannopoulos J, Mirny LA, Mirny LA, Lander ES, Lander ES, Dekker J, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293. https://doi.org/10.1126/science.1181369
5. Dame RT, Tark-Dame M (2016) Bacterial chromatin: converging views at different scales. Curr Opin Cell Biol 40:60–65. https://doi.org/10.1016/j.ceb.2016.02.015
6. Le TBK, Imakaev MV, Mirny LA, Laub MT (2013) High-resolution mapping of the spatial organization of a bacterial chromosome. Science 342(6159):731–734. https://doi.org/10.1126/science.1242059
7. Marbouty M, Le Gall A, Cattoni DI, Cournac A, Koh A, Fiche J-B, Mozziconacci J, Murray H, Koszul R, Nollmann M (2015) Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by hi-C and super-resolution imaging. Mol Cell 59(4):588–602. https://doi.org/10.1016/j.molcel.2015.07.020
8. van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES (2010) Hi-C: a method to study the three-dimensional architecture of genomes. J Vis Exp (39). https://doi.org/10.3791/1869

Chapter 2

Processing and Analysis of Hi-C Data on Bacteria

Andreas Hofmann and Dieter W. Heermann

Abstract

The study of three-dimensional genome organization has recently gained much attention in the context of novel techniques for detecting genome-wide contacts using next-generation sequencing. These genome-wide chromosome conformation capture-based methods, such as Hi-C, give a deep topological insight into the architecture of the genome inside the cell. This chapter reviews the steps to process next-generation Hi-C sequencing data to generate a final contact probability map. We describe these steps using publicly available Hi-C datasets of different bacteria. We also present strategies to assess the quality of Hi-C datasets.

Key words Hi-C, Hi-C data analysis, Bacteria

1 Introduction

In humans, nearly two meters of DNA must be folded to fit inside the micrometer-sized cell nucleus. How is DNA compacted to this level and how can it remain accessible for gene transcription, replication, and repair at the same time? Novel technologies, such as "chromosome conformation capture" (3C)-based methods that map genome-wide spatial interactions along the genome, have, during the last 15 years, allowed to shed light on this question. Massive improvements in the throughput of such methods produce ever-increasing amounts of data. Most of the raw data are deposited in repositories that are publicly available.

Mammalian interphase chromosomes are hierarchically organized [1, 2]. Fluorescence microscopy and genome-wide 3C studies, such as Hi-C, have revealed inter-chromosomal compartmentalization in the form of distinct chromosome territories [3, 4]. Individual chromosomes also exhibit compartmentalization to form domains [5–8]. These 3C studies indicate that eukaryotic genomes are partitioned into discrete structural units with highly increased frequency of internal contacts.

Besides studies of eukaryotic chromosomes of humans, mice and *Drosophila melanogaster*, recently, the circular chromosomes of

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_2,

model bacterial species such as *Caulobacter crescentus*, *Bacilus subtilis*, and *Mycoplasma pneumoniae* have been shown to analogously be composed of domains with the help of Hi-C analyses [9–12]. Taken together, these results suggest that intra-chromosomal compartmentalization is a fundamental building block of chromosome structure of organisms.

3C was invented in 2002 by Dekker et al. [13]. It allows for focused quantification of contact frequencies between selected regions. All 3C-based techniques aim to generate a two-dimensional library of three-dimensional chromosome contacts. The first step in the procedure is cross-linking of chromatin by addition of formaldehyde. This causes interacting chromatin segments to be covalently linked together. The fixed chromatin is then digested with a restriction enzyme. The ends of the fragmented cross-linked DNA are thereafter re-ligated under diluted conditions to favor intramolecular ligation of the cross-linked fragments. The ligated DNA molecules form a hybrid of two DNA fragments consisting of the two segments that were cross-linked. Next-generation sequencing is used to quantify the number of such hybrid DNA-molecules.

Hi-C is a genome-wide 3C-based technology introduced by Lieberman-Aiden et al. in 2009 [4]. Realizing a genome-wide quantification of interactions, it constitutes a major breakthrough in the study of chromatin architecture. The Hi-C protocol differs from the standard 3C protocol in that there is an extra step needed before ligation. It consists of filling in the restriction digest of the chromosome with biotin-labeled nucleotides. After purification and shearing/fragmentation of the Hi-C library, the biotin labeled material is pulled down to ensure that only ligation junctions are selected for further analysis.

"Chromosome conformation capture carbon copy" (5C) captures interactions between all restriction fragments within a selected region [14]. For example, it was used to study the spatial organization of the bacterial *Caulobacter crescentus* genome [15], the regulatory landscape of mouse X inactivation [6], and the long-range interaction landscape of gene promoters in the human genome [16]. A technique similar to Hi-C, called genome conformation capture (GCC), has been applied for mapping yeast chromosome interactions [17] as well as for studying the spatial organization of the *Escherichia coli* nucleoid [18].

All 3C-based methods, contrary to microscopy-based techniques, allow for both a more systematic and quantitative characterization of genome topology and a higher resolution. The essential drawback, however, is that the conventionally ensemble 3C-based methods are mostly performed on large populations of cells, leading to loss of information at the single-cell level.

In this chapter, we give an overview of the data analysis involved in the framework of a Hi-C experiment. Moreover, we

present and discuss publicly available Hi-C datasets of bacterial genomes and present possibilities to assess and compare them in terms of data quality.

2 Hi-C Data Processing

This section covers the main steps involved in the data processing of a genome-wide 3C-based study. Since the focus of this chapter is mainly on genome-wide methods, such as Hi-C or GCC, methods relevant for these technologies are discussed in this section. Hi-C data processing can be subdivided into the following four main steps: (1) Mapping to the reference genome, (2) Quality control, (3) Binning and contact matrix generation, (4) Balancing (*see* Fig. 1). Each of these steps is discussed in the following section.

2.1 Mapping to the Reference Genome

The first step of genome-wide 3C data analysis consists of mapping reads back to the reference genome. The Hi-C method quantifies an interaction by a ligation product formed between two restriction fragments. By using paired-end sequencing and mapping both ends of each paired sequence to the reference genome, the two restriction fragments in the ligation product can be determined. However, if the read length is larger than the length of one of the restriction fragments, the mapping will not work. To solve this problem, the mapping procedure can be refined by means of an iterative mapping scheme that involves truncating reads to a smaller

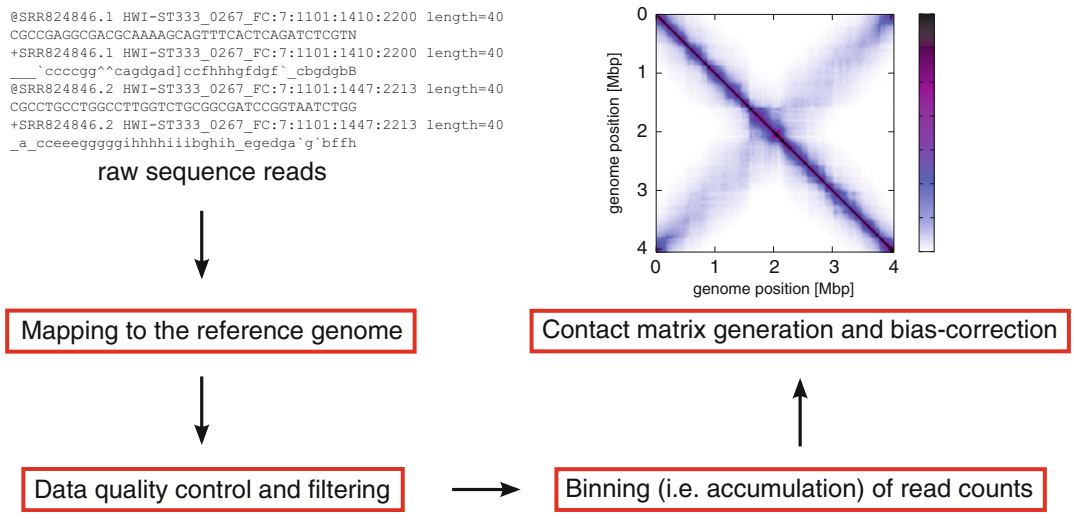


Fig. 1 Schematic overview of the general workflow for analyzing genome-wide 3C data such as Hi-C. The input (at the top left) is raw sequencing data and as a typical output we illustrated a bias-corrected contact matrix (at the top right). The dataset used here for illustration purposes is that of the study of Le et al. [9] (SRR824846)

length prior to mapping [19]. Reads that are not aligned uniquely at both ends are then re-aligned by iteratively increasing their portions. This process is repeated until either all reads uniquely map or until the read is extended to its entirety. Only paired-end reads with both sides being uniquely mapped to the reference genome contribute to the set of Hi-C interactions. All other paired-end reads are discarded.

2.2 Quality Control

The next step is quality control to ensure that the aligned sequence reads are likely to be the result of proximity-based ligation of digested fragments, and that they are reflecting long-range chromatin interaction rather than just random collision. Self-circularized or un-ligated (dangling-end) products will result in reads that map with both ends on the same restriction fragment. These reads should be removed. Also reads from neighboring fragments that map to the same strand should be removed since they are likely the result of incomplete digestion. Furthermore, reads that map multiple times at the exact same location on the reference genome are often the result of biased PCR-amplification and should also be removed. Hi-C sequencing reads can be compared to randomly generated control sequencing reads, whereby the Hi-C reads should be significantly closer to the chosen restriction sites than random reads [20]. The Hi-C reads should also be in the correct orientation with respect to the restriction site.

2.3 Binning and Contact Matrix Generation

After the alignment of the sequence reads and quality control the next step is the construction of contact matrices of the interaction data. To produce a contact matrix, the genome is divided into equally sized loci, so called bins. The result of this aggregation of read-counts across bins is a symmetric matrix composed of interaction frequencies between bins covering the entire genome. The size of these bins used to represent the meaningful contacts between pairs of genomic loci can be referred to as the resolution of a Hi-C experiment (*see* Fig. 2 for a comparison of a contact matrix depicted at different bin sizes).

A linear increase of resolution requires a quadratic increase in total sequencing depth; the size of the bins and effectively the resolution is limited by sequencing depth. Moreover, the resolution also depends on the restriction enzyme used for the Hi-C experiment (*see* Subheading 3.4). To illustrate this relationship between resolution and the total number of collected read pairs, we want to highlight that for a resolution of 100 kbp in the first Hi-C study of the human genome 8.4 million reads were collected [4]. Increasing the resolution to 1 kbp required 4.9 billion reads [8]. Thus, increasing the resolution by two orders of magnitude required increasing the total number of reads by three orders of magnitudes.

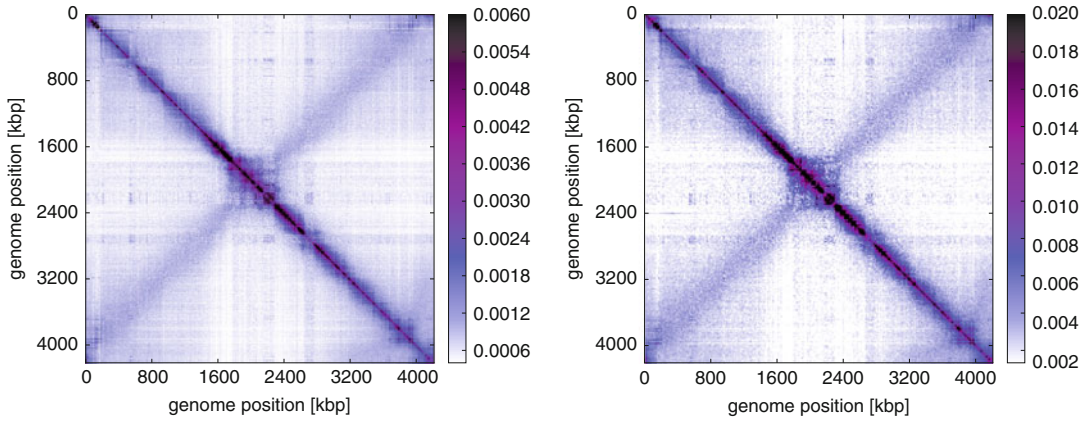


Fig. 2 Hi-C contact map depicted at two different resolutions (left: 4kbp and right: 16kbp). The contact map of higher resolution reveals much more detailed structures, especially on the interactions within the domains along the diagonal. Data on *B. subtilis* from Marbouty [10] (SRR2772323)

2.4 Balancing

Besides the bias introduced by individual reads or restriction fragments, binning generates biases, as well. Yaffe and Tanay identified the origin of some of these biases, such as the nonuniform distribution of the length of restriction fragments with respect to ligation efficiency, the nucleotide composition of the genome under investigation and issues with uniquely mapping the interactions back onto the reference genome. They proposed an integrated probabilistic model [20] for eliminating these known systematic biases from the "raw" contact maps. This procedure is referred to as normalization.

Several other models for normalizing Hi-C contact maps have been proposed [19, 21, 22]. However, these approaches do not explicitly incorporate the aforementioned biases on the grounds that it is not possible to know each and every bias. Since most of these approaches are based on the Sinkhorn-Knopp (SK) balancing algorithm [23], they can be more precisely referred to as balancing instead of normalization. Explicit bias correction and balancing yield comparable results [8].

In this section, we focus on the SK balancing algorithm [23] that transforms a symmetric non-negative matrix $\mathbf{A} = (a_{ij})$, $\mathbf{A} \in \mathbb{R}_{+}^{n \times n}$, into a doubly stochastic matrix $\mathbf{S} = (s_{ij})$, that is, a matrix whose rows and columns sum up to 1, i.e., $\sum_i s_{ij} = \sum_j s_{ij} = 1$.

The SK algorithm is an iterative process that consists in solving

$$\mathbf{S} = \mathbf{D}_1 \mathbf{A}\, \mathbf{D}_2, \tag{1}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are unique up to a scalar factor diagonal matrices with positive main diagonal. The matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are obtained by alternatingly normalizing columns and rows of.

Applied to our situation, **A** constitutes the raw matrix of contact frequencies, and the diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ contain the biases for the bins involved in the contacts between bins i and j. **S** is then the matrix of unbiased relative contact frequencies, which is defined such that each row and column of the upper triangular matrix sums to 1. The biases, and therefore the matrix **S**, can be found by using the SK algorithm that converges to the solution of Eq. (1).

Regardless of the method used, it is important to compare the raw and the bias-corrected contact map to check the effect of the procedure. Figure 3 depicts two such comparisons for two different

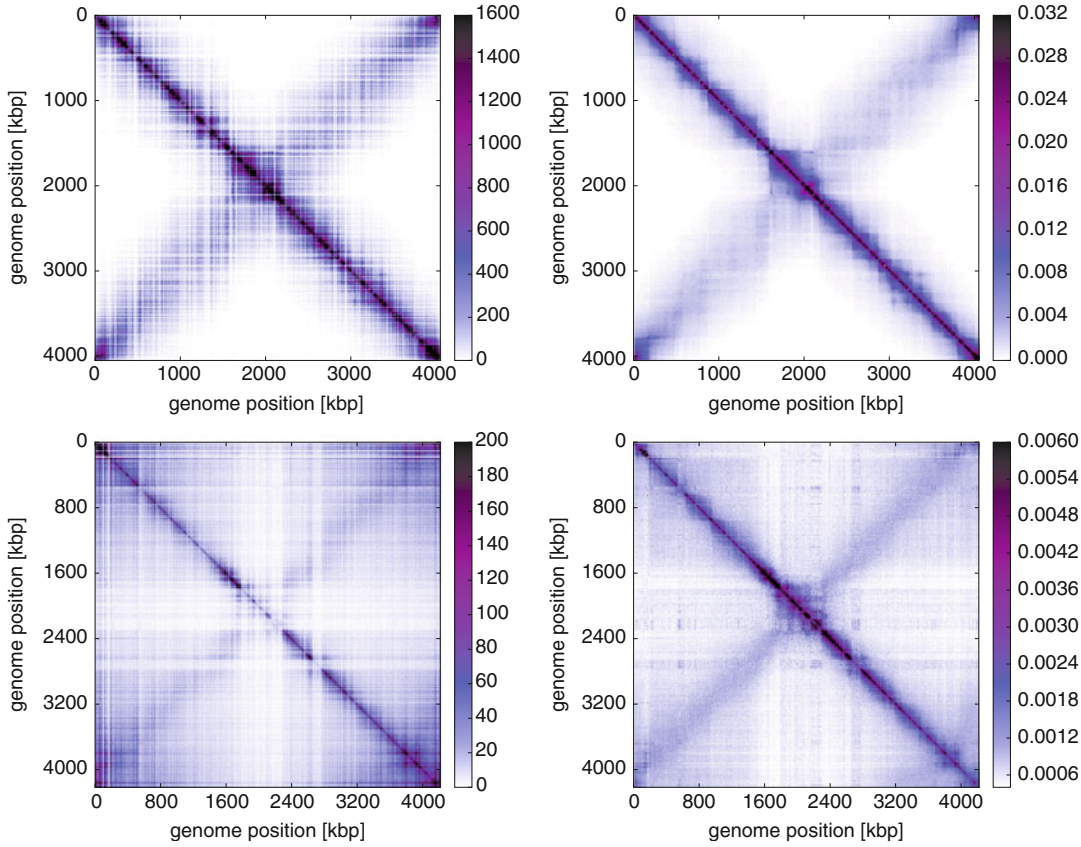


Fig. 3 Comparison of "raw" Hi-C contact frequency matrices on the left and the respective balanced equivalents, i.e., contact probability matrices, on the right. Data on *C. crescentus* of the study of Le et al. [9] (SRR824846) in the upper row and on *B. subtilis* of the study of Marbouty et al. [10] (SRR2772323) in the bottom row. We balanced the matrices by applying the Sinkhorn-Knopp (SK) algorithm. These two examples illustrate that the balancing procedure leads to very different results depending on the input "raw" Hi-C contact frequency matrix. In the case of the upper matrix on *C. crescentus*, the coverage of reads is homogenous along the genome and the contact frequency matrix already indicates a domain structure along the diagonal. In contrast, the coverage of reads is considerably more heterogeneous in the case of the lower matrix on *B. subtilis*. The region from 1.7 up to 2.3 Mbp lacks reads. Therefore, the structure of the balanced contact map in this region is rather an artifact of the balancing procedure than of biological relevance

Hi-C datasets. In both cases, balancing has the effect of smoothing the contact map such that no obvious high or low rows and columns remain. In extreme cases (compare the two lower maps in Fig. 3), large regions with a strikingly low number of captured reads are completely reorganized by balancing. The resulting contact map should be regarded with caution because it is likely to contain artifacts caused by balancing, especially in and near the aforementioned regions.

2.5 Concluding Remarks

It is important to be aware of the limitations of the experimental elements that impact the computational data processing. In fact, 3C experiments are never truly genome-wide since the usage of a restriction enzyme is biased to where its sites are located in the genome. Furthermore, the restriction enzyme chosen has to balance, among others, the frequency of cutting, fragment size, and size uniformity across the studied genome.

Because of the necessity of balancing contact maps constructed from 3C data, it is crucial to differentiate between raw contact maps representing all the captured contacts and their raw frequencies and the balanced contact maps representing contact probabilities. Regardless, balanced contact maps should be compared with their raw equivalent in order to check the effects of the balancing procedure.

3 Hi-C Data Assessment

In this section, we focus on Hi-C and other genome-wide 3C experiments performed in bacteria. We also present a procedure to assess data quality for comparison of datasets. Finally, we highlight the importance of the choice of restriction enzymes for Hi-C experiments in terms of data analysis.

3.1 Hi-C Data Availability

The first Hi-C study was carried out by Lieberman-Aiden et al. [4] and addressed the folding of the human genome. In 2011, Umbarger et al. [15] described the first 3C-based study, here 5C, of a bacterial genome. To date there have been several studies of bacterial genomes using genome-wide 3C-based experiments. We have summarized them in Table 1 in a chronological order. There are several datasets for each study that correspond to Hi-C libraries that differ, for example, in bacterial growth conditions, the restriction enzyme used or the sequencing method. We listed studies with publicly available Hi-C datasets.

3.2 How Many Reads?

Sequencing depth is decisive for the resolution of a Hi-C experiment. More precisely, it is the number of valid reads, i.e., those that remain after the filtering step discussed in Subheading 3.4. Therefore, given a Hi-C dataset the relevant question to ask is: How

Table 1
Overview of genome-wide 3C-based (xC) datasets of bacterial species for different conditions

Bacterial Species	xC	Res_{max}	RE	Datasets	Reference
Caulobacter crescentus	5C	13 kbp	BglII	3	Umbarger et al. [15]
Escherichia coli	GCC	20 kbp	HhaI	8	Cagliero et al. [18]
Caulobacter crescentus	Hi-C	10 kbp	BglII, NcoI	23	Le et al. [9]
Bacillus subtilis	Hi-C	4 kbp	HindIII, HpaII	15	Marbouty et al. [10]
Bacillus subtilis	Hi-C	10 kbp	HindIII	41	Wang et al. [11]
Mycoplasma pneumoniae	Hi-C	3 kbp	HindIII, HpaII	8	Trussart et al. [12]

We list the restriction enzymes (RE) that have been used in the datasets as well as the highest achieved resolution (Res_{max}) and the number of datasets that have been published

many "valid" reads remain after the filtering process? Although in the end the absolute number of valid reads is decisive from the data processing point of view, the relative fraction of the number of valid and total reads is also very important from an experimental point of view since it determines something like the efficiency of a Hi-C experiment. There are large variations in the number of valid reads in the datasets under discussion. While it exceeds 50% for the Hi-C data on *C. crescentus* (Le et al. [9]), it amounts to less than 5% for that on *M. pneumoniae* (Trussart et al. [12]). The latter means that, in the extreme case, albeit 187 million total reads have been sequenced, only 3 million valid reads contribute to the contact map (ERR1413594). In most cases, the key limiting factor is ligation efficiency, i.e., unligated fragments (dangling-ends), self-ligated fragments (self-circles), and poor biotin label removal.

3.3 What Kind of Reads?

The result of a Hi-C experiment is a contact probability matrix of the binned genomic interactions. It is visualized as a heat map of relative interaction frequencies (a "contact map") encoding the interaction data using a color map. This graphical representation gives the observer an immediate impression of the Hi-C data since it highlights prominent features, such as domains along the diagonal or the presence of a secondary diagonal. The graphical representation is highly dependent on the choice of the color map and on its scaling that can either be linear or logarithmic. An example of how different the same contact probability matrix can be visualized is illustrated in Fig. 4. The heat map with the rainbow color map and the appropriate lower limit on the color bar on the right highlights the hardly varying contact probabilities much better than the map on the right.

For this reason, it is important to evaluate a Hi-C dataset objectively in addition to heat map visualization. One possibility is to regard the number of the captured reads as a function of the

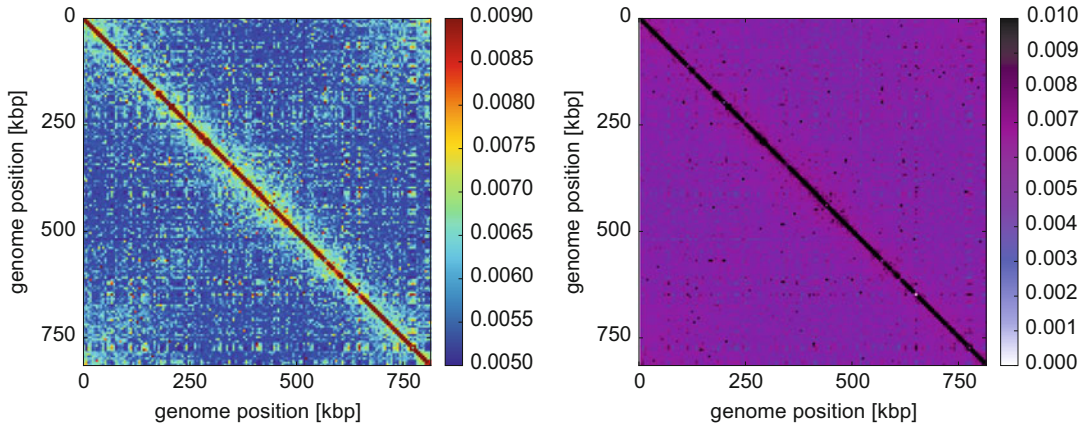


Fig. 4 Hi-C contact map depicted for two different color maps and different lower limits of the color bar. Since the Hi-C data contains a high level of noise a lower limit on the color bar needs to be set. Moreover, a rainbow color map highlights the hardly varying contact probabilities much better than the colormap in the heat map on the right. Data on *M. pneumoniae* from Trussart [12] (ERR1413595)

genomic distance between the interacting loci (*see* Fig. 5). In all Hi-C experiments a lot of reads are detected between neighboring genomic sites. This is, of course, due to the fact that these sites are intrinsically linked to each other. Thus, these reads do not contain relevant information. The read distribution shows how large this fraction of trivial reads is compared to more interesting short-, intermediate-, and long-range interactions related to the distinct three-dimensional shape of the regarded genomes. It can therefore be used as an indicator for the quality of the reads of a given Hi-C dataset.

The read distribution also shows that the captured interactions decrease with increasing genomic distance, such that short-range interactions will typically have higher coverage and thus higher effective resolution. In the contact map this is reflected by a gradual decrease of the average interaction probability the further one moves away from the diagonal. This finding follows the intuition that topologically close loci interact frequently assuming random motion in 3D space. Also, polymer models predict a power-law decrease of the contact probability as a function of the genomic distance [24].

3.4 Resolution of Hi-C Data and the Selection of Restriction Enzymes

Restriction enzymes are proteins that cut the DNA at specific sites. There is a wide range of restriction enzymes and in addition to the biochemical criteria referred to in Chapter 1 the choice of a certain restriction enzyme is also relevant for data analysis since it inevitably determines the resolution of a Hi-C map. This is due to the fact that in a Hi-C experiment interactions between genomic loci are measured in terms of restriction fragments. Thus, choosing an

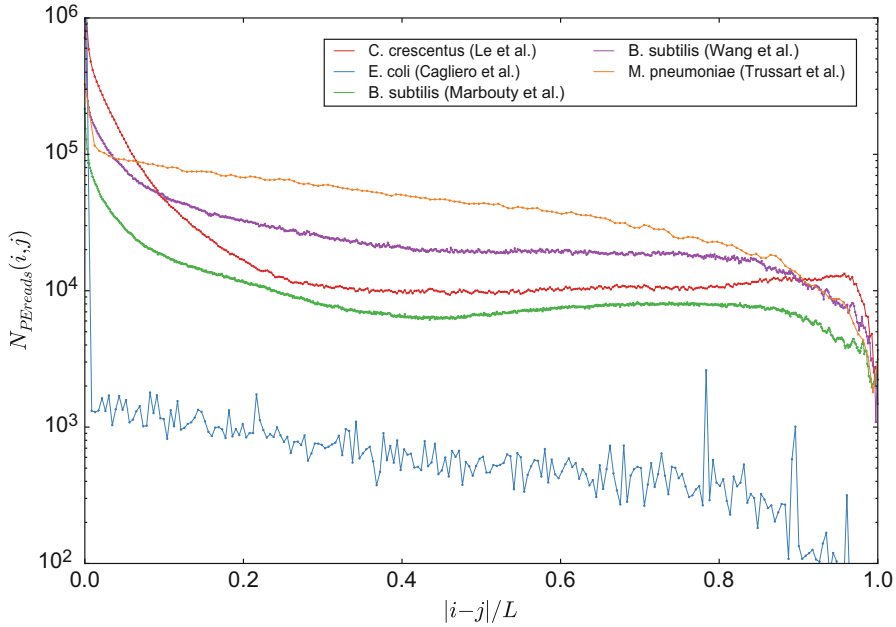


Fig. 5 Number of valid counts between different genomic loci *i* and *j* as a function of their genomic distance (which has been normalized by the length *L* of the respective genome), depicted for different bacteria. Most of the counts represent interactions between neighboring or near genomic loci and with increasing genomic length between the interacting loci the counts decrease exponentially. This exponential decay is different for the various datasets. This is a consequence of the distinct three-dimensional shape of the different genomes that is reflected in distinct interaction patterns. Moreover, also experimental factors have an impact on the interaction profile as can be seen from the interaction profiles of the two *B. subtilis* Hi-C datasets. The two curves have approximately the same slope but one is shifted upward (Wang et al.) compared to the other one (Marbouty et al.) because of the larger number of valid counts stemming from deeper sequencing of the corresponding Hi-C library. Furthermore, the read distribution also reveals what kind of reads have been captured in the Hi-C experiment, such as in the case of the GCC dataset of the study by Cagliero et al. where a sharp decline of contacts indicates that it is highly biased toward contacts of neighboring restriction fragments (99.5% of the valid read counts occur between neighboring bins)

appropriate bin size for a given Hi-C dataset very much depends on the distribution of the lengths of the restriction fragments.

In Fig. 6 we have contrasted the distribution of the length of a high-frequency restriction enzyme (frequent cutter) with a low-frequency cutter (rare cutter) for the *E. coli* genome. Clearly, we could choose the frequent cutter HhaI in order to obtain an *E. coli* contact map of 1 kbp or even 500 bp resolution, whereas the rare cutter BglII would limit the maximum possible resolution to around 10 kbp. It is important to remark that the choice of a frequent cutter predetermines the need of a high sequencing depth. A summary of various restriction enzymes and the statistics of the length of the respective restriction fragments for the *E. coli* genome can be found in Table 2.

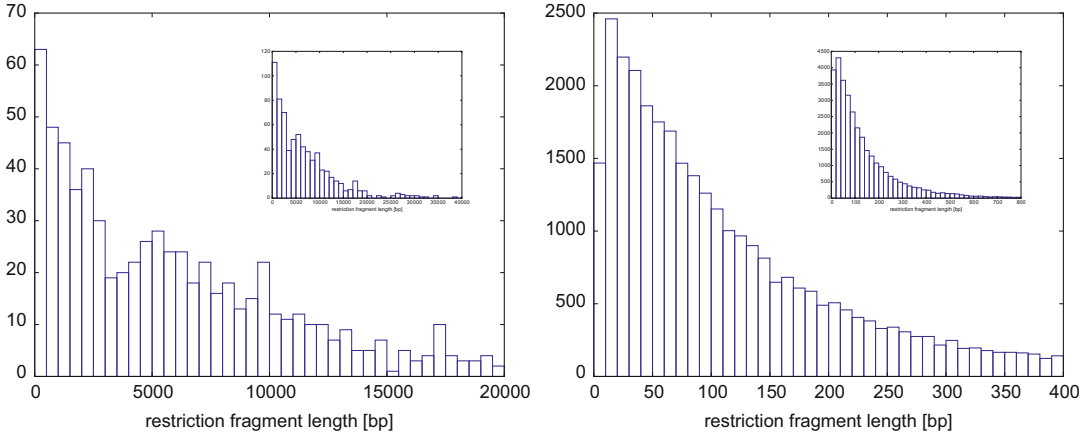


Fig. 6 The distribution of the length of the restriction fragments for two different restriction enzymes (A: BglII | B: HhaI) using the example of the *E. coli* genome. The main graphs show the interesting part of the distribution and the inset graphs show the overall distribution. The two distributions show the huge difference between the restriction fragments generated by the frequent cutter HhaI and the rare cutter BglII as regards the total number of generated fragments as well as their typical lengths

Table 2
Overview of the statistics of the length of the restriction fragments for different restriction enzymes using the *E. coli* genome as an example

Restriction enzyme	Recognition site	No. of RFs	L	$\tilde{L}$	L_{min}	L_{max}
HindIII	A^AGCTT	557	8333	4963	9	62,667
NcoI	C^CATGG	614	7560	5163	21	47,514
BglII	A^GATCT	702	6612	5048	16	42,359
MluI	A^CGCGT	1329	3493	2444	6	33,370
BstYI	R^GATCY	3191	1455	958	6	18,605
HpaII	C^CGG	24,312	191	122	4	3376
HhaI	GCG^C	32,795	142	90	4	4100

Listed are all restriction enzymes that have been used in the discussed Hi-C experiments. Besides its recognition site we listed the number of restriction fragments (RFs) generated by the respective restriction enzyme for the *E. coli* genome (MG1655, NC_000913.3) as reference. We also included the mean, median, smallest and biggest restriction fragment length (L)

4 Summary

Genome-wide chromosome conformation capture-based methods are now widely used. The resulting data, most of which is deposited in publicly available repositories, is the starting point for the computational modeling of the three-dimensional architecture of a multiplicity of genomes. Besides the results of the modeling

approaches, it is also interesting to know how reliable the underlying Hi-C data is. It is not only the number of captured reads, but also the distribution of these reads which is decisive for this question. Moreover, contact maps before and after balancing can differ significantly. It is therefore necessary to compare them and understand possible deviations.

Acknowledgment

We would like to thank Remus T. Dame, Frédéric Crémazy and Fatema Zahra Rashid for the stimulating and fruitful discussions. This work was supported by a grant from the International Human Frontier Science Program Organization (RGP0014/2014).

References

1. Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. Mol Cell 49(5):773–782. https://doi.org/10.1016/j.molcel.2013.02.011
2. Bickmore WA, van Steensel B (2013) Genome architecture: domain organization of interphase chromosomes. Cell 152(6):1270–1284. https://doi.org/10.1016/j.cell.2013.02.001
3. Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2 (4):292–301. https://doi.org/10.1038/35066075
4. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326 (5950):289–293. https://doi.org/10.1126/science.1181369
5. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398):376–380. https://doi.org/10.1038/nature11082
6. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation Centre. Nature 485 (7398):381–385. https://doi.org/10.1038/nature11049
7. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148 (3):458–472. https://doi.org/10.1016/j.cell.2012.01.010
8. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at Kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021
9. Le TBK, Imakaev MV, Mirny LA, Laub MT (2013) High-resolution mapping of the spatial organization of a bacterial chromosome. Science 342(6159):731–734. https://doi.org/10.1126/science.1242059
10. Marbouty M, Le Gall A, Cattoni DI, Cournac A, Koh A, Fiche J-B, Mozziconacci J, Murray H, Koszul R, Nollmann M (2015) Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by Hi-C and super-resolution imaging. Mol Cell 59 (4):588–602. https://doi.org/10.1016/j.molcel.2015.07.020
11. Wang X, Le TBK, Lajoie BR, Dekker J, Laub MT, Rudner DZ (2015) Condensin promotes the juxtaposition of DNA flanking its loading site in Bacillus subtilis. Genes Dev 29 (15):1661–1675. https://doi.org/10.1101/gad.265876.115

12. Trussart M, Yus E, Martinez S, Baù D, Tahara YO, Pengo T, Widjaja M, Kretschmer S, Swoger J, Djordjevic S, Turnbull L, Whitchurch C, Miyata M, Marti-Renom MA, Lluch-Senar M, Serrano L (2017) Defined chromosome structure in the genome-reduced bacterium mycoplasma pneumoniae. Nat Commun 8:14665. https://doi.org/10.1038/ncomms14665
13. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. https://doi.org/10.1126/science.1067799
14. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16(10):1299–1309. https://doi.org/10.1101/gr.5571506
15. Umbarger MA, Toro E, Wright MA, Porreca GJ, Baù D, Hong S-H, Fero MJ, Zhu LJ, Marti-Renom MA, McAdams HH, Shapiro L, Dekker J, Church GM (2011) The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. Mol Cell 44(2):252–264. https://doi.org/10.1016/j.molcel.2011.09.010
16. Sanyal A, Lajoie BR, Jain G, Dekker J (2012) The long-range interaction landscape of gene promoters. Nature 489(7414):109–113. https://doi.org/10.1038/nature11279
17. Rodley CDM, Bertels F, Jones B, O'Sullivan JM (2009) Global identification of yeast chromosome interactions using genome conformation capture. Fungal Genet Biol 46(11):879–886. https://doi.org/10.1016/j.fgb.2009.07.006
18. Cagliero C, Grand RS, Jones MB, Jin DJ, O'Sullivan JM (2013) Genome conformation capture reveals that the Escherichia coli chromosome is organized by replication and transcription. Nucleic Acids Res 41 (12):6058–6071. https://doi.org/10.1093/nar/gkt325
19. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods 9(10):999–1003. https://doi.org/10.1038/nmeth.2148
20. Yaffe E, Tanay A (2011) Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet 43(11):1059–1065. https://doi.org/10.1038/ng.947
21. Cournac A, Marie-Nelly H, Marbouty M, Koszul R, Mozziconacci J (2012) Normalization of a chromosomal contact map. BMC Genomics 13:436. https://doi.org/10.1186/1471-2164-13-436
22. Li W, Gong K, Li Q, Alber F, Zhou XJ (2015) Hi-Corrector: a fast, scalable and memory-efficient package for normalizing large-scale Hi-C data. Bioinformatics 31(6):960–962. https://doi.org/10.1093/bioinformatics/btu747
23. Sinkhorn R, Knopp P (1967) Concerning nonnegative matrices and doubly stochastic matrices. Pac J Math 21(2):343–348
24. Bohn M, Heermann DW (2010) Diffusion-driven looping provides a consistent framework for chromatin organization. PLoS One 5(8): e12218. https://doi.org/10.1371/journal.pone.0012218

Chapter 3

GeF-seq: A Simple Procedure for Base Pair Resolution ChIP-seq

Onuma Chumsakul, Kensuke Nakamura, Shu Ishikawa, and Taku Oshima

Abstract

Nucleotide sequences recognized and bound by DNA-binding proteins (DBPs) are critical to control and maintain gene expression, replication, chromosome segregation, cell division, and nucleoid structure in bacterial cells. Therefore, determination of the binding sequences of DBPs is important not only to study DBP recognition mechanisms, but also to understand the fundamentals of cell homeostasis. While ChIP-seq analysis appears to be an effective way to determine DBP-binding sites on the genome, the resolution is sometimes not sufficient to identify the sites precisely. Here, we introduce a simple and effective method named *Ge*nome *f*ootprinting with high-throughput *seq*uencing (GeF-seq) to determine binding sites of DBPs at single base-pair resolution. GeF-seq detects binding sites of DBPs as sharp peaks and thus makes it possible to identify the recognition sequence in each "binding peak" more easily and accurately than using ChIP-seq.

Key words GeF-seq, High resolution ChIP-seq, DNaseI footprinting, Genome footprinting, DNA-binding proteins, Consensus sequence

1 Introduction

Chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) analysis is an effective method to identify the binding sites of DNA-binding proteins (DBPs) on genomic DNA [1, 2]. This method has been adopted for the analysis of nucleoid proteins in bacterial cells, and to identify their binding sites along the bacterial genome [3–7]. In this method, DBPs are initially crosslinked to the genomic DNA through treatment of living cells with a chemical crosslinker (e.g., formaldehyde). To purify the DBPs crosslinked to DNA, the genomic DNA is first fragmented, in general, by the use of sonication. Then, the protein-DNA complexes are purified by chromatin immunoprecipitation (ChIP) using a specific antibody to the target protein, or alternatively by affinity precipitation (ChAP) using an affinity resin for a peptide tag fused to the target protein (e.g., nickel resin for a histidine tag). After decrosslinking of the protein-DNA complexes,

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_3,

the DNA fragments are purified. The purified DNA fragments are then sequenced using next-generation sequencing (NGS) and the sequence reads are mapped to the genome to identify binding sites of the DBPs [1, 2].

Due to the methodological limitations of ChIP-seq, the protein-binding peak is detected as being much broader than the actual binding site. In the general ChIP-seq procedure, genomic DNA is physically fragmented by sonication. Assuming that the genomic DNA is randomly fragmented, the sequence reads, which are mapped around the binding site of the DBP, form two separate peaks on the plus and minus strands of the DNA fragments, which represent distribution of the 5′ and 3′ ends of the DNA fragments (Fig. 1A). Therefore, theoretically, the actual binding site of the DBP is located centrally between the two peaks (Fig. 1A). Many programs to predict the real binding sites of DBPs from ChIP-seq data have been developed based on this theory [8]. Occasionally the prediction is confounded, for instance when two binding sites are closely located, and the forward and reverse peaks of the two binding sites are hardly separated, the peaks are merged (Fig. 1B). Therefore, in such cases, incorrect locations of the binding sites are predicted.

From ChIP-seq data, the DNA recognition sequence of the DBP is also hard to identify at each binding site. Identification of the recognition sequence is based on the simple idea that a DBP recognizes and binds to conserved DNA sequences [9, 10]. Therefore, to identify recognition sequences, we look for homologous DNA sequences in the DBP-binding regions, and if we can find a consensus sequence in a significant number of DBP-binding regions, it is considered a recognition sequence for the cognate DBP. However, because the binding sites are predicted by the method described above using peaks wider than the real binding region, and sometimes overlapping peaks are acquired by ChIP-seq (Fig. 1A, B), the predictions of the binding site sequences are not always accurate, and thus no conserved sequence is found at many binding sites. In contrast, in vitro DNA footprinting is able to precisely identify the regions bound by DBPs by the use of the DBP to protect its cognate-binding site from DNase I digestion [11]. DNA footprinting could therefore be used to directly determine DBP recognition sequences at a genome-wide scale, if it could be conducted.

To accurately and comprehensively identify the binding sequences of DBPs on genomic DNA, we developed a technique combining DNase I footprinting and ChIP-seq [12, 13]: GeF-seq (*Ge*nome *f*ootprinting with high-throughput *seq*uencing). This method is very simple and straightforward, but detects the binding sites of DBPs as very sharp and narrow peaks, with a resolution comparable to that of in vitro DNase I footprinting (Fig. 2) and can thus be used to identify the DBP recognition sequence far more accurately and comprehensively than generic ChIP-seq.

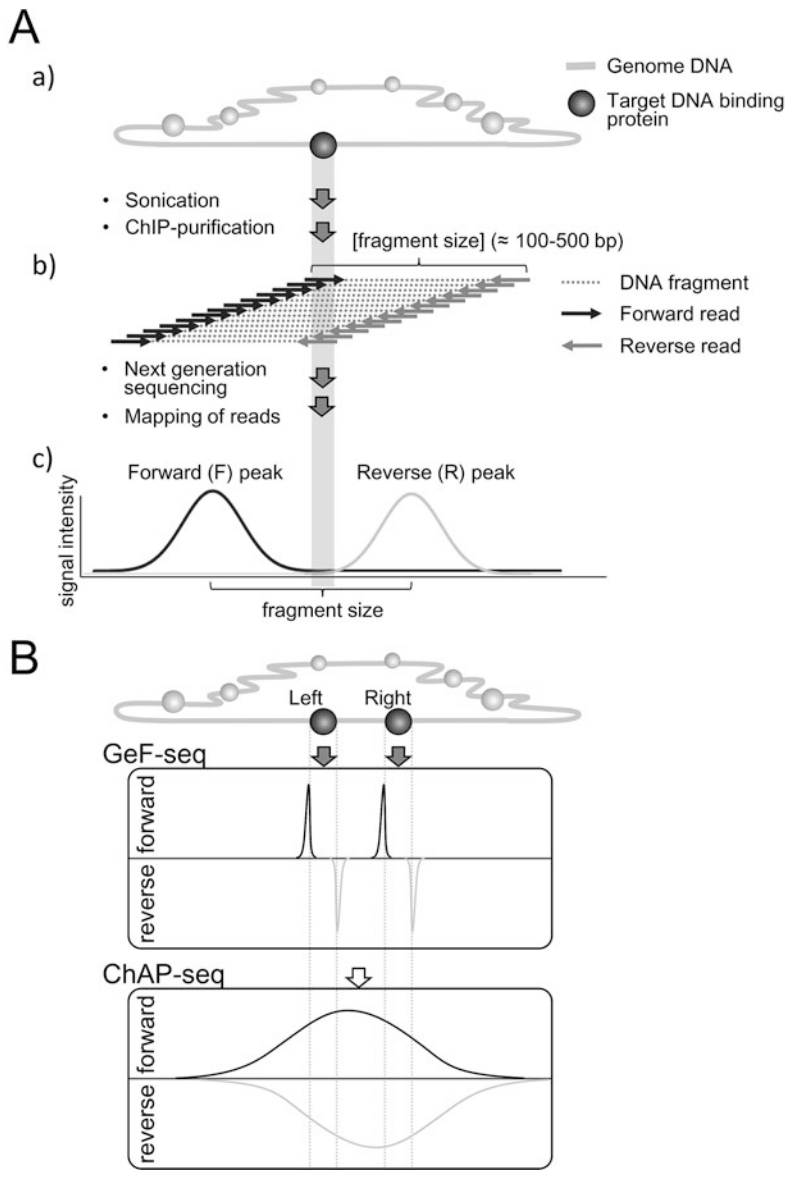


Fig. 1 Identification scheme of binding sites of DBP from ChIP-seq mapping data A: (**a**) The DBP is crosslinked to genomic DNA. The genomic DNA is fragmented by sonication and purified with chromatin immunoprecipitation as protein-DNA complexes. (**b**) Both ends of the fragments (black and gray arrows) are sequenced by high-throughput sequence. (**c**) Schematic diagram of the mapping result of sequencing. Sequence reads of the plus (black arrows) and the minus (gray arrows) strands of the DNA fragments mapped with mapping programs developed for ChIP-seq yield peaks separated with the distance as long as the average size of the DNA fragments purified by ChIP-seq procedures. The middle of two peaks is the location of the real binding site of the DBP (indicated by a gray thick vertical line). B: schematic representation of false binding sites when two real binding sites are closely colocated. Since peaks of the plus and the minus strands form wider peaks, these peaks overlap to form two merged peaks. In this case, at the middle of the two merged peaks, the false binding site (gray arrow) is identified. The real binding sites (white arrows) should be located to the left or right of the false binding site, which would be predicted by the mapping results, when real plus and minus peaks (black and gray dotted peaks) are mapped by GeF-seq. (reproduced from ref. 21 with permission from YODOSHA)

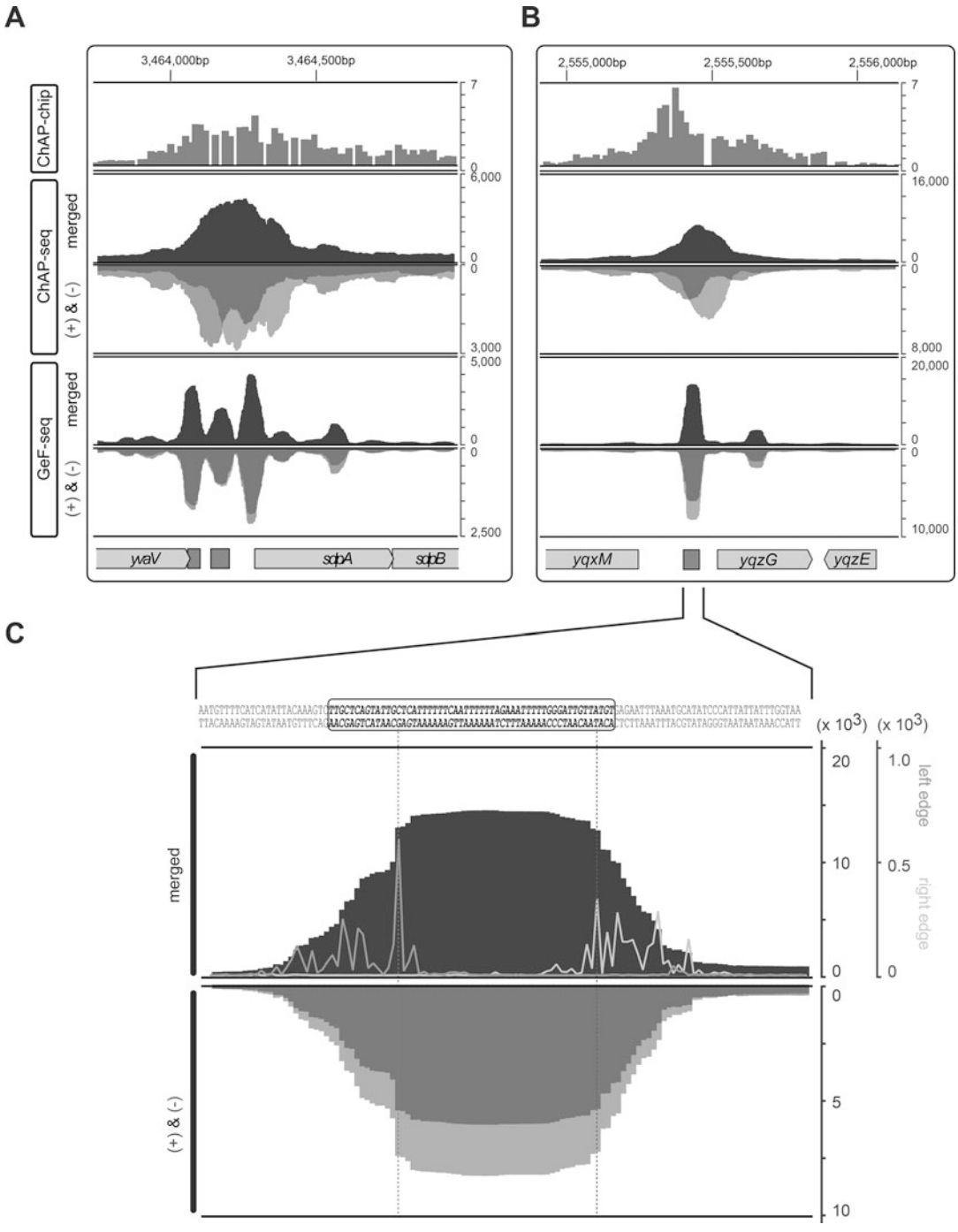


Fig. 2 The higher resolution of GeF-seq compared to ChAP-chip and ChAP-seq. (**a**) ChAP-chip, ChAP-seq and GeF-seq results of AbrB (a possible nucleoid protein and grobal transcriptional regulator in *B. subtilis* [22]) binding sites shown for the *sdpA* locus. The lower panel of the ChAP-seq and GeF-seq results indicates plus (dark gray) and minus (light gray) strand reads separately, whereas the upper panel shows merged images of the plus and minus binding peaks. As previously shown in Fig. 1, the plus and minus strand peaks are gapped in ChAP-seq, whereas they are consistent in GeF-seq. (**b**) This panel represents ChAP-chip, ChAP-seq and

2 Materials

We use bacterial strains expressing a DBP with a 12 × Histidine-tag, under the control of the original promoter on the genome. To fuse the His-tag to the C-terminal end of the target DBP in *B. subtilis*, the forward sequence for recombination, the His-tag coding sequence, the antibiotic gene and the backward sequence for recombination were connected in this order by overlap extension PCR (OE-PCR), and inserted via double crossover event by natural competence for DNA uptake [14]. For *E. coli*, DNA fragments were similarly prepared and introduced by phage lambda Red recombinase-mediated homologous recombination [15]. The following protocol was optimized for *B. subtilis* and *E. coli* and may be applicable (with small modifications) to other bacteria.

2.1 Solutions

1. 3 M glycine: Dissolve glycine with H_2O.
2. Tris-buffered saline (TBS) buffer: 8 g of NaCl, 0.2 g of KCl, 3 g of Tris hydroxymethyl aminomethane are dissolved in 1 L H_2O. The pH of the solution is adjusted with HCl at 7.4.
3. SMM buffer: 0.02 M maleic acid, 0.5 M sucrose, 0.02 M $MgCl_2$. Prepare the appropriate volume of the solution and adjust to pH 6.5 with NaOH.
4. Phenylmethylsulfonyl fluoride (PMSF): 100 mM PMFS dissolved in methanol.
5. King2 buffer: 0.1 M Tris-Cl (pH 7.5), 0.2 M NaCl, 1% Triton X-100, 0.1% Na-deoxycholate, 0.2% Brij 58, 20% glycerol. Prepare the appropriate amount of solution and store at −25 °C.
6. Mg-Ca solution: 0.1 $MgCl_2$, 0.05 M $CaCl_2$. Prepare the solution using a 1 M $MgCl_2$ and a 1 M $CaCl_2$ stock solution.
7. UT buffer: 0.1 M HEPES, 8 M Urea, 0.5 M NaCl, 10 mM imidazole, 1% Triton X-100, 10 mM β-mercaptoethanol. Store at −25 °C.
8. Elution buffer: 0.1 M Tris–HCl (pH 7.5), 0.5 M imidazole (pH 7.5), 1% SDS, 10 mM DTT. Store at −25 °C.
9. M-wash buffer: 0.1 M Tris–HCl (pH 7.5), 1% SDS, 10 mM DTT. Store at −25 °C.

Fig. 2 (continued) GeF-seq results of AbrB-binding site for the *yqxM* locus. (**c**) A magnified image of panel B. The DNA sequence shown above the GeF-seq shows the corresponding DNA sequence of the window. The AbrB-binding regions determined by in vitro DNase I footprinting are indicated by a rectangle [23]. The GeF-seq resolution is similar to that of the in vitro DNase I footprint. (Reproduced from ref. 12 with permission from Oxford University Press.)

10. Phenol/chloroform/isoamyl alchohol solution: 25(Phenol):24(chloroform):1(isoamyl alchohol) to purify fragmented genomic DNAs.
11. 3 M sodium acetate (pH 5.2).
12. 20 mg/mL glycogen.
13. TAE buffer: 0.04 M Tris-acetate, 0.001 M EDTA.
14. LB medium: 10 g/L tryptone, 5 g/L yeast extract, 5 or 10 g/L NaCl (pH 7.0).

2.2 Equipments

1. Sonicator (suitable for rupture of cells in 3.5 mL solution for protein purification).
2. Magnetic stand (suitable for 1.5 mL and 2.0 mL microcentrifuge tubes).
3. Instruments to determine the length and amount of short DNA fragments (~150 bp) to confirm average length and quantity of DNA library for illumina sequencing (*see* **Note 1**).
4. Centrifuge tubes; 1.5, 2.0, and 15 mL.

2.3 Enzymes, Kit, and Other Materials

1. Lysozyme.
2. RNase A: 10 mg/mL, DNase-free.
3. DNase I.
4. EDTA-free protease inhibitor: to inhibit degradation of proteins.
5. Magnetic affinity resin for His-tag purification.
6. DNA Library generation kit: any Illumina sequencer compatible kit is applicable.
7. Low range agarose to prepare agarose gel to confirm the efficiency of DNaseI treatment and the average length of DNA fragments purified from cells.
8. 20 bp and 50 bp ladder marker.

2.4 Software

1. mpsmap, mkindx, pair_map: Programs to map short read sequences to a reference sequence and identify binding sites of DBPs [16]. This software and a detailed description of its usage is available at the GeF-seq website http://metalmine.mydns.jp/maps/pair_map/.
2. mapsview (available from our GeF-seq website shown above): Viewers of mapping results (*see* **Note 2**).
3. MEME, Bipad: Programs to identify consensus sequences across the binding sites of the DBP. Both programs are available at http://meme-suite.org/[17] (MEME), BiPad:http://bipad.cmh.edu/ [18] (Bipad).

3 Methods (*See* Note 3)

3.1 DNA-Protein Crosslinking for ChAP

1. Cultivate *B. subtilis* cells expressing His-tagged DBP in LB media (*see* **Note 4**) until any desired OD corresponding to ~160 OD units. For the first experiment, 2 extra samples should be prepared for optimization of DNase I digestion as described in the next section.
2. Add formaldehyde to a final concentration of 1%.
3. Incubate at 37 °C for 30 min with shaking to crosslink protein-DNA complexes (*see* **Note 5**).
4. Stop the crosslinking reaction by adding 60 mL of 3 M glycine (scalable, keeping the final concentration of glycine to 390 mM; *see* **Note 6**).
5. Centrifuge at 6000 × *g* for 5 min at room temperature, then discard the supernatant.
6. Add 2 mL of TBS buffer, mix by pipetting and transfer the cell suspension to a 15 mL tube.
7. Add an additional 10 mL of TBS buffer and mix well.
8. Centrifuge at 6000 × *g* for 5 min at room temperature and discard the supernatant.
9. Store the cell pellet at −80 °C until use.

3.2 Optimization of DNase I Digestion Condition

1. Use 2 of samples prepared in 3.1, resuspend cells in each tube with 3 mL of TBS, and mix the samples together.
2. Divide cell suspension to six fractions in 2 mL tubes.
3. Centrifuge at 6000 × *g* for 5 min at room temperature and then discard the supernatant.
4. Add 1 mL of SMM buffer (supplemented with 5 mg/mL lysozyme; *see* **Note 7**) to the pellet.
5. Add 10 μL of 100 mM PMSF to prevent degradation of proteins.
6. Incubate at 37 °C for 20 min with mixing.
7. Centrifuge at 6000 × *g* for 5 min at 4 °C and then discard the supernatant.
8. Add 1 mL of SMM buffer to the cell pellet. Gently dissolve the cell pellet by pipetting.
9. Centrifuge at 6000 × *g* for 5 min at room temperature and then discard the supernatant.
10. Add 126 μL of King2 buffer, 6.8 μL of 25 × protease inhibitor cocktail and 3.4 μL of RNase A (10 mg/mL).
11. Add 17 μL of Mg-Ca solution (final concentration: 10 mM $MgCl_2$, 5 mM $CaCl_2$).

12. Prepare a DNase I dilution series at 10 × concentrations (for example: 15, 12.5, 10, 5, 2.5, and 0 units/mL).
13. Add 17 μL of diluted DNase I solution (total reaction volume is 170 μL, final concentration of the DNaseI is 1.5, 1.25, 1.0, 0.5, 0.25, and 0 units/mL, *see* **Note 8**).
14. After thoroughly mixing the sample by gentle pipetting, incubate at 37 °C for 30 min for DNase I digestion.
15. Add 1 mL of UT buffer and 11.7 μL of 100 mM PMSF to stop the DNase I reaction.
16. Break cells with sonication.
17. Centrifuge at 16,000 × *g* at room temperature for 10 min.
18. Purify DNA fragment as described in Subheading 3.6, **step 2–12**.
19. Dissolve the DNA pellet in 30 μL nuclease-free water.
20. To confirm DNA digestion efficiency, electrophorese 1 μL of the supernatant DNA using a 2% low-range agarose gel in TAE buffer (*see* Fig. 3b).
21. Select optimum concentration of DNase I that generates DNA fragments around 50 bp for further experiments.

3.3 DNase I Digestion of Genomic DNA for ChAP

1. Add 3 mL of SMM buffer (supplemented with 5 mg/mL lysozyme; *see* **Note 7**) to the frozen pellet.
2. Add 30 μL of 100 mM PMSF to prevent degradation of proteins.
3. Incubate at 37 °C for 20 min with mixing.
4. Centrifuge at 6000 × *g* for 5 min at 4 °C and then discard the supernatant.
5. Add 3 mL of SMM buffer to the cell pellet. Gently dissolve the cell pellet in SMM buffer by pipetting.
6. Centrifuge at 6000 × *g* for 5 min at room temperature and then discard the supernatant.
7. Add 370 μL of King2 buffer, 20 μL of 25 × protease inhibitor cocktail and 10 μL of RNase A (10 mg/mL).
8. Add 50 μL of Mg-Ca solution (final concentration: 10 mM $MgCl_2$, 5 mM $CaCl_2$) and 50 μL of DNase I (0.01 units/μL, final concentration 1 unit/mL).
9. After thoroughly mixing the sample by gentle pipetting, incubate at 37 °C for 30 min for DNase I digestion.
10. Add 3 mL of UT buffer and 35 μL of 100 mM PMSF to stop the DNase I reaction.

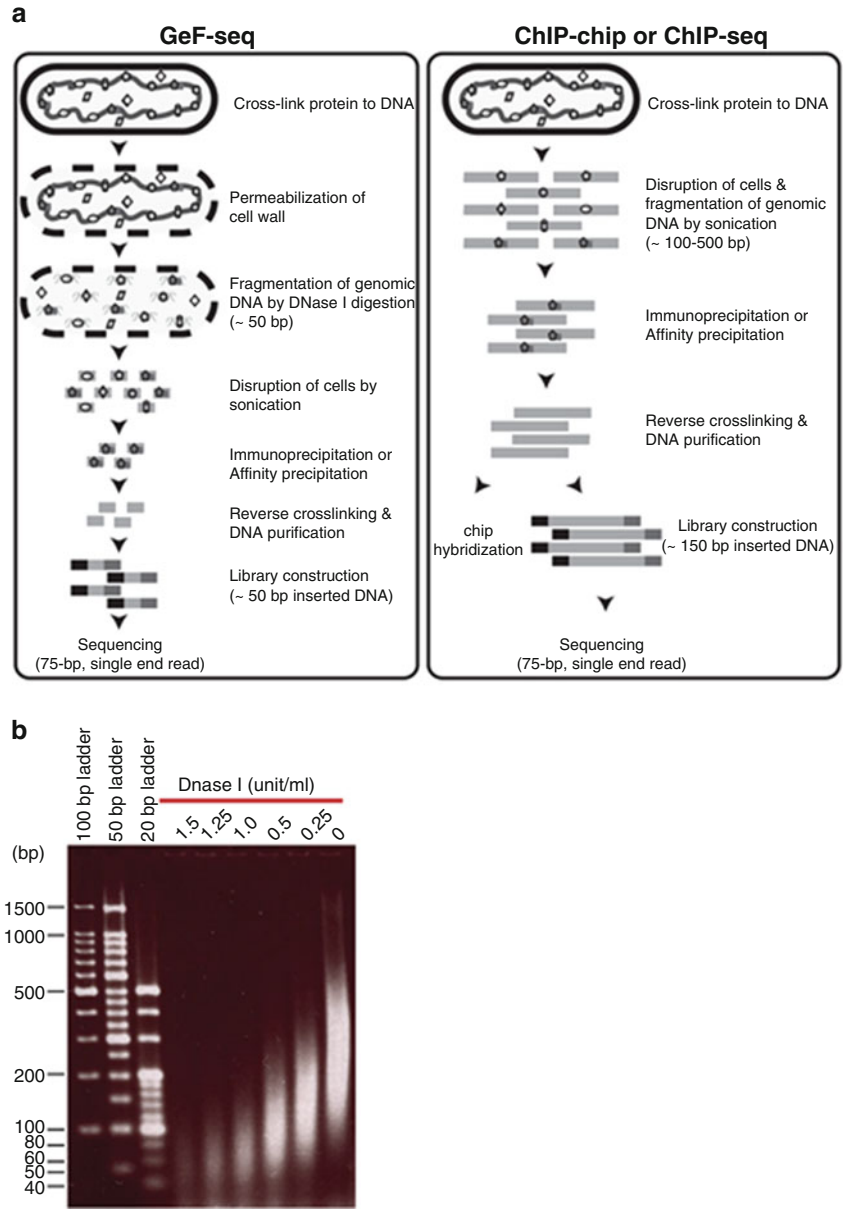


Fig. 3 The difference between ChIP-chip/ChiP-seq and GeF-seq methods and fragment size after DNase I digestion. (**a**) Schematic diagrams of the GeF-seq and ChIP-chip/ChIP-seq methods. The largest difference between the procedures is at the fragmentation step. With this fragmentation step, the DNA fragment size is close to the size of the binding region of DBP (~50 bo, *see* Fig. **b**). (**b**) Electrophoresis of the decrosslinked DNA fragments after the DNase I digestion step in the GeF-seq procedure. Fragments of ~50 bp are optimal for mapping of DBP-binding sites; they are suitable to acquire sharp peaks and to determine the recognition sequence accurately and comprehensively. (Reproduced from ref. 12 and ref. 21 with permission from Oxford University Press and YODOSHA.)

3.4 Sonication

1. Rupture the cells by sonication (*see* **Note 9**).
2. Centrifuge the sample at 6000 × *g* for 10 min and recover the supernatant.

3.5 Purification of the DNA-Protein Complex for ChAP

1. Before the purification of the DNA-protein complex, separate 200 μL of the supernatant from the sonicated sample (*see* above, Subheading 3.6, **step 2**), and store at −25 °C (*see* **Note 10**).
2. Transfer the remaining supernatant to two 2.0 mL tubes.
3. Add 50 μL of magnetic affinity resin for His-tag purification to each tube.
4. Gently mix the supernatant with the beads overnight at room temperature (*see* **Note 11**).
5. Centrifuge at 2000 × *g* for 1 min at room temperature.
6. Set the tubes on the magnetic stand to separate the beads and the supernatant.
7. Discard the supernatant using aspirator (or pipette).
8. Wash the magnetic beads with 1.5 mL of UT buffer. Separate the supernatant and beads using the magnetic stand and discard the supernatant.
9. Repeat washing **step 5** times.
10. After the final wash, remove as much of the supernatant as possible, then centrifuge at 16,000 × *g* for 3 min at room temperature.
11. Set the tubes on the magnetic stand and completely remove the supernatant using the pipette.
12. Add 55 μL of Elution buffer to the pellet of magnetic beads in each tube and gently mix for 10 min.
13. Centrifuge at 16,000 × *g* for 3 min at room temperature, then set the tubes on the magnetic stand and leave for about 1 min at room temperature.
14. Collect the supernatant as the purified fraction.
15. The total amount of the purified fraction becomes ~110 μL. The presence of protein included in the purified fraction should be confirmed by SDS-PAGE or Western blotting. (1) Mix 10 μL of the purified fraction with 2.5 μL of 5× SDS-PAGE loading buffer. (2) Heat at 90 °C for 45 min to remove crosslinking between the DNA and proteins. (3) Perform SDS-PAGE and/or Western blotting. If the target DBP forms a protein complex by this procedure, one can also determine the proteins included in the complex by protein complex analysis using, e.g., mass spectrometry.

3.6 Decrosslinking of the DNA-Protein Complex and Purification of the DNA Fragments

1. Transfer 85 μL of the purified fraction to a PCR tube and mix with 15 μL of M-wash buffer.
2. Transfer 30 μL of the supernatant fraction to a PCR tube and mix with 70 μL of M-wash buffer.
3. Incubate the purified and supernatant fractions at 65 °C overnight (15~18 h) to decrosslink protein-DNA complexes.
4. To purify DNA fragments, Phenol/chloroform/isoamyl alchohol treatment was used (*see* **Note 12**). Transfer the purified and supernatant fractions to 1.5 mL tubes and mix with 200 μL of Phenol/chloroform/isoamyl alchohol solution.
5. Add 100 μL of distilled water, 20 μL of 3 M sodium acetate (pH 5.2) and 1 μL of 20 mg/mL glycogen, then mix well.
6. Centrifuge at 16,000 × *g* for 5 min at room temperature and transfer 160 μL of supernatant (upper phase) to a new 1.5 mL tube.
7. Add 400 μL of ethanol (2.5×) and mix well, followed by incubation at –80 °C for 20 min.
8. Centrifuge at 16,000 × *g* for 5 min at 4 °C and then discard the supernatant.
9. Add 500 μL of cold 70% ethanol and gently mix by inverting the tube.
10. Centrifuge at 16,000 × *g* for 5 min at 4 °C and discard the supernatant completely.
11. Dissolve the DNA pellet in 30 μL nuclease-free water. The DNA recovered from the purified fraction is "ChAP DNA" and that from the supernatant fraction is "supernatant DNA."
12. To confirm DNA digestion efficiency, electrophorese 1 μL of the supernatant DNA using a 2% low-range agarose gel in TAE buffer (*see* **Note 8**).
13. Prepare the libraries for Illumina paired-end sequencing using the NEBnext DNA Library Prep Reagent kit according to the manufacturer's instructions.

3.7 GeF-Seq Data Analysis (See Note 12)

The detailed procedure for informatics analysis to determine the DNA recognition sites of DBPs on the genome, based on GeF-seq analysis, is described at our web site: http://metalmine.mydns.jp/maps/pair_map/, and programs needed for the analysis can be downloaded from the web site (*see* Subheading 2.4).

Here, we briefly describe the procedure to identify the binding sites with basepair resolution using "pair_map" programs (*see* **Note 13**).

1. Perform mapping of sequence reads on the reference genome using "mpsmap":

mkindx −11 genome.fasta (index preparation).
mpsmap –a 8 –i 11 –r 36 –h 35 –map –job jobname genome.fasta gefseq.fastq (*see* **Note 14**).

2. (Optional) Bias caused by copy number variation, as known as *ori-ter* bias can be adjusted with data from regular genome sequencing as a control. A detailed description of the procedure can be found in http://metalmine.mydns.jp/maps/gefseq/.
3. Detect the binding peak positions using "pair_map". The predicted protein recognition sequence segments are written in gefseq.fastq_jobname_edge.fasta (*see* **Note 14**).
4. Search the conserved sequence motif among the sequence segments using software such as MEME or Bipad [17, 18].
5. Visualize the mapping results and predicted binding sites using mapsview.

4 Notes

1. We use the Agilent Bioanalyzer (Agilent). However, other instruments to measure precise length and amount of DNA fragments are also applicable, for instance, the Tapestation (Agilent).
2. Mapsview can only be used with mpsmap. If you use other well-known mapping program (for instance, Bowtie2 [19]), you can use IGV [20]. Even in this case, mpsmap will be needed for the determination of the edges of the DBP-binding sites.
3. While we examplified the GeF-seq of the DBP (AbrB) highly expressed in the *B. subtilis* cells cultivated in the LB medium, this protocol can be applied to other DBPs lower or higher expressed in *B. subtilis* and *E. coli* cells. Growth conditions and reagent volumes can be adjusted according to the expression level of the target DBP. Our standard cultivation medium is LB. However, other media (for instance, 2 × YT) can also be applied for GeF-seq.
4. The volume of the bacterial culture can be variable. This protocol is optimized for 160 OD units of cells.
5. Temperature, reaction time, and duration of the mixing for the crosslinking reaction can be varied to optimize the crosslinking efficiency, which may depend on the properties of the target protein, bacterial species, and culture conditions.
6. At **step 6**, Tris contained in the TBS buffer would stop the reaction. Therefore, even if this step is omitted, the reaction should be stopped at a later step.

7. Lysozyme should be added to SMM buffer immediately before use. When *E. coli* is used for GeF-seq, it is necessary to add 1 mM EDTA into SMM buffer and leave the cell suspension for 15 min at room temperature to remove outer-membrane, because Gram-negative bacteria is resistant to lysozyme by their outer-membrane.
8. The fragment size produced by the DNase I digestion can be controlled by adjusting the amount of enzyme and the incubation time. Because DNase I activity can vary between suppliers or different lots of the product, experimental conditions have to be determined at the beginning of the analysis.
9. We use the Astrason Ultrasonic Processor XL2020 (Misonix, NY, USA). The sonication conditions used are: level 5, with a 4 s ON period followed by a 10 s OFF period (which is the interval required to avoid overheating of samples). The ON (4 s) and OFF (10 s) cycle is repeated until the total ON time is 10 min.
10. The supernatant fraction will be used to confirm the sizes of fragmented DNA from DNase I treatment, and to prepare the control library for sequencing.
11. After inactivation of protease by PMSF in earlier steps, proteins and DNA fragments are stable under this denatured condition.
12. Any methods that can recover small fragments of DNA also can be used, e.g., silica-based DNA purification systems.
13. GeF-seq is based on DNase I footprinting. The accuracy of the mapped positions for both ends of the DNA fragments co-purified with the DNA-binding proteins is crucial to the quality of the binding site predictions. Programs "pmapsr" and "pair_map" have been developed to identify the edges of the protein-binding sites with single base resolution using Illumina sequencing data.
14. To identify the protein-binding sequences, first, our mapping program, "mpsmap", assigns short reads including the adapter-oligo sequence by permitting the presence of a large mismatch region. Then the "pmapsr" program identifies the 5′ and 3′ edges of the DNaseI digested DNA fragments based on the positions as the first position of the read, and the mismatch start position, from each read. Alternatively, the "pair_map" program identifies the both end of the digested DNA fragments as the genome regions in between the initial positions of the read pairs facing each other.

Acknowledgments

We thank Dr. Jon Hobman for critical reading of the manuscript. This work was supported by JSPS KAKENHI JP15K07359 (to SI), JP26430199 (to KN), JP26450090 and JP26106004 (to TO).

References

1. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y et al (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4:651–657
2. Spiro S (2012) Genome-wide mapping of the binding sites of proteins that interact with DNA. Methods Mol Biol 881:137–156
3. Prieto AI, Kahramanoglou C, Ali RM, Fraser GM, Seshasayee AS et al (2012) Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. Nucleic Acids Res 40:3524–3537
4. Kahramanoglou C, Seshasayee AS, Prieto AI, Ibberson D, Schmidt S et al (2011) Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. Nucleic Acids Res 39:2073–2091
5. Blasco B, Chen JM, Hartkoorn R, Sala C, Uplekar S et al (2012) Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein. PLoS Pathog 8: e1002621
6. Myers KS, Yan H, Ong IM, Chung D, Liang K et al (2013) Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. PLoS Gen 9: e1003565
7. Perkins TT, Davies MR, Klemm EJ, Rowley G, Wileman T et al (2013) ChIP-seq and transcriptome analysis of the OmpR regulon of *Salmonella enterica* serovars Typhi and Typhimurium reveals accessory genes implicated in host colonization. Mol Microbiol 87:526–538
8. Pepke S, Wold B, Mortazavi A (2009) Computation for ChIP-seq and RNA-seq studies. Nat Methods 6:S22–S32
9. Li H, Rhodius V, Gross C, Siggia ED (2002) Identification of the binding sites of regulatory proteins in bacterial genomes. Proc Natl Acad Sci U S A 99:11772–11777
10. Robison K, McGuire AM, Church GM (1998) A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. J Mol Biol 284:241–254
11. Galas DJ, Schmitz A (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res 5:3157–3170
12. Chumsakul O, Nakamura K, Kurata T, Sakamoto T, Hobman JL et al (2013) High-resolution mapping of in vivo genomic transcription factor binding sites using in situ DNase I footprinting and ChIP-seq. DNA Res 20:325–338
13. Chumsakul O, Anantsri DP, Quirke T, Oshima T, Nakamura K et al (2017) Genome-wide analysis of ResD, NsrR, and Fur binding in *Bacillus subtilis* during anaerobic fermentative growth by in vivo Footprinting. J Bacteriol 199:e00086-17
14. Spizizen J (1958) Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. Proc Natl Acad Sci U S A 44:1072–1078
15. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645
16. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H et al (2011) Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res 39:e90
17. Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite. Nucleic Acids Res 43:W39–W49
18. Bi C, Rogan PK (2004) Bipartite pattern discovery by entropy minimization-based multiple local alignment. Nucleic Acids Res 32:4979–4991
19. Langmead B, Salzberg S (2012) Fast gapped-read alignment with bowtie 2. Nat Methods 9:357–359
20. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192
21. Oshima T, Ishikawa S, Chumsakul O, Nakamura K (2014) High resolution determination methods of the DNA binding protein binding site. (Japanese). Standard protocols on Next Generation Sequencing (ZIKKEN IGAKU

extra issue), Yodosha co. ltd., Japan (Japanese); p. 131–142

22. Chumsakul O, Takahashi H, Oshima T, Hishimoto T, Kanaya S et al (2011) Genome-wide binding profiles of the *Bacillus subtilis* transition state regulator AbrB and its homolog Abh reveals their interactive role in transcriptional regulation. Nucleic Acids Res 39:414–442
23. Strauch MA, Bobay BG, Cavanagh J, Yao F, Wilson A et al (2007) Abh and AbrB control of *Bacillus subtilis* antimicrobial gene expression. J Bacteriol 189:7720–7732

Chapter 4

Genomic SELEX Screening of Regulatory Targets of *Escherichia coli* Transcription Factors

Tomohiro Shimada, Hiroshi Ogasawara, and Akira Ishihama

Abstract

The genome of *Escherichia coli* K-12 is transcribed by a single species of RNA polymerase. The selectivity of its transcriptional targets is modulated via two-steps of protein-protein interaction: at the first step, seven species of the sigma subunit are involved, at the second step, a total of approximately 300 species of transcription factor (TFs). For the identification of the regulatory targets of these two groups of regulatory proteins, we developed two in vitro approaches, "Genomic SELEX" (currently designated as gSELEX) and "PS (promoter-specific)-TF" screenings. Here, we describe a detailed protocol of the genomic SELEX screening system which uses purified regulatory proteins and fragments of genomic DNA from *E. coli*.

Key words Genomic SELEX, Transcription factor: Sigma factor, RNA polymerase, DNA-protein interaction, Regulatory target, *Escherichia coli*

1 Introduction

Bacteria constantly monitor the physical, chemical and biological conditions of their environment, and respond for adaptation and survival by modifying their gene expression pattern. Transcription of genes is carried out by a single species of RNA polymerase (RNAP). The model organism *Escherichia coli* K-12 harbors about 4500 genes on its genome, but the total amount of RNAP is only about 2000 molecules per genome [1, 2]. The pattern of genome transcription is determined by controlling the use of this limited number of RNAP along the genome [2, 3]. The RNAP core enzyme with subunit composition $\alpha_2\beta\beta'\omega$ carries the activity required for RNA polymerization, but lacks the ability of promoter recognition and transcription initiation. The holoenzyme, which is formed after binding the σ subunit, is capable of performing these steps. *E coli* K-12 expresses seven species of the σ subunit, each capable of recognizing a specific set of promoters [1, 4–6] (*see* Fig. 1a). The promoter selectivity of the RNAP holoenzyme is further modulated via protein-protein interaction with another

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_4,

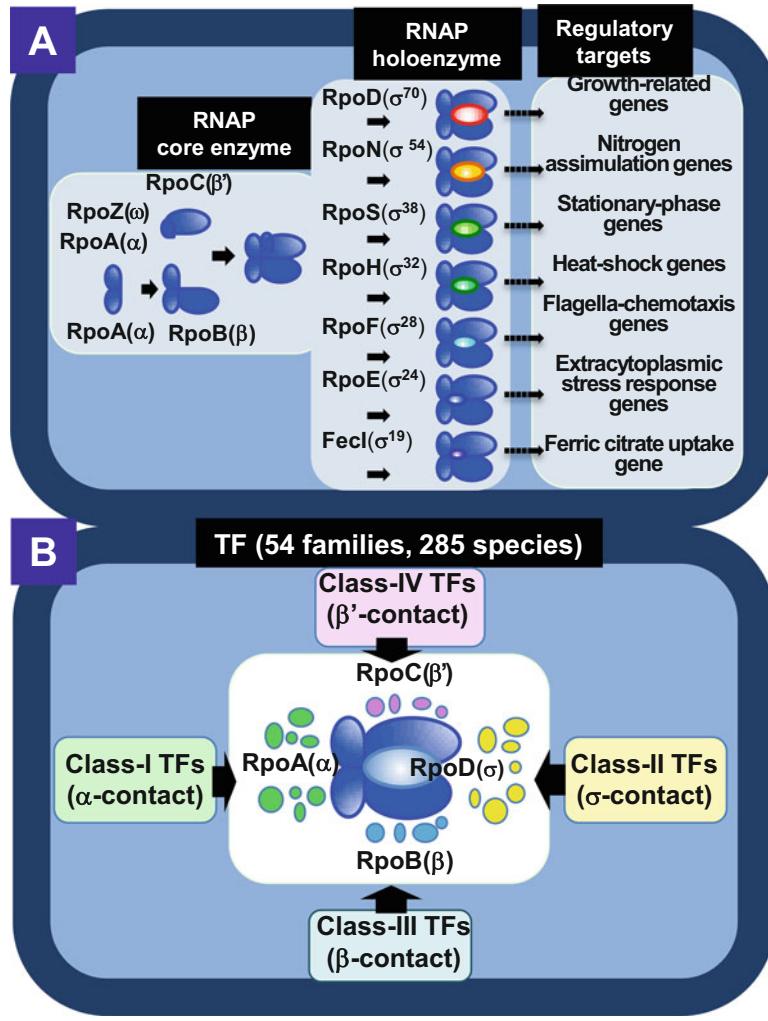


Fig. 1 Functional differentiation of RNA polymerase. *E. coli* RNA polymerase core enzyme with the subunit composition of $\alpha_2\beta\beta'\omega$ is assembled in a sequential manner. The core enzyme lacks the activities of promoter recognition and transcription initiation, but gains these activities by binding one of seven species of the σ subunit, forming the holoenzyme. Each holoenzyme recognizes a specific set of the regulatory target promoters of genes that are involved in specific cell functions as indicated [2, 3]

group of regulatory proteins, referred to as transcription factors (TF) [2, 3]. *E. coli* K-12 expresses approximately 300 species of TF, most of which associate with DNA targets, usually located near promoters and interact with the promoter-bound RNAP (*see* Fig. 1b). DNA-bound TFs interact directly with promoter-bound RNAP via one of its subunits [7–9]. Identification of the regulatory targets for all seven σ subunits and all 300 TFs is important in order to understand global gene regulation in *E. coli*.

Several experimental approaches have been established for the identification of regulatory targets of the σ subunits and TFs, among which microarray-based transcriptome analysis for monitoring mRNA levels of the full set of genes in the presence and absence of test factor [10–12]. Transcriptome analysis has also been performed under an endless variety of conditions for monitoring the genes involved in stress response. Currently, the microarray technology for transcriptome analysis has been replaced by direct sequencing of all transcripts (RNA-Seq analysis) [13, 14] (for details *see* **Note 1**). In parallel, TF-binding sites on the *E. coli* genome have been mapped by using the ChIP-chip (chromatin-immunoprecipitation) method [15, 16]. This method has been further improved by replacing microarray detection by deep sequencing (ChIP-Seq), which enables true genome-wide mapping of TF-binding sites. A huge amount of transcription data has been collected in databases such as EcoCyc [17] [https://EcoCyc.org] and RegulonDB [18] [http://regulondb/ccg.unam.mx]. The data set identified so far, however, does not necessarily represent the regulatory targets under the direct control of the test σ factor and test TF, but instead include large numbers of indirectly affected genes [2, 3, 19]. Generally the direct targets of TFs represent only minor fractions of the genes detected by Microarray and RNA-Seq analyses, because downstream of a specific regulator, genes encoding other transcription regulators are organized, altogether forming a complex regulatory network. Moreover, caution is needed when selecting in vivo transcription datasets as unexpected differences have been found between *E. coli* strains due to differences in genetic background, even between lab stocks of the same *E. coli* K-12 strain (for details *see* **Note 2**).

To avoid the complexity inevitably associated with in vivo systems, we developed two in vitro approaches, one aimed at identifying the regulatory target promoters of each sigma factor and each TF, and the other aimed at identifying all sigma factors and TFs involved in regulation of each promoter. "Genomic SELEX (systematic evolution of ligands by exponential enrichment)" allows identification of target promoters, genes and operons under the direct control of each sigma factor and each TF ([19, 20]; also *see* TEC database [http://www.shigen.nig.ac.jp/ecoli/tec/]). The "PS-TF (promoter-specific TF)" screening system allows identification of all TFs involved in regulation of each promoter [24]. By using Genomic SELEX screening system, we have identified all "constitutive promoters" under the direct control of each sigma factor in the absence of other activators and repressors [21, 22]. Here, we describe details of the Genomic SELEX procedure for identification of regulatory targets of TFs.

2 Materials

2.1 Bacterial Strains

Escherichia coli K12 W3110 type-A [23].
Escherichia coli DH5α.
Escherichia coli BL21(DE3).
Escherichia coli JM109.

2.2 Chemicals

Ni-NTA Agarose.
Protein marker.
Cy3-dCTP.
Cy5-dCTP.
Human Cot-1 DNA.

2.3 DNA

pBR322 plasmid.
pET21a plasmid.
100 bp DNA ladder marker.

2.4 Oligonucleotides

F-primer (100 pmol/μL): 5′ CTTGGTTATGCCGGTACTGC 3′.
R-primer (100 pmol/μL): 5′ GCGATGCTGTCGGAATGGAC 3′.

2.5 Tiling Array

E. coli K12 4×44 Universal array (Oxford Gene Technology)
Backing plates (Oxford Gene Technology)
Hybridization chamber

2.6 Enzymes

EcoRV.
HindIII.
BamHI.
T4 polynucleotide kinase.
CIAP (Calf intestine alkaline phosphatase).
Proteinase K.
Ex Taq polymerase.

2.7 Culture Media

LB medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl).

SOC medium (20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 186 mg/L KCl, 2 mM $MgCl_2$, 8 mM D-glucose).

LB plate (LB medium with 15 g/L agar).

2.8 Gels

5% Polyacrylamide (1.67 mL 30% acrylamide solution (29:1), 0.1 mL 10% APS, 0.01 mL TEMED, 1.0 mL 10×TBE buffer, and H_2O to make 10 mL).

Agarose gel (0.8~1.0 g agarose in 100 mL TAE buffer for 0.8~1.0% gel).

SDS-polyacrylamide lower gel (30% acrylamide solution (29.2:0.8), 0.1 mL 10% APS, 0.01 mL TEMED, and Laemmli lower-gel buffer (*see* Subheading 2.9) to make 10 mL gel).

SDS-Polyacrylamide upper gel (30% acrylamide solution (29.2:0.8), 0.1 mL 10% APS, 0.01 mL TEMED, and Laemmli upper-gel buffer (*see* Subheading 2.9) to make a 10 mL gel).

2.9 Buffers

TAE buffer (40 mM Tris-base, 20 mM acetate, 1 mM EDTA, and adjust pH to 8.0).

TBE buffer (500 mM Tris-base, 485 mM boric acid, 20 mM EDTA, and adjust pH to 8.0).

10× Binding buffer (10 mL 1 M Tris–HCl, pH 7.8 at 37 °C, 6 mL 0.5 M Mg acetate, 30 mL 5 M NaCl, 1 mL 25 mg/mL BSA, and H_2O to a final volume of 100 mL).

Protein elution buffer (appropriate concentration of imidazole in binding buffer).

Imidazole elution buffer (in Binding buffer).

Cell wash buffer (0.16 M NaCl, 10 mM Tris–HCl, pH 8.0 at 4 °C).

Cell suspension buffer (0.10 M NaCl, 50 mM Tris–HCl, pH 8.0 at 4 °C).

4× Laemmli lower-gel buffer (181.7 mg/mL Tris-base, 4.0 mg/mL SDS, pH 8.8).

4× Laemmli upper-gel buffer (60.6 mg/mL Tris-base, 4.0 mg/mL SDS, pH 6.8).

10× Laemmli running buffer (30.28 mg/mL Tris-base, 144.14 mg/mL glycine, 10 mg/mL SDS).

5× SDS dye (1 M Tris–HCl, pH 6.8 at 25 °C, 0.5 g SDS, 2.5 mL 2-mercaptoethanol, 0.1 mL 1% BPB, 0.5 mL glycerol).

Storage buffer (50% glycerol, 10 mM Tris–HCl, pH 7.5 at 4 °C, 20 mM $MgCl_2$, 200 mM KCl, 5 mM EDTA).

2.10 Reagents

30% Acrylamide solution [29:1] (290 mg/mL acrylamide, 10 mg/mL N′, N′-methylene bis-acrylamide).

30% Acrylamide solution [29.2:0.8] (292 mg/mL acrylamide, 8 mg/mL N′, N′-methylene bis-acrylamide).

6× loading dye (5 mg bromophenol blue, 3 mL glycerol, 0.1 mL 0.5 M EDTA, 6.9 mL H_2O).

Tris-saturated phenol (500 g crystallized phenol, 400 mL 0.5 M Tris–HCl, pH 8.0, 0.5 g 8-quinolinol).

Phenol-Chloroform solution (50 mL Tris-saturated phenol + 50 mL chloroform).

10× Blocking agent (Agilent).

2× HiRPM hybridization buffer (Agilent).

2.11 Reaction Kits

DNeasy blood & tissue Kit (Qiagen).
Blunting kit (T4 DNA polymerase) (Takara).
DNA ligation kit (Takara).
Plasmid extraction kit (QIAGEN-tip 500).
PCR purification kit (QIAGEN QIA quick).
CBB (Coomassie brilliant blue) staining kit.

Protein assay kit (Biorad).
BioPrime DNA labeling system (Invitrogen).
Oligo αCGH/ChIP-on-chip hybridization kit (Agilent).
Oligo αCGH wash-buffer-1 and -2 sets (Agilent).

2.12 Equipment

Microcentrifuge.
Centrifugal concentration spin dryer.
UV transilluminator.
SUPREC™-01 tube (TaKaRa).
Muromac Column (5 × 50 mm) (Muromachi Technos).
Photometer.
Sonicator.
Thermal cycler.
Hybridization oven (Agilent).
Glass container (Agilent).
High resolution microarray scanner (G2565CA Agilent).
Image analyzer (GE Healthcare).
DNA sequencer.

3 Construction of the Genomic DNA Library

3.1 Preparation of the Genomic DNA Segments

1. Purify genomic DNA from *E. coli* cells using DNeasy blood and tissue kit. Solubilize the purified DNA in nuclease-free water, and measure its DNA concentration.
2. Check DNA purity by 0.7% agarose gel electrophoresis.
3. Sonicate 0.5 mL aliquot of the genomic DNA (0.2 mg/mL) using a sonicator at 10 s interval for a total of 120 s.
4. Check the fragmentation of genomic DNA by 5% PAGE (polyacrylamide gel electrophoresis). Use 100 bp DNA ladder as a reference. The size distribution of major DNA segments should be in the range 200~300 bp. [Note: if the size is longer than this level, repeat the sonication].
5. Excise the part of the gel containing 200–300 bp-long DNA segments with a surgical knife.
6. Elute DNA from smashed gel pieces by soaking at 4 °C in 400 μL of elution buffer overnight.
7. Centrifuge the gel suspension for 17,800 × *g* for 10 min at room temperature.
8. Remove the supernatant and keep for further processing.
9. Repeat the DNA elution by adding 400 μL of elution buffer to the gel.
10. Combine supernatant with that obtained at **step 8** and transfer into SUPREC™-01 tube to remove remaining gel pieces.

11. Centrifuge the tube for 9100 × *g* for 5 min at room temperature.
12. Add 1 mL of 100% chilled ethanol to the flow-through fraction for DNA precipitation.
13. Centrifuge the tube for 17,800 × *g* for 20 min at 4 °C.
14. After the removal of the supernatant, add 200 μL of 70% ethanol.
15. Centrifuge the tube for 17,800 × *g* for 5 min at 4 °C.
16. Remove the supernatant with microcentrifuge and dry DNA precipitates under centrifugal concentration spin dryer.
17. Dissolve DNA in 100 μL of nuclease-free water and measure the DNA concentration.

3.2 Blunt-End Formation and Phosphorylation of the Genomic DNA Fragments

1. Treat 10× 1 μg of DNA isolated at step x.17 with Takara blunting kit to obtain 10 μg of blunt-ended DNA.
2. Purify the DNA with phenol/chloroform solution and precipitate it with ethanol.
3. Dissolve the DNA in 80 μL of nuclease-free water.
4. Add 10 μL of 10× polynucleotide kinase buffer, 10 μL of 20 mM ATP solution, and 1 μL (10 units) of T4 polynucleotide kinase.
5. Incubate for 30 min at 37 °C.
6. Purify the DNA with phenol/chloroform solution, and precipitate it with ethanol in the presence of 2 M ammonium acetate for the removal of ATP.
7. Dissolve the DNA in 0.1 mL nuclease-free water and determine the DNA concentration by measuring OD_{260nm}.

3.3 Preparation of EcoRV-Digested and Dephosphorylated pBR322 Vector

1. Digest 5 μg of pBR322 with 15 units EcoRV restriction enzyme.
2. Purify the plasmid with phenol-chloroform solution, and precipitate DNA with ethanol.
3. Dissolve the DNA in 90 μL of nuclease-free water.
4. Add 10 μL of 10× alkaline phosphatase buffer and 0.5 μL (10 units) of calf intestine alkaline phosphatase (CIAP).
5. Incubate for 30 min at 37 °C.
6. Add 0.5 μL (10 units) of CIAP.
7. Incubate for 30 min at 37 °C.
8. Add 5 μL of 10% SDS solution for inactivation of CIAP.
9. Add 1 μL of 500 mM of EDTA solution, and then 0.5 μL of proteinase K (20 mg/mL).
10. Incubate at 56 °C for 30 min.

11. Extract the DNA with phenol/chloroform solution.
12. Precipitate the DNA with ethanol, and then dissolve in 20 μL of nuclease-free water.
13. Determine the DNA concentration of EcoRV-digested and dephosphorylated plasmid solution by measuring OD_{260nm}.

3.4 Ligation and Transformation

1. Mix 15 ng of blunt-ended genomic DNA segments, and 100 ng of EcoRV-digested and dephosphorylated vector DNA in a volume of 5 μL so as to keep the molar ratio of insert DNA:vector DNA at 3:1.
2. Add 5 μL of DNA Ligation Kit and incubate overnight at 16 °C.
3. Transform 10 μL of ligated DNA solution into 100 μL of the competent cells.
4. Incubate on ice for 30 min.
5. Incubate at 42 °C for 45 s for heat shock
6. Cool-down on ice for 3 min.
7. Add 300 μL SOC medium.
8. Incubate for 30 min at 37 °C.
9. Spread 50 μL each of the transformant solution on LB-ampicillin plate using cell spreader.
10. Incubate the plate at 37 °C overnight.
11. Check the number of colonies [Note: the number should be more than 1000 on each plate].

3.5 Quality Check of the Plasmid Library

1. Pick up few dozens of colonies in 2 mL of LB-ampicillin medium and cultivate at 37 °C up to log phase.
2. Harvest cells and isolate plasmid DNA with phenol-chloroform
3. Digest each plasmid DNA by treatment with HindIII and BamHI restriction enzymes.
4. Run on 1.0% agarose gel, in parallel with 200 bp DNA ladder marker, and stain using ethidium bromide.
5. Check the fragment size of the digested plasmid [Note: when 200–300 bp-long DNA segments were cloned into plasmid, 550–650 bp-long fragments can be observed].
6. Prepare a total number of 1×10^5 transformants [Note: the total amount of genomic DNA segments should be 5~6 times more than the size of *E. coli* K-12 W3110 genome of 4.6 Mbp].
7. Suspend individual colonies in plasmid extraction kit (QIAGEN-tip 500).
8. Purify the plasmid DNA following the procedure of supplier.

4 Preparation of DNA Substrates for SELEX

4.1 Regeneration of DNA Segments from the Genomic DNA Library

1. Mix 3 ML of the plasmid library of genomic DNA (1 ng/μL), 1 μL of F-primer (100 pmol/μL), 1 μL of R-primer (100 pmol/μL), 10 μL of 10× Ex Taq buffer, 8 μL of dNTP mixture, 76.5 μL of ddH_2O, and 0.5 μL of Ex Taq polymerase.
2. Subject to PCR under the conditions: 94 °C for 5 min treatment; and then [94 °C-30 s -> 50 °C-1 min -> 70 °C-1 min] for 20 cycles followed by 70 °C treatment for 3 min. [Note: Optimization is required depending on the type of thermal cycler].
3. Add 20 μL of 6× loading dye to the PCR product and run it, together with 100 bp DNA ladder marker, on 5% polyacrylamide gel for 40 min at a constant voltage of 150 V.
4. Stain using ethidium bromide solution and visualize on UV-transilluminator to verify synthesis of products of the expected length.
5. Cut the gel piece containing 200–400 bp-long DNA segments with a surgical knife.
6. Extract the DNA from gel pieces as noted above (1–1).
7. Dissolve DNA in 40 μL nuclease-free water and measure the DNA concentration. Store the DNA sample at −20 °C.

5 Expression and Purification of TFs

5.1 Construction of TF Expression Plasmids

1. For cloning of TF, design both 5′- and 3′-primers, each hybridizing the flanking sequences of 5′- and 3′-termini of the coding sequence of the TF of interest. For the insertion into the expression plasmid pET21a, add one unique restriction enzyme sequence to each primer (for details *see* [25]).
2. Amplify TF-coding sequences by PCR using the 5′- and 3′-primers thus prepared under the PCR conditions: 94 °C for 5 min treatment; and then [94 °C-30 s > 50 °C-1 min > 70 °C-1 min] for 20 cycles followed by 70 °C treatment for 3 min.
3. Purify the PCR-amplified DNA with phenol-chloroform.
4. Digest the purified PCR-amplified DNA fragment with two kinds of restriction enzyme, each introducing a single cleavage within one of the primer pairs, and insert it into pET21a vector between the same restriction sites as used for the preparation of insert DNAs.

5.2 Preparation of Transformants Carrying TF Expression Plasmids

1. Add 0.1 ng of pTF plasmid DNA into 1.5 mL tube including 50 μL of *E. coli* BL21 competent cells.
2. Incubate on ice for 10 min and then expose to 42 °C for 45 s. After cooling down on ice, add 120 μL of LB medium.
3. Incubate at 37 °C for 40 min, and then spread onto LB plate containing Amp for isolation of transformants.

5.3 Check the Small-Scale Expression of His-Tagged TF

1. Pick a single colony and inoculate into 3 mL of LB medium.
2. Incubate at 30 °C in a water bath shaker at 150 rpm until an $OD_{600} = 0.7 \sim 1.0$, and then store at 4 °C.
3. Add an aliquot of this pre-culture into 20 mL of fresh LB medium containing 0.1 mg/mL ampicillin.
4. Incubate at 30 °C in a water bath shaker at 150 rpm until an $OD_{600} = 0.5 \sim 0.7$.
5. Add 0.5 M IPTG as an inducer to achieve a final concentration of 1 mM IPTG, and continue the incubation.
6. Take 1 mL aliquots at one-hour intervals to measure the expression level of TF.
7. Harvest cells by centrifugation by using a refrigerated centrifuge.
8. Suspend the cells in more than 0.3 mL of cell suspension buffer.
9. Add 100 mM PMSF solution to achieve a final concentration of 0.2 mM PMSF.
10. After storage on ice for 20 min, sonicate the cell suspension using the following setting: each 0.5 s a pulse at amplitude of 14% for a total of 30 s [Note: the condition of sonication should be controlled depending on the equipment used].
11. Take 0.1 mL cell lysate and centrifuge at 4 °C for 20 min at 20,400 × *g*, and save supernatant (S1).
12. Suspend cell pellets in 0.1 mL lysis buffer containing 2% TritonX-100, centrifuge at 4 °C for 20 min at 20,400 × *g*, and save the supernatant (S2).
13. Resuspend cell pellets in 0.1 mL lysis buffer containing 2 M guanidine, centrifuge at 4 °C for 20 min 20,400 × *g*, and save the supernatant (S3).
14. Suspend the final cell pellet in 0.1 mL of lysis buffer.
15. Take 12 μL each of the supernatant fractions (S1–3) and the cell pellet fraction and add 3 μL of 5× SDS dye. Run all four samples alongside a protein size maker on 6–10% of SDS-PAGE at a constant current of 30 mA for appropriate time.

16. Stain the gel using CBB (Coomassie brilliant blue) staining kit for 1 h. Rinse the gel and detect with use of Image analyzer.
17. Compare the gel patterns for samples collected at different time points after the addition of IPTG to confirm TF induction, and among supernatant and precipitate fractions to find the best condition of TF extraction.

5.4 Large-Scale Expression and Purification of His-Tagged TF

1. For large-scale expression of TFs, follow the culture conditions as described for small-scale culture and harvest cells at the time following IPTG addition that resulted in maximum induction of TF.
2. Suspend TF-expressing cells in 30 mL of lysis buffer and add 0.06 mL of 100 mM PMSF solution to obtain a final concentration of 0.2 mM PMSF. [Note: PMSF is a protease-inhibitor]
3. Add 0.9 mL of 10 mg/mL lysozyme (final 0.3 mg/mL) and incubate on ice for 20 min.
4. Sonicate the cell suspension using the same settings as employed for small-scale expression.
5. Centrifuge the cell lysate at 4 °C for 30 min at 20,400 × *g*.
6. Collect the supernatant.
7. In case a TF is found to be present in the precipitate during the small-scale test, extract the TF with use of TF suspension buffer containing 1% TX-100 and/or 1 mM guanidine.
8. Into the combined TF extract, add 4 mL of Ni-NTA agarose suspension, and mix well using a rotator at 4 °C for 1 h.
9. Load the Ni-NTA agarose suspension onto a column and collect the flow-through fraction as the unbound TF sample. [Note: His-tagged TFs bind to agarose-attached Nickel].
10. Wash the column with 20 mL of lysis buffer (containing TX-100 and/or guanidine).
11. Wash the column with 20 mL of lysis buffer containing 25 mM Imidazole.
12. Elute TF using by applying wash buffer containing increasing Imidazole concentrations: (1) 2 mL of wash buffer containing 100 mM Imidazole, (2) 2 mL of wash buffer containing 200 mM Imidazole and (3) 4 mL of wash buffer containing 400 mM Imidazole.
13. Take a 12 μL aliquot from each fraction and add 3 μL of 5×SDS dye.
14. Run samples, alongside 3 μL of protein marker, on a 6~10% SDS-PAGE at a constant current of 30 mA for an appropriate time estimated from the mobility of marker dye

15. Stained the gel with CBB stained kit for 1 h. Based on the location of TF bands. Choose the fraction(s) containing high-level TF.
16. Combine the imidazole eluates containing high-level TF and dialyze against the storage buffer at 4 °C.
17. Measure the protein concentration by any protein staining reagent and store the purified TF at −80 °C.

6 Genomic SELEX Screening

Genomic SELEX was developed to identify regulatory targets of *E. coli* TFs using a library of genomic DNA segments (*see* Subheading 4) and purified His-tagged TFs (*see* Subheading 5) [20] (*see* Fig. 2). For enrichment of specific targets, SELEX is often repeated several cycles.

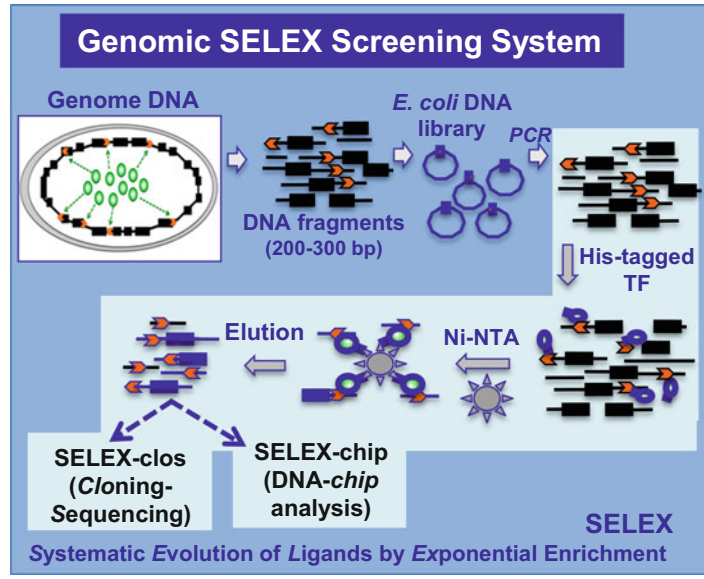


Fig. 2 Genomic SELEX screening system. The genome DNA of *E. coli* K-12 W3110 was sonicated into a mixture of 200~300 bp-long fragments, which were inserted into a multi-copy plasmid pBR322. For genomic SELEX screening, these DNA fragments can be regenerated by PCR. The encoding sequence of each transcription factor (TF) was PCR-amplified and inserted into protein-expression plasmid pET21a. His-tagged TF was over-expressed in transformants, and affinity-purified. Formation of His-tagged TF and DNA segments carrying its recognition target sequence was formed and isolated by using Ni-NTA affinity column [20]. DNA was isolated from TF-DNA complexes and the mapping of TF-bound DNA can be performed by either SELEX-clos or SELEX-chip method as described in text

1. Mix 10 μL of 10× Binding buffer, 5 pmol DNA fragment mixture and 10 pmol His-tagged TF in a final volume of 0.1 mL.
2. Incubate DNA-TF mixture for 30 min at 37 °C.
3. Load the mixture onto a Ni-NTA agarose column [Note: Ni-NTA column should be prepared as follows: put 0.4 mL Ni-NTA suspension into a column (5 × 50 mm); wash the column three times with 0.2 mL 5 mM imidazole]
4. Wash 5 times with 0.2 mL of 10 mM Imidazole buffer.
5. Elute DNA-TF complexes with 1 mL of 200 mM imidazole buffer.
6. Take an aliquot of 0.01 mL DNA-TF complex fraction and subject to PCR by adding (in a total volume of 0.1 mL): 1 μL of F-primer (100 pmol/μL), 1 μL of R-primer (100 pmol/μL), 10 μL of 10 × Ex Taq buffer, 8 μL of dNTP mixture, 0.5 μL of Ex Taq polymerase, and 69.5 μL of redistilled H_2O.
7. Perform PCR under the conditions: 94 °C-3 min for complete denaturation followed by [94 °C-30 s > 50 °C-1 min > 70 °C-1 min] cycles. To find the best PCR cycle to give the optimum level of DNA amplification, interrupt PCR at cycles 16, 18 and 20 and check the level of DNA amplification as follows.
8. For the detection of PCR products, run a 5 μL aliquot together with 1 μL of 6 × loading dye on a 5% PAGE at a constant voltage of 150 V for 40 min alongside a 100 bp DNA-ladder marker.
9. After staining with ethidium bromide, elute DNA bands of 200 to 400 bp in length, and after ethanol precipitation, store DNA samples for mapping using SELEX-clos or SELEX-chip (*see* below).

7 Mapping of SELEX Library

To identifiy the DNA segments isolated by SELEX and contain a TF-binding sequence two approaches can be used: SELEX-clos (cloning and sequencing) (*see* Fig. 3) and SELEX-chip (tiling chip analysis) (*see* Fig. 4).

7.1 SELEX-clos (Cloning and Sequence) System

1. Mix 5 pmol of SELEX DNA fragment mixtures, and 100 ng (about 0.5 pmol) of TA-cloning vector (pT7 Blue-T vector, Novagen) in a total volume of 5 μL. [Note: the ratio of insert DNA to vector DNA should be approximately 10:1].
2. Add 5 μL of DNA Ligation Kit.
3. Incubate at 16 °C for more than 3 h.

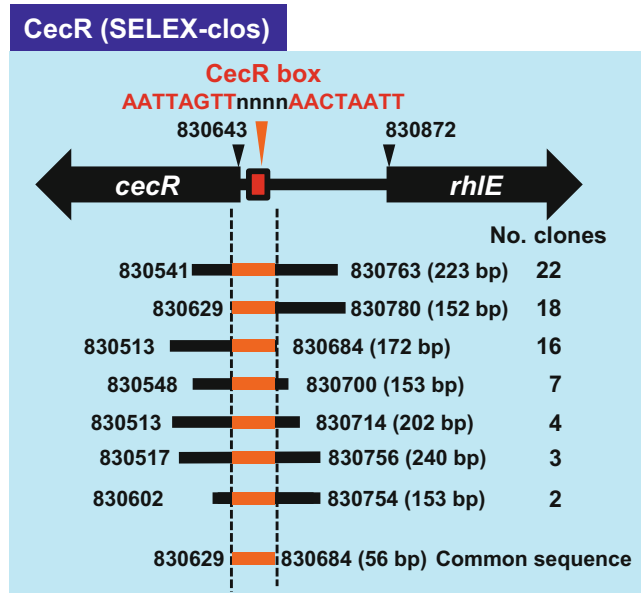


Fig. 3 SELEX-clos analysis for mapping TF-binding sequences: Representative result. SELEX-clos is one of the methods for mapping DNA segments isolated by genomic SELEX screening. SELEX DNA segments were cloned into a multi-copy plasmid, and the recombinant plasmids were isolated from transformed *E. coli* cells. DNA sequences were determined for inserts from as many as recombinant plasmids. By alignment the sequences, the binding sequence of test TF can be predicted. The number of same sequences correlates with the affinity of test TF to each DNA segment. The data shown represent the SELEX-clos analysis for CecR (regulator of cefoperazone and chloramphenicol sensitivity) [29]

4. Add 10 μL of ligated DNA solution to 0.1 mL *E. coli* competent cell suspension.
5. After incubation on ice for 10 min
6. Expose the cell suspension to heat shock at 42 °C for 45 s and then cool down on ice for 3 min.
7. Add 0.3 mL SOC medium.
8. Incubate at 37 °C for 30 min.
9. Centrifuge at 4 °C for 1 min at 9,100 × *g*, and suspend the cell pellet in 0.1 mL of LB medium.
10. Spread 0.1 mL each of these transformant solutions on to LB-ampicillin-X gal plate using cell spreader.
11. Check the number of white colonies [Note: Insertion of DNA inside the *lacZ* gene results in forming while colonies. The number of white colonies should be more than 100 on a single plate].
12. Pick up a single white colony, inoculate into 1–5 mL LB medium containing ampicillin, and incubate at 37 °C

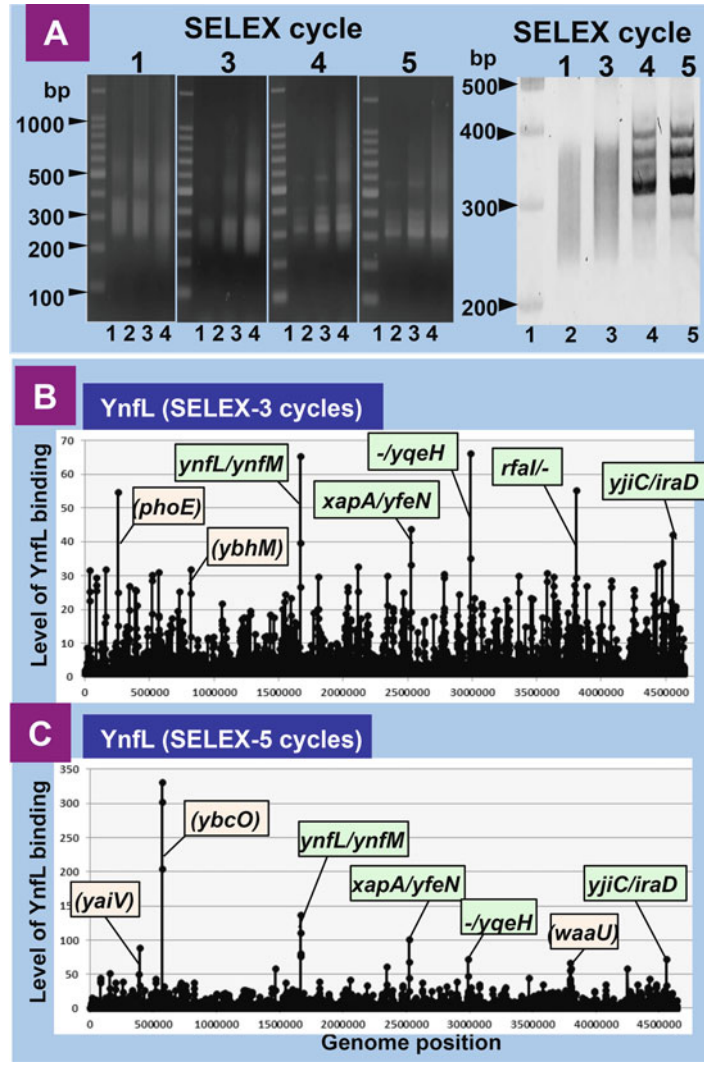


Fig. 4 SELEX-chip analysis for mapping TF-binding sequences: Representative result. SELEX-chip is one of the methods for mapping DNA segments isolated by genomic SELEX screening. SELEX DNA segments were Cy3-labeled and, together with Cy5-labeled, subjected to DNA-chip analysis [19, 20]. Mapping regulatory targets of uncharacterized LysR-type YnfL is shown as an example of SELEX-chip analysis. (**a**) SELEX samples at SELEX-cycle 3, 4 and 5 were analyzed by agarose gel electrophoresis (left) and PAGE (right). Enrichment of specific DNA segments with binding affinity to YnfL is observed. (**b**) SELEX-chip pattern of SELEX-cycle 3. (**c**) SELEX-chip pattern of SELEX-cycle 5. The binding sites of YnfL are classified as shown in Fig. 5

overnight [Note: for accurate mapping, more than 100 independent colonies should be examined].

13. Harvest cells by centrifugation and purify plasmid DNA by QIAprep Spin Miniprep kit protocol.
14. For PCR amplification of DNA segments inserted into plasmids, approximately 100 ng plasmid DNA is mixed with 3.75 μL of 5× Sequencing buffer, 1 μL 2T7P primer (3.2 pmol/μL), 0.5 μL BigDye Terminator v3.1 and nuclease-free doubly distilled water to make a total volume of 20 μL.
15. Run the PCR reaction under the conditions: 95 °C-20 s > 50 °C-15 s > 60 °C-1 min (25 cycles).
16. After the PCR reaction, purify PCR products by using the QIAGEN PCR purification kit.
17. Sequence the purified PCR products and map the sequence on the genome.
    Note that: the combined size of substrate DNA segments used for SELEX is approximately 5~6 fold longer than that of *E. coli* genome (4,650 Kbp), and thus a number of different DNA segments can be isolated by SELEX, each binding to a specific test TF and carrying the common TF-binding sequence (*see* Fig. 3).

7.2 SELEX-Chip System

For a faster identification of DNA segment isolated by SELEX they can be analyzed by using a tilling array of *E. coli* K-12 genome. SELEX-isolated DNA segments are labeled with Cy3, while the original DNA library is labeled with Cy5. After simultaneous hybridization of two fluorescent labeled DNA samples onto a single tiling array, the genome DNA fragment that formed complexes with test TF can be easily identified by measuring the Cy3/Cy5 ratio (*see* Fig. 4).

1. Prepare both SELEX DNA samples and the original library DNA at concentrations of approximately 20 ng/μL.
2. Mix 20 μL each of these DNA (about 400 μg each) and 2.5× Random primer (BioPrime kit).
3. Incubate at 94 °C for 3 min for DNA denaturation.
4. Add 5 μL dNTP mix (Bioprime kit), 3.75 μL Cy3-dCTP (or Cy5-dCTP) and 1 μL Klenow (Bioprime kit).
5. Incubate the mixture at 37 °C for 6 h.
6. Purify DNA by using the PCR amplification kit and dissolve purified DNA in nuclease-free water.
7. Mix 21 μL Cy3 labelled DNA, 21 μL Cy5 labelled DNA, 5.5 μL of 1 mg/mL Human Cot-1 DNA, 12 μL of 10× Blocking Agent and 59.5 μL of 2× HiRPM hybridization

buffer. [Note: Prepare the 10× Blocking Agent by adding 1.35 mL water to the 10× Blocking Agent tube. After standing at room temperature for 1 h, resuspend it by vortexing and store at −20 °C].

8. Incubate the mixture at 94 °C for 3 min for DNA denaturation, and then incubate at 37 °C for 30 min.
9. Pipette 110 μL each of the first hybridization mix onto the top chamber of SureHyb gasket slide (Backing plate) [Note: the gasket slide contains four chambers to be used simultaneously for four sets of the sample].
10. Place a 4×44 universal array onto each chamber of the gasket slide so as to make contact with the hybridization mix.
11. Place the clamp assembly (Hybridization chamber) on the slide and tighten the thumbscrew. [Note: Remove air bubbles on top of slides]
12. For hybridization of DNA samples with the tilling array, treat the gasket at 65 °C for 24 h by rotating at 20 rpm with use of a hybridization oven. [Note: the *E. coli* K12 4×44 K tilling array (OGT technology) contains a total of probes, each 60 base-long probe aligned at 105 bp interval in the order of *E. coli* genome sequence].
13. Scan of the tilling array as follows: (1) Place wash buffer-1 into the glass container-1 and disassembly glass dishes; (2) place a flea and slide rack in the container-1; (3) take the jig from hybridization oven, remove slide/gasket sandwich, place them in disassemble bath; (4) put the gasket slide from the array under the surface of the buffer; (5) transfer the slide quickly to rack in container-1 bath; (6) wash it at room temperature for 5 min with moderate stirring; (7) place pre-warmed wash buffer-2 into the glass container-2; (8) remove the rack place quickly into glass container-2; (9) wash it at 37 °C for 1 min with moderate stirring; and (10) remove the rack slowly [Note: Wash buffers and glass containers should be pre-warmed at 37 °C overnight. All glass containers, forceps, fleas and slide rack should be cleaned up with use of deionized distilled water].
14. Scan the slide using a high-resolution microarray scanner. Analyze the scan image by Agilent feature extraction software. The grid file can be downloaded from the array supplier (OGT technology).
15. Normalize the fluorescent intensity of the Cy3-labeled SELEX sample for each probe with respect to that of the Cy5-labeled original library sample. Plot the Cy3/Cy5 ratio versus genome position. [Note: approximately 300 bp-long SELEX fragments should bind to two or more consecutive 60 bp-long probes

aligned at 105 bp intervals. This criterion should be employed to avoid the background noise of nonspecific binding of test TF to non-target DNA sites. Usually the cut-off level of 4 is employed to identify specific binding.]

8 Prediction and Confirmation of the Regulatory Targets of TFs

1. Predict the regulatory target genes of test TF based on its sites along the map of *E. coli* genome (*see* Fig. 4).
2. Classify the type of TF-binding: type-A binding site, inside spacer of bidirectional transcription unit; type-B binding site, upstream of one transcription unit but downstream of another transcription unit; type-3 binding site, inside spacers of both of transcription units; and type-4, inside open reading frames (*see* Fig. 5). [Note: The list of regulatory targets of *E. coli* TFs thus predicted based on SELEX patterns are assembled in TEC (Transcription Profile of *Escherichia coli*) database (www.shigen.nig.ac.jp/ecoli/tec/)] [19].
3. For confirmation of the prediction of regulatory target by a test TF, perform the following experiments:
   (a) In vitro assays: Gel shift assay of TF-target DNA complexes; direct observation of TF-DNA complexes by

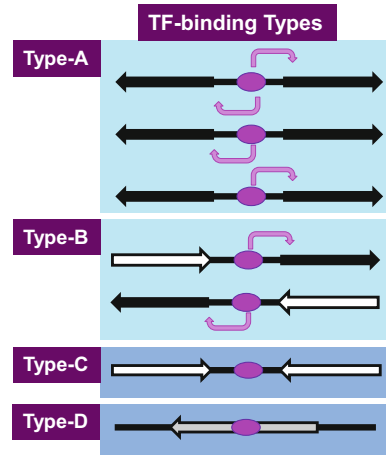


Fig. 5 Classification of TF-binding sites and the prediction of regulatory targets. The binding sites of TF on the genome can be classified into four types: type-A spacer between bidirectional transcription units; type-B, upstream of a transcription unit but downstream of another transcription unit; type-C, spacer downstream of two transcription units; and type-D, upon open reading frames. Number of the regulatory targets are estimated to be one or two for type-A and one for type-B

AFM; DNase-I footprinting assay of TF-binding sites; in vitro transcription of target DNA in the presence of TFs

(b) In vivo assays: transcriptome analysis in the presence and absence of TFs; ChIP-chip or ChIP-seq analyses; reporter assay of the promoter activity; Northern blot analysis of target RNAs; pRT-PCR assay of target RNAs; S1-nuclease assay of target RNAs; primer extension assay of target RNAs.

(c) In silico analysis: Search of the consensus sequence of TF binding among regulatory targets.

9 Notes

1. Limitation of transcription analysis in vivo

   Identification in vivo of the full set of regulatory targets for two classes of regulatory proteins, the σ factors and TFs, in *E. coli* is impossible because (1) regulatory proteins are not at all times present in *E. coli* cells, (2) some regulatory proteins are not always functional, but their activities are controlled by post-translational modification such as phosphorylation and acetylation, or through interaction of effector ligands of small molecules, (3) some regulatory proteins function as heteromers with other proteins and (4) regulatory proteins compete for binding to the RNAP core enzyme (in the case of σ subunit) or to overlapping DNA sites (in the case of DNA-binding TFs).

2. Problems associated with the published databases

   Transcription-related data in databases such as RegulonDB [http://regulondb/ccg.unam.mx] are obtained using different *E. coli* strains grown under various conditions. Recent findings of the difference in genome sequences between varieties of *E. coli* strains indicate the presence of huge amounts of difference in genetic background between *E. coli* strains [26, 27], including the difference in the set of genes for regulatory proteins of transcription. Care should be taken to use the data bases for modeling the genome regulation [19].

3. Usefulness of Genomic SELEX and PS-TF systems in vitro

   To overcome the problems associated with in vivo analysis of transcription, we have developed two in vitro approaches: Genomic SELEX screening of regulatory target promoters, genes and operons by a specified regulatory protein [19, 20] and PS-TF screening for search of TFs involved in regulation of a specified promoter [19, 24]. The SELEX system has been applied for the identification of the regulatory targets of all seven σ factors [21, 22]. In the case of TF analysis, the SELEX screening system is, in particular, useful for identification of

Identification of Regulatory Functions of Uncharacterized TFs by SELEX	
YagI => XynR	Regulator of xylonate catabolism
YbaO => DecR	Regulator of cysteine detoxification
YbiH => CecR	Regulator of cefoperazone and chloramphenicol sensitivity
YbjK => RcdA	Regulator of *csgD* (master regulator of biofilm formation)
YcdC => RutR	Regulator of pyrimidine metabolism and purine degradation
YcjZ => PgrR	Regulator of peptidoglycan (PG) recycling
YdeO => PhhR	Regulator of intracellular pH homeostasis
YdcN => SutR	Regulator of sulfur utilization
YdhM => NemR	Regulator of N-ethylmaleimide reductase
YeaM => NimR	Regulator of resistance to 2-nitroimidazole
YedW => HprR	HprSR TCS response regulator of H_2O_2 sensitivity
YgiP => Dan	Nucleoid-associated regulator for anaerobic growth

Fig. 6 Use of genomic SELEX for identification of regulatory functions of uncharacterized TFs. The genomic SELEX is most useful for short-cut identification of regulatory targets of uncharacterized TFs. So far we have identified the regulatory functions for more than hitherto uncharacterized TFs

regulatory targets of hitherto uncharacterized TFs [19], of which the regulatory targets have been identified for more than ten species using the SELEX system (*see* Fig. 6).

4. Limitation and improvement of transcription analysis in vitro
For successful application of these in vitro systems, however, it is essential to prepare the regulatory proteins in functional forms. At present, however, the functional forms have not determined for a number of TFs. A systematic search for effectors affecting TF activities is in progress in our research team, including screening in vitro of natural and synthetic chemical compounds affecting the DNA-binding activity of TFs [28]. Phenotype microarray (PM) system is one selection for screening in vivo of effectors affecting the functions of TFs [29, 30].

References

1. Ishihama A (2000) Functional modulation of *Escherichia coli* RNA polymerase. Annu Rev Microbiol 54:499–518
2. Ishihama A (2010) Prokaryotic genome regulation: multi-factor promoters, multi-target regulators and hierarchic networks. FEMS Microbiol Rev 34:628–645
3. Ishihama A (2012) Prokaryotic genome regulation: a revolutionary paradigm. Proc Jpn Acad Ser B Phys Biol Sci 88:485–508
4. Helmann J, Chamberlin M (1988) Structure and function of bacterial sigma factors. Annu Rev Biochem 57:839–872
5. Gruber TM, Gross CA (2003) Multiple sigma subunits and the participating of bacterial transcription space. Annu Rev Microbiol 57:441–466
6. Gourse RL, Ross W, Rutherford ST (2006) General pathway for turning on promoters transcribed by RNA polymerase containing alternative sigma subunits. Mol Microbiol 63:1296–1306
7. Ishihama A (1992) Role of the RNA polymerase alpha subunit in transcription activation. Mol Microbiol 6:3283–3288

8. Ishihama A (1993) Protein-protein communication within the transcription apparatus. J Bacteriol 175:2483–2489
9. Busby S, Ebright RH (1999) Transcription activation by catabolite activator protein. J Mol Biol 293:199–213
10. Selinger DW, Cheung KJ, Mei R et al (2000) RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. Nat Biotechnol 18:1262–1268
11. Wei Y, Lee JM, Richmond C et al (2001) High-density microarray-mediated gene expression profiling of *Escherichia coli*. J Bacteriol 183:545–556
12. Hatfield GW, Hung S, Baldi P (2003) Differential analysis of DNA microarray gene exprerssion data. Mol Microbiol 47:871–877
13. Cho BK, Zengler K, Qiu Y et al (2009) The transcription unit architecture of the *Escherichia coli* genome. Nat Biotechnol 27:1043–1049
14. Conway T, Creecy JP, Maddox SM et al (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. MBio 5:e01442-14
15. Bulyk ML (2006) DNA microarray technologies for measuring protein-DNA interactions. Curr Opin Biotechnol 17:422–430
16. Wade JT, Struhl K, Busby SJW et al (2007) Genomic analysis of protein–DNA interactions in bacteria: insights into transcription and chromosome organization. Mol Microbiol 65:21–26
17. Keseler IM, Collado-Vides J, Santos-Zavaleta A et al (2011) EcoCyc: a compehensive database of *Escherichia coli* biology. Nucleic Acids Res 39:D583–D590
18. Gama-Castro S, Salgado H, Santos-Zavaleta A et al (2016) RegulonDB v9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res 44: D133–D143
19. Ishihama A, Shimada T, Yamazaki Y (2016) Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. Nucleic Acids Res 44:2058–2074
20. Shimada T, Fujita N, Maeda M et al (2005) Systematic search for the Cra-binding promoters using genomic SELEX systems. Genes Cells 10:907–918
21. Shimada T, Yamazaki Y, Tanaka K et al (2014) The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of *Escherichia coli*. PLoS One 9:e90447
22. Shimada T, Tanaka K, Ishihama A (2017) The whole set of the constitutive promoters recognized by four minor sigma subunits of *Escherichia coli* RNA polymerase. PLoS One 12: e0179181
23. Jishage M, Ishihama A (1997) Variation in RNA polymerase sigma subunit composition within different stocks of *Escherichia coli* W3110. J Bacteriol 179:959–963
24. Shimada K, Ogasawara H, Yamada K et al (2013) Screening of promoter-specific transcription factors: multiple regulators for the sdiA gene involved in cell division and quorum sensing. Microbiology 159:2501–2512
25. Yamamoto H, Hirao K, Oshima T et al (2005) Functional characterization *in vitro* of all two-component signal transduction systems from *Escherichia coli*. J Biol Chem 280:1448–1456
26. Land M, Hauser L, Jun S-R et al (2015) Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 15:141–161
27. Dunne KA, Chaudhuri RR, Rossiter AE (2017) Sequencing a piece of history: complete genome sequence of the original *Escherichia coli* strain. Microbial Genomics 3. https://doi.org/10.1099/mgen.0.000106
28. Shimada T, Shimada K, Matsui M et al (2014) Roles of cell division control factor SdiA: recognition of quorum sensing signals and modulation of transcription regulation targets. Genes Cells 19:405–418
29. Yamanaka Y, Shimada T, Yamamoto K et al (2016) Transcription factor CecR (YbiH) regulates a set of genes affecting the sensitivity of *Escherichia coli* against cefoperazone and chloramphenicol. Microbiology 162:1253–1264
30. Ogasawara H, Ohe S, Ishihama A (2015) Role of transcription factor NimR (YeaM) in sensitivity control of *Escherichia coli* to 2-nitroimidazole. FEMS Microbiol Lett 362:1–8

Chapter 5

Modular Assembly of Synthetic Secondary Chromosomes

Celine Zumkeller, Daniel Schindler, and Torsten Waldminghaus

Abstract

The development of novel DNA assembly methods in recent years has paved the way for the construction of synthetic replicons to be used for basic research and biotechnological applications. Questions of how chromosomes need to be constructed to maintain the genetic information can now be answered by a learning-by-building approach. Here, we describe an efficient pipeline for the design and assembly of synthetic, secondary chromosomes in *Escherichia coli* based on the popular Modular Cloning system (MoClo).

Key words Synthetic chromosomes, *Escherichia coli*, Modular cloning, synVicII, Genome engineering, DNA libraries, Oligo design

1 Introduction

Chromosomes harbor the genetic information in cells of all domains of life. This storage of genetic information implies more than simply getting genes into a linear order. The genetic information needs to be copied before cell division, it needs to be distributed into daughter cells and errors and damage needs to be repaired. All these processes have underlying molecular mechanisms that are a central part of chromosomes to function as such. The question of what actually makes a chromosome a chromosome has fascinated researchers for decades, but is now asked from a new perspective with synthetic biology becoming a scientific discipline. It is now possible to actually construct entire bacterial or eukaryotic chromosomes from synthetic DNA oligonucleotides [1–3]. This technical revolution raises the question what engineering rules one has to follow when constructing a chromosome from scratch. Which genes need to be inserted and how should they be arranged, which replication origin(s) could be used and which functional DNA motifs are relevant to support chromosome maintenance [4–6]. The upcoming synthetic genomics makes it possible to try and answer chromosome biology questions with a learning-by-

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_5,

building approach. If one thinks that this or that part or factor is needed to make a functional chromosome one could go ahead and construct a synthetic chromosome to verify this hypothesis. Such an experimental chromosome construction is not entirely new since so called mini-chromosomes have been used for many years to better understand chromosome biology [7–9]. However, the efficient assembly of larger replicons provides unprecedented opportunities to answer basic research questions on mechanisms that maintain chromosomes. Many different DNA assembly methodologies have been developed over the last years [1–3, 10, 11]. The method of choice will largely depend on the specific replicon to be constructed. One important set of DNA assembly approaches omits the classical use of specific restriction digests and ligation and is instead based on short homologous regions on the ends of the DNA fragments to be assembled. Such homology can be used in vitro in combination with an exonuclease, a DNA polymerase, and a DNA ligase to assemble multiple DNA fragments in the so-called Gibson assembly [1]. The same fragments with homology can alternatively be transformed into the yeast *Saccharomyces cerevisiae* to make use of its powerful recombination system to join the respective DNA parts [12]. Assembly in yeast is popular in synthetic chromosome construction may it be bacterial chromosomes or actually synthetic yeast chromosomes [2, 13, 14]. Another DNA assembly approach making use of short sequence homologies is the ligase cycling reaction (LCR). Here, the homology is brought in by short bridging oligonucleotides [15]. Besides allowing cloning of multiple DNA fragments within a single reaction, the described methods are able to construct larger assemblies without so-called scars. Such scars occur if cloning is based on the use of conventional restriction enzymes because here specific sequences need to be inserted that will remain after assembly. Alternative assembly methods make the actual assembly reaction and the reuse of fragments easier. One example is the Modular Cloning system (MoClo) developed by Marillonnet and coworkers [16, 17]. This method is based on the activity of type IIS restriction enzymes, which unlike conventional restriction enzymes, used for classical cloning, cleave DNA outside their recognition site. Correct positioning of the recognition sites on a DNA fragment and within a vector will lead to a loss of the respective recognition sites in case the desired assembly takes place (*see* Fig. 1a).

Such a reaction can therefore be carried out in one-pot with a combination of the type IIS restriction enzyme and the DNA ligase. Four base-pair overhangs mediate the fusion of the fragment and the vector. The MoClo system comprises a set of seven such receptor vectors with the second four-base overhang always matching the first of the following vector (*see* Fig. 1b). This design determines the order of the fragments to be assembled. Importantly, such an assembly of multiple fragments is performed with whole

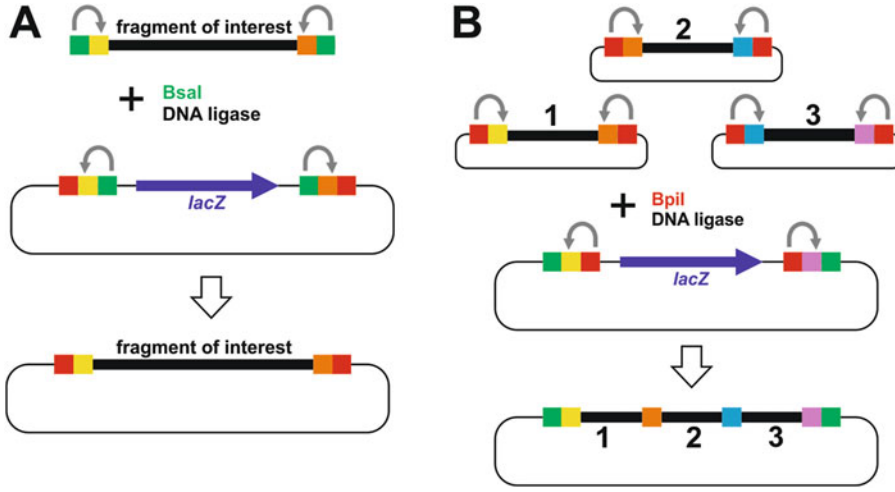


Fig. 1 The working principles of the MoClo system. (**a**) Directional cutting with the type IIS restriction enzyme BsaI leads to loss of the recognition sites on the fragment and the backbone following subsequent *lacZ*-gene release. The reaction mix contains both the restriction enzyme and the DNA ligase because the correct fusions will not be cleaved anymore. Fragments can be released by cutting with a second type IIS endonuclease (here BpiI) allowing the assembly of multiple fragments. (**b**) Matching overhang sequences determine the order of fragments (color-coded rectangles). Mixing of respective plasmids with a suitable acceptor vector results in efficient assembly of multiple fragments within the vector without the need for prior fragment isolation

plasmids carrying the fragments to be assembled and the respective receptor vector. No prior cleavage and purification of the fragments is required. In addition, the plasmids are designed for hierarchical assemblies with alternating restriction enzymes and antibiotic resistance markers allowing reuse of vectors by continuous switching.

A rise in popularity of the MoClo system has recently been observed, and diverse sets of specialized genetic parts for different organisms have been described [16, 18–21]. The modularity of the system allows reuse of parts and easy sharing between research groups. Here, we outline how the MoClo system can be used to construct synthetic secondary chromosomes. A clear-cut distinction between secondary chromosomes and plasmids may actually not be possible [22, 23]. One well-studied example of a secondary chromosome is found in *Vibrio cholerae* [24, 25]. The secondary replication origin has been shown to replicate a mini-chromosome in *E. coli* [25, 26]. Based on this work we have developed the core replicon synVicII and characterized its replication in the heterologous host *E. coli* [22, 27] (*see* Fig. 2). The most recent version is compatible with the MoClo system introduced above. This protocol outlines the entire assembly of synthetic secondary chromosomes including the final assembly with the core synVicII. However, the protocol could also be adjusted to combine the assembled sequence with any other plasmid or chromosomal replication origin. The only requirement would be that a respective core replicon is compatible with the MoClo system (*see* **Note 1**).

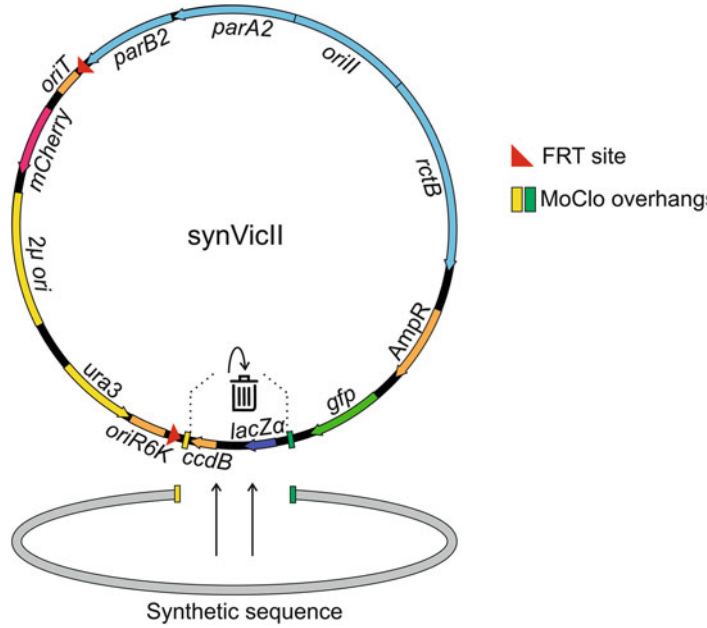


Fig. 2 Schematic overview of synVicII and its most relevant features. SynVicII carries the origin of replication of the *Vibrio cholerae* chromosome II (*oriII*), genes encoding the respective replication initiator *rctB* and the native *parAB2* segregation system (all light blue). Selection for the core replicon is maintained by the bla gene (AmpR) (orange), mediating ampicillin resistance. GFP (green) might be used for the analysis of the replicon, e.g., regarding stability [22]. Remaining features are required for different cloning procedures such as yeast recombination (*2 µ ori*, *ura3*) or conjugational transfer (*oriT*). MoClo assembly of the final chromosome building block and the core replicon synVicII results in the loss of the MoClo cloning cassette and allows easy selection for correct clones. Cloning features enclosed by FRT sites (red triangles) can be removed by a flippase-dependent, site-directed recombination reaction after final assembly

2 Materials

1. Standard reagents and instrumentation for PCR reactions
   (a) DNA polymerase and corresponding reaction buffer.
   (b) Dinucleoside triphosphate set: dATP, dCTP, dGTP, dTTP (10 mM).
   (c) PCR tubes.
   (d) PCR Thermal Cycler.
2. Standard reagents and equipment for gel electrophoresis
   (a) Electrophoresis chamber.
   (b) Gel casting tray and gel comb.
   (c) Power supply.

(d) Agarose for nucleic acid electrophoresis.
(e) DNA ladder.
(f) DNA gel stain.
(g) Gel documentation system.

3. Standard PCR purification kit.
4. Vortex mixer.
5. Mini-centrifuge.
6. Chemically competent *E. coli* cells.
7. LB medium: 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) sodium chloride.
8. 0.5 mL PCR tubes.
9. Plasmid purification kit.

2.1 Software

1. MARSeG (http://synmikro.com/research/chassis-and-genomes/torsten-waldminghaus/marseg/).

2.2 Plasmids

Please contact the authors to request the pMA vectors (*see* Fig. 5). The pICH vectors described in this article were built by Weber et al. and are available from Addgene [4].

2.3 DNA Oligonucleotides

1. Subheading 3.3, **step 1**: Custom designed forward and reverse oligonucleotides for amplification of sequences by PCR.
2. Subheading 3.3, **step 1**: Designed mixture of oligonucleotides for annealing reaction.
3. Subheading 3.3, **step 4**: lib_fw: 5′-TTTTAGGNNGGTCTC NGGAG-3′.
4. Subheading 3.3, **step 4**: lib_rv: 5′-TTAATCCNNGGTCTC NAGCG-3′.

2.4 Enzymes

1. Taq DNA Polymerase (5 U/μL).
2. Phusion DNA Polymerase (2 U/μL).
3. BsaI (10 U/μL).
4. T4 DNA-Ligase (5 U/μL) and T4 DNA-Ligase (HC) (20 U/μL).
5. BpiI (10 U/μL).
6. RNase A (10 mg/mL).

2.5 Antibiotics and Supplements

1. Ampicillin in EtOH (100 μg/mL).
2. Spectinomycin in H_2O (100 μg/mL).
3. Kanamycin in H_2O (35 μg/mL).
4. Chloramphenicol in EtOH (35 μg/mL).

5. Isopropyl β-D-1-thiogalactopyranoside (IPTG) in H_2O (100 μg/mL).
6. 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) in Dimethylformamide (DMF) (20 μg/mL).

2.6 Chemicals

1. TAE buffer: 40 mM Tris, 20 mM acetic acid, 1 mM EDTA.
2. Resuspension buffer: 50 mM Glucose, 10 mM EDTA, 10 mM Tris–HCl (pH 8.0).
3. Lysis Solution: 0.2 M NaOH, 1% SDS.
4. 7.5 M ammonium acetate.
5. Chloroform.
6. Precipitation solution: 30% PEG 8000, 1.5 M NaCl.
7. Sodium dodecyl sulfate (SDS).
8. DNA dye (e.g., Ethidium Bromide or SYBR Green).

3 Methods

3.1 Overall Design of Synthetic Chromosomes

The design of a synthetic chromosome will depend on the specific research question being asked or the practical application being desired. Replicon construction is based on a number of basic building blocks within a set of plasmids, which are stepwise assembled in a hierarchical scheme. Even though the chromosome assembly is highly modular, the overall principle is defined by the 4-base overhangs generated during the MoClo assembly rounds. Figure 3 gives an overview of the building blocks and overhangs that constitute the synthetic chromosome and serves as guideline for overall design considerations such as positioning of the sequences of interest. The desired position of a specific DNA sequence within the final replicon will be determined by the initial order of building blocks to be assembled.

The assembly shown in Fig. 3 is based on 125 basic building blocks, which can be extended or reduced based on individual design ideas (*see* **Note 2**).

1. Plan the overall design of your synthetic secondary chromosome and fill out a table analogous to the example shown in Fig. 4 to define your level-1 building blocks. Consider different aspects such as the nature of your building blocks (defined or variable sequences; synthetic or natural), the position of those sequences within the chromosome, the presence of scars between building blocks and possible utilization of the modularity conferred by the MoClo system (*see* **Note 3**).
2. Familiarize yourself with the consecutive MoClo cloning steps and the different vectors available for various requirements (*see* Fig. 5).

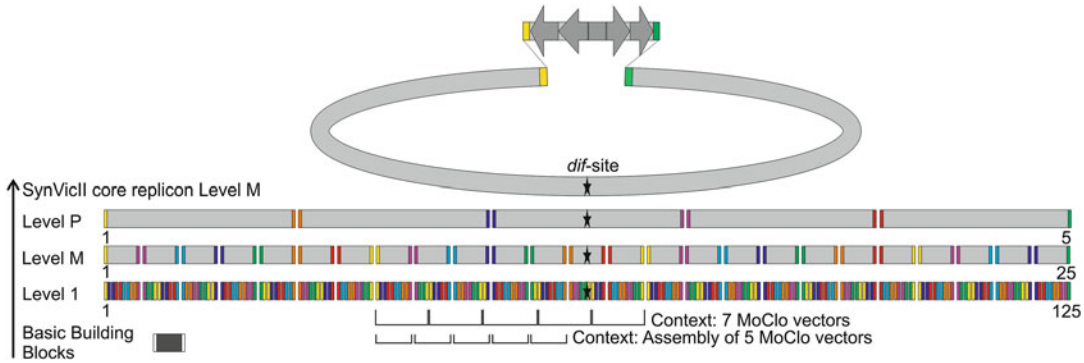


Fig. 3 Overall design principle of the synthetic secondary chromosomes based on the overhangs generated during consecutive MoClo rounds. Synthetic chromosome assemblies start with basic building blocks (dark gray, bottom panel) with universal overhangs (white rectangles) used for cloning into seven different level-1 vectors. Dependent on the level-1 vector, this process adds different four-base overhangs (colored rectangles, compare Fig. 5a). Identical overhangs (same color) are complementary and can be fused together. With each consecutive MoClo round (black arrow), five building blocks will constitute a new building block on a higher MoClo assembly level. Consequently, two different contexts exist within the MoClo assembly scheme, the 7 different MoClo vectors and the assembly of 5 of them at a time to generate the next building block. This alternating order is essential for one-pot MoClo assembly rounds. Finally, five level-P-building blocks are assembled with the synVicII core resulting in the synthetic secondary chromosome. Positioning of desired DNA sequences as level-1 building blocks will determine the position in the final construct, as shown here for the *dif*-site of *E. coli* (black star). Chromosome building blocks can be reused and used at different positions with some restrictions. An example is given on level M where every seventh building block has the same four-base overhangs (color-coded) and is therefore interchangeable

Synthetic chromosome assembly starts with level-1 MoClo vectors (*see* Fig. 5b), where basic building blocks are cloned into (*see* Fig. 5c) (*see* **Notes 4** and **5**).

(a) Decide whether to use the lacZ MoClo version or the ccdB version for standard cloning of basic building blocks (*see* **Note 6**).

(b) Preferably use the *ccdB* MoClo vectors for your library-based basic building blocks (*see* **Note 6**).

(c) Check if your basic building blocks should be assembled in a certain orientation. Choose your level-1 vector versions accordingly, to get either a forward (fw) or reverse (rv) orientation (*see* **Note 7**).

(d) For the assembly of building blocks from level M onward use the required end-linker to bridge the ends of the desired building block to the respective MoClo vector carrying the universal "H" overhang (*see* Fig. 5, black rectangle). The end-linkers are removed in subsequent assembly steps.

3. The assembly process continues with level-M MoClo vectors (*see* Fig. 5b) which confer a spectinomycin resistance to allow counterselection against level-1 vectors.

Level-P BB	Level P Vector & EL	Level-M BB	Level M Vector & EL	Level-1 BB	Level 1 Vector	Characteristics
Level P-1	A-H: pMA67	Level M-1	A-H: pMA60	Level 1-1	A-B: pMA53	Random Spacer
				Level 1-2	B-C: pMA54	Promoter A
				Level 1-3	C-D: pMA55	CDS A
				Level 1-4	D-E: pMA56	Terminator A
			EL_F-H: pICH50914	Level 1-5	E-F: pMA57	Random Spacer
		Level M-2	F-H: pMA65	Level 1-6	F-G: pMA58	Random Spacer
				Level 1-7	G-A: pMA59	Random Spacer
				Level 1-8	A-B: pMA53	Random Spacer
				Level 1-9	B-C: pMA54	Promoter B
			EL_D-H: pICH50892	Level 1-10	C-D: pMA55	CDS B
		Level M-3	D-H: pMA63	Level 1-11	D-E: pMA56	Terminator B
				Level 1-12	E-F: pMA57	Random Spacer
				Level 1-13	F-G: pMA58	Random Spacer
				Level 1-14	G-A: pMA59	Random Spacer
			EL_B-H: pICH50872	Level 1-15	A-B: pMA53	Random Spacer
		Level M-4	B-H: pMA61	Level 1-16	B-C: pMA54	Promoter C
				Level 1-17	C-D: pMA55	CDS C
				Level 1-18	D-E: pMA56	CDS D
				Level 1-19	E-F: pMA57	CDS E
			EL_G-H: pICH50927	Level 1-20	F-G: pMA58	Terminator C
		Level M-5	G-H:pMA66	Level 1-21	G-A: pMA59	Random Spacer
				Level 1-22	A-B: pMA53	Random Spacer
				Level 1-23	B-C: pMA54	Promoter F
				Level 1-24	C-D: pMA55	CDS F
	EL_E-H: pICH79289		EL_E-H: pICH50900	Level 1-25	D-E: pMA56	Terminator F
Level P-2	E-H: pMA71	Level M-6	E-H: pMA64	Level 1-26	E-F: pMA57	$LacO_4$-$TetO_4$
				Level 1-27	F-G: pMA58	$LacO_4$-$TetO_4$
				Level 1-28	G-A: pMA59	$LacO_4$-$TetO_4$
				Level 1-29	A-B: pMA53	$LacO_4$-$TetO_4$
			EL_C-H: pICH50881	Level 1-30	B-C: pMA54	$LacO_4$-$TetO_4$
		Level M-13	E-H: pMA64	Level 1-61	E-F: pMA57	Random Spacer
				Level 1-62	F-G: pMA58	Random Spacer
				Level 1-63	G-A: pMA59	dif site
				Level 1-64	A-B: pMA53	Random Spacer
			EL_C-H: pICH50881	Level 1-65	B-C: pMA54	Random Spacer
		Level M-25	B-H: pMA61	Level 1-121	B-C: pMA54	Random Spacer
				Level 1-122	C-D: pMA55	Random Spacer
				Level 1-123	D-E: pMA56	Random Spacer
				Level 1-124	E-F: pMA57	Random Spacer
	EL_G-H: pICH79300		EL_G-H: pICH50927	Level 1-125	F-G: pMA58	Random Spacer

Fig. 4 Exemplary extract of a list of the chromosome building blocks and MoClo vectors. The table names the different building blocks (BB) on level-1, M and P (column: Level-1 building block, Level-M building block and Level-P building block). It describes the characteristics of the level-1 building block (column: Characteristics) and shows which MoClo vectors and end-linkers (EL) must be used for consecutive MoClo rounds (Column: Level-1 Vector, Level-M vector & end-linker, Level-P Vector & end-linker). Note that building blocks are partly interchangeable. For example, vectors level1–9 and level1–23 could be exchanged in the assembly (or an alternative assembly) to switch the inserted promotors. Capital letters refer to the four-base overhangs according to Fig. 5

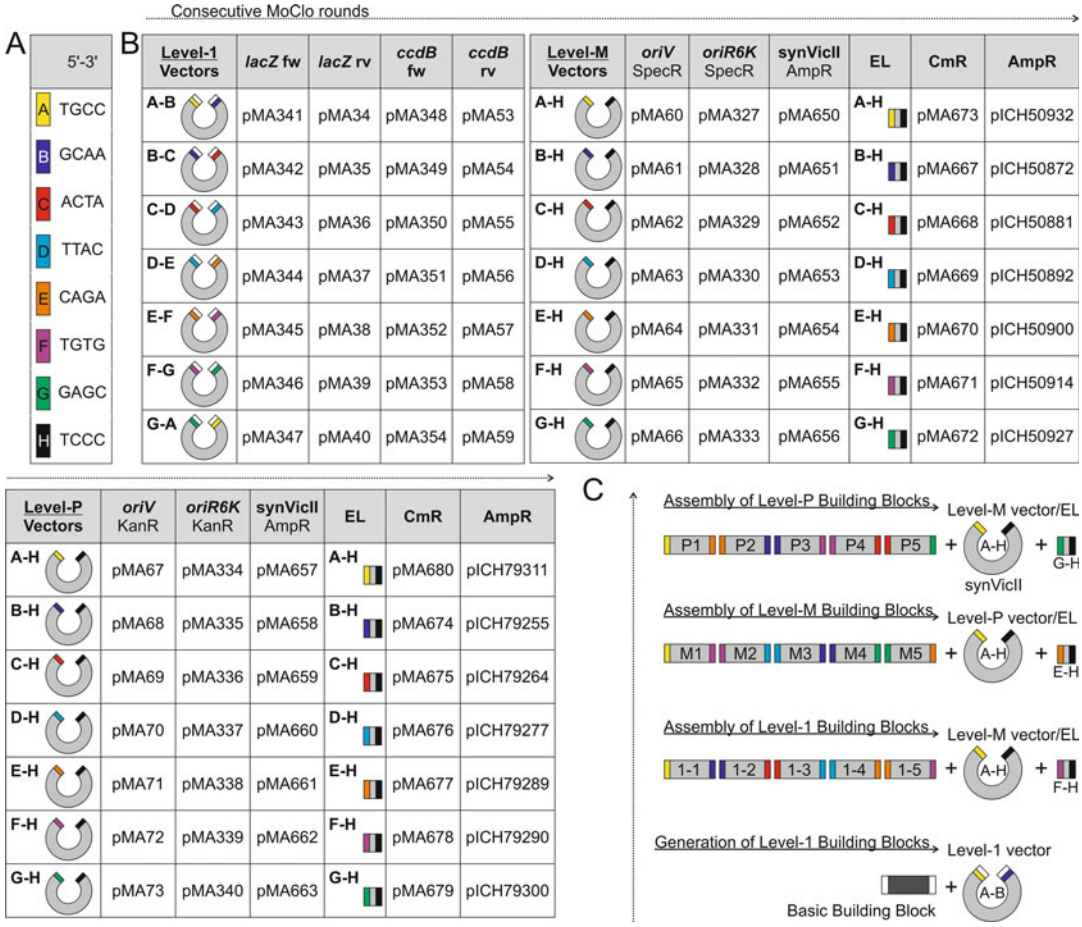


Fig. 5 (**a**) Overview of specific overhangs and their respective sequence. (**b**) Overview of MoClo vectors used for chromosome assembly. Tables show available MoClo vectors of level-1, level-M, and level-P, together with end-linkers (EL) required for level-M and level-P cloning. Overhangs are shown as colored rectangles, colors are equivalent to the ones used in Fig. 3. According to the present overhangs, vectors are designated alphabetically, e.g., "A-B" for the first level-1 vector. Alternative types of vectors are available for each MoClo level with the same overhangs as indicated. Vectors differ regarding their genetic features and hence confer a different application range (cf. Subheading 3.1). Vectors denoted as pMA plasmids can be requested from the authors, pICH vectors were created by Weber et al. [16] and are available on Addgene. (**c**) Exemplary assembly of level-1 building blocks and consecutive assembly steps. Building blocks and basic building blocks are shown as gray and dark gray rectangles respectively, containing the specific (colored rectangles) and universal level-1 overhangs (white rectangle). They are designated according to their position in the chromosome (cf. Fig. 3). Together with the correct MoClo vectors and end-linkers (where necessary), consecutive one-pot cloning reactions result in the final construct

(a) Use the *oriV* dependent vectors for routine assembly of building blocks.

(b) Consider using vectors dependent on the conditional replication origin *oriR6K* if integration of building

blocks into the primary chromosome is desired (*see* **Note 8**).

(c) Choose the MoClo-compatible synVicII for the final assembly of a synthetic secondary chromosome (*see* Subheading 1 and Fig. 2) [22, 27].

(d) On the next hierarchy level level-P vectors will be used. An encoded Kanamycin resistance enables counter-selection against level-M vectors. Except for the selection marker level-P vectors share all features of level-M vectors.

3.2 Sequence- and Oligo Design for the Generation of Basic Chromosome Building Blocks

The nature of the basic building blocks will determine how the first design and assembly steps will take place. Basic building blocks might be fully defined sequences of interest such as transcription units. However, a fraction of the building blocks for synthetic chromosomes might constitute random spacer or scaffolding sequences. The most efficient and cost-effective solution for the generation of such random sequences is working with DNA libraries generated from mixtures of oligonucleotides (*see* **Note 9**). Regardless of the building blocks' nature, it is essential to remove or avoid any BsaI and BpiI recognition sites, which would interfere with the MoClo system.

3.2.1 Primer Design for Fully-Defined Basic Building Blocks

One type of chromosome part will typically be some sort of defined sequence such as coding sequences or transcription units that can be amplified from a template DNA. We describe here how such sequences are made compatible to the MoClo system by PCR amplification with specific primers.

1. Identify the sequences of interest for amplification by PCR.
2. For sequences of interest without internal BsaI and BpiI recognition site, design Primer 1 and 4 as illustrated in Fig. 6. Extend the 25–30 bp homology region to include a BsaI recognition site and level-1 universal overhangs (black frame) as shown.

For sequences of interest with internal BsaI or BpiI recognition sites, design two more primers (Primers 2 and 3) that will introduce a single nucleotide exchange to remove the recognition site while maintaining the native function of the sequence (e.g., the amino acid sequence of a protein) (*see* Fig. 6) (*see* **Note 10**). Extend primers 2 and 3 for BsaI recognition sites that will allow BsaI-dependent cloning of the two fragments in the respective level-1 destination vector (Subheading 3.3) [16].

3.2.2 Design of DNA Oligonucleotides for Synthetic Sequence Libraries

The basic principle of generating a library of synthetic sequences used here is the annealing of two DNA oligonucleotides based on pairing of a homologous sequence region at their respective 3′ and 5′ ends (*see* Fig. 7). Elongation of such annealed oligonucleotides will then result in a mixture of DNA sequences based on the diversity

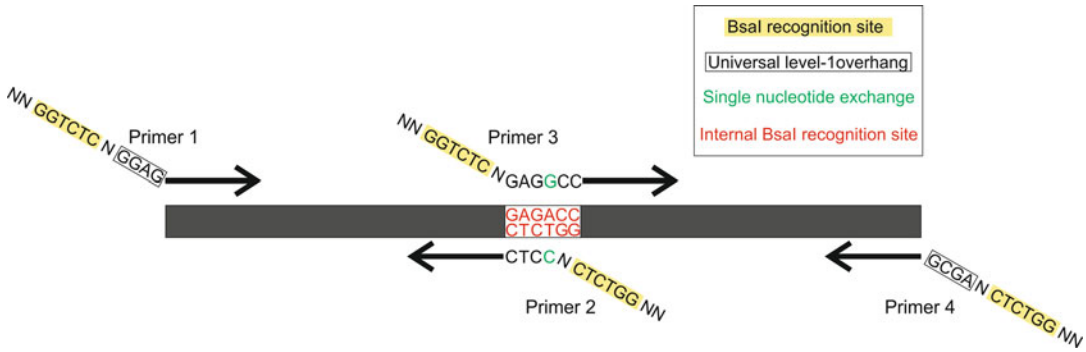


Fig. 6 Strategy for the generation of specified building blocks with modification of internal BsaI or BpiI recognition sites. The coding gene X (dark gray) is carrying an internal BpiI recognition site (red letters). Primers 1–4 consist of 25–30 bp homology (black arrow) to allow annealing to the coding gene X. Additionally, Primers 1 and 4 carry the recognition site (highlighted in yellow) and universal overhangs required for the generation of level-1 building blocks (black frame). Primers 2 and 3 anneal on the internal BsaI recognition site, thereby introducing a single nucleotide exchange (green letters). They also contain a BsaI recognition site, to allow a single-pot MoClo reaction of the two resulting building blocks into a level-1 MoClo vector, which then constitutes a level-1 building block free of internal restriction sites [16]

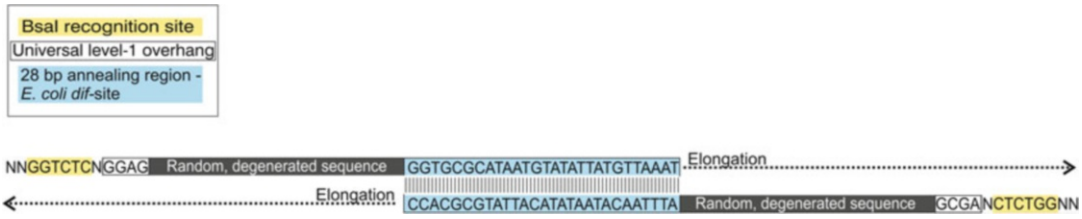


Fig. 7 Use of oligonucleotides for the creation of basic building block libraries by annealing and elongation. Annealing will take place at the 28 bp-long annealing region (highlighted in light blue) (Here, the *dif*-site of *E. coli* was used as an example. The oligonucleotides also contain BsaI recognition sites (highlighted in yellow) and level-1-specific overhangs (black frame) for subsequent MoClo rounds. Elongation at the 3′ end of the annealed oligonucleotides will give the library of basic building blocks

determined by the design of the oligonucleotides. When working with DNA libraries it is necessary to avoid certain DNA sequences with biological function such as promoters, replication termination sites, or restriction sites that would otherwise affect function or qualitative analysis of the synthetic chromosomes. The computer program MARSeG (Motif Avoiding Randomized Sequence Generator) fulfils this requirement and generates degenerated sequences with a high degree of diversity while excluding a list of DNA sequences provided by the user (*see* Fig. 8) (*see* **Note 11**) [21].

1. Download MARSeG from the website http://synmikro.com/research/chassis-and-genomes/torsten-waldminghaus/marseg/ and follow instructions given in the detailed manual on the same web page to obtain your custom, degenerated sequence.

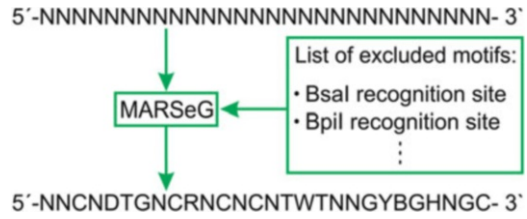


Fig. 8 Operating principle of MARSeg: The program requires an input sequence (with or without fully defined sequence regions) and a list of DNA motifs to be excluded. It then returns a degenerated sequence that excludes the respective DNA motifs

2. Extend the 5′ or 3′ end of the MARSeG-designed oligonucleotides to include a BsaI recognition site and level-1 universal overhangs according to Fig. 7.
3. Optional: If not previously included in MARSeG (as prefix or suffix), add homologous sequences to the 3′ or 5′ end of the oligonucleotides as shown in Fig. 7 to allow annealing of oligonucleotides (*see* **Notes 12** and **13**).
4. Order the designed oligonucleotides from your manufacturer of choice (*see* **Note 14**).

3.3 Generation of Level-1 Building Blocks

Based on the two different DNA oligonucleotide designs described above (Subheading 3.2) the generation of level-1 building blocks begins with two distinct procedures. For one, DNA libraries are annealed, elongated, enriched, and extracted (**steps 1–8**). For another, defined sequences of interest are amplified by PCR (**steps 10–12**). These DNA fragments are then treated similarly in subsequent **steps 13–15**.

1. For DNA library construction based on annealing of synthetic DNA oligonucleotides prepare the following 50 μL reaction on ice (*see* **Notes 15** and **16**):

   0.5 μL Taq-Polymerase.
   5 μL 10× reaction buffer (delivered with corresponding enzyme).
   0.5 μL dNTPS (10 mM).
   2.5 μL oligonucleotide 1 (10 μM).
   2.5 μL oligonucleotide 2 (10 μM).
   39 μL ddH_2O.
2. Mix and spin down.
3. Use the following cycling conditions to anneal and extend oligonucleotides:

Cycle Step	Temperature (°C)	Time (min:s)	Cycles
Initial denaturation	95	3:00	1 ×
Denaturation Annealing Elongation	95 50–65 (*see* **Note 16**) 72	3:00 0:20 Template length in kb × 1 min	5 ×
Final elongation	72	10:00	1 ×
Hold	10	∞	1 ×

4. Purify the double-stranded DNA products (*see* **Note 17**) with a standard PCR purification kit (elute two times with 100 µL H_2O each) and amplify the library by standard PCR with primers 5′-TTTTAGGNNGGTCTCNGGAG-3′ (lib_fw) and 5-′-TTAATCCNNGGTCTCNAGCG-3′ (lib_rev) (*see* **Note 18**).
5. To do this prepare the following 50 µL reaction on ice:

   1 µL Phusion Polymerase.
   10 µL 5× polymerase reaction buffer (delivered with corresponding polymerase).

   1 µL dNTPS (10 mM).
   1 µL Template DNA.
   1.25 µL lib_fw (10 µM).
   1.25 µL lib_rev (10 µM).
   34.5 µL ddH_2O.
6. Mix and spin down reaction.
7. Use the following PCR cycling conditions:

Cycle Step	Temperature (°C)	Time (min:secs)	Cycles
Initial denaturation	95	3:00	1 ×
Denaturation Annealing Elongation	95 55 72	0:30 0:20 Template length in kb × 1 min	35 ×
Hold	10	∞	1 ×

8. Run the whole PCR product on an agarose gel and extract the band of the expected size with a standard gel extraction kit (elute two times with 25 µL H_2O each) to obtain the DNA library (*see* **Note 19**).

9. Continue with **step 13** using the DNA libraries you produced.
10. For all chromosome parts with defined sequences that can be amplified by PCR use primers 1 and 4 (*see* Fig. 6) for full amplification. For sequences with internal BsaI or BpiI sites amplify the sequence in parts (*see* Fig. 6, Primers 1 and 2; Primers 3 and 4) by PCR. Prepare the PCR reaction mix as mentioned in **steps 4** and **5** with primers described in Subsection 3.2.1. Use the PCR cycling conditions given under point 7.
11. Run 3 μL of the PCR product on a gel to confirm successful amplification.
12. Purify the PCR products with a standard purification kit.

    In the second stage of building, extracted DNA library fragments and PCR products (= basic building blocks) are cloned into the appropriate MoClo vector to convey position-dependent overhangs for the level-1 MoClo assembly step (cf. Subheading 3.1 and Fig. 5).
13. Perform a BsaI-dependent MoClo reaction using the basic building blocks as inserts with the appropriate level-1 MoClo vectors (E.G., pMA53–59) [16]. To this end, prepare the following 25 μL reaction.

    1 μL T4 DNA-Ligase (5 U/μL) (*see* **Note 20**).

    2.5 μL 10 × T4 DNA-ligase buffer (delivered with corresponding enzyme).

    40 fmol level-1 vector (*see* **Note 21**).

    40 fmol of each basic building block.

    1 μL BsaI.

    Add up to 25 μL with ddH_2O.
14. Run the MoClo reaction using the following conditions:

Temperature (°C)	Time (min)
37	300
50	20
80	10
10	∞

15. Transform an *Escherichia coli* cloning strain with the MoClo reactions and plate 50 μL of cell culture on ampicillin-LB plates containing IPTG, and X-Gal, where appropriate (*see* **Notes 22–24**). For cloned libraries, use the rest of the cell culture to inoculate 5 mL liquid ampicillin-LB medium (*see* **Note 25**). Incubate the liquid culture and the plates overnight at 37 °C.

3.4 Consecutive MoClo Assembly Steps

1. Evaluate the bacterial colonies on the plates. For working with sequence libraries, the presence of many (several hundreds) colonies is not only expected, but also essential to confirm high diversity of the plasmid library. For cloned PCR products of defined sequences, isolation of a single, white colony must be possible (*see* **Note 26**).
2. For building block libraries, purify the plasmid library from 5 mL of the liquid culture using a standard plasmid purification kit.
3. For fully defined PCR building blocks, pick a white colony from the plate and inoculate a liquid ampicillin-LB culture. Incubate overnight at 37 °C. Purify the plasmid from the culture using a standard plasmid purification kit.
4. Check isolated plasmids for the presence of the desired building block by either restriction digest or PCR combined with gel electrophoresis or sequencing.
5. For sets of five building blocks to be assembled according to your assembly scheme, perform a BpiI-dependent level-M MoClo reaction with five level-1 vectors containing either one specific DNA building block or a library of building blocks (see above) and the respective level-M vectors and end-linkers (cf. Subheading 3.1 and Fig. 5) (*see* **Note 27**).

   Prepare the following 25 μL reaction:

   - 2.5 μL 10 × T4 DNA-ligase buffer.

     40 fmol level-M vector.

     40 fmol of each level-1 vector with building blocks.

     40 fmol of the respective end-linker vector.

     1 μL T4 DNA-Ligase (5 U/μL).

     1 μL BpiI.

   Add up to 25 μL with ddH_2O.
6. Run the MoClo reaction using the parameters given in Subheading 3.3, **step 14**.
7. Transform *E. coli* with the MoClo reaction and plate 50 μL of cell culture on spectinomycin plates with IPTG and X-Gal where appropriate (*see* **Note 23**).
8. Isolate plasmids from individual colonies and check for the presence of the building block by either restriction digest and gel analysis or sequencing.
9. Perform sets of BsaI-dependent level-P MoClo reactions with five level-M building block vectors and the respective level-P vectors and end-linkers (cf. Subheading 3.1 and Fig. 5) (*see* **Note 28**).

Prepare the reaction as described in Subheading 3.3, **step 13** and run the MoClo reaction using the parameters as given in Subheading 3.3, **step 14**.

10. Transform *E. coli* cells with the MoClo reactions and plate cells on ampicillin-plates with IPTG and X-Gal where appropriate (*see* **Note 23**). Incubate plates for 12 h at 37 °C.
11. Isolate and check plasmids as described above.
12. Perform a BpiI-dependent level-M MoClo reaction with five level-P building block vectors, the MoClo-compatible synVicII backbone and end-linker as described in **step 5** (cf. Subheading 3.1 and Fig. 5) (*see* **Note 29**). In this step, it is critical to only use the end-linker vectors encoding chloramphenicol resistance (*see* **Note 30**). Use a 0.5 mL PCR Tube for the preparation of the reaction to leave space for the addition of competent *E. coli* cells (*see* **Note 31**).
13. Transfer 400 µL competent *E. coli* cells to the MoClo reaction. Follow the standard transformation protocol.

3.5 Extraction of Synthetic Secondary Chromosomes

Extraction of large replicons is difficult due to not only their size but also their low copy number. The procedure described below is based on a protocol first described by Rondon et al. [28] and later adapted by Heringa et al. [29]. This procedure is best suited for analysis of synthetic chromosomes of 40–100 kb. Extracted DNA can be used for restriction analyses and transformation of *E. coli*.

1. Grow cells harboring the synthetic secondary chromosome in liquid medium with ampicillin.
2. Harvest 5 mL of cells by centrifugation and discard the culture medium.
3. Resuspend cells by vortexing in 100 µL resuspension buffer (*see* **Note 32**).
4. Add 200 µL Lysis solution and mix by inverting several times. Incubate at room temperature for 5 min (*see* **Note 33**).
5. Add 150 µL ammonium acetate (7.5 M) and 150 µL chloroform, quickly afterward. Invert several times. Let it chill on ice for 10 min. Centrifuge for 10 min at maximum speed.
6. Pipette 200 µL precipitation solution into a reaction tube. Add the upper aqueous phase of the supernatant. Invert several times and let chill on ice for 15 min.
7. Centrifuge at maximum speed to pellet DNA. Remove the supernatant and resuspend in 25 µL ddH_2O (*see* **Note 34**). Store DNA on ice and proceed promptly with downstream procedures.

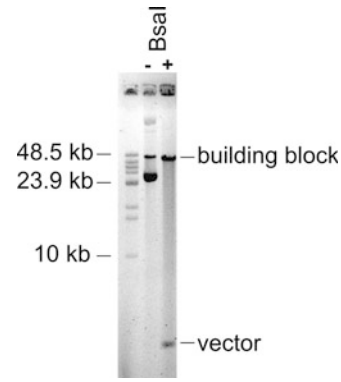


Fig. 9 Cut and uncut DNA of a 42,500 bp building block on a 0.4% agarose gel. DNA of a building block cloned into a level-M MoClo vector was separated by gel electrophoresis on a 0.4% agarose gel containing SDS either cut or uncut as indicated. The DNA separates into a 42,500 bp fragment presenting the building block and the MoClo vector of 1800 bp upon BsaI digest

3.6 Gel Electrophoretic Analysis of Synthetic Secondary Chromosomes

One way to analyze constructed synthetic chromosomes is a restriction digest combined with a standard agarose gel (1% agarose in TAE buffer) at standard parameters (120 V for one hour). However, this is difficult if restriction sites are unknown and respective building blocks exceed 10 kb. In other cases, a size verification of the final building block into the core replicon might be desired. Both demands require a specialized gel electrophoretic separation as follows:

1. Incubate 12 μL of isolated DNA (cf. Subheading 3.5) in a total volume of 20 μL with either 5 U BsaI or BpiI (the enzyme that releases the building block from the core replicon) for 1 h according to the manufacturer's instructions.
2. Run restriction digested DNA on a 0.4% agarose gel in TAE buffer containing 0.2% SDS (cf. Chemicals) at 30 V overnight (*see* **Note 35**). Use a DNA ladder that sufficiently resolves the size of the synthetic chromosome.
3. Stain DNA with your dye of choice and visualize under UV light. An example is shown in Fig. 9.

4 Notes

1. To construct a MoClo compatible core replicon one could follow the procedure using homologous recombination in yeast as described for synVicII [27].
2. The final size of the synthetic secondary chromosome will depend on the number and size of individual basic building blocks. Level-P and level-M vectors can be used in alternating

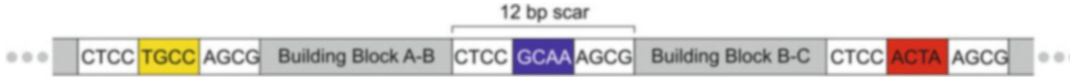


Fig. 10 Fusion of level-1 building blocks leaves a 12 bp scar. Building Blocks A-B and B-C (gray) are fused at the B overhang (dark purple), which will consequently remain between the two building blocks. Moreover, the universal basic building block overhangs "CTCC" and "AGCG" (black frame) acquired during generation of basic building blocks, add additional 8 bp to the scar

assembly steps basically infinitely. The size limit might however be set by the decreasing transformation efficiency of *E. coli* with increasing size of DNA fragments as well as shearing forces which lead to DNA damage by pipetting. We routinely assemble and transform replicons up to 100,000 bps.

3. The length, as well as the nature of the level-1 building blocks, can be varied according to the respective replicon design. It might be advantageous to work with DNA libraries for example if spacers are constructed where the exact sequence is not essential. On the other hand, one might use specific sequences of interests such as genes, promoters, or entire transcription units that can be amplified by PCR from template DNA. The position of single building blocks within the final synthetic chromosome is dependent on the four-base overhangs that are generated in the first assembly step with level-1 vectors (cf. Generation of Level-1 Building Blocks 3.3). With consecutive assembly steps, a 12 bp long scar sequence will remain between the basic building blocks due to the MoClo assembly system (*see* Fig. 10). Note that the length of those scars is dependent on the exact MoClo vectors being used. Level-1 pMA vectors shown in Fig. 5b will create a 12 bp scar but the use of other MoClo vectors can result in longer scars encoding specific restriction sites to facilitate assembly analyses [16, 21]. Building blocks containing the same overhangs are interchangeable. This can reduce the total number of basic building blocks required for chromosome assembly especially when using DNA libraries. Furthermore, it makes shuffling of building blocks (efficient creation of different synthetic chromosome versions) possible. This is, for example, relevant when studying position-dependent effects on gene expression or chromosome maintenance.
4. Chromosome assembly starts with level-1 vectors instead of level-0 vectors (used in the original MoClo system [16]) because all seven level-1 vectors carry the same overhangs for the insertion of DNA building blocks (depicted as black frame in Fig. 10 and white rectangle in Fig. 5b, c). Consequently, all basic building blocks (independent of how exactly they were constructed) contain the universal level-1 overhangs (depicted as black frame in Figs. 6 and 7 and white rectangle in Fig. 5b, c) and can be cloned into any level-1 vector.

5. All level-1 vectors described here, differ from the original MoClo vectors [16] in such a way that a 12 bp long sequence between the BpiI and BsaI recognition sites has been deleted to reduce the danger of homologous recombination events taking place in the final constructs.
6. Level-1 vectors are available either as lacZ version, where selection of building block carrying vectors is dependent on blue white screening or as *ccdB* version, where cells harboring the vector without building block insertion will consequently die because of the expression of the CcdB toxin [21]. The *ccdB* version is particularly suited for library cloning where it is critical to recover only positive clones to prevent the presence of any MoClo vector parts in the final construct.
7. In addition, level-1 vectors are available as forward (fw) and reverse (rv) versions. The difference between the two versions is the switched position of the universal overhangs "GGAG" and "AGCG." By choosing the vector, one determines the direction of the respective basic building block (e.g., the orientation of a gene).
8. Replication of oriR6K requires specially engineered E. coli strains carrying the *pir* gene. By using these vectors for the assembly of chromosomal integration cassettes, one omits the presence of false positive clones, harboring the plasmid.
9. Many companies offer the production of mixtures of oligonucleotides based on a single degenerated sequence determined by the customer (mixtures of nucleotides are used, e.g., Y for pyrimidines). DNA library building blocks might be used interchangeably at different positions within the chromosome as long as the MoClo vectors share the same overhangs. This saves costly working time.
10. Multiple restriction sites can be removed this way in a single cloning step. However, if a very high number of internal sites need to be changed it may be cheaper or more convenient to order synthetic DNA fragments.
11. The degree of diversity of the DNA library will decrease with increasing number of motifs to be excluded by the user [21]. Consider this when using the MARSeG program.
12. The length of the region to be annealed should not be too short to allow efficient pairing. On the other hand, long annealing regions raise the risk of homologous recombination events taking place within the final replicon if building blocks are used at multiple positions. One solution is to introduce degenerated nucleotides (such as Ns) into the annealing region. This is possible as far as the variability is not increased too much since each oligonucleotide needs to anneal to a matching oligonucleotide and this likelyhood decreases with

increasing diversity of the annealing region. We have successfully worked with annealing regions of 30 bps including five Ns at various positons leading to 1024 different annealing region variations. Annealing sites might also consist of defined sequences such as DNA motifs (e.g., *dif*-site of *E. coli* shown in Fig. 7) or promoters, where the use of degenerated nucleotides is not possible.

13. To save time by avoiding a subsequent extension step, oligonucleotides can be designed in such a way that annealing occurs over the entire length. However, full annealing is limited by the annealing efficiency that decreases with an increasing number of degenerated sequences (cf. **Note 11**) as well as by the practical limitations of chemical oligonucleotide synthesis (200 bps).
14. 4 nmol of standard desalted oligonucleotides with a size of 200 bps is sufficient for annealing and subsequent elongation.
15. The use of Taq-Polymerase is highly advisable, because it tolerates mismatches.
16. The annealing temperature is dependent on the oligonucleotides. Oligonucleotides consisting of defined sequences should be annealed according to their melting temperature. Degenerated oligonucleotide mixtures should be annealed in 12 parallel reactions with different annealing temperatures ranging for example from 57 to 62 °C (depending on specific design).
17. Double-stranded DNA products of library fragments run as gradient at different temperatures should be pooled before purification.
18. Primers bind annealed and elongated reaction products within the universal level-1 MoClo overhangs. The PCR product expected will not change in size, but will have a higher yield. This step is required because the concentration of gel-extracted DNA tends to be relatively low.
19. Annealing, extension, and enrichment of the oligonucleotides may result in unwanted reaction products of different sizes. This step is critical to obtain DNA library building blocks of correct size.
20. For library cloning, efficiency of the MoClo reaction can be increased by using a highly concentrated T4 DNA-Ligase (20 U/μL).
21. 40 fmol is equivalent to about 100 ng of a 4 kb vector based on the average molecular weight of a nucleotide pair being 660 g/mol.
22. *Escherichia coli* cloning strains including but not limited to DH5*α*, *DH5αλpir*, *Top10*, *XL1-Blue* are suitable because of the lack of recombination systems that would interfere with cloning and plasmid maintenance.

23. Addition of IPTG and X-Gal is only useful when using the *lacZ* MoClo vectors.
24. When working with the ccdB-based MoClo vectors, make sure to use a CcdB-sensitive *E. coli* strain such as *E. coli* Top 10. For cloning with the original *lacZ*-based MoClo vectors, only use Δ*lacZ E. coli* strains to allow blue-white-screening on plates containing IPTG and X-Gal (e.g., *E. coli* XL1-Blue). On these plates, clones that do not have the plasmid-borne *lacZ* anymore because of a successful cloning event can be identified by their white color, while clones without insert will express *lacZ* and consequently show a blue color.
25. Inoculation of a liquid culture is required to obtain a highly diverse plasmid library and not individual clones.
26. Several hundreds of colonies are expected. The number of colonies is dependent on the quality of the competent cells rather than on the efficiency of the MoClo reaction.
27. For library cloning only use the level-M *ccdB* vectors (pMA60–66) (cf. **Note 6**). In most cases, assembly of five level-1 vectors will contain at least one fully defined subsequence. In these cases, the assembly is continued with individual level-M vectors carrying the desired sequence arrangement. However, if the five level-1 building blocks assembled into one level-M vector are all derived from sequence libraries, one can continue with further assemblies based on level-M vector libraries if desired. This could for example apply if larger spacer sequences are constructed from smaller building blocks.
28. For library cloning only use the level-P ccdB vectors, (pMA67–73) (cf. **Note 6**).
29. To increase efficiency of large-building block assemblies use highly concentrated T4 DNA-Ligase (20 U/μL).
30. End-linker vectors mediating ampicillin resistance (but not the end-linker vectors mediating chloramphenicol) would result in *E. coli* colonies carrying these end-linker vectors only, because counter-selection against the *amp* encoding synVicII is not possible.
31. Addition of competent cells to the MoClo reaction will prevent plasmid shearing during pipetting of the DNA. This improves overall transformation efficiency. It is advisable to eventually transform a conjugation-proficient *E. coli* strain such as MFD-pir [30] to allow future transfer of the mobilizable synthetic chromosome to other *E. coli* strains.
32. To eliminate RNA in this DNA isolation it is advised to add 1 μL RNase A to the resuspension buffer as done in conventional plasmid isolation procedures. However, in our experience some RNA remains. If a restriction digest is performed

later, RNase A can alternatively be added to the reaction mix (0.1 μL per reaction).

33. The Lysis buffer should always be mixed freshly from stock solutions of NaOH and SDS.
34. Purified DNA (100 kb synthetic chromosome) of a 5 mL culture is of sufficient quality and concentration for two restriction digests. When used for next-generation sequencing, residual *E. coli* chromosomal DNA was detected. As suggested by Heringa et al. DNA can be further purified by gel electrophoresis and subsequent gel extraction.
35. SDS must only be added after heating up the agarose!

References

1. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi Z-Q, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56. https://doi.org/10.1126/science.1190719
2. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, Cai Y, Zeller K, Agmon N, Han JS, Hadjithomas M, Tullman J, Caravelli K, Cirelli K, Guo Z, London V, Yeluru A, Murugan S, Kandavelou K, Agier N, Fischer G, Yang K, Martin JA, Bilgel M, Bohutskyi P, Boulier KM, Capaldo BJ, Chang J, Charoen K, Choi WJ, Deng P, DiCarlo JE, Doong J, Dunn J, Feinberg JI, Fernandez C, Floria CE, Gladowski D, Hadidi P, Ishizuka I, Jabbari J, Lau CYL, Lee PA, Li S, Lin D, Linder ME, Ling J, Liu J, Liu J, London M, Ma H, Mao J, McDade JE, McMillan A, Moore AM, Oh WC, Ouyang Y, Patel R, Paul M, Paulsen LC, Qiu J, Rhee A, Rubashkin MG, Soh IY, Sotuyo NE, Srinivas V, Suarez A, Wong A, Wong R, Xie WR, Xu Y, Yu AT, Koszul R, Bader JS, Boeke JD, Chandrasegaran S (2014) Total synthesis of a functional designer eukaryotic chromosome. Science 344:55–58. https://doi.org/10.1126/science.1249252
3. Schindler D, Waldminghaus T (2015) Synthetic chromosomes. FEMS Microbiol Rev 39:871–891. https://doi.org/10.1093/femsre/fuv030
4. Touzain F, Petit MA, Schbath S, El Karoui M (2011) DNA motifs that sculpt the bacterial chromosome. Nat Rev Microbiol 9:15–26. https://doi.org/10.1038/nrmicro2477
5. Milbredt S, Farmani N, Sobetzko P, Waldminghaus T (2016) DNA replication in engineered escherichia coli genomes with extra replication origins. ACS Synth Biol 5:1167–1176. https://doi.org/10.1021/acssynbio.6b00064
6. Hutchison CA 3rd, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC (2016) Design and synthesis of a minimal bacterial genome. Science 351:aad6253. https://doi.org/10.1126/science.aad6253
7. Yu W, Han F, Gao Z, Vega JM, Birchler JA (2007) Construction and behavior of engineered minichromosomes in maize. Proc Natl Acad Sci U A 104:8924–8929. https://doi.org/10.1073/pnas.0700932104
8. Birchler JA (2015) Promises and pitfalls of synthetic chromosomes in plants. Trends Biotechnol 33:189–194. https://doi.org/10.1016/j.tibtech.2014.12.010
9. Dasgupta S, Lobner-Olesen A (2004) Host controlled plasmid replication: Escherichia coli minichromosomes. Plasmid 52:151–168. https://doi.org/10.1016/j.plasmid.2004.08.001
10. Casini A, Storch M, Baldwin GS, Ellis T (2015) Bricks and blueprints: methods and standards for DNA assembly. Nat Rev Mol Cell Biol 16:568–576. https://doi.org/10.1038/nrm4014
11. Karas BJ, Suzuki Y, Weyman PD (2015) Strategies for cloning and manipulating natural and

synthetic chromosomes. Chromosom Res 23:57–68. https://doi.org/10.1007/s10577-014-9455-3

12. Ma H, Kunes S, Schatz PJ, Botstein D (1987) Plasmid construction by homologous recombination in yeast. Gene 58:201–216
13. Benders GA, Noskov VN, Denisova EA, Lartigue C, Gibson DG, Assad-Garcia N, Chuang RY, Carrera W, Moodie M, Algire MA, Phan Q, Alperovich N, Vashee S, Merryman C, Venter JC, Smith HO, Glass JI, Hutchison CA 3rd (2010) Cloning whole bacterial genomes in yeast. Nucleic Acids Res 38:2558–2569. https://doi.org/10.1093/nar/gkq119
14. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CLV, Chandrasegaran S, Cai Y, Boeke JD, Bader JS (2017) Design of a synthetic yeast genome. Science 355:1040–1044. https://doi.org/10.1126/science.aaf4557
15. de Kok S, Stanton LH, Slaby T, Durot M, Holmes VF, Patel KG, Platt D, Shapland EB, Serber Z, Dean J, Newman JD, Chandran SS (2014) Rapid and reliable DNA assembly via ligase cycling reaction. ACS Synth Biol 3:97–106. https://doi.org/10.1021/sb4001992
16. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6:e16765. https://doi.org/10.1371/journal.pone.0016765
17. Werner S, Engler C, Weber E, Gruetzner R, Marillonnet S (2012) Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. Bioeng Bugs 3:38–43. https://doi.org/10.4161/bbug.3.1.18223
18. Lee ME, DeLoache WC, Cervantes B, Dueber JE (2015) A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth Biol 4:975–986. https://doi.org/10.1021/sb500366v
19. Engler C, Youles M, Gruetzner R, Ehnert TM, Werner S, Jones JD, Patron NJ, Marillonnet S (2014) A golden gate modular cloning toolbox for plants. ACS Synth Biol 3:839–843. https://doi.org/10.1021/sb4001504
20. Duportet X, Wroblewska L, Guye P, Li Y, Eyquem J, Rieders J, Rimchala T, Batt G, Weiss R (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. Nucleic Acids Res 42:13440–13451. https://doi.org/10.1093/nar/gku1082
21. Schindler D, Milbredt S, Sperlea T, Waldminghaus T (2016) Design and assembly of DNA sequence libraries for chromosomal insertion in bacteria based on a set of modified MoClo vectors. ACS Synth Biol 5(12):1362–1368. https://doi.org/10.1021/acssynbio.6b00089
22. Messerschmidt SJ, Kemter FS, Schindler D, Waldminghaus T (2015) Synthetic secondary chromosomes in Escherichia coli based on the replication origin of chromosome II in Vibrio cholerae. Biotechnol J 10:302–314. https://doi.org/10.1002/biot.201400031
23. Harrison PW, Lower RPJ, Kim NKD, Young JPW (2010) Introducing the bacterial "chromid": not a chromosome, not a plasmid. Trends Microbiol 18:141–148. https://doi.org/10.1016/j.tim.2009.12.010
24. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM (2000) DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. Nature 406:477–483. https://doi.org/10.1038/35020000
25. Egan ES, Waldor MK (2003) Distinct replication requirements for the two vibrio cholerae chromosomes. Cell 114:521–530. https://doi.org/10.1016/S0092-8674(03)00611-1
26. Pal D, Venkova-Canova T, Srivastava P, Chattoraj DK (2005) Multipartite regulation of rctB, the replication initiator gene of vibrio cholerae chromosome II. J Bacteriol 187:7167–7175. https://doi.org/10.1128/JB.187.21.7167-7175.2005
27. Messerschmidt SJ, Schindler D, Zumkeller CM, Kemter FS, Schallopp N, Waldminghaus T (2016) Optimization and characterization of the synthetic secondary chromosome synVicII in Escherichia coli. Front Bioeng Biotechnol 4:96. https://doi.org/10.3389/fbioe.2016.00096
28. Rondon MR, Raffel SJ, Goodman RM, Handelsman J (1999) Toward functional genomics in bacteria: analysis of gene expression in Escherichia coli from a bacterial artificial chromosome library of Bacillus cereus. Proc Natl Acad Sci U S A 96:6451–6455
29. Heringa SD, Monroe JD, Herrick JB. A simple, rapid method for extracting large plasmid DNA from Bacteria. Nat Preced. 2007. doi: https://doi.org/10.1038/npre.2007.1249.1

30. Ferrières L, Hémery G, Nham T, Guérout A-M, Mazel D, Beloin C, Ghigo J-M (2010) Silent mischief: bacteriophage Mu insertions contaminate products of Escherichia coli random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range rp4 conjugative machinery. J Bacteriol 192:6418–6427. https://doi.org/10.1128/JB.00621-10

Chapter 6

High-Resolution Characterization of DNA/Protein Complexes in Living Bacteria

Nicole A. Becker, Justin P. Peters, and L. James Maher III

Abstract

The occurrence of DNA looping is ubiquitous. This process plays a well-documented role in the regulation of prokaryotic gene expression, such as the *Escherichia coli* lactose (*lac*) operon. Here, we present two complementary methods for high-resolution in vivo detection of DNA/protein binding within the bacterial nucleoid by using either chromatin immunoprecipitation combined with phage λ exonuclease digestion (ChIP-exo) or chromatin endogenous cleavage (ChEC), coupled with ligation-mediated polymerase chain reaction (LM-PCR) and Southern blot analysis. As an example we apply these in vivo protein-mapping methods to *E. coli* to show direct binding of architectural proteins in the Lac repressor-mediated DNA repression loop.

Key words Architectural proteins, Chromatin endogenous cleavage (ChEC), Chromatin immunoprecipitation (ChIP), Phage lambda exonuclease, High-resolution mapping, Ligation-mediated PCR (LM-PCR), Polymerase chain reaction (PCR), Lac repression loop, Southern blot

1 Introduction

A common mechanism of gene repression in bacteria is to sequester a gene promoter inside a tightly bent DNA loop. This type of gene repression depends on many factors, including the mechanical properties of the looped DNA and the concentration and binding strength of the protein that anchors the repression loop [1–4]. A classic, well-studied example is the Lac repressor protein anchoring *lac* repression loops [5–11]. However, many aspects of this looped system remain difficult to access experimentally. In particular, many groups have speculated about the participation of other, sequence-nonspecific "architectural" DNA-binding proteins in repression loop formation and stabilization [12–15], but the lack of high-resolution protein-mapping techniques in vivo has prohibited such analyses. We recently sought to test whether tight looping of

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_6,

DNA in *lac* promoter repression is aided or facilitated by direct binding of architectural proteins within the DNA loop [16]. Here, we describe methodology for two complementary high-resolution protein-mapping techniques and their utility in showing that an exogenous architectural protein can directly bind tightly looped DNA to facilitate gene repression by the Lac repressor in living *E. coli* cells.

The first method combines chromatin immunoprecipitation (ChIP), which identifies where a DNA-binding protein of interest is located on the genome, with phage λ exonuclease digestion (exo), which degrades each 5′-phosphorylated DNA strand in the 5′→3′ direction until a protein-DNA cross-linking point is encountered, followed by ligation-mediated Polymerase Chain Reaction (LM-PCR) and Southern blot analysis, which allows for mapping of these protein-DNA boundary sites at single base resolution in a chosen genomic region of interest. A schematic overview of this method is shown in Fig. 1. The chromatin immunoprecipitation-exonuclease (ChIP-exo) analysis method was adapted for bacterial analysis from [17, 18] and LM-PCR was adapted from standard methods [19–22]. Since ChIP-exo is an immunoprecipitation-based method, it requires either an antibody to the targeted protein or a derivative of this protein which is tagged and can be targeted for immunoprecipitation.

The second method utilizes chromatin endogenous cleavage (ChEC), in which specifically designed endonucleases are cross-linked in vivo and then activated with Ca^{2+} ions to locally introduce double-stranded DNA breaks at the sites of protein-DNA cross-linking, followed by LM-PCR and Southern blot analysis. A schematic overview of this method is shown in Fig. 2. The chromatin endogenous cleavage (ChEC) analysis method was adapted for bacterial analysis from previous publications [23, 24]. ChEC requires that a protein fusion be created between the desired mapping protein and micrococcal nuclease (MNase), which can be appended to the N or C terminus of the fusion (*see* **Note 1**).

2 Materials

2.1 Equipment

1. Microcentrifuge.
2. Floor shaker.
3. Spectrophotometer.
4. Sonicator.
5. Paramagnetic stand.
6. Water baths.
7. Thermocycler.

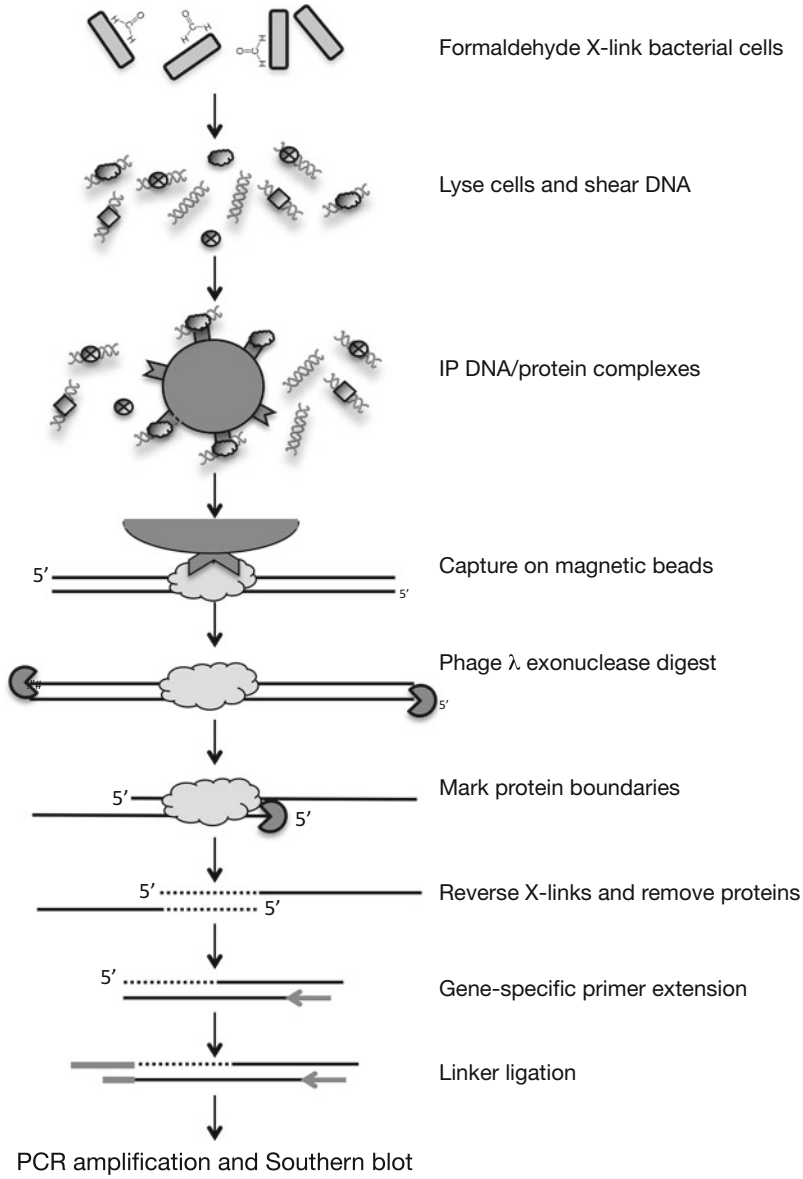


Fig. 1 Schematic overview of ChIP-exo-based method of high-resolution protein mapping

8. Lyophilizer.
9. Sequencing gel apparatus.
10. Owl semi-wet transfer system.
11. UV Stratalinker.
12. Hybridization oven.
13. Phosphorimager.

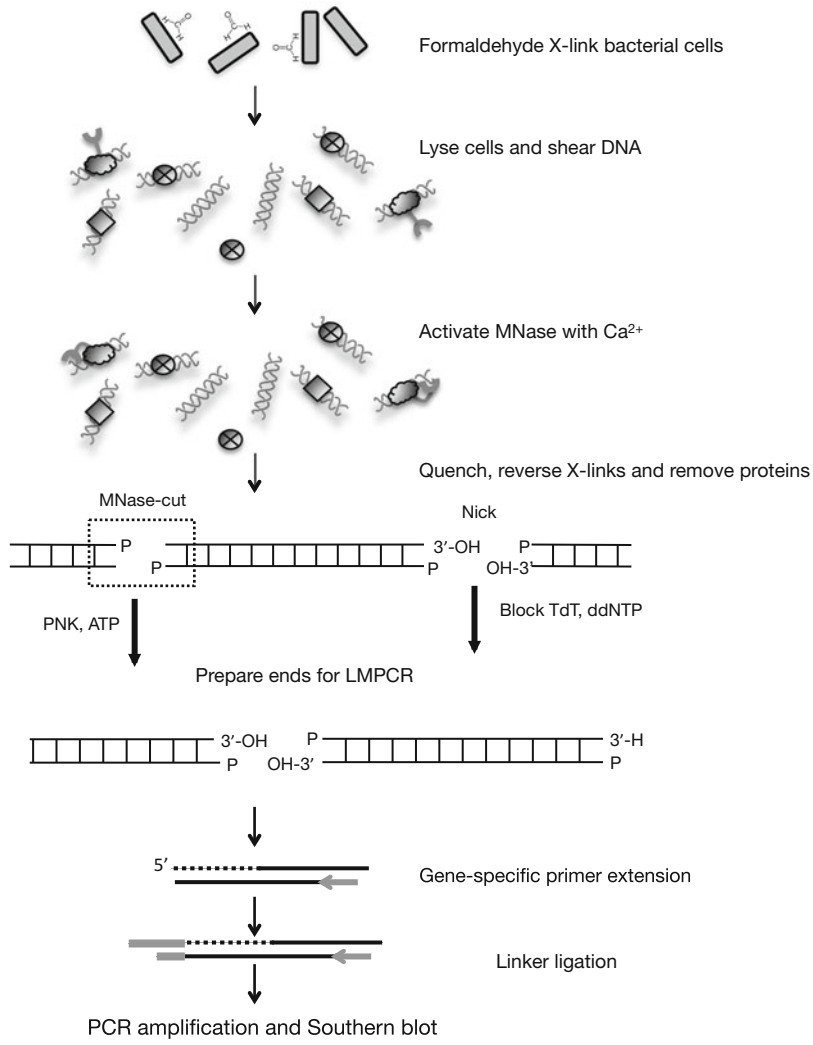


Fig. 2 Schematic overview of ChEC-based method of high-resolution protein mapping

2.2 Reagents

1. Bacterial strain of interest and necessary antibiotic(s).
2. Necessary antibody reagents for ChIP-exo, e.g., antibodies against IgG (AbCam, polyclonal antibody), LacI (LSBio, polyclonal antibody), and Myc epitope tag (Sigma, polyclonal antibody).
3. Autoclaved ultrapure (18 MΩ-cm) water.
4. Miller's Luria-Bertani (LB) media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.0).
5. 1× DPBS, no calcium, no magnesium (ThermoFisher).
6. Fresh 37% formaldehyde.
7. 2 M Tris–HCl, pH 8.0.
8. 20% SDS.

9. 5 M NaCl.
10. 0.5 M EDTA.
11. 100 mM EGTA.
12. 10% Triton-X.
13. 1% sodium deoxycholate.
14. FA lysis buffer (50 mM HEPES-KOH, pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton-X, 0.1% sodium deoxycholate).
15. 100 mM PMSF (phenylmethylsulfonyl fluoride).
16. Protease inhibitor cocktail (Roche).
17. 10 mg/mL ribonuclease A (RNaseA).
18. Protein A/G magnetic beads (Pierce).
19. FA high salt wash buffer (50 mM HEPES-KOH, pH 8.0, 2 mM EDTA, 1 M NaCl, 1% Triton-X, 0.1% sodium deoxycholate).
20. FA low salt wash buffer (50 mM HEPES-KOH, pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% Triton-X, 0.1% sodium deoxycholate).
21. FA wash buffer 3 (10 mM Tris–HCl, pH 8.0, 2 mM EDTA, 25 mM LiCl 1% Igepal, 1% sodium deoxycholate).
22. TE + Triton-X buffer, pH 8.0 (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 0.1% Triton-X).
23. TE + Triton-X buffer, pH 8.9 (10 mM Tris–HCl, pH 8.9, 1 mM EDTA, 0.1% Triton-X).
24. TE + Triton-X buffer, pH 7.4 (10 mM Tris–HCl, pH 7.4, 1 mM EDTA, 0.1% Triton-X).
25. 10× Lambda (λ) exonuclease reaction buffer, pH 9.4 (New England Biolabs).
26. 5 U/μL λ exonuclease (New England Biolabs).
27. 10× NEB buffer 2, pH 7.4 (New England Biolabs).
28. 30 U/μL $RecJ_f$ (New England Biolabs).
29. Elution buffer (50 mM Tris–HCl, pH 8.0, 10 mM EDTA, 1% SDS).
30. 10 mg/mL proteinase K.
31. 3 M NaOAc, pH 5.2.
32. 10 mg/mL glycogen.
33. 1:1 phenol:chloroform.
34. 100% ethanol.
35. 70% ethanol.
36. ChEC lysis buffer (50 mM HEPES-KOH, pH 8.0, 10 mM EGTA, 1 mM EDTA, pH 8.0, 150 mM NaCl, 1% Triton-X, and 0.1% sodium deoxycholate).
37. 2× stop solution (2% SDS, 400 mM NaCl, 20 mM EDTA, 5 mM EGTA, and 0.2% glycogen).

38. 100 mM $CaCl_2$.
39. 20 U/μL terminal (deoxynucleotidyl)transferase (New England Biolabs).
40. 10× terminal (deoxynucleotidyl)transferase buffer (New England Biolabs).
41. 2.5 mM $CoCl_2$.
42. 2.5 mM ddNTP mix.
43. 10× phi29 DNA polymerase reaction buffer (New England Biolabs).
44. 10 U/μL phi29 DNA polymerase (New England Biolabs).
45. 2.5 mM dNTP mix.
46. 1 mg/mL BSA (bovine serum albumin).
47. 50 mM ATP.
48. 10× T4 DNA ligase buffer (New England Biolabs).
49. 10 U/μL T4 polynucleotide kinase (New England Biolabs).
50. 400 U/μL T4 DNA ligase (New England Biolabs).
51. 10 mg/mL tRNA.
52. Ligation stop mix (1.2 M NaOAc, pH 5.2 with 20 μg/μL tRNA).
53. 10× *Taq* polymerase buffer.
54. 5 U/μL *Taq* polymerase.
55. 50 mM $MgCl_2$.
56. *Taq* stop mix (0.53 M NaOAc, pH 5.2, 70 μg/μL tRNA, and 18 mM EDTA).
57. Deionized formamide dye mix (99.5% deionized formamide, 1 mg/mL xylene cyanol, 1 mg/mL bromophenol blue, and 10 mM EDTA).
58. DMS (dimethyl sulfate).
59. DMS buffer (50 mM sodium cacodylate, pH 6.5 with 1 mM EDTA).
60. DMS stop buffer (1.5 M NaOAc, pH 7.0 with 1 M β-mercaptoethanol).
61. Formic acid.
62. Hydrazine.
63. Hydrazine stop buffer (0.3 M NaOAc, pH 7.5 with 0.1 M EDTA).
64. 1 M piperidine.
65. 0.1 M piperidine.
66. 40% 19:1 acrylamide:bisacrylamide solution.
67. 0.5× TBE buffer (40 mM Tris–HCl, pH 8.3, 45 mM boric acid, and 1 mM EDTA).

68. 6% Denaturing polyacrylamide gel mix (6% 19:1 acylamide/bisacrylamide, 7.5 M Urea, 0.5× TBE).
69. 2.5 mM dCTP.
70. 2.5 mM dGTP.
71. 2.5 mM dTTP.
72. α-^{32}P dATP (3000 Ci/mmol; 10 mCi/mL).
73. Chroma-spin + TE columns (Clontech).
74. Zeta-probe nylon membrane (Bio-Rad).
75. 17 mm Whatman paper.
76. Hybridization buffer (0.25 M Na_2HPO_4, pH 7.3, 1 mM EDTA, 7% SDS, and 10 mg/mL BSA).
77. Wash buffer 1 (20 mM Na_2HPO_4, pH 7.3, 1 mM EDTA, 2.5% SDS, and 2.5 mg/mL BSA).
78. Wash buffer 2 (20 mM Na_2HPO_4, pH 7.3, 1 mM EDTA, and 1% SDS).

3 Methods

3.1 In Vivo Cross-Linking with Formaldehyde

1. Grow desired bacterial strain(s) to saturation in 5 mL of LB with appropriate antibiotic(s).
2. Start a 40 mL LB subculture, with appropriate antibiotic(s), from 500 μL of the saturated overnight bacterial culture. Grow at 37 °C with constant agitation at 280 rpm, until cells are at log phase with an OD_{600} in a 1 cm path cuvette between 0.4 and 0.8.
3. In a 50 mL conical tube, collect each cell pellet by centrifugation at 4000 × *g* for 10 min at room temperature. Discard the supernatant and completely resuspend each pellet in 20 mL of room temperature DPBS.
4. Add 37% formaldehyde (*see* **Note 2**) to a final concentration of 0.75%. For 20 mL of DPBS suspension add 405 μL 37% formaldehyde. Incubate samples at room temperature for 20 min with constant rotation.
5. Stop the formaldehyde cross-linking with 6 mL 2 M Tris–HCl, pH 8.0, for a final concentration 260 mM (*see* **Note 3**). Further incubate at room temperature with constant rotation for an additional 5 min.
6. Harvest cross-linked cells by centrifugation at 4000 × *g* for 10 min at 4 °C. Resuspend in 10 mL of cold DPBS. Wash the cell pellets 3 times with 10 mL of cold DPBS by centrifugation at 4000 × *g* for 10 min at 4 °C. Following the final wash and centrifugation, decant the supernatant and store the cross-linked pellets at −80 °C until ready to proceed with lysis.

3.2 Chromatin Immunoprecipitation-Exonuclease (ChIP-Exo) Analysis

1. Thaw cross-linked bacterial pellets (*see* Subheading 3.1, **step 6**) on ice for 10 min, then resuspend the cell pellets in 1 mL cold FA lysis buffer supplemented with 1 mM PMFS, protease inhibitor cocktail, 100 μg/mL RNaseA, and 0.2% SDS. Transfer the cell suspension from 50 mL conical tube to a 1.7 mL microfuge tube.
2. Lyse cells and shear cellular DNA with sonication on ice with four 10-s bursts at 10 W (*see* **Note 4**). Place samples on ice between pulses for a minimum of 30 s to allow the lysates to remain cold.
3. Remove cellular debris with centrifugation at 14,000 × *g* for 3 min at 4 °C. Transfer clarified supernatant to a fresh 1.7 mL microfuge tube. A small amount, ~10 μL, of lysate should be checked on a 0.8% agarose gel (*see* Fig. 3) to check the quality and size distribution of the sheared DNA (*see* **Note 5**).
4. Aliquots of the lysate can be frozen at −80 °C, to be thawed on ice for future immunoprecipitations. The clarified lysate can also be used immediately in the immunoprecipitation protocol.

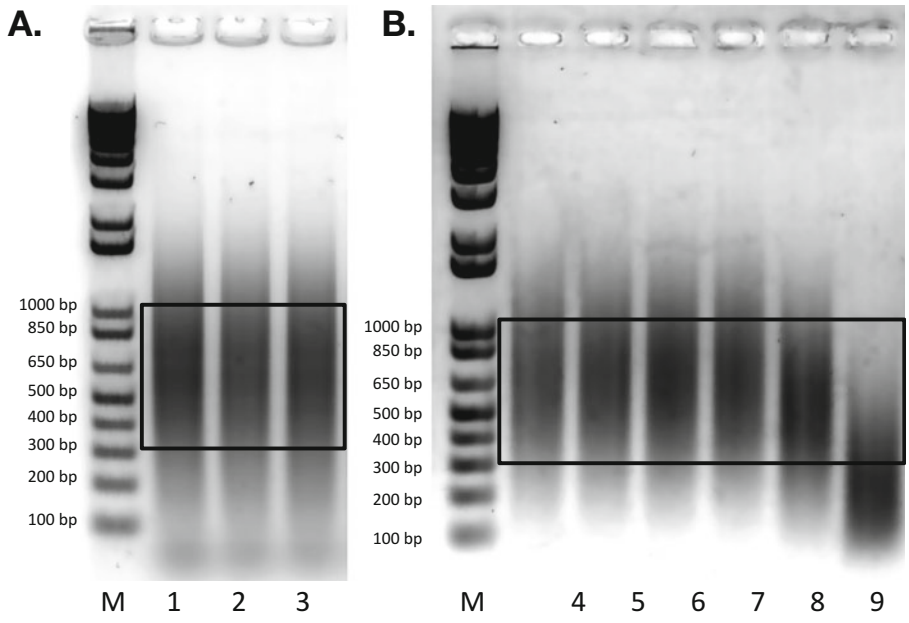


Fig. 3 Quality and size distribution of sheared DNA. (**a**) Representative ChIP-exo lysate samples following formaldehyde cross-linking and lysis. Lanes 1–3 show three different bacterial strains grown, treated, and lysed in parallel. (**b**) Representative ChEC DNA samples following cross-linking, lysis, and MNase cleavage. Lanes 4–9 represent a single bacterial strain with MNase activation time points of 0, 5, 15, 30, 45, and 60 min, respectively. Ideal size distribution for the majority of the DNA for both ChIP-exo and ChEC DNA is between 300 bp and 1000 bp. A 1 kb Plus DNA ladder, with fragment sizes (in base pairs) is indicated, M

5. For a standard immunoprecipitation, on ice combine 75 μL lysate sample (equivalent to 3 mL original bacterial culture) and 675 μL FA lysis buffer. Add 6 μg of antibody specific to protein of interest. Incubate immunoprecipitations overnight with gentle rotation at 4 °C.
6. Capture the desired antibody/DNA complexes on protein A/G magnetic beads (*see* **Note 6**). Prepare 20 μL of protein A/G magnetic beads by washing 2 times with cold FA lysis buffer. Wash by adding 20 μL of beads to 500 μL of cold FA lysis buffer with a 1 min capture on a paramagnetic stand. Then, aspirate unbound solution using care to avoid the pellet. Add the washed A/G beads to the overnight lysate antibody mix and then incubate at room temperature for 2 h with gentle rotation.
7. Wash the magnetic complexes following each reaction below using the steps as described here. Perform each wash with 500 μL of cold buffer followed by gentle agitation for 3 min at room temperature, capture on a magnetic stand for 1 min, and then aspirate of unbound solution.
8. Wash samples once with FA lysis buffer, twice with FA high salt wash buffer, twice with FA low salt wash buffer, twice with FA wash buffer 3, and once with TE + Triton-X buffer, pH 8.0. Perform one final wash with TE + Triton-X buffer, pH 8.9 to prepare the samples for exonuclease treatment.
9. Map the position of the protein/DNA cross-links in 60 μL digests containing 1× λ exonuclease reaction buffer, pH 9.4 and 10 U λ exonuclease. Incubate for 30 min at 37 °C.
10. Wash the magnetic beads samples once with TE + Triton-X buffer, pH 8.0, once with FA high salt wash buffer, once with FA low salt wash buffer, and once with FA wash buffer 3. Perform one final wash with TE + Triton-X buffer, pH 7.4 to prepare the samples for $RecJ_f$ treatment.
11. Remove contaminating DNA in a 60-μL $RecJ_f$ endonuclease digestion, containing 1× NEB buffer #2, pH 7.4 and 30 U $RecJ_f$. Incubate for 30 min at 37 °C.
12. Wash the magnetic bead samples once with TE + Triton-X buffer, pH 8.0, once with FA high salt wash buffer, once with FA low salt wash buffer, once with FA wash buffer 3, and one final additional wash with TE + Triton-X buffer, pH 8.0.
13. Recover the DNA samples from the magnetic beads in 200 μL elution buffer at 65 °C for 30 min. Capture the magnetic beads with the magnetic stand for 1 min. Transfer the DNA/protein containing supernatant to a fresh 1.7 mL microfuge tube.
14. Add proteinase K to 0.5 mg/mL final concentration, and incubate each sample overnight at 65 °C to reverse DNA-protein cross-links and digest proteins.

15. Following the cross-link reversal, purify the 200 μL DNA sample with the addition of 22 μL 3 M NaOAc, pH 5.2, 1 μL 10 mg/mL glycogen and an equal volume 1:1 phenol:chloroform. Mix samples thoroughly after the addition of phenol:chloroform. Subject samples to centrifugation at 14,000 × *g* for 5 min to separate phases. Transfer 220 μL of upper aqueous layer to a fresh tube, add 2.5 volumes 100% ethanol, and place samples on dry ice for ~15 min. Pellet DNA with centrifugation at 14,000 × *g* for 15 min, decant the supernatant, and wash the pellet with 750 μL 70% ethanol. After final centrifugation, remove all ethanol traces by pipetting. Allow the pellet to air dry at room temperature for ~10 min. Resuspend the DNA pellet in 30 μL of ultrapure water. The ChIP-exo DNA is now ready to be analyzed by LM-PCR.

3.3 Chromatin Endogenous Cleavage (ChEC) Analysis

1. Thaw cross-linked bacterial pellets on ice for 10 min. Resuspend the cell pellets on ice in 1 mL cold ChEC lysis buffer (*see* **Note 7**) supplemented with 1 mM PMFS, protease inhibitor cocktail, 100 μg/mL RNaseA, and 0.2% SDS. Transfer the cell suspension from a 50 mL conical tube to a 1.7 mL microfuge tube.
2. Lyse the cells and shear cellular DNA by sonication on ice with four 10-s bursts at 10 W. Place samples on ice between pulses for a minimum of 30 s to allow the lysates to remain cold.
3. Remove cellular debris with centrifugation at 14,000 × *g* for 3 min at 4 °C. Transfer the clarified supernatant to a fresh 1.7 mL microfuge tube.
4. For endogenous MNase cleavage, transfer 600 μL of cell lysate to a new 1.7 mL microfuge tube. Place samples at room temperature for 5 min to allow the sample to equilibrate. As a negative control, transfer 100 μL of cell lysate to a fresh tube containing an equal volume of 2× stop solution. To the remaining 500 μL of cell lysate add $CaCl_2$ to 13 mM final (65 μL of 100 mM $CaCl_2$) to activate the MNase fusion protein. Remove 100 μL samples at desired time points, typically 5, 15, 30, 45, and 60 min, and transfer to an equal volume of 2× stop solution (*see* **Note 8**).
5. Add proteinase K to 0.5 mg/mL final concentration and incubate each sample overnight at 65 °C to reverse DNA-protein cross-links and digest proteins.
6. Following cross-link reversal, purify the 200 μL DNA sample with the addition of 1 μL 10 mg/mL glycogen and an equal volume of 1:1 phenol:chloroform. Mix samples thoroughly after the addition of phenol:cholorform. Subject samples to centrifugation at 14,000 × *g* for 5 min to separate phases.

Transfer 200 μL of upper aqueous layer to a fresh tube, add 2.5 volumes 100% ethanol, and place samples on dry ice for ~15 min. Pellet the DNA with centrifugation at 14,000 × *g* for 15 min, decant the supernatant, and wash the pellet with 750 μL 70% ethanol. After final centrifugation, remove all ethanol traces with pipetting. Allow the pellet to air dry at room temperature for ~10 min. Resuspend the DNA pellet in 500 μL of ultrapure water.

7. Quantitate the DNA using a spectrophotometer, and check ~1 μg of total DNA on a 0.8% agarose gel for the quality and size distribution of cleavage products (*see* **Note 5**).
8. Block nicked DNA termini with terminal (deoxynucleotidyl) transferase (TdT) in a 50 μL reaction containing 2 μg ChEC DNA, 1X TdT buffer, 250 μM $CoCl_2$, 250 μM ddNTP mix, and 10 U TdT enzyme. Incubate the samples at 37 °C for 30 min.
9. To each 50 μL sample add 129 μL ultrapure water, 20 μL 3 M NaOAc (pH 5.2), and 1 μL 10 mg/mL tRNA.
10. Extract samples with equal volume, 200 μL, of 1:1 phenol: chloroform. Mix samples thoroughly after the addition of phenol:chloroform. Subject samples to centrifugation at 14,000 × *g* for 5 min to separate phases. Transfer 200 μL of upper aqueous layer to a fresh tube, add 2.5 volumes 100% ethanol, and place samples on dry ice for ~15 min. Pellet DNA with centrifugation at 14,000 × *g* for 15 min, decant the supernatant, and wash the pellet with 750 μL 70% ethanol. After final centrifugation, remove all ethanol traces by pipetting. Allow the pellet to air dry at room temperature for ~10 min and then resuspend in 30 μL ultrapure water. The ChEC samples are now ready to proceed with LM-PCR.

3.4 Ligation-Mediated Polymerase Chain Reaction (LM-PCR)

3.4.1 Primer Design

1. When designing primers for LM-PCR, select sequences ~150 base pairs upstream and downstream from the genomic region of interest (*see* Fig. 4). Primers 1 and 4 should have T_m values of ~50 °C, to allow for complete annealing prior to phi29 DNA polymerase extension. Primers 2 and 5 should overlap 1 and 4, respectively, by ~10 bp and have T_m values of 61–67 °C, for optimal PCR amplification with *Taq* polymerase. Primers 3 and 6 should overlap 2 and 5, respectively, by 6–10 bp and have T_m values of 58–63 °C. Primers 3 and 6 amplify the genomic DNA to create a template for radioactive probe synthesis (*see* **Note 9**).

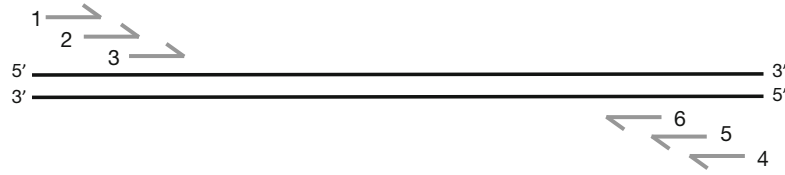


Fig. 4 Schematic of primer design in genomic region of interest. Primers 1 and 4 should have T_m values of ~50 °C, to allow for complete annealing prior to phi29 DNA polymerase extension. Primers 2 and 5 should overlap 1 and 4, respectively, by ~10 bp and have T_m values of 61–67 °C, for optimal PCR amplification with *Taq* polymerase. Primers 3 and 6 should overlap 2 and 5, respectively, by 6–10 bp and have T_m values of 58–63 °C

3.4.2 Phi29 Extension and Linker Ligation for ChIP-Exo and ChEC Samples

1. Use one third of the processed ChIP-exo or ChEC DNA samples as templates for gene-specific primer extension using phi29 DNA polymerase (*see* **Note 10**). In a 0.2 mL PCR tube, combine 11 μL of the cleaved sample (ChIP-exo or ChEC) DNA with 1× phi29 buffer in a final volume of 20 μL that also contains the following components at the indicated final concentrations: 0.2 mg/mL BSA, 125 μM dNTPs, and 1 μM primer 4.
2. In a thermal cycler, denature at 95 °C for 5 min, anneal at 60 °C for 5 min, cool the sample to room temperature, and finally add 10 U of phi29 polymerase (*see* **Note 11**). Extend the annealed DNA primer for 20 min at 30 °C, followed by heat inactivation at 65 °C for 10 min.
3. For the ChEC sample only, to the 0.2 mL PCR tube containing the 20 μL phi29 extension product, add 17 μL ultrapure water, 2 μL 10× T4 DNA ligase buffer, 0.5 μL 50 mM ATP, and 0.5 μL PNK (10 U/μL). Incubate the sample at 37 °C for 30 min. Inactivate the PNK enzyme by heating the sample for 20 min at 65 °C.
4. Cool all samples (ChIP-exo and ChEC) on ice and consolidate by brief centrifugation.
5. To each 20 μL phi29 extension reaction from ChIP-exo, add 30 μL of ChIP-exo linker ligation mix (1X T4 DNA ligase buffer, 1.7 μM unidirectional linker, 0.8 mM ATP, and 6.7 UT4 DNA ligase). To each 40 μL phi29/PNK reaction from ChEC, add 10 μL of ChEC linker ligation mix (1× T4 ligase buffer, 5 μM unidirectional linker, 1 mM ATP, and 20 U T4 DNA ligase). The linker ligation mixes each contain unidirectional linker (s*ee* **Note 12**).
6. Mix samples well and incubate overnight at 16 °C.
7. Following the overnight incubation, heat-inactivate the T4 DNA ligase for 10 min at 70 °C.

8. Transfer the ligation samples to fresh 1.7 mL microfuge tubes. Add 50 μL of ligation stop mix and 100 μL ultrapure water for a final sample volume of 200 μL.
9. To ethanol precipitate the samples add 600 μL of 100% ethanol, mix well, chill on dry ice for 15 min, and subject to centrifugation for 15 min at 15,000 × *g* in a 4 °C microcentrifuge. Carefully remove the supernatant. Wash one time with 750 μL of 70% ethanol, subject to centrifugation for 5 min at 15,000 × *g*. Carefully remove the supernatant, being careful to not disturb the pellet, and allow the samples to air dry for 5 min.
10. Resuspend the DNA pellets in 25 μL of ultrapure water. Flick gently and let it stand at room temperature for 10 min.

3.4.3 PCR Amplification of ChIP-Exo and ChEC Samples

1. For both ChIP-exo and ChEC samples, perform 19 cycles of PCR in the presence of *Taq* DNA polymerase using gene-specific primer 5, here a nested *lac* promoter-specific primer (5′-$AT_2A_2GT_2G_3TA_2CGC_2AG$), and the unidirectional linker primer (5′-$GCG_2TGAC_3G_3AGATCTGA_2T_2C$).
2. Transfer 25 μL of ligated DNA sample from a 1.7 mL microfuge tube to a fresh 0.2 mL PCR tube.
3. Prepare 2× *Taq* polymerase mix (25 μL per sample) containing 2× *Taq* polymerase buffer, 0.2 mg BSA, 4 mM $MgCl_2$, 0.4 mM dNTPs, 0.8 μM gene-specific primer 5, 0.8 μM linker primer, and 3 U *Taq* polymerase, adding the polymerase as the last step.
4. Before the addition of *Taq* polymerase, mix gently and transfer 60 μL of 2× polymerase mix to a fresh tube. To this tube add an equal volume of ultrapure water. This is now a 1× *Taq* mix and will be used as a booster mix after the initial amplification to ensure all extensions are complete. Place this 1× *Taq* mix on ice until ready for use.
5. Once *Taq* polymerase has been added to 2× polymerase mix, transfer 25 μL of this mix to each PCR sample and mix gently.
6. Perform PCR using an annealing temperature that is optimal for gene-specific primer 5. After an initial denaturation step at 95 °C for 3 min, perform 19 cycles of amplification at 95 °C for 30 s, 57 °C for 30 s, and 72 °C for 1 min.
7. When PCR amplification is complete add 5 μL of 1× *Taq* mix with added *Taq* polymerase to each sample. This booster mix contains 1 U of *Taq* polymerase per 5 μL of 1× *Taq* mix. Further incubate the reactions at 72 °C for 10 min to ensure completion of all extensions.
8. To each reaction add 70.5 μL *Taq* stop mix. Transfer the samples to fresh 1.7 mL microfuge tubes and add 190 μL of

1:1 phenol:chloroform, followed by vortex mixing for 5 s. Subject samples to centrifugation at 15,000 × *g* for 5 min to separate phases. Transfer 125 μL of the upper aqueous layer to a fresh 1.7 mL microfuge tube and add 330 μL 100% ethanol. Ethanol precipitate by mixing well, chilling on dry ice for 15 min, and subjecting to centrifugation in a microcentrifuge at 4 °C for 15 min at 15,000 × *g*. Carefully remove the supernatant. Wash one time with 750 μL of 70% ethanol, subject to centrifugation for 5 min at 15,000 × *g*. Carefully remove the supernatant, being careful to not disturb the pellet, and let the samples air dry at room temperature for 5 min.

9. Resuspend each sample in 6 μL deionized form dye mix.

3.5 Southern Blot Analysis

3.5.1 Generation of Sequencing Ladders as Markers Within the Gel Blot

1. Prior to Maxam-Gilbert base-specific reactions [25], ethanol precipitate ~40 μg of purified genomic DNA: add 600 μL of 100% ethanol, mix well, chill on dry ice for 15 min, and subject to centrifugation at 4 °C for 15 min at 15,000 × *g*. Carefully remove the supernatant. Wash one time with 750 μL of 70% ethanol, and subject to centrifugation for 5 min at 15,000 × *g*. Carefully remove the supernatant, being careful to not disturb the pellet, and allow the samples to air dry for 5 min. Resuspend the dried pellets in an appropriate amount of ultrapure water for the base-specific reactions that follow.
2. *G reaction chemistry:* Resuspend DNA pellet in 5 μL ultrapure water and let it stand at room temperature for 10 min. On ice, combine 5 μL of genomic DNA solution (~40 μg), 200 μL of DMS buffer, and 1 μL DMS (stored at −80 °C under argon). Incubate at room temperature for 4 min. Terminate the reaction by adding 50 μL of DMS stop buffer. Mix by pipetting a few seconds and then add 750 μL of cold 100% ethanol. Place the sample on dry ice.
3. *G + A reaction chemistry:* Resuspend DNA pellet in 11 μL ultrapure water and let it stand at room temperature for 10 min. On ice, combine 11 μL of genomic DNA solution (~40 μg) and 25 μL of cold formic acid (stored at 4 °C). Incubate at room temperature for 10 min. Add 200 μL of DMS stop buffer. Mix by pipetting a few seconds and then add 750 μL of cold 100% ethanol. Place the sample on dry ice.
4. *C > T reaction chemistry:* Resuspend DNA pellet in 5 μL ultrapure water and let it stand at room temperature for 10 min. Mix on ice 5 μL of genomic DNA (~40 μg) and 15 μL 5 M NaCl. Mix well before adding 30 μL of cold hydrazine (keep frozen at −20 °C and minimize exposure to light). Incubate at room temperature for 7.5 min. Terminate reaction by adding 200 μL

of hydrazine stop buffer. Mix by pipetting a few seconds and then add 750 μL of cold 100% ethanol. Place the sample on dry ice.

5. Keep all samples in dry ice for at least 15 min.
6. Warm samples to room temperature. Subject to centrifugation for 15 min at ~15,000 × *g* at room temperature.
7. Remove the supernatant, subject to centrifugation briefly, and carefully remove all remaining liquid.
8. Resuspend the pellets in 270 μL of ultrapure water. Make sure pellets are completely dissolved.
9. Ethanol precipitate as follows. Add 1/10 volume 3 M NaOAc, pH 5.2 (30 μL) and then 2.5 volumes (750 μL) of cold 100% ethanol. Mix the samples and place on dry ice for at least 15 min. Warm the samples to room temperature. Subject to centrifugation for 15 min at 15,000 × *g* at room temperature. Remove supernatants with a pipette. Wash pellets with 1 mL 70% cold ethanol. Subject samples to centrifugation at 15,000 × *g* for 5 min at room temperature. Aspirate the supernatant with a pipette, subject to centrifugation briefly, and carefully remove all remaining liquid. Let the pellets air dry at room temperature for ~5 min.
10. For *G reaction chemistry*, dissolve pellets in 100 μL 1 M piperidine (10%). For *G + A reaction chemistry*, dissolve pellets in 100 μL 0.1 M piperidine (1%). For *C > T reaction chemistry*, dissolve pellets in 100 μL 1 M piperidine (10%).
11. Heat the samples at 90 °C for 30 min (either apply cap locks to tubes or cover with a heavy object to prevent the lids from popping open due to increased pressure).
12. Cool samples completely and subject to centrifugation briefly to consolidate droplets. Apply parafilm to the open tubes. Puncture the parafilm 4–6 times to give ventilation holes. Freeze samples on dry ice.
13. Place samples in lyophilizer overnight to remove all traces of piperidine.
14. Resuspend samples to a concentration of 1 μg/μL. Assume a 10% loss of sample due to treatment and processing when determining amount of ultrapure water to use to resuspend.
15. Before use for LM-PCR, electrophorese ~1 μg of the piperidine-treated sample on a 1.5% agarose gel along with 100 bp ladder to confirm the cleavage profile. Optimal samples for LM-PCR have the bulk of the cleavage in the range of 300–600 nucleotides.
16. Quantify treated DNA and dilute to a final concentration of 20 ng/μL. For Maxam-Gilbert reference ladders, use 11 μL during a standard LM-PCR method.

3.5.2 Gel Electrophoresis of ChIP-Exo and ChEC Samples and Probe Synthesis

1. Resolve the amplified DNA samples on a 6% denaturing polyacrylamide (19:1 acrylamide:bisacrylamide) 45 cm × 35 cm sequencing gel.
2. Pre-run the gel in 0.5× TBE buffer at 80 W until gel temperature reaches ~45 °C.
3. Denature the formamide dye DNA samples at 95 °C for 3 min, followed by immediate placement on ice.
4. Load the samples and sequencing ladders (*see* **Note 13**).
5. Electrophorese the samples at 85 W, making sure that the gel temperature does not go above 55 °C. The size of the samples will determine electrophoresis time. For the *lac* promoter sequence the 6% gel electrophoresis continues for one additional hour after the xylene cyanol dye reaches the bottom of the gel.
6. During electrophoresis of the DNA samples, the radioactive probe for the Southern blot can be prepared using single-strand DNA extension with gene-specific primer 6, 5′-$G_2TA_2CGC_2AG_3T_4C$ for the *lac* promoter region.
7. In a 0.2 mL PCR tube, assemble a 100 μL probe synthesis reaction containing 0.8 ng DNA template from primer 3/6 amplification product, 2 mM $MgCl_2$, 10 μM dCTP, 10 μM dGTP, 10 μM dTTP, 1× *Taq* polymerase buffer, 1 μM gene-specific primer 3, 0.03 U *Taq* polymerase, and 10 μL α-^{32}P dATP (3000 Ci/mmol; 10 mCi/mL). Perform the following PCR protocol for the labeling reaction using an annealing temperature equal to the T_m of gene-specific primer 6. Cycle at 95 °C for 1 min, 61 °C for 1 min, and 75 °C for 2 min for 35 cycles.
8. When probe synthesis is complete, split the radiolabeled probe into two samples and pass over chroma-spin + TE columns to remove unincorporated radioactive nucleotides.

3.5.3 DNA Transfer and Southern Blotting Protocol

1. Transfer the sequencing gel to a Zeta-probe nylon membrane using an Owl semi-wet transfer system. To the center of the bottom metal plate, add ~20 mL of 0.5× TBE buffer (same as gel running buffer).
2. Onto this plate place two layers of pre-wet 17 mm Whatman paper. Use a rolling pin to remove all the air bubbles. Add an additional 40 mL of buffer to the top of the blot paper.
3. Add one additional pre-wet piece of 17 mm Whatman paper, starting with contact in the center and then out to the edges. Roll again to remove air bubbles.
4. Evenly distribute 50 mL 0.5× TBE buffer to top of blot paper.
5. Place gel onto blot paper. Carefully add from center out to edges, taking care that bubbles are not introduced.

6. Evenly add 20 mL of buffer on center of gel.
7. Place pre-wet membrane (center out to edges) without introducing air bubbles.
8. Add one sheet pre-wet 17 mm Whatman blot paper. Use a rolling pin with minimal pressure to remove air bubbles.
9. Add 50 mL of buffer to the top of blot paper and then add two sheets pre-wet 17 mm Whatman blot paper, and roll out air bubbles with rolling pin using moderate pressure (always roll from center out).
10. Perform transfer at 1.6 amp and ~12 V for 50 min.
11. Following the transfer, cross-link the DNA to the membrane using the auto-crosslink setting of a UV Stratalinker unit.
12. Roll the cross-linked membrane, DNA side in, into a hybridization tube containing 14 mL of hybridization buffer, warmed to 55 °C. Allow the membrane to block and pre-hybridize for 1 h at 62.4 °C.
13. After pre-hybridization, decant the pre-hybridization buffer and add 6 mL of fresh 55 °C hybridization buffer and 100 μL of gene-specific radiolabeled probe. Allow the membrane and probe to hybridize overnight at 62.4 °C.

3.5.4 Blot Washing and Exposure

1. Following the overnight hybridization, remove the membrane from the hybridization tube and place in a wash container that can hold ~500 mL of liquid. The membrane should also be able to lay completely flat in the container.
2. Wash the radioactive membrane, one time with 400 mL wash buffer 1 and two times with 400 mL wash buffer 2. Perform each wash at room temperature, with buffer warmed to 65 °C, for 5 min with gentle agitation (*see* **Note 14**).
3. When the last membrane wash is complete, allow the membrane to air dry for ~5 min. Wrap the membrane in plastic-wrap and expose to a storage phosphor screen overnight.
4. Perform imaging on a Typhoon FLA 7000 (GE LifeSciences). Representative images from both ChIP-exo and ChEC are shown in Fig. 5.

4 Notes

1. When creating MNase fusion proteins it is important to confirm both the retained function/binding properties of the desired mapping protein and MNase functionality. We have always added MNase to the C terminus of mapping proteins with success.

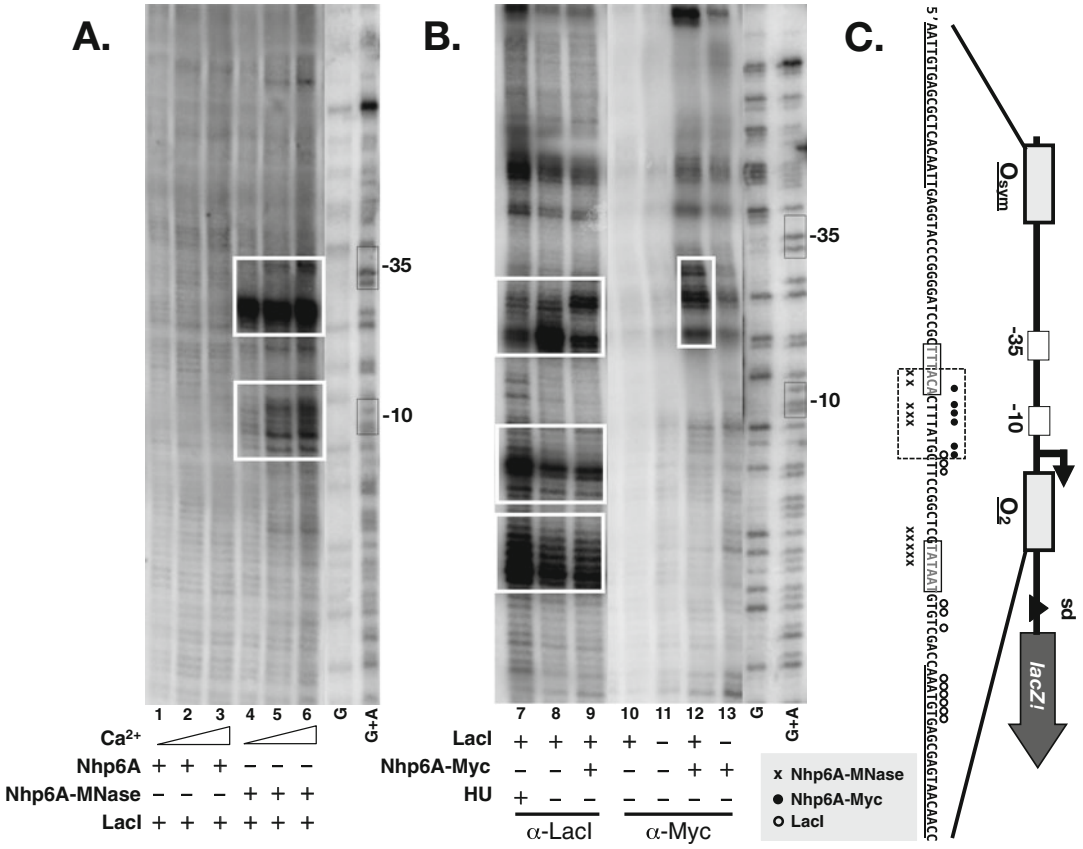


Fig. 5 ChEC and ChIP-exo in vivo protein-mapping Southern blots of engineered *E. coli* strains exogenously expressing the yeast architectural protein Nhp6A. Schematic map of the engineered *lac* promoter region analyzed with high-resolution mapping, far right. DNA operators, $\underline{O_{sym}}$ and $\underline{O_2}$, are sequences that are bound by Lac repressor (LacI) in the absence of allolactose analog, IPTG, to allow repression loop formation. Cis elements include -35 and -10 (open boxes), $+1$ (bent arrow), and Shine-Dalgarno (black triangle). (**a**) Nhp6A-MNase fusion mapping around the *lac* promoter region with ChEC-LM-PCR. Bacterial lysates expressing either Nhp6A (lanes 1–3) or Nhp6-MNase fusion (lanes 4–6) were given 10 mM Ca^{2+} for 0, 5, or 10 min. Nhp6A without the MNase fusion is added as a control for potential background cleavage, which as seen in lanes 1–3 is minimal. MNase cleavage regions of particular interest are indicated (white boxes). The flexibility of the fusion allows for some MNase cleavage on either side of the binding site, though preferential cleavage is seen in the upper white box. (**b**) ChIP-exo-LM-PCR mapping of protein binding sites for LacI, and c-Myc tagged Nhp6A. Bacterial lysates were immunoprecipitated with either α-LacI (lanes 7–9) or α-Myc (lanes 10–12). ChIP-exo protein-binding boundaries, of interest, are indicated (white boxes). Endogenous LacI robustly binds to the $\underline{O_2}$ operator, as expected, and this binding is attenuated in the absence of the endogenous architectural protein, HU, deleted in the indicated samples (middle and lower white boxes). The signal detected at the Nhp6A binding site, upper white box, is likely a cross-linking bystander effect, since this signal is increased in the presence of Nhp6A-Myc (lane 9). Immunoprecipitation with α-Myc definitively shows the Nhp6A-Myc binding site (white box at right). Only samples in lanes 11 and 13 were grown in the presence of IPTG, inhibiting LacI binding to operator DNA sequences, as indicated. Nhp6A-Myc binding appears largely dependent on the formation of the Lac repression loop (lane 12). C. Summary of high-resolution data from ChEC and ChIP-exo mapped onto the engineered *lac* promoter sequence. Protein-binding boundaries are identified for ChEC (Nhp6A-MNase, crosses) and ChIP-exo (LacI, open circles and Nhp6A-Myc, closed circles). The Nhp6A binding site is indicated within the dashed black box. All data shown are representative of at least three replicates [16]. Sequencing ladders (G and G + A) were created by standard Maxam-Gilbert chemical modifications of genomic DNA in vitro

2. Use of fresh formaldehyde is best to ensure proper cross-linking with monomeric formaldehyde. Over time formaldehyde can form paraformaldehyde polymers, which are not as reactive or cell-permeable. Do not use a formaldehyde solution that contains a white precipitate.
3. It is necessary to quench unreacted formaldehyde after the cross-linking reaction. Glycine has been the typical quenching agent, but Tris has recently been shown to be more efficient [26].
4. We recommend using 50% power on a model 60 sonic dismembrator (Fisher Scientific).
5. Following lysis and sonication of samples for both ChIP-exo and ChEC it is important to check the quality and size distribution of the harvested bacterial DNA. A size distribution of DNA fragments between 300 bp and 1000 bp offers best results for LM-PCR analysis with a desired resolution of ~100–150 bp.
6. The use of magnetic beads allows for a rapid and easy method to change buffers and purify samples between enzymatic steps.
7. Addition of EGTA is necessary to control for early activation of MNase during the initial lysis steps.
8. For best results with ChEC, the 2× stop solution should be made fresh prior to each use.
9. Primer design considerations are important for the success of LM-PCR. All primers used for LM-PCR should be gel purified to ensure that the oligonucleotides used for all extension/amplification steps are full length and that any shorter, failed syntheses are not present in the oligonucleotide stocks. It is important to perform a trial PCR reaction with the designed PCR pairs to ensure that each pair generates the single desired PCR product from a genomic DNA template. Heterogeneous PCR products will degrade LM-PCR results. We find that primers designed to a region of ~300 bp will yield ~100 bp or ~150 bp resolution on 8% or 6% denaturing gels, respectively.
10. The choice of which primer set, 1–3 or 4–6, depends on which strand of the DNA is desired for analysis. Each strand can be analyzed in separate LM-PCR processing steps from the same initial ChIP-exo or ChEC samples. The present example analysis uses primers 4–6, where primer 4 (5′-GCA_2G_2C$GAT_2A_2GT_2$G) is gene-specific to the *lac* promoter region. DNA cleavage with MNase leaves 5′ hydroxyl and 3′ phosphate termini so that in order to proceed with LM-PCR, we treat the phi29 extension with polynucleotide kinase (PNK) to create phosphorylated 5′ termini. Otherwise, the steps are identical for the two sample types.

11. To ensure that the gene specific region is fully extended to the cleavage site, phi29 DNA polymerase is used because it has good strand displacement and high processivity.
12. The unidirectional linker is annealed by combining 20 μM of LJM-648 (5′-GCG$_2$TGAC$_3$G$_3$AGATCTGA$_2$T$_2$C), 20 μM of LJM-649 (5′-GA$_2$T$_2$CAGATC), and 250 mM Tris–HCl, pH 7.6. Heat the mixture to 95 °C for 5 min, then transfer to a 70 °C water bath. Allow the water bath to cool to room temperature over 1 h and leave at room temperature for an additional hour. Further cool the linker to 4 °C over 1 h and leave at 4 °C overnight in a water bath. The annealed linker can now be frozen at −20 °C in aliquots. Avoid repeated freeze thaw cycles of the linker aliquots and always thaw on ice, to ensure the unidirectional linker remains annealed.
13. At the start of gel electrophoresis it is important to include sequencing marker samples (G and G + A ladders) created by standard Maxam-Gilbert chemical modifications of purified genomic DNA. This allows for the orientation of the genomic region within samples and identification of cleavage locations.
14. Care must be taken when working with radioactivity and in properly disposing of radioactive liquid waste.

Acknowledgments

This work was supported by the Mayo Foundation and by National Institutes of Health grant GM75965 to LJM.

References

1. Adhya S (1989) Multipartite genetic control elements: communication by DNA loop. Annu Rev Genet 23:227–250
2. Garcia HG, Grayson P, Han L, Inamdar M, Kondev J, Nelson PC, Phillips R, Widom J, Wiggins PA (2007) Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. Biopolymers 85 (2):115–130
3. Peters JP 3rd, Maher LJ (2010) DNA curvature and flexibility in vitro and in vivo. Q Rev Biophys 43(1):23–63
4. Peters JP, Becker NA, Rueter EM, Bajzer Z, Kahn JD, Maher LJ 3rd (2011) Quantitative methods for measuring DNA flexibility in vitro and in vivo. Methods Enzymol 488:287–335
5. Bellomy GR, Mossing MC, Record MT Jr (1988) Physical properties of DNA in vivo as probed by the length dependence of the lac operator looping process. Biochemistry 27 (11):3900–3906
6. Muller J, Oehler S, Muller-Hill B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. J Mol Biol 257(1):21–29
7. Becker NA, Kahn JD, Maher LJ 3rd (2005) Bacterial repression loops require enhanced DNA flexibility. J Mol Biol 349(4):716–730
8. Bond LM, Peters JP, Becker NA, Kahn JD, Maher LJ 3rd (2010) Gene repression by minimal lac loops in vivo. Nucleic Acids Res 38 (22):8072–8082
9. Becker NA, Peters JP, Maher LJ 3rd, Lionberger TA (2013) Mechanism of promoter repression by lac repressor-DNA loops. Nucleic Acids Res 41(1):156–166
10. Becker NA, Greiner AM, Peters JP, Maher LJ 3rd (2014) Bacterial promoter repression by DNA looping without protein-protein binding competition. Nucleic Acids Res 42 (9):5495–5504

11. Mogil LS, Becker NA, Maher LJ 3rd (2016) Supercoiling effects on short-range DNA looping in E. coli. PLoS One 11(10):e0165306
12. Becker NA, Kahn JD, Maher LJ 3rd (2007) Effects of nucleoid proteins on DNA repression loop formation in Escherichia coli. Nucleic Acids Res 35(12):3988–4000
13. Becker NA, Kahn JD, Maher LJ 3rd (2008) Eukaryotic HMGB proteins as replacements for HU in E. coli repression loop formation. Nucleic Acids Res 36(12):4009–4021
14. Sebastian NT, Bystry EM, Becker NA, Maher LJ 3rd (2009) Enhancement of DNA flexibility in vitro and in vivo by HMGB box A proteins carrying box B residues. Biochemistry 48 (10):2125–2134
15. Czapla L, Peters JP, Rueter EM, Olson WK, Maher LJ 3rd (2011) Understanding apparent DNA flexibility enhancement by HU and HMGB architectural proteins. J Mol Biol 409 (2):278–289
16. Becker NA, Maher LJ 3rd (2015) High-resolution mapping of architectural DNA binding protein facilitation of a DNA repression loop in Escherichia coli. Proc Natl Acad Sci U S A 112(23):7177–7182
17. Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147(6):1408–1419
18. Rhee HS, Pugh BF (2012) ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Curr Protoc Mol Biol Chapter 21:Unit 21 24
19. Pfeifer GP, Steigerwald SD, Mueller PR, Wold B, Riggs AD (1989) Genomic sequencing and methylation analysis by ligation mediated PCR. Science 246(4931):810–813
20. Mueller PR, Wold B (1989) In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246(4931):780–786
21. Becker NA, Maher LJ 3rd (1999) LMPCR for detection of oligonucleotide-directed triple helix formation: a cautionary note. Antisense Nucleic Acid Drug Dev 9(3):313–316
22. Mueller PR, Wold B, Garrity PA (2001) Ligation-mediated PCR for genomic sequencing and footprinting. Curr Protoc Mol Biol Chapter 15:Unit 15 13
23. Schmid M, Durussel T, Laemmli UK (2004) ChIC and ChEC; genomic mapping of chromatin proteins. Mol Cell 16(1):147–157
24. Goetze H, Wittner M, Hamperl S, Hondele M, Merz K, Stoeckl U, Griesenbeck J (2010) Alternative chromatin structures of the 35S rRNA genes in Saccharomyces cerevisiae provide a molecular basis for the selective recruitment of RNA polymerases I and II. Mol Cell Biol 30(8):2028–2045
25. Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65(1):499–560
26. Hoffman EA, Frey BL, Smith LM, Auble DT (2015) Formaldehyde crosslinking: a tool for the study of chromatin complexes. J Biol Chem 290(44):26404–26411

Chapter 7

Imaging of Transcription and Replication in the Bacterial Chromosome with Multicolor Three-Dimensional Superresolution Structured Illumination Microscopy

Carmen Mata Martin, Cedric Cagliero, Zhe Sun, De Chen, and Ding Jun Jin

Abstract

Superresolution imaging technology has contributed to our understanding of the subnucleoid organization in *E. coli* cells. Multicolor superresolution images revealing "bacterial nucleolus-like structure or organization," "nucleolus-like compartmentalization of the transcription factories," and "spatial segregation of the transcription and replication machineries" have enhanced our understanding of the dynamic landscape of the bacterial chromatin. This chapter provides a brief introduction into multicolor three-dimensional superresolution structured illumination microscopy (3D-SIM) used to study the spatial organization of the transcription machinery and its spatial relationship with replisomes from a microbiological research perspective. In addition to a detailed protocol, practical considerations are discussed in relation to (1) sampling and treatment of cells containing fluorescent fusion proteins, (2) imaging the transcription and replication machineries at single-cell levels, (3) performing imaging experiments to capture the spatial organization of the transcription machinery and the nucleoid, and (4) image acquisition and analysis.

Key words Bacterial nucleolus-like structure or organization, Transcription machinery, RNA polymerase, Replisomes, Three-dimensional structured illumination microscopy, Superresolution imaging, *E. coli*

1 Introduction

Shortly following the first applications of green fluorescent protein (GFP) technology in eukaryotes in the early 1990s, this important tool was used in bacteria to image the transcription and replication machineries at the single-cell level [1–4]. *Escherichia coli* (*E. coli*) as a simple model system has many advantages for imaging studies, mainly because it is one of the most thoroughly studied organisms, particularly in terms of genetics, physiology, and biochemistry. Due to its rapid response it is a powerful tool to investigate the effects

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-1-4939-8675-0_7) contains supplementary material, which is available to authorized users.

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_7,

of environmental signals and system perturbations. First, an *E. coli* chromosomal *rpoC-gfp* fusion was constructed and used for studies of RNA polymerase (RNAP) during rapid growth and stress response [5, 6]. Since then, many derivatives of the RNAP-GFP reporter have been constructed to study different aspects of RNAP in *E. coli* cells [7–11]. Fluorescent fusions to components of *E. coli* replisomes at replication forks, such as SSB and SeqA, have also been constructed to study the replication machinery in the cell [12, 13].

Until recently, imaging the transcription machinery and replisomes relied on widefield fluorescent microscopy, which is restricted by a diffraction limit of ~250 nm in the lateral dimension and ~500 nm in the axial dimension for GFP. Because *E. coli* cells are small (~1 μm [short axis] × ~1 to 5 μm [long axis] on average in growing cells with different growth media and temperatures) and the nucleoids inside cells grown in different media and environments are even smaller [14], it is challenging and/or impossible to study subnucleoid organization in detail using conventional widefield microscopy. Recently, superresolution microscopy has been established as a powerful tool for microbial cell biology studies [15]. For instance, photoactivated localization microscopy (PALM) or stochastic optical reconstruction microscopy (STORM) has been used to determine the distribution and movement of RNAP at the single-molecule level in living cells [7, 9, 11]. However, many PALM and STORM instruments currently are limited to one-color imaging in bacteria. In contrast, three-dimensional structured illumination microscopy SIM (3D-SIM), which increases the lateral resolution by a factor of two compared to classical widefield fluorescent microscopy, is capable of simultaneous multiple-color imaging [16, 17]. 3D-SIM has proven to be particularly useful in studying the spatial relationship between the transcription and replication machineries in the nucleoid of *E. coli* [18, 19].

Obtaining high resolution images of fast-growing bacterial cells is a challenge inherent to *E. coli* physiology. Both the distribution of *E. coli* RNAP and the nucleoid structure are extremely sensitive to perturbations in the environment. It has been shown that procedures as short as 5–12 min involved in sampling of fast-growing cells from a shaking flask in a water bath to mounting the cells onto a microscopy slide followed by imaging will result in changes in growth environments and cause perturbation of bacterial physiology. Consequently, the resultant live-cell images will be an adapted state, rather than images of the fast-growing cells as intended [6, 10]. This problem has been underappreciated in the field and it is important to consider potential changes in bacterial physiology when performing live-cell imaging or when interpreting live-cell images. To overcome this problem, a procedure using formaldehyde to fix cells immediately after sampling was developed to capture snapshots of the dynamic organization of transcription machinery and the nucleoid in changing environments. The fixed-

cell procedure is able to "freeze" or immobilize RNAP and other subcellular structures before being imaged [6]. The fixed-cell images are consistent with extensive studies of *E. coli* genetics and physiology [10, 14].

In this chapter, we provide a detailed protocol for the application of three-color 3D-SIM to study the spatial organizations of the transcription machinery and replisomes in fast-growing cells. We describe the strains used for these studies, followed by a detailed description of cell sampling procedures and image acquisition by SIM. Finally, we describe image reconstitution and analysis using imaging tools such as Nis-Elements Ar/Nis-Elements C from Nikon, Matlab, and Fiji J.

2 Materials

2.1 Chromosomal Fluorescent Protein Fusions to the Transcription Machinery and Replisomes

2.1.1 Bacterial Strains

All strains are derivatives of E. coli MG1655.

1. CC72: MG1655 *rpoC*-Venus [8].
2. CC341: CC72 *nusA*-mCherry::Kan^R [18].
3. CC376: CC72 *ssB*-mCherry:: Kan^R [18].

2.2 Bacterial Culture and Fixation Reagent

1. Luria Broth (LB) medium: tryptone 10 g/L, yeast extract 5 g/L, NaCl 5 g/L. Adjust the pH to 7.0.
2. Formaldehyde solution for molecular biology (Sigma-Aldrich F8775), 36.5–38% in H_2O.
3. Heated water bath with agitation (37 °C, 225 rpm).
4. Culture flasks.
5. Tubes for bacterial growth (15 mL culture tube).
6. UV/Visible spectrophotometer.

2.3 Materials for Microscope Slide and Cover Glass

1. Wash medium (Supplemented M63 medium): 100 mL 10× M63 medium, 1 mL 0.1 M CaCl2, 1 mL 1 M $MgSO_4$, distilled H_2O (dH_2O) up to 1 L. Sterilize with autoclave. Recipe for 1 L 10× M63 medium [20]: 20 g $(NH_4)_2SO_4$, 136 g KH_2PO_4, 5 mg $FeSO_4.7H_2O$, dH_2O up to 1 L and adjust to pH 7 with KOH. Aliquot to 100 mL/bottle. Sterilize with autoclave.
2. Poly-L-lysine solution, 0.1% (w/v) in H_2O.
3. 70% (v/v) solution of ethanol in water.
4. Microscope slides (Globe Scientific Inc. 1301), 25 × 75 × 1 mm ½ gross.
5. Microscope cover glass high precision (Denville Scientific Inc. M1100-15), size 22 × 22 mm, tolerance 170 ± 5 μm, No.1.5H.

6. Hoechst 33342 10 μg/mL.
7. Mounting medium: 9 mL glycerol, 1 mL Tris–HCl (1 M pH 8.5) and 1 mg/mL poly (diallyldimethylammonium chloride (PDDA)). Add H_2O up to 10 mL. Aliquot and store in dark at −80 °C.
8. 0.1 μm TetraSpeck microspheres fluorescent blue/green/orange/dark red (Life technologies, Molecular Probes 7279), used for calibration instrument optics in 3D-SIM.
9. Nail polish for sealing edges of cover glass on microscope slides.

2.4 Microscope

1. Nikon N-SIM Ti-2E inverted microscope.
2. LU-NV series laser unit, 405 nm, 488 nm, 561 nm, 640 nm.
3. Objective CFI SR HP Apochromat TIRF 100XC Oil (NA 1.49).
4. Camera ORCA-Flash 4.0 sCMOS camera (Hamamatsu Photonics K.K).
5. Software NIS-Elements Ar/NIS-Elements C with additional software modules NIS-A 6D and N-SIM Analysis.
6. Operations conditions, maintaining the temperature in a range 20 °C–28 °C (±0.5 °C) at the working room.
7. Inmersion oil NF type 2 Nikon (1.515 reflective index, 40.8 dispersion index, 4500 st viscosity, 23 °C).
8. 100% pure spun cellulose fiber optical lens tissue (Electron Microscopy Sciences 71712-01).
9. Nikon lens cleaner solution.

2.5 Alignment and Reconstruction

1. Software NIS-Elements Ar/NIS-Elements C.

2.6 Data

1. MATLAB (Mathworks).
2. ImageJ/Fiji and Adobe Photoshop.

3 Methods

3.1 Cell Culture, Sampling and Fixation

1. Inoculate the strain (from the frozen stock at −80 °C) in 3 mL LB and grow overnight with vigorous agitation at 37 °C.
2. Dilute the overnight culture 1:200 into fresh LB medium in a culture flask at 37 °C in a water bath with vigorous and constant agitation.
3. At early exponential growth phase point (~OD_{600} 0.2, no more than 0.3) transfer 900 μL cell culture from the flask into a 1.5 mL Eppendorf tube containing 100 μL of Formaldehyde 37% (final concentration 3.7%) to fix cells. Keep at room temperature for at least 20–30 min.

3.2 Slide Preparation

Unless otherwise mentioned, the procedures are performed at room temperature (23 °C).

1. After 20–30 min fixation, spin the cells at 3300 × *g* for 1 min.
2. Wash the pellet once with 1 mL supplemented M63 medium.
3. Spin the cells at 3300 × *g* for 1 min.
4. Resuspend the cells in 20 μL supplemented M63 medium.
5. Add 2 μL Hoechst 33,342 10 μg/mL and incubate in dark for ~10–15 min.
6. Add appropriate amount of supplemented M63 medium; the amount should be adjusted according to the cell concentration in the sample. Use 200 μL supplemented M63 medium for cells from OD_{600} 0.2 cultures.
7. Wash the microscope slides, first with ethanol 70% and then with water. Let slides dry.
8. Dilute the 0.1 μm TetraSpeck fluorescent microspheres 1:1000 in dH_2O.

   It is important that the microsphere beads are uniformly suspended by mixing on a vortex mixer for at least 1 min before use.
9. Drop 8–10 μL of poly-L-lysine onto a cover glass. Let dry passively for 20 min.
10. Drop 2 μL of sample of cells on top of the poly-L-lysine. Let dry passively for 15 min.
11. Drop 1 μL of TetraSpeck microspheres dilution at one corner of the cover glass. Let dry passively for 15 min.
12. Drop 6–8 μL of mounting medium at the center of the microscope slide.
13. Put the cover glass (with cells sample face down) on top of the mounting medium on the microscope slide. Press gently but firmly on a flat surface.
14. Seal all sides of cover glass on the microscope slide with nail polish.
15. Visualize on microscope, or keep in dark at −20 °C

3.3 Microscope Instrumentation

1. Turn on the camera, lasers, microscope and computer at least 30 min before the beginning of image acquisition to allow the camera to reach the targeted temperature (−80 °C).
2. Add a drop of immersion oil on the top of cover glass of the microscope slide and then place the cover glass of the slide on top of the lens.

3. Focus the cells.
4. Open the software and adjust the settings.

   Optimized acquisition settings are as follows:

   – mCherry, 60 ms exposure with 561 nm laser (556–566 nm excitation range, 30% transmission). Emission range: 570–640 nm. EM Gain Multiplier 250.
   – Venus, 200 ms exposure with 488 nm laser (470–490 nm excitation range, 30% transmission). Emission range: 500–545 nm. EM Gain Multiplier 250.
   – Hoechst 33342, 40 ms exposure with 405 nm laser (395–415 nm excitation range, 99% transmission). Emission range: 435–485 nm. EM Gain Multiplier 250.

   For 3D-SIM imaging, image stacks should be composed of at least 15 Z-sections of 0.1 μm to acquire to whole cell (sample thickness of 1.4 μm).
5. After acquisition clean the objective lens using optical lens tissue. Sample slides can be cleaned by removing the immersion oil from the cover glass using Nikon lens cleaner solution and then stored at −20 °C in dark.

3.4 Alignment and Reconstruction

1. Images are processed for optimal alignment using calibration microsphere beads in the images to correct for chromatic aberration with the Software NIS-Elements Ar/NIS-Elements C. The TetraSpeck fluorescent microspheres facilitate the calibration process. It is important to verify the colocalization and resolve objects emitting different wavelengths of light in the same optical plane. The calibration profile of one image can be used to realign all the images from the same acquisition day.
2. Image reconstruction is carried out using Software NIS-Elements Ar/NIS-Elements C. To avoid introducing artifacts, it is important to carefully set the following parameters:
   – Illumination Modulation Contrast.

     Typical value setting 0.70.
   – High Resolution Noise Suppression.

     Typical value setting 1.
   – Out-of-Focus Blur Suppression.
     Typical value setting 0.05.

   This should be determined by the user because the quality of the acquired images is different and is affected by different factors. The user should set slice reconstruction parameters on the N-SIM Pad tab sheet.

3.5 Image Processing and Editing

Use ImageJ/Fiji to crop representative isolated cells. Analysis requires at least ~100–150 cells, but the more, the better.

For each cell, three formats of image are necessary:

- TIFF format photo 16-bit, with all channels (c) and all slices per channels (z), used for image analysis. Do not merge the channels and slices.
- TIFF format photo 16-bit maximum intensity projection, one slice (z) per each channel (c). Useful for editing. Do not merge the channels.
- TIFF format photo 16-bit best slices (z) per each channel (c). Useful for editing. Do not merge the channels.

Three sets of representative 2D images of the transcription and replication machineries in the bacterial chromatin in fast-growing cells from the three strains are described in the procedure (Fig. 1a–c). Because 3D-SIM acquires images in Z-stack, the information in XYZ dimensions can be used to build volume and obtain rotation projection information with movie editor software. Using the software NIS-Elements Ar/NIS-Element C with the navigation axis or navigation cube option, 3D model movies can be generated. Three 3D model Videos (A, B, and C), generated from 3D-SIM images of the three strains are presented in supplement.

3.6 Analysis

Images are processed and analyzed using Image Processing ToolBox™ MathWorks written in MATLAB context, an image analysis program that has been written for the purpose of analyzing 3D microscope images by foci detection and colocalization.

Foci detection. Clustering algorithms identify the foci localizations that represent the spatial organization of the proteins in cells, including k-means, nearest neighbor clustering, and point-correlation analysis.

Protein colocalization analysis. The colocalization frequency of a particular protein-mCherry fusion and RNAP-Venus in each cell is measured using a subcellular colocalization analysis where the positions of the weighted centroids of the foci in the two channels are determined and the distance between each mCherry-weighted centroid and its closest RNAP-Venus-weighted centroid is recorded.

Two foci are scored as colocalized if the distance between the two centroids is below the theoretical microscope resolution for mCherry (140 nm).

The colocalization frequency is represented as the fraction of mCherry foci which colocalizes with at least one of the transcription foci (RNA-Venus) in a cell and the central tendency of the distribution in a population of cells is represented by the median.

Colocalization of mCherry foci and RNAP-Venus foci is also represented as a cumulative distribution of pairwise distance.

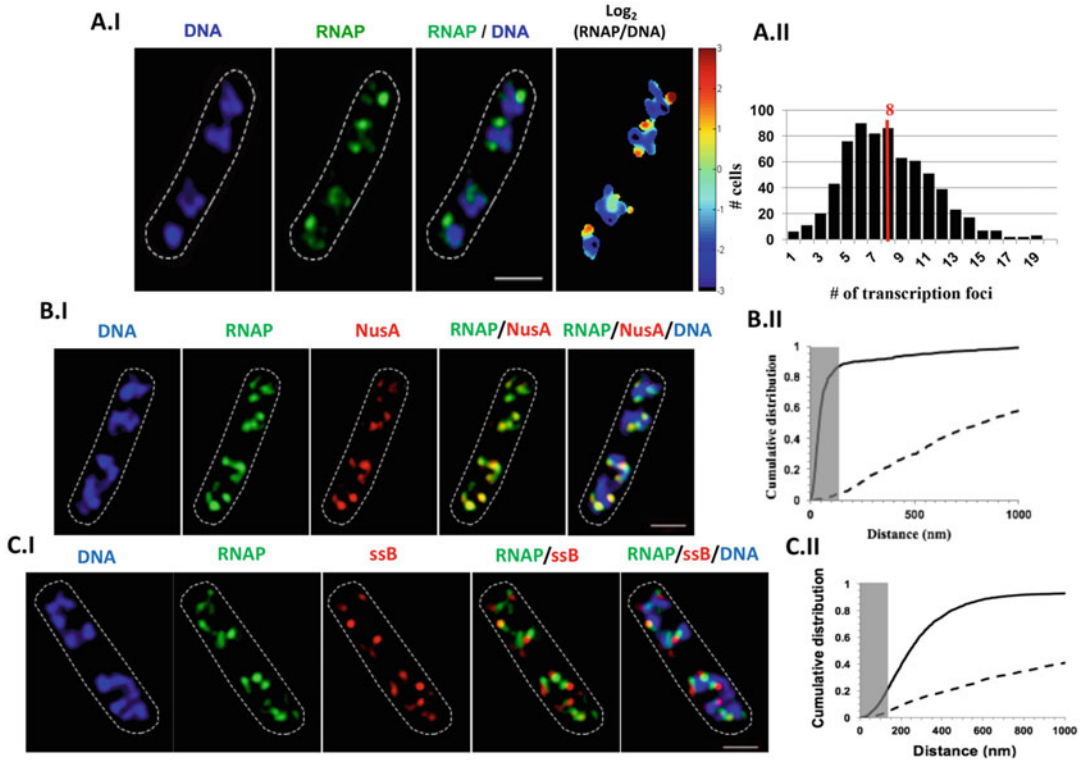


Fig. 1 2D images from 3D-SIM, coimages, and analysis (**A**) Coimaging of RNAP and DNA reveal spatial compartmentalization of transcription foci in fast-growing cells. (A.I) Images of DNA (nucleoid), RNAP, and overlay of DNA (blue) and RNA (green) from a representative fast-growing CC72 single cell (LB, 37 °C). The cell shape is outlined by a dotted line. The scale bar represents 1 μm. Images were reconstructed using Software NIS-Elements Ar/NIS-Elements C, processed using ImageJ/Fiji and edited with Adobe Photoshop. The $\log_2$(-RNAP/DNA) image (heat map) is a quantitative representation of the relationship between RNA Pol and DNA, which is represented by a color scale bar with values ranging from −3 to 3. Note that regions enriched in RNAP up to eightfold over DNA are at the periphery of the nucleoid (red foci) and regions enriched in DNA up to eightfold over RNAP are in the center of the nucleoid (blue regions). (A.II) Histogram of a number of apparent RNAP foci in fast-growing cells processed. The red line indicates the median number of transcription foci in the population of cells. Similar results were reported by Cagliero et al. [18]. A 3D model movie is shown in supplement Video A. (**B**) Transcription elongation factor NusA forms foci and colocalizes with transcription foci in fast-growing CC341 cells. (B.I) Images of DNA, RNAP, NusA, overlay of RNAP (green) and NusA (red), overlay of RNAP (green), NusA (red), DNA (blue). The cell shape is outlined by a dotted line. The scale bar represents 1 μm. Images have been processed as described in Fig. a.I. NusA image indicates that NusA forms foci and perfectly colocalizes with RNAP foci (overall yellow color on the RNAP/NusA overlay). (**B.II**) Colocalization analysis: the cumulative distribution of the distances between NusA foci and their closest RNAP foci in the population of cells (—) NusA-mCherry RNAP-Venus and (- - -) NusA-mCherry RNAP-Venus random. The gray rectangle represents the colocalization area (≤140 nm), as the theoretical SIM microscope resolution for the mCherry is 140 nm. 87.1% of the NusB foci are within 140 nm of the closest transcription foci. Similar results were reported by Cagliero et al. [18]. A 3D model movie is shown in supplement Video B. (C) Spatial segregation of transcription foci and replisome tracked by SSB protein using the strain CC376. (**C.I**) Images of DNA, RNAP, SSB, overlay of RNAP (green) and SSB (red), overlay of RNAP (green), SSB (red), DNA(blue). The cell shape is outlined by a dotted line. The scale bar represents 1 μm. SSB forms foci, which are separate from RNAP. (**C.II**) Colocalization analysis: the cumulative distribution of the distances between SSB foci and their

To generate the random data set, the regions of interest from different channels are combined. Within the combined region of interest, the position of each mCherry centroid is randomly shuffled and similar analyses are performed. Only the distances ranging from 0 to 1000 nm are included in the plots to avoid scaling issue and to increase the readability.

4 Notes

1. Notes and general considerations on the bacterial strains used in this work:

   CC72: This strain contains a chromosomal *rpoC*-Venus fusion and is used to coimage RNAP and DNA to reveal nucleolus-like compartmentalization of the transcription machinery in fast-growing cells.

   CC341: This strain has both *rpoC*-Venus and *nusA*-mCherry fusions in the chromosome; the latter is linked to a kanamycin resistant marker. This strain is used to coimage RNAP, NusA, and DNA to reveal colocalization of RNAP and NusA.

   CC376: This strain contains both chromosomal *rpoC*-Venus and *ssB*-mCherry fusions; the *ssB*-mCherry fusion is linked to a kanamycin resistant marker. This strain is used to coimage RNAP, SSB, and DNA to reveal spatial segregation (different territories) of the transcription and replication machineries in the landscape of the bacterial chromosome.

   It is important to verify that the growth phenotypes of the strains carrying the fluorescent fusion protein(s) are similar to that of the isogenic wild type. It is known that some fluorescent fusion protein(s) cause aggregation and artifacts [21, 22].

2. General considerations about Materials used in the procedure:
   - The poly-L-lysine is commonly used as an attachment factor which improves cell adherence. This phenomenon is based on the interaction between the positively charged polymer and negatively charged cells.
   - 70% ethanol used for cleaning microscope slides.
   - Hoechst is a fluorescent stain of DNA and nucleoid.
   - Mounting media holds cells in place between the cover glass and the microscope slide. It enables the transfer of the specimen into a medium which matches the refractive index of the

Fig. 1 (continued) closest RNAP foci in the population of cells Only 22.3% of the SSB foci are within 140 nm of the closest transcription foci. Similar results were reported by Cagliero et al. [18]. A 3D model movie is shown in supplement Video C

objective ($n = 1.515$ in the case of oil immersion lenses). Mounting media often scavenge free radicals and thus preserve the brightness and retard photobleaching.

- 0.1 μm TetraSpeck microspheres fluorescent as daily calibration and instrumentation adjustment are required for high-precision imaging, particularly in multicolor applications. Mix the TetraSpeck dilution with the cells samples and drop them together on a cover glass.
- The Formaldehyde is used because it is a small four-atom molecule that rapidly penetrates cells and cross-links stably bound complexes of proteins to proteins and proteins to DNA [23].

3. General considerations for sample preparation:
   - Make sure that the sample is properly fixed and mounted onto the slide.
   - The sample should be as bright as possible. Dim samples may require long acquisition times, which increases the risk of vibration-induced artifacts and higher EM gain levels, which could introduce more noise (noise = artifacts). Note that 3D-SIM takes 15 images for each time-point/z slice. This can lead to excessive photobleaching.
   - The cells should be flat, and cover glass should be made of high quality 1.5 optical glass. Lack of flatness will affect the illumination modulation contrast.
   - The cover glass with correct thickness is critical for SIM. Use "High Performance" No. 1.5H cover glass (thickness tolerance = 170 μm ± 5 μm, compared to ± 15 μm for standard cover glass).

     It is critical that the cover glass and the objective lens are kept clean for optimal imaging.
   - The correct density of single cells on the slide ($n \geq 100$) is important for image analysis. A high density of cells causes cells touch each other, resulting in poor resolution of single cells; however, a low density of cells requires more pictures to be taken, which is time-consuming.
   - Slides with samples should be kept on ice and in dark. Before imaging, however, the slides should be returned to room temperature to avoid condensation, during which the slides are covered with foil to avoid light exposure.
   - The samples can be stored in dark at −20 °C (no more than 1 month).

4. General considerations for image acquisition:
   - The objective correction collar should be adjusted under the same conditions as used for sample imaging, i.e., the same

mounting medium, cover glass thickness, temperature, etc. If the collar is not correctly adjusted, the Point Spread Function (PSF) of any diffraction limited object will be distorted, and reconstruction of the resulting 3D-SIM images will be affected. The PSF should always be as close to symmetrical as possible.

- Cells should be focused using transmitted light illumination and not using high-energy wavelengths to prevent photobleaching of samples.
- Optimize signal-to-noise ratio of images using a short exposure time and a maximum intensity value that does not exceed ~12.000 counts for 16-bit mode and ~4000 for 14-bit mode for the camera.
- Recommended acquisition camera settings:

Format: No binning.

Conversion Gain: 1.0X.

Readout Mode: EM Gain 1 MHz, 16-bit or 10 MHz, 14-bit recommended for speed.

EM Gain Multiplier: Keep as low as possible to reduce noise (normal range 200 to 300).

Maximum pixel value: The reconstruction process involves additive functions. To avoid oversaturation and artifacts in the final reconstructed image, use the "histogram window" to help adjust settings to achieve the following intensity values.

- There are no standard acquisition parameters; thus, optimum parameters have to be empirically determined by the user for each specific fluorescent fusion construction and experimental condition. The quality of the acquired images not only depends on the quality of the microscope components but also on numerous factors such as the brightness, the dynamics and the quantity of fluorescent molecules present in the cells.

  3D-SIM imaging requires multiple exposures which means that photobleaching of the fluorescent molecule can be a serious issue. To compromise, it is necessary to reduce the exposure time to the minimum that allows for satisfying fluorescent signal collection.
- At least a couple of microsphere beads per image should be captured, because the beads are essential for alignment.

  An ideal field should have about 20 single cells and a few calibration microsphere beads in the same focal plane.

  Remember analysis requires at least ~100–150 cells to be statistically significant

– Keep the system in a stable environment. Make sure that the room hosting the system is isolated from environmental vibrations (check antivibration table) and, critically, from temperature fluctuations. Optimal system operation is at 23 °C.

5. General considerations for image reconstruction:
   Image reconstruction is carried out using Software NIS-Elements Ar/NIS-Elements C. To avoid introducing artifacts, it is important to carefully set the following parameters:
   – Illumination Modulation Contrast.

     The purpose of this parameter is to improve image reconstruction when the contrast of the illumination pattern is low. This can happen in highly scattered amples or where there is a large amount of out-of-focus light.
   – High Resolution Noise Suppression.

     This parameter is to remove noise from the image and eliminate artifacts that can be introduced during the process.
   – Out-of-Focus Blur Suppression.
     This parameter removes the out-of-focus light and corrects the intensity distribution spread.

Acknowledgment

The authors would like to thank Valentin Magidson and Stephen Lockett at OMAL Facility (NCI, NIH) for their help and discussions. This research was supported by the Intramural Research Program of the National Cancer Institute (The Center for Cancer Research) NIH.

References

1. Gordon GS, Sitnikov D, Webb CD, Teleman A, Straight A, Losick R, Murray AW, Wright A (1997) Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. Cell 90:1113–1121
2. Hiraga S, Ichinose C, Niki H, Yamazoe M (1998) Cell cycle-dependent duplication and bidirectional migration of SeqA-associated DNA-protein complexes in *E. coli*. Mol Cell 1:381–387
3. Lemon KP, Grossman AD (1998) Localization of bacterial DNA polymerase: evidence for a factory model of replication. Science 282:1516–1519
4. Lewis PJ, Thaker SD, Errington J (2000) Compartmentalization of transcription and translation in Bacillus subtilis. EMBO J 19:710–718
5. Cabrera JE, Jin DJ (2003a) Construction, purification, and characterization of *Escherichia coli* RNA polymerases tagged with different fluorescent proteins. Methods Enzymol 370:3–10
6. Cabrera JE, Jin DJ (2003b) The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. Mol Microbiol 50:1493–1505
7. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC (2012) Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. Mol Microbiol 85:21–38
8. Cagliero C, Jin DJ (2013) Dissociation and re-association of RNA polymerase with DNA

during osmotic stress response in *Escherichia coli*. Nucleic Acids Res 41:315–326

9. Endesfelder U, Finan K, Holden SJ, Cook PR, Kapanidis AN, Heilemann M (2013) Multiscale spatial organization of RNA polymerase in *Escherichia coli*. Biophys J 105:172–181
10. Jin DJ, Cagliero C, Martin CM, Izard J, Zhou YN (2015) The dynamic nature and territory of transcriptional machinery in the bacterial chromosome. Front Microbiol 6:497
11. Stracy M, Lesterlin C, Garza de Leon F, Uphoff S, Zawadzki P, Kapanidis AN (2015) Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. Proc Natl Acad Sci U S A 112:E4390–E4399
12. Molina F, Skarstad K (2004) Replication fork and SeqA focus distributions in *Escherichia coli* suggest a replication hyperstructure dependent on nucleotide metabolism. Mol Microbiol 52:1597–1612
13. Reyes-Lamothe R, Possoz C, Danilova O, Sherratt DJ (2008) Independent positioning and action of *Escherichia coli* replisomes in live cells. Cell 133:90–102
14. Jin DJ, Cagliero C, Zhou YN (2013) Role of RNA polymerase and transcription in the organization of the bacterial nucleoid. Chem Rev 113:8662–8682
15. Coltharp C, Xiao J (2012) Superresolution microscopy for microbiology. Cell Microbiol 14:1808–1818
16. Fiolka R, Shao L, Rego EH, Davidson MW, Gustafsson MG (2012) Time-lapse two-color 3D imaging of live cells with doubled resolution using structured illumination. Proc Natl Acad Sci U S A 109:5311–5315
17. Gustafsson MG (2000) Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J Microsc 198:82–87
18. Cagliero C, Zhou YN, Jin DJ (2014) Spatial organization of transcription machinery and its segregation from the replisome in fast-growing bacterial cells. Nucleic Acids Res 42:13696–13705
19. Jin DJ, Mata Martin C, Sun Z, Cagliero C, Zhou YN (2017) Nucleolus-like compartmentalization of the transcription machinery in fast-growing bacterial cells. Crit Rev Biochem Mol Biol 52:96–106
20. Elbing, K., and Brent, R. (2002) Media preparation and bacteriological tools. Curr Protoc Mol Biol *Chapter 1*, Unit 1 1
21. Langhorst ME, Schaffer J, Goetze B (2009) Structure brings clarity: structured illumination microscopy in cell biology. Biotechnol J 4:858–8665
22. Wang S, Moffitt JR, Dempsey GT (2014) Characterization and development of photoactivatable fluorescent proteins for single-molecule-ased superresolution imaging. Proc Natl Acad Sci USA 111:8452–8457
23. Schmiedeberg L, Skene P, Deaton A, Bird A (2009) A temporal threshold for formaldehyde crosslinking and fixation. PLoS One 4:e4636

Chapter 8

Genetic Approaches to Study the Interplay Between Transcription and Nucleoid-Associated Proteins in *Escherichia coli*

Karin Schnetz

Abstract

Bacterial nucleoid-associated proteins are important in nucleoid-structuring, homeostasis of DNA supercoiling, and in regulation of transcription. Vice versa, transcription influences DNA supercoiling and possibly DNA-binding of nucleoid-associated proteins. Here, I describe genetic tools to study the interplay between transcription and nucleoid-associated proteins such as H-NS in *Escherichia coli*. These standard methods include construction of genomic promoter reporter gene fusions to study regulation of promoters, genome insertion of promoter cassettes to drive expression of a gene of interest, and construction of isogenic *hns* mutants and precautions when doing so.

Key words Transcription regulation, Transduction, Phage T4, Recombineering, Chromosomal integration, Attachment site

1 Introduction

Genetic analysis of transcription regulation includes a plethora of experimental approaches and assays. A common principle of these approaches is comparing the regulatory output in wild-type and isogenic mutants. In *Escherichia coli* genetic tools to generate isogenic mutants include classical methods such as transduction of markers with the help of phages, but also various methods to mutate genes in the genome. Further, reporter gene fusions are still an invaluable tool for studying transcription regulation, despite current methods for the quantification of transcripts like qRT-PCR and RNA-seq.

For deletion and replacement of genes in the genome Lambda Red recombineering is commonly used [1]. However, recombineering with this method leaves a Flp recombinase target-site (*frt*-site) as a *scar* in the genome. Such a scar may have no impact in many cases, but may be perturbing in other cases. Methods for

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_8,

introducing mutations, including point-mutations, that are scarless have been described; these methods rely on dual, positive and negative, selection markers including pairs of selectable loci such as *tetA-sacB* or *neo-ccdB* [2]. Recombineering with such a dual selection system involves two rounds of recombination. Recently, a straightforward direct approach for scarless recombineering has been described in which a CRISPR/cas selection system is combined with Lambda Red recombineering [3, 4]. Lambda Red recombineering can also be used to insert fragments at specific sites in the genome. This includes the insertion of constitutive or regulated promoter cassettes upstream of a gene of interest, as described here. Insertion of promoter cassettes allows to uncouple expression of a gene from its natural regulation and to constitutively express or to control expression of this gene.

In general, for studying regulation of transcription initiation the use of fusions of the promoter to be studied to a reporter gene is very useful. A classical reporter gene is *lacZ* encoding β-galactosidase which can be monitored qualitatively (e.g., on X-Gal plates) or quantitatively [5]. Plasmid encoded promoter-reporter gene fusions are useful to study regulation of a promoter. However, to achieve physiological conditions, to avoid titration of regulatory proteins by high copy numbers of the promoter region, and to avoid fluctuation of plasmid-copy number, it can be an advantage to integrate the promoter reporter gene fusion into the genome. In principle promoter-reporter gene fusions can be integrated anywhere in the genome, e.g., by Lambda Red recombineering. Another possibility is using phage-based site-specific recombination systems such as the phage lambda Integrase system that catalyzes recombination between a chromosomal phage lambda attachment site *attB* and the phage *attP* site. Diederich et al. [8] have established a plasmid-based *attP-attB* system, of which a slightly modified version is described here.

The genetic analysis of transcription regulation of a gene or a promoter-reporter gene fusion by a specific transcription regulator or a pleiotropic transcription regulator, such as the nucleoid-associated H-NS, involves the use of a wild-type and isogenic mutant (s) in which the regulator encoding gene is deleted or carries other mutations. This genetic comparison requires that the mutant allele encoding the regulator is transferred to the strain carrying the reporter or that this mutation is constructed in the strain of interest. In case the required allele is present in another strain background it can be transferred by transduction, i.e., by phage-mediated transfer of DNA between strains. Selection of clones that carry a transferred allele can be a difficult task for alleles that are not coupled to an easily selectable marker. However, often alleles are used that have been tagged with a selectable antibiotic resistance gene by Lambda Red recombineering. In addition, by recombineering a selectable marker can easily be inserted next to

the allele that needs to be transferred. The transfer of such a tagged allele to a recipient strain can be selected on antibiotic plates. After the selection of the transductants on selective antibiotic plates the presence and integrity of the transduced allele and the presence of other alleles of the recipient strain need to be carefully validated, e.g., by PCR and by sequencing. Transduction by phage P1*vir* is a standard method in *E. coli* genetics [5]. Here, I describe transduction by a phage T4 mutant (T4*GT7*) which is likewise a general transducing phage, as described [6].

Of the plethora of genetic tools available for the analysis of transcription regulation and the interplay of transcription and regulation by nucleoid-associated proteins, here three methods are described. These include the construction and genome integration of promoter-reporter gene fusions, the insertion of promoter cassettes to control expression of a specific gene, and generalized transduction by phage T4*GT7*.

2 Materials

2.1 Antibiotics and Media

1. 50 mg/mL spectinomycin stock solution in 30% ethanol stored at −20 °C, for selection use 50 μg/mL final concentration.
2. 10 mg/mL kanamycin stock in water solution stored at 4 °C, for selection use 25 μg/mL final concentration.
3. 50 mg/mL ampicillin stock solution in 50% ethanol stored at −20 °C, for selection use 50 μg/mL final concentration.
4. LB medium. Dissolve 10 g tryptone, 5 g yeast extract, 5 g NaCl per liter, for plates add in addition 15 g bacto agar. Autoclave.
5. SOB medium. Dissolve 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.25 mL 2 M KCl, adjust pH to 7.0 with NaOH. Autoclave. After autoclaving, prior to use add 10 mL 1 M $MgCl_2$ per liter.
6. T4 top-agar. Weigh in 6 g agar, 10 g tryptone, 8 g NaCl, 2 g trisodium citrate dehydrate, 3 g glucose per liter. Autoclave and store in approximately 50 mL aliquots.

2.2 Plasmids and Strains

1. Plasmids pKES148, pKES263, pKES304, pKES305, pKEAR3, pKES219, and pKESK22 (available upon request).
2. Plasmid pLDR8 is available from ATCC (https://www.lgcstandards-atcc.org) as ATCC® 77357™.
3. Plasmid pKD46 (available from CGSC as CGSC#7669, http://cgsc2.biology.yale.edu).
4. *E. coli* strains BW23473 (*pir*$^+$, CGSC#7837) and BW23474 (*pir-116*, CGSC 7838) that are needed for the propagation of Pir-dependent plasmids carrying an R6Kγ origin of replication (available from CGSC, http://cgsc2.biology.yale.edu).

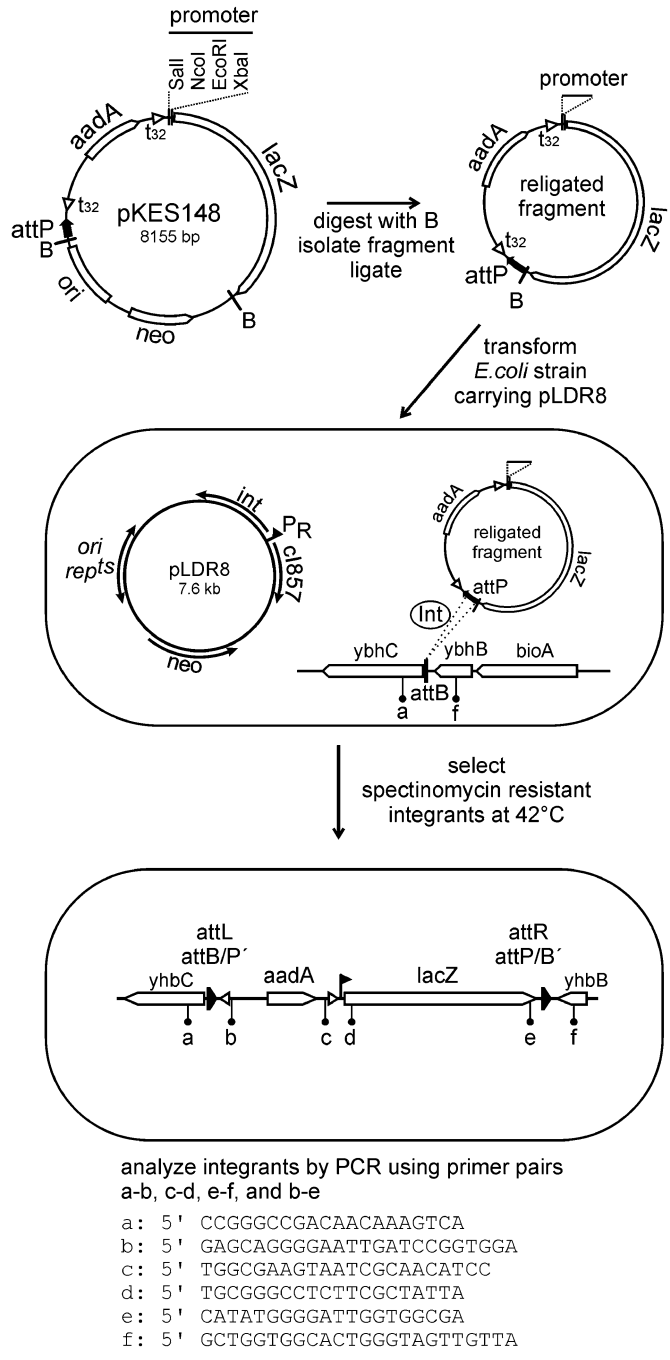


Fig. 1 Principle of the construction and genomic integration of promoter *lacZ* reporter gene fusions. Plasmid pKES148 is a vector for cloning of promoter-*lacZ* fusions and subsequent integration into the chromosomal *attB*-site. For integration plasmid derivatives carrying the promoter-*lacZ* fusion are digested with BamHI (B), the promoter-*lacZ* the fragment is isolated and self-ligated. In a next step the *E. coli* strain of interest, which harbors pLDR8, is made competent at

5. *E. coli* strain U65 Δ(*araC-araBAD*) Δ(*lacI-lacZYA*) $_{FRT}P_{cp8}$-*araE* Δ(*araH-araF*)$_{FRT}$ [7] (available upon request).

3 Methods

3.1 Generation of Chromosomal Promoter Reporter Gene (lacZ) Fusions

Integration of reporter gene fusions into the chromosomal phage lambda attachment site *attB* of *E. coli* K12 using plasmids has been established by Diederich et al. [8]. The principle of the method is site-specific integration of origin-less circular DNA carrying a phage Lambda *attP*-site, a promoter *lacZ* fusion, and a resistance gene cassette for selection (*see* Fig. 1). The site-specific recombinase is provided by a *helper* plasmid, pLDR8 [8]. This plasmid carries the *int* gene, encoding phage lambda integrase, under control of the phage lambda P_R promoter. In addition, lambda cI repressor allele cI_{857} is carried on the plasmid, which encodes a temperature-sensitive cI repressor protein mutant. Therefore, *int* expression is repressed at low temperature (28 °C or 30 °C) and induced at high temperature 42 °C (or 37 °C). Further, the pSC101-derived pLDR8-plasmid replicates only at low temperature (28 °C or 30 °C), since the plasmid-encoded replication protein, RepA, is temperature-sensitive. Thus, at low temperature the plasmid is maintained and *int* is repressed, while upon a shift to high temperature *int* expression is induced and integrase for catalyzing site-specific integration is synthetized, and the *helper* plasmid pLDR8 is lost. Accordingly, cultures for preparing bacterial cells competent for uptake of DNA and integration at *attB* are grown at elevated temperature (37 °C), and integrants are selected on spectinomycin plates at 42 °C and then validated by PCR. Here, I describe a slightly modified version [9] of the original method [8]. For integration we use plasmids carrying *attP*, an omegon-spectinomycin-resistance cassette (spectinomycin-resistance encoded by *aadA)* [10], a polylinker-site for cloning of promoter fragments, and the *lacZ* reporter gene (*see* Fig. 1).

1. Clone a promoter fragment into vector pKES148 or one of its derivatives carrying the *lacZ* reporter gene (*see* Fig. 1).
2. Digest DNA of the plasmid that carries the promoter-*lacZ* fusion with BamHI (or BglII depending on the vector), separate the fragments by agarose gel electrophoresis, and elute the fragment carrying the promoter-*lacZ* fusion, the *attP*-site, and

Fig. 1 (continued) 37 °C for induction of *int* and then transformed with the religated fragment. Integration of the re-ligated fragment at *attB* is catalyzed by Int (integrase) and integrants are selected by plating on LB spectinomycin plates at 42 °C. Integrants are analyzed by PCR using the indicated primer pairs. Primer pair b-e is used to exclude integration of dimeric re-ligated fragments

the *aadA* gene (coding for spectinomycin resistance), but lacking the origin of replication and the *neo* gene (kanamycin resistance). Gel purification of the origin-less fragment reduces contamination by plasmid DNA.

3. Carefully determine the concentration of the isolated fragment. Use maximally 10 ng of this fragment for self-ligation in a 20 μL ligation sample (0.5 ng/μL) to create an origin-less circular DNA. Higher amounts are disadvantageous for self-ligation, because the likelihood of intermolecular ligations increases (*see* **Note 1**). This self-ligated circular DNA is used for integration into the *attB*-site (**step 9**).
4. In parallel, use a standard method to prepare competent cells of the *E. coli* K12 strain of interest, which needs to be a *lacZ* mutant. In general, chemically ($CaCl_2$) competent cells work fine.
5. Transform the strain with the *helper* plasmid pLDR8 [8]. Select the transformants on a LB kanamycin plate, and incubate the plate overnight at 28 °C (or 30 °C).
6. Use the transformants of your *E. coli* strain with pLDR8 to inoculate a small culture in LB kanamycin (3–4 mL) and grow overnight at 28 °C (or 30 °C).
7. Use the fresh overnight culture to inoculate a culture in LB kanamycin medium for preparing competent cells. For example, inoculate 25 mL of LB kanamycin medium with 500 μL of the fresh over-night culture. Grow the culture for 90 min at 37 °C to induce expression of *int* and then harvest it on ice.
8. Of this harvested culture prepare chemically competent cell.
9. For the integration of the promoter-*lacZ* fusion, transform the competent cells of the *E. coli* strain harboring pLDR8 with the self-ligated circular fragment (from **step 3**). Half of the ligation sample is sufficient and the ligation sample may be used directly without any need for the purification of the DNA. Plate the transformants on LB spectinomycin plates that have been pre-warmed at 42 °C, and incubate the plates at 42 °C overnight.
10. To analyze clones for the integration of the promoter *lacZ* fusions pick several individual colonies (usually 4–8 are sufficient) and re-streak them on LB spectinomycin plates, incubate the plates overnight at 37 °C. In case colonies differ in size, select representatives of these size variants for analysis.
11. Characterize the selected clones by analytical scale PCRs using primer pairs specific for the *attB/attP* fusion-sites *attL* and *attR* (*see* Fig. 1). This analysis includes PCR reactions to exclude integration of dimers, and to confirm the integrity of the promoter *lacZ* fusion, respectively (*see* Fig. 1). For setting

up the PCRs pick a tiny aliquot of bacteria from the selected clone or from a single colony of the re-streaked clone and suspend this aliquot in 100 µL H_2O. Use 1 µL of this suspension as a template in the PCR reactions. The PCR works best if only very few bacteria are used (*see* **Note 2**).

12. To test for loss of the helper plasmid pLDR8 and to make sure that not the intact plasmid carrying the *lacZ* reporter plasmid was transformed, replica-plate the clones by re-streaking on LB kanamycin plates. The colonies need to be kanamycin-sensitive.
13. Store two independent validated clones as glycerol or DMSO stocks at −80 °C. For storing as a glycerol stock, to 1 mL of a fresh overnight culture add glycerol to a final concentration of 15–25%, and freeze at −80 °C. For storing as DMSO stock, add 50 µL DMSO to 1 mL of fresh overnight culture and freeze at −80 °C.

3.2 Integration of Promoter Cassettes into the Genome

Genomic insertion of inducible or constitutive promoter cassettes is a tool to control expression of a gene of interest. Here, plasmids are described for integration of such promoter cassettes by Lambda Red recombineering, as described [1]. One of these plasmids (pKEAR3) carries an *araC* P_{BAD} cassette for arabinose inducible expression of a gene of interest. In case, gradual induction of the P_{BAD} promoter is required use strain U65 [7], which does not ferment arabinose (*see* **Note 3**). The second set of plasmids carries a constitutive variant of the *lacUV5* promoter (pKES304) and the phage Lambda P_L promoter (pKES263), respectively. A further plasmid (pKES305) carries the constitutive *lacUV5* promoter followed by phage Lambda terminator *tR1*. In this case expression can be induced by Lambda antiterminator protein N that is provided *in trans* by plasmid pKES219. In Fig. 2 schemes of the plasmids, primer sequences for PCR amplification of the fragment used for recombination into the genome, as well as the basic strategy for integration and validation of strains are shown, as described [1]. It is recommended that the allele that has been generated by Lambda Red recombineering is transferred by transduction to the strain background of interest, at least in case the strain of interest carries additional alleles that were generated by Lambda Red recombineering, since such alleles may serve as off-targets in recombineering (*see* **Note 4**). As a last step the selection marker that is flanked by Flp recombinase target sites (*frt*) may be deleted by Flp recombinase catalyzed site-specific recombination, as described [1].

1. For Lambda Red recombineering amplify the fragment encompassing the FRT-flanked resistance gene and promoter cassette by PCR (*see* Fig. 2). The 5′ end of the primers used for amplification needs to carry a 30–50 bp sequence homologous to the

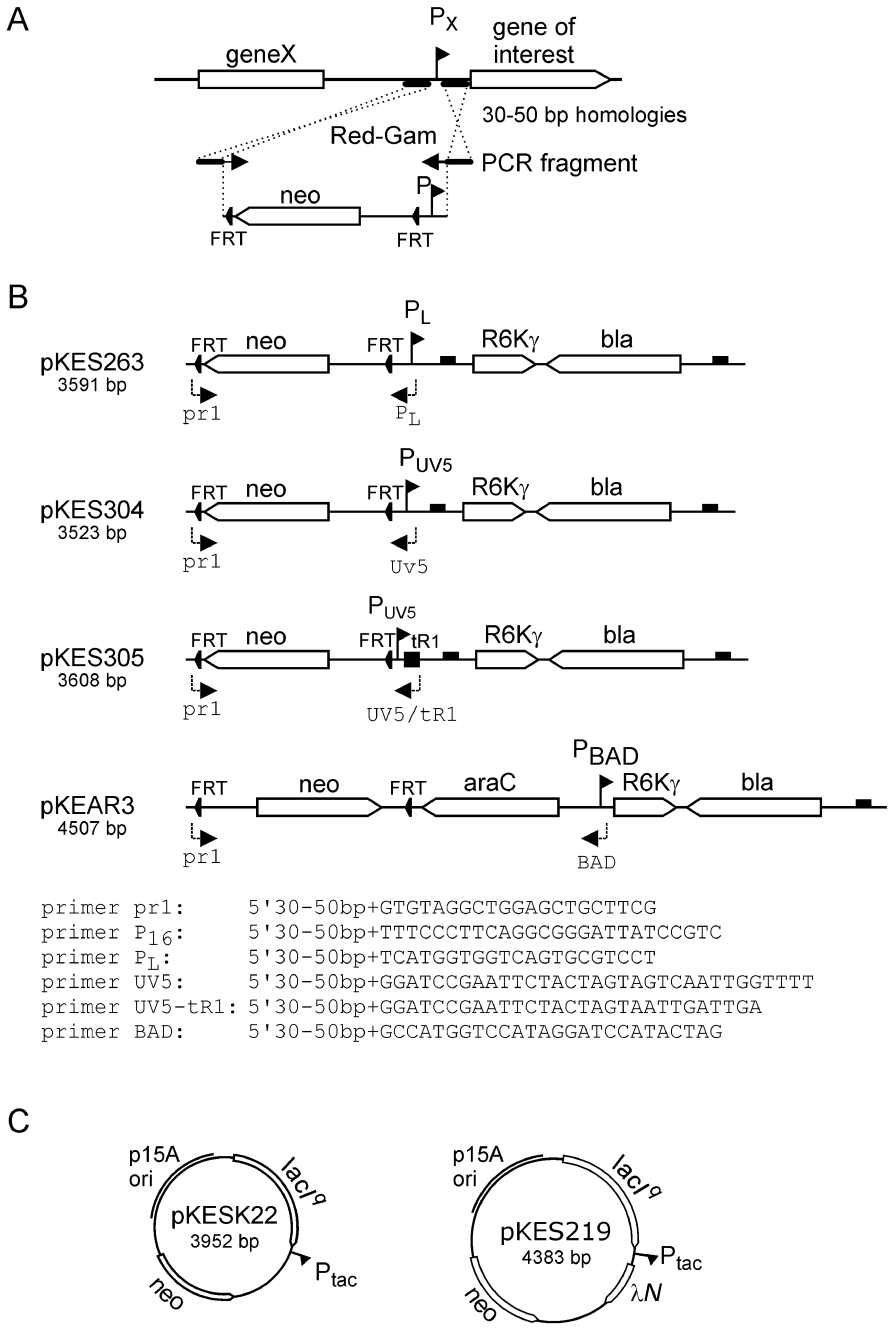


Fig. 2 Plasmids for targeted genome integration of promoters by recombineering. (**a**) Principle of integration of fragments carrying promoters and the *frt*-flanked *neo* gene (kanamycin resistance) by recombineering. (**b**) Plasmids that serve as templates for generating the fragments for integration. Plasmids pKES262, pKES263 [14], pKES304, and pKES305 are derivatives of pKD13, while pKEAR3 is a derivative of pKD4 [1]. Indicated are primer sequences for amplification of fragments. (**c**) In plasmid pKES219 the Lambda *N* gene is expressed under control of the *tac* promoter. This plasmid is needed to induce expression by the *lacUV5* promoter and *tR1* terminator cassette that has been inserted to direct expression of a gene of interest. Plasmid pKESK22 [15] serves as control

genomic target locus. It is recommended that the PCR fragment is purified by gel electrophoresis followed by gel elution. It is important that the fragment is eluted in H_2O and that the concentration of the fragment is 100 ng/μL or higher.

2. For Lambda Red recombineering transform the *E. coli* strain of choice (*strain X*) with plasmid pKD46 [1] which carries the *red-gam-exo* genes under control of the arabinose inducible P_{BAD} promoter. Any standard transformation method may be used. Transformants are selected on LB ampicillin plates at 28 °C (or 30 °C), since replication of plasmid pKD46 is temperature-sensitive.
3. Grow *strain X* carrying pKD46 bacteria in SOB medium with ampicillin and 10 mML-arabinose at 28 °C (or 30 °C) and prepare electro-competent cells.
4. Electro-transform *strain X*/pKD46 with the linear purified PCR fragment. Use 1 μL (>100 ng/μL) of DNA in H_2O for transformation. Select the integrants by plating half of the sample on LB plates containing the appropriate antibiotic (kanamycin) and incubate the plates at 37 °C overnight. If no clones grow overnight, plate the remainder of the transformation sample, which has been kept at room temperature until the next day.
5. Re-streak resistant clones on selective plates.
6. Analyze the re-streaked clones using primers specific for the target locus. The PCR fragment may be used for sequencing.
7. In addition, re-streak the clones on ampicillin plates and incubate the plates at 37 °C to check for loss of the helper plasmid pKD46. The clones need to be ampicillin sensitive.
8. Store at least two independent ampicillin-sensitive clones which carry the cassette inserted as glycerol or DMSO stocks at −80 °C (as described above in Subheading 3.1, **step 13**).
9. Recommended: transduce the newly generated allele into the *E. coli* strain of interest.
10. To delete the *frt* flanked resistance gene, transform the strains with temperature-sensitive plasmid pCP20, and select transformants at 28 °C (or 30 °C) on LB ampicillin plates. Plasmid pCP20 carries the *flp* gene under control of the P_L promoter which is repressed by the temperature-sensitive cI_{857} repressor at low temperature only. In addition, replication of pCP20 is temperature-sensitive.
11. Re-streak 2–4 transformants on LB plates and incubate overnight at 42 °C. Screen by replica-plating for loss of the kanamycin resistance cassette as well as for loss of pCP20, i.e., loss of ampicillin resistance. Incubate the replica-plates at 37 °C overnight.

12. Analyze kanamycin sensitive clones by PCR using the target locus-specific primers. The PCR fragment may be sequenced for validating proper recombineering.
13. Store at least two independent clones as glycerol or DMSO stocks at −80 °C.

3.3 Construction of Isogenic Mutants (hns and Others) by Transduction

Transduction is a convenient method to transfer *hns* alleles or any other allele encoding nucleoid-associated proteins from a donor strain, and to create isogenic mutants of strains. The use of isogenic mutants is important in genetic analyses such as analyses of expression levels. In the following a protocol for transduction by phage T4*GT7* is given [6]. Alternatively, for transduction the commonly used phage P1*vir* [5] may be used. By routine, after transduction, the transduced allele and all other alleles should be validated by PCR or by sequencing. This is important to exclude false positives. False positives may include failure of transduction of the allele although the selected marker was present or selection of clones that are not of the expected strain background.

Furthermore, *E. coli* strain carrying mutations in *hns* have a tendency to accumulate secondary mutations, including mutations in *rpoS* or *stpA* [11]. Therefore, *hns* mutant strains should be immediately stored when constructed, several independent isolates should be kept for comparative studies, and for each experiment the mutants should be freshly streaked from frozen stocks. Sequencing of *stpA* and *rpoS* is recommended. In our hands, methods to manipulate the genome can work less efficiently for *hns* mutants. Therefore, it is recommended that *hns* alleles are transferred as a last step in strain construction.

3.3.1 Preparation of a Phage T4GT7 Lysate on the Donor Strain

1. To prepare a lysate on the donor strain (e.g., an *hns* mutant, carrying an *hns* allele that is tagged with a resistance gene), set up a serial dilution series of a T4*GT7* lysate. Depending on the titer of the lysate set up 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} dilutions. Use 100 μL of each dilution and add to a 100 μL aliquot of an overnight culture of the donor strain. Incubate for 10–20 min at room temperature for adsorption of the phage.
2. Add 1 mL of LB medium to each bacteria/phage sample and transfer to 10 mL culture tubes.
3. Add 3 mLT4-Topagar (melted by boiling and kept liquid at 45 °C in a water-bath). Mix shortly and carefully by rolling the tube (avoid air bubbles and work quickly). Immediately pour the sample onto fresh LB plates so that a homogenous layer of top agar is formed on the plate. The plates that are used for plating should be freshly made on the same day or the day before.

4. Incubate the plates with the top agar facing up at 37 °C for 8–14 h. Longer incubation results in reduced transduction rates.
5. Use the plate that shows an almost confluent lysis. Pipette 1 mL of LB onto this plate, scratch off the top-agar off, and transfer to a 10 mL glass tube or 15 mL plastic tube.
6. Extract the sample at least twice with pure chloroform. To this end, add approximately half volume of pure chloroform, mix well (e.g., using a glass pistil), centrifuge, and take off the aqueous upper layer. Repeat the chloroform extraction until no interphase is apparent.
7. Store the T4*GT7* lysate at 4 °C. Storage is possible for several weeks to months. However, upon storage the titer of the lysate decreases and therefore older lysates may not work well for transduction.
8. Optional: determine the phage titer by setting up serial dilutions of the lysate and plating with bacteria as described above. The expected titer is 10^{10} to 10^{11} phages/mL.

3.3.2 Transduction of the Recipient Strain

1. For transduction prepare 200 μL aliquots of a fresh overnight culture of the recipient strain. Add 2 μL, 1 μL, 0.5 μL, and 0.2 μL of the T4*GT7* lysate that was grown on the donor strain. Incubate the samples for 10–20 min at room temperature for adsorption of the phage. In addition, set up control samples that lack bacteria or phage.
2. Plate the samples onto appropriate plates for selection of transductants and incubate the plates overnight (up to 24 h).
3. Re-streak colonies immediately on the same type of plates. Re-streaking must be repeated several times (3–4 times) to get rid of contaminating T4*GT7* phages. It is important to re-streak colonies as soon as possible. Old colonies may contain T4 resistant mutants which cannot be transduced further.
4. Analyze the transductants by PCR using primer pairs specific for the transduced allele. In addition, analyze other genetic markers of the recipient by PCR or by genetic methods such as validation of the phenotype.
5. Store at least 2 independent validated clones as glycerol or DMSO stocks at −80 °C.

4 Notes

1. For ligation of origin-less DNA circles and for other cloning procedures accurate determination of the concentration of the DNA fragment is required. As a standard, an aliquot of the

fragment should be separated on an agarose gel, which allows evaluating both the size of the fragment and the concentration. Measurement of the DNA concentration by OD_{260} (for example using a Nanodrop instrument) is valuable when used in addition, but may be misleading when used as the only method.

2. Analysis of colonies by analytical scale PCR can be set up by adding bacteria to the PCR tubes and does not require isolation of DNA or lysis of the bacterial cells. However, for *E. coli* colony PCR works best if only very few bacteria are added to the PCR reactions. To achieve this, a tiny aliquot of a colony may be resuspended in 100 μL H_2O, and then only 1 μL of this suspension is added as a template to the PCR reaction. Usually, the yield of the specific product drops and unspecific PCR products are obtained when too many bacteria are added.
3. The arabinose-inducible promoter P_{BAD} has the advantage that it is tightly regulated [12]. Induction of the promoter by arabinose can be achieved in any medium and strain background if a high concentration of arabinose (1–10 mM) is used. However, the P_{BAD} is catabolite regulated and arabinose is consumed as carbon source by *E. coli*. Therefore, induction works better when cells are not grown in LB medium, but when they are grown in a medium that lacks glucose, such as tryptone medium or minimal media. Further, for homogenous and gradual induction of P_{BAD} by arabinose it is recommended to use a strain with a modified arabinose regulon [13]. We commonly use strain U65 that is derived from MG1655 *rph*$^+$ strain BW30270 [7]. In strain U65 the *araC-araBAD* and *araFGH* loci are deleted. In addition, the expression of *araE* encoding an arabinose transporter is constitutive. In strain U65 gradual induction of P_{BAD} using a range of 2 μM–100 μM arabinose is achievable.
4. Lambda Red-mediated integration of PCR fragments generally is very specific. However, if the target strain already carries one or more FRT sites elsewhere in the genome, the Lambda-Red system may catalyze homologous recombination into one of these FRT sites. As a consequence false positives are selected. To avoid this, an allele may by mutated in the wild-type *E. coli* background and then transferred by transduction to the strain background of interest.

Acknowledgment

I thank members of the laboratory, in particular Susann Fragel, Aathmaja A. Rangarajan, and Cihan Yilmaz, for comments and discussion. I thank Merlin Muhr and Aathmaja A. Rangarajan for construction of plasmids.

References

1. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645 https://doi.org/10.1073/pnas.120163297
2. Li X-T, Thomason LC, Sawitzke JA, Costantino N, Court DL (2013) Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. Nucleic Acids Res 41:e204 https://doi.org/10.1093/nar/gkt1075
3. Reisch CR, Prather KLJ (2015) The no-SCAR (Scarless Cas9 assisted recombineering) system for genome editing in Escherichia coli. Sci Rep 5:15096 https://doi.org/10.1038/srep15096
4. Reisch CR, Prather KLJ (2017) Scarless Cas9 assisted recombineering (no-SCAR) in Escherichia coli, an easy-to-use system for genome editing. In: Current protocols in molecular biology, vol 117. Wiley, p 31.38.31–31.38.20. https://doi.org/10.1002/cpmb.29
5. Miller JH (1992) A short course in bacterial genetics. A laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, Plainview
6. Wilson GG, Young KYK, Edlin GJ, Konigsberg W (1979) High-frequency generalised transduction by bacteriophage T4. Nature 280:80–82 https://doi.org/10.1038/280080a0
7. Breddermann H, Schnetz K (2016) Correlation of antagonistic regulation of leuO transcription with the cellular levels of BglJ-RcsB and LeuO in Escherichia coli. Front Cell Infect Microbiol 6:106. https://doi.org/10.3389/fcimb.2016.00106
8. Diederich L, Rasmussen LJ, Messer W (1992) New cloning vectors for integration into the lambda attachment site *attB* of the *Escherichia coli* chromosome. Plasmid 28:14–24 https://doi.org/10.1016/0147-619X(92)90032-6
9. Dole S, Kühn S, Schnetz K (2002) Post-transcriptional enhancement of Escherichia coli *bgl* operon silencing by limitation of BglG-mediated antitermination at low transcription rates. Mol Microbiol 43:217–226 https://doi.org/10.1046/j.1365-2958.2002.02734.x
10. Prentki P, Krisch HM (1984) In vitro insertional mutagenesis with a selectable DNA fragment. Gene 29:303–313. https://doi.org/10.1016/0378-1119(84)90059-3
11. Singh K, Milstein JN, Navarre WW (2016) Xenogeneic silencing and its impact on bacterial genomes. Annu Rev Microbiol 70:199–213 https://doi.org/1.1146/annurev-micro-102215-095301
12. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177:4121–4130
13. Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD (2001) Homogeneous expression of the PBAD promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiol 147:3241–3247 https://doi.org/10.1099/00221287-147-12-3241
14. Pannen D, Fabisch M, Gausling L, Schnetz K (2016) Interaction of the RcsB response regulator with auxiliary transcription regulators in *Escherichia coli*. J Biol Chem 291:2357–2370 https://doi.org/10.1074/jbc.M115.696815
15. Stratmann T, Madhusudan S, Schnetz K (2008) Regulation of the *yjjQ-bglJ* operon, encoding LuxR-type transcription factors, and the divergent *yjjP* gene by H-NS and LeuO. J Bacteriol 190:926–935 https://doi.org/10.1128/JB.01447-07

Part II

In Vitro Approaches

Chapter 9

Atomic Force Microscopy Imaging and Analysis of Prokaryotic Genome Organization

Ryosuke L. Ohniwa, Hugo Maruyama, Kazuya Morikawa, and Kunio Takeyasu

Abstract

This protocol describes the application of atomic force microscopy for structural analysis of the prokaryotic and organellar nucleoids. It is based on a simple cell manipulation procedure that enables step-wise dissection of the nucleoid. The procedure includes (1) on-substrate-lysis of cells, and (2) enzyme treatment, followed by atomic force microscopy. This type of dissection analysis permits analysis of nucleoid structure ranging from the fundamental units assembled on DNA to higher order levels of organization. The combination with molecular-genetic and biochemical techniques further permits analysis of the functions of key nucleoid factors relevant to signal-induced structural re-organization or building up of basic structures, as seen for Dps in *Escherichia coli*, and TrmBL2 in *Thermococcus kodakarensis*. These systems are described here as examples of the successful application of AFM for this purpose. Moreover, we describe the procedures needed for quantitative analysis of the data.

Key words Nucleoid structure, On-substrate lysis, Reconstitution of nucleoid, Atomic force microscopy, Data analysis

1 Introduction

In general prokaryotic genomes are circular. The mm scale sized genomic DNA is packed in a cell only a few μm in diameter in the form of a "nucleoid" [1, 2]. Electron microscopy observations of isolated bacterial nucleoids have revealed that the circular genome is fibrous and bundled at the core to form a rosette-like structure with interwound loops emanating from the core [3–5].

Atomic Force Microscopy (AFM) has been applied to understand nucleoid architecture via two different approaches. The in vitro approach consists of analysis of re-constituted DNA-protein complexes to understand structural units of the nucleoid. Hundreds of proteins are associated with the *Escherichia coli* nucleoid, including proteins such as HU (heat-unstable nucleoid protein), H-NS (histone-like nucleoid structuring protein), FIS (factor

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837, https://doi.org/10.1007/978-1-4939-8675-0_9,

for inversion stimulation), and Dps (DNA-binding protein from starved cells) [6–10]. AFM analyses demonstrated that HU binding induces multiple forms of DNA-protein complexes such as compacted, rigid filamentous, or corkscrew DNA structures [11, 12]. H-NS "zips up" the double-stranded DNA to form a bridge between adjacent strands [13, 14]. FIS preferably binds to A/T-rich regions with a highly degenerated dyad symmetry, bends DNA, and stabilizes DNA loops at high concentration [15]. This approach has contributed to our understanding of the basic principles that shape nucleoid architecture. However, these nucleoid proteins constitute only a small fraction of all proteins associated with the nucleoid [9, 16], giving limited information on overall nucleoid structure.

We have developed another approach to analyze nucleoid architecture. We "dissect" cells by removing cellular components step by step, and analyze the structure of the whole nucleoid by AFM. In this chapter, we introduce the sample preparation and image analysis methods in the "dissection approach" with practical notes on application to *E. coli*, *Thermococcus kodakarensis*, chloroplasts in spinach cells, and mitochondria in HeLa cells.

2 Materials

2.1 Culture Medium

Lysogeny (Luria) Broth: 10 g/L peptone, 5 g/L yeast extract, 5 g/L sodium chloride, pH 7.0.

Brain Heart Infusion Broth: 5 g/L beef heart (infusion from 250 g), 12.5 g/L calf brains (infusion from 200 g), 2.5 g/L Na_2HPO_4, 2 g/L D(+)-glucose, 10 g/L peptone, 5 g/L NaCl.

0.8× ASW-YT-S0: 0.8× ASW (16 g/L NaCl, 2.4 g/L $MgCl_2$•$6H_2O$, 4.8 g/L $MgSO_4$•$7H_2O$, 0.8 g/L $(NH_4)_2SO_4$, 0.16 g/L $NaHCO_3$, 0.24 g/L $CaCl_2$•$2H_2O$, 0.4 g/L KCl, 0.34 g/L KH_2PO_4, 40 mg/L NaBr, 16 mg/L $SrCl_2$•$6H_2O$, 8 mg/L Fe(NH4) citrate), supplemented with 5 g/L of yeast extract, 5 g/L of tryptone, and 0.2% elemental sulfur).

Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum.

2.2 Buffers

Phosphate-buffered saline (PBS): 100 mM NaCl, 10 mM phosphate buffer, pH 7.2.

Lysis buffer A: 10 mM Tris–HCl pH 8.2, 1 mM NaN_3, 0.1 M NaCl, milliQ water.

Topo I buffer: 35 mM Tris–HCl, pH 8.0, 72 mM KCl, 5 mM $MgCl_2$, 5 mM DTT, 5 mM spermidine, 0.01% bovine serum albumin.

Extraction buffer: 25 mM HEPES-NaOH pH 7.0, 15 mM $MgCl_2$, 100 mM NaCl, 0.4 M sorbitol, and 0.5% Triton X-100).

MNase buffer: 20 mM Tris–HCl pH 7.4, 5 mM NaCl, 2.5 mM $CaCl_2$.

Sucrose gradient (SG) buffer: 20 mM Tris–HCl pH 7.4, 5 mM NaCl, 50 mM EDTA.

AFM fixation buffer: 10 mM Tris–HCl pH 8.0, 5 mM NaCl, 0.3% glutaraldehyde.

Chloroplast isolation buffer: 350 mM sorbitol, 25 mM HEPES-NaOH pH 7.0, 2 mM EDTA, 2 mM sodium L-ascorbate.

50% percoll solution: 50% percoll, 1.5% PEG 6000, 0.5% ficoll 400, 0.3% BSA in chloroplast isolation buffer.

Chloroplast nucleoid isolation buffer: 20 mM Tris–HCl pH 7.0, 2% Nonidet P-40, 0.5 mM EDTA, 1.5 mM spermidine, 7 mM 2-mercaptoethanol, 0.4 mM PMSF.

2.3 Chemicals and Enzymes

5 mg/mL Lysozyme.
5 mg/mL Brij 58 (polyoxyethylene hexadecyl ether).
1 mg/mL Sodium deoxycholate.
1 mg/mL Spermidine.
RNase A (DNase free).
Protease K (RNA grade, DNase free).
Topoisomerase I (DNase free).
Micrococcal nuclease (MNase).

2.4 Equipment and Other Materials

Atomic Force Microscope; Bruker Nanoscope/MultiMode 3–8® (Bruker Corporation, USA) Tapping Mode™, Seiko SP3800N/SPI400® (Seiko Instruments, Japan) Dynamic Mode™.
AFM probe; OMCL-AC160TS (Olympus, Japan).
Nitrogen-gas blower with a yellow tip at the nozzle.
Mica (Ted Pella, Inc., Pelco Mica Sheets, Grade V5).

Gold particles (with diameters 10, 30, and 80 nm; BBInternational, Cardiff, UK).
pBluescript II (Stratagene).

3 Methods

3.1 Dissection of Bacterial Cells and Structural Analysis of Nucleoid by AFM

1. Harvest *E. coli* cells (100 μL aliquot of 5 mL culture in LB) by centrifugation (13,000 × *g*, 1 min, 4 °C), and wash the cells once with 1 mL of phosphate-buffered saline (PBS, pH 7.2).
2. Resuspend the cell pellet in 250 μL of PBS. Place a 50 μL aliquot on a round cover glass (15 mm in diameter) at room temperature for 5 min (*see* **Note 1**).
3. Remove excess liquid by gentle nitrogen gas blow from a micropipette tip at the end of tubing. Control the strength of gas by pressure regulator.
4. Immerse the cover glass in 2 mL of Lysis buffer A for 5 min, and add a lysozyme solution to the final conc. of 25 μg/mL (*see* **Note 2**).

5. Incubate the sample for 2 min at 25 °C, and add Brij 58 and sodium deoxycholate to final concentrations of 0.25 mg/mL and 0.1 mg/mL, respectively (*see* **Note 3**).
6. Incubate the sample for 10 min, and dry the cover glass under nitrogen gas.
7. Wash the sample surface gently with distilled water and dry it again.
8. Examine the sample by AFM in air with tapping mode (*see* **Note 4**).
9. Analyze Data (*see* **Notes 5** and **6**). In the case of nucleoid fibers, the sample (fiber diameter) is much smaller than the tip of the cantilever. Thus the following equation according to the circular cone model [17] can be used for the estimation of real diameter.

   $S = A \times W + B$, where S is the real width of the sample, W is the apparent width of the sample in the image, and A and B are constants determined by the tip characters. To obtain the values of A and B, gold particles (with diameters 10, 30, and 80 nm) can be used as the standards.
10. Plot the estimated size of all analyzed particles as histogram using software such as Origin and Excel (Figs 1g–j, 2c, d, g, h, k–n).
11. Determine the average size of the analyzed particles by Gaussian fitting to the obtained histogram (Figs 1g–j, 2c, g, h, k–n).

3.2 Further Enzymatic Treatment of Dissected Nucleoid Fibers

1. (Optional) Additional enzymatic treatment can be performed after **step 6** in Subheading 3.1. Treat the sample (lysed *E. coli* cells) with 5 μg/mL RNase A (more than 5 min) at room temperature, 1 mg/mL protease K (more than 5 min) at room temperature, and/or 4 U Topoisomerase I (more than 30 min) at 37 °C (*see* **Note 7**).
2. Dry the sample under nitrogen gas. Then move to **step 7** in Subheading 3.1 (*see* **Note 8**).

3.3 Dissection of Archaea Cells and Structural Analysis of Nucleoid by AFM

The dissection approach is also applicable to analysis of other nucleoids such as those of archaea, chloroplasts, and mitochondria. Archaea constitute a distinct domain of life, along with bacteria and eukarya [18]. Although both archaea and bacteria are prokaryotes, they use a totally different set of proteins for genome folding [19]. Most species in *Euryarchaeota*, a major phylum within archaea, encode proteins homologous to eukaryotic histones.

1. Culture *T. kodakarensis* KOD1 cells anaerobically in ASW-YT-S^0 medium at 85 °C and harvest by centrifugation for 1 min at 6,000 × *g* at 4 °C. Wash cells in 0.8× ASW buffer.

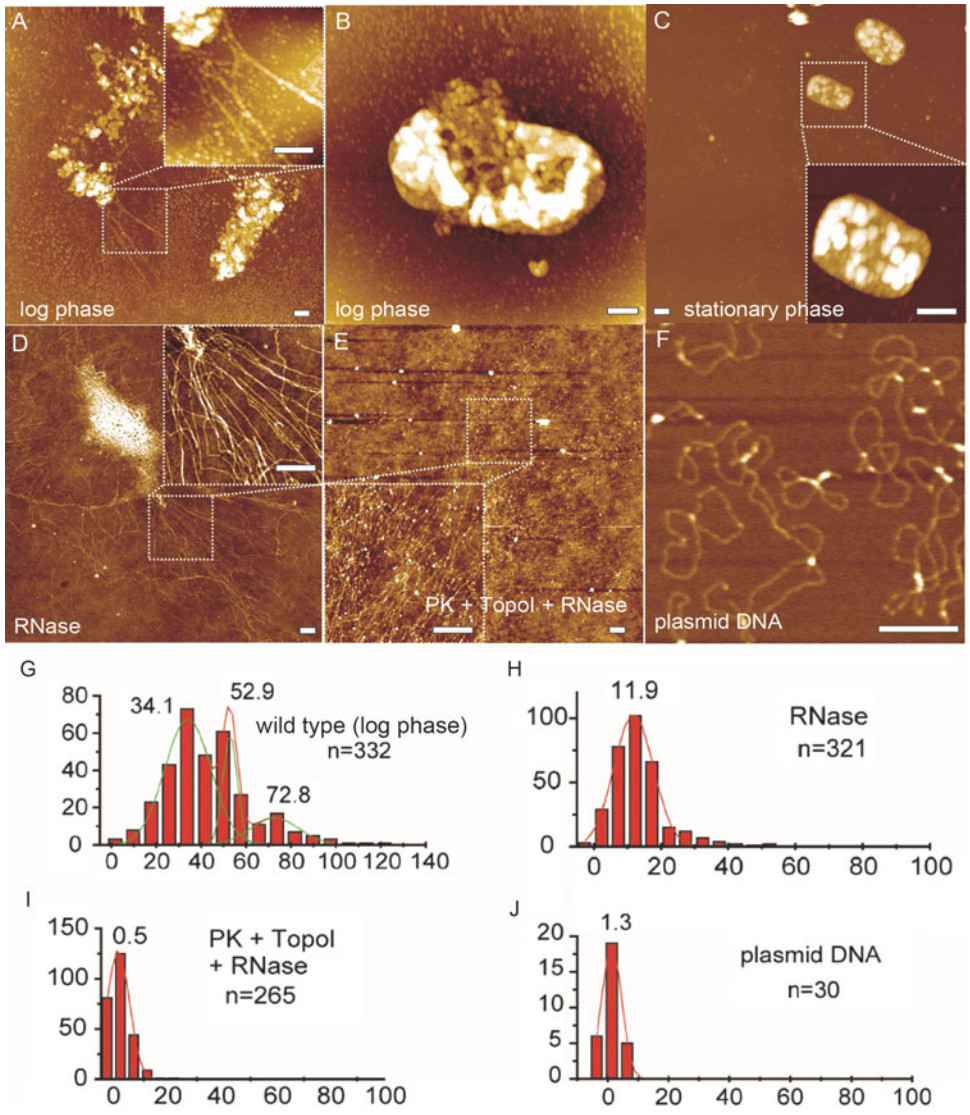


Fig. 1 AFM imaging and analysis of *E. coli* nucleoid structure. The lysed *E. coli* cells are observed by AFM: (**a**, **b**) in log phase, (**c**) in stationary phase, (**d**) after treatment with RNase A in log phase, (**e**) after treatment with RNase A, Protease K and Topoisomerase I in log phase, and (**f**) plasmid DNA (pRSFDuet-1, 3829 bp, Novagen). (**g**-**j**) Population distribution of the width of released fibers (*see* Methods 3.1. Step 9). Scale bars represent 500 nm

2. Make a cell suspension of $OD_{660} = 0.5$ by adding appropriate volume of 0.8× ASW. The cell suspension (50 μL) is applied onto a cover glass and left for 10 min at 25 °C (*see* **Note 9**).
3. Remove excess liquid by gentle blowing of nitrogen gas using a micropipette tip as the nozzle.
4. Lyse the cells on the cover glass through addition of Milli-Q water, and incubate it for 10 min at 25 °C (*see* **Note 10**).

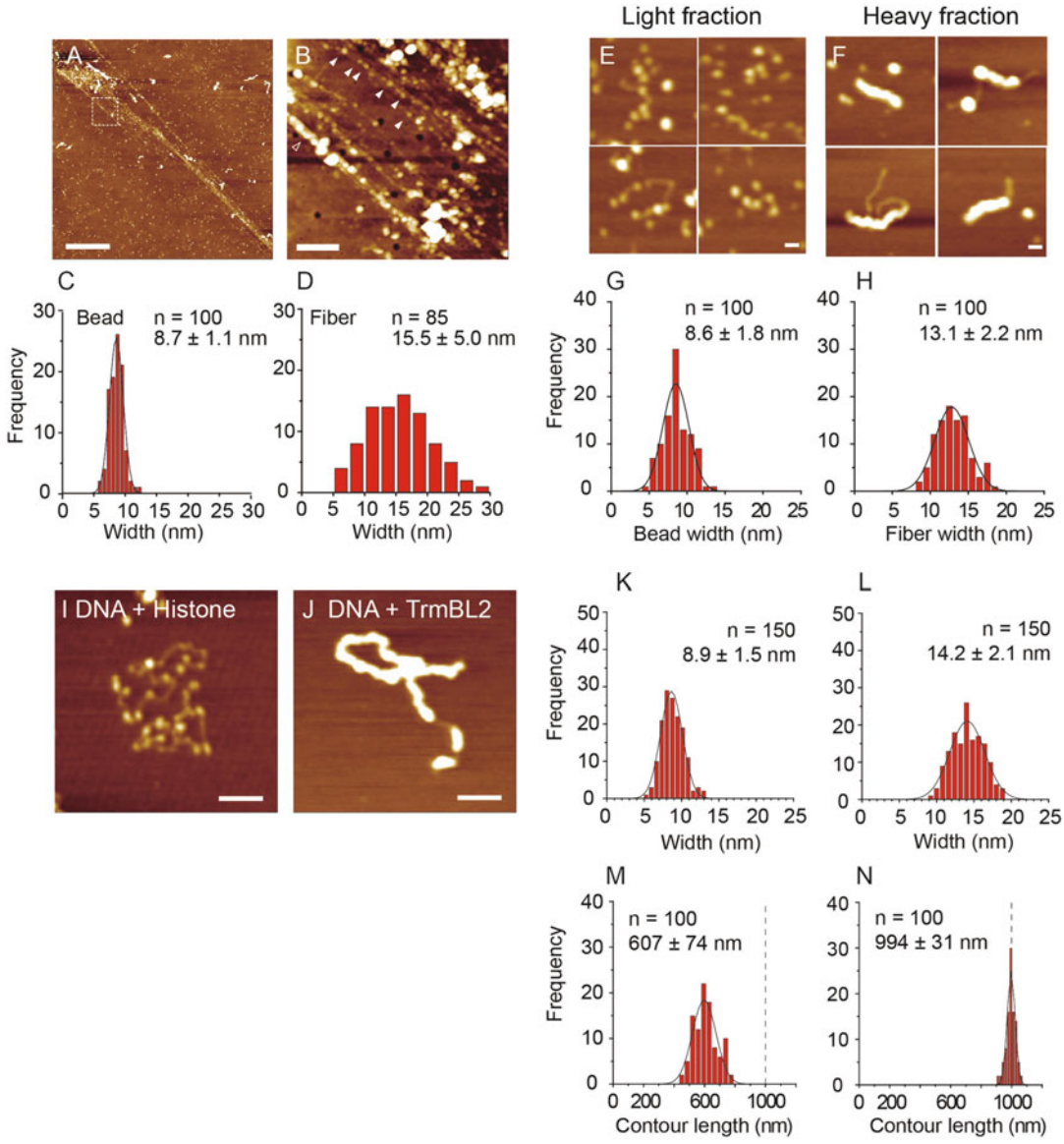


Fig. 2 AFM imaging and analysis of *T. kodakarensis* nucleoid Structure. (**a**–**d**) Chromosome fiber of *T. kodakarensis* analyzed through AFM (**a**) Chromosome released from *T. kodakarensis* in the log phase by on-substrate lysis. Scale bar, 2 μm. (**b**) Detailed AFM analysis of the chromosome fiber. Beads-on-a-string structures (filled triangle) as well as thick structures (open triangle) were observed in areas where chromosomes were well spread out (boxed area in (**a**)). Scale bar, 200 nm. (**c**–**d**) Histograms indicate the diameters of beads (**c**) and fibrous structures (**d**). The means ± s.d. of the distributions are indicated. (**e**–**h**) Separation of chromosome fragments with different structures and protein components by sucrose-density gradient sedimentation. (**e**–**f**) AFM images of chromosome fragments in the lighter fraction (**e**) and the heavier fraction (**f**). Scale bars, 50 nm. (**g**–**h**) Histograms below the images indicate the diameter of each structure. The means ± s.d. of the distributions are indicated. (**i**–**n**) Recombinant chromatin proteins form distinct structures on dsDNA in vitro. (**i**–**j**) AFM images of 3-kbp linear DNA incubated with recombinant Histone A (**i**), TrmBL2 (**j**) at a protein: DNA ratio of 10: 1 (wt: wt). Scale bars, 100 nm. (**k**–**l**) Diameters of the structures formed in the DNA by

5. Remove liquid gently by gentle blowing of nitrogen gas, and subject the sample to AFM examination in air with tapping mode. Fibers released from the cells can be analyzed by scanning areas where chromosome is well-spread (*see* **Note 4**).
6. Analyze data as in Subheading 3.1, **step 9** (*see* **Note 11**).

3.4 AFM Analysis of Distinct Chromosome Structures Separated through Sucrose Gradient Sedimentation

On-substrate lysis followed by AFM analysis gives a clue on what kind of structural units exist in the nucleoid. However, it does not provide information regarding the protein(s) giving rise to these different structures. One way to answer such a question is to use deletion strains which lack genes encoding specific nucleoid proteins and analyze the resulting chromosome structure. Another way is to combine AFM analysis with sucrose-density gradient sedimentation of Micrococcal nuclease (MNase)-digested chromosome fragments, enabling individual analysis of different structures depending on their densities, and mass spectrometry to identify the constituent proteins.

1. Culture *T. kodakarensis* cells anaerobically in ~60 mL of ASW-YT-S^0 medium at 85 °C, and then harvest and wash the cells with 0.8× ASW as in Subheading 3.3, **step 1**.
2. Make cell suspensions with 50 μL 0.8× ASW, and disrupt the cells by the addition of 450 μL Extraction buffer containing sorbitol and a detergent Triton X-100 (*see* **Note 12**).
3. After incubating for 10 min at 4 °C, separate soluble proteins (supernatant) and a chromosomal DNA-enriched insoluble fraction (pellet) by centrifugation at 14,000 × *g* for 20 min. Wash the insoluble chromatin fraction, containing DNA and chromatin-associated protein, with the extraction buffer. Chromatin can be frozen at this step for later use (*see* **Note 13**).
4. Digest chromatin fraction extracted from ~60 mL of culture with 1 unit of MNase in 100 μL of MNase buffer for 30 min at 37 °C in the presence of 10 μg/μL RNase. Stop the reaction by adding EDTA to 50 mM.
5. Apply the digested chromatin onto 5 mL of 5–10% sucrose density gradient made in SG buffer. Perform ultracentrifugation at 28,000 × *g* for 16–24 h at 4 °C in a swinging bucket until sedimentation equilibrium has been attained.
6. Collect 500 μL each of ten fractions from the surface by means of careful pipetting (*see* **Note 14**).

Fig. 2 (continued) Histone A (**k**), TrmBL2 (**l**) are shown in histograms. (**m–n**) Contour length of the 3-kbp DNA molecule when bound with recombinant Histone A (**m**), TrmBL2 (**n**) are indicated in histograms. Dotted lines indicate the length of a 3 kbp double-stranded DNA (~1000 nm). The means ± s.d. of the distributions are indicated

7. Pass 50 μL of each fraction through a size exclusion spin column pre-equilibrated with SG buffer (*see* **Note 15**).
8. Fix the samples containing chromosome fragments with 0.3% glutaraldehyde for 30 min at 25 °C, dilute the samples to concentration suitable for AFM analysis, and apply the samples onto a mica surface. Mica is pretreated with 10 mM spermidine for 10 min at room temperature and washed with water, dried with nitrogen gas before sample application. Analyze chromosome fragments by AFM (*see* **Notes 16** and **17**).

3.5 Reconstitution of Nucleoid Fibers.

Reconstitution of chromosome structure with defined factors offers the possibility of proving more directly which protein is responsible for forming each structural unit (*see* **Note 18**).

1. Digest a 3 kbp plasmid pBluescript II (Stratagene) with the restriction enzyme *Hin*dIII, which is a single-cutting enzyme on this plasmid. Purify the linear DNA and mix the DNA with each recombinant protein (in 10 mM Tris–HCl pH 8.0, 200 mM NaCl) in DNA-to-protein ratios (wt:wt) of 1:0, 1:0.3, 1:1, 1:3, or 1:10 and incubated for 20 min at 50 °C (*see* **Note 19**).
2. Dilute the protein–DNA complexes to 1:20 in AFM fixation buffer. After incubation for 30 min at 25 °C, deposit 10 μL of the mixture (containing ~2.5 ng of DNA) on mica that had been pretreated with 10 mM spermidine for 10 min.
3. After 10 min of incubation, wash the mica with 1 mL of pure water, dry the sample with nitrogen gas, and subject the samples to AFM analysis (*see* **Note 20**).

3.6 Dissection of Chloroplasts to Analyze Nucleoid Structure by AFM

Chloroplasts can be isolated from plant cells such as spinach by the following procedure. Genomic DNA of isolated chloroplast is released by a treatment with hypotonic solution.

1. Cut 50–100 leaves of Spinach (*Spinacia oleracea*) into pieces on ice with scissors in 50 mL of chloroplast isolation buffer. Then, further crush them by homogenization.
2. Filter the sample through a double layer of absorbent gauze and a double layer of Miracloth.
3. Centrifuge the filtered sample at 1000 × *g* for 5 min, and resuspend the pellet in 500 μL of chloroplast isolation buffer.
4. Pre-centrifuge 5 mL of 50% percoll solution in a fixed angle rotor at 43,000 × *g* for 10 min.
5. Load the resuspended sample onto the linear percoll gradient, centrifuge the sample in a swing-out rotor at 7800 × *g* for 10 min, and collect intact chloroplast (lower green band in percoll gradient) (*see* **Note 21**).

6. Incubate the isolated chloroplasts in chloroplast nucleoid isolation buffer at 25 °C for 30 min.
7. Place the collected chloroplasts onto a coverglass, and remove the extra liquid under nitrogen gas.
8. Immerse the sample in milli-Q water for 5 min at room temperature, and dry the sample by nitrogen gas (*see* **Note 22**).
9. Observe the sample by AFM in air with tapping mode and analyze the data (*see* **Note 23**).

3.7 Dissection of Mitochondria to Analyze Nucleoid Structure by AFM

Mitochondria can be isolated from cultured cells such as HeLa cell using mitochondria isolation kits. Genomic DNA of isolated mitochondria is released by a treatment with hypotonic solution.

1. Culture HeLa S3 cells in Dulbecco's modified Eagle medium (Sigma) supplemented with 10% fetal bovine serum (Bio Whitaker) in 5% CO_2 at 37 °C. Harvest cultured cells with 70% confluency in a 100 mm dish, and centrifuge the cells at 1000 × *g* for 5 min at 4 °C.
2. Isolate HeLa cell mitochondria by mitochondria isolation kits (*see* **Note 24**), and suspend the mitochondria in 200 μL PBS with protein inhibitor cocktail.
3. Place the collected mitochondria onto a coverglass, and remove the extra liquid under nitrogen gas.
4. Immerse the sample in milli-Q water for 5 min at room temperature, and dry the sample by nitrogen gas (*see* **Note 25**).
5. Observe the sample by AFM in air with tapping mode.

4 Notes

1. Bacterial cells easily attach on the cover glass.
2. *E. coli* is a Gram-negative bacterium, and removal of the cell wall and membrane is necessary to observe the nucleoid. In this procedure, we use an enzyme (lysozyme) and detergent (Brij 58 and sodium deoxycholate) to remove cell wall and membrane, respectively. This procedure is applicable for investigation of nucleoids of Gram-positive *Bacillus subtilis*. For the analysis of the Gram-positive *Staphylococcus aureus*, lysozyme must be replaced with Lysostaphin (80 μg/mL) and N-Acetylmuramidase (170 μg/mL).
3. The nucleoid is dispersed from the cell after removing cell wall and membrane.
4. We use Nanoscope III, IV or V (Tapping Mode™) or SP3800N (Dynamic Mode™) with OMCL-AC160TS probe. Images are analyzed using Nanoscope software. Bead diameter

and fiber width are calculated based on apparent bottom-to-bottom distance of particles, using the circular cone model to estimate real dimension of sample width (Subheading 3.1, **step 9**).

5. The apparent horizontal dimensions of the samples measured by AFM are generally much larger than the real dimensions owing to the effects of the edge curvature and point angle of the cantilever [20, 21]. Therefore, the real dimensions need to be estimated from the apparent dimensions. AFM image data consists of a set of pixel data. If the image is taken in 2 μm × 2 μm with 512 × 512 pixel, each pixel represents about 5 nm × 5 nm. Thus, in this case, it is not reasonable to discuss the size difference less than 5 nm. Researchers should consider appropriate image size and pixel to analyze the sample size.
6. When *E. coli* cells are harvested in the log phase, and lysed under physiological salt conditions, fibers with widths of 30–80 nm (thicker than the 2 nm naked DNA) are released (Fig. 1a). Loop structures composed of the 80 nm fibers are also detectable inside the cell (Fig. 1b). These fibers are further condensed into beaded structures toward the stationary phase (Fig. 1c). This implies the existence of step-wise folding mechanisms of DNA, in achieving the higher-order architectures. According to our analyses of *E. coli* strains lacking a series of the nucleoid proteins, the lack of none of these nucleoid proteins disrupts the 30–80 nm fiber structures [17]. Interestingly, the 30–80 nm fiber structures are also commonly found in *S. aureus* and *Clostridium perfringens* [22]. *S. aureus* does not encode the genes for nucleoid proteins found in *E. coli* such as H-NS and FIS. Shot-gun MS/MS analyses showed that just a few percent of total proteins in the isolated nucleoid was orthologous between *E. coli* and *S. aureus* [9].
7. RNase A and Topoisomerase I can be added simultaneously, but protease K treatment should be separate, as protease K degrades both RNase A and Topoisomerase I. Topoisomerase I works in Topo I buffer. RNase A and Protease K work both in milli-Q and Topo I buffer.
8. Enzymatic treatments enable further dissection. RNase A treatment of the lysed cells converts thicker fibers into 10 nm fibers in both *E. coli* and *S. aureus* (Fig. 1c), suggesting that the 10 nm fibers are the thinnest fiber unit without RNA [23]. In the case of *E. coli*, the 10 nm fiber is the thinnest fiber unit even in the mutants lacking any one of the major nucleoid proteins. These 10 nm fibers are disrupted into naked-DNA with a diameter of 2 nm by additional treatment with protease K and topoisomerase I, suggesting that nucleoid proteins and DNA supercoiling are required to build up the 10 nm fiber (Fig. 1d) [17].

9. Archaeal cells are easily attached on a cover glass.
10. On-substrate lysis and dissection of the hyperthermophylic archaeaon *T. kodakarensis* can be performed in a similar way as for *E. coli*. Enzymatic treatment is not required because this cell is easily disrupted by hypotonic solution. However, this method may not be applicable to all other types of archaea. The cell rupture method needs to be determined for each type of archaea.
11. On-substrate lysis of a histone possessing hyperthermophilic archaeaon *T. kodakarensis* exposed both "beads-on-a-string" structures which has 8–10 nm diameter and fiber structures with width of 10–20 nm, suggesting the existence of several different structural units in its nucleoid (Fig. 2a–d) [24].
12. As described above, enzymatic treatment is not necessary to disrupt the cell. Before adding Extraction buffer, cells are resuspended in small volume of buffer to achieve homogeneous cell lysis.
13. This method is based on a method previously reported for a *Pyrococcus* species, which is closely related to *Thermococcus* [25]. Cells are disrupted by detergent. Sorbitol is added to prevent osmotic shock. This simple separation method may only work well for species closely related to *Thermococcus*. Other methods should be developed for other types of archaea. Separation of chromosomal DNA can be tested by agarose gel electrophoresis. If antibodies specific to proteins with known cellular localization are available, they can be used in Western blotting to prove a successful separation of the chromosome and other cellular components.
14. A fraction collecting device can be used if available.
15. Size exclusion (gel filtration) is performed to remove free (unbound) proteins and sucrose which might interfere with AFM analysis. For example, Micro Bio-Spin Tris column (Bio-Rad) can be used. Protein is concentrated from the rest of the fraction for SDS-PAGE, Western blotting, or mass spectrometry analysis.
16. Glutaraldehyde treatment could be omitted if overfixation is a concern. Dilution factor to achieve appropriate concentration of chromosome fragment usually would be enough to terminate glutaraldehyde fixation. A given fraction can be submitted to either structure analysis by AFM or the identification of protein composition by mass spectrometry.
17. In the case of *T. kodakarensis*, AFM analysis shows that "beads-on-a-string" and fibrous structures with diameters comparable to those of the structures observed in the on-substrate lysis (Fig. 2a–d) are enriched in the lighter and heavier fractions of a

sucrose-density gradient, respectively (Fig. 2e–h). Mass spectrometry identifies histone in the lighter fraction and TrmBL2, a transcription factor-like protein, in the heavier fraction, suggesting that these proteins are the building blocks of each structure. When a deletion strain of TrmBL2 gene is subjected to the same analysis, fibrous structure is no longer detected in the heavier fraction, further supporting that TrmBL2 is the component of the fibrous structure [24].

18. In the case of *E. coli*, reconstitution of higher-order nucleoid fibers has not been successful because proteins involved in the fiber structures seem to be heterogeneous. Although researchers identified so-called nucleoid proteins such as HU, H-NS, FIS, etc., each of them is not essential to sustain the higher-order fiber structures thicker than 10 nm [17].
19. Instead of using a bacterial plasmid, using its own genomic DNA sequence might be preferable to reconstitute physiological structures. For this purpose, an appropriate genomic region can be amplified by PCR, purified and subjected to incubation with recombinant proteins. Methods for protein expression and purification are described elsewhere [24].
20. Reconstitution of DNA-protein complex with recombinant Histone A and B, which were identified in *T. kodakarensis*, successfully exhibited "beads-on-a-string" structures with a diameter of approximately 9 nm (Fig. 2i, k). Recombinant TrmBL2 binds on DNA and forms relatively smooth fiber structures with a diameter of 15 nm (Fig. 2j, l). These structures are similar to those found in on-substrate lysis of the cell (Fig. 2a–d), or in sucrose- density gradient fractions (Fig. 2e–h). AFM analysis also gives an indication of how these proteins interact with genomic DNA. By measuring the contour length of the DNA molecule on the AFM images, it is possible to tell whether DNA is wrapped around the protein as in the case of histone protein (Fig. 2m), or the protein simply binds and covers the DNA molecule like TrmBL2 (Fig. 2n). In addition to re-constituting fiber structures with a diameter of 10 nm by using DNA fragments and Alba, which is a well-known DNA-binding protein in archaea, we revealed the fundamental nucleoid structure existing in the cell of *T. kodakarensis* [24].
21. In **step 4**, after centrifugation, two separated green bands may be found. The upper band mainly consists of partially disrupted chloroplast.
22. Isolated chloroplasts can be burst by hypotonic liquid such as water. Nucleoid is observable around the debris of chloroplasts. Enzymatic treatment is applicable as is the case of bacterial nucleoid.

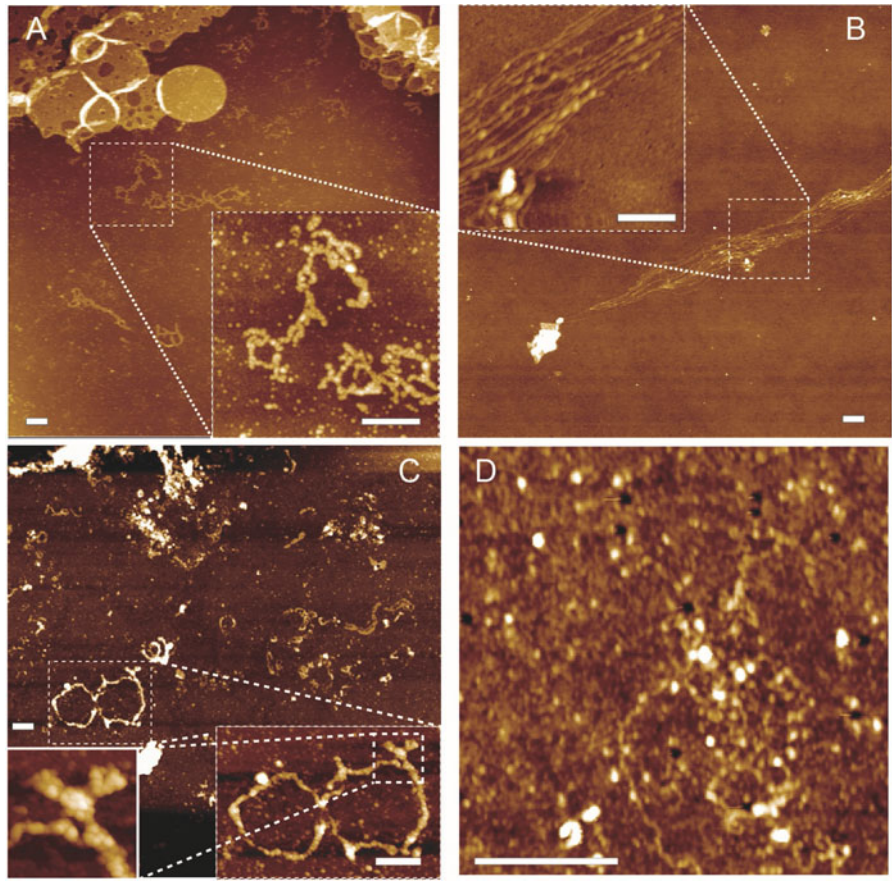


Fig. 3 AFM images of spinach chloroplast and human mitochondria nucleoids. (**a**) The lysed chloroplasts and (**b**) lysed chloroplasts subsequently treated by RNase A observed by AFM. (**c**) The lysed human mitochondria and (**d**) subsequently treated lysed human mitochondria observed by AFM. Scale bars represent 500 nm

23. In the case of chloroplast and mitochondria, the lyses of isolated chloroplasts and mitochondria exposed the 30 nm beaded structures which were, in turn, disrupted into 10 nm fibers by the treatment of RNase A (Fig. 3) [22].
24. Mitochondria isolation kits such as MITOISO2 SIGMA.
25. Isolated mitochondria can be burst by hypotonic solution. Nucleoid is observable around the debris of mitochondria. Enzymatic treatment is applicable as is the case of bacterial nucleoid and chloroplast.

References

1. Poplawski A, Bernander R (1997) Nucleoid structure and distribution in thermophilic Archaea. J Bacteriol 179(24):7625–7630
2. Robinow C, Kellenberger E (1994) The bacterial nucleoid revisited. Microbiol Rev 58 (2):211–232
3. Kavenoff R, Bowen BC (1976) Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*. Chromosoma 59 (2):89–101
4. Kavenoff R, Ryder OA (1976) Electron microscopy of membrane-associated folded

chromosomes of *Escherichia coli*. Chromosoma 55(1):13–25

5. Sloof P, Maagdelijn A, Boswinkel E (1983) Folding of prokaryotic DNA. Isolation and characterization of nucleoids from *Bacillus licheniformis*. J Mol Biol 163(2):277–297
6. Azam TA, Ishihama A (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. J Biol Chem 274 (46):33105–33113
7. Drlica K, Rouviere-Yaniv J (1987) Histonelike proteins of bacteria. Microbiol Rev 51 (3):301–319
8. Kundu TK, Kusano S, Ishihama A (1997) Promoter selectivity of *Escherichia coli* RNA polymerase sigmaF holoenzyme involved in transcription of flagellar and chemotaxis genes. J Bacteriol 179(13):4264–4269
9. Ohniwa RL, Ushijima Y, Saito S, Morikawa K (2011) Proteomic analyses of nucleoid-associated proteins in *Escherichia coli*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, and *Staphylococcus aureus*. PLoS One 6(4):e19172. https://doi.org/10.1371/journal.pone.0019172
10. Rouviere-Yaniv J, Yaniv M, Germond JE (1979) *E. coli* DNA binding protein HU forms nucleosomelike structure with circular double-stranded DNA. Cell 17(2):265–274
11. van Noort J, Verbrugge S, Goosen N, Dekker C, Dame RT (2004) Dual architectural roles of HU: formation of flexible hinges and rigid filaments. Proc Natl Acad Sci U S A 101 (18):6969–6974
12. Maurer S, Fritz J, Muskhelishvili G (2009) A systematic in vitro study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of DNA organization. J Mol Biol 387(5):1261–1276. https://doi.org/10.1016/j.jmb.2009.02.050
13. Dame RT, Wyman C, Goosen N (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Res 28(18):3504–3510
14. Dame RT, Noom MC, Wuite GJ (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. Nature 444(7117):387–390
15. Schneider R, Lurz R, Luder G, Tolksdorf C, Travers A, Muskhelishvili G (2001) An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA. Nucleic Acids Res 29(24):5107–5114
16. Murphy LD, Zimmerman SB (1997) Isolation and characterization of spermidine nucleoids from *Escherichia coli*. J Struct Biol 119 (3):321–335
17. Ohniwa RL, Muchaku H, Saito S, Wada C, Morikawa K (2013) Atomic force microscopy analysis of the role of major DNA-binding proteins in organization of the nucleoid in *Escherichia coli*. PLoS One 8(8):e72954. https://doi.org/10.1371/journal.pone.0072954
18. Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87 (12):4576–4579
19. Peeters E, Driessen RP, Werner F, Dame RT (2015) The interplay between nucleoid organization and transcription in archaeal genomes. Nat Rev Microbiol 13(6):333–341. https://doi.org/10.1038/nrmicro3467
20. Bustamante C, Rivetti C, Keller DJ (1997) Scanning force microscopy under aqueous solutions. Curr Opin Struct Biol 7(5): 709–716
21. Nettikadan S, Tokumasu F, Takeyasu K (1996) Quantitative analysis of the transcription factor AP2 binding to DNA by atomic force microscopy. Biochem Biophys Res Commun 226 (3):645–649
22. Takeyasu K, Kim J, Ohniwa RL, Kobori T, Inose Y, Morikawa K, Ohta T, Ishihama A, Yoshimura SH (2004) Genome architecture studied by nanoscale imaging: analyses among bacterial phyla and their implication to eukaryotic genome folding. Cytogenet Genome Res 107(1–2):38–48
23. Ohniwa RL, Morikawa K, Takeshita SL, Kim J, Ohta T, Wada C, Takeyasu K (2007) Transcription-coupled nucleoid architecture in bacteria. Genes Cells 12(10):1141–1152. https://doi.org/10.1111/j.1365-2443.2007.01125.x
24. Maruyama H, Shin M, Oda T, Matsumi R, Ohniwa RL, Itoh T, Shirahige K, Imanaka T, Atomi H, Yoshimura SH, Takeyasu K (2011) Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon *Thermococcus kodakarensis*. Mol Biol Cell 22 (3):386–398. https://doi.org/10.1091/mbc.E10-08-0668
25. Matsunaga F, Forterre P, Ishino Y, Myllykallio H (2001) *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proc Natl Acad Sci U S A 98 (20):11152–11157. https://doi.org/10.1073/pnas.191387498

Chapter 10

Dynamic Light Scattering of DNA-Ligand Complexes

Guangcan Yang and Yanwei Wang

Abstract

Dynamic Light Scattering (DLS) enables the characterization of sizes and electrokinetic properties of colloids, polymers, macromolecules. DNA is a charged semiflexible polyelectrolyte, which is condensed or compacted by counterions, proteins, and other condensing agents in processes such as chromosome compaction and gene therapeutic applications. DNA condensation is closely related to charge screening, since packaging requires effective neutralization of its surface negative charges. In this chapter, we describe in detail the protocol for DLS of DNA-ligand complexes. As an example, we describe data for condensation of DNA by chitosan and the measurement of size, zeta potential, and electrophoretic mobility of the DNA-ligand complex by DLS.

Key words Dynamic light scattering, DNA condensation, Chitosan, Zeta potential, Electrophoretic mobility

1 Introduction

Light scattering is a consequence of the interaction of light with small particles or molecules in medium. When particles become larger than the tenth of wavelength of light, the scattering changes from being isotropic to a distortion in the forward scattering direction. When the size of the particles increases further, being equivalent to or greater than the wavelength of the laser, the scattering becomes a complex function with maxima and minima along the scattering angle. The Dynamic Light Scattering (DLS) technique was developed for analyzing the size distribution profile of small particles in suspension or polymers in solution [1, 2].

In practice, DLS signal is obtained by the extraction of spectral information derived from time-dependent fluctuations of the light scattered from a limited volume within the sample. More precisely, when a suspension of particles is illuminated with a monochromatic coherent beam of light, the corresponding scattered light waves spread out in all directions. The interference of scattered waves in the far field region generates a scattered light intensity angular

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_10,

distribution. Due to the random motion of the suspended particles within the sample the interference can be stochastically either constructive or destructive. The stochastic interference light intensity can be recorded by photomultipliers and is usually analyzed by means of the intensity or photon auto-correlation function. The correlation function method is also referred to as photon correlation spectroscopy or quasi-elastic light scattering. In the time domain, the autocorrelation function (ACF) usually starts decaying from zero delay time, and depends on particle size. The faster dynamics due to smaller particles lead to faster decorrelation of the scattered intensity trace. Actually, the intensity ACF is the Fourier transformation of the power spectrum. Therefore, we can perform DLS measurements in the spectral domain, and then transform them into the time domain.

From the signal of DLS, we can extract speed information of particles undergoing Brownian motion, influenced by particle size, medium viscosity, and temperature. According to the Stokes-Einstein theory, we can calculate the particle size from these measured quantities. When an oscillating electric field is applied to the sample, an electrophoretic light scattering is developed for surface charged particles. Thus, we can obtain additional electrokinetic properties such as zeta potential and electrophoretic mobility of particles.

1.1 Particle Size

In the section, we describe the measurement of particle sizes by DLS. As we have mentioned before, particles in medium scatter the light and thereby imprint information about their motion in the scattered light. Because the fluctuation of the scattered light is related to the stochastic motion of particles, its analysis yields physical information of the particles. Experimentally one investigates intensity fluctuations by computing the intensity correlation function, whose analysis and decomposition provide the diffusion coefficient of the particles D, relating to the radius R of the particles.

More specifically, the suspended particles of the colloidal dispersion in measurement naturally undergo Brownian motion. This stochastic motion leads to fluctuations of the distances between the particles and hence also in fluctuations of the phase relations of the scattered light. Meanwhile, the number of particles within the scattering volume may vary in time. Both the factors yield a fluctuating scattered intensity. The corresponding measured normalized intensity correlation function can be expressed as

$$g_2(q,\tau) = \frac{\langle I_s(q,t) I_s(q,t+\tau)\rangle}{\langle |I_s(q,t)|^2 \rangle} \tag{1}$$

where I_s is the scattered intensity, τ is the time lag, and q is the scattering vector. The electric field correlation function $g_1(q, \tau)$ is defined as

$$g_1(q,\tau) = \frac{\langle E_s(q,t)E_s^*(q,t+\tau)\rangle}{\langle |E_s(q,t)|^2\rangle} \tag{2}$$

where E_s is the electric field of scattered light. The two correlation functions are connected by the expression

$$g_2(q,\tau) = 1 + \beta|g_1(q,\tau)|^2 \tag{3}$$

where β is a factor representing the degree of spatial coherence of the scattered light over the detector. The field correlation function decays with time and can be used to determine the diffusion coefficient D of the scattering particles. For a monodisperses ample, we can fit it to an exponential function to extract the decay rate Γ, which is related to the diffusion coefficient by the expression

$$\Gamma = q^2 D \tag{4}$$

Therefore, one obtains the diffusion coefficient once we have the decay rate of the field correlation function. Furthermore, by using the Stokes-Einstein equation one can obtain the hydrodynamic radius as

$$R = \frac{kT}{6\pi\eta D} \tag{5}$$

where k is the Boltzmann constant, T is the temperature of the suspending medium in Kelvins, and η is its viscosity.

The setup of DLS is shown schematically in Fig. 1. A monochromatic light source, usually a laser, is shot through an attenuator and into a sample. The scattered light then goes through a polarizer and focusing lens and is next collected by a photomultiplier. All of the molecules in the solution are being hit with the light and all of the molecules diffract the light in all directions. The diffracted light from all of the molecules can either interfere constructively or destructively. This process is repeated at short time intervals and the resulting patterns are analyzed by an autocorrelator that compares the intensity of light at each spot over time. The output of the autocorrelator is fed into a computer and analyzed to give information about particle sizes.

1.2 Electrophoretic Mobility and Zeta Potential of Particle

Colloids or polyelectrolytes with dissociating groups in aqueous solution usually have an electric surface charge, resulting in an electrostatic Coulomb force when an external electric field is applied. According to the double layer theory [3], all surface charges in solution are screened by a diffuse layer of ions, which has the same absolute charge but opposite sign with respect to that

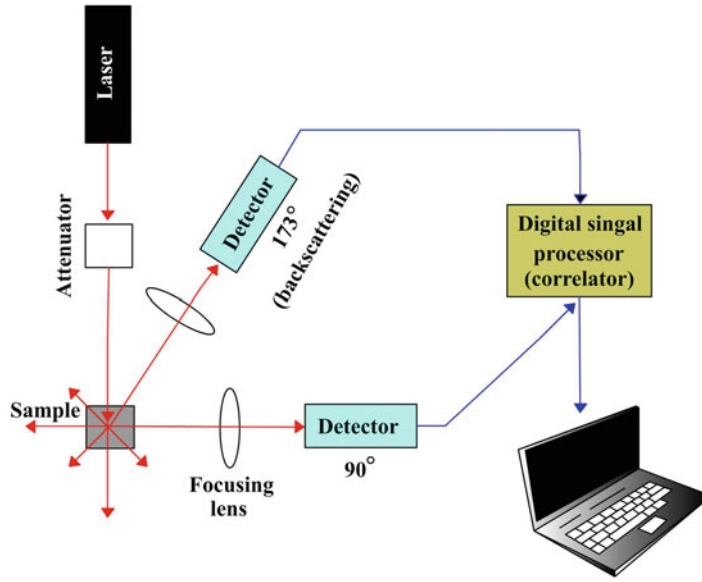


Fig. 1 Schematic illustration of dynamic light scattering

of the surface charge. The electric field also exerts a force on the ions in the diffuse layer which has direction opposite to that acting on the surface charge. This latter force is not actually applied to the particle, but to the ions in the diffuse layer located at some distance from the particle surface, and part of it is transferred all the way to the particle surface through viscous stress. This part of the force is also called electrophoretic retardation force. In the electric field the charged particle is at steady movement through the diffuse layer with zero total resulting force. The electrophoretic technique can be incorporated in DLS to probe electrokinetic properties of colloids or polyelectrolytes.

In an instrument using electrophoretic light scattering, a sample is subjected to light while its particles move under the influence of an oscillating electric field. In DLS, the phase of laser light scattered from the particles is monitored over time; particles drifting at constant velocity in an electric field yield a phase that evolves linearly in time at a rate proportional to their mobility. Analysis of shifting wavelength, or Doppler shift, in the resulting scattered light allows determination of electrophoretic mobility—from which zeta potential can be calculated.

Figure 2 shows the capillary sample cell for zeta potential and electrophoretic mobility measurement. In the cell, we can find two electrodes at the terminals of U channel, where an alternating electric field is applied to particles in solution, inducing electrophoretic motion of particles. When a beam of laser enters the sample cell through the bottom part of cell, as shown in the inset of Fig. 2,

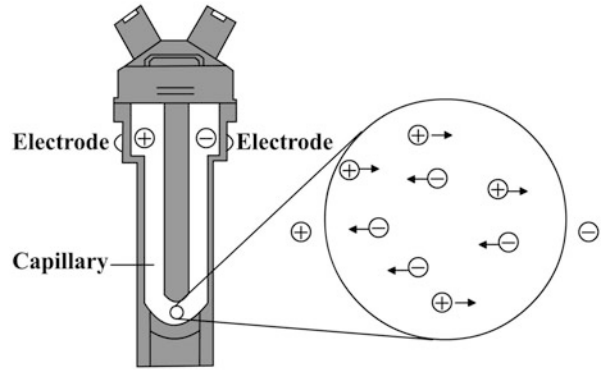


Fig. 2 Schematic illustration of a capillary sample cell for zeta potential measurement

the frequency of the scattered light is shifted due to the directional motion of the particles with the applied electric field. The scattered light is coherent with the reference light, producing a very small modulated light source, whose frequency is equal to the difference between the reference and scattered light frequencies. The frequency difference is related to the velocity of the particles and can be expressed as

$$\Delta f = 2v \sin(\theta/2)/\lambda \tag{6}$$

where v is the velocity of the particles to be measured, λ is the wavelength of the incident laser, θ is scattering angle. A second beam of light (the reference beam) is mixed with the scattered beam in order to sensitively extract the frequency shift in the scattered light. The scattered and reference beams interfere to form a beat, which is used to extract the magnitude of the frequency shift, then being used to determine the particle velocity. The electrophoretic mobility of colloids can be expressed as $\mu = \nu/E$, where v is particle velocity and E is the applied electric field. Thus the zeta potential of particles can be calculated by the Smoluchowski equation

$$\xi = \frac{\mu\eta}{\varepsilon} \tag{7}$$

where ε is the dielectric constant of the dispersion medium, and η is dynamic viscosity of the dispersion medium.

In this protocol, electrophoretic-mobility measurements were carried out using a Malvern Zetasizer NanoZS90 instrument using a combination of laser Doppler velocimetry and phase analysis light scattering (PALS). A typical measuring result of zeta potential is shown in Fig. 3, where the Tris buffered DNA sample was used.

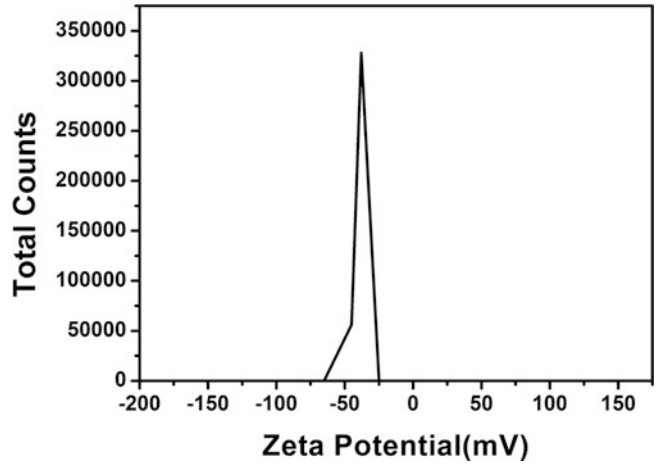


Fig. 3 A typical DNA zeta potential distribution by DLS

1.3 DNA Condensation and Electrophoretic Mobility of DNA-Ligand Complex

DNA is one of the most important biological polyelectrolytes, and is highly negatively charged in solution. It is tightly packaged from viruses to eukaryotic cells in order to store, transport, and preserve the genetic material. In vitro, the highly-charged stiff polymer can be condensed into compact structures by multivalent ions and many other condensing agents or ligands [4–6]. The understanding of DNA compaction is not only important for the study of fundamental biological processes such as chromosome compacting, but also for the development of new gene carriers in therapeutic applications [7–9]. DNA condensation refers to the process of compacting DNA molecules in vitro or in vivo. Meanwhile, DNA condensation has many potential applications in medicine and biotechnology [10]. Mechanistic details of DNA packing are essential for its functioning in the process of gene regulation in living systems [11].

The diameter of DNA is about 2 nm, while the length of a stretched single molecule may be up to several dozens of centimeters depending on the organism. Many features of the DNA double helix contribute to its large stiffness, including the mechanical properties of the sugar-phosphate backbone, electrostatic repulsion between phosphates, stacking interactions between the bases of each individual strand, and strand-strand interactions. Obviously, an unpacked DNA would randomly occupy a much larger volume than when it is orderly packed. To cope with volume constraints, DNA can pack itself in the appropriate solution conditions with the help of ions and condensing agents. When DNA condensation occurs, extended DNA chains collapse into compact and orderly particles containing only one or a few molecules. On the other hand, DNA compaction is closely related to its charge neutralization, because structural packaging requires an effective screening of the negative charges on DNA. In some conditions,

overcompensation or charge inversion occurs when the charge of counterions surrounding the DNA surface is greater than the bare charge of polyelectrolyte itself [12–14]. Charge inversion of DNA is a counterintuitive phenomenon in which the effective charge of DNA switches its sign from negative to positive in the presence of some counterions or ligands. However, the underlying microscopic mechanism of attraction between like-charged macroions and their charge inversion is still controversial although these effects have been observed in various systems from simple colloids to intricate protein-DNA complexes by various experimental approaches. Two driving mechanisms have been proposed: pure electrostatic interaction from a strong correlation effect and specific chemical adsorption of counterions. In some cases, the strong correlation effect plays a dominant role as shown in mixing counterions, changing pH value of solution or lowering dielectric constant of solution [15–17].

DLS is a reasonable and noninvasive method to study DNA collapse transition because the substantial reduction in its gyration radius in the compact form leads to a large increase in the scattered intensity. It can also provide the information of zeta potential or electrophoretic mobility of DNA-ligand complexes when a specific designed capillary sample cell with electrodes is used. They characterize the charge screening or charge inversion of DNA induced by counterions, ligands, or other condensing agents.

In this protocol, we use chitosan as a ligand to condense lambda DNA. Chitosan is a linear polysaccharide composed of deacetylated unit and acetylated unit. It is made by treating the chitin shells of shrimp and other crustaceans with an alkaline substance, like sodium hydroxide. It is highly positively charged in solution. Therefore, the linear copolymer chitosan has recently emerged as an attractive gene delivery vehicle because of its nontoxic and biodegradable nature [18–21]. Furthermore, DNA–chitosan complexes also have recently been reported to have high transfection efficiency [22–24].

2 Materials and Experimental Conditions

1. Double-stranded λ-phage DNA (48,502 bp) (New England Biolabs) diluted in 1 × TE buffer (10 mM Tris–HCl and 1 mM EDTA, pH = 8.0) at a DNA concentration of 500 ng μL^{-1}.
2. Chitosan (MW = 5000, hydrosoluble).
3. Measurement buffer: 10 mM Tris–HCl, pH 8.0.

   All solutions are made with 18.2 MΩ deionized water purified through the Milli-Q water purification system (Millipore

Corporation, USA). All reagents are prepared and stored at room temperature (unless indicated otherwise).

4. ZetasizernanoZS90 (Malvern, UK0 or NanoBrook Omni (Brookhaven, USA).
5. The laser source is a He-Ne gas laser ($\lambda = 633$ nm). Light scattering by the avalanche photodiode mounted on the goniometric arm in the direction of the incident radiation.
6. DTS1070 sample cell for zeta potential measurement or ZE0118 sample cell for particle size measurement.

3 Methods

3.1 Particle Size Measurement

1. Turn on power of the DLS instrument, then wait for 30 min for preheating. When the indicator light is changed from yellow to green, it is ready for measurement.
2. Start up the Zetasizer Software, then click File--New--Measurement file to select or setup a storage path and input a file name to store for measurement conditions and data.
3. Open Measure menu, click Manual, and use the right key single-click measurement type, select Size to set measuring conditions.
4. Click Sample to input the name of sample.
5. In terms of sample, select material and medium.
6. The temperature is set to be 25 °C, and the equilibration time is set to 90 s (*see* **Note 1**).
7. Click the measurement cell.
8. Single-click measurement, choose measurement angle as 90; choose measure duration as automatic or manual (*see* **Note 2**).
9. Click Data Processing to choose the granularity calculation model, General Purpose option is used.
10. After all settings are completed, click OK to confirm all the options.
11. Before loading the sample, the sample cell must be rinsed thoroughly by ultrapure water (*see* **Notes 3** and **5**).
12. The DNA molecules and chitosan were diluted in Tris buffer (10 mM, pH = 8.0). The final DNA concentration is 1 ng μL^{-1} and the concentration of chitosan varies from 0.001 to 2 mM. All measurements were carried out after 5 min incubation at room temperature. Gently load 50 μL sample into measurement cell using a pipette.
13. Open the sample chamber cover according to the manufacturer's instructions. Place the sample cell into the sample chamber (make sure the symbol ▾ points toward the operator.

Table 1
Particle sizes of condensed DNA at different concentrations of chitosan

Concentration (mM)	*N*	Particle sizes (nm)	Mean value	Error SD
	1	450		
0.001	2	480	456.67	20.81
	3	440		
	1	477		
0.01	2	395	427.33	43.66
	3	410		
	1	329		
0.05	2	255	294.33	37.22
	3	299		
	1	311		
0.1	2	187	257.33	34.38
	3	223		
	1	131		
0.5	2	225	240.33	63.79
	3	144		
	1	450		
1	2	480	166.67	50.93
	3	440		

14. In the popup Manual measurement window of Zetasizer Software, click Start to start the measurement.
15. All experiments were repeated at least three times to ensure consistent results while taking the standard deviation as the error bar.

As an example the particle sizes of DNA-chitosan complexes are shown in Table 1 and Fig. 4. The particle sizes of condensed DNA gradually decrease with increasing concentration of chitosan. This behavior implies DNA condensation and compaction by chitosan, corresponding to a change in DNA conformation from being loosely extended to compact. The trend of particle sizes of condensed DNA is shown in Fig. 4. Three typical measuring results of particle size distribution of condensed DNA by chitosan are presented in Fig. 5.

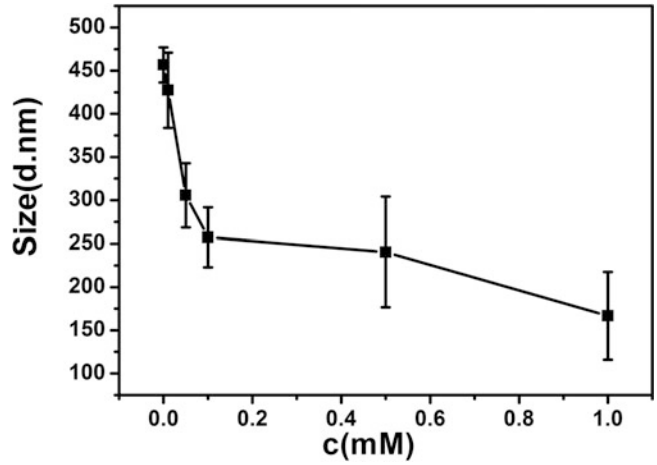


Fig. 4 The particle sizes of condensed DNA as a function of chitosan concentration. The value is the average value for three times. The error is the standard deviation

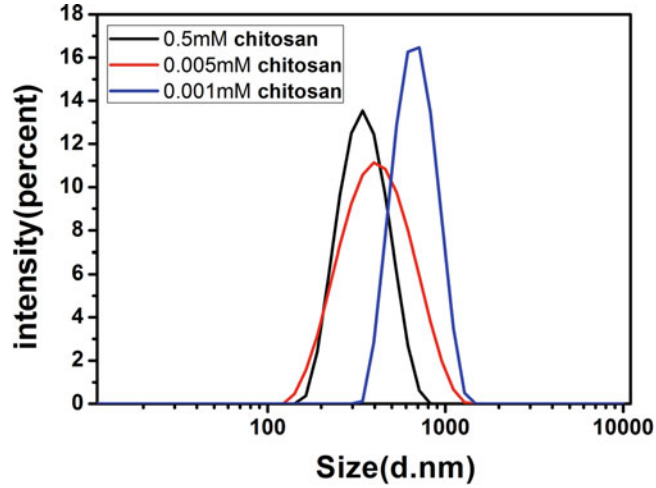


Fig. 5 The particle size distributions of condensed DNA at three concentrations of chitosan

3.2 Zeta Potential and Electrophoretic Mobility Measurement

The zeta potential and electrophoretic mobility measurements of DNA ligand use the same procedure since they are related by the Smoluchowski Eq. (7). They are presented simultaneously by the equipped measurement software. The procedure for measuring zeta potential and electrophoretic mobility is basically the same as the one for particle size described above except some measuring options. For completeness and consistency, we still present it as follows.

1. Turn on power of the DLS instrument, then wait for 30 min for preheating. When the indicator light is turned from yellow to green, it is ready for measurement.

2. Double-click desktop icon of Zetasizer Software, then click. File--New--Measurement file to choose or set up a path to save data, and input a file name of saving measurement conditions and data.
3. Open Measure menu, click Manual, and use the right key single-click measurement type, select Zeta potential to set measuring conditions.
4. Click Sample to input the name of sample.
5. Select Material and medium.
6. The temperature is set to be 25 °C, and the equilibration time is set to 90 s (*see* **Note 1**).
7. Choose the measurement cell.
8. Single-click Measurement, choose measurement angle as 90, choose measure duration as automatic or Manual (*see* **Note 2**).
9. Click Data Processing to choose the granularity calculation model, General Purpose option is used.
10. After all settings are completed, click OK to confirm all of the options.
11. Before loading the sample, the sample cell must be rinsed thoroughly by ultrapure water (*see* **Notes 4** and **5**).
12. The DNA molecules and chitosan were diluted in Tris buffer (10 mM, pH = 8.0). The final DNA concentration is 1 ng μL^{-1} and the concentration of chitosan varies from 0.001 to 2 mM. All measurements were carried out after 5 min incubation at room temperature. Gently load 1 mL sample into measurement cell using a pipette.
13. Open the sample chamber cover according to instruction of the instrument, place the sample cell into the sample chamber (make sure symbol Malvern pointing to the operator).
14. In the popup Manual measurement window of Zetasizer Software, click Start to start the measurement.
15. All experiments are repeated at least three times to ensure consistent results while taking the standard deviation as the error bar.

Typical zeta potentials of condensed DNA as a function of chitosan concentration are listed in Table 2. As the concentration of chitosan increases, the Zeta potential of the complex gradually increases and finally switches from a negative to a positive value. The trend of zeta potential of the DNA-ligand complex can be seen in Fig. 6. For instance, when the concentration of chitosan increases from 0.05 mM to 0.1 mM, the zeta potential of the complex changes from about −4.7 to 4.22 mV, corresponding to DNA charge inversion in addition to DNA compaction.

Table 2
Zeta potential of condensed DNA at different concentrations of chitosan in 10 mM Tris

Concentration (mM)	*N*	Zeta potential (mV)	Mean value	Error SD
	1	−20.70		
0.01	2	−18.80	−20.27	1.31
	3	−21.30		
	1	−4.72		
0.05	2	−5.16	−4.70	1.28
	3	−4.32		
	1	6.36		
0.1	2	2.88	4.22	1.87
	3	3.42		
	1	20.70		
0.5	2	19.33	19.43	1.23
	3	18.25		
	1	22.70		
1	2	21.32	22.57	1.19
	3	23.69		

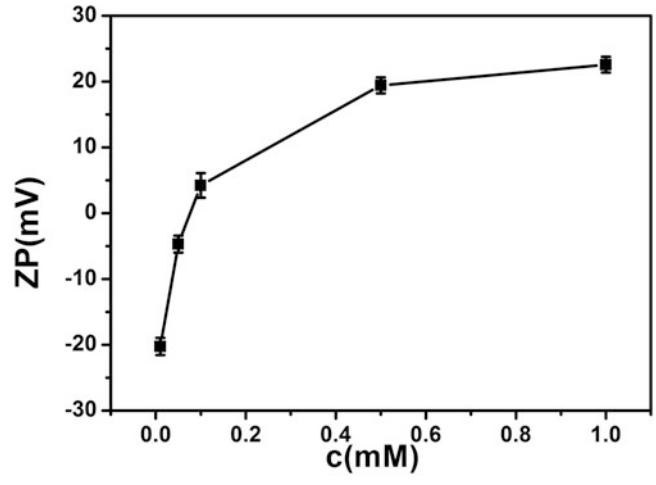


Fig. 6 The zeta potential ξ, of condensed DNA as a function of chitosan concentration. The value is the average value for three times. The error is the standard deviation

Table 3
Electrophoretic mobility of condensed DNA at different concentrations of chitosan

Concentration (mM)	*N*	Mobility ($10^{-4}cm^2V^{-1}\ s^{-1}$)	Mean value	Error SD
	1	−1.335		
0.01	2	−1.572	−1.4	0.12
	3	−1.404		
	1	−0.332		
0.05	2	−0.215	−0.4	0.28
	3	−0.755		
	1	0.113		
0.1	2	0.236	0.32	0.25
	3	0.611		
	1	1.132		
0.4	2	0.687	0.91	0.22
	3	0.932		
	1	1.315		
0.5	2	1.576	1.4	0.13
	3	1.443		
	1	1.402		
0.6	2	1.913	1.6	0.25
	3	1.664		
	1	1.923		
2	2	2.161	1.98	0.15
	3	1.881		

The measured electrophoretic mobility of DNA-chitosan complexes is shown in Table 3. The electrophoretic mobility of the complex changes from a negative value to a positive value with increasing chitosan concentration, implying DNA charge inversion. The trend in electrophoretic mobility is shown in Fig. 7.

3.3 Conclusions

In summary, DLS is a powerful and easy tool to measure particle sizes and electrokinetic properties of biomacromolecules and their complexes. In the present protocol, we outlined the sample preparation and measurement procedure for the DNA-chitosan system. We note that the particle sizes of DNA-chitosan complexes decrease with increasing concentration of chitosan. In mild

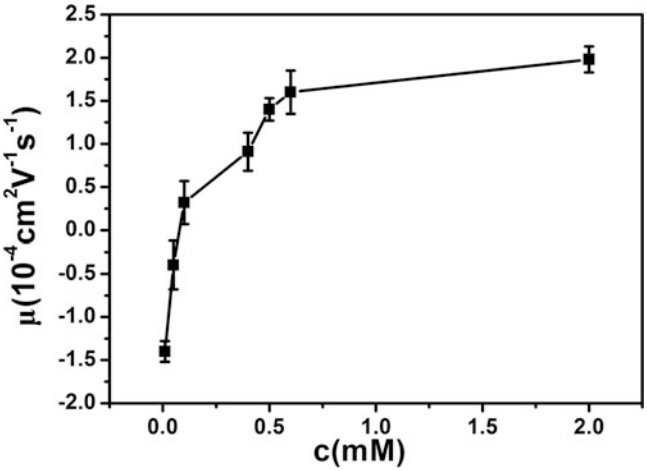


Fig. 7 The electrophoretic mobility μ, of condensed DNA as a function of chitosan concentration. The value is the average value for three times. The error is the standard deviation

solution condition, their typical zeta potential varies from negative to positive, while the corresponding electrophoretic mobility varies accordingly. These results imply the compaction of DNA by chitosan, the neutralization of DNA, and the even inversion of its surface charge.

4 Notes

1. The temperature is set at 25 °C and the equilibration time is set to 90 s. The equilibration time can be set to 120 s or more for iced samples for better equilibration at 25 °C.
2. Using the automatic option, the number of measurement cycles is selected according to the quality of the sample automatically by the system. In the Manual option, the number of measurement cycles can be input manually.
3. The sample cell needs to be rinsed by ethanol prior to use and then rinsed by ultrawater before loading the sample.
4. It is recommended that the cell is flushed with ethanol or methanol to facilitate wetting. A syringe or a wash bottle may be used. Fill one of the syringes with deionized water, or the dispersant being used for the measurement. Place the full syringe in one of the sample ports on the cell and the empty syringe into the other. Flush the contents of the full syringe, through the capillary, into the empty syringe. Repeat the flushing process 5 times, flushing the liquid back and forth between the syringes. After this the cell is ready for use.

5. Never attempt to clean the optical surface of the folded capillary cell as this will cause small surface scratches that will give inaccurate results.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 11574232, 11274245, 11304232).

References

1. Bhattacharjee S (2016) DLS and zeta potential–what they are and what they are not? J Control Release 235:337–351
2. Goldburg W (1999) Dynamic light scattering. Am J Phys 67(12):1152–1160
3. Shaw DJ, Costello B (1993) Introduction to colloid and surface chemistry. Butterworth-Heinemann, Oxford
4. Bloomfield VA (1997) DNA condensation by multivalent cations. Biopolymers 44 (3):269–282
5. Teif VB, Bohinc K (2011) Condensed DNA: condensing the concepts. Prog Biophys Mol Biol 105(3):208–222
6. Nguyen T, Rouzina I, Shklovskii B (2000) Reentrant condensation of DNA induced by multivalent counterions. J Chem Phys 112 (5):2562–2568
7. Vijayanathan V, Thomas T, Thomas T (2002) DNA nanoparticles and development of DNA delivery vehicles for gene therapy. Biochemistry 41(48):14085–14094
8. Thomas T, Tajmir-Riahi H, Thomas T (2016) Polyamine–DNA interactions and development of gene delivery vehicles. Amino Acids 48(10):2423–2431
9. Tan X, Lu X, Jia F, Liu X, Sun Y, Logan JK, Zhang K (2016) Blurring the role of oligonucleotides: spherical nucleic acids as a drug delivery vehicle. J Am Chem Soc 138 (34):10834–10837
10. Gelbart WM, Bruinsma RF, Pincus PA, Parsegian VA (2000) DNA-inspired electrostatics. Phys Today 53(9):38–44
11. Strey HH, Podgornik R, Rau DC, Parsegian VA (1998) Dna-dna interactions. Curr Opin Struct Biol 8(3):309–313
12. Besteman K, Van Eijk K, Lemay S (2007) Charge inversion accompanies DNA condensation by multivalent ions. Nat Phys 3 (9):641–644
13. Grosberg AY, Nguyen T, Shklovskii B (2002) Colloquium: the physics of charge inversion in chemical and biological systems. Rev Mod Phys 74(2):329
14. Nguyen TT, Shklovskii BI (2002) Model of inversion of DNA charge by a positive polymer: fractionalization of the polymer charge. Phys Rev Lett 89(1):018101
15. Nilsson LG, Guldbrand L, Nordenskiöld L (1991) Evaluation of the electrostatic osmotic pressure in an infinite system of hexagonally oriented DNA molecules: a Monte Carlo simulation study. Mol Phys 72(1):177–192
16. Guo Z, Wang Y, Yang A, Yang G (2016) The effect of pH on charge inversion and condensation of DNA. Soft Matter 12 (31):6669–6674
17. Wang Y, Wang R, Cao B, Guo Z, Yang G (2016) Single molecular demonstration of modulating charge inversion of DNA. Sci Rep 6:38628
18. Salas E, Martinez P, Godinez R, Del Nunez AA, Miranda C (2008) Evaluation of humoral and cellular immune response in amurine model using vesicles outer membrane of Brucella ovis in chitosan nanoparticles. Asian Chitin J 4:59–66
19. Jayakumar R, Nwe N, Tokura S, Tamura H (2007) Sulfated chitin and chitosan as novel biomaterials. Int J Biol Macromol 40 (3):175–181
20. Kai E, Ochiya T (2004) A method for oral DNA delivery with N-acetylated chitosan. Pharm Res 21(5):838–843
21. Ivanov V, Martemyanova J, Muller M, Paul W, Binder K (2008) Conformational changes of a single semiflexible macromolecule near an adsorbing surface: a Monte Carlo simulation. J Phys Chem B 113(12):3653–3668
22. MacLaughlin FC, Mumper RJ, Wang J, Tagliaferri JM, Gill I, Hinchcliffe M, Rolland AP (1998) Chitosan and depolymerized chitosan

oligomers as condensing carriers for in vivo plasmid delivery. J Control Release 56 (1):259–272

23. Fang N, Chan V, Mao H-Q, Leong KW (2001) Interactions of phospholipid bilayer with chitosan: effect of molecular weight and pH. Biomacromolecules 2(4):1161–1168

24. Richardson SW, Kolbe HJ, Duncan R (1999) Potential of low molecular mass chitosan as a DNA delivery system: biocompatibility, body distribution and ability to complex and protect DNA. Int J Pharm 178(2):231–243

Chapter 11

Microscale Thermophoresis Analysis of Chromatin Interactions

Ivan Corbeski, Velten Horn, Ramon A. van der Valk, Ulric B. le Paige, Remus T. Dame, and Hugo van Ingen

Abstract

Architectural DNA-binding proteins are key to the organization and compaction of genomic DNA inside cells. The activity of architectural proteins is often subject to further modulation and regulation through the interaction with a diverse array of other protein factors. Detailed knowledge on the binding modes involved is crucial for our understanding of how these protein-protein and protein-DNA interactions shape the functional landscape of chromatin in all kingdoms of life: bacteria, archaea, and eukarya.

Microscale thermophoresis (MST) is a biophysical technique that has seen increasing application in the study of biomolecular interactions thanks to its solution-based nature, its rapid application, modest sample demand, and the sensitivity of the thermophoresis effect to binding events. Here, we describe the use of MST in the study of chromatin interactions, with emphasis on the wide range of ways in which these experiments are set up and the diverse types of information they reveal. These aspects are illustrated with four very different systems: the sequence-dependent DNA compaction by architectural protein HMfB; the sequential binding of core histone complexes to histone chaperone APLF; the impact of the nucleosomal context on the recognition of histone modifications; and the binding of a LANA-derived peptide to nucleosome core. Special emphasis is given to the key steps in the design, execution, and analysis of MST experiments in the context of the provided examples.

Key words MST, HMf, Nucleosome, Histones, DNA, Binding affinity

1 Introduction

Biophysical characterization of functional chromatin interactions has typically relied thus far on band-shift assays (electrophoretic mobility shift assays, EMSA) for protein-DNA interaction, as well as on common biophysical techniques such as surface plasmon resonance (SPR), isothermal calorimetry (ITC), and nuclear magnetic resonance (NMR). Thorough characterization of binding modes and binding affinities is often a critical step preceding

Ivan Corbeski and Velten Horn contributed equally to this work.

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_11,

subsequent structural and functional studies. This calls for a flexible technique which is fast and can characterize interactions in solution with reasonable throughput and modest sample demands. Microscale thermophoresis (MST) fulfills these criteria and thus provides an efficient option for the analysis of biological interactions.

1.1 MST

Analogous to electrophoresis, thermophoresis is the flow or directed movement of molecules along a temperature gradient [1]. Technological advances have made it possible to use small temperature gradients (typically 2–6 °C) and detect the resulting micrometer-scale movements, allowing the application in molecular biology as microscale thermophoresis (MST) [2]. MST typically requires fluorescent labeling of the protein of interest for high sensitivity of detection. Samples are loaded onto glass capillaries and a specific spot is heated by an infra-red laser (*see* Fig. 1a). The resulting temperature gradient causes thermophoresis of the labeled molecule, typically away from the heated spot, which is observed through a decrease in fluorescence in the heated region (*see* Fig. 1b). Since thermophoresis is sensitive to molecular size, charge, and hydration shell [3], changes in these properties due to binding will cause changes in thermophoresis. For analysis of interactions, MST curves are recorded as titration series with increasing amounts of ligand and normalized with respect to their initial equilibrium fluorescence (F_{norm}) (Fig. 1b). A binding curve is extracted by plotting the F_{norm} values at the end of the laser on –period (phase IV, Fig. 1b), which captures binding induced changes both in thermophoresis and in the intrinsic temperature dependence of the fluorophore (indicated by the arrow "Thermophoresis + T-Jump" in Fig. 1b). Both effects can also be analyzed separately. Since the observed F_{norm} values are the population weighted averages of the unbound and bound molecules, standard methods can be used to fit the binding curve and extract the binding affinity. Furthermore, MST can be performed in virtually every buffer [4] and other characteristics such as thermodynamic properties, binding stoichiometry, and enzyme kinetics can be extracted with customized experimental designs [5].

1.2 MST of Chromatin Systems

The study of chromatin function is strongly connected to DNA-protein and protein-protein interactions that modulate the chromatin state. An increasing number of studies has employed MST to investigate binding events in chromatin related systems (*see* for instance [6–10]), and some of the earliest examples have been included in reviews [11–14]. A variety of labeling strategies have emerged, in particular for the study of nucleosome-protein interactions. Fluorescent nucleosomes have been constructed using Cy5-labeled DNA [14], or AlexaFluor647-labeled histone H3 [15]. Furthermore, MST has been used to derive insights in protein-binding mechanisms, e.g., demonstration of cooperative

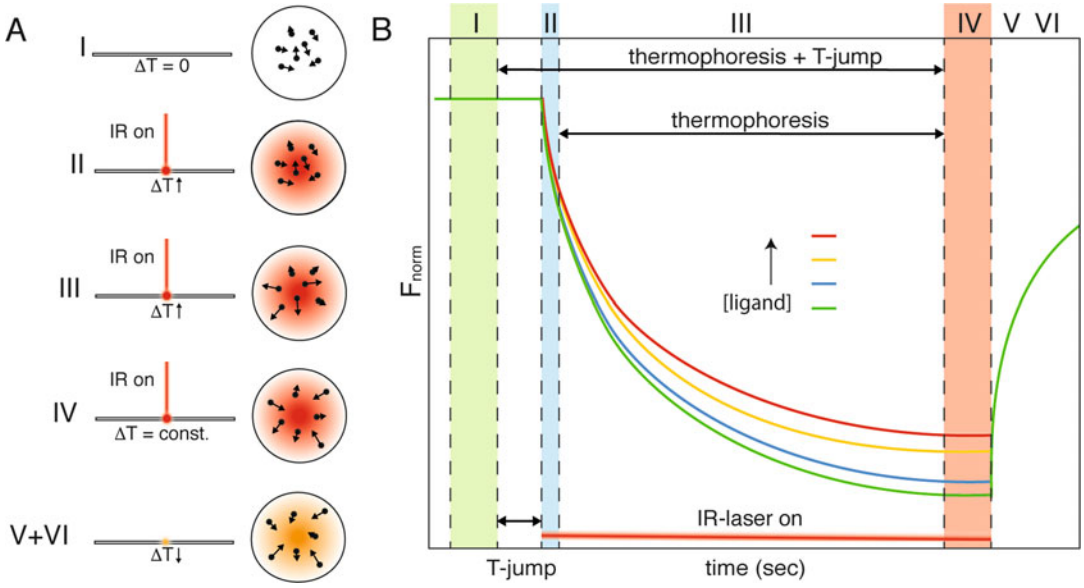


Fig. 1 Principle of MST. Schematic depiction of the MST experiment (**a**) and MST time trace (**b**). In (**a**), the capillaries (left) and particles in a cross-section thereof (right) during different stages of the experiment (I–VI) are shown. Starting from equilibrium (phase I), laser irradiation is started (phase II), causing particles to move out or into the heated volume (thermophoresis, phase III). When thermophoresis is counter balanced by mass diffusion a steady state is reached (phase IV). When the laser is switched off, the particle concentration re-equilibrates (phase VI). In (**b**), a four-sample titration series is shown, ligand concentration indicated. Each sample is characterized by a distinct thermophoresis curve. The rapid change in normalized fluorescence in phases II and V is caused by the temperature dependence in fluorescence (T-jump). Binding curves can be extracted by plotting F_{norm} values for regions "Thermophoresis + T-jump" (I vs. IV), Thermophoresis (start vs. end of III), or T-jump (I vs. II)

binding [16] or determination of binding sites from a comparison of different deletion mutants [17].

Here, we describe in detail the use of MST to study chromatin interactions. We put particular emphasis on the step-by-step optimization of experimental conditions. We present examples on the interaction of architectural proteins with both DNA and other proteins, which illustrate the additional information MST can provide on the binding mechanism, the sensitivity of the thermophoretic effect, and the merits of custom data analysis. With these, we provide first-hand reports on assay issues observed for various chromatin-related samples and strategies to detect and overcome them.

2 Materials

2.1 Fluorescent Labeling

1. Monolith NT™ Protein Labeling Kit RED/GREEN/BLUE, either NT-647-NHS, which reacts with solvent exposed primary amines, or NT-647-MALEIMIDE which reacts with

sulfhydryl groups to form dye-protein conjugates (NanoTemper Technologies) (*see* **Note 1**).

2. Variable speed benchtop microcentrifuge.
3. 1.5–2 mL microcentrifuge tubes.
4. 10 mL assay buffer (*see* **Note 2**).
5. 100% dimethylsulfoxide (DMSO).
6. Eppendorf heating block capable of reaching 95 °C (for the preparation of labeled dsDNA only).

Prepare all solutions using water (resistivity 18.2 MΩ × cm and organic content less than five parts per billion) and analytical grade reagents.

2.2 MST

1. MST instrument Monolith NT.115 equipped with "Red" channel (NanoTemper Technologies) (*see* **Note 3**).
2. Capillaries: NT.115 standard, hydrophobic, hydrophilic (*see* **Note 4**) or premium treated capillaries, or the NT.115™ Capillary Selection Set (NanoTemper Technologies).
3. PC with dedicated NT Control and MO Affinity Analysis software (version 2.1.5). Custom analysis scripts are available upon request from the corresponding author.
4. Small volume reaction tubes (e.g., as found in the labeling kit or 200 μL PCR tubes).
5. Calibrated pipettes in the range 2–1000 μL.
6. NanoDrop™ Spectrophotometer (Thermo Scientific).
7. Aluminum foil.

2.3 Stock Solutions

1. At least 100 μL of 20 μM of biomolecule to be labeled (*see* **Note 5**).
2. 100 μL of biomolecule to be titrated with a concentration 20 times the expected dissociation constant (*see* **Note 5**).
3. 10 mL of assay buffer.
4. 10–100 mg/mL bovine serum albumin (BSA).
5. 5–10% Tween-20.
6. 4 M NaCl.

3 Methods

3.1 Design of MST Experiment: Choice of Fluorescent Labeling

The MST experiment can be performed with either of the interaction partners fluorescently labeled (*see* **Note 6**). Proteins and peptides can be labeled either with the manufacturer's labeling kits or with other widely available fluorophores and coupling strategies (*see* also **Note 1**). DNA molecules are readily labeled using custom

oligo synthesis using commercially available labeled nucleotides, or using modification of the termini for coupling of dyes. For the applications described below, proteins were labeled using the manufacturer supplied labeling kits, as detailed in Subheading 3.1.1, labeling of DNA is described in Subheading 3.1.2.

3.1.1 Protein Labeling

1. Prepare a solution of pure protein at a concentration of 20 μM in a volume of 100 μL.
2. Prepare the spin column for buffer exchange into labeling buffer, using the manufacturer supplied spin columns and instructions (*see* **Note 7**). Resuspend the dried labeling buffer in 3.0 mL water. Prepare column A by resuspending the slurry. Remove excess storage solution by placing the column in a 1.5–2 mL microcentrifuge tube and centrifuging at 1500 × *g* for 1 min. Wash the column three times with 300 μL labeling buffer.
3. Exchange the protein to labeling buffer by placing the protein solution from **step 1** at the center of the resin. Be careful not to disturb the resin. Place the column in a new microcentrifuge tube and centrifuge at 1500 × *g* for 2 min.
4. Dissolve the solid fluorescent dye in 30 μL DMSO (yielding a ~470 μM solution) and mix thoroughly by vortexing (*see* **Note 8**). Prepare 100 μL 20–60 μM dye solution in Labeling Buffer (*see* **Note 9**) and take 100 μL 20 μM protein solution in Labeling Buffer. Add the dye to the protein in a 1:1 volume ratio for a final 1:1–3:1 molar ratio of dye to protein in a 200 μL volume. Incubate the reaction for 30 min at room temperature and in the dark (*see* **Note 10**). Proceed with **step 5** in the meantime.
5. Prepare the gravity flow column for purification of labeled protein and removal of unreacted dye (*see* **Note 11**). Pour off the storage solution in column B and wash the column three times with 3 mL assay buffer (*see* **Note 2**) in a 15 mL tube using the supplied adapter through gravity flow.
6. Separate the labeled protein obtained at **step 4** from unreacted dye. Apply the labeling reaction mixture to the center of column B from **step 5**. Let the sample enter the bed completely, then add 300 μL assay buffer and discard the flow-through (*see* **Note 12**). Place the column in a new 15 mL tube. Add 600 μL assay buffer and collect the eluate in ~50 μL fractions (one drop at a time) in appropriate tubes, e.g., 1.5 mL microcentrifuge tubes.
7. Verify the presence of labeled protein in elution fractions by determining their fluorescence intensity and their capillary scan signal shape (*see* Subheading 3.2) in the MST instrument. According to the gel filtration principle, larger particles will

elute prior to smaller particles (*see* **Note 13**). At 20% LED power, 10 nM labeled protein should yield fluorescence intensities of approximately 100–200 counts (*see* **Note 14** and Subheading 3.2).

8. Pool the fractions that contain labeled protein and shield them from light.
9. Determine the protein and dye concentrations and derive the labeling efficiency by measuring the absorbance at 280 (A_{280}) and 650 nm (A_{650}) in a suitable spectrophotometer, e.g., using a NanoDrop instrument (Thermo Scientific) and applying the Lambert-Beer law (*see* **Note 15**).
10. Aliquot the labeled protein as 10 μL aliquots (e.g., into 200 μL PCR tubes), flash-freeze in liquid nitrogen, and store for several weeks to months at −80 °C (*see* **Notes 5** and **16**).

3.1.2 DNA Labeling

1. Design a DNA polynucleotide sequence according to the requirements of the experiment. In this example, a polynucleotide was obtained from a commercial supplier.
2. Design a second complementary polynucleotide sequence, with the addition of a 5' Cy5-label (*see* **Note 17**).
3. In the following steps it is imperative that the samples be shielded from light (kept in the dark) as much as possible to prevent photo bleaching.
4. Combine the single-stranded DNA sequences by mixing 10 nmol of each strand and increase the volume to a total of 100 μL (*see* **Note 18**).
5. Heat the sample to 95 degrees Celsius and let the DNA strands anneal by slowly returning to room temperature. You now have a 100 μM stock of fluorescent double-strand DNA.
6. Verify the integrity of the DNA by running it on a 1% agarose gel, for very short polynucleotides on a 5% polyacrylamide gel (*see* **Note 19**).

3.2 Optimization of Experimental Conditions

Optimization of experimental conditions is paramount to obtain high-quality data and derive accurate binding parameters, which is in particular due the sensitivity of the MST experiment to protein adsorption to exposed surfaces and protein aggregation. To avoid such experimental artifacts, the correct capillary type has to be chosen, and the buffer composition needs to be optimized to ensure a homogeneous state of the sample, free from aggregation. Here we outline this procedure step-by-step (*see* Fig. 2), but we note that some parameters are interrelated and that addition of ligand may result in the need for further optimization.

1. Set the machine to the desired temperature and wait for temperature equilibration (*see* **Notes 20** and **21**).

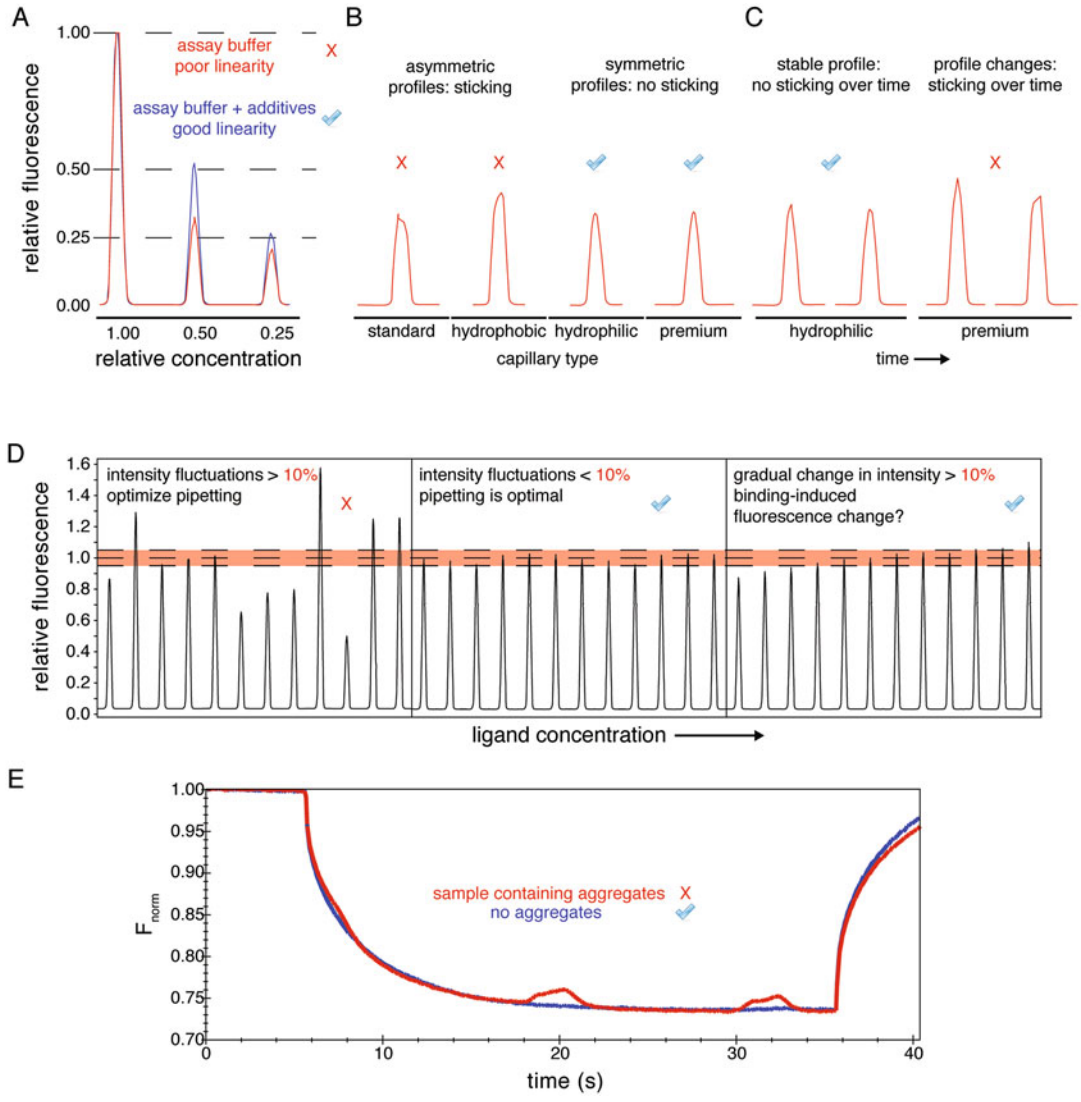


Fig. 2 Optimization of assay conditions (**a**) Capillary scans of a dilution series of the fluorescently labeled molecule. The addition of additives (0.5 mg/mL BSA and 0.05% Tween) prevents sticking to reaction tubes and leads to a consistent dilution series. (**b**) Capillary scans of different types of capillaries loaded with the same labeled molecule. Asymmetric peaks are a sign of adsorption to the capillary wall. (**c**) Time-dependent changes in the capillary scans for the same molecule. While both premium and hydrophilic coatings show no adsorption initially, only hydrophilic capillaries remain free of adsorption. (**d**) Capillary scans of a titration series of the same molecule. Addition of 0.5 mg/mL BSA and 0.05% Tween-20 together with diligent pipetting improves the reproducibility of the fluorescence intensity to within the required limits (compare the left and middle panels). Gradual fluorescence intensity changes are indicative of a binding reaction and can be used for analysis (*see* **Note 36**). (**e**) Aggregates in the sample led to irregularly shaped MST-traces (red), which were prevented by spinning the sample for 20 min at 20,000 × *g* and 4 °C to remove aggregates (*see* **Note 26**) (blue). All data are acquired on NT-467 labeled $APLF^{AD}$ and its interaction with core histone complexes (*see* Subheading 3.6.2)

2. Make a calibration curve of the dye alone at the set temperature. For this purpose prepare 20 μL 200 nM dye solution in assay buffer in a capped, small volume reaction tube to avoid evaporation (e.g., 200 μL PCR tubes) and label this tube 1. Label 8 more PCR tubes 2 through 9. Add 10 μL assay buffer in each tube 2–9. Add 10 μL from tube 1 to 2 and mix by pipetting up and down. Then add with a new pipette tip 10 μL from tube 2 to 3 and mix by pipetting up and down. Continue this series through to tube 9. Fill all samples in standard capillaries (*see* **Note 22**) and perform a capillary scan (button "Start CapScan" in NT Control) at 50% LED power to measure the fluorescence intensity at each concentration. Prepare a calibration curve with fluorescence intensity on the *y*- and dye concentration on the *x*-axis.
3. Determine the optimal concentration of the fluorescently labeled molecule. To this purpose prepare a dilution series using the stock of labeled protein as in **step 2**. Fill the samples in standard capillaries and start the capillary scan with 50% LED power. If the fluorescence is much lower than expected compared to the calibration curve or not linear over the dilution series, the protein likely sticks to the reaction tube or pipette tips (*see* also **step 4**). In that case add of 0.05% Tween-20 or another detergent (*see* Fig. 2a) and repeat the experiment. If the sample behaves well, adjust the concentration to be lower or at most in the order of the expected K_D, while still resulting in a fluorescence signal with high signal-to-noise (*see* **Notes 23** and **24**).
4. Determine the optimal type of capillary coating to ensure a homogenous sample. For this purpose load capillaries of each type (Monolith NT™ Standard, Hydrophobic, Hydrophilic (*see* **Note 4**) and Premium Treated) with working concentration of the labeled biomolecule determined in **step 3** and perform a capillary scan. Inspect the shape of the fluorescence signal, which reflects the distribution of the labeled molecule in a cross-section of the capillary. If a "U"- or "M"-shaped peak is observed instead of a smooth Gaussian-shaped fluorescence peak, the sample adsorbs to the capillary wall (Fig. 2b). Also verify that no sticking occurs over time by running a second capillary scan ~15 min after the first one (*see* Fig. 2c and **Note 25**). For further experiments select capillaries with minimal or no sticking. In case of sticking in all capillary types, proceed to **step 6** and optimize the assay buffer.
5. Use the capillary profiles to also judge the reproducibility of the fluorescence intensity from the four replicates, or from a binding experiment. Random variations larger than 10% can be caused by inaccurate pipetting or sticking to the walls of reaction tubes and pipette tips (Fig. 2d). In that case test whether addition of detergents like Tween-20, passivating agents like

BSA, or higher salt concentrations in the assay buffer improve the results (*see* **Note 26**).

6. Check the thermophoresis signal for sample aggregation. For this purpose start a thermophoresis measurement with the following settings: labeled molecule at working concentration; LED power adjusted to yield at least 200 fluorescence counts; 40% MST power; 30/5 s MST power on/off time. Aggregates, when present, will be transported in and out of the measurement volume, causing sharp increase and decrease in fluorescence over time and a bumpy appearance of the MST curve (Fig. 2e). Make sure to use stocks and buffers that are free from aggregates or particulate matter, and adjust the assay buffer composition to prevent later aggregation (*see* **Note 27**).

3.3 Preparation of Dilution Series

To determine the dissociation constant (K_D) of a molecular interaction, a dilution series of up to 16 titration points is prepared, where the concentration of the fluorescent binding partner is kept constant and the concentration of the unlabeled binding partner is varied (*see* **Notes 28** and **29**).

1. Prepare 200 µL of the fluorescently labeled biomolecule (from Subheading 3.1.1, **step 10**) in the optimized buffer with double the concentration of the final reaction to account for two-fold dilution with titrant (*see* **Note 30**). Here, and in subsequent steps, use small volume reaction tubes to avoid evaporation (e.g., 200 µL PCR tubes).
2. Prepare the titrant stock concentration to be 40-fold the expected K_D (*see* **Note 31**) in assay buffer (labeled tube 1) (*see* **Note 32**) and determine the concentration (*see* **Note 33**).
3. Prepare 15 tubes labeled 2 through 16 with 15 µL of the assay buffer. With a clean tip, transfer 15 µL from tube 1 to tube 2 and mix well by pipetting. Continue this serial dilution until tube 16 (*see* **Note 34**).
4. Transfer 10 µL from tubes 1–16 to new reaction tubes and add 10 µL of your fluorescently labeled sample stock to the tubes (*see* **Note 35**). Mix very well by pipetting up and down. After an adequate incubation time (typically 5 min, *see* **Note 36**), place the capillaries in the tubes to load the samples (*see* **Note 22**).

3.4 MST Measurement

1. Load the capillaries of the previous step in the MST machine (*see* also **Note 21**) and perform a full MST measurement using the LED power setting determined earlier, and two consecutive measurements using 20% and 40% MST power (button "Start CapScan + MST Measurement") (*see* also **Note 25**).
2. Analyze the outcome of the measurement carefully to ensure the data is of sufficient quality. Inspect results of the capillary scan to see if there is ligand-induced sticking; inspect the

reproducibility of the fluorescence intensity to see if the variation is larger than 10% (Fig. 2d, *see* **Note 37**); inspect the MST traces to see if there is ligand-induced aggregation (*see* Fig. 2e). If any of these issues are observed, carefully re-evaluate the previous steps. A new round of assay condition optimization, this time including the ligand, can help to solve these issues. When no issues are encountered, proceed to **step 3**.

3. Analyze the MST-derived binding curve in the MO Affinity Analysis software. If a transition is observed, estimate its amplitude. Estimate the noise from the scatter in the data points at the lowest concentration of ligand where no binding is expected. Minimum amplitude should be 5 units (5‰ normalized fluorescence intensity change) and minimum signal-to-noise should be 3. If either the amplitude or signal-to-noise is too low, increase the MST power to 60–80% (*see* **Notes 38** and **39**).
4. Once the final conditions have been established, perform the serial dilution (Subheadings 3.3, **steps 3** and **4**) in triplicates each with two MST powers of the same dilution series. Next, proceed to data analysis (*see* Subheading 3.5).

3.5 MST Data Analysis

The manufacturer's MO Affinity Analysis software provides multiple options to extract a binding curve from the raw MST traces (*see* **Note 40**). The resulting binding curves can be fit directly in the software for a 1:1 binding model, or the data can be exported to be used in third-party software. Here we describe the default procedure using the instrument software, together with options for custom analysis (*see* **Note 41**).

1. In a new analysis set, click and drag the three replicates in a single experiment for a combined analysis.
2. If binding-induced changes in fluorescence intensity are observed, fit the binding curve using the "Initial Fluorescence" button (*see* also **Note 37**). Otherwise, proceed to **step 3**.
3. Extract replicate-averaged binding curves using the two default settings: (1) "Thermophoresis + T-Jump"; (2) "Thermophoresis," and fit these using the thermodynamic model to extract the K_D (*see* **Note 42**). Compare the extracted values for consistency.
4. Use the "Temperature Jump" method to see if there is a binding-induced change in the fluorophore, which may hold structural information if the location of the fluorophore is known.
5. Export the data for further analysis, error estimation of fit--parameters or fitting to custom-binding models using, e.g., MatLab (scripts available upon requests), Python, or the PALMIST program of Scheuermann et al. [18].

3.6 Description of Examples

3.6.1 Different DNA Compaction Modes of HMfB

HMfB is an archaeal histone protein from *Methanothermus fervidus* with a sequence-dependent ability to either bend or wrap DNA [19, 20]. Here, we analyzed the HMfB-DNA interaction using two DNA sequences (78 bp), either with a specific HMfB-binding site ("specific sequence") or without ("aspecific sequence"). Premium capillaries were used as reduced affinity was observed for the regular and hydrophobic capillaries, likely due to the protein sticking to the capillaries. The F_{norm} values derived from the MST traces were normalized to obtain ΔF_{norm} values to facilitate comparison (Fig. 3a, b). Note that a custom dilution series was made to better sample the transitions regions.

For the aspecific sequence a single binding transition is observed that can be fit yielding a K_{D} of 1 ± 0.2 μM (Fig. 3b). For the specific DNA sequence two transitions are observed. The ΔF_{norm} decreases at low protein concentrations, indicating that the bound DNA is more mobile than free DNA. At higher protein concentrations however, the ΔF_{norm} increases indicating a less mobile, and possibly larger protein-DNA complex. This second binding mode occurs at concentrations identical to those found for aspecific DNA, indicating that this is an aspecific DNA-binding mode. Together, this suggests that at high protein concentrations, HMfB forms the same structure independent of DNA sequence, while at low protein concentrations it forms a more compact structure at specific sequences. This hypothesis is supported by data from complementary techniques (such as TPM described in Chapter 14), and summarized in Fig. 3c.

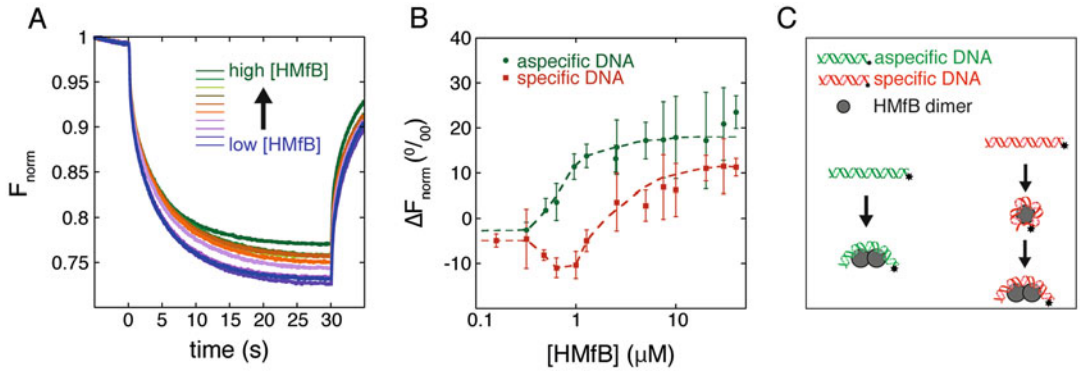


Fig. 3 MST analysis of HMfB interaction with two DNA substrates (**a**) MST-thermophoresis curves for a range of HMfB concentrations. Data acquired at 22 °C, 40 nM of Cy5-labeled DNA in 50 mM Tris–HCl pH 8.0, 75 mM KCl, premium capillaries, 20% laser power, 40% MST power. (**b**) Change in normalized fluorescence (ΔF_{norm}) for HMfB binding to a specific (red) or aspecific DNA sequence. Error bars indicate the standard deviation in a triplicate of experiments. (**c**) Model depicting the possible conformations of the HMfB-DNA complex. Without a specific binding site in the DNA substrate, HMfB binds as a multimer, forming a large structure. If a preferred DNA sequence is present, HMfB forms first a compact nucleosome-like structure. At higher HMfB concentrations similar structures as seen for the aspecific DNA sequence can be formed

3.6.2 Histone Binding by Histone Chaperone APLF

Histone chaperones are involved in the assembly and disassembly of the nucleosome for DNA replication, transcription, and repair [21]. Aprataxin and Polynucleotide kinase Like Factor (APLF) is a DNA repair protein with histone chaperone function [22]. Here, we studied the interaction of the APLF acidic domain ($APLF^{AD}$) with histone complexes using MST. $APLF^{AD}$ was labeled with the manufacturer's red dye NT-647 according to Subheading 3.1.1. Given the low labeling efficiency (~15%) and the expected high affinity, 25 nM of $APLF^{AD}$ was used with 100% LED power to arrive at an optimal fluorescence intensity of 400 counts. The assay buffer was supplemented with both 0.5 mg/mL BSA and 0.05% Tween-20 and experiments were conducted in hydrophilic-treated capillaries to prevent sticking to the reaction tubes and the capillaries (*see* Fig. 2).

The MST data show that $APLF^{AD}$ binds with high and comparable affinity to both H2A-H2B and $(H3\text{-}H4)_2$, suggesting that APLF is a generic histone chaperone (Fig. 4). Interestingly, the binding curves show in both cases two transitions, suggesting two separate binding events (*see* **Note 43**), one with affinity in the higher nanomolar range and one in the micromolar range. The data were fitted to a sequential binding site model using an in-house written MatLab script. Errors in the best-fit parameters represent the 95% confidence interval based on statistical F-tests [18]. The additional binding mode of histones to $APLF^{AD}$ may be relevant in its chaperoning mechanism, promoting the retention of multiple copies of histone complexes.

3.6.3 Binding of a Nucleosome-Mimicking Peptide to a Reader Protein

Posttranslational histone modifications are key regulators of chromatin function, mostly through their specific interaction with so-called reader proteins. We recently found that the recognition of trimethylated lysine 36 on H3 (H3K36me3) by the PSIP1-PWWP domain is driven by the nucleosomal context of this modification [23]. PSIP1-PWWP binds with very low affinity (K_D 17 mM) to a H3K36me3 peptide, but with 10,000-fold enhanced affinity to modified nucleosomes [23]. To address the importance of electrostatics, we here investigate by MST the binding of PSIP1-PWWP to a $H3K_C36me3$ peptide modified with a stretch of glutamate residues ($H3K_C36me3\text{-}E_7$) to mimic nucleosomal DNA. NanoTemper dye NT-647 was coupled to a cysteine mutant of PSIP1-PWWP (N86C). Premium capillaries and 0.05% Tween-20 were used to avoid fluorescence loss due to sticking.

The capillary scan of the titration series showed a ligand-dependent decrease in initial fluorescence, only for the trimethylated version of the peptide and not for an unmodified peptide (Fig. 5a). Additionally, a denaturing test (*see* **Note 37**) showed full fluorescence recovery, proving that the changes are binding

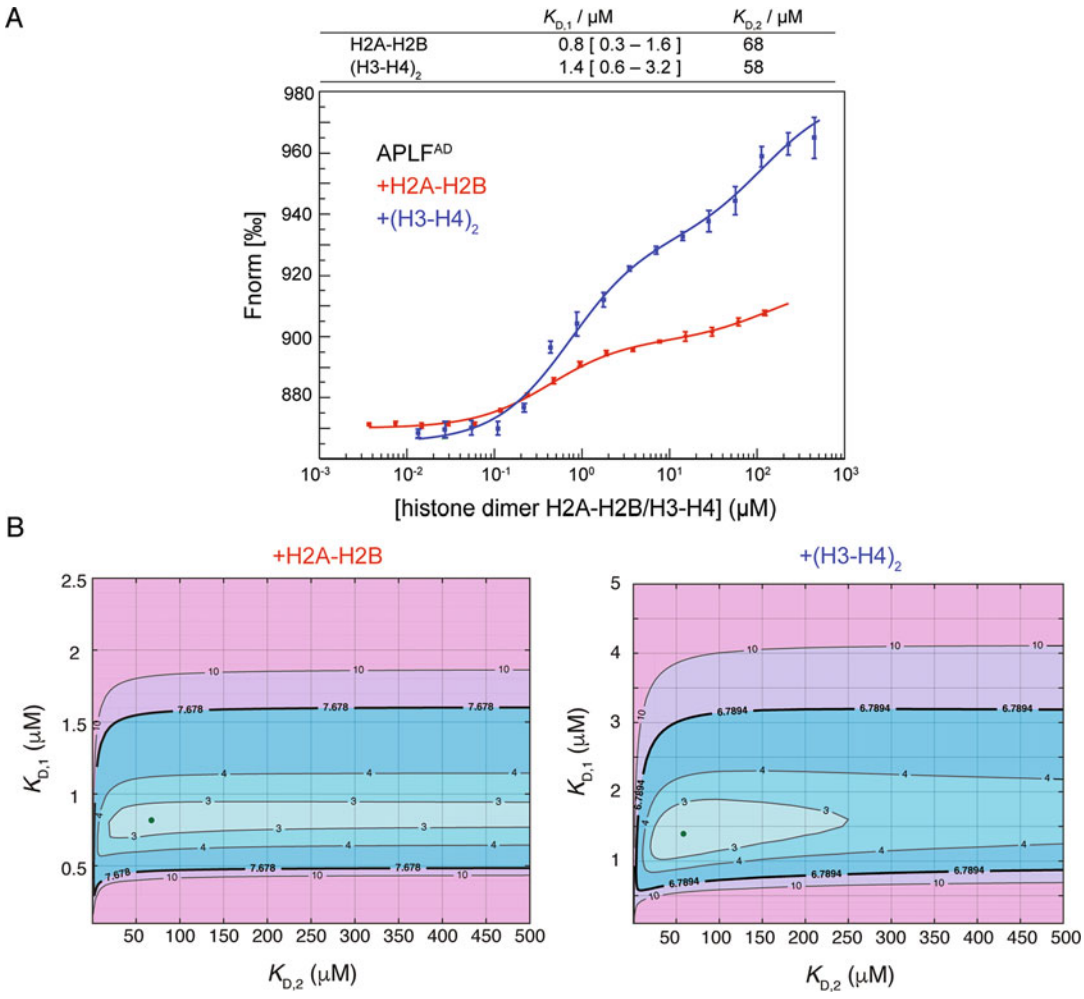


Fig. 4 APLF^{AD} binds with similar affinities to H2A-H2B and (H3-H4)$_2$. (**a**) MST-binding curves of H2A-H2B (red) or (H3-H4)$_2$ (blue) with fluorescently labeled APLF^{AD} (25 nM) in assay buffer (25 mM NaPi, pH 7.0, 300 mM NaCl, supplemented with 0.5 mg/mL BSA and 0.05% Tween-20), recorded at 25 °C, 20% MST, and 100% LED power with 30/5 s laser-on/off time. The binding curve represents the average from three measurements with standard deviation. Best-fit values for the corresponding affinities and 95% confidence limits are listed in the table. (**b**) Plots showing the reduced χ^2-surface that expresses the quality of fit in contour-mode as a function of the high-affinity ($K_{D,1}$) and the low-affinity ($K_{D,2}$) dissociation constants for H2A-H2B (left) and (H3-H4)$_2$ (right) binding to APLF^{AD}. The best-fit values are indicated by a green dot. The 95% confidence critical value for the reduced χ^2 is indicated with a thick black line. Figure reproduced from ref. 30 with permission from the authors

induced. Fits of the initial fluorescence, T-jump, and thermophoresis + T-jump all give comparable and consistent K_D values of ~1.6 mM (Fig. 5b, c, d). The tenfold increase in binding affinity compared to the native H3K36me3 peptide underscores the importance of the nucleosomal context, in which both electrostatic and geometric factors are critical.

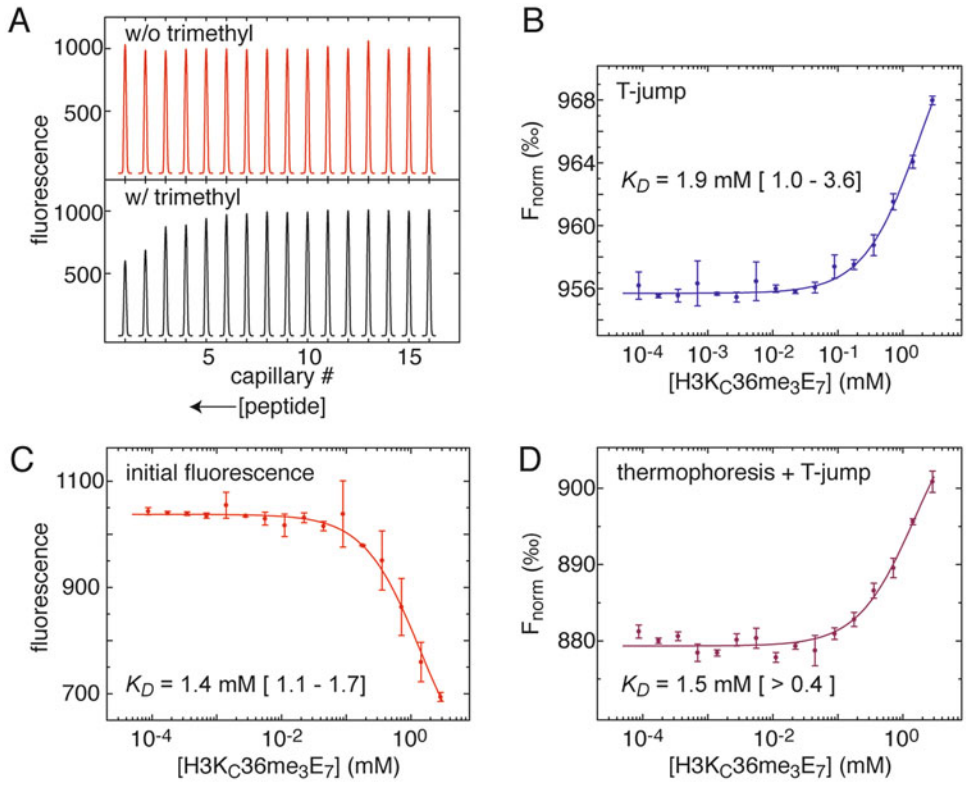


Fig. 5 MST of nucleosome-mimicking H3K$_C$36me3-peptides binding to the PSIP1-PWWP domain. (**a**) Capillary scans for H3K$_C$36-E$_7$ peptides (SAPATGGVCKEEEEEEE) with (red) and without (black) cysteine-based trimethyl lysine analogue [31]. (**b**, **c**, **d**) Analysis of the MST measurements was done by fitting the data for T-jump (**b**), initial fluorescence (**c**) and Thermophoresis + T-jump (**d**). Best-fit values for the K_D and 95% confidence limits are indicated in the plots. Data recorded on 0.35 μM NT-467 labeled N86C PWWP domain titrated with peptide in 10 mM Tris pH 7.5, 100 mM KCl, supplemented with 0.05% Tween-20, 25 °C, 20%/20% LED/MST power, with 30/5 s laser-on/off time

3.6.4 Analysis of a Nucleosome-Peptide Interaction

The latency-associated nuclear antigen (LANA) of Kaposi's sarcoma-associated herpes virus (KHSV) binds to nucleosomes to ensure the persistence of the KHSV cosmid in both daughter cells during host cell division [24]. A short peptide sequence from LANA binds strongly and specifically to the acidic patch on the H2A-H2B surface of the nucleosome [25]. Here, we analyzed the binding of a peptide containing the residues 2–22 of LANA (2.1 kDa) to nucleosomes (210 kDa), which were labeled using Nano Temper's NT467-Red-NHS dye after reconstitution [26]. Labeling efficiency was on the order of 10%, sufficient to yield 600 fluorescence counts with a 135 nM solution and 100% LED power. Labeling and purification did not interfere with nucleosome integrity as shown by native PAGE analysis (data not shown). Optimization of the conditions led to the choice of 0.1 mg/mL BSA as additive—0.01% Triton X-100 or 0.01%

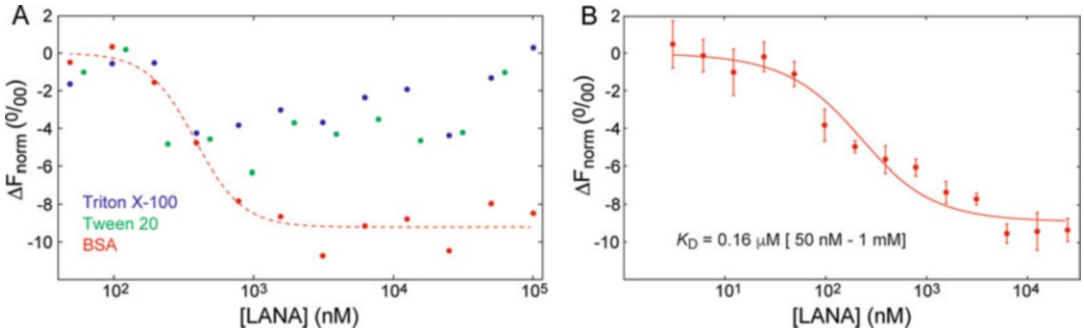


Fig. 6 MST of mononucleosomes binding to LANA. (**a**) Buffer optimization: comparison of the binding curves of 200 nM NT-467 labeled nucleosomes titrated with LANA in 10 mM Tris pH 7.5, 100 mM KCl, supplemented with either BSA, Tween-20 or Triton X-100. A fit to the BSA curve is shown to guide the eye. (**b**) MST-binding curve (average of three independent measurements, with standard deviations) of 135 nM of NT-467 labeled mononucleosomes titrated with LANA. Measurements were performed at 22 °C, 50% MST and 100% LED power with 30/5 s laser-on/off time. Best-fit values and 95% confidence intervals are shown in the figure

Tween 20 resulted in comparatively weaker thermophoretic changes upon LANA binding (Fig. 6a). Hydrophobic capillaries were selected to prevent binding-induced sticking.

The MST-binding curves show a clear transition that was fitted using an in-house written MatLab script to a single binding event with K_D of 0.16 μM, which agrees well with the previously published value of 0.24 μM under slightly different conditions [27] (Fig. 6b). These data demonstrate the possibility of reliably determining binding parameters of small molecule effectors to fluorescently labeled nucleosomes. While the thermophoretic changes may be larger using the reverse labeling setup, the current strategy allows screening of several effectors with minimum nucleosome consumption.

4 Notes

1. One is by no means restricted to use the manufacturer's fluorescent labeling kit. There are many commercially available fluorescent compounds that can be coupled to free amino or thiol groups. Make sure to use compounds compatible with the excitation and detection wavelengths of your instrument.
2. Assay buffer refers to the buffer of choice in which the interaction is investigated. The protein should be stable and well behaved in this buffer. Typical buffer conditions are 50 mM Tris–HCl pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween-20.
3. We limit the description of the method to the Monolith NT.115 instrument from NanoTemper Technologies

(Munich, Germany). Instruments with either a higher fluorescence sensitivity, a setup to excite and detect intrinsic tryptophan fluorescence, or a high-throughput automated screening setup are also available.

4. Hydrophilic-treated capillaries were used in this study, but these are no longer available. Premium capillaries are advised as replacement.
5. Higher stock concentrations are useful to have optimum flexibility to adjust exact assay buffer composition during optimization of the experimental conditions (*see* Subheading 3.2) or in follow-up experiments.
6. In case no binding is detected with one binding partner fluorescently labeled, it is useful to try and test labeling the other binding partner.
7. Buffer exchange to labeling buffer (**steps 2** and **3**) can be skipped if the protein sample is purified directly into a suitable buffer (good buffers are HEPES, PBS, Na-Ac) with $6.5 < \text{pH} < 8.5$. Buffers should not contain reducing agents dithiothreitol (DTT) and β-mercaptoethanol (BME), since these substances significantly reduce labeling efficiency. Tris (2-carboxyethyl)phosphine (TCEP) can be used as a reducing agent if required. Additionally, the buffer for NHS-labeling must be free of primary amines, e.g., ammonium ions, Tris, glycine, ethanolamine, glutathione or imidazole. Buffers with protein impurities or protein carriers like bovine serum albumin (BSA) should not be used. If a labeling buffer is used other than the one supplied, dye concentrations required for optimum labeling efficiency may need to be re-established.
8. The dye can be used for a few hours after resuspending it according to the manufacturer's manual. For longer storage of stock solutions, dye may be frozen and aliquoted in DMSO under anhydrous conditions to prevent hydrolysis.
9. For some samples labeling efficiency may be increased by using a higher fold excess of dye.
10. The reaction can be incubated in a drawer or cupboard, or wrapped in aluminum foil.
11. For optimal MST results, unreacted dye needs to be removed.
12. When using 200 μL labeling reaction, the volume must be adjusted to 500 μL after the sample has entered the bed by adding 300 μL assay buffer. If a scale up or scale down of the reaction is necessary, make sure the total volume loaded on the column is of 500 μL. Ensure the whole labeling mix has entered the column bed before completing to 500 μL
13. Use the early fractions that contain higher amounts of labeled protein. Depending on the assay buffer composition, later

fractions might contain free dye. It can also be helpful to look at the thermophoresis signal and check for "bumps," which are indicative of aggregates in the sample (*see* Fig. 2d).

14. If required, after fluorescent labeling, the protein can be concentrated using a device like an Amicon® Ultra Centrifugal Filter (Millipore).
15. The molar absorbance of the dye at 650 nm is 250,000 M^{-1} cm^{-1}. Protein absorption at 280 nm has to be adjusted according to $A_{protein} = A_{280}–0.028 \cdot A_{650}$ due to absorption of the dye. Typically, the labeling ratio is between 0.5 and 1.1 (according to the manufacturer).
16. If the stored protein is to be used for a new interaction study, repetition of previous experiments can be conducted to assess the stability and quality of the sample.
17. It is possible to generate much larger labeled DNA sequences for MST using Polymerase Chain Reaction (PCR), by following the protocol described in Chapters 12 and 14 using fluorescently labeled DNA sequences as primers.
18. You can make the DNA stock at any concentration. Making a higher concentration stock allows you to reduce the effects of incidental photo bleaching.
19. If in your control step you do not observe a single band for your double-stranded DNA, you should redo the annealing of the two strands while reducing the rate at which the two DNA strands anneal.
20. The experimental temperature in the MST instrument can be set between 22 and 45 °C or left unspecified for room temperature. The temperature is best set to a defined value and reported, together with the used MST power.
21. After opening of the instrument's front door for sample insertion, close the door and wait until the set temperature is reached as shown on the instrument's display.
22. Both tube and capillary can be tilted to ease filling. To have air on both sides, the capillary may be inverted for 1 or 2 s after loading.
23. Typically, 5–100 nM of the fluorescently labeled molecule is sufficient to obtain optimal fluorescence, with intensity above 400 and below 1500, with a minimum of 200. The LED power can be varied between 15% and 95% to achieve this. For very high-affinity interactions, use the lowest possible concentration in which you get 200 fluorescence counts with 95% LED power.
24. If the concentration of labeled molecule is on the order of the K_D, the dilution series (Subheading 3.3) is better prepared

linearly to optimally sample the binding curve, in which case the data is best plot using a linear *x*-axis resulting in a hyperbole-binding curve.

25. Capillary scans of each new titration should be performed at the start of the experiment to assess the quality of the samples and again at the end of the experiment to check for changes in the fluorescence peak shape and intensity that can occur over time.
26. Additives should be used at the lowest functional concentration possible, as they may affect the binding.
27. Aggregates can be removed by centrifuging sample stocks (20 min, 20,000 × *g*, 4 °C) and filter buffers through 0.22–0.45 μm filters. Detergents such as Tween 20 or Triton X-100 (0.01–0.1%) or changes in assay buffer conditions (different ionic strength, pH) can also help to prevent aggregation or binding-induced aggregation.
28. A control experiment can be conducted with an unrelated nonbinding molecule or a binding-deficient mutant of the binding partner.
29. Do not use less than 10 titration points per experiment.
30. Using a two-fold serial dilution, which is easily performed with minimum source of error, optimally spaced data points along the sigmoidal-binding curve are obtained (*see* also **Note 24**).
31. The "Concentration Finder" tool within the NT Control software can be used to simulate binding curves and determine the required titrant concentration ranges. In case the dissociation constant is unknown, a 3–5 fold dilution series starting from a high (~100 μM) concentration will allow monitoring of binding events within a ligand concentration range of nM to μM. If a binding transition is observed, the titrant stock concentration and dilution series can be adjusted accordingly.
32. If the sample buffer is different from the assay buffer, adjust the composition of the buffers in order to obtain same composition (e.g., DMSO, glycerol, salt, detergent, BSA etc.) by adding the components in the stock preparations to the final concentration needed.
33. As in any quantitative essay it is essential to accurately determine protein concentration, work with calibrated pipettes, and perform precise and reproducible protein dilution.
34. Change the pipet tip after each transfer or, alternatively, pre-wet the tip properly. For some pipettes it may be necessary to reset the pipetting volume after placing a new pipet tip.

35. After having obtained practical experience and confidence with the pipetting procedure, the pipetting volumes can be reduced to half to save material.
36. The incubation time is necessary to establish equilibrium which is determined by the binding kinetics of complex formation and dissociation. In sporadic cases with very high affinity and very low dissociation rates, equilibration may take hours to days [18]. Equilibration can be verified by repeating a titration series after different incubation times, e.g., 1, 5, and 10 min.
37. A systematic ligand-dependent increase or decrease in fluorescence intensity may be caused by binding if the fluorophore is close to the binding site. To determine if this is the case, spin down the reaction tubes for 10 min at 15,000 × *g*, remove 10 μL of supernatant, and add this to 10 μL of denaturing buffer (4% SDS, 40 mM DTT), then heat for 5 min at 95 °C, load the samples into capillaries and perform a capillary scan. If the fluorescence intensity is now constant within 10%, the effect was binding-induced. If the effect remains, the sample may be lost due to aggregation and a new round of optimization has to be started.
38. Reduce the laser-on time at high MST powers to 15 s to reduce effects from sample-heating.
39. Compare refs. 18, 28, 29 for MST curves with a range of signal-to-noise ratios.
40. Sign and amplitude of the thermophoresis effect are typically not analyzed since they depend in a complex manner on the changes in conformation, size, and charge.
41. The use of custom software permits more realistic determination of errors in fit-parameters and flexibility in choosing binding models.
42. Avoid the use of the Hill equation, since the reported EC50 values are protein concentration dependent and the fitted cooperativity coefficient may be larger than the number of binding sites.
43. In case multiple binding transitions are observed in an MST titration, it can be useful to confirm the result by using complementary techniques.

Acknowledgments

We thank the NanoTemper team for their support and advice. We thank prof. Dr. G.A. van der Marel and N. Meeuwenoord (Leiden University) for support in peptide synthesis and purification. This work was supported by the Netherlands Organisation for Scientific

Research (NWO) (VIDI 864.08.001, VICI 016.160.613 to RTD and VIDI 723.013.010 to HvI) and NanonextNL of the Government of the Netherland and 130 partners.

References

1. Ludwig C (1856) Diffusion zwischen ungleich erwärmten Orten gleich zusammengesetzter Lösungen. Sitzungsber Akad Wiss Wien Math-Naturwiss 20:539
2. Wienken CJ, Baaske P, Rothbauer U et al (2010) Protein-binding assays in biological liquids using microscale thermophoresis. Nat Commun 1:100. https://doi.org/10.1038/ncomms1093
3. Duhr S, Braun D (2006) Why molecules move along a temperature gradient. Proc Natl Acad Sci U S A 103:19678–19682. https://doi.org/10.1073/pnas.0603873103
4. Seidel SAI, Dijkman PM, Lea WA et al (2013) Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. Methods 59:301–315. https://doi.org/10.1016/j.ymeth.2012.12.005
5. Jerabek-Willemsen M, André T, Wanner R et al (2014) MicroScale Thermophoresis: interaction analysis and beyond. J Mol Struct 1077:101–113. https://doi.org/10.1016/j.molstruc.2014.03.009
6. Schubert T, Pusch MC, Diermeier S et al (2012) Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. Mol Cell 48:434–444. https://doi.org/10.1016/j.molcel.2012.08.021
7. Zillner K, Filarsky M, Rachow K et al (2013) Large-scale organization of ribosomal DNA chromatin is regulated by Tip5. Nucleic Acids Res 41:5251–5262. https://doi.org/10.1093/nar/gkt218
8. Su X-C, Wang Y, Yagi H et al (2014) Bound or free: interaction of the C-terminal domain of Escherichia coli single-stranded DNA-binding protein (SSB) with the tetrameric core of SSB. Biochemistry 53:1925–1934. https://doi.org/10.1021/bi5001867
9. Silva APG, Ryan DP, Galanty Y et al (2016) The N-terminal region of Chromodomain helicase DNA-binding protein 4 (CHD4) is essential for activity and contains a high mobility group (HMG) box-like-domain that can bind poly(ADP-ribose). J Biol Chem 291:924–938. https://doi.org/10.1074/jbc.M115.683227
10. Yamagata K, Kobayashi A (2017) The cysteine-rich domain of TET2 binds preferentially to mono- and dimethylated histone H3K36. J Biochem 161:327–330. https://doi.org/10.1093/jb/mvx004
11. Zillner K, Jerabek-Willemsen M, Duhr S et al (2012) Microscale thermophoresis as a sensitive method to quantify protein: nucleic acid interactions in solution. Methods Mol Biol 815:241–252. https://doi.org/10.1007/978-1-61779-424-7_18
12. Zhang W, Duhr S, Baaske P, Laue E (2014) Microscale thermophoresis for the assessment of nuclear protein-binding affinities. Methods Mol Biol 1094:269–276. https://doi.org/10.1007/978-1-62703-706-8_21
13. Flores JK, Kariawasam R, Gimenez AX et al (2015) Biophysical characterisation and quantification of nucleic acid-protein interactions: EMSA, MST and SPR. Curr Protein Pept Sci 16:727–734
14. Schubert T, Längst G (2015) Studying epigenetic interactions using MicroScale Thermophoresis (MST). AIMS Biophysics 2 (3):370–380
15. Willhoft O, McCormack EA, Aramayo RJ et al (2017) Crosstalk within a functional INO80 complex dimer regulates nucleosome sliding. elife 6:e25782. https://doi.org/10.7554/eLife.25782
16. Schrader A, Gross T, Thalhammer V, Längst G (2015) Characterization of Dnmt1 binding and DNA methylation on nucleosomes and Nucleosomal arrays. PLoS One 10:e0140076. https://doi.org/10.1371/journal.pone.0140076
17. Zocco M, Marasovic M, Pisacane P et al (2016) The Chp1 chromodomain binds the H3K9me tail and the nucleosome core to assemble heterochromatin. Cell Discovery 2:16004. https://doi.org/10.1038/celldisc.2016.4
18. Scheuermann TH, Padrick SB, Gardner KH, Brautigam CA (2016) On the acquisition and analysis of microscale thermophoresis data. Anal Biochem 496:79–93. https://doi.org/10.1016/j.ab.2015.12.013
19. Bailey KA, Marc F, Sandman K, Reeve JN (2002) Both DNA and histone fold sequences contribute to archaeal nucleosome stability. J Biol Chem 277:9293–9301. https://doi.org/10.1074/jbc.M110029200

20. Henneman B, Dame RT (2015) Archaeal histones: dynamic and versatile genome architects. AIMS Microbiol 1:72–81
21. Hammond CM, Strømme CB, Huang H, et al (2017) Histone chaperone networks shaping chromatin function. Nat Publ Group 1–19. doi: https://doi.org/10.1038/nrm.2016.159
22. Mehrotra PV, Ahel D, Ryan DP et al (2011) DNA repair factor APLF is a histone chaperone. Mol Cell 41:46–55. https://doi.org/10.1016/j.molcel.2010.12.008
23. van Nuland R, van Schaik FM, Simonis M et al (2013) Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. Epigenetics Chromatin 6:12. https://doi.org/10.1186/1756-8935-6-12
24. Ballestas ME, Chatis PA, Kaye KM (1999) Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen. Science 284:641–644
25. Barbera AJ, Chodaparambil JV, Kelley-Clarke B et al (2006) The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. Science 311:856–861. https://doi.org/10.1126/science.1120541
26. Dyer PN, Edayathumangalam RS, White CL et al (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. Meth Enzymol 375:23–44
27. Beauchemin C, Moerke NJ, Faloon P, Kaye KM (2014) Assay development and high-throughput screening for inhibitors of Kaposi's sarcoma-associated Herpesvirus N-terminal latency-associated nuclear antigen binding to nucleosomes. J Biomol Screen 19:947–958. https://doi.org/10.1177/1087057114520973
28. Allam R, Scherbaum CR, Darisipudi MN et al (2012) Histones from dying renal cells aggravate kidney injury via TLR2 and TLR4. J Am Soc Nephrol 23:1375–1388. https://doi.org/10.1681/ASN.2011111077
29. van der Berg JP, Madoori PK, Komarudin AG et al (2015) Binding of the Lactococcal drug dependent transcriptional regulator LmrR to its ligands and responsive promoter regions. PLoS One 10:e0135467. https://doi.org/10.1371/journal.pone.0135467
30. Corbeski I, Dolinar K, Wienk H, et al (2018) DNA repair factor APLF acts as a H2A-H2B histone chaperone through binding its DNA interaction surface. Nucleic Acid Res. https://doi:10.1093/nar/gky507
31. Simon MD, Chu F, Racki LR et al (2007) The site-specific installation of methyl-lysine analogs into recombinant histones. Cell 128:1003–1012. https://doi.org/10.1016/j.cell.2006.12.041

Chapter 12

Quantitative Determination of DNA Bridging Efficiency of Chromatin Proteins

Ramon A. van der Valk, Liang Qin, Geri F. Moolenaar, and Remus T. Dame

Abstract

DNA looping is important for genome organization in all domains of life. The basis of DNA loop formation is the bridging of two separate DNA double helices. Detecting DNA bridge formation generally involves the use of complex single-molecule techniques (atomic force microscopy, magnetic, or optical tweezers). Although DNA bridging can be qualitatively described, quantification of DNA bridging and bridging dynamics using these techniques is challenging. Here, we describe a novel biochemical assay capable of not only detecting DNA bridge formation, but also allowing for quantification of DNA bridging efficiency and the effects of physico-chemical conditions on DNA bridge formation.

Key words DNA bridging, DNA looping, DNA-DNA interactions, DNA-DNA cross-linking, DNA bridging proteins, Pull-down assay

1 Introduction

Three-dimensional organization of genomes affects and is affected by DNA transactions such as transcription regulation, replication, and recombination. In cells, a family of DNA-binding proteins, called chromatin proteins, is involved in the organization of the genome. These proteins wrap DNA around themselves, bend it, or bridge DNA, forming loops. DNA loops play a variety of roles in genome organization. These loops may operate locally with regulatory functions at specific single genes [1, 2], or over longer distances, enabling the organism to co-regulate genes that are in terms of genomic position far apart [3, 4]. Although studies involving DNA looping have a rich history [3, 5–12], in recent years, numerous new insights have become available through the application of new biochemical and biophysical techniques.

Classically, DNA loops (DNA bridges) were studied through the use of electron microscopy and atomic force microscopy [13–16]. These techniques permit visualization of DNA bridges. However, these static images are incapable of resolving the

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_12,

formation of DNA bridges or its modulation. The advent of biophysical techniques such as magnetic and optical tweezers has made it possible to stretch bridged DNA molecules by applying force [17, 18] and determine biophysical properties of the DNA bridges, but it is difficult to quantitate the protein(s)-DNA bridging efficiency.

Here, we describe a novel method for the quantification of protein-DNA bridging efficiency and its modulation by environmental conditions and other proteins. In this "bridging assay", we use streptavidin-coated paramagnetic beads coupled to 5′ biotin-labeled DNA (bait DNA). The DNA-coated beads are then incubated in the presence of ^{32}P radioactively labeled DNA and a DNA bridging protein (or any di- or multivalent DNA-binding ligand). The beads are pulled down by using their magnetic properties and the amount of recovered ^{32}P radiolabeled DNA (prey DNA) is detected through liquid scintillation. The recovered ^{32}P radiolabeled DNA is a direct measurement of the amount of DNA bridges formed under these conditions.

2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water, to attain a sensitivity of 18 MΩ-cm at 25 °C) and analytical grade reagents. Prepare and store all reagents at −20 °C (unless indicated otherwise). You also need access to some routine biochemical techniques [19].

2.1 Stock Solutions

The following stock solutions are required to perform this experiment

1. Phosphate-buffered saline (PBS): 12 mM $NaPO_4$ pH 7.4, 137 mM NaCl.
2. Renaturation buffer 10× (RB 10×): 200 mM Tris–HCl pH 9.5, 10 mM Spermidine, 1 mM EDTA.
3. Labeling buffer (LB): 500 mM Tris–HCl pH 9.5, 100 mM $MgCl_2$, 40% Glycerol.
4. Coupling buffer (CB): 20 mM Tris–HCl pH 8.0, 2 mM EDTA, 2 M NaCl, 2 mg/mL Acetylated BSA, 0.04% Tween20.
5. Incubation buffer 10× (IB 10×): 100 mM Tris–HCl pH 8.0, 0.2% Tween20, 10 mg/mL Acetylated BSA.
6. DNA storage buffer: 10 mM Tris–HCl pH 8.0, 50 mM KCl, 10 mM $MgCl_2$.
7. Stop buffer: 10 mM Tris–HCl pH 8.0, 1 mM EDTA, 200 mM NaCl, 0.2% SDS.

2.2 Generation of DNA Substrates Using PCR

To generate a DNA substrate for TPM, it is advised to use Polymerase Chain Reaction. This reaction requires:

1. A DNA template containing the sequence of interest (*see* **Note 1**).
2. A Forward primer.
3. A Reverse primer.
4. Dream-Taq DNA polymerase 5 U/μL.
5. 2 mM Deoxyribose Nucleotide Triphosphate (dNTP).
6. 10× Dream Taq-polymerase reaction buffer.
7. Gene Elute PCR cleanup kit (Sigma-Aldrich).
8. Eppendorf® PCR tubes.
9. Biorad T100 Thermocycler or any other available PCR machine.
10. 1% agarose gel in 1× TBE.
11. Nanodrop® (Thermo Fisher).
12. DNA ladder.

2.3 Bridging Assay Equipment

1. Magnetic Eppendorf rack.
2. Eppendorf shaker.
3. Eppendorf rack.
4. Eppendorf pipettes.
5. Streptavidin coated Dynabeads.

2.4 Quantifying DNA Bridging Through Radioactivity

1. Liquid scintillator (HIDEX 300SL).
2. Counting vials.
3. 37 °C heat block.
4. 80 °C water bath.
5. Eppendorf® PCR tubes.
6. ATP, gamma ^{32}P.
7. Tabletop Eppendorf centrifuge.
8. T4 Polynucleotide Kinase 10 U/μL.
9. Mini G50 columns (GE Healthcare).

3 Methods

3.1 Generation of DNA Substrates Using PCR

These reagents are combined in an Eppendorf® PCR tube according to the scheme below. These reactions must be done for both the standard and biotinylated primers.

Reagent	Final quantity
DNA template	1 ng
Forward primer	10 pmol
Reverse primer	10 pmol
2 mM dNTP	5 μL
10× Taq polymerase buffer	5 μL
5 U/μL dream Taq polymerase	0.2 μL
H_2O	Add to total volume of 50 μL

1. Keep this reaction mix on ice as much as possible and initiate the PCR using the following protocol (*see* **Note 2**).

	Cycles	Temperature	Duration
Initial denaturation	1	95 °C	5 min
Denaturation	35	95 °C	30 s
Annealing		62 °C	30 s
Elongation		72 °C	4 min
Final elongation	1	72 °C	10 min
	1	15 °C	∞

2. Purify the PCR product using the GeneElute PCR cleanup kit.
3. Load 2 μL of the purified PCR product on a 1% agarose gel in TBE buffer alongside a DNA molecular weight marker for verification that a product of the expected length is formed. An example of a successful PCR and purification of the obtained PCR product is shown in Fig. 1.
4. Finally, the concentration of purified PCR-generated DNA needs to be determined accurately. Determine the concentration of the purified DNA by measuring UV absorbance at 260 and 280 nm (*see* **Note 3**). If no other method is available the concentration of DNA can also be approximated using a DNA dilution series run on an agarose gel compared to a reference marker. Store the DNA solution at −20 °C.

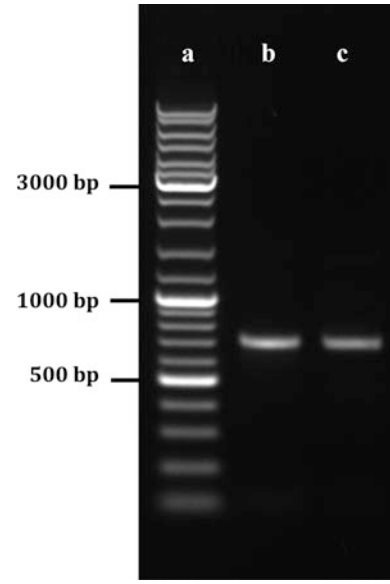


Fig. 1 Visualization of PCR product size by agarose gel electrophoresis. (**a**) 2 μL of the GeneRuler DNA molecular weight marker. (**b**) 2 μL of the purified PCR-generated unlabeled DNA (685 bp, ready for ^{32}P labeling). (**c**) 2 μL of the purified PCR-generated biotin-labeled DNA (685 bp)

3.2 Radio-labeling DNA

1. Add 1.5 μL of RB 10× to 2 pmol of the purified DNA and fill to a final volume of 15 μL using H_2O.
2. Prepare the Kinase mix according to the following scheme:

Kinase mix component	Added volume per DNA labeling (μL)
LB 10×	2.5
50 mM DTT	2.5
ATP, gamma ^{32}P	2
10 U/μL polynucleotide kinase	1
H_2O	2

3. Incubate the DNA mix at 80 °C for 2 min and immediately put the sample on ice (*see* **Note 4**).
4. Add 10 μL of the Kinase mix to the DNA sample and incubate at 37 °C for 30 min.
5. Stop the reaction by adding 1 μL of 0.5 M EDTA. Incubate the sample at 75 °C for 15 min to deactivate the kinase.
6. Quickly spin the sample down using a tabletop centrifuge.
7. Prepare the mini G50 column by pre-incubating it in DNA storage buffer as described by the column manual.

8. Purify the labeled DNA using the G50 column.
9. Assess the volume of the purified DNA and fill it to 100 μL using DNA storage buffer. The DNA should now have a final concentration of approximately 20 pmol/μL.
10. Fill a counting vial with 7 mL of H_2O.
11. Prepare 2 μL of the labeled DNA for liquid scintillation, by transferring it to a PCR tube and submerging it in the counting vial.
12. Determine the amount of counts per minute per vial.
13. Quantify the counts per μL of labeled DNA

3.3 Bridging Assay

The DNA-bridging assay relies on the immobilization of bait DNA on magnetic microparticles and the capture and detection of ^{32}P labeled prey DNA if DNA-DNA bridge formation occurs (*see* Fig. 2 for a schematic depiction of the assay).

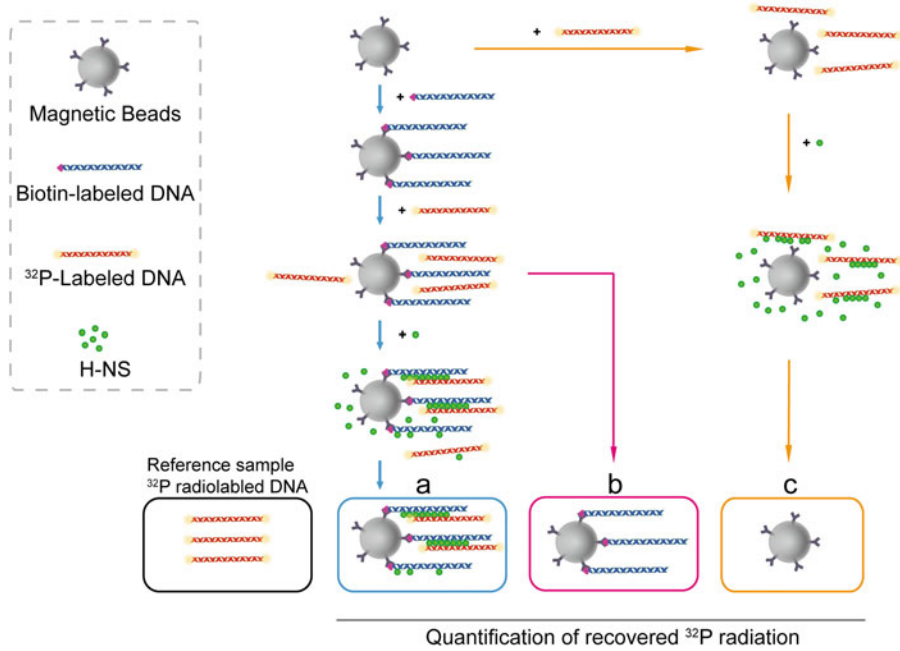


Fig. 2 Schematic depiction of the DNA bridging assay with the H-NS DNA bridging protein. A standard DNA bridging assay is shown by the blue arrows. Here streptavidin-coated paramagnetic beads are coupled to the bait DNA, 5′ labeled with biotin. The beads bound with DNA are then incubated in the presence of ^{32}P radiolabeled DNA strand and H-NS. Next, using a magnetic rack, the beads are pulled down and the amount of recovered ^{32}P radiolabeled DNA (prey DNA) is quantified based on a reference ^{32}P radiolabeled DNA sample. (**b**) The pink arrows indicate a standard negative control for the DNA-bridging assay, in which no H-NS is added. This control checks the stability of both prey and bait DNA. No ^{32}P radiolabeled DNA should be recovered for this sample (*see* **Note 5** if this is the case). (**c**) The orange arrows indicate a standard negative control for the assay in which the DNA bridging assay is performed in the absence of the bait DNA to test the stability of the protein. No ^{32}P radiolabeled DNA should be recovered for this assay (*see* **Note 6** if this is the case)

1. Wash 3 μL of streptavidin-coated paramagnetic beads (henceforth referred to as "beads") per condition you wish to test with 50 μL of PBS on the magnetic rack (*see* **Note 7**).
2. Wash the beads with 50 μL of CB twice.
3. Resuspend the beads in 3 μL of CB.
4. Dilute 100 pmol of biotinylated DNA in a total volume of 3 μL using DNA storage buffer (one per sample).
5. Add the biotinylated DNA solution to the washed and resuspended beads.
6. Gently vortex the sample to ensure that the beads are resuspended.
7. Incubate the samples at 25 °C for 20 mins in the Eppendorf shaker at 1000 rpm.
8. Wash the beads with 16 μL of 1× IB twice.
9. Resuspend the beads in 16 μL of 1× IB.
10. Add 2 μL of the protein of interest.
11. Add 2 μL of radiolabeled DNA (with a minimum of 5000 counts per minute).
12. Gently vortex the sample to ensure that the beads are resuspended.
13. Incubate the samples at 25 °C for 20 min in the Eppendorf shaker at 1000 rpm.
14. Gently wash the beads with bridged protein-DNA complexes with 20 μL of IB.
15. Resuspend the beads in Stop buffer.
16. Transfer the sample to the liquid Cherenkov-scintillation counter.

4 Results

4.1 DNA-Bridging Efficiency as a Function of Protein Concentration.

The protein concentration used in the assay determines the amount of DNA bridging observed. It is therefore essential to test a range of protein concentrations whenever a previously uncharacterized DNA-bridging protein is investigated using the bridging assay. Here, we show an example (Fig. 3) from our recent study investigating the DNA-bridging efficiency of the Histone-like Nucleoid Structuring protein (H-NS) [20]. Using this assay, it was demonstrated that the DNA-bridging efficiency of H-NS is highly dependent on protein concentration.

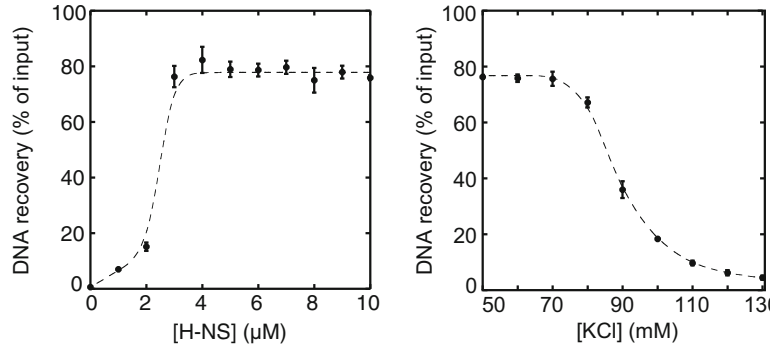


Fig. 3 DNA bridging as a function of H-NS concentration [20]. The experiments were performed in the presence of 10 mM Tris–HCl pH 8.0, 50 mM KCl, 10 mM $MgCl_2$, 5% (w/v) glycerol. Error bars indicate the standard deviation of a triplicate of experiments

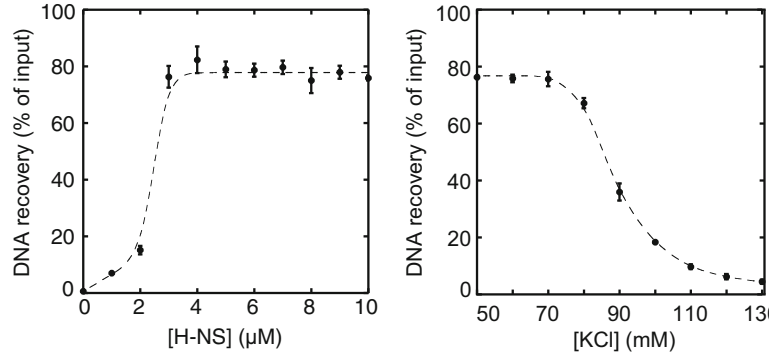


Fig. 4 Modulation of H-NS-DNA bridging by $[K^+]$ [20]. The experiments were performed with 3,3 μM H-NS in 10 mM Tris–HCl pH 8.0, 10 mM $MgCl_2$, 5% w/v glycerol. Error bars indicate the standard deviation of a triplicate of experiments

4.2 DNA Bridging Efficiency of H-NS as a Function of Physico-Chemical Conditions

The DNA-bridging assay allows for facile testing of the effect of altered physico-chemical conditions. It has been shown previously that H-NS-mediated DNA bridging is strongly modulated by environmental factors such as osmotic stress [17, 20]. The DNA-bridging assay revealed that increasing the amount of KCl in the buffer indeed effectively abolishes DNA bridging by H-NS (Fig. 4) [20].

This strong dependence of H-NS activity on environmental factors underlines the necessity to test different buffer conditions when testing new proteins. It is, however, important to verify that the DNA-binding activity of the protein is still intact under conditions that no DNA bridging is observed (*see* **Note 8**).

5 Notes

1. The length of the DNA substrate used can affect the efficiency of the assay. A 685 base pair DNA substrate was used in the experiments described here.
2. This PCR protocol was optimized for a 685 bp DNA substrate using a Biorad T100 Thermocycler. Some optimization may be required for different thermocyclers or substrates of different lengths.
3. When determining the concentration of the DNA it may be advantageous to use a Nanodrop® as this technique requires a very small volume.
4. Snap-chilling your DNA before labeling the DNA remains single-stranded and increases the efficiency of the Kinase.
5. Recovery of ^{32}P labeled DNA in the absence of DNA-bridging proteins may be an indication of DNA aggregation.

   In these cases, it is best to:

   (a) Check the integrity of DNA on a 1% agarose gel.

   (b) Re-evaluate the experimental buffer as the absence of salt may cause larger DNA substrates to aggregate. Similar effects may occur at extreme pH's.

   (c) Use new beads as the streptavidin coating may decay over time, leading to inconclusive experiments.
6. If ^{32}P Labeled DNA is recovered in the absence of biotin labeled DNA, and not in the situation described in **Note 5**, it is likely caused by precipitation or aggregation of the protein. In these cases it is best to:

   (a) Optimize the experimental buffer. Some proteins precipitate in suboptimal conditions. The conditions can vary greatly from protein to protein so it is best to test a wide array of conditions (ion concentrations, pH, ion composition, etc.) and detergents until a suitable buffer has been found.

   (b) Use new beads, *see* **Note 5c**.
7. When washing the beads on the magnetic rack pay attention to the following:

   (a) Keep the Eppendorf tubes in the magnetic rack and incubate for at least 1 min to ensure that the beads are pelleted.

   (b) When removing the supernatant make sure to pipette slowly and not to disturb the pelleted beads with the pipette tip.

   (c) Use a 0.5–10 Eppendorf micropipette to ensure that all liquid is removed from the sample.

   (d) Gently pipette the new liquid onto the pelleted beads.

8. DNA binding of proteins is best confirmed with additional solution-based experiments such as Microscale thermophoresis (*see* Chapter 11) or Tethered particle motion (*see* Chapter 14).

Acknowledgments

This work was supported by grants from the Netherlands Organization for Scientific Research [VICI 016.160.613] and the Human Frontier Science Program (HFSP) [RGP0014/2014].

Reference

1. Besse M, von Wilcken-Bergmann B, Müller-Hill B (1986) Synthetic lac operator mediates repression through lac repressor when introduced upstream and downstream from lac promoter. EMBO J 5(6):1377–1381
2. Becker NA, Peters JP, Lionberger TA, Maher IIILJ (2013) Mechanism of promoter repression by lac repressor–DNA loops. Nucleic Acids Res 41(1):156–166. https://doi.org/10.1093/nar/gks1011
3. Schleif R (1992) DNA Looping. Annu Rev Biochem 61(1):199–223. https://doi.org/10.1146/annurev.bi.61.070192.001215
4. Rao Suhas SP, Huntley Miriam H, Durand Neva C, Stamenova Elena K, Bochkov Ivan D, Robinson James T, Sanborn Adrian L, Machol I, Omer Arina D, Lander Eric S, Aiden Erez L (2014) A 3D map of the human genome at Kilobase resolution reveals principles of chromatin looping. Cell 159 (7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021
5. Matthews KS (1992) DNA looping. Microbiol Rev 56(1):123–136
6. Rippe K, von Hippel PH, Langowski J (1995) Action at a distance: DNA-looping and initiation of transcription. Trends Biochem Sci 20 (12):500–506
7. Schleif R (2000) Regulation of the L-arabinose operon of Escherichia coli. Trends in genetics : TIG 16(12):559–565
8. Chambeyron S, Bickmore WA (2004) Does looping and clustering in the nucleus regulate gene expression? Curr Opin Cell Biol 16 (3):256–262. https://doi.org/10.1016/j.ceb.2004.03.004
9. Saiz L, Vilar JM (2006) DNA looping: the consequences and its control. Curr Opin Struct Biol 16(3):344–350. https://doi.org/10.1016/j.sbi.2006.05.008
10. Dorman CJ, Kane KA (2009) DNA bridging and antibridging: a role for bacterial nucleoid-associated proteins in regulating the expression of laterally acquired genes. FEMS Microbiol Rev 33(3):587–592. https://doi.org/10.1111/j.1574-6976.2008.00155.x
11. van der Valk RA, Vreede J, Cremazy F, Dame RT (2014) Genomic looping: a key principle of chromatin organization. J Mol Microbiol Biotechnol 24(5–6):344–359. https://doi.org/10.1159/000368851
12. Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, Zerbino D, Schoenfelder S, Javierre B-M, Osborne C, Fraser P, Spivakov M (2016) CHiCAGO: robust detection of DNA looping interactions in capture hi-C data. Genome Biol 17:127. https://doi.org/10.1186/s13059-016-0992-2
13. Lyubchenko YL, Shlyakhtenko LS, Aki T, Adhya S (1997) Atomic force microscopic demonstration of DNA looping by GalR and HU. Nucleic Acids Res 25(4):873–876. https://doi.org/10.1093/nar/25.4.873
14. Dame RT, Wyman C, Goosen N (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Res 28(18):3504–3510. https://doi.org/10.1093/nar/28.18.3504
15. Dame RT, Luijsterburg MS, Krin E, Bertin PN, Wagner R, Wuite GJ (2005) DNA bridging: a property shared among H-NS-like proteins. J Bacteriol 187(5):1845–1848. https://doi.org/10.1128/JB.187.5.1845-1848.2005
16. Murugesapillai D, McCauley MJ, Huo R, Nelson Holte MH, Stepanyants A, Maher LJ,

Israeloff NE, Williams MC (2014) DNA bridging and looping by HMO1 provides a mechanism for stabilizing nucleosome-free chromatin. Nucleic Acids Res 42 (14):8996–9004. https://doi.org/10.1093/nar/gku635

17. Liu Y, Chen H, Kenney LJ, Yan J (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. Genes Dev 24(4):339–344. https://doi.org/10.1101/gad.1883510
18. Dame RT, Noom MC, Wuite GJ (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. Nature 444(7117):387–390. https://doi.org/10.1038/nature05283
19. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, vol 2. Cold spring harbor laboratory press, Cold spring harbor
20. van der Valk RA, Vreede J, Qin L, Moolenaar GF, Hofmann A, Goosen N, Dame RT (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-bridging activity. elife 6: e27369. https://doi.org/10.7554/eLife.27369

Chapter 13

Approaches for Determining DNA Persistence Length Using Atomic Force Microscopy

Justin P. Peters and L. James Maher III

Abstract

Atomic force microscopy (AFM) is widely used to image and study biological molecules. As an example, we have utilized AFM to investigate how the mechanical properties of DNA polymers depend on electrostatics and the strength of DNA base stacking by studying double-stranded DNA molecules incorporating several different neutral and charged base modifications. Here, we describe ten complementary approaches for determining DNA persistence length by AFM imaging. The combination of different approaches provides increased confidence and statistical reliability over existing methods utilizing only a single approach.

Key words Atomic force microscopy, DNA mechanics, Persistence length, Thymidine analogs, Wormlike chain model

1 Introduction

DNA molecules and other biopolymers are routinely visualized using atomic force microscopy (AFM). The strength of AFM lies in its ability to rapidly scan and visualize multiple molecules on a surface. A common objective of these types of experiments is to determine mechanical properties such as persistence length.

The most widespread experimental method for imaging individual DNA molecules involves depositing the sample onto mica. The high charge density of DNA allows for its immobilization onto a planar two-dimensional surface by virtue of ionic interactions between DNA phosphates and surface charges. Adsorption of DNA onto mica surfaces is typically promoted using either divalent cations or polyamines; these agents bridge the negative charges of the phosphate backbone and the negatively charged mica surface. Under conditions of weak adsorption, DNA molecules are believed to undergo thermal conformational fluctuations prior to immobilization and therefore achieve chain statistics that represent equilibrium behavior in two dimensions [1, 2]. The conformational state of the deposited molecules is generally analyzed by fitting

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_13,

experimental datasets with formulae accounting for statistical properties of equilibrated molecules, typically derived from theoretical predictions of the wormlike chain (WLC) model. The WLC model has proven to be a good model of DNA mechanics and to accurately describe conformations of DNA molecules in two or three dimensions [3–6].

Validation and support of a molecule deposition and image analysis approach comes from both an agreement with literature data from independent techniques as well as conformity in the calculated values for persistence length based on theoretical expectations. Here, we discuss ten approaches for determining DNA persistence length. We show how these ten approaches can be applied to AFM images of a series of DNA molecules incorporating thymidine analogs. The end results of such analysis have been previously published (Table 3 in [7]), but the step-by-step details, associated figures, and tables appear for the first time below.

2 Materials

2.1 Equipment

1. Thermocycler.
2. Microcentrifuge.
3. Spectrophotometer capable of measuring UV absorbance at 260 nm.
4. Atomic Force Microscope (with instrument placed on vibration isolation table).

2.2 PCR Reagents

1. Plasmid DNA containing desired sequence.
2. Forward and reverse primers.
3. dTTP analogs desired in modified DNA.
4. dNTP mixture lacking dTTP: 2.5 mM dATP, 2.5 mM dCTP, and 2.5 mM dGTP.
5. 1 mg/mL bovine serum albumin (BSA).
6. 50 mM $MgCl_2$.
7. Taq DNA polymerase (Invitrogen).
8. 10× Taq DNA polymerase buffer (Invitrogen).
9. 5 M betaine.
10. 5× PrimeSTAR GC buffer (Takara).
11. PrimeSTAR HS DNA polymerase (Takara).
12. 5× GC-rich solution (Roche).
13. Pwo PCR buffer (Roche).
14. Pwo SuperYield DNA Polymerase (Roche).
15. QIAquick PCR Purification Kit (Qiagen).

2.3 AFM Reagents

1. Autoclaved ultrapure (18 MΩ-cm) water.
2. 10× AFM buffer: 50 mM Tris–HCl, pH 7.5 supplemented with 100 mM NaCl and 50 mM $MgCl_2$.
3. Pure DNA samples.
4. AFM-grade bare mica.
5. Silicon cantilevers designed for tapping mode.

3 Methods

3.1 Sample Preparation

Figure 1 shows chemical structures for seven dTTP analogs (*see* **Note 1**).

1. Perform polymerase chain reaction (PCR) to synthesize DNA fragments 753 base pairs (bp) in length from a pUC19-based plasmid (pJ1506) containing a central 200 bp intrinsically straight sequence [8, 9], as shown in Fig. 2 (*see* **Note 2**).
2. For analog 1, include in the PCR reactions (100 μL) 20 ng plasmid template (*see* **Note 3**), 0.4 mM forward and reverse primers, 100 μg/mL BSA, Taq DNA polymerase buffer, 2 mM $MgCl_2$, 0.2 mM each dNTP (with dTTP completely replaced by the analog triphosphate) and 5 U Taq DNA polymerase. Cycle using conditions: 98 °C (3 min), 30 cycles of [94 °C (30 s), 60 °C (30 s) and 72 °C (45 s)], followed by 72 °C (5 min) and a 4 °C hold.
3. For analogs 2, 3, 4, and 5, include in the PCR reactions (100 μL) 20 ng template, 0.4 mM forward and reverse primers,

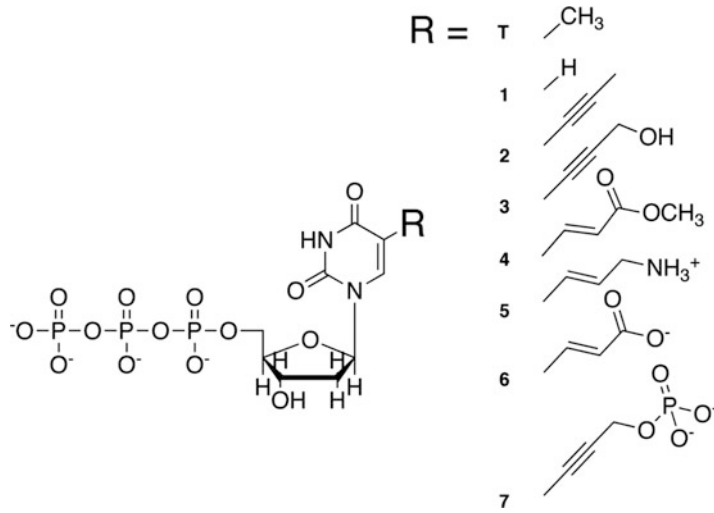


Fig. 1 Chemical structure of 2′-deoxythymidine triphosphate (dTTP) variants. The site of thymine modification (C5 position) is shown with a large R. Structures of the functional groups for the seven variants are shown at right

Sequence amplified from plasmid pJ1506

```
  1  CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
     GCCACTACTG CCACTTTTGG AGACTGTGTA CGTCGAGGGC CTCTGCCAGT GTCGAACAGA CATTCGCCTA CGGCCCTCGT

 81  GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA
     CTGTTCGGGC AGTCCCGCGC AGTCGCCCAC AACCGCCCAC AGCCCCGACC GAATTGATAC GCCGTAGTCT CGTCTAACAT

161  CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT
     GACTCTCACG TGGTATACGC CACACTTTAT GGCGTGTCTA CGCATTCCTC TTTTATGGCG TAGTCCGCGG TAAGCGGTAA

241  CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG
     GTCCGACGCG TTGACAACCC TTCCCGCTAG CCACGCCCGG AGAAGCGATA ATGCGGTCGA CCGCTTTCCC CCTACACGAC

321  CAAGGCGATT AAGTTGGGTA ACGCCAGGGT TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGAATT CGAGCTCGGT
     GTTCCGCTAA TTCAACCCAT TGCGGTCCCA AAAGGGTCAG TGCTGCAACA TTTTGCTGCC GGTCACTTAA GCTCGAGCCA

401  ACCCGGGGAT CCTCTCGCGG CGCCCGCGAC TCGAGCCTAG CCTATGACAT GACACGTTAC GTTAGTCGAG TCGATCAGAT
     TGGGCCCCTA GGAGAGCGCC GCGGGCGCTG AGCTCGGATC GGATACTGTA CTGTGCAATG CAATCAGCTC AGCTAGTCTA

481  CAGACGCTAC GCTAGCTGAG CTGACTGTAC TGTATGCAAT GCAACCTCAC CTCAGGACAG GACACGTGAC GTGATGCTAT
     GTCTGCGATG CGATCGACTC GACTGACATG ACATACGTTA CGTTGGAGTG GAGTCCTGTC CTGTGCACTG CACTACGATA

561  GCTACCAGAC CAGCTGCACT GCAGACTGGA CTGACGCTAC GCTATCGCAT CGCAGATGAG ATGAAGCCGG GCGCCGCCAT
     CGATGGTCTG GTCGACGTGA CGTCTGACCT GACTGCGATG CGATAGCGTA GCGTCTACTC TACTTCGGCC CGCGGCGGTA

641  GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG CCGGAAGCAT AAAGTGTAAA
     CCAGTATCGA CAAAGGACAC ACTTTAACAA TAGGCGAGTG TTAAGGTGTG TTGTATGCTC GGCCTTCGTA TTTCACATTT

721  GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACA
     CGGACCCCAC GGATTACTCA CTCGATTGAG TGT
```

Fig. 2 Sequence of the double-stranded DNA (dsDNA) fragment amplified from plasmid pJ1506. Forward (LJM-4762, bold magenta) and reverse (LJM-3223, bold cyan) primers are used to PCR amplify 753-bp fragments for AFM which contain central 5-bp direct repeats (bold) to eliminate long-range sequence-directed curvature

PrimeSTAR GC buffer, 0.2 mM each dNTP (dTTP completely replaced with analog), 2 M betaine and 5 U PrimeSTAR HS DNA polymerase. Cycle using conditions: 98 °C (3 min), 30 cycles of [98 °C (15 s), 60 °C (5 s) and 72 °C (45 s)], followed by 72 °C (5 min) and a 4 °C hold.

4. For analogs 6 and 7, include in the PCR reactions (100 μL) 20 ng template, 0.4 mM forward and reverse primers, Pwo PCR buffer, GC-rich solution, 0.2 mM each dNTP (dTTP completely replaced with analog), 2 M betaine and 5 U Pwo SuperYield DNA Polymerase. Cycle using conditions: 98 °C (3 min), 30 cycles of [98 °C (1 min), 60 °C (2 min), and 72 °C (8 min)], followed by 72 °C (5 min) and a 4 °C hold.
5. Following PCR, purify the reactions using a PCR Purification Kit and quantify using a spectrophotometer (*see* **Note 4**).

3.2 AFM Imaging

1. Dilute each DNA sample to a concentration of ~2 μg/mL in AFM buffer and briefly mix to ensure even dispersal in solution (*see* **Note 5**).
2. Freshly cleave mica discs to serve as a support for sample adsorption. Mica is composed of layers of minerals separated

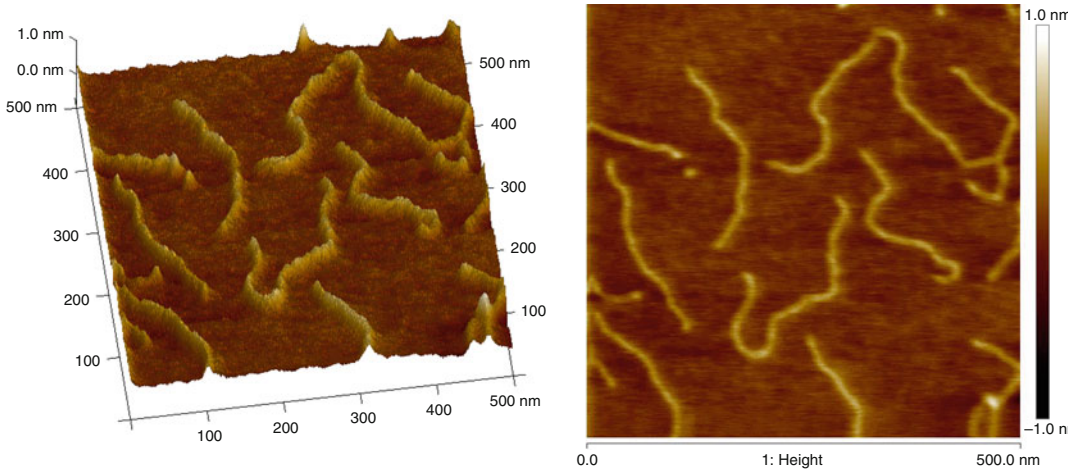


Fig. 3 Representative 500 × 500 nm (512 × 512 pixel) AFM image of 753 bp dsDNA (containing variant 6) deposited on mica in 5 mM Tris–HCl, pH 7.5 supplemented with 10 mM NaCl and 5 mM $MgCl_2$. The three-dimensional representation (left), which visualizes several 2D equilibrium conformations of these DNA fragments when subjected to thermal fluctuations, is represented in two dimensions (right) with a color gradient that reflects the sample height

by intercalating ions, and each layer can be peeled off when adhered to double-sided tape. Repeat this step three to four times to ensure complete removal of the upper layer, exposing a freshly cleaved surface (*see* **Note 6**).

3. Deposit a 10 μL droplet of DNA in the center of a mica disc, ensuring that the pipette tip does not contact the substrate. Wait for 2–3 min to ensure adequate coverage (*see* **Note 7**).
4. Carefully rinse the DNA solution with 2–3 mL of autoclaved ultrapure water. Perform this step by holding the mica over a sink using tweezers, so that excess DNA solution is rinsed from the mica substrate.
5. Dry the substrate under a gentle stream of nitrogen gas for ~1 min, making sure that any excess water is removed. Optionally, repeat the water rinse and nitrogen flow drying steps prior to imaging (*see* **Note 8**).
6. Perform AFM imaging at ambient temperature and humidity using tapping-mode AFM in air with silicon cantilevers operating near their resonance frequency and set point of 0.6–1.2 V (*see* **Note 9**).
7. Collect topographic images (512 × 512 pixels) with a scan size of 500 nm and scan rate near 5 Hz (*see* **Note 10**). Flatten AFM images by subtracting from each scan line a least-squares-fitted third-order polynomial using software available with the AFM instrument. Apply no additional background correction to the images. Figure 3 shows a representative AFM image of DNA fragments on mica. Specifically, the three-dimensional

representation (left), which visualizes several 2D equilibrium conformations of these DNA fragments when subjected to thermal fluctuations, can also be represented in two dimensions (right) with a color gradient that reflects the sample height (*see* **Note 11**).

3.3 Image Processing and Data Analysis

1. From a large set of images collected for each DNA variant, digitize the DNA molecules into pixel skeletons using an algorithm previously described [10]. Although the detection and thinning of the molecules is automated (*see* **Note 12**), a human supervisor can reject erroneously segmented skeletons or those not meeting set criteria during interactive steps. Figure 4 illustrates this method.
2. Align the generated sets of pixel coordinates (skeletons) along the positive *x*-axis to graphically illustrate the bending stiffness, Fig. 5a.

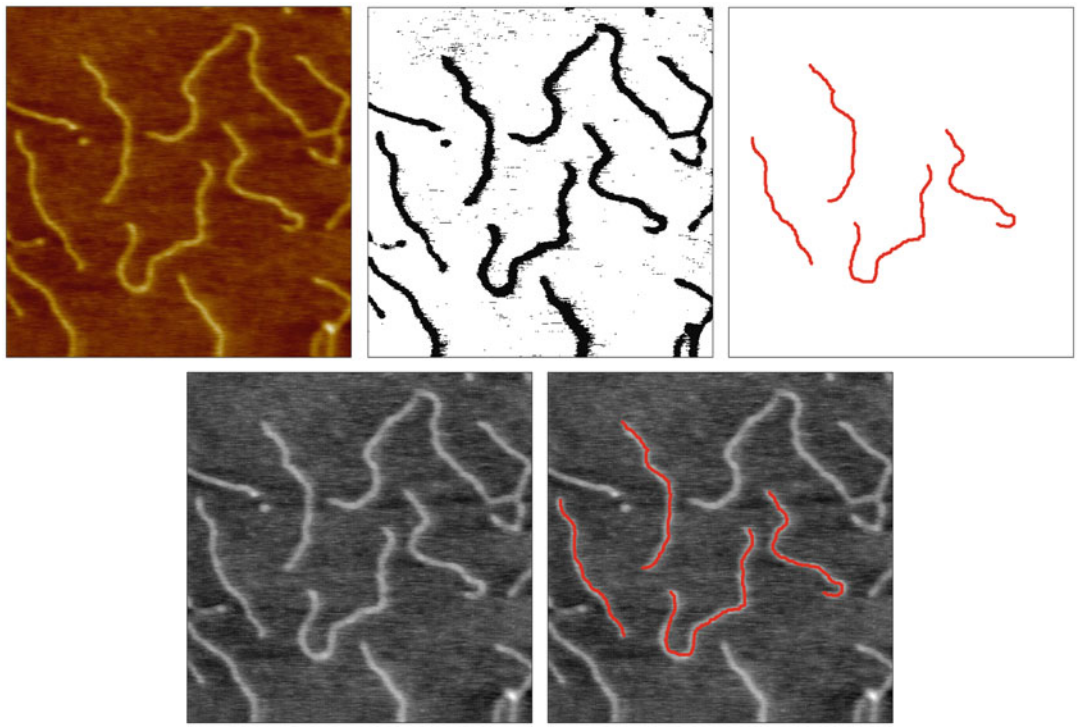


Fig. 4 Image processing algorithm. AFM images (top left) were transformed into bi-color maps (top center) with a threshold. Pixels above the threshold were then iteratively eroded from the edge of the DNA segments, if the removal of a pixel did not severe the segment. This thinning process was repeated until no more pixels could be removed, leaving behind DNA skeletons only one pixel wide (top right) for subsequent analysis. As a final step, a pixelated grayscale version of the image (bottom left) and one with the DNA skeletons superimposed (bottom right) were evaluated by a human supervisor

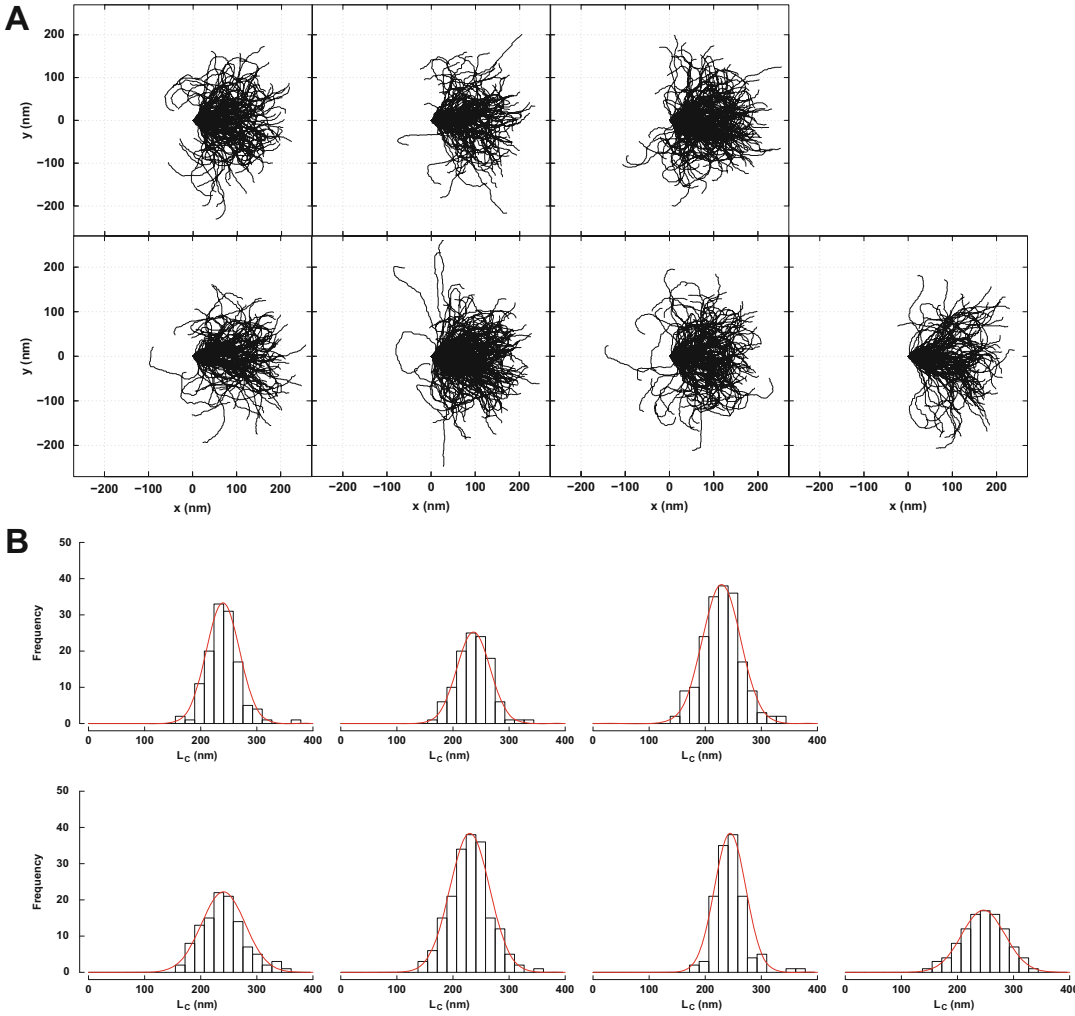


Fig. 5 Analysis of DNA pixel skeletons. Data are shown (left to right, top to bottom) for DNA variants 1–7, respectively. This convention will be continued for all subsequent figures. (**a**) Alignment of DNA pixel skeletons along the positive *x*-axis. The predominance of DNA skeletons in the first and fourth quadrants visually demonstrates the DNA bend stiffness. (**b**) Distribution of DNA pixel skeleton contours. Histograms were fit with Gaussian distributions (red lines)

3. Analyze the DNA skeletons (*see* **Note 13**), to determine the distribution of DNA contour lengths using the (n_e, n_o, n_c)-based corner chain estimator [11] (*see* **Note 14**), given by

$$L_C = 0.980n_e + 1.406n_o - 0.091n_c \quad (1)$$

The values of L_C appear normally distributed, as shown in Fig. 5b, with mean and standard deviation reported in Table 1 for the indicated number of skeletons, *N*.

Table 1
Analysis of DNA skeletons

Thymidine variant	*N*	L_C (nm)	*h* (Å/bp)
Variant 1	126	240 ± 30	3.18 ± 0.39
Variant 2	113	237 ± 30	3.14 ± 0.39
Variant 3	187	229 ± 34	3.04 ± 0.45
Variant 4	113	240 ± 39	3.19 ± 0.52
Variant 5	188	230 ± 35	3.05 ± 0.47
Variant 6	131	244 ± 28	3.25 ± 0.37
Variant 7	101	246 ± 39	3.27 ± 0.51

N is the number of molecules used in the analysis for a given DNA incorporating the thymidine variants shown in Fig. 1, L_C is the estimated contour length, and *h* is the estimated DNA helical rise, each presented as mean ± standard deviation. Data appear in Table 3 of [7].

4. Determine the value for DNA helical rise (h) by dividing the average estimated contour length by the expected total number of base pairs, Table 1 (*see* **Note 15**).
5. After detection of DNA skeletons, extract the trajectories of DNA centerlines using a routine developed with the particular goal of analyzing local bend angles [12] (*see* **Note 16**). Perform this procedure for step sizes l ranging from 3 to 17 nm, as shown in Fig. 6a.
6. From the generated sets of coordinates (which can be thought of as representations of the DNA axis) compute statistical quantities such as the Euclidean distance between the first and the last points of a DNA representation with contour length separation distance d, the so called end-end distance R (*see* **Note 17**). Use nonoverlapping logarithmic sampling of the DNA representations, Fig. 7, to avoid intrinsic correlations from sampling and to allow for the estimation of the statistical reliability of the measurements (*see* **Note 18**).

3.4 WLC Fitting and Persistence Length Determination

The wormlike chain (WLC) model captures the entropic elastic behavior of a semi-flexible biopolymer through idealization as an intrinsically straight, inextensible elastic rod [13]. In fact, the WLC model is the continuous limit of the discrete freely rotating chain, as illustrated in Fig. 8. For a chain of contour length L ($L \equiv nl$ for the discrete chain of n segments of length l), the end-end distance vector $\mathbf{R}$ is given from the unit tangent vector $\mathbf{u}(s)$ at arc length position s by

$$\int_0^L \mathbf{u}(s)\,ds = \mathbf{R} \tag{2}$$

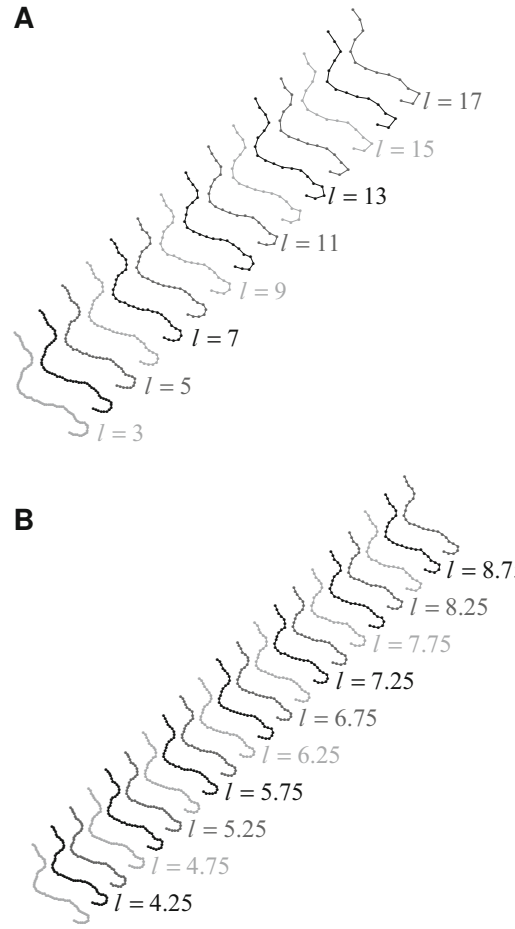


Fig. 6 Generating DNA centerline representations. Using the algorithm of Wiggins et al. [12] along with start and 'end points' defined from the pixel skeletons, (*x*,*y*) coordinate pairs separated by step size *l* were generated for each DNA molecule for step sizes $l = 3, 4, 5, \ldots, 17$ nm (**a**) and $l = 4, 4.25, 4.5, \ldots, 9$ nm (**b**). The true end point actually lies within a circular region centered on the pixel skeleton end point with radius equal to 1% of the chosen step size

The WLC model attributes bending deformations induced by thermal fluctuations (k_B is the Boltzmann constant and T is the absolute temperature) to a classical (obeying Hooke's law, *see* **Note 19**) elastic energy function

$$\frac{E_{\mathrm{WLC}}}{k_B T} = \frac{P\theta^2}{2d} \tag{3}$$

where the elastic bend constant P controls the exponential decay of the orientational correlation between unit tangent vectors separated by curvilinear (arc length) distance d with angle θ between them (Fig. 8).

$$\langle \mathbf{u}(s') \cdot \mathbf{u}(s'+d) \rangle = e^{-\frac{d}{P}} \tag{4}$$

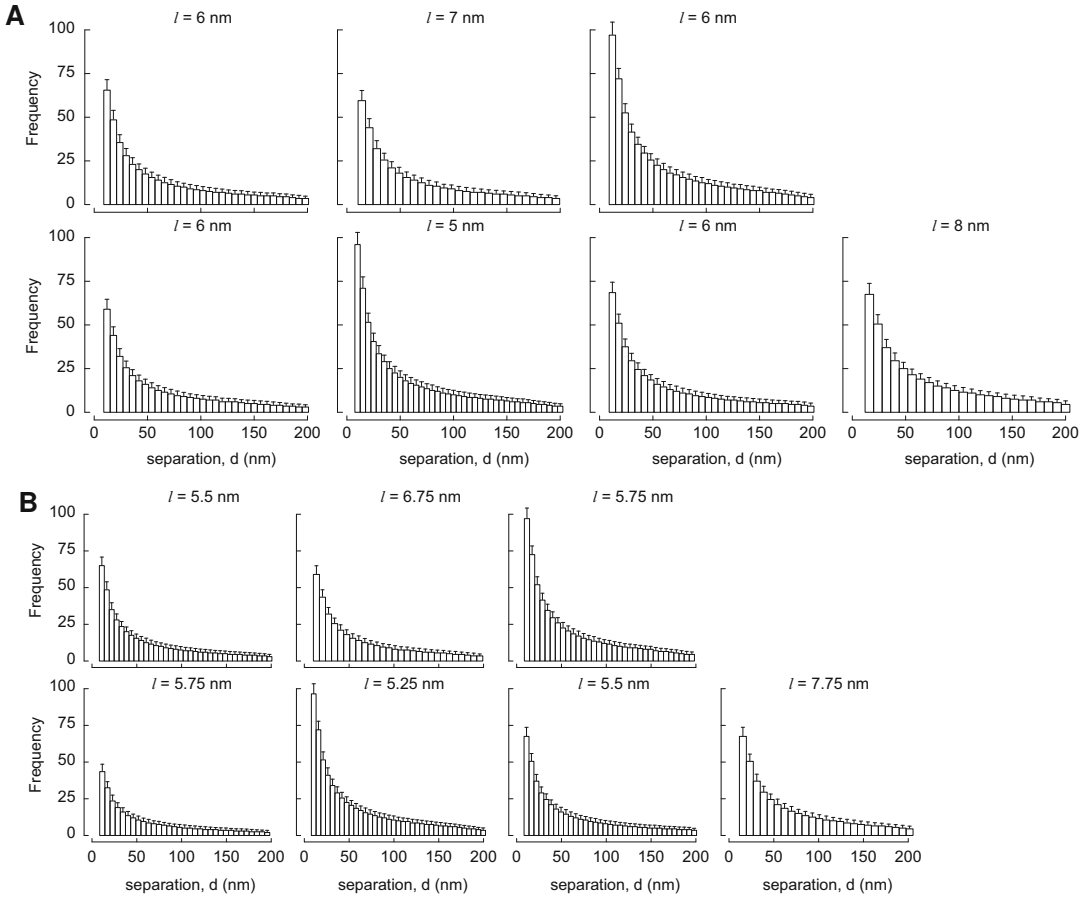


Fig. 7 Non-overlapping logarithmic sampling of DNA representations generated with the indicated step size *l*. Data are presented for the optimal step size from $l = 3, 4, 5, \ldots, 17$ nm (**a**) and $l = 4, 4.25, 4.5, \ldots, 9$ nm (**b**). How these optimal step sizes were chosen is discussed later in the text. This convention of showing data at the indicated optimal step size will be continued for all subsequent figures

and the $\langle\ \rangle$ notation indicates averaging over thermal realizations. Equation 4 can be reinterpreted using the definition of scalar projection as

$$\langle \cos\theta \rangle_{2D} = e^{-\frac{d}{(2P)}} \tag{5}$$

The elastic bend constant P is referred to as the *persistence length* (length through which the memory of the initial orientation of the chain persists) and quantifies the chain's resistance to bending (*see* **Note 20**).

The probability distribution of θ (angular distribution function) is given by Boltzmann statistics

$$G(\theta;d,P) = \frac{e^{-\frac{p\theta^2}{2d}}}{\int_{-\infty}^{\infty} e^{-\frac{P(\theta')^2}{2d}} d\theta'} = \sqrt{\frac{P}{2\pi d}} e^{-\frac{p\theta^2}{2d}} \tag{6}$$

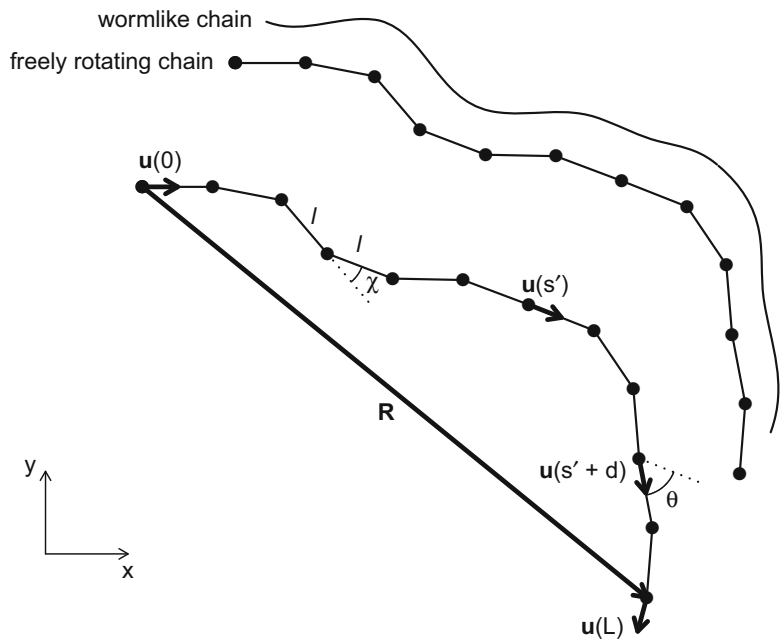


Fig. 8 Comparison of the continuous wormlike chain (WLC) model with the discrete freely rotating chain (FRC) model and illustration of the geometric quantities discussed in the text

Thus, the harmonic energy function makes the WLC angular distribution Gaussian (*see* **Note 21**). The prediction that $G(\theta)$ is normally distributed can be explored by examining the moments of θ, which are given by

$$\langle \theta^m \rangle = \int_{-\infty}^{\infty} \theta^m G(\theta) d\theta \tag{7}$$

All odd moments are zero (regardless of dimensionality of the molecules)

$$\langle \theta \rangle = \langle \theta^3 \rangle = \langle \theta^5 \rangle = \ldots = 0 \tag{8}$$

The first four odd moments of θ are shown in Figs. 9, 10, 11, and 12, respectively, after being normalized

$$\frac{\langle \theta \rangle}{\langle \theta^2 \rangle^{\frac{1}{2}}} \approx \frac{\langle \theta^3 \rangle}{\langle \theta^2 \rangle^{\frac{3}{2}}} \approx \frac{\langle \theta^5 \rangle}{\langle \theta^2 \rangle^{\frac{5}{2}}} \approx \frac{\langle \theta^7 \rangle}{\langle \theta^2 \rangle^{\frac{7}{2}}} \approx 0 \tag{9}$$

The even moments give nonzero predictions and each is a function of the three dimensional persistence length, i.e., $P \equiv P_{3D}$ (As a convention, the subscript will be omitted when referring to the 3D case). As discussed earlier, the process of DNA adsorption on a surface diminishes the number of possible DNA

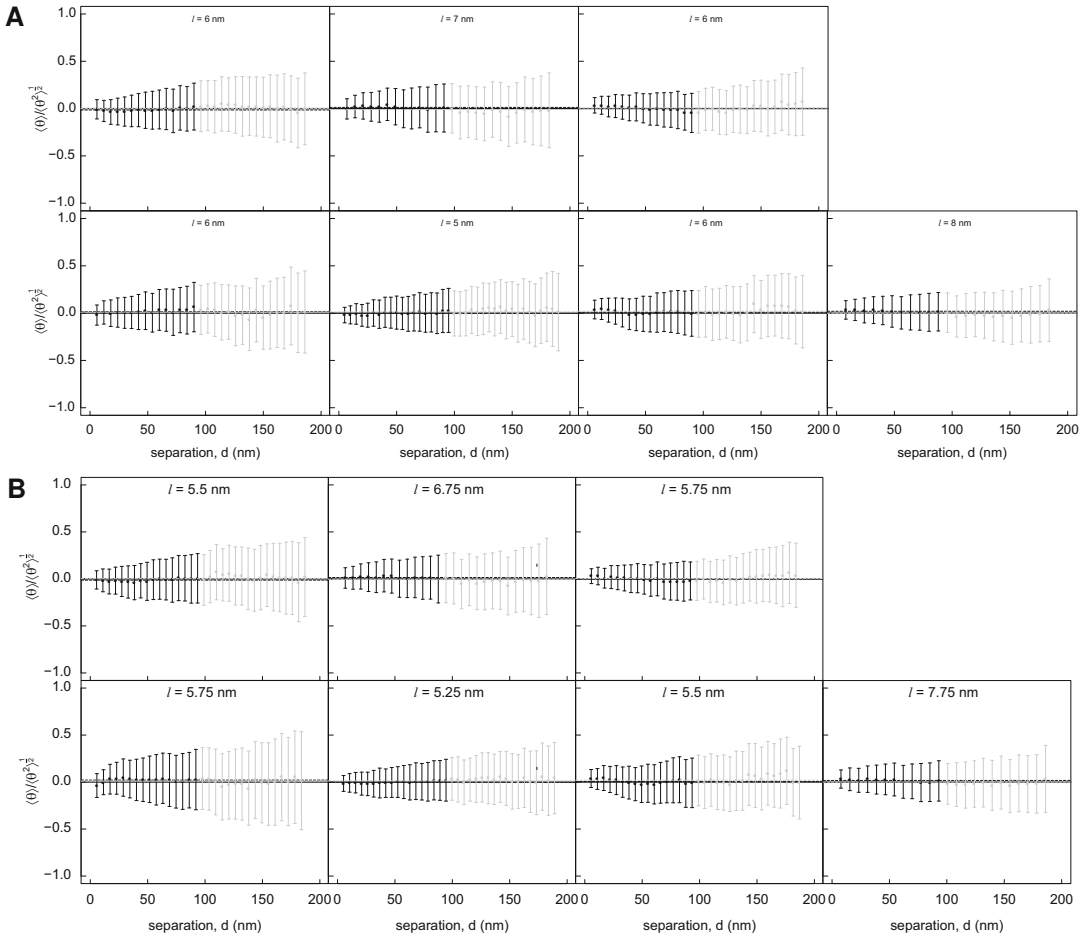


Fig. 9 Normalized first moment of θ. The mean of the black symbols is indicated with a solid line, while the mean of all symbols (black and gray) is indicated with a dashed line

conformations; therefore, any quantitative description of the apparent DNA conformation confined to a plane requires two-dimensional (2D) reformulation of the existing models for three-dimensional (3D) conformations. Consider

$$\langle \theta^2 \rangle = \frac{2d}{P} \tag{10}$$

This equation has an identical mathematical form in 2D

$$\langle \theta^2 \rangle_{2D} = \frac{2d}{P_{2D}} \tag{11}$$

Due to the loss of one degree of freedom moving 3D → 2D, there exists a scaling relationship between the two persistence lengths, so that the substitution $P_{2D} = 2P$ gives this 2D conformational statistic in terms of the desired 3D persistence length

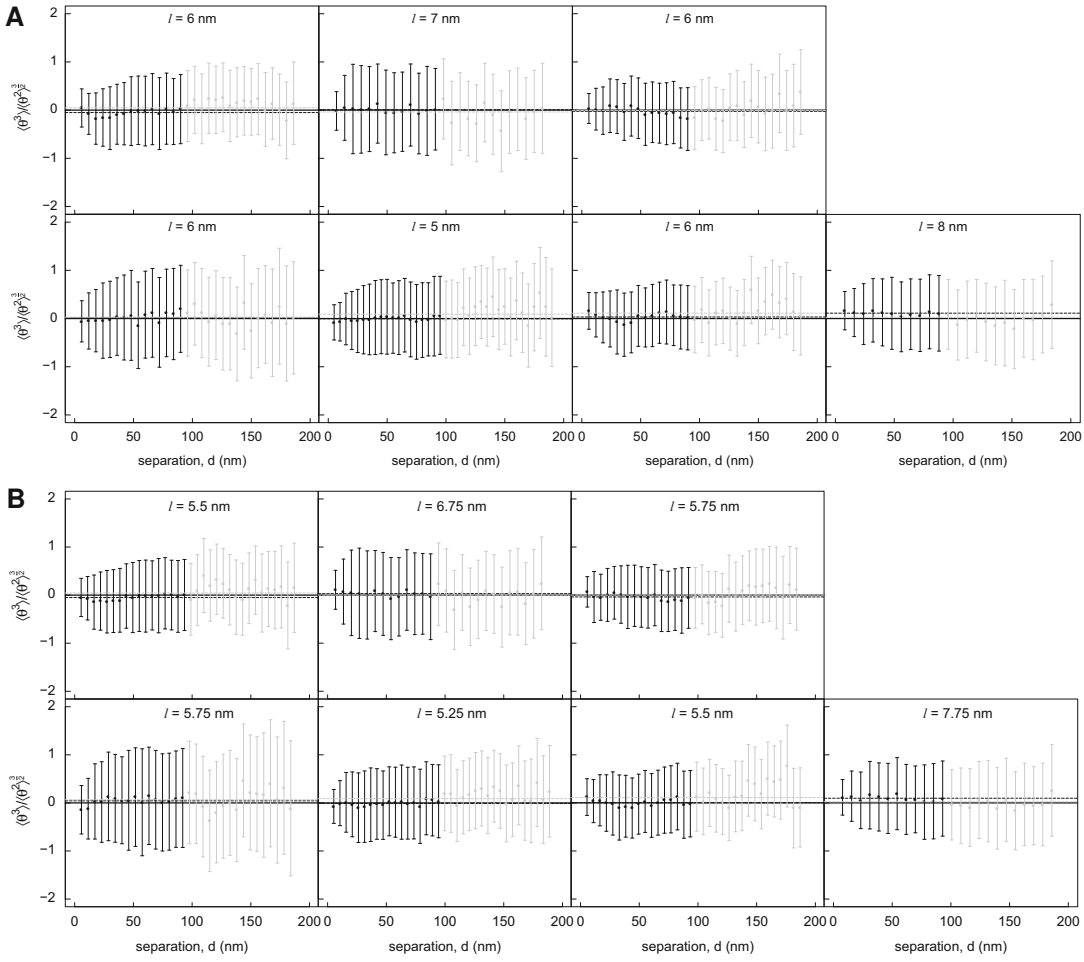


Fig. 10 Normalized third moment of θ. The mean of the black symbols is indicated with a solid line, while the mean of all symbols (black and gray) is indicated with a dashed line

$$\langle \theta^2 \rangle_{2D} = \frac{d}{P} \tag{12}$$

Figure 13 shows the second moment of θ (*see* **Note 22**). We obtain a persistence length estimate from this function (Eq. 12), or more generically the function $f(d;P)$, by employing a weighted nonlinear least squares method, i.e., minimizing the cost function $C(P)$ with respect to the free parameter P

$$\chi^2 = \min_P C(P) = \min_P \sum_{d=0}^{l_{\max}} \left(\frac{\mu_d - f(d;P)}{\sigma_d} \right)^2 \tag{13}$$

where μ_d and σ_d are experimentally determined mean values and standard deviations, respectively, from sampling and $f(d;P)$ are theoretical predictions of the WLC model. The fitted value of P (and its uncertainty) along with the average persistence length

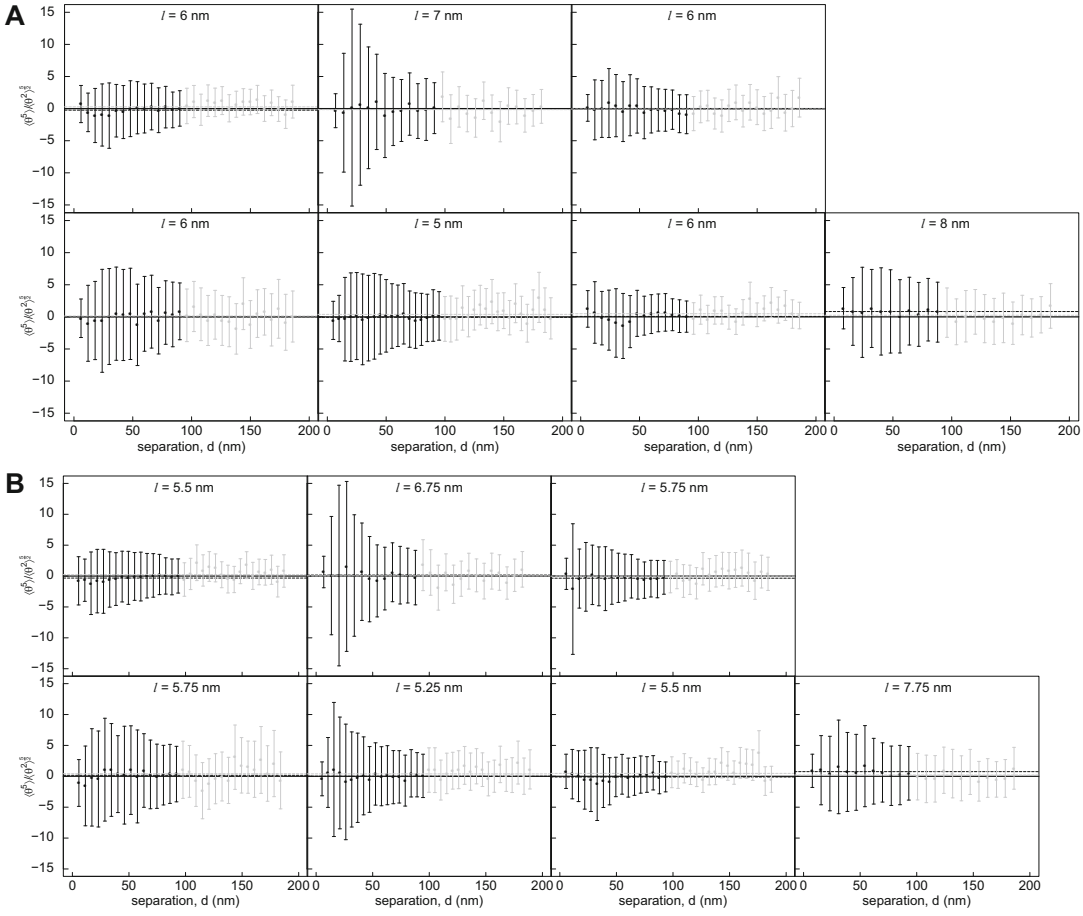


Fig. 11 Normalized fifth moment of θ. The mean of the black symbols is indicated with a solid line, while the mean of all symbols (black and gray) is indicated with a dashed line

value from ten distinct estimates, P_{ave}, are also shown in Fig. 13. The latter average will be included in all subsequent figures for comparison with individual values.

1. Perform fitting to the measured quantities with corresponding equations from the WLC theory. Begin with the second moment of θ.
2. Perform fitting to the fourth moment.

$$\langle \theta^4 \rangle_{2D} = 3\frac{d^2}{P^2} \tag{14}$$

Figure 14 shows the fourth moment of θ.

3. Perform fitting to the sixth moment.

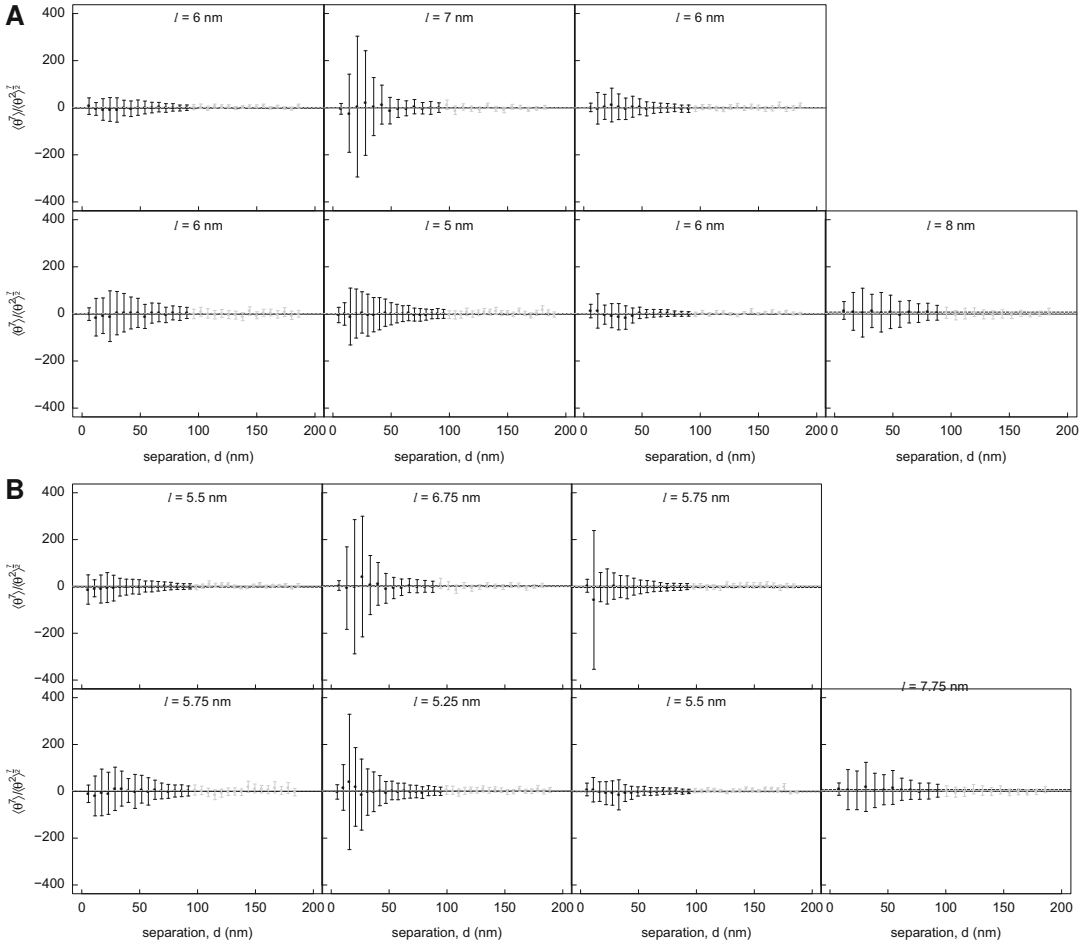


Fig. 12 Normalized seventh moment of θ. The mean of the black symbols is indicated with a solid line, while the mean of all symbols (black and gray) is indicated with a dashed line

$$\left\langle \theta^6 \right\rangle_{2D} = 15\frac{d^3}{P^3} \tag{15}$$

Figure 15 shows the sixth moment of θ.

4. Examine the ratio of the fourth moment to the square of the second moment, the so-called kurtosis (k), which satisfies the following.

$$k = \frac{\left\langle \theta^4 \right\rangle}{\left\langle \theta^2 \right\rangle^2} = 3 \tag{16}$$

Figure 16 shows k, which serves as an additional check of the normality prediction (*see* **Note 23**).

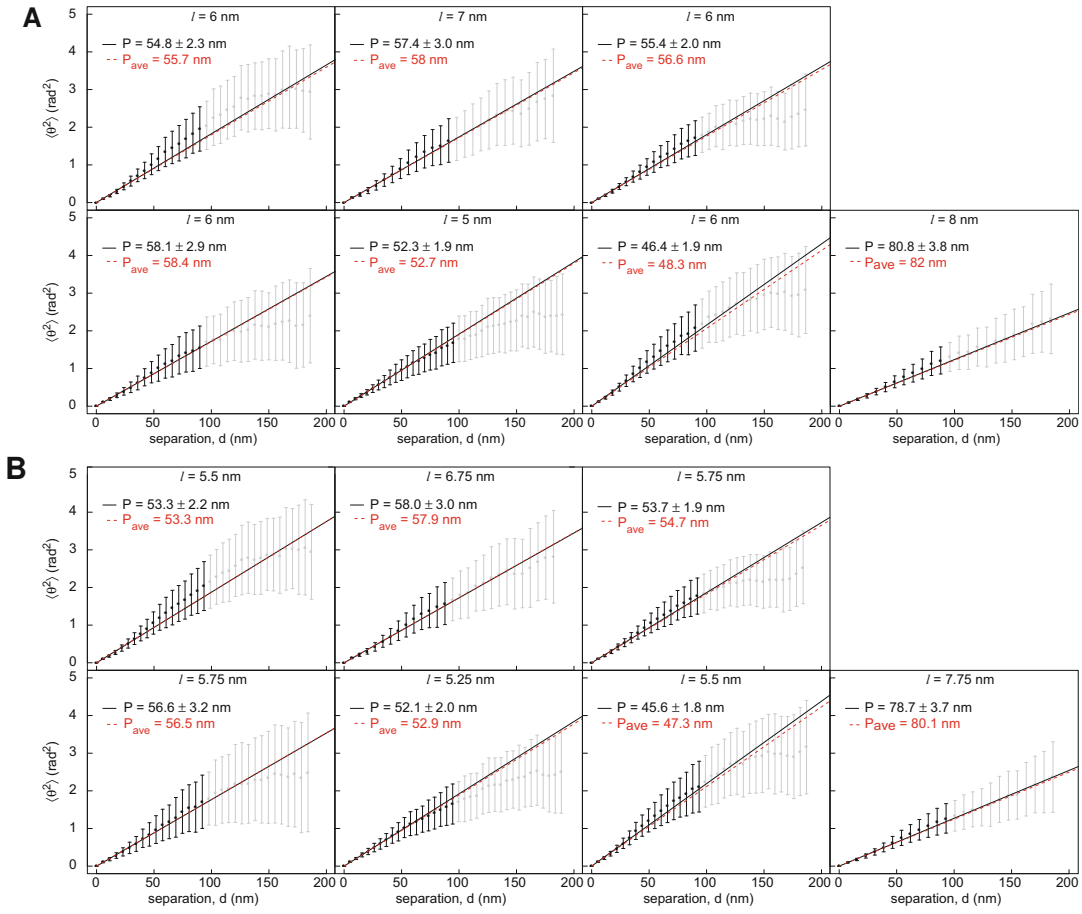


Fig. 13 The second moment of θ yields an estimate of P. The fitted value of P (and its uncertainty) along with the average persistence length value from ten distinct estimates, P_{ave}, are listed. The curve corresponding to P_{ave} is shown as a red dashed line. The value of P_{ave} and its associated curve will be included in all subsequent figures with estimates of P for comparison with the individual estimates

5. Examine one final check of normality of the angular distribution function, which comes from the ratio.

$$\frac{\langle \theta^6 \rangle}{\langle \theta^2 \rangle^3} = 15 \tag{17}$$

Figure 17 shows this final check of the normality prediction.

6. Perform fitting to cos θ. Recall the prediction of Eq. 5 (adjusted for dimensionality).

$$\langle \cos\theta \rangle_{2D} = e^{-\frac{d}{P}} \tag{18}$$

The fitted value of P from this approach is shown in Fig. 18.

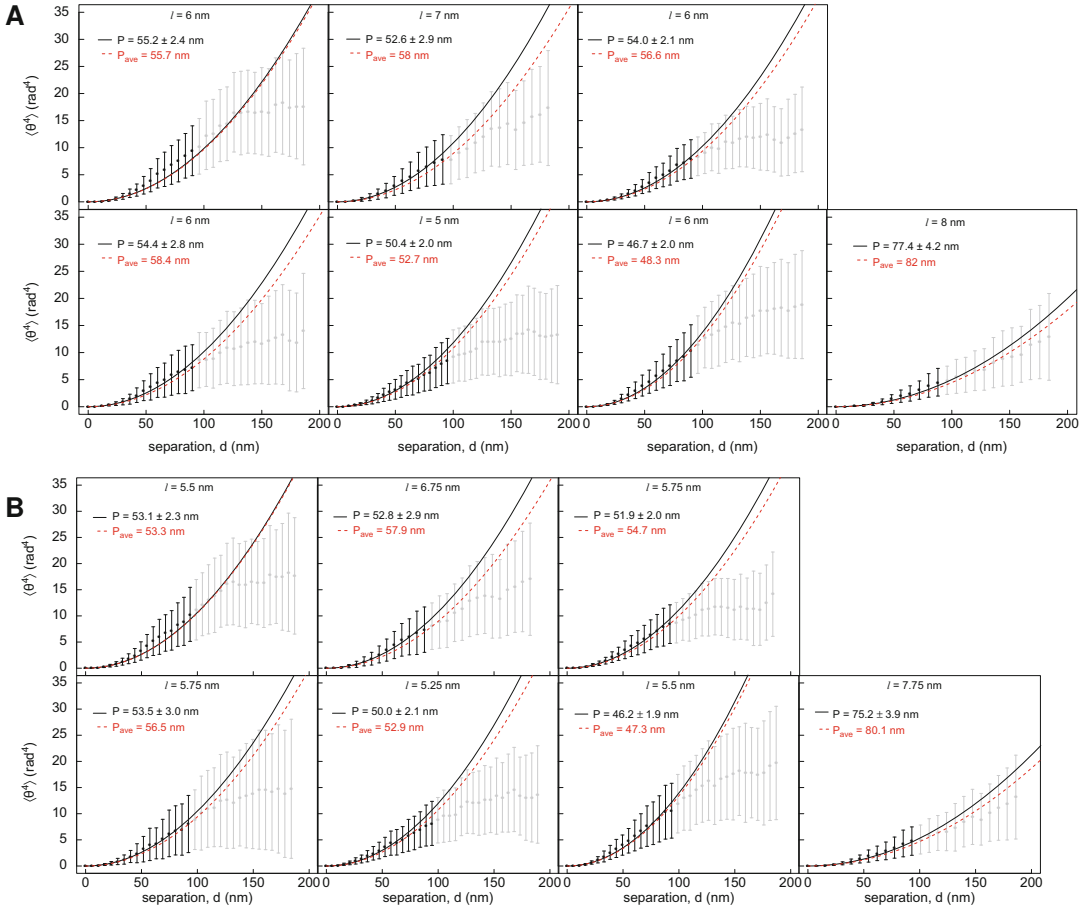


Fig. 14 The fourth moment of θ yields an estimate of P. See legend of Fig. 13

7. Perform fitting to the projection of the end-end distance vector $\mathbf{R}$ onto the initial tangent vector $\mathbf{u}_0$

$$\langle \mathbf{R} \cdot \mathbf{u}_0 \rangle_{2D} = \int_0^d \langle \mathbf{u}(s) \cdot \mathbf{u}_0 \rangle ds = \int_0^d e^{-\frac{s}{2P}} ds = 2P\left(1 - e^{-\frac{d}{2P}}\right) \quad (19)$$

Figure 19 shows $\langle \mathbf{R} \cdot \mathbf{u}_0 \rangle$.

8. Perform fitting to the second moment of R given by

$$\begin{aligned} \langle R^2 \rangle_{2D} &= \int_0^d \int_0^d \langle \mathbf{u}(s) \cdot \mathbf{u}(s') \rangle ds ds' \\ &= \int_0^d \int_0^d e^{-\frac{|s-s'|}{2P}} ds ds' \\ &= \int_0^d 2 \int_0^{s'} e^{-\frac{s-s'}{2P}} ds ds' \\ &= 4dP\left[1 - \frac{2P}{d}\left(1 - e^{-\frac{d}{2P}}\right)\right] \end{aligned} \quad (20)$$

The same result is reached by observing that $d\langle R^2 \rangle_{2D} = 2\langle \mathbf{R} \cdot \mathbf{u}_0 \rangle_{2D} ds$

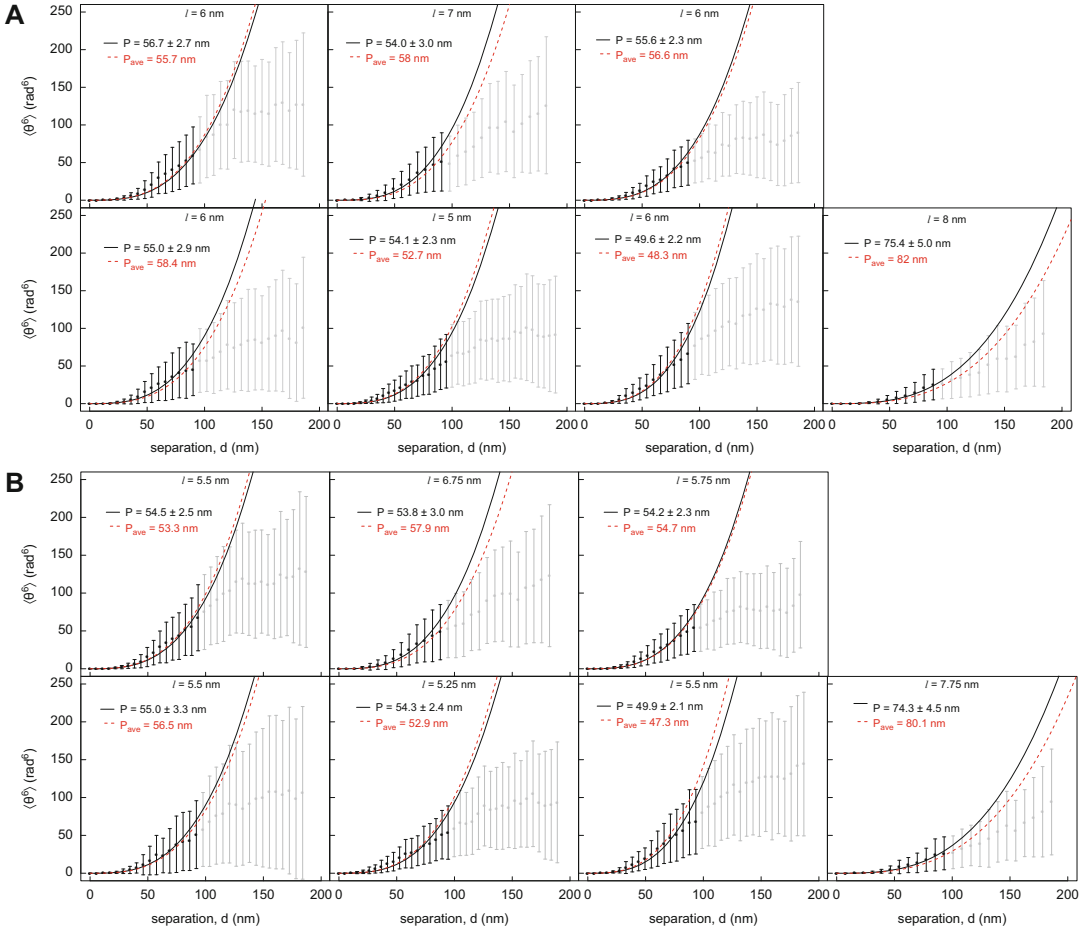


Fig. 15 The sixth moment of θ yields an estimate of P. See legend of Fig. 13

$$\begin{aligned}\left\langle R^2\right\rangle_{2D} &= \int_0^d 2\langle \mathbf{R}\cdot\mathbf{u}_0\rangle_{2D} ds = \int_0^d 4P\left(1-e^{-\frac{s}{2P}}\right) ds \\ &= 4dP\left[1-\frac{2P}{d}\left(1-e^{-\frac{d}{2P}}\right)\right] \end{aligned} \tag{21}$$

Figure 20 shows the second moment of R.

9. Perform fitting to the fourth moment of R.

$$\begin{aligned}\left\langle R^4\right\rangle_{2D} &= 32d^2P^2 - 240dP^3 + 696P^4 \\ &\quad -\frac{320}{3}dP^3e^{-\frac{d}{2P}} - \frac{6272}{9}P^4e^{-\frac{d}{2P}} + \frac{8}{9}P^4e^{-\frac{2d}{P}}\end{aligned} \tag{22}$$

Figure 21 shows the fourth moment of R assuming 2D chain statistics (*see* **Note 24**).

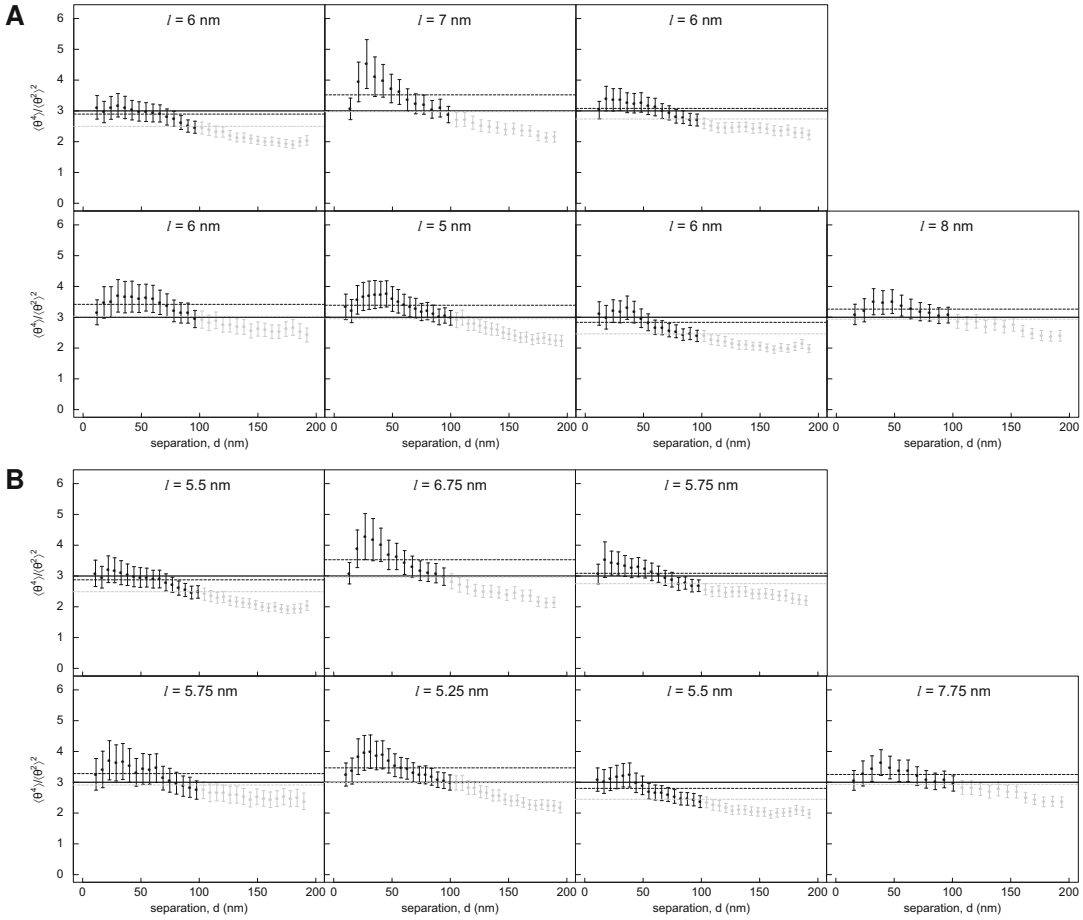


Fig. 16 Kurtosis. The kurtosis prediction of three is indicated with a bold black line. The mean of the black symbols is indicated with a dashed line, while the mean of all symbols (black and gray) is indicated with a gray dashed line

10. Perform fitting to the normalized difference $(\langle R^4\rangle - \langle R^2\rangle^2)/d^4$, which does not vanish for large separations. Fits using these equations are shown in Fig. 22.
11. Perform fitting to—ln $G(\theta)$. Recall the WLC angular distribution (Eq. 6). Binning of the measured angles is necessary to calculate the histogram—ln $G(\theta;d,P)$ [14]. For a bin size of $\Delta\theta$ around a bin center of θ, the probability is evaluated using the error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^{\pi} e^{-t^2} dt \tag{23}$$

so that

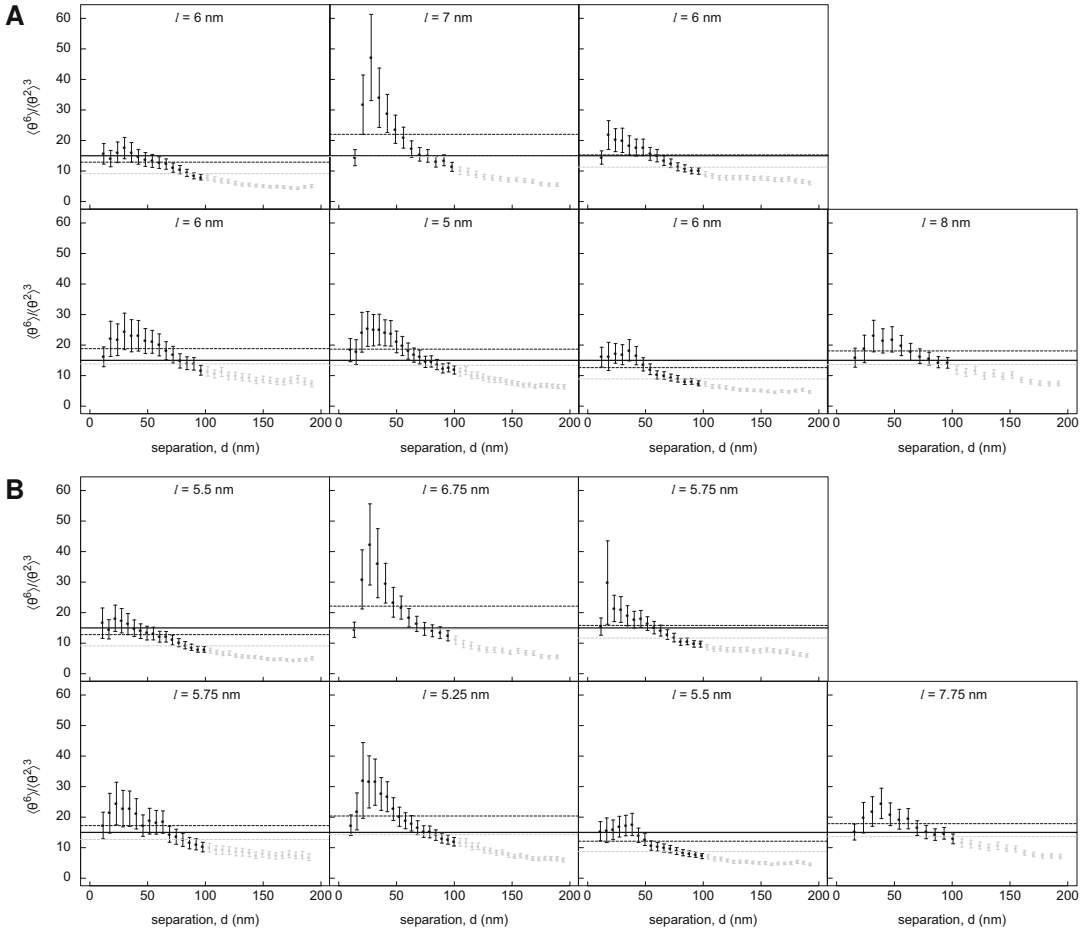


Fig. 17 Normality prediction. The prediction of 15 is indicated with a bold black line. The mean of the black symbols is indicated with a dashed line, while the mean of all symbols (black and gray) is indicated with a gray dashed line

$$G(\theta; d, P) = \operatorname{erf}\left(\sqrt{\frac{P}{2d}}\left(\theta + \frac{\Delta\theta}{2}\right)\right) - \operatorname{erf}\left(\sqrt{\frac{P}{2d}}\left(\theta - \frac{\Delta\theta}{2}\right)\right) \tag{24}$$

The negative logarithm of this equation was used for the fitting shown in Fig. 23.

12. Determine P using the maximal $\cos\chi$. Figure 8 shows that the angle between adjacent segments is denoted as χ (a special case of θ when $d = l$). Equation 18 is still applicable

$$\langle \cos\chi \rangle_{2D} = e^{-\frac{1}{2P}} \tag{25}$$

and can be rearranged to yield

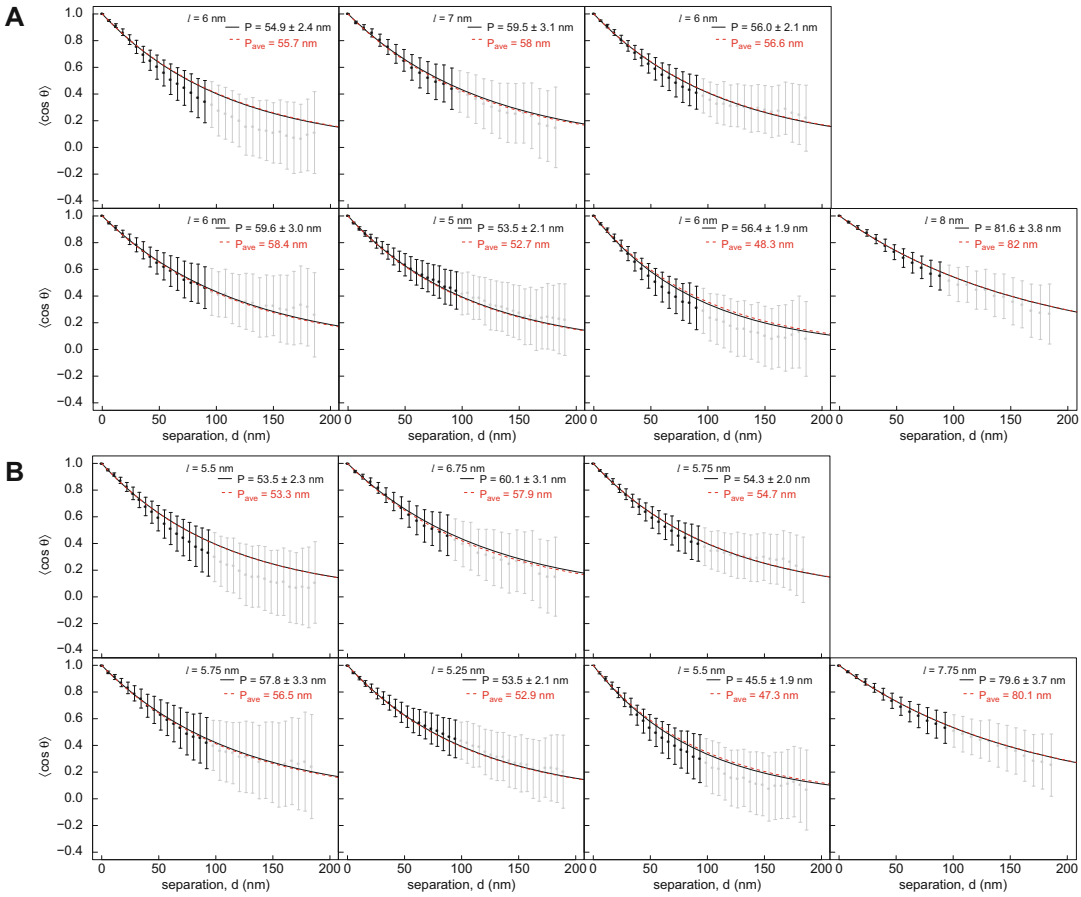


Fig. 18 Estimate of P from cos θ. See legend of Fig. 13

$$P = \frac{-l}{2\ln\left(\langle \cos\chi \rangle_{2D}\right)} \tag{26}$$

The quantity $\langle \cos\chi \rangle_{\max}$ provides a reliable means for determining the optimal segment length for analysis and provides an additional estimate of P at this optimal segment length (*see* **Note 25**). This approach is illustrated in Fig. 24.

13. Assemble the fitted values of P into summary tables. The complete analysis for each DNA incorporating a thymidine analog is given for variant 1 to variant 7 in Tables 2, 3, 4, 5, 6, 7, and 8, respectively. Step sizes l ranging by nanometer from 3 nm to 17 nm appear at the top of each table, and the optimal step size, l_{optimal}, is indicated in bold (*see* **Note 26**).

    We repeated the analysis of persistence length for step sizes ranging from 4 nm to 9 nm (*see* **Note 27**). This analysis appears in the bottom half of Tables 2, 3, 4, 5, 6, 7, and 8 for variant 1 to variant 7, respectively. For each variant, l_{optimal} is again indicated in bold. Importantly, the estimates of P from

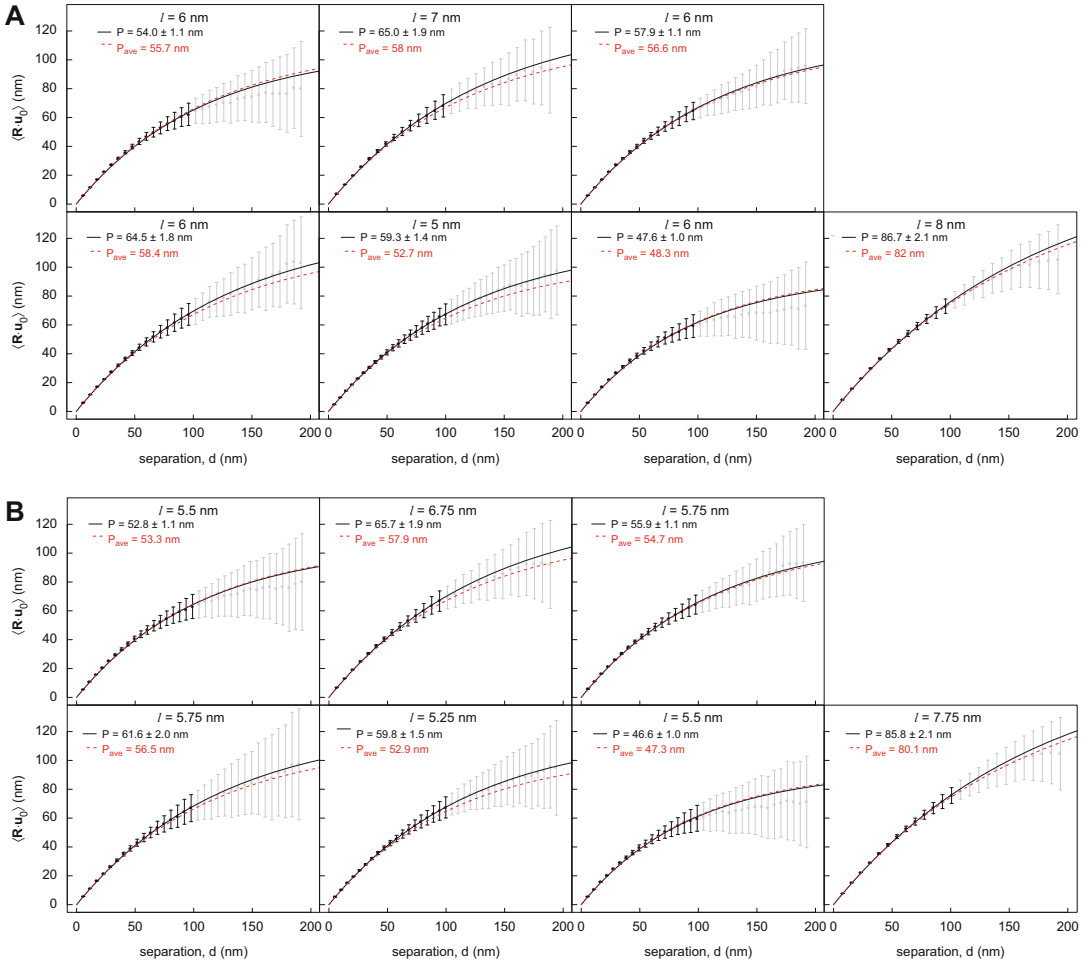


Fig. 19 Estimate of P from $\langle \mathbf{R} \cdot \mathbf{u}_0 \rangle$. See legend of Fig. 13

the fractional optimal step sizes l_{optimal} were consistently within one half nanometer of the previous estimates. This is well within the uncertainty from fitting. These results indicate that discretization of l (to integral values) has little effect on the estimates of P (Table 9) and renders a more exact determination of l_{optimal} largely unnecessary. Overall, the experimental data show good agreement with the theoretical expectations and exhibit consistency in the values of persistence length calculated using different formulae. This conformity in the calculated values for DNA persistence length derived from ten approaches (Table 9) provides strong support for using the framework of the WLC model to analyze AFM data.

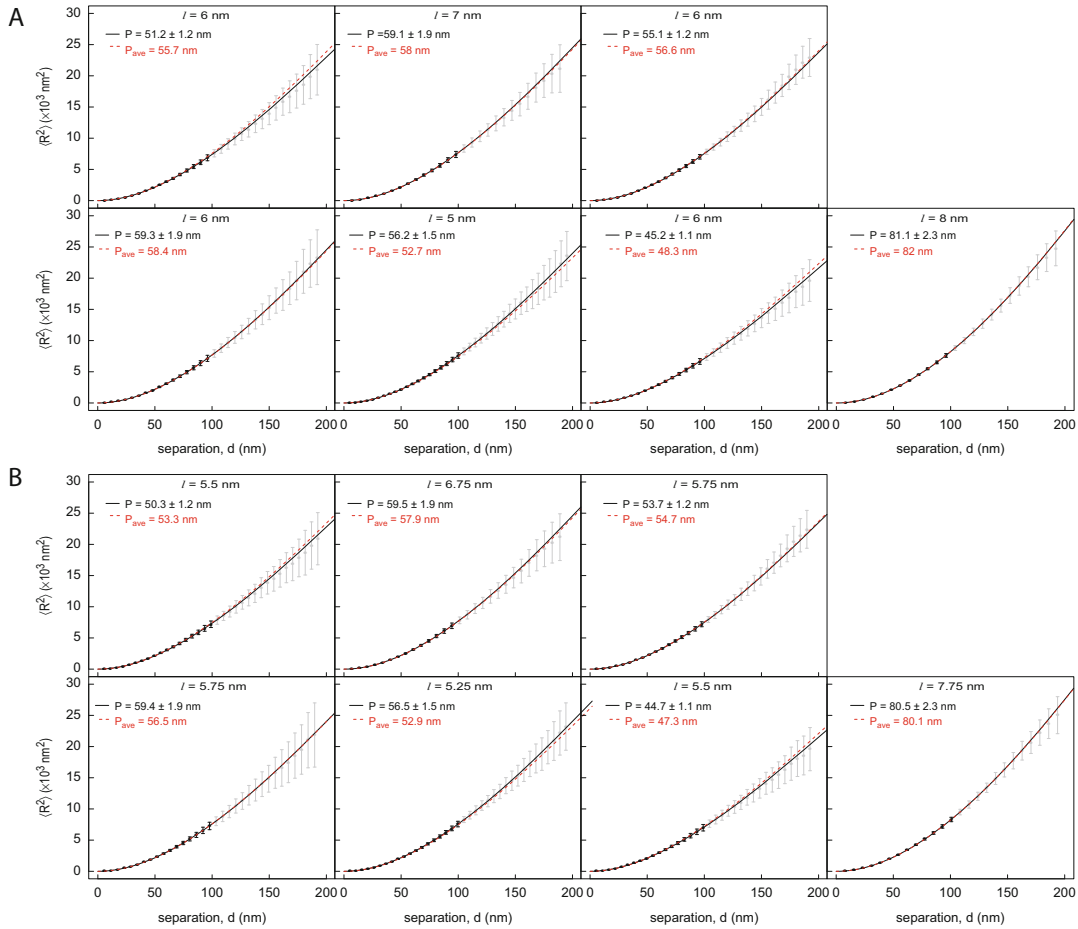


Fig. 20 Estimate of *P* from the second moment of *R*. See legend of Fig. 13

4 Notes

1. Analogs 1, 2, and 3 were purchased commercially (TriLink BioTechnologies), while analogs 4, 5, 6, and 7 were custom synthesized as described previously [8].
2. We amplify PCR products with primers LJM-4762 (5′-CG_2TG $ATGACG_2TGA_4$) and LJM-3223 (5′-$TGTGAGT_2AGCTCA$ $CTCAT_2AG_2$) using the reported conditions [8].
3. Alternatively 20 ng of purified PCR product from a previous reaction can serve as the template instead of plasmid DNA. This may help increase yield.
4. In addition to quantification using a NanoDrop 1000 Spectrophotometer (Thermo Scientific), we always recommend confirming the homogeneity of the purified DNA by gel

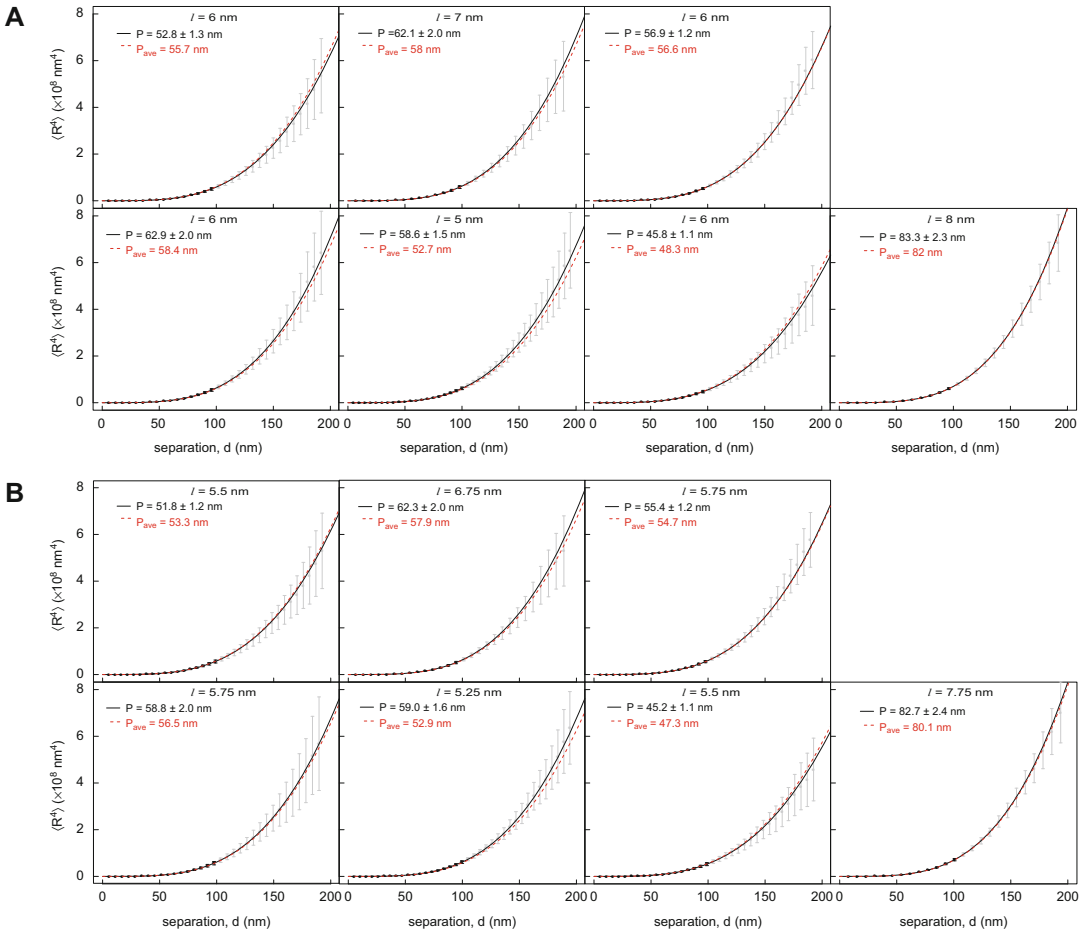


Fig. 21 Estimate of *P* from the fourth moment of *R*. See legend of Fig. 13

electrophoresis prior to AFM analysis to ensure a homogenous population of the expected size.

5. Since AFM is a surface technique, the way in which the DNA molecules adhere to the surface must be considered. Use of Mg^{2+} ions at a concentration lower than 6 mM allows DNA molecules to adhere to the mica surface weakly enough that interactions between the DNA and mica do not affect the chain statistics. Specifically, when low millimolar amounts of divalent cations (Mg^{2+}, Ni^{2+}) are used to promote adsorption, experiments have shown that DNA molecules are able to freely equilibrate in 2D conformations before being immobilized on the surface [1]. In contrast, strong adsorption without equilibration (i.e., kinetic trapping) leads to conformations reflecting a projection of the DNA chains in solution onto a plane.
6. We recommend always using the highest quality mica available, e.g., grade V1 (Ted Pella). Mica is the substrate of choice for

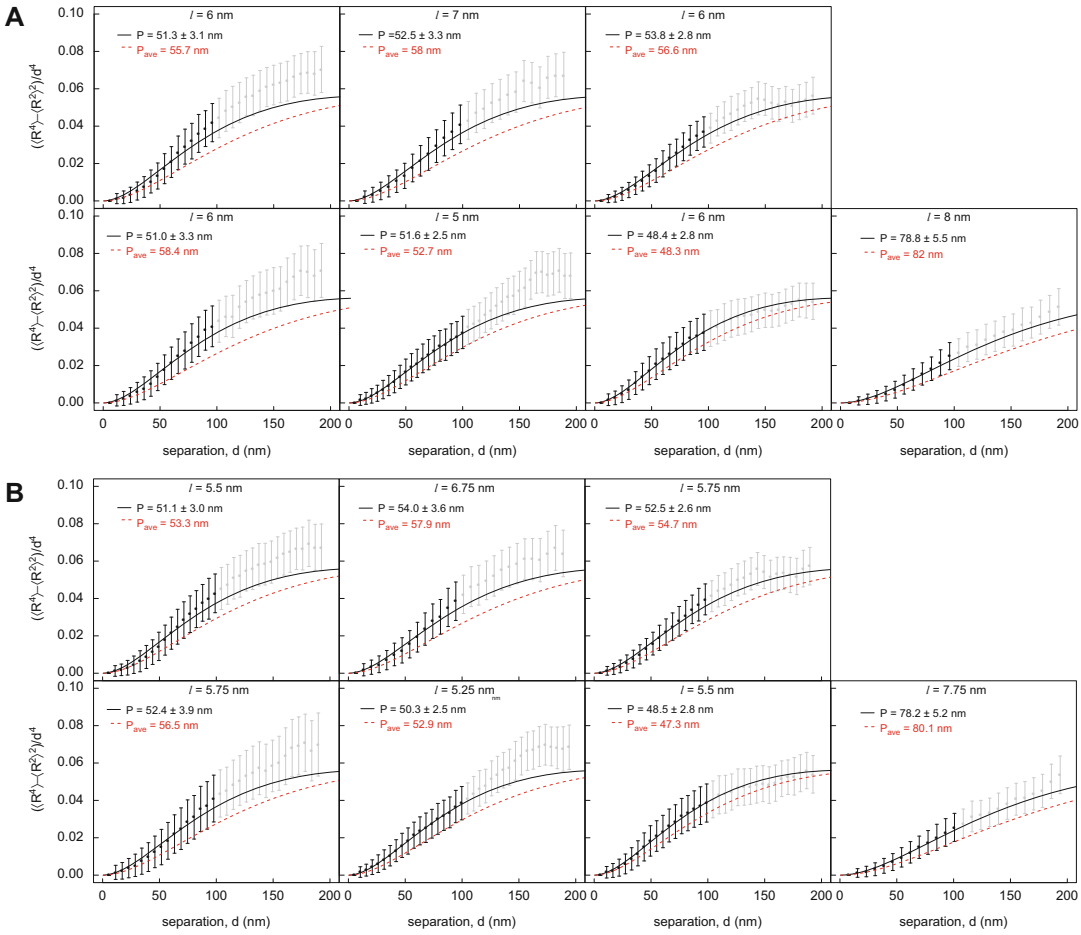


Fig. 22 Estimate of P from the difference $(\langle R^4\rangle - \langle R^2\rangle^2)/d^4$. *See* legend of Fig. 13

DNA studies because it is atomically flat, with a layer height of 0.37 nm ± 0.02 nm, and is easily cleaved to expose a clean surface.

7. The deposition time varies depending on the concentration of DNA used as well as the desired density of DNA fragments per frame.
8. Rinsing of the sample and its drying, steps that are required for immobilization of DNA onto the mica surface for imaging in air, may alter DNA conformation in a way that depends on the forces occurring between pre-equilibrated molecules and the mica surface. In typical solvent conditions, the Debye screening length is less than 5 nm. Since the rinsing buffer is pure water, the Debye length increases, potentially leading to an increase in the apparent persistence length. Thus, a completely dry substrate is necessary for faithful image collection.

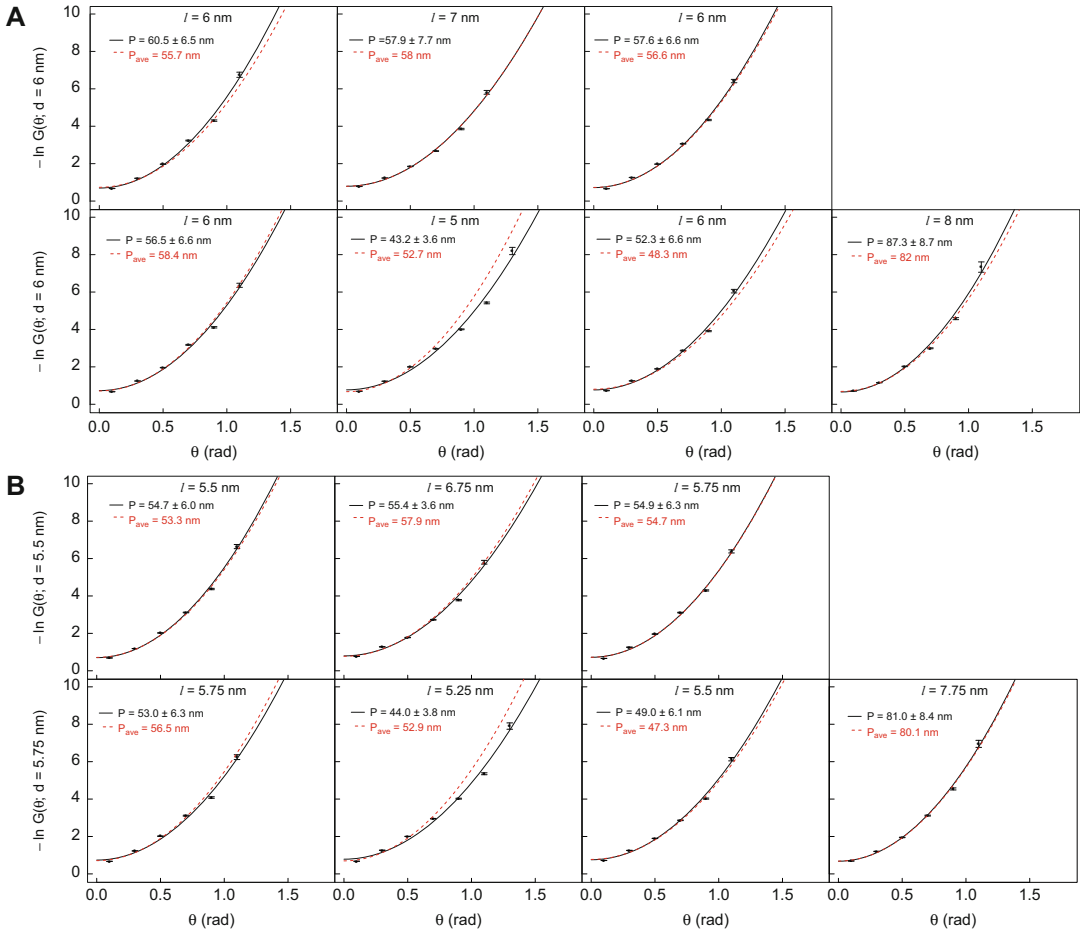


Fig. 23 Estimate of *P* from — ln *G*. See legend of Fig. 13

9. We use a NanoScope IV controlled Multimode equipped with a type E scanner (Bruker) operating in tapping mode. Tapping mode ensures that minimal lateral force is applied, compared to contact mode. The imaging parameters are dependent on the instrument and the choice of probe. We select AFM probes for tapping mode in air (FESP, Bruker), with a nominal resonant frequency of ~75 kHz, force constant of ~3 N/m, and tip radius <10 nm. The use of ultrasharp silicon probes allows a very high resolution to be obtained. The ability to image at near molecular resolution requires the tapping force to be sufficiently large to provide significant contrast in the AFM images, however, not so large that it will deform the sample.
10. We recommend using a separation for the AFM height measurements of ~1 nm, which is achieved when the scan size (in nanometers) roughly equals the number of pixels per line. Owing to experimental limitations of AFM (e.g., pixelation, tip

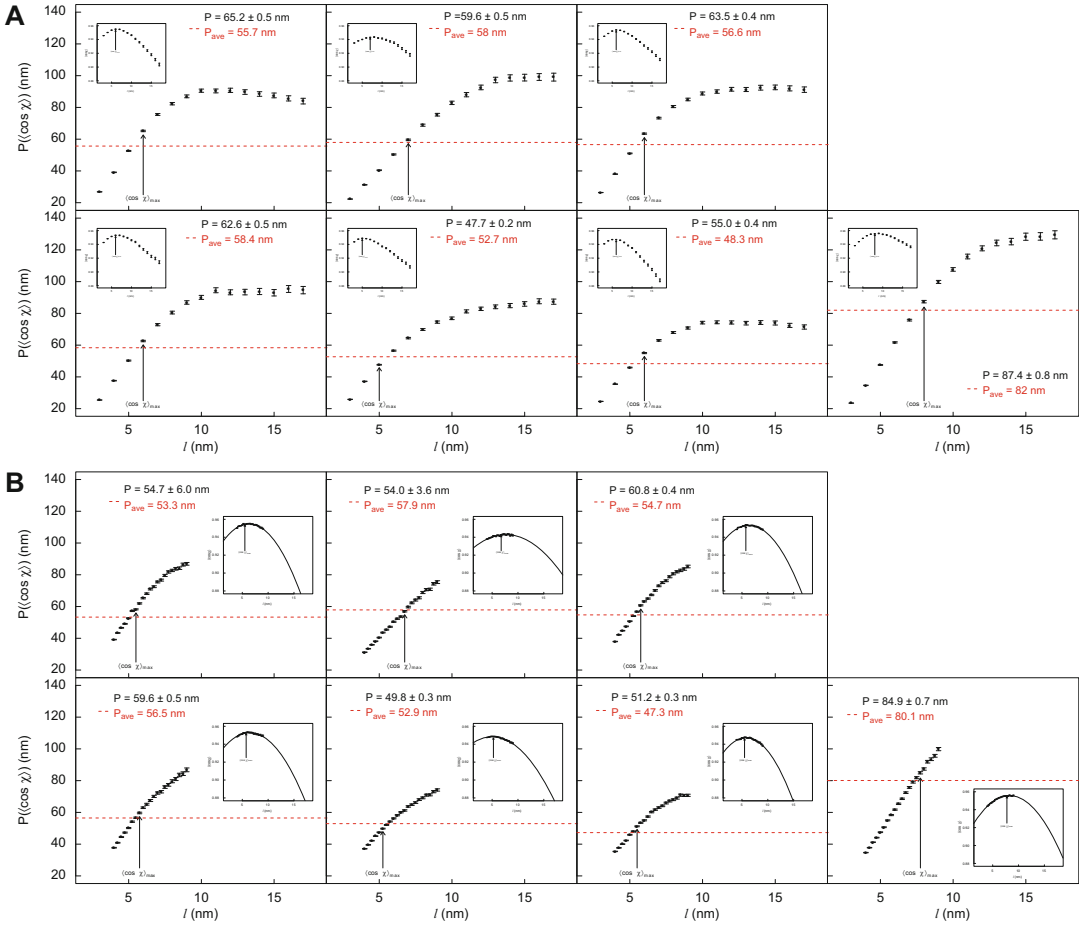


Fig. 24 Estimate of P from $P(\cos\chi_{max})$. The value of P (and its uncertainty) is determined at the step size corresonding to $\cos\chi_{max}$ for l ranging by nanometer from 3 nm to 17 nm (**a**) or l ranging by quarter nanometer from 4 nm to 9 nm (**b**). The value of the average persistence length from 10 distinct estimates, P_{ave}, is also listed. The curve corresponding to P_{ave} is shown as a red dashed line. The maximum value of $\langle\cos\chi\rangle$ is determined by fitting with a quadradic function (inset)

radius, and noise), experimental DNA traces only correspond to the physical conformation above a resolution limit. In Subheading 3.3, **step 5** we analyze the statistics of DNA at a 3 nm length scale minimum (even though the AFM height measurements are separated by ~1 nm), because, beyond this minimum resolution (typically near 6 nm, *see* **Note 25** and Tables 2, 3, 4, 5, 6, 7, and 8), the measured chain statistics reflect the underlying conformation of the chain rather than tracing artifacts.

11. Clean images of DNA are obtained by ensuring from visual inspection that the surface is free of any salt deposits or other impurities. Interaction of DNA with salt deposits could potentially introduce large distortions or otherwise disrupt chain statistics.

Table 2
Summary of AFM analysis for variant 1: the optimal *l* for each set of step sizes is indicated in bold

l (nm)	*P* ± SEM (nm)									*P* ± SD (nm)	
	$\langle \cos\theta \rangle$	$\langle \theta^2 \rangle$	$\langle \theta^4 \rangle$	$\langle \theta^6 \rangle$	$\langle \mathbf{R}\cdot\mathbf{u}_0 \rangle$	$\langle R^2 \rangle$	$\langle R^4 \rangle$	$\langle R^4 \rangle - \langle R^2 \rangle^2$	$-\ln G(\theta)$	$\langle \cos\chi \rangle$	Average
3	43.1 ± 1.7	43.7 ± 1.6	45.9 ± 1.8	50.4 ± 2.2	45.7 ± 1.2	45.6 ± 1.2	46.4 ± 1.3	52.8 ± 3.1	25.1 ± 3.3	26.8 ± 0.1	42.6 ± 9.2
4	47.5 ± 2.0	47.4 ± 1.9	47.9 ± 2.0	50.9 ± 2.3	48.9 ± 1.2	48.1 ± 1.2	49.2 ± 1.3	52.1 ± 3.1	37.3 ± 2.9	39.1 ± 0.2	46.8 ± 4.8
5	51.9 ± 2.2	51.8 ± 2.1	52.5 ± 2.2	54.8 ± 2.5	51.4 ± 1.1	49.4 ± 1.1	50.8 ± 1.2	51.2 ± 2.9	48.1 ± 5.5	52.6 ± 0.3	51.5 ± 1.8
6	**54.9 ± 2.4**	**54.8 ± 2.3**	**55.2 ± 2.4**	**56.7 ± 2.7**	**54.0 ± 1.1**	**51.2 ± 1.2**	**52.8 ± 1.3**	**51.3 ± 3.1**	**60.5 ± 6.5**	**65.2 ± 0.5**	**55.7 ± 4.3**
7	58.1 ± 2.6	57.8 ± 2.6	55.7 ± 2.7	54.5 ± 2.7	55.6 ± 1.1	52.0 ± 1.2	54.0 ± 1.3	50.5 ± 2.9	71.6 ± 7.6	75.6 ± 0.6	58.5 ± 8.3
8	60.3 ± 2.8	60.0 ± 2.8	57.6 ± 3.0	55.1 ± 2.9	57.3 ± 1.1	53.2 ± 1.2	55.4 ± 1.3	50.4 ± 3.0	76.6 ± 8.8	82.3 ± 0.8	60.8 ± 10.4
9	63.4 ± 2.9	63.3 ± 3.0	62.7 ± 3.3	59.6 ± 3.3	58.0 ± 1.1	53.2 ± 1.2	55.3 ± 1.3	51.2 ± 2.9	78.9 ± 9.9	87.0 ± 0.9	63.3 ± 11.3
10	64.2 ± 3.1	64.1 ± 3.2	64.5 ± 3.4	64.7 ± 3.7	59.4 ± 1.1	54.2 ± 1.3	56.5 ± 1.4	50.9 ± 2.9	83.1 ± 11.1	90.5 ± 1.0	65.2 ± 12.5
11	63.9 ± 3.2	63.7 ± 3.3	64.5 ± 3.5	66.0 ± 3.9	61.2 ± 1.3	55.1 ± 1.4	57.7 ± 1.5	50.7 ± 3.0	79.5 ± 8.0	90.4 ± 1.2	65.3 ± 11.7
12	64.9 ± 3.4	64.9 ± 3.5	64.8 ± 3.8	64.7 ± 4.1	63.6 ± 1.4	56.7 ± 1.5	59.1 ± 1.6	52.3 ± 3.2	78.0 ± 8.8	90.8 ± 1.3	66.0 ± 11
13	63.1 ± 3.3	62.9 ± 3.3	63.8 ± 3.5	65.8 ± 3.8	62.2 ± 1.4	55.4 ± 1.5	58.0 ± 1.6	50.4 ± 2.9	78.4 ± 9.6	89.7 ± 1.4	65.0 ± 11.4
14	64.7 ± 3.4	64.6 ± 3.5	64.0 ± 3.7	63.6 ± 3.8	65.3 ± 1.5	57.3 ± 1.6	59.9 ± 1.7	52.0 ± 3.1	75.7 ± 10.5	88.5 ± 1.4	65.6 ± 10.1
15	64.0 ± 3.5	63.9 ± 3.5	63.4 ± 3.7	62.7 ± 3.7	64.9 ± 1.5	56.6 ± 1.6	59.2 ± 1.7	51.5 ± 3.0	73.0 ± 7.8	87.6 ± 1.5	64.7 ± 9.8
16	64.0 ± 3.6	64.1 ± 3.7	64.7 ± 3.9	65.9 ± 4.2	68.0 ± 1.7	58.4 ± 1.8	61.1 ± 1.9	52.3 ± 3.2	72.6 ± 8.4	85.7 ± 1.6	65.7 ± 8.9
17	62.2 ± 3.5	61.9 ± 3.5	61.3 ± 3.5	61.0 ± 3.5	67.8 ± 1.8	57.8 ± 1.8	60.7 ± 1.9	50.9 ± 3.0	72.4 ± 9.0	84.0 ± 1.8	64.0 ± 9.0

4	47.5 ± 1.9	47.4 ± 1.9	47.8 ± 2.0	50.6 ± 2.2	48.7 ± 1.1	47.9 ± 1.2	48.9 ± 1.2	52.2 ± 3.1	37.3 ± 2.9	39.2 ± 0.2	46.7 ± 4.8
4.25	49.2 ± 2.0	49.2 ± 1.9	49.7 ± 2.0	52.0 ± 2.3	49.5 ± 1.1	48.5 ± 1.2	49.8 ± 1.2	51.8 ± 3.0	39.0 ± 4.7	43.3 ± 0.2	48.2 ± 4.0
4.5	49.9 ± 2.1	49.9 ± 2.0	50.2 ± 2.1	52.6 ± 2.4	50.5 ± 1.1	49.1 ± 1.2	50.5 ± 1.2	51.5 ± 3.0	45.0 ± 4.9	46.6 ± 0.2	49.6 ± 2.2
4.75	50.7 ± 2.1	50.6 ± 2.0	51.0 ± 2.1	53.1 ± 2.3	50.6 ± 1.1	48.6 ± 1.1	49.9 ± 1.2	51.3 ± 3.0	43.7 ± 5.2	49.1 ± 0.3	49.9 ± 2.5
5	51.9 ± 2.2	51.7 ± 2.1	52.2 ± 2.2	54.3 ± 2.4	51.3 ± 1.1	49.6 ± 1.2	51.2 ± 1.2	51.2 ± 3.0	48.4 ± 5.5	52.6 ± 0.3	51.5 ± 1.6
5.25	53.2 ± 2.2	53.0 ± 2.2	53.3 ± 2.3	55.3 ± 2.5	52.0 ± 1.1	50.0 ± 1.2	51.6 ± 1.2	51.0 ± 2.9	53.7 ± 5.7	57.2 ± 0.4	53.0 ± 2.1
5.5	**53.5 ± 2.3**	**53.3 ± 2.2**	**53.1 ± 2.3**	**54.5 ± 2.5**	**52.8 ± 1.1**	**50.3 ± 1.2**	**51.8 ± 1.2**	**51.1 ± 3.0**	**54.7 ± 6.0**	**58.2 ± 0.3**	**53.3 ± 2.2**
5.75	54.8 ± 2.3	54.6 ± 2.3	54.5 ± 2.3	55.4 ± 2.5	53.6 ± 1.1	50.9 ± 1.2	52.4 ± 1.2	51.2 ± 3.0	53.0 ± 4.1	62.0 ± 0.4	54.2 ± 3.1
6	55.0 ± 2.3	54.8 ± 2.3	55.0 ± 2.4	56.4 ± 2.6	54.2 ± 1.1	51.3 ± 1.2	53.1 ± 1.3	51.3 ± 3.0	60.5 ± 6.5	65.3 ± 0.4	55.7 ± 4.3
6.25	56.3 ± 2.4	56.1 ± 2.4	56.0 ± 2.4	57.0 ± 2.6	53.9 ± 1.0	51.0 ± 1.1	52.7 ± 1.2	50.7 ± 2.9	64.7 ± 6.8	68.1 ± 0.5	56.7 ± 5.6
6.5	56.6 ± 2.5	56.2 ± 2.5	54.8 ± 2.6	54.4 ± 2.6	54.2 ± 1.1	51.3 ± 1.2	53.1 ± 1.3	50.7 ± 3.0	68.6 ± 7.1	70.9 ± 0.5	57.1 ± 7.0
6.75	57.0 ± 2.5	56.6 ± 2.5	55.1 ± 2.6	54.4 ± 2.8	55.5 ± 1.1	52.2 ± 1.2	54.1 ± 1.3	51.2 ± 3.1	59.8 ± 4.9	72.5 ± 0.6	56.8 ± 6.0
7	58.3 ± 2.6	57.8 ± 2.6	55.5 ± 2.7	54.3 ± 2.7	55.2 ± 1.1	51.9 ± 1.2	53.9 ± 1.2	50.4 ± 2.9	70.6 ± 7.6	75.5 ± 0.6	58.3 ± 8.2
7.25	58.8 ± 2.6	58.5 ± 2.7	56.7 ± 2.8	54.9 ± 2.8	56.6 ± 1.1	52.6 ± 1.2	54.6 ± 1.3	51.0 ± 3.1	73.1 ± 7.9	76.6 ± 0.6	59.3 ± 8.6
7.5	60.1 ± 2.7	59.8 ± 2.8	58.3 ± 2.9	56.2 ± 2.9	56.2 ± 1.1	52.4 ± 1.2	54.4 ± 1.3	50.5 ± 3.0	73.6 ± 8.2	79.6 ± 0.7	60.1 ± 9.3
7.75	60.0 ± 2.7	59.6 ± 2.7	57.7 ± 2.8	56.1 ± 2.8	56.4 ± 1.1	52.1 ± 1.2	54.0 ± 1.2	50.2 ± 2.9	70.4 ± 8.5	81.8 ± 0.7	59.8 ± 9.5
8	60.6 ± 2.8	60.2 ± 2.9	57.7 ± 3.0	55.4 ± 2.9	57.8 ± 1.2	53.3 ± 1.3	55.4 ± 1.3	50.6 ± 3.1	75.6 ± 8.8	82.7 ± 0.7	60.9 ± 10.2
8.25	60.3 ± 2.7	60.0 ± 2.8	59.6 ± 2.9	58.2 ± 3.0	57.3 ± 1.1	52.8 ± 1.2	54.9 ± 1.3	50.4 ± 2.9	73.5 ± 9.1	83.8 ± 0.8	61.1 ± 10.1
8.5	61.4 ± 2.9	61.3 ± 2.9	60.8 ± 3.2	59.1 ± 3.3	58.6 ± 1.2	53.9 ± 1.3	56.0 ± 1.4	51.8 ± 3.1	82.4 ± 9.3	84.2 ± 0.8	63.0 ± 11.2
8.75	61.6 ± 2.9	61.3 ± 3.0	58.9 ± 3.1	56.7 ± 3.1	58.4 ± 1.2	53.7 ± 1.3	56.0 ± 1.3	50.8 ± 3.0	74.7 ± 9.7	86.4 ± 0.9	61.9 ± 10.7
9	62.8 ± 2.9	62.7 ± 3.0	62.0 ± 3.2	59.3 ± 3.2	58.3 ± 1.1	53.7 ± 1.2	56.2 ± 1.3	50.8 ± 2.9	78.2 ± 9.9	86.9 ± 0.9	63.1 ± 11.1

Table 3
Summary of AFM analysis for variant 2: the optimal *l* for each set of step sizes is indicated in bold

l (nm)	*P* ± SEM (nm)									*P* ± SD (nm)	
	$\langle\cos\theta\rangle$	$\langle\theta^2\rangle$	$\langle\theta^4\rangle$	$\langle\theta^6\rangle$	$\langle \mathbf{R}\cdot\mathbf{u}_0\rangle$	$\langle R^2\rangle$	$\langle R^4\rangle$	$\langle R^4\rangle - \langle R^2\rangle^2$	$-\ln G(\theta)$	$\langle\cos\chi\rangle$	Average
3	44.8 ± 2.1	44.5 ± 1.9	46.0 ± 2.3	51.9 ± 3.0	53.2 ± 2.2	49.4 ± 2.0	50.6 ± 1.9	55.4 ± 3.5	21.4 ± 2.2	22.3 ± 0.1	44.0 ± 12.2
4	50.2 ± 2.4	49.0 ± 2.3	48.1 ± 2.5	52.3 ± 3.1	57.1 ± 2.1	52.7 ± 1.9	54.3 ± 1.9	54.4 ± 3.5	29.1 ± 4.5	31.2 ± 0.1	47.8 ± 9.7
5	53.6 ± 2.6	52.0 ± 2.5	49.5 ± 2.5	53.0 ± 2.8	59.3 ± 1.9	54.9 ± 1.8	57.1 ± 1.9	52.9 ± 3.3	35.8 ± 5.6	40.4 ± 0.2	50.9 ± 7.3
6	57.5 ± 2.9	55.5 ± 2.8	51.3 ± 2.9	53.2 ± 3.2	62.8 ± 1.9	57.2 ± 1.9	59.8 ± 1.9	53.0 ± 3.5	50.4 ± 4.3	50.4 ± 0.3	55.1 ± 4.2
7	**59.5 ± 3.1**	**57.4 ± 3.0**	**52.6 ± 2.9**	**54.0 ± 3.0**	**65.0 ± 1.9**	**59.1 ± 1.9**	**62.1 ± 2.0**	**52.5 ± 3.3**	57.9 ± 7.7	**59.6 ± 0.5**	**58.0 ± 4.0**
8	63.3 ± 3.4	61.2 ± 3.4	55.7 ± 3.3	55.2 ± 3.3	68.5 ± 2.0	61.7 ± 2.0	64.9 ± 2.1	53.1 ± 3.5	60.2 ± 5.8	68.8 ± 0.7	61.3 ± 5.4
9	66.1 ± 3.6	63.8 ± 3.6	57.2 ± 3.6	55.8 ± 3.5	70.0 ± 1.9	62.8 ± 2.0	66.6 ± 2.1	52.1 ± 3.3	68.8 ± 10.1	75.4 ± 0.8	63.9 ± 7.1
10	69.1 ± 3.9	66.9 ± 3.9	57.2 ± 3.8	54.8 ± 3.5	71.9 ± 1.9	63.8 ± 2.0	67.7 ± 2.1	52.6 ± 3.3	61.7 ± 5.1	82.9 ± 1.1	64.9 ± 9.0
11	70.7 ± 4.1	68.3 ± 4.2	56.4 ± 3.9	54.0 ± 3.5	73.7 ± 2.1	64.7 ± 2.1	69.0 ± 2.2	51.8 ± 3.2	63.7 ± 5.7	88.0 ± 1.3	66.0 ± 10.7
12	73.3 ± 4.5	71.3 ± 4.6	61.3 ± 4.5	56.3 ± 4.0	78.3 ± 2.3	67.4 ± 2.3	71.7 ± 2.4	54.8 ± 3.7	80.1 ± 13.7	92.6 ± 1.5	70.7 ± 11.5
13	74.1 ± 4.5	72.1 ± 4.6	62.5 ± 4.3	58.4 ± 3.9	79.2 ± 2.3	67.9 ± 2.3	72.6 ± 2.4	53.4 ± 3.3	78.1 ± 9.7	97.3 ± 1.7	71.6 ± 12.3
14	75.5 ± 4.6	73.5 ± 4.8	63.0 ± 4.6	58.0 ± 4.1	82.7 ± 2.5	69.8 ± 2.4	74.7 ± 2.5	54.9 ± 3.5	85.3 ± 10.4	98.6 ± 2.0	73.6 ± 13.2
15	75.2 ± 4.8	73.1 ± 4.9	63.2 ± 4.5	59.7 ± 4.1	81.7 ± 2.5	69.4 ± 2.5	74.3 ± 2.6	53.6 ± 3.3	71.7 ± 7.9	98.8 ± 2.1	72.1 ± 12.5
16	76.8 ± 5.1	74.4 ± 5.3	61.8 ± 4.8	56.8 ± 4.2	86.4 ± 2.8	71.7 ± 2.7	76.9 ± 2.9	55.4 ± 3.7	75.4 ± 8.4	99.2 ± 2.1	73.5 ± 13.3
17	77.2 ± 5.1	75.0 ± 5.3	64.7 ± 4.9	60.1 ± 4.4	86.6 ± 2.8	71.7 ± 2.7	76.9 ± 2.8	55.0 ± 3.5	74.4 ± 9.0	99.1 ± 2.5	74.1 ± 12.7

4	49.8 ± 2.4	48.7 ± 2.3	48.1 ± 2.5	52.6 ± 3.1	57.0 ± 2.1	52.8 ± 1.9	54.4 ± 1.9	54.5 ± 3.5	29.1 ± 4.5	31.2 ± 0.1	47.8 ± 9.7
4.25	50.5 ± 2.4	49.2 ± 2.3	47.9 ± 2.4	52.4 ± 2.9	57.7 ± 2.0	53.4 ± 1.9	55.3 ± 1.9	53.8 ± 3.4	30.7 ± 4.8	33.4 ± 0.2	48.4 ± 9.1
4.5	51.5 ± 2.5	50.2 ± 2.3	48.9 ± 2.5	53.6 ± 3.1	58.2 ± 2.0	54.0 ± 1.9	55.9 ± 1.9	53.6 ± 3.4	33.5 ± 5.0	35.9 ± 0.2	49.5 ± 8.3
4.75	52.6 ± 2.6	51.1 ± 2.5	49.1 ± 2.6	53.5 ± 3.1	59.1 ± 2.0	54.7 ± 1.9	56.8 ± 1.9	53.4 ± 3.4	34.5 ± 5.3	38.1 ± 0.2	50.3 ± 7.9
5	53.7 ± 2.6	52.1 ± 2.5	49.5 ± 2.5	53.1 ± 2.9	59.7 ± 1.9	55.2 ± 1.9	57.4 ± 1.9	53.3 ± 3.4	35.8 ± 5.6	40.5 ± 0.2	51.0 ± 7.4
5.25	54.6 ± 2.7	53.0 ± 2.5	50.1 ± 2.6	53.4 ± 2.9	60.7 ± 1.9	55.8 ± 1.8	58.1 ± 1.9	53.4 ± 3.4	40.4 ± 5.9	43.6 ± 0.3	52.3 ± 6.2
5.5	55.6 ± 2.7	53.9 ± 2.6	50.6 ± 2.6	53.4 ± 2.9	62.1 ± 2.0	56.9 ± 1.9	59.3 ± 1.9	53.4 ± 3.4	42.3 ± 6.1	45.4 ± 0.3	53.3 ± 6.0
5.75	55.7 ± 2.8	53.7 ± 2.6	50.5 ± 2.8	54.0 ± 3.3	61.8 ± 1.9	56.3 ± 1.9	58.9 ± 1.9	51.4 ± 3.2	46.5 ± 4.2	47.2 ± 0.3	53.6 ± 4.9
6	57.4 ± 2.9	55.4 ± 2.8	51.2 ± 2.8	53.3 ± 3.2	62.5 ± 1.9	57.3 ± 1.9	59.8 ± 1.9	53.1 ± 3.4	50.5 ± 4.3	50.4 ± 0.4	55.1 ± 4.1
6.25	57.4 ± 2.9	55.5 ± 2.8	51.2 ± 2.7	53.5 ± 2.9	63.0 ± 1.9	57.6 ± 1.9	60.2 ± 1.9	52.7 ± 3.3	50.5 ± 6.9	52.3 ± 0.4	55.4 ± 4.1
6.5	58.7 ± 3.0	56.8 ± 2.9	52.2 ± 2.8	53.5 ± 2.9	64.0 ± 1.9	58.5 ± 1.9	61.2 ± 1.9	53.1 ± 3.4	55.4 ± 7.2	53.7 ± 0.4	56.7 ± 3.9
6.75	**60.1 ± 3.1**	**58.0 ± 3.0**	**52.8 ± 2.9**	**53.8 ± 3.0**	**65.7 ± 1.9**	**59.5 ± 1.9**	**62.3 ± 2.0**	**54.0 ± 3.6**	**55.4 ± 7.5**	**57.0 ± 0.4**	**57.9 ± 4.1**
7	59.4 ± 3.1	57.3 ± 2.9	52.7 ± 2.8	54.2 ± 3.0	65.5 ± 1.9	59.9 ± 1.9	62.8 ± 2.0	53.0 ± 3.3	58.5 ± 7.7	59.6 ± 0.5	58.3 ± 4.1
7.25	60.6 ± 3.1	58.5 ± 3.0	53.2 ± 2.9	54.0 ± 3.0	66.3 ± 1.9	60.2 ± 1.9	63.3 ± 2.0	52.9 ± 3.4	58.2 ± 8.1	62.3 ± 0.5	58.9 ± 4.5
7.5	61.3 ± 3.2	59.2 ± 3.2	54.3 ± 3.0	54.9 ± 3.2	65.8 ± 1.9	59.7 ± 1.9	62.7 ± 2.0	52.5 ± 3.4	62.2 ± 8.3	64.1 ± 0.6	59.7 ± 4.5
7.75	62.2 ± 3.3	59.9 ± 3.2	54.1 ± 3.1	55.0 ± 3.3	66.2 ± 1.9	60.0 ± 1.9	63.2 ± 2.0	51.9 ± 3.2	60.9 ± 5.6	65.5 ± 0.6	59.9 ± 4.8
8	63.2 ± 3.4	61.1 ± 3.3	55.5 ± 3.3	55.1 ± 3.3	68.0 ± 1.9	61.5 ± 2.0	64.9 ± 2.1	52.9 ± 3.4	60.2 ± 5.8	68.8 ± 0.7	61.1 ± 5.4
8.25	63.7 ± 3.4	61.2 ± 3.4	54.5 ± 3.2	54.8 ± 3.2	67.7 ± 1.9	61.1 ± 1.9	64.4 ± 2.0	52.2 ± 3.3	58.9 ± 6.0	70.6 ± 0.7	60.9 ± 5.9
8.5	65.4 ± 3.5	63.3 ± 3.5	57.0 ± 3.5	55.4 ± 3.5	69.6 ± 1.9	62.2 ± 2.0	65.5 ± 2.0	53.6 ± 3.5	62.4 ± 6.2	70.9 ± 0.7	62.5 ± 5.8
8.75	65.2 ± 3.5	63.1 ± 3.5	57.1 ± 3.5	55.3 ± 3.5	69.6 ± 1.9	62.1 ± 2.0	65.7 ± 2.1	52.3 ± 3.4	66.3 ± 6.4	74.3 ± 0.8	63.1 ± 6.7
9	66.3 ± 3.7	63.9 ± 3.7	57.0 ± 3.6	55.7 ± 3.5	70.2 ± 2.0	62.5 ± 2.0	66.0 ± 2.1	52.3 ± 3.3	55.1 ± 5.4	75.4 ± 0.9	62.4 ± 7.4

Table 4
Summary of AFM analysis for variant 3: the optimal *l* for each set of step sizes is indicated in bold

l (nm)	*P* ± SEM (nm)									*P* ± SD (nm)	
	$\langle \cos\theta \rangle$	$\langle \theta^2 \rangle$	$\langle \theta^4 \rangle$	$\langle \theta^6 \rangle$	$\langle \mathbf{R}\cdot\mathbf{u}_0 \rangle$	$\langle R^2 \rangle$	$\langle R^4 \rangle$	$\langle R^4 \rangle - \langle R^2 \rangle^2$	$-\ln G(\theta)$	$\langle \cos\chi \rangle$	Average
3	43.3 ± 1.5	43.9 ± 1.4	46.6 ± 1.6	53.0 ± 2.1	48.6 ± 1.3	48.4 ± 1.3	49.3 ± 1.2	54.6 ± 2.7	25.5 ± 2.2	26.4 ± 0.1	44.0 ± 10.1
4	47.1 ± 1.7	47.1 ± 1.6	47.8 ± 1.7	52.4 ± 2.1	51.2 ± 1.2	50.8 ± 1.2	51.9 ± 1.2	53.5 ± 2.7	33.9 ± 4.4	38.0 ± 0.1	47.4 ± 6.5
5	51.4 ± 1.8	51.1 ± 1.8	50.4 ± 1.8	54.0 ± 2.2	53.9 ± 1.1	52.6 ± 1.2	54.4 ± 1.2	52.4 ± 2.5	45.1 ± 5.5	51.0 ± 0.2	51.6 ± 2.7
6	**56.0 ± 2.1**	**55.4 ± 2.0**	**54.0 ± 2.1**	**55.6 ± 2.3**	**57.9 ± 1.1**	**55.1 ± 1.2**	**56.9 ± 1.2**	**53.8 ± 2.8**	**57.6 ± 6.6**	**63.5 ± 0.4**	**56.6 ± 2.8**
7	58.1 ± 2.1	57.5 ± 2.1	55.8 ± 2.2	56.7 ± 2.4	59.8 ± 1.1	56.4 ± 1.2	58.5 ± 1.3	53.2 ± 2.6	63.5 ± 5.0	73.3 ± 0.5	59.3 ± 5.6
8	61.8 ± 2.4	61.2 ± 2.4	59.9 ± 2.6	60.1 ± 2.8	62.1 ± 1.1	57.2 ± 1.2	59.6 ± 1.3	52.7 ± 2.6	67.1 ± 5.8	80.5 ± 0.7	62.2 ± 7.4
9	63.5 ± 2.6	62.6 ± 2.6	56.6 ± 2.6	55.6 ± 2.6	63.5 ± 1.2	58.1 ± 1.3	60.7 ± 1.4	52.0 ± 2.6	67.0 ± 6.6	85.0 ± 0.8	62.5 ± 9.1
10	63.2 ± 2.7	62.1 ± 2.7	56.5 ± 2.6	55.6 ± 2.6	64.5 ± 1.3	58.6 ± 1.4	61.4 ± 1.4	51.8 ± 2.6	69.7 ± 7.3	88.8 ± 1.0	63.2 ± 10.3
11	63.8 ± 2.7	62.5 ± 2.8	55.5 ± 2.6	55.0 ± 2.6	65.9 ± 1.3	59.6 ± 1.4	62.5 ± 1.5	51.9 ± 2.6	63.3 ± 5.7	90.0 ± 1.1	63.0 ± 10.5
12	66.3 ± 3.0	64.7 ± 3.0	55.5 ± 2.8	54.2 ± 2.7	67.5 ± 1.4	60.4 ± 1.5	63.6 ± 1.5	51.9 ± 2.6	63.7 ± 6.2	91.4 ± 1.2	63.9 ± 11
13	65.7 ± 3.0	64.0 ± 3.0	56.1 ± 2.8	56.2 ± 2.8	68.5 ± 1.5	61.2 ± 1.5	64.6 ± 1.6	51.2 ± 2.5	66.1 ± 6.8	91.3 ± 1.3	64.5 ± 10.9
14	67.5 ± 3.1	66.5 ± 3.1	61.3 ± 3.1	58.9 ± 2.9	71.9 ± 1.5	63.3 ± 1.6	66.5 ± 1.7	54.4 ± 2.8	67.5 ± 7.3	92.4 ± 1.5	67.0 ± 10.2
15	66.9 ± 3.2	65.6 ± 3.2	60.4 ± 3.0	59.3 ± 2.9	72.5 ± 1.7	63.8 ± 1.8	67.3 ± 1.9	53.2 ± 2.6	69.9 ± 7.9	92.6 ± 1.5	67.2 ± 10.5
16	69.2 ± 3.4	68.1 ± 3.5	63.4 ± 3.4	61.9 ± 3.4	78.1 ± 1.9	66.7 ± 1.9	70.0 ± 2.0	56.0 ± 3.0	71.2 ± 8.5	92.0 ± 1.6	69.7 ± 9.8
17	68.7 ± 3.4	67.6 ± 3.4	62.6 ± 3.3	61.2 ± 3.2	77.7 ± 1.9	66.2 ± 1.9	69.7 ± 2.0	54.8 ± 2.8	72.3 ± 9.0	91.2 ± 1.7	69.2 ± 10

4	47.2 ± 1.7	47.2 ± 1.6	47.8 ± 1.7	52.4 ± 2.1	51.0 ± 1.1	50.6 ± 1.2	51.7 ± 1.2	53.4 ± 2.7	33.9 ± 4.4	38.0 ± 0.1	47.3 ± 6.5
4.25	48.4 ± 1.7	48.3 ± 1.6	48.4 ± 1.7	52.9 ± 2.1	51.9 ± 1.1	51.3 ± 1.2	52.8 ± 1.2	53.0 ± 2.6	39.1 ± 4.7	42.2 ± 0.2	48.8 ± 4.8
4.5	49.1 ± 1.7	48.8 ± 1.7	48.7 ± 1.8	53.5 ± 2.2	52.1 ± 1.1	51.4 ± 1.2	52.7 ± 1.2	52.7 ± 2.6	40.3 ± 5.0	44.8 ± 0.2	49.4 ± 4.1
4.75	50.2 ± 1.8	49.8 ± 1.7	49.3 ± 1.8	54.0 ± 2.3	52.9 ± 1.1	52.0 ± 1.2	53.5 ± 1.2	52.7 ± 2.6	43.9 ± 3.4	47.5 ± 0.2	50.6 ± 3.1
5	51.0 ± 1.8	50.6 ± 1.7	49.8 ± 1.9	53.8 ± 2.2	53.4 ± 1.1	52.1 ± 1.2	53.7 ± 1.2	52.0 ± 2.5	44.3 ± 5.5	50.7 ± 0.2	51.1 ± 2.8
5.25	52.3 ± 1.9	51.8 ± 1.8	50.5 ± 1.9	54.2 ± 2.3	54.4 ± 1.1	52.9 ± 1.2	54.7 ± 1.2	52.0 ± 2.5	47.7 ± 5.8	54.0 ± 0.3	52.4 ± 2.1
5.5	53.0 ± 1.9	52.4 ± 1.9	51.2 ± 1.9	54.2 ± 2.2	55.1 ± 1.1	53.4 ± 1.2	55.1 ± 1.2	52.4 ± 2.6	50.8 ± 3.9	56.8 ± 0.3	53.4 ± 1.9
5.75	**54.3 ± 2.0**	**53.7 ± 1.9**	**51.9 ± 2.0**	**54.2 ± 2.3**	**55.9 ± 1.1**	**53.7 ± 1.2**	**55.4 ± 1.2**	**52.5 ± 2.6**	**54.9 ± 6.3**	**60.8 ± 0.4**	**54.7 ± 2.5**
6	55.5 ± 2.0	54.8 ± 2.0	52.8 ± 2.1	54.3 ± 2.4	56.8 ± 1.1	54.6 ± 1.2	56.6 ± 1.2	52.8 ± 2.7	57.3 ± 6.6	63.3 ± 0.4	55.9 ± 3.0
6.25	55.7 ± 2.1	55.0 ± 2.0	52.6 ± 2.1	54.6 ± 2.4	56.8 ± 1.1	54.3 ± 1.2	56.3 ± 1.2	51.7 ± 2.5	55.9 ± 4.5	65.0 ± 0.4	55.8 ± 3.6
6.5	56.2 ± 2.1	55.4 ± 2.1	52.8 ± 2.2	54.3 ± 2.5	57.7 ± 1.1	54.8 ± 1.2	56.8 ± 1.3	52.1 ± 2.6	63.3 ± 7.1	67.3 ± 0.4	57.1 ± 4.7
6.75	57.5 ± 2.2	56.8 ± 2.2	53.4 ± 2.3	54.3 ± 2.5	58.9 ± 1.1	55.7 ± 1.2	57.7 ± 1.3	52.8 ± 2.7	60.8 ± 4.9	70.2 ± 0.5	57.8 ± 5.0
7	57.9 ± 2.2	57.1 ± 2.2	53.9 ± 2.2	54.8 ± 2.4	58.9 ± 1.1	55.7 ± 1.2	57.8 ± 1.2	52.4 ± 2.6	64.3 ± 5.0	73.0 ± 0.5	58.6 ± 6.0
7.25	58.2 ± 2.3	57.2 ± 2.2	53.0 ± 2.3	53.7 ± 2.4	59.7 ± 1.2	56.0 ± 1.2	58.2 ± 1.3	52.1 ± 2.6	68.1 ± 7.9	75.1 ± 0.6	59.1 ± 7.2
7.5	59.6 ± 2.3	58.8 ± 2.3	55.0 ± 2.4	55.0 ± 2.5	60.0 ± 1.1	56.1 ± 1.2	58.2 ± 1.3	52.0 ± 2.6	66.9 ± 8.3	76.2 ± 0.6	59.8 ± 7.0
7.75	60.5 ± 2.3	59.6 ± 2.3	55.1 ± 2.4	55.8 ± 2.5	60.2 ± 1.1	56.4 ± 1.2	58.7 ± 1.3	52.0 ± 2.5	62.6 ± 5.6	78.7 ± 0.6	60.0 ± 7.3
8	61.6 ± 2.4	60.9 ± 2.4	56.0 ± 2.6	55.1 ± 2.6	61.6 ± 1.1	57.2 ± 1.2	59.5 ± 1.3	52.9 ± 2.7	67.0 ± 5.8	80.6 ± 0.6	61.2 ± 7.9
8.25	61.8 ± 2.4	61.0 ± 2.5	56.6 ± 2.5	56.0 ± 2.6	61.7 ± 1.1	57.2 ± 1.2	59.6 ± 1.3	52.3 ± 2.6	63.8 ± 6.0	81.9 ± 0.7	61.2 ± 8.0
8.5	61.6 ± 2.5	60.8 ± 2.5	58.3 ± 2.6	57.7 ± 2.7	62.6 ± 1.2	57.5 ± 1.3	59.9 ± 1.3	52.1 ± 2.7	64.8 ± 6.2	82.8 ± 0.8	61.8 ± 8.1
8.75	63.0 ± 2.5	62.3 ± 2.6	59.8 ± 2.7	58.8 ± 2.9	63.1 ± 1.2	57.7 ± 1.3	60.2 ± 1.3	52.1 ± 2.6	68.9 ± 6.3	83.6 ± 0.8	62.9 ± 8.5
9	63.3 ± 2.6	62.5 ± 2.6	58.8 ± 2.7	57.7 ± 2.7	63.2 ± 1.2	57.8 ± 1.3	60.4 ± 1.3	52.3 ± 2.6	67.6 ± 6.5	85.2 ± 0.8	62.9 ± 8.9

Table 5
Summary of AFM analysis for variant 4: the optimal *l* for each set of step sizes is indicated in bold

l (nm)	*P* ± SEM (nm)									*P* ± SD (nm)	
	$\langle \cos\theta \rangle$	$\langle \theta^2 \rangle$	$\langle \theta^4 \rangle$	$\langle \theta^6 \rangle$	$\langle \mathbf{R}\cdot\mathbf{u}_0 \rangle$	$\langle R^2 \rangle$	$\langle R^4 \rangle$	$\langle R^4 \rangle - \langle R^2 \rangle^2$	$-\ln G(\theta)$	$\langle \cos\chi \rangle$	Average
3	47.4 ± 2.2	47.1 ± 2.1	48.1 ± 2.3	52.9 ± 2.8	54.7 ± 2.2	51.4 ± 2.0	53.5 ± 2.0	52.3 ± 3.3	26.5 ± 2.2	25.4 ± 0.1	45.9 ± 10.9
4	52.3 ± 2.5	51.2 ± 2.3	49.7 ± 2.4	52.5 ± 2.8	58.6 ± 1.9	54.6 ± 1.9	57.2 ± 1.9	51.6 ± 3.3	33.6 ± 4.4	37.6 ± 0.2	49.9 ± 8.1
5	56.1 ± 2.7	54.6 ± 2.6	51.8 ± 2.6	54.3 ± 2.9	61.3 ± 1.9	56.9 ± 1.8	60.0 ± 2.0	50.2 ± 3.1	44.2 ± 3.6	50.3 ± 0.3	54.0 ± 5.1
6	**59.6 ± 3.0**	**58.1 ± 2.9**	**54.4 ± 2.8**	**55.0 ± 2.9**	**64.5 ± 1.8**	**59.3 ± 1.9**	**62.9 ± 2.0**	**51.0 ± 3.3**	**56.5 ± 6.6**	**62.6 ± 0.5**	**58.4 ± 4.3**
7	62.9 ± 3.2	61.3 ± 3.1	56.7 ± 3.1	56.5 ± 3.2	66.8 ± 1.8	60.4 ± 1.8	64.3 ± 2.0	50.4 ± 3.1	59.8 ± 7.7	72.9 ± 0.7	61.2 ± 6.2
8	66.3 ± 3.5	64.6 ± 3.5	60.1 ± 3.5	59.2 ± 3.6	69.0 ± 1.8	61.8 ± 1.9	66.0 ± 2.0	50.1 ± 3.2	62.0 ± 5.8	80.5 ± 0.9	64.0 ± 7.8
9	69.2 ± 3.8	67.3 ± 3.8	60.8 ± 3.7	59.0 ± 3.6	71.5 ± 1.9	63.4 ± 1.9	67.9 ± 2.1	50.1 ± 3.1	63.0 ± 6.6	86.9 ± 1.1	65.9 ± 9.6
10	70.2 ± 4.0	68.3 ± 4.0	61.9 ± 3.8	60.8 ± 3.7	72.7 ± 2.0	64.1 ± 2.0	68.9 ± 2.2	50.0 ± 3.1	66.8 ± 7.4	90.1 ± 1.3	67.4 ± 10.2
11	72.1 ± 4.2	70.1 ± 4.2	61.8 ± 3.9	59.6 ± 3.7	76.5 ± 2.2	66.9 ± 2.2	71.7 ± 2.4	51.8 ± 3.2	61.5 ± 5.7	94.6 ± 1.6	68.7 ± 11.7
12	71.4 ± 4.3	69.3 ± 4.4	61.1 ± 4.1	58.8 ± 3.9	76.9 ± 2.3	66.5 ± 2.3	71.6 ± 2.5	50.4 ± 3.3	62.0 ± 6.3	93.3 ± 1.6	68.1 ± 11.7
13	72.2 ± 4.5	70.0 ± 4.5	61.9 ± 4.1	59.8 ± 3.8	77.0 ± 2.3	66.3 ± 2.3	71.5 ± 2.5	50.0 ± 3.1	62.8 ± 6.8	93.4 ± 1.8	68.5 ± 11.6
14	73.1 ± 4.6	70.9 ± 4.6	62.5 ± 4.2	60.0 ± 4.0	81.6 ± 2.6	69.2 ± 2.5	74.5 ± 2.7	52.0 ± 3.3	67.0 ± 7.3	93.8 ± 1.9	70.5 ± 11.6
15	72.5 ± 4.7	70.0 ± 4.6	61.4 ± 4.2	59.6 ± 3.9	81.4 ± 2.7	68.5 ± 2.6	74.2 ± 2.8	50.3 ± 3.1	65.4 ± 8.0	93.0 ± 2.0	69.6 ± 11.9
16	74.0 ± 4.8	71.7 ± 4.9	62.0 ± 4.5	58.7 ± 4.1	85.9 ± 2.9	71.3 ± 2.8	76.8 ± 2.9	52.6 ± 3.4	70.3 ± 8.5	95.4 ± 2.2	71.9 ± 12.6
17	72.5 ± 4.8	69.3 ± 4.7	58.7 ± 4.1	57.4 ± 3.9	85.3 ± 3.0	70.4 ± 2.8	76.1 ± 2.9	50.5 ± 3.1	67.5 ± 9.1	94.6 ± 2.3	70.2 ± 13.2

4	51.5 ± 2.8	50.8 ± 2.7	50.1 ± 2.8	53.5 ± 3.2	57.5 ± 2.2	53.5 ± 2.1	55.3 ± 2.1	53.6 ± 4.1	34.2 ± 4.4	37.7 ± 0.2	49.8 ± 7.6
4.25	52.5 ± 2.8	51.8 ± 2.7	50.8 ± 2.7	53.9 ± 3.1	58.1 ± 2.1	53.7 ± 2.0	55.5 ± 2.0	53.1 ± 4.0	36.8 ± 3.1	40.9 ± 0.2	50.7 ± 6.6
4.5	53.7 ± 3.0	52.9 ± 2.8	51.5 ± 2.8	54.6 ± 3.1	58.3 ± 2.0	54.0 ± 1.9	56.1 ± 2.0	52.7 ± 3.9	38.3 ± 5.0	44.5 ± 0.3	51.7 ± 5.9
4.75	54.2 ± 3.0	53.4 ± 2.8	52.0 ± 2.8	54.7 ± 3.1	58.9 ± 2.0	54.6 ± 1.9	56.7 ± 2.0	52.7 ± 3.8	41.7 ± 5.2	47.3 ± 0.3	52.6 ± 4.9
5	55.1 ± 3.1	54.1 ± 2.9	51.9 ± 2.9	54.2 ± 3.1	59.5 ± 2.0	55.2 ± 2.0	57.3 ± 2.0	52.5 ± 3.8	44.1 ± 5.5	50.1 ± 0.4	53.4 ± 4.2
5.25	56.0 ± 3.2	55.0 ± 3.0	52.6 ± 2.9	54.9 ± 3.1	60.4 ± 2.0	55.7 ± 2.0	58.2 ± 2.1	52.2 ± 3.8	47.8 ± 5.8	54.1 ± 0.4	54.7 ± 3.4
5.5	57.6 ± 3.2	56.4 ± 3.0	53.5 ± 3.0	54.9 ± 3.1	61.3 ± 1.9	56.4 ± 1.9	58.9 ± 2.0	52.6 ± 3.8	49.0 ± 6.1	56.6 ± 0.4	55.7 ± 3.5
5.75	**57.8 ± 3.3**	**56.6 ± 3.2**	**53.5 ± 3.0**	**55.0 ± 3.3**	**61.6 ± 2.0**	**56.4 ± 1.9**	**58.8 ± 2.0**	**52.4 ± 3.9**	**53.0 ± 6.3**	**59.6 ± 0.5**	**56.5 ± 3.0**
6	59.1 ± 3.3	58.0 ± 3.2	55.1 ± 3.1	56.1 ± 3.3	63.2 ± 2.0	57.5 ± 2.0	60.0 ± 2.1	53.3 ± 4.0	59.9 ± 6.5	62.9 ± 0.5	58.5 ± 3.2
6.25	59.0 ± 3.4	57.8 ± 3.3	54.7 ± 3.2	56.1 ± 3.4	62.7 ± 2.0	57.0 ± 2.0	59.6 ± 2.1	51.8 ± 3.8	63.4 ± 6.8	65.2 ± 0.6	58.7 ± 4.1
6.5	60.8 ± 3.6	59.6 ± 3.5	56.3 ± 3.3	57.2 ± 3.5	63.9 ± 2.0	57.9 ± 2.0	60.6 ± 2.1	52.2 ± 3.9	59.0 ± 7.1	67.5 ± 0.6	59.5 ± 4.2
6.75	61.7 ± 3.6	60.4 ± 3.5	56.5 ± 3.4	56.6 ± 3.6	65.7 ± 2.1	59.2 ± 2.0	62.0 ± 2.2	52.9 ± 4.0	62.1 ± 7.4	70.1 ± 0.7	60.7 ± 4.9
7	61.4 ± 3.6	60.2 ± 3.5	56.7 ± 3.3	57.3 ± 3.5	64.9 ± 2.0	58.5 ± 2.0	61.3 ± 2.1	52.4 ± 3.8	62.2 ± 7.7	72.2 ± 0.7	60.7 ± 5.3
7.25	62.8 ± 3.7	61.6 ± 3.6	57.1 ± 3.6	56.5 ± 3.7	65.9 ± 2.0	59.0 ± 2.0	61.9 ± 2.1	52.4 ± 3.9	56.7 ± 5.3	73.0 ± 0.8	60.7 ± 5.8
7.5	63.3 ± 3.8	62.1 ± 3.7	58.3 ± 3.7	57.9 ± 3.8	66.9 ± 2.0	59.5 ± 2.0	62.7 ± 2.2	52.5 ± 4.0	68.1 ± 8.2	75.9 ± 0.8	62.7 ± 6.5
7.75	63.7 ± 3.8	62.5 ± 3.8	60.2 ± 3.7	61.2 ± 3.9	66.2 ± 2.0	59.1 ± 2.0	62.4 ± 2.1	51.4 ± 3.7	69.7 ± 8.5	77.3 ± 0.8	63.4 ± 6.9
8	65.1 ± 4.0	63.8 ± 4.0	59.8 ± 3.8	59.2 ± 3.9	67.7 ± 2.1	60.1 ± 2.0	63.0 ± 2.2	52.5 ± 4.0	64.2 ± 8.9	79.6 ± 1.0	63.5 ± 7.0
8.25	65.4 ± 4.0	64.1 ± 3.9	59.6 ± 3.8	58.7 ± 3.9	67.7 ± 2.0	60.0 ± 2.0	63.2 ± 2.2	51.9 ± 3.7	65.1 ± 6.0	81.0 ± 1.0	63.7 ± 7.6
8.5	66.6 ± 4.1	65.3 ± 4.0	61.5 ± 3.9	60.9 ± 4.1	69.3 ± 2.1	61.4 ± 2.1	64.7 ± 2.2	52.7 ± 4.0	68.9 ± 9.5	83.6 ± 1.1	65.5 ± 8.0
8.75	67.9 ± 4.2	66.6 ± 4.2	60.7 ± 4.1	58.2 ± 4.0	69.1 ± 2.0	60.7 ± 2.0	63.9 ± 2.1	52.4 ± 3.9	68.4 ± 9.8	84.4 ± 1.1	65.2 ± 8.6
9	67.2 ± 4.1	65.9 ± 4.1	61.0 ± 4.0	60.1 ± 4.0	69.3 ± 2.1	61.2 ± 2.1	64.7 ± 2.2	52.2 ± 3.7	67.0 ± 6.6	86.8 ± 1.2	65.5 ± 8.9

Table 6
Summary of AFM analysis for variant 5: the optimal *l* for each set of step sizes is indicated in bold

l (nm)	*P* ± SEM (nm)									*P* ± SD (nm)	
	$\langle\cos\theta\rangle$	$\langle\theta^2\rangle$	$\langle\theta^4\rangle$	$\langle\theta^6\rangle$	$\langle \mathbf{R}\cdot\mathbf{u}_0\rangle$	$\langle R^2\rangle$	$\langle R^4\rangle$	$\langle R^4\rangle - \langle R^2\rangle^2$	$-\ln G(\theta)$	$\langle\cos\chi\rangle$	Average
3	45.4 ± 1.6	45.4 ± 1.5	47.6 ± 1.9	53.4 ± 2.4	52.7 ± 1.6	50.8 ± 1.6	52.3 ± 1.5	51.8 ± 2.5	22.8 ± 2.2	25.9 ± 0.1	44.8 ± 11.2
4	49.6 ± 1.9	48.9 ± 1.7	48.7 ± 1.9	53.4 ± 2.3	55.9 ± 1.5	54.0 ± 1.5	55.7 ± 1.5	52.8 ± 2.7	33.7 ± 2.9	37.2 ± 0.1	49.0 ± 7.7
5	**53.5 ± 2.1**	**52.3 ± 1.9**	**50.4 ± 2.0**	**54.1 ± 2.3**	**59.3 ± 1.4**	**56.2 ± 1.5**	**58.6 ± 1.5**	**51.6 ± 2.5**	**43.2 ± 3.6**	**47.7 ± 0.2**	**52.7 ± 4.9**
6	56.3 ± 2.2	54.8 ± 2.1	51.8 ± 2.2	54.3 ± 2.5	62.3 ± 1.4	58.3 ± 1.6	60.9 ± 1.6	51.8 ± 2.7	48.5 ± 4.3	56.5 ± 0.3	55.5 ± 4.3
7	59.3 ± 2.4	57.6 ± 2.3	53.3 ± 2.3	54.9 ± 2.5	64.1 ± 1.4	59.8 ± 1.6	62.8 ± 1.6	51.0 ± 2.5	55.2 ± 5.1	64.4 ± 0.5	58.3 ± 4.6
8	62.2 ± 2.6	60.2 ± 2.6	54.8 ± 2.5	54.8 ± 2.5	66.3 ± 1.5	61.3 ± 1.6	64.4 ± 1.7	51.4 ± 2.6	53.1 ± 4.1	69.9 ± 0.6	59.8 ± 6.1
9	63.6 ± 2.7	61.4 ± 2.7	54.8 ± 2.6	55.4 ± 2.7	68.4 ± 1.5	62.2 ± 1.6	65.6 ± 1.7	50.9 ± 2.5	55.5 ± 4.6	74.5 ± 0.7	61.2 ± 7.2
10	65.1 ± 2.9	62.8 ± 2.9	55.8 ± 2.8	55.8 ± 2.8	70.2 ± 1.6	63.6 ± 1.7	67.3 ± 1.8	51.1 ± 2.6	56.2 ± 5.2	76.9 ± 0.8	62.5 ± 7.9
11	66.5 ± 3.0	64.0 ± 3.0	56.4 ± 2.8	56.5 ± 2.8	73.4 ± 1.8	65.3 ± 1.8	69.3 ± 1.9	51.4 ± 2.5	56.1 ± 5.7	81.3 ± 1.0	64.0 ± 9.2
12	67.7 ± 3.3	65.0 ± 3.3	56.1 ± 3.0	55.7 ± 3.0	75.4 ± 1.9	66.3 ± 2.0	70.5 ± 2.0	50.9 ± 2.7	59.8 ± 6.3	82.9 ± 1.1	65.0 ± 9.8
13	67.7 ± 3.3	64.8 ± 3.3	57.0 ± 2.9	57.3 ± 2.8	75.4 ± 2.0	66.9 ± 2.0	71.3 ± 2.1	51.0 ± 2.6	60.5 ± 6.8	84.1 ± 1.2	65.6 ± 9.8
14	68.9 ± 3.5	65.7 ± 3.4	56.3 ± 3.0	56.2 ± 2.9	78.8 ± 2.1	68.4 ± 2.1	72.8 ± 2.2	51.0 ± 2.6	61.0 ± 7.4	84.9 ± 1.3	66.4 ± 10.6
15	69.4 ± 3.6	66.4 ± 3.5	58.1 ± 3.0	58.1 ± 2.9	79.0 ± 2.1	68.6 ± 2.2	73.6 ± 2.3	50.5 ± 2.5	63.3 ± 8.0	86.0 ± 1.4	67.3 ± 10.5
16	72.3 ± 4.0	69.0 ± 4.0	58.7 ± 3.5	57.5 ± 3.3	84.7 ± 2.5	72.0 ± 2.5	77.0 ± 2.6	52.7 ± 2.8	64.0 ± 8.6	87.6 ± 1.6	69.6 ± 11.6
17	72.2 ± 3.8	68.9 ± 3.8	58.6 ± 3.3	57.8 ± 3.1	84.8 ± 2.4	71.5 ± 2.3	77.0 ± 2.4	51.7 ± 2.6	68.0 ± 9.1	87.4 ± 1.6	69.8 ± 11.5

4	49.0 ± 1.9	48.1 ± 1.7	48.4 ± 2.1	53.4 ± 2.5	55.6 ± 1.5	53.1 ± 1.5	55.0 ± 1.5	50.4 ± 2.5	32.5 ± 2.9	37.1 ± 0.1	48.3 ± 7.7
4.25	50.0 ± 1.9	49.0 ± 1.8	48.3 ± 1.9	53.2 ± 2.4	56.1 ± 1.5	53.7 ± 1.5	55.7 ± 1.5	50.4 ± 2.5	34.6 ± 3.1	39.6 ± 0.2	49.1 ± 7.0
4.5	50.6 ± 1.9	49.5 ± 1.8	48.5 ± 1.9	53.5 ± 2.3	56.8 ± 1.4	54.4 ± 1.5	56.6 ± 1.5	50.3 ± 2.5	33.7 ± 2.3	42.2 ± 0.2	49.6 ± 7.1
4.75	51.6 ± 1.9	50.4 ± 1.8	48.9 ± 1.9	53.3 ± 2.3	57.6 ± 1.4	55.0 ± 1.5	57.3 ± 1.5	50.0 ± 2.5	39.5 ± 3.4	45.1 ± 0.2	50.9 ± 5.6
5	52.9 ± 2.0	51.4 ± 1.9	49.5 ± 2.0	53.7 ± 2.4	58.3 ± 1.4	55.3 ± 1.5	57.6 ± 1.5	50.1 ± 2.4	38.0 ± 2.5	47.5 ± 0.2	51.4 ± 5.9
5.25	**53.5 ± 2.1**	**52.1 ± 2.0**	**50.0 ± 2.1**	**54.3 ± 2.4**	**59.8 ± 1.5**	**56.5 ± 1.5**	**59.0 ± 1.6**	**50.3 ± 2.5**	**44.0 ± 3.8**	**49.8 ± 0.3**	**52.9 ± 4.7**
5.5	54.3 ± 2.1	52.7 ± 2.0	50.4 ± 2.2	54.4 ± 2.6	59.9 ± 1.4	56.6 ± 1.5	59.1 ± 1.5	50.1 ± 2.5	45.9 ± 4.0	52.1 ± 0.3	53.5 ± 4.3
5.75	55.0 ± 2.2	53.3 ± 2.1	50.2 ± 2.1	53.3 ± 2.3	60.2 ± 1.4	56.6 ± 1.5	59.3 ± 1.5	49.3 ± 2.4	45.6 ± 4.2	54.3 ± 0.3	53.7 ± 4.5
6	55.6 ± 2.2	53.8 ± 2.1	50.5 ± 2.2	53.5 ± 2.6	61.1 ± 1.4	57.3 ± 1.5	60.0 ± 1.6	49.8 ± 2.5	43.3 ± 3.0	56.2 ± 0.3	54.1 ± 5.3
6.25	56.3 ± 2.3	54.5 ± 2.2	51.1 ± 2.3	54.9 ± 2.6	61.5 ± 1.5	57.8 ± 1.6	60.6 ± 1.6	49.8 ± 2.5	49.6 ± 4.5	58.2 ± 0.4	55.4 ± 4.3
6.5	56.7 ± 2.3	54.8 ± 2.2	50.9 ± 2.3	54.2 ± 2.7	62.5 ± 1.5	58.4 ± 1.6	61.4 ± 1.6	49.4 ± 2.5	46.2 ± 3.3	60.1 ± 0.4	55.5 ± 5.4
6.75	57.7 ± 2.4	55.6 ± 2.3	50.9 ± 2.4	53.3 ± 2.7	62.9 ± 1.5	58.4 ± 1.6	61.3 ± 1.6	49.4 ± 2.5	51.8 ± 4.9	61.2 ± 0.4	56.3 ± 4.8
7	58.3 ± 2.4	56.2 ± 2.3	51.5 ± 2.3	53.9 ± 2.6	63.1 ± 1.4	58.8 ± 1.6	62.0 ± 1.6	49.0 ± 2.4	53.6 ± 5.1	63.8 ± 0.5	57.0 ± 5.0
7.25	59.1 ± 2.4	57.1 ± 2.4	51.5 ± 2.5	53.4 ± 2.7	64.3 ± 1.4	59.6 ± 1.6	62.5 ± 1.6	50.1 ± 2.6	53.4 ± 5.3	64.9 ± 0.5	57.6 ± 5.3
7.5	59.7 ± 2.5	57.7 ± 2.5	52.7 ± 2.4	54.3 ± 2.6	64.3 ± 1.5	59.4 ± 1.6	62.3 ± 1.6	49.6 ± 2.5	49.3 ± 3.8	66.2 ± 0.5	57.6 ± 5.9
7.75	61.1 ± 2.5	59.0 ± 2.5	53.3 ± 2.5	55.0 ± 2.6	65.0 ± 1.4	60.2 ± 1.6	63.5 ± 1.6	49.7 ± 2.5	48.3 ± 4.0	68.1 ± 0.6	58.3 ± 6.6
8	61.0 ± 2.7	58.7 ± 2.6	52.6 ± 2.6	53.6 ± 2.7	65.7 ± 1.5	60.3 ± 1.7	63.7 ± 1.7	49.2 ± 2.5	51.1 ± 4.1	69.2 ± 0.6	58.5 ± 6.7
8.25	61.7 ± 2.7	59.4 ± 2.6	53.2 ± 2.6	54.8 ± 2.7	66.8 ± 1.6	61.3 ± 1.7	64.8 ± 1.7	50.1 ± 2.5	51.6 ± 4.2	70.9 ± 0.6	59.4 ± 6.9
8.5	61.6 ± 2.7	59.3 ± 2.6	52.5 ± 2.6	53.0 ± 2.7	66.8 ± 1.5	60.9 ± 1.6	64.3 ± 1.7	49.2 ± 2.5	50.9 ± 4.4	71.1 ± 0.6	58.9 ± 7.3
8.75	61.9 ± 2.7	59.4 ± 2.7	53.0 ± 2.6	54.3 ± 2.8	67.6 ± 1.6	61.5 ± 1.7	65.1 ± 1.8	49.0 ± 2.5	56.2 ± 6.5	73.1 ± 0.7	60.1 ± 7.3
9	63.0 ± 2.8	60.6 ± 2.7	53.9 ± 2.6	54.5 ± 2.7	67.8 ± 1.6	61.4 ± 1.7	65.1 ± 1.7	49.3 ± 2.4	54.4 ± 4.6	74.1 ± 0.7	60.4 ± 7.5

Table 7
Summary of AFM analysis for variant 6. The optimal *l* for each set of step sizes is indicated in bold

l (nm)	*P* ± SEM (nm)									*P* ± SD (nm)	
	$\langle \cos\theta \rangle$	$\langle \theta^2 \rangle$	$\langle \theta^4 \rangle$	$\langle \theta^6 \rangle$	$\langle \mathbf{R}\cdot\mathbf{u}_0 \rangle$	$\langle R^2 \rangle$	$\langle R^4 \rangle$	$\langle R^4 \rangle - \langle R^2 \rangle^2$	$-\ln G(\theta)$	$\langle \cos\chi \rangle$	Average
3	37.1 ± 1.5	38.2 ± 1.4	41.4 ± 1.5	47.8 ± 1.9	41.5 ± 1.1	40.4 ± 1.1	40.4 ± 1.1	50.4 ± 2.9	24.6 ± 2.2	24.6 ± 0.1	38.6 ± 8.4
4	40.7 ± 1.6	41.3 ± 1.5	43.3 ± 1.6	48.1 ± 2.0	44.0 ± 1.1	42.4 ± 1.1	42.6 ± 1.1	49.4 ± 2.9	32.5 ± 4.4	35.4 ± 0.2	42.0 ± 5.1
5	43.8 ± 1.8	44.1 ± 1.7	45.4 ± 1.8	49.7 ± 2.1	45.5 ± 1.0	43.8 ± 1.1	44.2 ± 1.1	48.7 ± 2.8	41.1 ± 5.5	45.9 ± 0.2	45.2 ± 2.5
6	**46.4 ± 1.9**	**46.4 ± 1.9**	**46.7 ± 2.0**	**49.6 ± 2.2**	**47.6 ± 1.0**	**45.2 ± 1.1**	**45.8 ± 1.1**	**48.4 ± 2.8**	**52.3 ± 6.6**	**55.0 ± 0.4**	**48.3 ± 3.1**
7	48.5 ± 2.1	48.4 ± 2.0	49.1 ± 2.1	51.5 ± 2.3	48.8 ± 1.0	46.3 ± 1.1	47.2 ± 1.1	48.1 ± 2.8	62.7 ± 7.7	63.0 ± 0.5	51.4 ± 6.2
8	50.7 ± 2.2	50.6 ± 2.2	52.0 ± 2.3	53.6 ± 2.5	50.2 ± 1.0	47.3 ± 1.1	48.2 ± 1.2	47.9 ± 2.8	63.9 ± 8.9	68.0 ± 0.6	53.2 ± 7.0
9	51.3 ± 2.3	51.2 ± 2.3	51.9 ± 2.3	54.0 ± 2.5	50.9 ± 1.0	47.8 ± 1.1	48.8 ± 1.2	48.1 ± 2.7	64.0 ± 6.6	70.8 ± 0.7	53.9 ± 7.5
10	52.5 ± 2.5	52.4 ± 2.4	51.9 ± 2.4	52.9 ± 2.5	52.0 ± 1.1	48.4 ± 1.2	49.4 ± 1.2	47.7 ± 2.7	59.4 ± 5.1	74.2 ± 0.8	54.1 ± 7.8
11	52.9 ± 2.5	52.9 ± 2.5	53.8 ± 2.6	54.9 ± 2.7	53.2 ± 1.1	49.0 ± 1.2	50.1 ± 1.3	47.7 ± 2.7	65.7 ± 8.2	74.4 ± 0.9	55.4 ± 8.3
12	53.0 ± 2.6	52.9 ± 2.6	53.3 ± 2.7	54.3 ± 2.8	54.8 ± 1.2	49.9 ± 1.4	51.2 ± 1.4	47.9 ± 2.9	62.3 ± 6.2	74.3 ± 1.0	55.4 ± 7.6
13	52.7 ± 2.7	52.5 ± 2.6	52.0 ± 2.6	53.9 ± 2.6	54.1 ± 1.3	49.5 ± 1.4	51.0 ± 1.4	46.8 ± 2.7	60.3 ± 6.8	73.8 ± 1.1	54.7 ± 7.6
14	53.3 ± 2.7	53.2 ± 2.7	53.1 ± 2.7	53.9 ± 2.7	57.1 ± 1.4	51.1 ± 1.5	52.5 ± 1.5	47.7 ± 2.8	64.0 ± 7.4	74.2 ± 1.2	56.0 ± 7.7
15	53.2 ± 2.8	52.9 ± 2.7	52.3 ± 2.6	54.0 ± 2.6	56.5 ± 1.4	50.5 ± 1.5	52.0 ± 1.5	47.1 ± 2.7	62.9 ± 8.0	73.9 ± 1.3	55.5 ± 7.7
16	53.3 ± 3.0	53.1 ± 2.9	52.4 ± 2.8	53.8 ± 2.9	60.0 ± 1.6	51.9 ± 1.7	53.3 ± 1.7	47.8 ± 3.0	61.9 ± 8.6	72.4 ± 1.3	56.0 ± 7.0
17	53.2 ± 3.0	53.0 ± 2.9	52.9 ± 2.9	54.3 ± 2.9	59.5 ± 1.6	51.7 ± 1.6	53.4 ± 1.7	47.6 ± 2.8	60.3 ± 9.3	71.4 ± 1.3	55.7 ± 6.6

4	40.9 ± 1.7	41.5 ± 1.6	43.3 ± 1.6	47.9 ± 2.0	44.0 ± 1.1	42.5 ± 1.1	42.6 ± 1.1	49.4 ± 2.9	32.5 ± 4.4	35.4 ± 0.2	42.0 ± 5.1
4.25	41.8 ± 1.7	42.3 ± 1.6	44.4 ± 1.6	48.9 ± 2.0	44.4 ± 1.0	42.8 ± 1.0	42.9 ± 1.0	49.4 ± 2.8	35.0 ± 4.7	37.7 ± 0.2	43.0 ± 4.4
4.5	42.5 ± 1.7	42.9 ± 1.6	44.6 ± 1.7	49.3 ± 2.0	44.8 ± 1.0	43.2 ± 1.1	43.4 ± 1.1	49.3 ± 2.8	36.9 ± 5.0	40.0 ± 0.2	43.7 ± 3.8
4.75	43.1 ± 1.7	43.4 ± 1.7	44.9 ± 1.7	49.3 ± 2.0	45.2 ± 1.0	43.5 ± 1.1	43.8 ± 1.1	48.9 ± 2.8	37.9 ± 5.3	42.3 ± 0.2	44.2 ± 3.2
5	43.9 ± 1.8	44.2 ± 1.7	45.3 ± 1.8	49.6 ± 2.1	45.7 ± 1.0	44.0 ± 1.1	44.4 ± 1.1	48.8 ± 2.8	41.2 ± 5.5	45.9 ± 0.2	45.3 ± 2.5
5.25	44.5 ± 1.8	44.8 ± 1.7	46.0 ± 1.8	49.8 ± 2.1	45.7 ± 1.0	44.1 ± 1.0	44.4 ± 1.0	48.6 ± 2.8	42.6 ± 5.8	48.1 ± 0.3	45.9 ± 2.3
5.5	**45.5 ± 1.9**	**45.6 ± 1.8**	**46.4 ± 1.9**	**49.9 ± 2.1**	**46.6 ± 1.0**	**44.7 ± 1.1**	**45.2 ± 1.1**	**48.5 ± 2.8**	**49.0 ± 6.1**	**51.2 ± 0.3**	**47.3 ± 2.2**
5.75	45.9 ± 1.9	46.0 ± 1.8	47.1 ± 1.9	50.4 ± 2.1	47.1 ± 1.0	45.1 ± 1.1	45.6 ± 1.1	48.7 ± 2.8	51.2 ± 6.3	53.4 ± 0.3	48.1 ± 2.8
6	46.5 ± 2.0	46.5 ± 1.9	46.9 ± 1.9	49.9 ± 2.2	47.6 ± 1.0	45.4 ± 1.1	45.9 ± 1.1	48.7 ± 2.9	52.3 ± 6.6	55.0 ± 0.4	48.5 ± 3.1
6.25	47.1 ± 2.0	47.0 ± 1.9	47.4 ± 2.0	50.5 ± 2.2	47.1 ± 1.0	45.2 ± 1.1	45.7 ± 1.1	48.2 ± 2.8	53.2 ± 4.5	57.4 ± 0.4	48.9 ± 3.8
6.5	48.2 ± 2.1	48.2 ± 2.0	49.2 ± 2.1	52.1 ± 2.3	48.3 ± 1.0	45.9 ± 1.1	46.6 ± 1.1	48.3 ± 2.9	57.0 ± 4.7	60.2 ± 0.4	50.4 ± 4.7
6.75	49.0 ± 2.1	49.0 ± 2.1	49.6 ± 2.1	51.5 ± 2.3	48.8 ± 1.0	46.3 ± 1.1	47.0 ± 1.1	48.5 ± 2.9	60.6 ± 7.4	61.7 ± 0.4	51.2 ± 5.4
7	48.1 ± 2.1	48.1 ± 2.0	49.1 ± 2.1	51.5 ± 2.3	48.3 ± 1.0	46.1 ± 1.1	46.8 ± 1.1	48.0 ± 2.8	62.7 ± 7.7	63.1 ± 0.5	51.2 ± 6.3
7.25	49.6 ± 2.1	49.6 ± 2.1	50.6 ± 2.2	53.0 ± 2.4	49.2 ± 1.0	46.3 ± 1.1	47.0 ± 1.1	48.2 ± 2.9	59.1 ± 5.2	64.1 ± 0.5	51.7 ± 5.7
7.5	49.8 ± 2.2	49.8 ± 2.2	50.2 ± 2.2	52.4 ± 2.4	49.3 ± 1.0	46.8 ± 1.1	47.6 ± 1.1	48.1 ± 2.8	57.1 ± 5.4	65.9 ± 0.6	51.7 ± 5.8
7.75	49.8 ± 2.2	49.7 ± 2.1	51.1 ± 2.2	53.6 ± 2.4	48.9 ± 1.0	46.6 ± 1.1	47.4 ± 1.1	48.2 ± 2.7	62.7 ± 5.6	67.1 ± 0.6	52.5 ± 6.9
8	50.9 ± 2.3	50.8 ± 2.2	52.0 ± 2.3	53.7 ± 2.5	50.1 ± 1.0	47.2 ± 1.1	48.0 ± 1.2	48.0 ± 2.8	63.9 ± 8.9	68.0 ± 0.6	53.3 ± 7.0
8.25	50.4 ± 2.2	50.3 ± 2.2	50.7 ± 2.3	52.7 ± 2.4	50.1 ± 1.0	47.1 ± 1.1	47.9 ± 1.1	48.1 ± 2.7	63.5 ± 6.0	69.0 ± 0.6	53.0 ± 7.3
8.5	51.9 ± 2.3	51.7 ± 2.3	51.9 ± 2.4	53.0 ± 2.5	51.4 ± 1.0	48.0 ± 1.2	49.0 ± 1.2	48.2 ± 2.9	59.7 ± 6.2	70.8 ± 0.7	53.6 ± 6.9
8.75	51.3 ± 2.3	51.2 ± 2.3	52.3 ± 2.4	54.5 ± 2.6	51.4 ± 1.1	47.9 ± 1.2	48.9 ± 1.2	47.9 ± 2.8	63.4 ± 6.4	70.7 ± 0.7	54.0 ± 7.4
9	51.4 ± 2.3	51.3 ± 2.3	51.8 ± 2.4	54.0 ± 2.5	51.0 ± 1.0	48.0 ± 1.2	49.0 ± 1.2	48.0 ± 2.8	64.0 ± 6.6	70.8 ± 0.7	53.9 ± 7.5

Table 8
Summary of AFM analysis for variant 7: the optimal *l* for each set of step sizes is indicated in bold

l (nm)	*P* ± SEM (nm)									*P* ± SD (nm)	
	$\langle \cos\theta \rangle$	$\langle \theta^2 \rangle$	$\langle \theta^4 \rangle$	$\langle \theta^6 \rangle$	$\langle \mathbf{R}\cdot\mathbf{u}_0 \rangle$	$\langle R^2 \rangle$	$\langle R^4 \rangle$	$\langle R^4 \rangle - \langle R^2 \rangle^2$	$-\ln G(\theta)$	$\langle \cos\chi \rangle$	Average
3	56.7 ± 2.4	57.2 ± 2.2	59.7 ± 2.6	64.5 ± 3.5	66.8 ± 2.7	63.4 ± 2.4	63.6 ± 2.3	80.8 ± 5.5	26.7 ± 2.2	23.5 ± 0.1	56.3 ± 17.8
4	62.8 ± 2.7	62.5 ± 2.5	62.4 ± 2.9	64.9 ± 3.6	72.1 ± 2.4	68.2 ± 2.3	69.0 ± 2.2	79.0 ± 5.5	34.2 ± 4.4	34.7 ± 0.1	61.0 ± 14.9
5	66.9 ± 2.9	66.5 ± 2.8	65.8 ± 3.0	68.0 ± 3.7	76.7 ± 2.3	73.2 ± 2.3	74.5 ± 2.3	79.2 ± 5.3	47.8 ± 5.5	47.5 ± 0.2	66.6 ± 11
6	72.2 ± 3.2	71.4 ± 3.1	69.4 ± 3.4	69.7 ± 4.0	79.9 ± 2.2	75.8 ± 2.3	77.4 ± 2.3	78.3 ± 5.5	61.0 ± 6.5	61.7 ± 0.4	71.7 ± 6.5
7	76.9 ± 3.5	76.1 ± 3.4	73.2 ± 3.5	72.7 ± 3.9	84.2 ± 2.2	79.3 ± 2.3	81.2 ± 2.3	78.9 ± 5.4	69.2 ± 12.8	75.7 ± 0.5	76.7 ± 4.4
8	**81.6 ± 3.8**	**80.8 ± 3.8**	**77.4 ± 4.2**	**75.4 ± 5.0**	**86.7 ± 2.1**	**81.1 ± 2.3**	**83.3 ± 2.3**	**78.8 ± 5.5**	**87.3 ± 8.7**	**87.4 ± 0.8**	**82.0 ± 4.2**
9	83.7 ± 3.9	82.6 ± 4.0	77.9 ± 4.3	76.0 ± 4.9	88.6 ± 2.1	83.4 ± 2.4	85.8 ± 2.4	77.8 ± 5.1	94.5 ± 9.8	99.8 ± 1.0	85.0 ± 7.6
10	86.6 ± 4.3	85.5 ± 4.4	81.2 ± 4.8	80.1 ± 5.4	91.5 ± 2.3	85.4 ± 2.5	88.0 ± 2.6	78.0 ± 5.1	99.3 ± 10.9	107.7 ± 1.1	88.3 ± 9.1
11	90.4 ± 4.5	89.5 ± 4.6	84.7 ± 5.2	81.4 ± 5.9	94.5 ± 2.3	87.3 ± 2.5	90.4 ± 2.6	78.7 ± 5.3	101.5 ± 12.1	115.8 ± 1.5	91.4 ± 10.7
12	93.6 ± 4.9	92.7 ± 5.1	87.3 ± 5.9	83.1 ± 6.8	99.7 ± 2.5	90.5 ± 2.8	93.7 ± 2.8	79.5 ± 5.5	101.2 ± 13.3	121 ± 1.6	94.2 ± 11.5
13	93.0 ± 5.0	92.1 ± 5.1	85.0 ± 5.7	80.9 ± 6.1	98.9 ± 2.6	89.4 ± 2.7	92.7 ± 2.8	77.7 ± 5.0	102.3 ± 14.5	124.4 ± 1.8	93.6 ± 13.2
14	95.4 ± 5.2	94.7 ± 5.4	88.4 ± 6.0	83.4 ± 6.6	103.5 ± 2.8	91.9 ± 2.9	95.3 ± 3.0	79.7 ± 5.3	103.9 ± 10.2	125.2 ± 2.0	96.2 ± 12.8
15	97.3 ± 5.3	96.3 ± 5.5	87.9 ± 6.1	82.0 ± 6.4	104.9 ± 2.7	92.4 ± 2.9	96.3 ± 3.0	78.6 ± 5.0	113.2 ± 16.8	128.3 ± 2.3	97.7 ± 14.8
16	98.1 ± 5.6	97.2 ± 5.9	88.7 ± 6.6	81.6 ± 7.1	109.2 ± 3.0	94.9 ± 3.2	98.9 ± 3.3	79.0 ± 5.4	115.4 ± 18	128.5 ± 2.4	99.2 ± 15.2
17	99.6 ± 5.8	98.9 ± 6.0	94.9 ± 6.6	94.2 ± 7.7	110.2 ± 3.1	95.5 ± 3.2	99.5 ± 3.3	81.0 ± 5.4	105.3 ± 12.6	129.5 ± 2.6	100.8 ± 12.6

4	62.7 ± 2.7	62.5 ± 2.6	62.6 ± 2.9	65.1 ± 3.6	72.6 ± 2.5	68.6 ± 2.3	69.5 ± 2.3	79.5 ± 5.5	34.2 ± 4.4	34.7 ± 0.1	61.2 ± 15.1
4.25	64.2 ± 2.7	63.8 ± 2.6	63.5 ± 2.9	65.9 ± 3.6	73.3 ± 2.3	69.6 ± 2.3	70.6 ± 2.2	79.0 ± 5.4	36.2 ± 4.7	37.8 ± 0.2	62.4 ± 14.2
4.5	65.1 ± 2.8	64.7 ± 2.7	64.6 ± 3.0	67.1 ± 3.7	74.9 ± 2.4	71.1 ± 2.3	72.2 ± 2.3	79.2 ± 5.3	41.3 ± 4.9	41.2 ± 0.2	64.1 ± 13
4.75	66.3 ± 2.8	65.9 ± 2.7	65.2 ± 2.9	67.4 ± 3.6	76.0 ± 2.3	72.2 ± 2.3	73.5 ± 2.3	79.2 ± 5.3	44.2 ± 5.2	44.1 ± 0.2	65.4 ± 12.1
5	67.0 ± 2.9	66.5 ± 2.8	65.8 ± 3.0	67.6 ± 3.6	76.2 ± 2.2	72.4 ± 2.3	73.8 ± 2.3	78.5 ± 5.2	47.8 ± 5.5	47.5 ± 0.2	66.3 ± 10.7
5.25	68.8 ± 3.0	68.1 ± 2.9	66.6 ± 3.1	68.2 ± 3.8	77.5 ± 2.3	73.8 ± 2.3	75.2 ± 2.3	78.9 ± 5.3	51.7 ± 5.7	50.7 ± 0.3	68.0 ± 9.8
5.5	70.0 ± 3.0	69.3 ± 2.9	67.6 ± 3.2	68.7 ± 3.8	78.2 ± 2.2	74.8 ± 2.3	76.2 ± 2.3	79.4 ± 5.4	54.2 ± 6.0	55.0 ± 0.3	69.3 ± 8.8
5.75	71.8 ± 3.1	70.9 ± 3.1	68.5 ± 3.3	69.1 ± 3.9	79.4 ± 2.2	75.4 ± 2.2	76.9 ± 2.2	78.8 ± 5.3	56.7 ± 6.3	58.3 ± 0.4	70.6 ± 7.9
6	72.8 ± 3.2	72.0 ± 3.2	69.8 ± 3.4	70.1 ± 4.0	80.3 ± 2.2	76.3 ± 2.3	78.0 ± 2.3	78.5 ± 5.4	61.0 ± 6.5	61.7 ± 0.4	72.1 ± 6.7
6.25	73.6 ± 3.3	72.8 ± 3.2	70.8 ± 3.5	71.1 ± 4.0	80.6 ± 2.1	76.4 ± 2.3	78.0 ± 2.3	78.3 ± 5.3	65.8 ± 6.8	65.4 ± 0.4	73.3 ± 5.2
6.5	74.6 ± 3.3	73.7 ± 3.3	70.9 ± 3.5	70.6 ± 4.1	82.0 ± 2.2	77.3 ± 2.3	79.1 ± 2.3	78.0 ± 5.3	68.9 ± 7.1	68.6 ± 0.5	74.4 ± 4.6
6.75	75.9 ± 3.4	75.1 ± 3.3	72.1 ± 3.7	71.1 ± 4.2	84.0 ± 2.2	79.0 ± 2.3	80.9 ± 2.4	79.2 ± 5.6	65.3 ± 12.4	72.4 ± 0.5	75.5 ± 5.5
7	76.8 ± 3.4	76.0 ± 3.3	73.1 ± 3.5	72.7 ± 3.9	83.7 ± 2.1	79.2 ± 2.3	81.2 ± 2.3	78.6 ± 5.3	80.5 ± 7.6	75.7 ± 0.6	77.7 ± 3.5
7.25	78.3 ± 3.6	77.4 ± 3.5	73.9 ± 3.8	73.0 ± 4.2	85.1 ± 2.1	79.8 ± 2.3	81.8 ± 2.4	78.7 ± 5.6	78.2 ± 7.9	79.1 ± 0.6	78.5 ± 3.5
7.5	78.7 ± 3.6	77.6 ± 3.6	72.8 ± 3.9	71.2 ± 4.4	85.8 ± 2.2	80.2 ± 2.3	82.3 ± 2.4	77.6 ± 5.2	73.3 ± 13.8	81.9 ± 0.7	78.1 ± 4.7
7.75	**79.6 ± 3.7**	**78.7 ± 3.7**	**75.2 ± 3.9**	**74.3 ± 4.5**	**85.8 ± 2.1**	**80.5 ± 2.3**	**82.7 ± 2.4**	**78.2 ± 5.2**	**81.0 ± 8.4**	**84.9 ± 0.7**	**80.1 ± 3.8**
8	81.5 ± 3.8	80.7 ± 3.8	77.7 ± 4.2	75.9 ± 5.0	87.3 ± 2.2	81.6 ± 2.3	83.8 ± 2.4	79.0 ± 5.4	79.0 ± 14.7	87.3 ± 0.7	81.4 ± 3.8
8.25	81.6 ± 3.8	80.5 ± 3.8	75.5 ± 4.4	73.0 ± 5.1	88.2 ± 2.2	82.7 ± 2.4	85.1 ± 2.4	78.7 ± 5.3	87.5 ± 9.0	91.9 ± 0.8	82.5 ± 5.9
8.5	82.7 ± 3.9	81.7 ± 3.9	78.3 ± 4.3	77.2 ± 5.0	89.4 ± 2.2	83.6 ± 2.4	85.8 ± 2.4	79.3 ± 5.6	86.7 ± 9.3	93.5 ± 0.9	83.8 ± 5.1
8.75	83.5 ± 3.9	82.5 ± 4.0	78.9 ± 4.5	76.7 ± 5.1	89.1 ± 2.2	83.6 ± 2.4	86.2 ± 2.5	78.5 ± 5.3	88.2 ± 9.5	95.6 ± 0.9	84.3 ± 5.7
9	83.9 ± 4.0	82.8 ± 4.0	78.2 ± 4.4	76.2 ± 4.9	88.7 ± 2.2	83.5 ± 2.4	86.1 ± 2.4	78.0 ± 5.2	92.8 ± 16.4	99.9 ± 1.0	85.0 ± 7.3

Table 9
Comparison of AFM analysis for each DNA incorporating a thymidine variant at the optimal *l* for each set of step sizes

Thymidine variant – *l* (nm)	*P* ± SEM (nm)									*P* ± SD (nm)	
	$\langle\cos\theta\rangle$	$\langle\theta^2\rangle$	$\langle\theta^4\rangle$	$\langle\theta^6\rangle$	$\langle R\cdot u_0\rangle$	$\langle R^2\rangle$	$\langle R^4\rangle$	$\langle R^4\rangle - \langle R^2\rangle^2$	$-\ln G(\theta)$	$\langle\cos\chi\rangle$	Average
Variant 1–6	54.9 ± 2.4	54.8 ± 2.3	55.2 ± 2.4	56.7 ± 2.7	54.0 ± 1.1	51.2 ± 1.2	52.8 ± 1.3	51.3 ± 3.1	60.5 ± 6.5	65.2 ± 0.5	55.7 ± 4.3
Variant 1–5.5	**53.5 ± 2.3**	**53.3 ± 2.2**	**53.1 ± 2.3**	**54.5 ± 2.5**	**52.8 ± 1.1**	**50.3 ± 1.2**	**51.8 ± 1.2**	**51.1 ± 3.0**	**54.7 ± 6.0**	**58.2 ± 0.3**	**53.3 ± 2.2**
Variant 2–7	59.5 ± 3.1	57.4 ± 3.0	52.6 ± 2.9	54.0 ± 3.0	65.0 ± 1.9	59.1 ± 1.9	62.1 ± 2.0	52.5 ± 3.3	57.9 ± 7.7	59.6 ± 0.5	58.0 ± 4.0
Variant 2–6.75	**60.1 ± 3.1**	**58.0 ± 3.0**	**52.8 ± 2.9**	**53.8 ± 3.0**	**65.7 ± 1.9**	**59.5 ± 1.9**	**62.3 ± 2.0**	**54.0 ± 3.6**	**55.4 ± 7.5**	**57.0 ± 0.4**	**57.9 ± 4.1**
Variant 3–6	56.0 ± 2.1	55.4 ± 2.0	54.0 ± 2.1	55.6 ± 2.3	57.9 ± 1.1	55.1 ± 1.2	56.9 ± 1.2	53.8 ± 2.8	57.6 ± 6.6	63.5 ± 0.4	56.6 ± 2.8
Variant 3–5.75	**54.3 ± 2.0**	**53.7 ± 1.9**	**51.9 ± 2.0**	**54.2 ± 2.3**	**55.9 ± 1.1**	**53.7 ± 1.2**	**55.4 ± 1.2**	**52.5 ± 2.6**	**54.9 ± 6.3**	**60.8 ± 0.4**	**54.7 ± 2.5**
Variant 4–6	59.6 ± 3.0	58.1 ± 2.9	54.4 ± 2.8	55.0 ± 2.9	64.5 ± 1.8	59.3 ± 1.9	62.9 ± 2.0	51.0 ± 3.3	56.5 ± 6.6	62.6 ± 0.5	58.4 ± 4.3
Variant 4–5.75	**57.8 ± 3.3**	**56.6 ± 3.2**	**53.5 ± 3.0**	**55.0 ± 3.3**	**61.6 ± 2.0**	**56.4 ± 1.9**	**58.8 ± 2.0**	**52.4 ± 3.9**	**53.0 ± 6.3**	**59.6 ± 0.5**	**56.5 ± 3.0**
Variant 5–5	53.5 ± 2.1	52.3 ± 1.9	50.4 ± 2.0	54.1 ± 2.3	59.3 ± 1.4	56.2 ± 1.5	58.6 ± 1.5	51.6 ± 2.5	43.2 ± 3.6	47.7 ± 0.2	52.7 ± 4.9
Variant 5–5.25	**53.5 ± 2.1**	**52.1 ± 2.0**	**50.0 ± 2.1**	**54.3 ± 2.4**	**59.8 ± 1.5**	**56.5 ± 1.5**	**59.0 ± 1.6**	**50.3 ± 2.5**	**44.0 ± 3.8**	**49.8 ± 0.3**	**52.9 ± 4.7**
Variant 6–6	46.4 ± 1.9	46.4 ± 1.9	46.7 ± 2.0	49.6 ± 2.2	47.6 ± 1.0	45.2 ± 1.1	45.8 ± 1.1	48.4 ± 2.8	52.3 ± 6.6	55.0 ± 0.4	48.3 ± 3.1
Variant 6–5.5	**45.5 ± 1.9**	**45.6 ± 1.8**	**46.4 ± 1.9**	**49.9 ± 2.1**	**46.6 ± 1.0**	**44.7 ± 1.1**	**45.2 ± 1.1**	**48.5 ± 2.8**	**49.0 ± 6.1**	**51.2 ± 0.3**	**47.3 ± 2.2**
Variant 7–8	81.6 ± 3.8	80.8 ± 3.8	77.4 ± 4.2	75.4 ± 5.0	86.7 ± 2.1	81.1 ± 2.3	83.3 ± 2.3	78.8 ± 5.5	87.3 ± 8.7	87.4 ± 0.8	82.0 ± 4.2
Variant 7–7.75	**79.6 ± 3.7**	**78.7 ± 3.7**	**75.2 ± 3.9**	**74.3 ± 4.5**	**85.8 ± 2.1**	**80.5 ± 2.3**	**82.7 ± 2.4**	**78.2 ± 5.2**	**81.0 ± 8.4**	**84.9 ± 0.7**	**80.1 ± 3.8**

12. We perform image analysis using custom software written in R (version 3.3.2).
13. Only DNA molecules satisfying the following criteria were included in the data set for analysis: (1) both ends were visible and the molecule was not in touch with (or crossed over by) any other molecules and (2) the estimated contour length L_C fell within the range 100–400 nm (bounds chosen given the expected contour length of ~250 nm for natural DNA).
14. When the next pixel in the DNA skeleton only has one coordinate (x or y) different from the previous pixel, the segment between two pixels is considered even. If both coordinates (x and y) are different from the previous pixel, the segment is considered odd. If moving from one pixel to the next there is an odd-to-even or even-to-odd transition, the segment is treated as a corner. Therefore, (n_e, n_o, n_c) represents the number of even, odd, and corner segments in the DNA skeleton, respectively.
15. As reported in previous studies [1, 11, 14, 15] estimates of helical rise determined using tapping-mode AFM in air (e.g., $h = 3.11 \pm 0.34$ Å/bp for natural DNA) underestimate the 3.38 Å/bp value measured by crystallography.
16. Again this procedure occurs automatically but with human supervision. This initial analysis revealed the importance of step sizes in the range 4–9 nm so that subsequent coordinate representations were generated for each of $l = 4, 4.25, 4.5, \ldots,$ 9 nm.
17. Clearly, $R \equiv R(d)$ is a function of arc length separation d (where d is an obligate multiple of l), but only R is written for simplicity. Similarly, $\theta(d)$ and other length and angle moments are assumed implicit functions of d.
18. It is known that reusing data (i.e., computing a quantity for each separation distance along a DNA contour) yields highly correlated points [14]. To avoid this, randomly divide each DNA representation into shorter length segments such that no piece is used twice and the maximal segment length is restricted to 80% of the expected contour length (~200 nm). This sampling procedure effectively results in a double average, over both conformation and position. The error bars in Fig. 7 (and Figs. 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24) show one standard deviation computed from 1000 different random draws from the same set of DNA representations. For the quantities that are quotients we apply error propagation to determine the standard deviation.
19. The bending energy function E_{WLC} is called classical (or harmonic) because it is a quadratic function of the bending strain variable θ.

20. The persistence length of DNA is sensitive to many experimental parameters such as temperature, pH, ionic forces, and excluded volume effects for long DNA fragments. These factors can in part explain discrepancies between P values estimated by different groups using various techniques. In particular, we note that Mg^{2+} ion concentrations in the low millimolar range (as recommended here) have been found to reduce the persistence length of DNA only slightly from monovalent salt estimates. Additionally, the effects of long-range electrostatic interactions are assumed to be irrelevant for behavior on length scales greater than l, which in our experiments was as small as 3 nm. In typical solvent conditions, where the Debye screening length is less than 5 nm, this assumption is reasonable.
21. The energy function E_{WLC} is harmonic. However, even if E is chosen to be non-harmonic, the angular distribution will approach a Gaussian form at large separations d because the iterated convolution of any distribution with itself converges to a Gaussian form. That is, even if non-harmonic elastic behavior is present, it will be hidden on long length scales by thermal fluctuations.
22. Excluded volume effects, which result from interactions between different molecules or among segments of the same molecule (self-avoiding effects), increase with increasing contour length, increasing DNA concentration, and transfer from three to two dimensions [1, 16]. Excluded volume effects can perturb the observed DNA conformations (and subsequently determined statistical quantities), leading to misinterpretation of data. We restrict the fitting to $0 < d < l_{\max}$, with $l_{\max} \approx 2P$ (~100 nm), since fitting over the entire range can yield an overestimation of P [14]. Indeed, deviation from the expected straight line for separations >120 nm may be indicative of the contribution of excluded volume effects. Alternatively, the analysis in this range (separation distances larger than ~170 nm) might start being affected by finite sampling effects, c.f. Fig. 7. Our results are consistent with previous reports [1] that self-avoiding interactions affect θ at shorter separations than R.
23. Assuming a normal distribution for $G(\theta)$, as predicted by WLC theory, k should always equal three (regardless of dimensionality of the molecules).
24. Unlike $\langle \mathbf{R} \cdot \mathbf{u}_0 \rangle$ and $\langle R^2 \rangle$, the fourth moment of R does not possess the same mathematical form in 3D as in 2D, so that this conformational statistic provides a reliable means of determining the dimensionality of the deposited molecules. The fits in Fig. 21 confirm that, at least over length scales less than ~100 nm, the DNA contours reflect equilibrium two-dimensional chain conformations.

25. Abels et al. [17] looked at multiple segment lengths l and showed that the value of $\langle \cos \chi \rangle$ peaks at a particular value of l and decreases at lower and higher l. The authors reasoned that as l decreased below the pixel resolution for the AFM setup, discretization of χ angles led to an underestimation of $\langle \cos \chi \rangle$. Additionally, they proposed an undercount of the number of large and small χ angles as l became large enough to average over them, similarly leading to underestimation of $\langle \cos \chi \rangle$.
26. Several trends can be observed from these data. First, all but one of the ten WLC predictions strongly depends on choice of step size l. Generally, the estimates of P are strictly increasing functions of increasing l. In some cases, the dependence on l is so dramatic that P increases 400% from $l = 3$ nm to $l = 17$ nm. In contrast, the estimate of P from the $\langle R^4 \rangle - \langle R^2 \rangle^2$ prediction is a decreasing function of increasing l. However, the maximal observed decrease in P is less than 5%. Overall, this estimator appears to be the most reliable: it is the most insensitive to choice of step size and it provides a means of assessing the dimensionality of the molecules being imaged. We recommend that if choosing only one of the ten WLC predictions to estimate P, the $\langle R^4 \rangle - \langle R^2 \rangle^2$ prediction would certainly be the first choice.
27. Our initial analysis revealed the importance of identifying the proper step size to obtain the best possible estimates of P. In particular, we realized that the optimal step size l_{optimal} occurred between 4 nm and 9 nm. We worried that discretization of l might prevent us from identifying the true (non-integral) value of l_{optimal}, with possible negative ramifications. Therefore, we further examined step sizes from 4 to 9 nm at every quarter nanometer.

Acknowledgments

This work was supported by the Mayo Foundation, Mayo Edward C. Kendall Fellowship (JPP), and the National Institutes of Health (GM75965 to LJM).

References

1. Rivetti C, Guthold M, Bustamante C (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. J Mol Biol 264(5):919–932
2. Hansma HG (2001) Surface biology of DNA by atomic force microscopy. Annu Rev Phys Chem 52:71–92
3. Bustamante C, Smith SB, Liphardt J, Smith D (2000) Single-molecule studies of DNA mechanics. Curr Opin Struct Biol 10 (3):279–285
4. Bustamante C, Bryant Z, Smith SB (2003) Ten years of tension: single-molecule DNA mechanics. Nature 421(6921):423–427

5. Neuman KC, Nagy A (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat Methods 5(6):491–505
6. Murugesapillai D, Bouaziz S, Maher LJ 3rd, Israeloff NE, Cameron CE, Williams MC (2017) Accurate nanoscale flexibility measurement of DNA and DNA-protein complexes by atomic force microscopy in liquid. Nanoscale 9 (31):11327–11337
7. Peters JP, Mogil LS, McCauley MJ, Williams MC, Maher LJ 3rd (2014) Mechanical properties of base-modified DNA are not strictly determined by base stacking or electrostatic interactions. Biophys J 107(2):448–459
8. Peters JP, Yelgaonkar SP, Srivatsan SG, Tor Y, Maher LJ 3rd (2013) Mechanical properties of DNA-like polymers. Nucleic Acids Res 41 (22):10593–10604
9. Vologodskaia MY, Vologodskii AV (2002) Contribution of the intrinsic curvature to measured DNA persistence length. J Mol Biol 317(2):205–213
10. Wang H, Dodd IB, Dunlap DD, Shearwin KE, Finzi L (2013) Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor. Nucleic Acids Res 41 (11):5746–5756
11. Rivetti C, Codeluppi S (2001) Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B-to A-form transition on mica. Ultramicroscopy 87(1–2):55
12. Wiggins PA, van der Heijden T, Moreno-Herrero F, Spakowitz AJ, Phillips R, Widom J, Dekker C, Nelson PC (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. Nat Nanotechnol 1(2):137–141
13. Vologodskii A, Frank-Kamenetskii MD (2013) Survey and Summary strong bending of the DNA double helix. Nucleic Acids Res 41 (14):6785–6792
14. Faas FGA, Rieger B, Van Vliet LJ, Cherny DI (2009) DNA deformations near charged surfaces: electron and atomic force microscopy views. Biophys J 97(4):1148–1157
15. Moukhtar J, Faivre-Moskalenko C, Milani P, Audit B, Vaillant C, Fontaine E, Mongelard F, Lavorel G, St-Jean P, Bouvet P (2010) Effect of genomic long-range correlations on DNA persistence length: from theory to single molecule experiments. J Phys Chem B 114 (15):5125–5143
16. Joanicot M, Revet B (1987) DNA conformational studies from electron microscopy. I. Excluded volume effect and structure dimensionality. Biopolymers 26(2):315–326
17. Abels JA, Moreno-Herrero F, van der Heijden T, Dekker C, Dekker NH (2005) Single-molecule measurements of the persistence length of double-stranded RNA. Biophys J 88(4):2737–2744

Chapter 14

Quantitation of DNA-Binding Affinity Using Tethered Particle Motion

Bram Henneman, Joost Heinsman, Julius Battjes, and Remus T. Dame

Abstract

The binding constant is an important characteristic of a DNA-binding protein. A large number of methods exist to measure the binding constant, but many of those methods have intrinsic flaws that influence the outcome of the characterization. Tethered Particle Motion (TPM) is a simple, cheap, and high-throughput single-molecule method that can be used to reliably measure binding constants of proteins binding to DNA, provided that they distort DNA. In TPM, the motion of a bead tethered to a surface by DNA is tracked using light microscopy. A protein binding to the DNA will alter bead motion. This makes it possible to measure binding properties. We use the bacterial protein Integration Host Factor (IHF) as an example to show how specific binding to DNA can be measured. Moreover, we show a new intuitive quantitative approach to displaying data obtained via TPM.

Key words Single molecule, Tethered particle motion, Root mean square, Displacement, DNA binding, Nucleoid associated protein, IHF

1 Introduction

Determination of binding affinity is an important part of the characterization of DNA-binding proteins. Typically, the binding affinity is expressed in terms of the binding constant (K_D). The K_D is the ratio of the off-rate constant, k_{off}, and the on-rate constant, k_{on}. Operationally, the K_D is defined as the concentration at which half of the substrate is bound by ligand (here: DNA and protein, respectively). For protein-DNA binding, the K_D is conventionally determined using Electrophoretic Mobility Shift Assays (EMSA), Filter Binding Assays, and Fluorescence Anisotropy [1–3]. Additionally, Isothermal Titration Calorimetry (ITC), Surface Plasmon Resonance (SPR) and Microscale Thermophoresis (MST) are used [4–6]. These techniques measure protein-DNA affinity in bulk, and therefore cannot distinguish different populations of bound and unbound molecules. Some of these techniques separate the protein-DNA complexes from solution or use small sample

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_14,

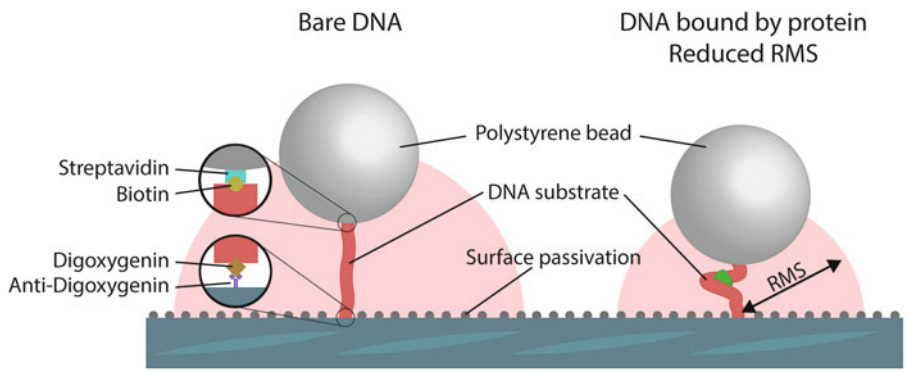


Fig. 1 Principle of Tethered Particle Motion (TPM). A bead is tethered to a glass surface by DNA. The position of the bead is tracked over time. Proteins that bind to the DNA and thereby deform it, change its apparent persistence length, which results in a change in the Root Mean Square displacement (RMS) of the bead

volumes, thus introducing large surface areas that may influence effective compound concentrations. Single-molecule techniques such as Optical Tweezers, Magnetic Tweezers, Acoustic Force Spectroscopy (AFS), and Atomic Force Spectrometry (AFM) can be used to estimate the K_D [7, 8], but require a high level of expertise and often expensive experimental setups. Single-molecule Förster Resonance Energy Transfer (smFRET) allows for measuring of DNA that is not bound to a surface and/or a bead [9], but requires fluorescently labeled DNA and protein and/or a sufficiently large and predictable conformational change of the DNA.

Tethered particle motion (TPM) is a simple, label-free single-molecule technique that can be used to characterize DNA-binding proteins in a high-throughput manner. The technique relies on the tracking of a bead in solution that has been tethered to a glass surface by a DNA molecule or other macromolecule (Fig. 1 and Fig. 2) [10]. The bead exhibits Brownian motion, which is restricted by the tether. In contrast to some of the above-mentioned techniques, no additional external force is applied to the bead. The xy-positions of the bead are recorded over time and are used to calculate the root mean square displacement (RMS) of the bead (Eq. 1). TPM has a high throughput (*see* **Note 1**), which makes it a powerful tool to study the characteristics of DNA-binding proteins. DNA-binding proteins that deform the DNA upon binding, change the apparent persistence length (L_p) of the DNA, a measure for rigidity, when bound, thereby altering the RMS [11]. Although the L_p cannot be directly calculated with TPM, the change in RMS is a reliable measure for protein binding. TPM has been used to measure DNA-binding by architectural proteins, proteins involved in DNA replication, transcription, DNA repair, and recombination [12–21].

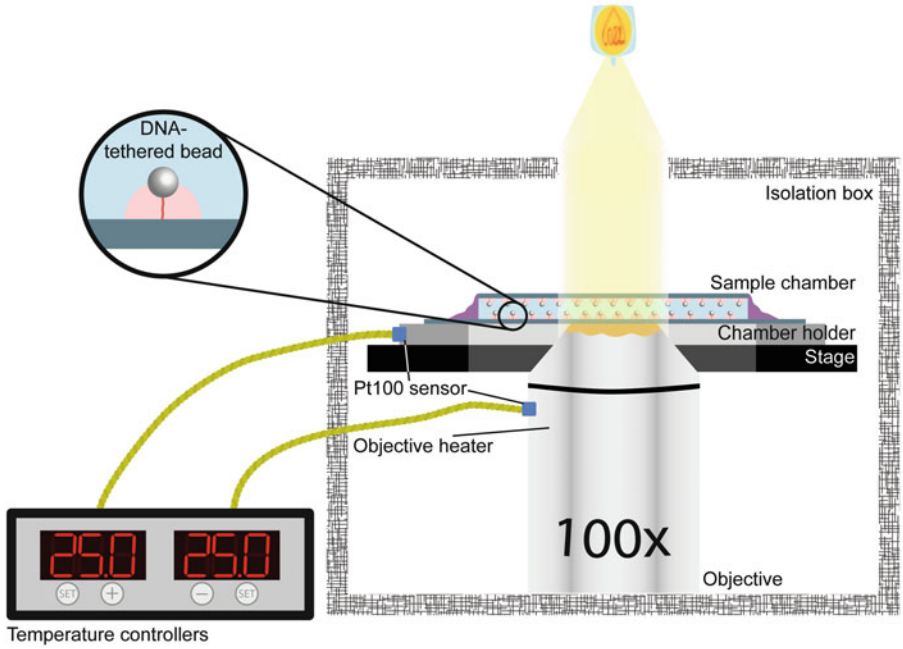


Fig. 2 Schematic of the Tethered Particle Motion setup. The sample chamber is heated by two heating elements, one placed in the sample chamber holder and one around the objective. An isolation box covers the stage, sample chamber and heating elements

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \bar{x})^2 + (y_i - \bar{y})^2\right]} \qquad (1)$$

Equation 1 Calculation of the Root Mean Square displacement. x and y are the coordinates of the two-dimensional position of the bead.

In this chapter we describe the use of TPM to measure the K_D of sequence-specific DNA-binding proteins. As an example we characterized the bacterial Integration Host Factor (IHF) [22] from *E. coli*. IHF is a nucleoid-associated protein (NAP) that is involved in shaping genome architecture in bacteria. IHF also plays a role in up- and downregulation of genes [23]. This function is attributed to the ability of IHF to bend DNA and bring regulatory elements together, as well as due to direct interactions with RNA polymerase [24, 25]. IHF is known to bend DNA when bound at its high affinity recognition sequence [26] and also bends DNA when bound to DNA nonspecifically [27]. Bound at its specific site, binding angles of 120–160° have been reported [28–30].

2 Materials

2.1 Parafilm Semi-Circles

1. Parafilm.
2. Office tape.
3. Laser cutter.
4. Computer-aided design (CAD) software (for example AutoCAD or CorelDRAW).
5. Tweezers.

2.2 Sample Chambers

1. Round cover glass, diameter 35 mm (VWR).
2. Round cover glass, diameter 28 mm (Thermo Scientific).
3. Parafilm semi-circles (see above).
4. Cover slip rack (Diversified Biotech).
5. Ethanol.
6. Acetone.
7. Beaker.
8. Sonication bath.
9. Heating block, capable of heating up to 100 °C, with hot plate for 35 mm cover glasses (*see* **Note 2**).
10. Tweezers with pointed tips.
11. KIMTECH SCIENCE* precision wipes tissue wipers (Kimberly-Clark).
12. Nitrile gloves (*see* **Note 3**).
13. Whatman paper.

2.3 Bead Suspension

1. Streptavidin-coated polystyrene beads, diameter 0.44 μm (Kisker Biotech).
2. Immersion sonicator with fine tip.
3. Bead buffer (BB) (10 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA pH 8.0, 1 mM DTT, 3% glycerol, 100 μg/mL BSA-acetylated [Ambion]).
4. Vortex.

2.4 Anti-Digoxigenin Solution

1. Anti-digoxigenin (anti-Dig) polyclonal antibodies 200 μg (Roche).
2. Bead buffer (BB) (see above).

2.5 DNA Substrate

1. Template for DNA substrate containing the binding site of interest. PCR amplification with compatible primers should yield a substrate of desired length and sequence (*see* **Note 4**).

2. Biotinylated forward primer, 10 μM in 10 mM Tris, pH 7.5, designed to be compatible with the template (*see* **Note 5**).
3. Digoxygenin-labeled reverse primer, 10 μM in 10 mM Tris, pH 7.5, compatible with the template (*see* **Note 5**).
4. Phusion master mix (New England Biolabs).
5. DMSO.
6. Milli-Q.
7. GenElute PCR clean-up kit (Sigma Aldrich).
8. PCR reaction tubes.
9. Biorad C1000 Thermal Cycler.

2.6 Substrate Quality Control

1. Agarose gel, 1% in Tris/Borate/EDTA (TBE) buffer.
2. GelRed nucleic acid gel stain (Biotium).
3. GeneRuler DNA marker.
4. Spectrophotometer to measure DNA concentrations (Nanodrop, Qubit, SimpliNano or comparable device).

2.7 Sample Chamber Reagents

1. Bead buffer (BB) (see above).
2. Anti-digoxigenin solution.
3. Passivation solution (4 mg/mL Blotting Grade Blocker [Bio-Rad] dissolved in SB).
4. Experimental buffer (EB) suitable for protein binding (*see* **Note 6**).
5. Protein of interest, preferably at high concentrations to avoid as much as possible components of the protein storage buffer on the EB (*see* **Note 7**).

2.8 Imaging

1. Isolation table.
2. Inverted microscope (Nikon Diaphot 300).
3. 100× oil-immersion objective (NA 1.25).
4. Immersion oil for microscopy (Merck).
5. Heat chamber (Bioscience tools TC-HLS-025 heating element for heating of the objective and a custom sample chamber heating element consisting of 40 parallel resistors); Pt100 sensors attached to the sample chamber holder and the objective, and two SA200 PID digital temperature controllers).
6. Climatized room.
7. Isolation box.
8. Thorlabs CMOS camera DCC1545M.
9. Lens paper.
10. Nacre-free nail polish.
11. Particle-tracking software (software available online [31]; *see* **Note 8**).

2.9 Data Analysis and Representation

1. Computer.
2. Custom MatLab routine for calculating RMS, standard deviation of the RMS and anisotropic ratio. The routine also selects the beads that meet the requirements for reliable beads (*see* Subheading 3.9).
3. Graphing and data analysis software, such as KaleidaGraph, Origin or IGOR Pro.
4. Adobe Illustrator.

3 Methods

3.1 Parafilm Semi-Circles

1. Cut a strip of parafilm at a length fitting the grid of the laser cutter.
2. Remove the protection layer from the parafilm and place the parafilm onto the grid of the laser cutter.
3. Load the template of the semi-circles in the laser cutter software of choice. Design has to be such that two parafilm semi-circles on a 28 mm coverglass create a 6 mm wide channel. Position the parafilm so that the template virtually overlays the parafilm.
4. Fix the parafilm on the grid with office tape.
5. Cut the parafilm (*see* **Note 9**).
6. Remove parafilm strip. The semi-circles will stick to the grid of the laser cutter. Take the semi-circles off the grid individually with tweezers.
7. Store the parafilm semi-circles at 4 °C to prevent sticking to each other.

3.2 Sample Chambers

The following steps should be carried out with gloves, both to protect the hands from exposure to chemicals and to keep the sample chambers free of contaminants.

1. Place the cover glasses in the slide holder using the tweezers with pointed tips (*see* **Note 10**).
2. Place the holder in a beaker and submerge in acetone. Place the beaker in a sonication bath and sonicate for one cleaning cycle (*see* **Note 11**).
3. Transfer the slide holder into a new beaker and submerge the slide holder in ethanol. Place the beaker in a sonication bath and sonicate for one cleaning cycle.
4. Remove the holder from the beaker. Place the cover glasses on a precision wipe and dry the cover glasses by gently wiping them with precision wipes. Turn the cover glasses over and leave exposed to air for 10 min to dry.

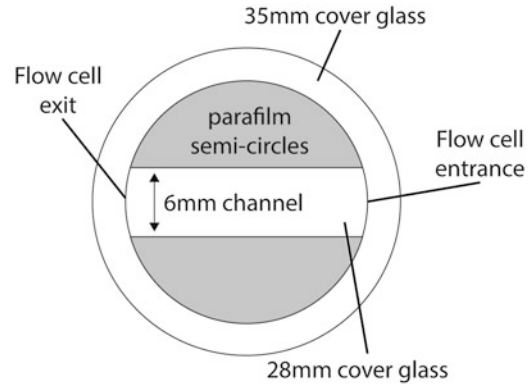


Fig. 3 Sample chamber for TPM. The sample chamber consists of two parafilm semi-circles sandwiched between two cover glasses. The sample chamber is agglutinated by heating to 100 °C

5. Using tweezers place two parafilm semi-circles in parallel on opposing sides of a 28 mm cover glass, so that a straight channel is created (Fig. 3). The parafilm should align with the glass on the round edges. Push the parafilm firmly but carefully onto the cover glass with the tweezers so that the parafilm adheres to the cover glass (*see* **Note 12**).
6. Place the 35 mm cover glass on top of the parafilm-covered side of the 28 mm cover glass using tweezers, carefully centered onto the 28 mm cover glass. Push firmly but carefully with the tweezers so that the slides stick together.
7. Place the sample chamber on the hot plate (preheated to 100 °C) using tweezers. Cover the sample chamber with the lid and press firmly and evenly on the lid for 10 s. Remove the lid from the sample chamber (*see* **Note 13**).
8. Lift the sample chamber off the hot plate using tweezers and let it cool down on a precision wipe. Store the sample chamber in a small plastic storage box at room temperature.

3.3 Bead Suspension

1. Dilute the stock of streptavidin-coated beads to 0.01% w/v in 5 mL SB (*see* **Note 16**).
2. Vortex thoroughly for 5 min.
3. Sonicate the bead suspension on ice for 3 min (active), 2.5 s on and 9.9 s off, at 25% of the maximum power of the immersion sonicator. The total run time will be around 15 min. The tip of the immersion sonicator should be submerged by roughly 3 cm. Never sonicate the bead suspension for more than 5 min of active time, as this may damage the beads and its coating.

4. Flow 30 μL of bead suspension into an empty sample chamber. If less than 1% of the beads is clustered, it is considered a good bead suspension.
5. Store the bead suspension at 4 °C.

3.4 DNA Substrate

1. Add the following components to a PCR reaction tube on ice:

Component	Quantity
DNA template in 10 mM Tris pH 7.5	10 ng
Biotinylated forward primer	25 pmol
Digoxygenin-labeled reverse primer	25 pmol
Phusion master mix	25 μL
DMSO	1.5 μL
Milli-Q	To total volume of 50 μL

2. Place the PCR reaction tube in the thermal cycler and use the following touchdown protocol (*see* **Note 14**):

Temperature	Duration	Cycles
98 °C	30 s	
98 °C 72 °C (−1 °C per cycle) 72 °C	10 s 20 s 60 s	15×
98 °C 57 °C 72 °C	10 s 20 s 60 s	25×
72 °C	600 s	
12 °C	∞	

3. Purify the PCR product using a PCR clean-up kit (*see* **Note 15**).
4. Run 2 μL of PCR product on a 1% agarose gel, prestained with GelRed, alongside a suitable DNA marker. Verify the length of the DNA substrate and estimate the concentration using a spectrophotometer. Store the DNA substrate at −20 °C at a concentration of 100–500 pM.

3.5 Anti-Digoxygenin Solution

1. Dissolve 200 μg of anti-digoxygenin in 1 mL of BB.
2. Aliquot per 20 μL and store the anti-digoxygenin at −20 °C.
3. When using the anti-digoxygenin solution, thaw an aliquot and dilute 10 times with BB.

3.6 DNA-Tethered Bead in Sample Chamber

1. Take a clean sample chamber and flow in 20 μg/mL anti-dig solution by holding the tip of a pipette against the opening on the right side of the sample chamber at a 45° angle and carefully pressing the plunger button (*see* **Note 17**). Incubate for 10 min at room temperature.
2. Wash the sample chamber with 100 μL passivation solution and incubate for 10 min at room temperature. The excess liquid flowing out of the sample chamber is collected using Whatman paper. This promotes an undisturbed and uninterrupted flow in the channel.
3. Wash the sample chamber with 100 μL BB (*see* **Note 18**).
4. Flow in 100 μL 100 pM DNA substrate (*see* **Note 19**). Incubate for 10 min at room temperature.
5. Wash the sample chamber with 100 μL BB.
6. Pipette the bead suspension stock up and down and flow in 100 μL bead suspension. Incubate for 10 min at room temperature.
7. Wash the sample chamber with 100 μL EB.
8. Flow in 100 μL protein solution, diluted to the desired concentration in EB. Incubate for 10 min at room temperature.
9. Wash the sample chamber with 100 μL protein solution of the same concentration as in **step 8**.
10. Seal the sample chamber on both sides by applying nail polish to the opening of the sample chamber. When the nail polish is dry, the sample chamber is ready to be used.

3.7 Dilution Series

1. Calculate a dilution series with a range from 0 to 1000 nM using twofold dilution steps (*see* **Note 20**). Determine the concentration range in which DNA binding is to be expected. Literature values can be used if available; if not, preliminary experiments may be required.
2. Dilute a protein stock aliquot to a concentration that can be used to make further protein dilutions with EB (*see* **Note 21**). Store this stock on ice (*see* **Note 22**).
3. Using the working stock, make a dilution series by diluting part of the working stock with EB to the desired concentration. Dilute part of this sample with EB in order to create the next sample (*see* **Note 21**). Store samples on ice (*see* **Note 22**).
4. Measure at least two, but preferably three dilution series (*see* **Note 23**).

3.8 Imaging

1. An hour before imaging, turn on the lamp and the heating stage.
2. Place immersion oil on the objective and start imaging software.

3. Place a sample chamber in the sample chamber holder.
4. Raise the objective until the beads are in focus (*see* **Note 24**).
5. Close the isolation chamber. The temperature of the chamber and the objective will now rise to a set temperature (*see* **Note 25**).
6. Re-focus on the beads and ensure that no focal drift is observed (*see* **Note 26**).
7. Select a field of view by moving the stage of the microscope. Select the beads and track them for at least 1500 frames. Images are taken at 25 Hz, with an exposure time of 20 ms. At least 100 "good" bead-tether combinations are needed for quantification after quality check (explained in Subheading 3.9; *see* **Note 27**).
8. Use the tracking software to determine x- and y-positions of the beads (*see* **Note 8**).

3.9 Data Analysis

1. Calculate the anisotropic ratio (Eq. 2), the root mean square displacement (RMS) and the standard deviation of the RMS for all measurements at a given concentration combined (*see* **Note 28**). Discard tethers with an anisotropic ratio of above 1.3. Discard measurements with a standard deviation of the RMS larger than 6% of the RMS. We thereby exclude beads sticking to the surface during the measurement (*see* **Note 29**).
2. Plot all RMS values that meet the requirements as explained in **step 1** a in a histogram.
3. Fit a Gaussian curve to the histogram to determine the average RMS and standard deviation for each condition.
4. Calculate the standard error of the mean for each condition.

$$\alpha = \frac{l_{\text{major}}}{l_{\text{minor}}} \tag{2}$$

Equation 2 Calculation of the anisotropic ratio (α). l_{major} and l_{minor} represent the length of the major and minor axis of the xy-scatter plot, respectively.

3.10 Analysis of Specific Protein-DNA Binding

1. Determine the RMS of the DNA-tethered bead in the absence of protein. This is the position of the peak in the Gaussian fit at 0 nM (Fig. 4a).
2. Determine the RMS at saturation levels of specific binding. This is the concentration at which the peak seen at 0 nM has completely disappeared and a peak at a different RMS has appeared (Fig. 4f).
3. Concentrations at which two peaks are observed, represent a situation in which a subpopulation of the DNA is bound by protein (Fig. 4b–e).

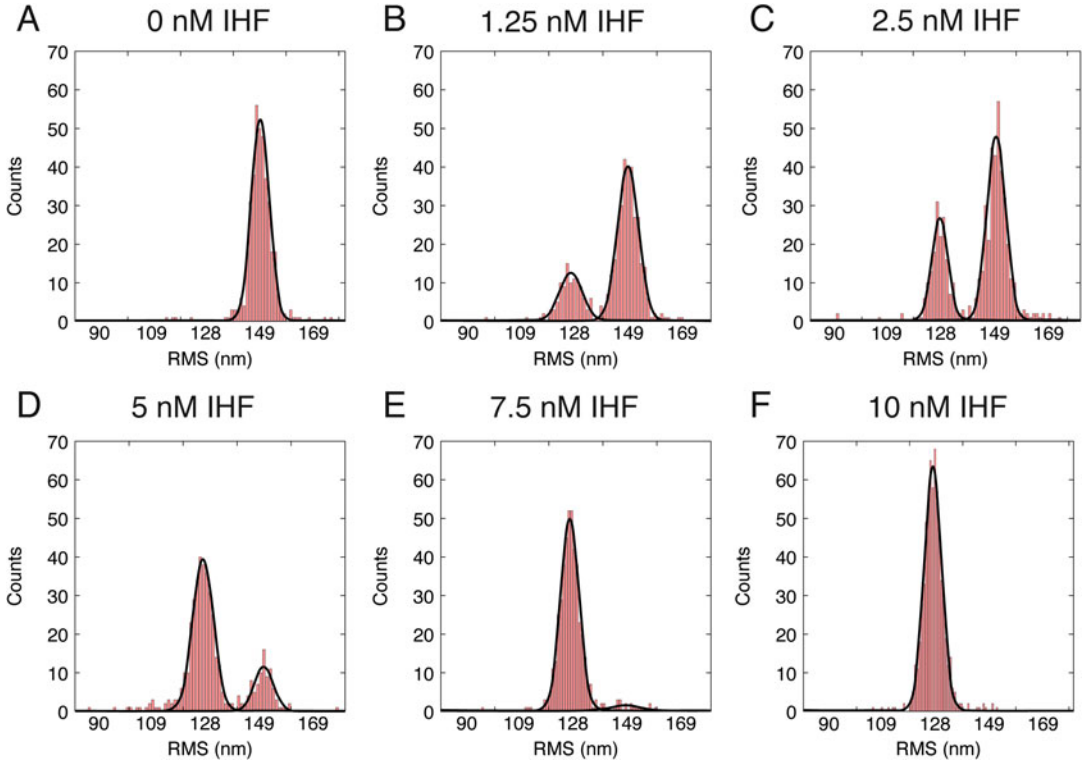


Fig. 4 Specific binding of IHF to DNA at 0–10 nM. (**a–f**) histograms of RMS values obtained for DNA-tethered beads at IHF concentrations of 0, 1.25, 2.5, 5, 7.5, and 10 nM. The histograms have been fitted with a Gaussian function in KaleidaGraph (black, semi-opaque line)

4. For the concentrations at which subpopulations of bound and unbound states can be observed, calculate the occupancy (Eq. 3). Occupancy is the ratio between bound and unbound DNA substrates.
5. Find the K_D and Hill coefficient (n) by solving the Hill binding model (Eq. 4) for K_D and n (*see* **Note 30**).
6. Fit the occupancies using the Hill-binding model (Fig. 5; *see* **Note 31**). Use graphing and data analysis software to display the occupancies per concentration and the Hill fit in a graph.

3.11 Example of K_D Estimation for DNA-Binding Protein IHF

We used IHF as an example to illustrate how to characterize specific DNA binding proteins. In our example, we observe a reduction in RMS at low (1.25 nM) IHF concentrations from 149 to 127 nm (Fig. 4). This reduction in RMS is caused by DNA bending as a consequence of IHF binding at its specific binding site (*see* **Note 32**). Saturation of specific binding is achieved at 10 nM, the concentration at which only the population with reduced RMS is observed. Fitting the data using the Hill binding model (χ^2: 4.6) yields a binding affinity (K_D) of 3.0 ± 0.3 nM, which is in

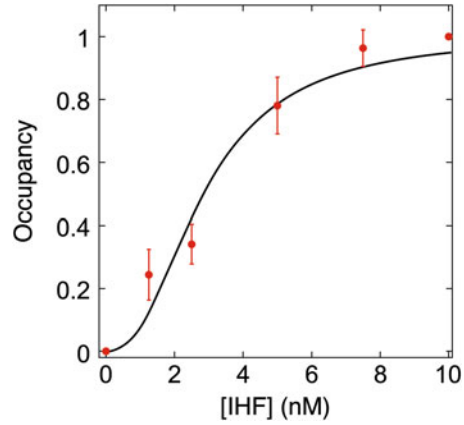


Fig. 5 Binding curve for specific binding of IHF to DNA. The data points were fit using the Hill-binding model. The Hill coefficient (n) is 2.5 ± 0.5, and the binding affinity (K_D) is 3.0 ± 0.3 nM. Error bars represent the standard error of the mean

agreement with values reported using EMSA on different variants of the binding consensus sequence [32]. A Hill coefficient (n) of 2.5 ± 0.5 was found. A Hill coefficient >1 indicates positively cooperative binding; $n < 1$ indicates negatively cooperative binding (*see* **Note 33**).

$$\theta = \frac{h_{\text{peak,bound}}}{h_{\text{peak,bound}} + h_{\text{peak,unbound}}} \tag{3}$$

Equation 3 Calculation of occupancy. The occupancy is the fraction of substrate bound by protein and can be calculated from the area of the peaks of the fitted Gaussian distributions.

$$\theta = \frac{1}{\left(\frac{K_D}{[L]}\right)^n + 1} \tag{4}$$

Equation 4 Hill binding model. Occupancy θ, binding constant K_D, ligand concentration $[L]$, Hill coefficient n.

3.12 Data Representation

TPM is a single molecule method with a high throughput. As a consequence data are often reported in terms of an average RMS value. However, displaying the full population of RMS values can be insightful to illustrate distribution shape and population heterogeneity. The new representation approach presented here accommodates both average values, as well as unprocessed population information.

1. Plot all data points as black squares using the graphing and analysis software.
2. In the same plot, include the values of the average RMS at all concentrations and display in a color different from that of the data points.

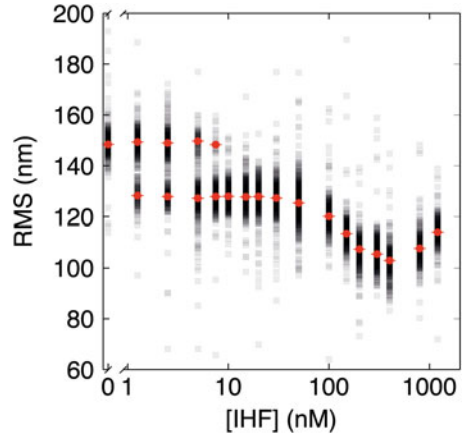


Fig. 6 Compaction of DNA by IHF. Black, partially transparent squares represent the individual RMS values of DNA-bead complexes after application of anisotropic ratio and standard deviation of the RMS as selection criteria. Red dots and bars represent the average RMS that resulted from fitting with a Gaussian function

3. Export the plot as a vector image.
4. Open the plot in Adobe Illustrator and select all black squares. Change the opacity from 100% to 5% (Fig. 6; *see* **Note 34**).

In our example, specific and nonspecific binding by IHF can be observed. Specific binding of IHF is observed between 0 and 10 nM, where the RMS decreases from 149 ± 5 nm to 127 ± 5 nm and is characterized by the two populations. Intensity profiles of the populations are equivalent to the histograms of Fig. 4. In the nonspecific binding regime, at IHF concentrations >30 nM, we find a single population of which the RMS continues to reduce as a function of IHF concentration up to 400 nM. This further reduction in RMS is reduced to sequence unspecific binding of the protein. At concentrations >400 nM the RMS increases, which is attributed to stiffening of DNA by IHF bound at high density along the DNA, similar to the effects of observed for the HU protein [33].

4 Notes

1. Up to 100 beads were measured for 1 min.
2. A custom-made hot plate was used to heat the sample chambers. The hot plate is made from a circular massive metal block in which a shallow cavity is carved (diameter 35 mm). A sample chamber fits into the cavity. A metal lid (diameter 35 mm) with a plastic handle covers the sample chamber. The hot plate is

placed on a heating block and heats the sample chamber to 100 °C.

3. Nitrile gloves were used, as the powder that is often found in rubber gloves may contaminate the sample chamber.
4. As a template the plasmid pGp1 was used [34], which contains part of the bacteriophage mu genome. A 685 bp fragment that contains 1 IHF-binding site exactly at the center was amplified, which is expected to result in the highest effect on end-to-end distance [30]. Our experience is that a DNA substrate in the order of 700 bp is ideal, as protein binding results in a clearly detectable change in RMS.
5. The forward and reverse primers that we used are 5' BIO-TGATTAAGGTTGTGGTTAATTTGTTATCAGTTCC-G3' and 5' DIG-ATTACTTCCGGTTTAGTTTCTAAGGCG 3', respectively.
6. The experimental buffer contained 10 mM Tris–HCl pH 8.0 and 100 mM KCl, which is suitable for specific binding of IHF [35]. The osmolarity of the buffer should be sufficiently high to prevent nonspecific binding of specific binding proteins. A different protein of interest may require buffer optimization.
7. Recombinant *E. coli* IHF at a concentration of 120 μM was used here. This stock was diluted 100-fold at minimum. The number of freeze-thaw cycles was minimized by aliquoting the protein stock, to guarantee protein activity.
8. Software written for Acoustic Force Spectrometry was used [31], which is freely available online (http://figshare.com/articles/AFS_software/1195874). Other particle tracing software has been described and compared by Chenouard et al. [36]. Our lab also has positive experience with Poly-ParticleTracker [37].
9. Here, the computer-aided design software AutoCAD was used to design the cutting template in combination with the Versa-Laser tabletop Laser cutter. We cut the parafilm at 55% laser power, 30% of the maximum speed, 1000 pulse and 1.00 mm Z-axis. Different laser cutters may require different settings, allowing the parafilm to be cut straight and without melting or burning.
10. Ensure that each slot is occupied by only one cover glass. The cover glasses tend to stick together, which prevents them from being cleaned properly.
11. For sonication of the glass slides, the Emag EMMI 4 sonicator was used, which is pre-programmed in cycles of 9 min at 42.5 kHz.
12. Note that the parafilm leaves grease marks on the cover glass, so avoid sliding after placing them.

13. Heating the sample chamber for too long may result in excessive melting of the parafilm, which causes a non-straight sample chamber. This prevents a good flow in the sample chamber. As a result, coating of the surface with anti-digoxygenin and/or passivation agent may be heterogenous, which reduces throughput and results in lower effective protein concentrations. Especially experiments at low concentration may then suffer from reduced reproducibility.
14. Conventional protocols with fixed annealing temperatures may be used as well, but need to be individually optimized for different primer-template combinations and synthesized fragment lengths.
15. Purification using the GenElute PCR clean-up kit separates the PCR product from other components that were added to the reaction tube to facilitate the PCR reaction, such as excess primers, nucleotides, and salts. The result of this purification is the labeled DNA substrate in 10 mM Tris–HCl pH 8.5 (provided with the kit).
16. The bead stock is vortexed and sonicated to homogenize the suspension. Beads tend to stick together, which will distort measurements.
17. Check if the sample chamber has no cracks in either of the cover glasses, as defects may create a flow inside the sample chamber while measuring. This will affect the RMS by giving it a directional bias. Also, check if the channel is straight; *see* Fig. 3. Left-handed people may prefer using the opening on the left side of the sample chamber for sample loading.
18. The volumes used for washing depend on the volume of the sample chamber. As a rule of thumb, we wash with approximately three times the volume of the sample chamber.
19. The DNA substrate concentration is 100 pM, but when we find that this concentration is not sufficient for a particular experiment, we raise the concentration up to 500 pM.
20. A dilution series with concentrations from 0 to 1000 nM is used, with twofold dilution steps. Based on results and literature values, an additional range and/or different dilution steps may be used, which may result in a more reliable calculation of the K_D.
21. When measuring many samples, consider breaking up the dilution series in two or more overlapping ranges. Here, we diluted a 120 μM IHF stock to a working stock of 1.2 μM. From that, we made a serial dilution which included IHF concentrations of 800, 400, 300, 200, 150, 100, and 50 nM: the high range concentrations. For the low range concentrations, we diluted the 1.2 μM working stock directly to 50 nM, from which we

subsequently made a serial dilution which included IHF concentrations of 30, 20, 15, 10, 7.5, 5, 2.5, and 1.25 nM. We verified that the RMS values at 50 nM from the high range dilution series were identical to the RMS values at 50 nM from the low range dilution series. We also measured the RMS in the absence of IHF, indicated with 0 nM.

22. Do not freeze the protein solutions, since the activity of the protein may be affected.
23. For reliable results and to minimize the effect of pipetting errors, experiments should be carried out at least in duplo, but preferably in triplo, and independently of each other. Especially when the concentration range over which binding takes place is narrow, small changes in sample concentrations can be of significant influence on the outcome of the binding curve.
24. Due to a sedimentation effect, there may be a slight difference in RMS between beads tethered to the bottom surface and beads tethered to the top surface. Therefore, always choose the same surface for measurements. Also, the focus for top and bottom surfaces may differ.
25. It may take 5–20 min, depending on the temperature, before the isolation chamber has completely reached the desired temperature.
26. While the temperature of the immersion oil on the objective rises, the refractive index of the oil changes slightly. As a result, focal drift is observed. Wait until the immersion oil has reached the desired temperature and then refocus. When no focal drift is observed, the immersion oil has the desired temperature.
27. At least 100 good bead-tether combinations are needed, but the actual amount of measured bead-tether combinations to reach this limit may differ. For IHF, 300 measured bead-tether combinations always resulted in more than 100 good combinations. The amount of bead-tether combinations that need to be measured, depends on the protein that is used.
28. Verify that individual experiments done at the same concentration give similar outcomes. If so, RMS values can be combined and can be fitted together. If one measurement is a clear outlier due to a known cause, discard the measurement. In other cases, re-measure the sample and check if the results align with any previous measurements.
29. Discard any bead with an anisotropic ratio >1.3, which is considered non-symmetrical. Non-symmetrical beads indicate for example that a bead is attached to two DNA tethers. An extensive study that links motion patterns to setup defects has been done by Visser et al. [38]. Also discard any bead with a standard deviation of the RMS of $>6\%$, since this indicates

rapid protein dissociation and sticking and/or releasing of the bead to or from the sample chamber surface.

30. The Hill equation can be solved using IGOR Pro, which was done here, but MatLab and other data analysis programs are also suitable for this.
31. As error bars, the standard error of the mean of the individual measurement series per concentration has been used.
32. If desired, the effective persistence length of the DNA can be determined for a single protein bound to DNA, which could be used to calculate the DNA bending angle as described by Kulić and coworkers [11]. However, this requires an extensive simulation series, which reveals the relation between RMS and L_p. The simulations are highly dependent on the length and sequence of the DNA substrate, buffer conditions, and temperature of the experiment. This means that for every experiment in which one or more of these parameters are changed, a new simulation series is required. The simulations have been described by Driessen et al. [39].
33. Here, the Hill coefficient is 2.5, which indicates positively cooperative binding. However, in this example, only one IHF dimer binds to DNA. The apparent positively cooperative binding is possibly a result of the dimeric nature of the protein. Both components of the dimer have to bind to the DNA in order to bend the DNA.
34. For representation of all data points, squares are most suitable, because squares are easy distinguishable when data points overlap. Circles may make it harder to distinguish individual data points in the plot. The opacity that suits your data best depends on the number of data points. When using more data points than shown in the example, it may be useful to use a lower opacity. The Gaussian distribution of the data points should be visible from the intensity of the combination of data points.

Acknowledgments

This work was supported by grants from the Netherlands Organization for Scientific Research [VICI 016.160.613] and the Human Frontier Science Program (HFSP) [RGP0014/2014].

References

1. Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc 2 (8):1849–1861. https://doi.org/10.1038/nprot.2007.249
2. Gilman AG (1970) A protein binding assay for adenosine 3′:5′-cyclic monophosphate. Proc Natl Acad Sci U S A 67(1):305–312
3. Heyduk T, Ma Y, Tang H, Ebright RH (1996) Fluorescence anisotropy: rapid, quantitative

assay for protein-DNA and protein-protein interaction. Methods Enzymol 274:492–503

4. Ziegler A, Seelig J (2007) High affinity of the cell-penetrating peptide HIV-1 tat-PTD for DNA. Biochemistry 46(27):8138–8145. https://doi.org/10.1021/bi700416h
5. Hart DJ, Speight RE, Cooper MA, Sutherland JD, Blackburn JM (1999) The salt dependence of DNA recognition by NF-kappaB p50: a detailed kinetic analysis of the effects on affinityand specificity. Nucleic Acids Res 27 (4):1063–1069
6. Jerabek-Willemsen M, Wienken CJ, Braun D, Baaske P, Duhr S (2011) Molecular interaction studies using microscale thermophoresis. Assay Drug Dev Technol 9(4):342–353. https://doi.org/10.1089/adt.2011.0380
7. Yang Y, Sass LE, Du C, Hsieh P, Erie DA (2005) Determination of protein-DNA binding constants and specificities from statistical analyses of single molecules: MutS-DNA interactions. Nucleic Acids Res 33(13):4322–4334. https://doi.org/10.1093/nar/gki708
8. McCauley MJ, Williams MC (2011) Measuring DNA-protein binding affinity on a single molecule using optical tweezers. Methods Mol Biol 749:305–315. https://doi.org/10.1007/978-1-61779-142-0_21
9. Coats JE, Lin Y, Rueter E, Maher LJ 3rd, Rasnik I (2013) Single-molecule FRET analysis of DNA binding and bending by yeast HMGB protein Nhp6A. Nucleic Acids Res 41 (2):1372–1381. https://doi.org/10.1093/nar/gks1208
10. Schafer DA, Gelles J, Sheetz MP, Landick R (1991) Transcription by single molecules of RNA polymerase observed by light microscopy. Nature 352(6334):444–448. https://doi.org/10.1038/352444a0
11. Kulic IM, Mohrbach H, Thaokar R, Schiessel H (2007) Equation of state of looped DNA. Phys Rev E Stat Nonlinear Soft Matter Phys 75 (1 Pt 1):011913. https://doi.org/10.1103/PhysRevE.75.011913
12. van der Valk RA, Vreede J, Qin L, Moolenaar GF, Hofmann A, Goosen N, Dame RT (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-bridging activity. elife 6:e27369. https://doi.org/10.7554/eLife.27369
13. Driessen RP, Lin SN, Waterreus WJ, van der Meulen AL, van der Valk RA, Laurens N, Moolenaar GF, Pannu NS, Wuite GJ, Goosen N, Dame RT (2016) Diverse architectural properties of Sso10a proteins: evidence for a role in chromatin compaction and organization. Sci Rep 6:29422. https://doi.org/10.1038/srep29422
14. Chintakayala K, Sellars LE, Singh SS, Shahapure R, Westerlaken I, Meyer AS, Dame RT, Grainger DC (2015) DNA recognition by Escherichia coli CbpA protein requires a conserved arginine-minor-groove interaction. Nucleic Acids Res 43(4):2282–2292. https://doi.org/10.1093/nar/gkv012
15. Wang H, Yehoshua S, Ali SS, Navarre WW, Milstein JN (2014) A biomechanical mechanism for initiating DNA packaging. Nucleic Acids Res 42(19):11921–11927. https://doi.org/10.1093/nar/gku896
16. Mack AH, Schlingman DJ, Salinas RD, Regan L, Mochrie SG (2015) Condensation transition and forced unravelling of DNA-histone H1 toroids: a multi-state free energy landscape. J Phys Condens Matter 27 (6):064106. https://doi.org/10.1088/0953-8984/27/6/064106
17. Wu HY, Lu CH, Li HW (2017) RecA-SSB interaction modulates RecA nucleoprotein filament formation on SSB-wrapped DNA. Sci Rep 7(1):11876. https://doi.org/10.1038/s41598-017-12213-w
18. Lu CH, Li HW (2017) DNA with different local torsional states affects RecA-mediated recombination progression. ChemPhysChem 18(6):584–590. https://doi.org/10.1002/cphc.201601281
19. Fan HF, Cheng YS, Ma CH, Jayaram M (2015) Single molecule TPM analysis of the catalytic pentad mutants of Cre and Flp site-specific recombinases: contributions of the pentad residues to the pre-chemical steps of recombination. Nucleic Acids Res 43(6):3237–3255. https://doi.org/10.1093/nar/gkv114
20. Fan HF, Hsieh TS, Ma CH, Jayaram M (2016) Single-molecule analysis of varphiC31 integrase-mediated site-specific recombination by tethered particle motion. Nucleic Acids Res 44(22):10804–10823. https://doi.org/10.1093/nar/gkw861
21. Chen YF, Lu CY, Lin YC, Yu TY, Chang CP, Li JR, Li HW, Lin JJ (2016) Modulation of yeast telomerase activity by Cdc13 and Est1 in vitro. Sci Rep 6:34104. https://doi.org/10.1038/srep34104
22. Mangan MW, Lucchini S, Danino V, Croinin TO, Hinton JC, Dorman CJ (2006) The integration host factor (IHF) integrates stationary-phase and virulence gene expression in Salmonella enterica serovar Typhimurium. Mol Microbiol 59(6):1831–1847. https://doi.org/10.1111/j.1365-2958.2006.05062.x

23. Peacock S, Weissbach H, Nash HA (1984) In vitro regulation of phage lambda cII gene expression by Escherichia coli integration host factor. Proc Natl Acad Sci U S A 81 (19):6009–6013
24. Goosen N, van de Putte P (1995) The regulation of transcription initiation by integration host factor. Mol Microbiol 16(1):1–7
25. Khodr A, Fairweather V, Bouffartigues E, Rimsky S (2015) IHF is a trans-acting factor implicated in the regulation of the proU P2 promoter. FEMS Microbiol Lett 362(3):1–6. https://doi.org/10.1093/femsle/fnu049
26. Wang S, Cosstick R, Gardner JF, Gumport RI (1995) The specific binding of Escherichia coli integration host factor involves both major and minor grooves of DNA. Biochemistry 34 (40):13082–13090
27. Swinger KK, Rice PA (2004) IHF and HU: flexible architects of bent DNA. Curr Opin Struct Biol 14(1):28–35. https://doi.org/10.1016/j.sbi.2003.12.003
28. Rice PA, Yang S, Mizuuchi K, Nash HA (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell 87 (7):1295–1306
29. Travers A (1997) DNA-protein interactions: IHF--the master bender. Curr Biol 7(4): R252–R254
30. Dame RT, van Mameren J, Luijsterburg MS, Mysiak ME, Janicijevic A, Pazdzior G, van der Vliet PC, Wyman C, Wuite GJ (2005) Analysis of scanning force microscopy images of protein-induced DNA bending using simulations. Nucleic Acids Res 33(7):e68. https://doi.org/10.1093/nar/gni073
31. Sitters G, Kamsma D, Thalhammer G, Ritsch-Marte M, Peterman EJ, Wuite GJ (2015) Acoustic force spectroscopy. Nat Methods 12 (1):47–50. https://doi.org/10.1038/nmeth.3183
32. Yang SW, Nash HA (1995) Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target. EMBO J 14(24):6292–6300
33. van Noort J, Verbrugge S, Goosen N, Dekker C, Dame RT (2004) Dual architectural roles of HU: formation of flexible hinges and rigid filaments. Proc Natl Acad Sci U S A 101 (18):6969–6974. https://doi.org/10.1073/pnas.0308230101
34. Giphart-Gassler M, Goosen T, van Meeteren A, Wijffelman C, van de Putte P (1979) Properties of the recombinant plasmid pGP1 containing part of the early region of bacteriophage mu. Cold Spring Harb Symp Quant Biol 43(Pt 2):1179–1185
35. Holbrook JA, Tsodikov OV, Saecker RM, Record MT Jr (2001) Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. J Mol Biol 310(2):379–401. https://doi.org/10.1006/jmbi.2001.4768
36. Chenouard N, Smal I, de Chaumont F, Maska M, Sbalzarini IF, Gong Y, Cardinale J, Carthel C, Coraluppi S, Winter M, Cohen AR, Godinez WJ, Rohr K, Kalaidzidis Y, Liang L, Duncan J, Shen H, Xu Y, Magnusson KE, Jalden J, Blau HM, Paul-Gilloteaux P, Roudot P, Kervrann C, Waharte F, Tinevez JY, Shorte SL, Willemse J, Celler K, van Wezel GP, Dan HW, Tsai YS, Ortiz de Solorzano C, Olivo-Marin JC, Meijering E (2014) Objective comparison of particle tracking methods. Nat Methods 11(3):281–289. https://doi.org/10.1038/nmeth.2808
37. Rogers SS, Waigh TA, Zhao X, Lu JR (2007) Precise particle tracking against a complicated background: polynomial fitting with Gaussian weight. Phys Biol 4(3):220–227. https://doi.org/10.1088/1478-3975/4/3/008
38. Visser EW, van ILJ, Prins MW (2016) Particle motion analysis reveals nanoscale bond characteristics and enhances dynamic range for biosensing. ACS Nano 10(3):3093–3101. https://doi.org/10.1021/acsnano.5b07021
39. Driessen RP, Sitters G, Laurens N, Moolenaar GF, Wuite GJ, Goosen N, Dame RT (2014) Effect of temperature on the intrinsic flexibility of DNA and its interaction with architectural proteins. Biochemistry 53(41):6430–6438. https://doi.org/10.1021/bi500344j

Chapter 15

Observing Bacterial Chromatin Protein-DNA Interactions by Combining DNA Flow-Stretching with Single-Molecule Imaging

HyeongJun Kim and Joseph J. Loparo

Abstract

Nucleoid-associated proteins bind to DNA specifically and nonspecifically to perform various roles in chromosome organization and segregation. In this chapter, we describe how the interaction between nucleoid-associated proteins and flow-stretched DNAs can be visualized on the single-molecule level. We describe three different experimental schemes that allow one to directly observe how these proteins that make up bacterial chromatin, associate with and act on DNAs. First, we describe how to visualize the diffusion of fluorescently labeled proteins on flow stretched DNAs. Second, we describe how the binding of bacterial chromatin proteins can be correlated with DNA condensation. Lastly, we describe the DNA motion capture assay, which allows one to probe the mechanism of DNA condensation by tracking how different segments of a flow stretched DNA are compacted by bacterial chromatin proteins.

Key words DNA-protein interactions, Single-molecule, Fluorescence imaging, Flow-stretching, DNA motion capture

1 Introduction

Across all domains of life, organisms compact their genomes by orders of magnitude to fit them within the volume of the cell, and to provide functional organization. In bacteria, different kinds of DNA-binding proteins, including nucleoid-associated proteins (NAPs), are involved in condensing, organizing, and segregating DNAs. Two classes of NAPs can be distinguished based on the mechanism by which they condense DNA: (1) DNA benders, which bend DNA, and (2) DNA bridgers which can span distal DNA segments, generating DNA loops on the order of kilobases [1, 2].

Directly probing how these proteins carry out their functions is difficult with conventional ensemble approaches. While well-established approaches, such as electrophoretic mobility shift assay (EMSA) or surface plasmon resonance, can be used to determine

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_15,

protein-DNA affinities, they cannot be used to establish how DNA conformation changes upon protein binding. Over the past two decades, single-molecule technologies have provided new opportunities to investigate bacterial chromatin organization with enhanced experimental detail. These approaches generally are force- or fluorescence-based. In force-based experiments, DNA is often tethered to a bead or a coverslip and stretched by force; the action of DNA-binding proteins is observed as changes in DNA extension upon protein binding. Here, force itself is also an experimental parameter. In fluorescence-based experiments, the position and distribution of proteins on DNA is directly visualized. Despite the success of force and fluorescence-based approaches, employing them individually limits the information that one is able to obtain. For instance, single-molecule magnetic tweezers experiments can detect DNA compaction by chromatin proteins, but cannot give insight into the distribution of the protein along the DNA. In this chapter, we present protocols focusing on combining DNA flow-stretching (force-based) with protein visualization (fluorescence) to directly correlate protein binding with DNA structuring.

First, we describe how to determine the diffusive properties of individual fluorescently labeled proteins on flow-stretched DNA. Second, we present a two-color imaging approach that enables simultaneous observations of labeled proteins bound to DNA and measurements of DNA length. Lastly, we describe a DNA motion capture assay where specific sites of lambda DNA are bound by a catalytically dead mutant of EcoRI (EcoRI E111Q) labeled with quantum dots. Tracking of the quantum dot position enables one to observe how different segments within the flow stretched DNA are remodeled by chromatin proteins. This assay exploits the differential tension along the DNA length created by flow to distinguish between DNA bending and bridging [3, 4]. DNA bridgers trap DNA loops that spontaneously form. Given that loop formation occurs more readily at low DNA tensions, DNA compaction occurs from the free end of the DNA. On the other hand, since DNA benders/wrappers are relatively insensitive to the force that DNA experiences and the effect of DNA reconfiguration by each protein occurs locally, DNA compaction occurs simultaneously along the length of DNA.

2 Materials

2.1 Surface-Passivated Coverslips

1. Coverslip (VWR 16004-312).
2. Polypropylene coplin staining jar (Bel-Art F44208-1000).
3. Ethanol (200 proof).
4. 1 M potassium hydroxide.

5. Sonicator.
6. Laboratory oven.
7. (3-Aminopropyl)triethoxysilane (Sigma-Aldrich A3648 or equivalent).
8. mPEG (Laysan, mPEG-SVA-5000 or equivalent).
9. Biotin-PEG (Laysan, biotin-PEG-SVA-5000 or equivalent).
10. 100 mM sodium bicarbonate.

2.2 Various DNA Substrate

2.2.1 Bacteriophage Lambda DNA with a Biotin at One End

1. Lambda-BL1Biotin: agg tcg ccg ccc/3′BioTEG/ (Stock concentration: 100 μM).
2. Lambda-BL2Biotin: ggg cgg cga cct/3′BioTEG/ (Stock concentration: 100 μM).
3. 10× ligase reaction buffer (NEB B0202S or equivalent).
4. T4 polynucleotide kinase (PNK) (NEB M0201S or equivalent).
5. λ-DNA (NEB N3013S or equivalent).
6. T4 DNA ligase (NEB M0202S or equivalent).
7. 100 mM ATP stock.
8. Heat block.

2.2.2 Bacteriophage Lambda DNA with a Biotin at One End and Digoxigenin at the Other End

1. Lambda-BL2Biotin: ggg cgg cga cct/3′BioTEG/ (Stock concentration: 100 μM).
2. Lambda-Dig: agg tcg ccg ccc aaa aaa aaa aaa/3′Digoxigenin/ (Stock concentration: 100 μM).
3. Same as Subheading 2.2.1, **items 3–8**.
4. Agarose.
5. 3 M Sodium acetate (NaOAc).
6. Ethanol (70% and 100%).
7. TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA).
8. Gel apparatus and gel loading dye.
9. Centrifuge.
10. NanoDrop UV-Vis spectrophotometer (Thermo Fisher Scientific or equivalent instrument).

2.3 Antibody-Conjugated Quantum Dot

1. Anti-digoxigenin fab fragments: Roche 11214667001 or equivalent.
2. Anti-His-tag antibody: MCA1396 (Bio-Rad) or equivalent.
3. Quantum dot conjugation kit: Thermo Fisher Scientific Q22001MP (for Qdot 605) or Q22061MP (for Qdot 705) or equivalent.

2.4 His6-EcoRI E111Q Mutant

1. Plasmid pEQ111M: Prepared by the laboratory of Paul Modrich, and is available from Addgene (plasmid #: 40190).
2. EcoRI_NheI_F: ac gctagc tctaataaaaaacagtcaaataggctaactg (NheI restriction site: gctagc).
3. EcoRI_EcoRI_R: ac gaattc tcacttagatgtaagctgttcaaacaag (EcoRI restriction site: gaattc, a stop codon is included).
4. PCR apparatus.
5. pET28b (Novagen).
6. Ligase buffer and ligase.
7. BL21 cells.
8. Buffer R: 50 mM Tris–HCl, pH 7.5, 300 mM NaCl, 10% sucrose, 0.1 mM PMSF (phenylmethylsulfonyl fluoride), 5 mM 2-Mercaptoethanol, 10 mM imidazole, protease inhibitor cocktail (Roche).
9. Buffer S: 60 mM Tris–HCl, pH 7.5, 450 mM NaCl, 15 mM 2-Mercaptoethanol, 0.15 mM EDTA, 10% glycerol, 0.21% Triton X-100.
10. Ni-NTA resin.
11. Glycerol.

2.5 Protein Labeling with a Fluorophore

1. TCEP (tris(2-carboxyethyl)phosphine).
2. Cy3 maleimide.

2.5.1 Labeling of Cysteine Residues

2.5.2 Labeling of Primary Amines

1. Cy3 NHS-Ester dyes.

2.6 Microscope Setup

1. Microscope body frame (Olympus IX-71 or equivalent).
2. EMCCD camera or CMOS camera.
3. 532 nm laser (Coherent Sapphire 532-75 CW CDRH or equivalent), 641 nm laser (Coherent Cube 640-100C (part number: 1150205) or equivalent).
4. Long-pass dichroic mirror that allows transmission for 641 nm laser but blocks 532 nm laser.
5. Beam expander: Three pairs of two lenses that generate expanded collimated beam. Two ACN127-025-A, two AC254-200-A, an AC254-040-A, and an AC508-150-A (Thorlabs).
6. Translation stages.
7. Two shutters (Uniblitz VS14 or equivalent) and a shutter controller (Uniblitz VMM-D3 or equivalent).

8. A filter cube consisting of a dichroic mirror, excitation, and emission filters. The dichroic mirror reflects the incoming laser beam but transmits fluorescence emission signal from the sample. Excitation and emission filters are needed to select only desired wavelength ranges for incoming laser light and outgoing fluorescence signal, respectively.

2.7 Air Spring

1. 50 mL tube.
2. Drill and drill bit (1.2 mm Flame, 3/32″ Shank).
3. PE60 tube.
4. Epoxy.

2.8 Other Buffers

1. EcoRI Binding Buffer (EBB buffer): 10 mM Tris pH 8.0, 150 mM NaCl, 10 mM $MgCl_2$.
2. EBB+BSA buffer: 10 mM Tris pH 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 0.2 mg/mL bovine serum albumin.
3. EcoRI Binding Buffer 2 (EBB2 buffer): 20 mM Tris, pH 7.5, 100 mM NaCl.

3 Methods

Here we describe three different single-molecule assays to study the interaction between bacterial chromatin proteins and DNA. All single-molecule experiment schemes mentioned in this chapter require a total internal reflection fluorescence (TIRF) microscope and rely on stretching tethered DNAs by flow. Depending on the specific research purpose, refer to the relevant sections as suggested in Table 1.

3.1 Coverslips Surface Passivation

3.1.1 Coverslip Cleaning

1. If dust is visible on the coverslip surface, gently blow it away using either a dust remover spray or air flow before placing it into a jar (*see* **Note 1**).
2. Place coverslips into a polypropylene coplin staining jar. The jar design is such that the interior space of the jar is separated into multiple sub-spaces via spacers (*see* **Note 2**). Make sure that there is not more than one coverslip in each sub-space.
3. Fill the polypropylene jar with ethanol (200 proof), and sonicate it for 30 min.
4. Discard ethanol into an appropriate chemical waste bottle. Thoroughly rinse the polypropylene jar with MilliQ water five times. Discard water.
5. Fill the polypropylene jar with 1 M potassium hydroxide (KOH), and sonicate it for 30 min.

Table 1
Overview of research schemes

Aim	Measuring diffusive properties of single proteins on DNA	Measuring DNA remodeling activity by chromatin proteins	DNA-motion capture to resolve DNA compaction
Suggested DNA substrate	*See* Subheading 3.2.1	*See* Subheading 3.2.2	*See* Subheading 3.2.1
Methods	*See* Subheading 3.8	*See* Subheading 3.9	*See* Subheading 3.10
Protein labeled?	Yes (*see* Subheading 3.5)	Yes (*see* Subheading 3.5) but not required	No
DNA labeled?	No	Yes (at the free end with a quantum dot)	Yes (at specific internal positions with multiple quantum dots)
Common	(1) Coverslip surface passivation: *see* Subheading 3.1 (2) Flowcell construction: *see* Subheading 3.6 (3) Microscope setup: *see* Subheading 3.7		

6. Discard 1 M KOH into an appropriate chemical waste bottle (*see* **Note 3**). Thoroughly rinse the polypropylene jar with MilliQ water five times. Discard water.
7. Repeat the ethanol (**steps 3** and **4**) and KOH (**steps 5** and **6**) wash steps. Store washed coverslips in MilliQ water (*see* **Note 4**).

3.1.2 Coverslip Silanization

1. Preheat a laboratory oven to 110 °C.
2. It is necessary to remove any traces of water from the polypropylene jars storing the cleaned coverslips prior to silanizing the coverslip surfaces (*see* **Note 5**). First, decant water and fill the jar with acetone. Discard acetone into a chemical waste bottle by tilting the jar. At this time, slowly rotate the jar to let the acetone touch the whole interior area of the jar, ensuring complete water removal. Separately, rinse the jar lid with acetone. Repeat this step two more times.
3. During the third wash, sonicate the coverslips in acetone for 10 min. Discard acetone.
4. Prepare ~2% (v/v) silane solution by mixing 1.1 mL of (3-Aminopropyl)triethoxysilane with 55 mL of acetone (for each jar) (*see* **Note 6**). Pour the solution immediately into the jar, close the lid, and move the jar horizontally in all directions on a bench top surface for 2 min to ensure complete silanization.
5. Discard the silane solution into an appropriate chemical waste bottle, and quench the reaction by rapidly pouring 1–2 L of MilliQ water into the jar.

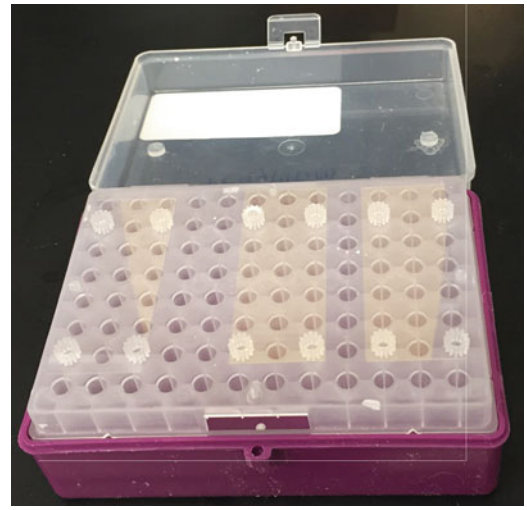


Fig. 1 Silane-treated coverslips on a pipet tip box. Add some water inside the pipet tip box (in the purple box under white pipet tip platform)

6. Rinse each coverslip thoroughly with MilliQ water before placing them into the lab oven. Let the coverslips dry for 30 min at 110 °C.
7. Place each coverslip onto a pipet tip box (*see* Fig. 1) in preparation for the PEGylation step.

3.1.3 Coverslip PEGylation

1. mPEG and biotin-PEG bottles have been stored at −20 °C since their receipt. Before their use, let the bottles reach ambient temperature, which typically takes 20–30 min.
2. Prepare fresh 100 mM sodium bicarbonate buffer (210 mg powder for 25 mL water).
3. Prepare a tube containing 6.8 mg of biotinylated PEG and 170.0 mg of mPEG (*see* **Notes 6** and **7**). This step can be done while silane-treated coverslips are being baked at 110 °C for 30 min.
4. Add 1100 μL of 100 mM sodium bicarbonate buffer to the PEG mixture and thoroughly mix it right before applying it to the silane-treated coverslips to minimize hydrolysis. Drop 100 μL of the PEG solution onto each silane-treated coverslip and gently place another silane-treated coverslip on the top of it, leading to the PEG solution being sandwiched between two coverslips.
5. In a dark place, leave the PEG-treated coverslips in the closed pipet tip box for 3–4 h. Make sure that a small amount of water is in the bottom of the box to ensure that the coverslips do not dry out.

6. Gently peel off the sandwiched coverslips and wash them using MilliQ water. Next, dry the PEGylated coverslips with compressed air.
7. Store the PEG-functionalized coverslips under vacuum. They should be stable for at least a month (*see* **Note 8**).

3.2 DNA Substrate Preparation

3.2.1 Bacteriophage Lambda DNA with a Biotin at One End

Bacteriophage lambda DNA has 12-base complementary single-strand 5′ overhangs. In this section, we describe how one of these overhangs can be labeled with a biotin-containing oligo, enabling tethering of the DNA onto the PEGylated coverslip.

1. Depending upon which end of the lambda DNA is intended to be labeled with a biotin, either Lambda-BL1Biotin or Lambda-BL2Biotin oligo will be used in the subsequent steps.
2. Phosphorylate either Lambda-BL1Biotin or Lambda-BL2Biotin oligo.
   (a) Combine the following:
   15.5 μL ddH_2O.
   2.0 μL 10× ligase reaction buffer.
   0.5 μL T4 polynucleotide kinase (PNK).
   2.0 μL of 100 μM oligo (Final concentration of the phosphorylated oligo: 10 μM).
   (b) Incubate the mixture at 37 °C for 3 h.
3. Annealing the phosphorylated oligo with λ-DNA
   (a) Combine the following:
   200 μL ddH_2O.
   25 μL 10× ligase reaction buffer.
   0.4 μL of the phosphorylated 10 μM oligo (10× more oligo than λ-DNA).
   (b) Gently add 25 μL λ-DNA (*see* **Note 9**).
   (c) Heat the mixture at 65 °C for 10 min, and then slowly cool it down to the room temperature by turning off the heat block and leaving the tube in it.
4. Ligation step
   (a) To the λ-DNA with annealed oligo prepared in **step 3**, add 1.5 μL of T4 DNA ligase and 2.0 μL of 100 mM ATP.
   (b) Leave the mixture at room temperature for 2 h.
   (c) The DNA substrate is now ready to be used in single-molecule experiments (*see* **Note 10**).

3.2.2 Bacteriophage Lambda DNA with a Biotin at One End and Digoxigenin at the Other End

The biotinylated end will be used in tethering DNA onto a glass surface via biotin-neutravidin (or biotin-streptavidin) interaction. The end with digoxigenin will be used for labeling with a quantum dot.

1. Phosphorylate Lambda-BL2Biotin and Lambda-Dig oligos in separate tubes as described in Subheading 3.2.1, **step 2**.
2. Annealing the phosphorylated Lambda-Dig oligo with λ-DNA (Annealing step #1).
   (a) Combine the following:
   375 μL ddH_2O.
   50 μL 10× ligase reaction buffer.
   1.80 μL of the phosphorylated 10 μM Lambda-Dig oligo. (15× more Lambda-Dig oligo than λ-DNA).
   (b) Gently add 75 μL λ-DNA (*see* **Note 9**).
   (c) Heat the mixture at 65 °C for 10 min, and then slowly cool it down to the room temperature by turning off the heat block and leaving the tube in it.
3. Ligation step #1
   (a) Add 3.0 μL of T4 DNA ligase and 4.0 μL of 100 mM ATP.
   (b) Leave the mixture at room temperature for 2 h.
4. Annealing the phosphorylated Lambda-BL2Biotin oligo with λ-DNA (Annealing step #2)
   (a) Add 7.20 μL of the phosphorylated 10 μM Lambda-BL2Biotin oligo to the mixture from **step 3** (60× more Lambda-BL2Biotin oligo than λ-DNA).
   (b) Heat the mixture at 45 °C for 30 min, and then slowly cool it down to the room temperature by turning off the heat block and leaving the tube in it.
5. Ligation step #2
   (a) Add 3.0 μL of T4 DNA ligase and 4.0 μL of 100 mM ATP to the reaction mixture.
   (b) Leave it at room temperature for 2 h.
6. Removing unconjugated excess oligos

   The Lambda-BL2Biotin oligo will be used for DNA-surface tethering and the Lambda-Dig oligo will be used for a quantum-dot labeling. Since these two oligos are complementary, it is crucial to remove excess oligos prior to introducing the DNA substrate into the flowcell. Otherwise, quantum-dot-bound short oligo duplexes will be tethered to the flowcell surface, obstructing imaging of the full-length DNA products. The long length of the λ-DNA makes it susceptible to shearing. Therefore, we suggest the gentle dialysis bag-based oligo removal procedure described below.

(a) Combine multiple teeth of a gel comb by taping them together, making a ~5 cm tooth. Place a thin layer of epoxy on the tape to make the tooth of the gel comb thicker. The resultant gel well should accommodate at least 600 μL sample (500 μL of DNA sample + 100 μL of gel loading dye).

(b) Prepare 0.4% agarose gel containing ethidium bromide. Use the gel comb prepared in (**step 1**).

(c) Add 1/5 volume of gel-loading dye to the sample and mix.

(d) Load the sample to the gel well.

(e) Apply high voltage (~90 V) for 10 min to allow all DNA in the well to get into the gel. Then, apply low voltage (15–25 V) overnight.

(f) Using a razor blade, cut the lambda DNA band, and put it in a dialysis bag. (The dialysis bag might need to be hydrated before use.) Add the same buffer for gel electrophoresis to the bag until the gel is submerged. Clip the bag and put it into the gel apparatus. Apply voltage (35–50 V) until all the DNA elutes from the gel. Then, reverse voltage polarity and apply ~50 V for 2 min (*see* **Note 11**).

(g) Gently draw up the solution from the bag and put it in a 15 mL centrifuge tube (*see* **Note 9**).

7. Recovery of purified DNA

(a) Add to the DNA solution 1/10 volume of 3 M NaOAC and then 2.5 volume of ethanol.

(b) Leave it either in a −20 °C freezer or in a 4 °C refrigerator overnight (*see* **Note 12**).

(c) Spin it with maximum speed (~7200 × *g* in case of Eppendorf 5430R centrifuge) for 10 min. Discard solution.

(d) Wash the DNA precipitate with 70% ethanol and spin it for 5 min with maximum speed. Discard solution.

(e) Air-dry the DNA and resuspend it with TE buffer or a buffer of your choice. Measure the DNA concentration. We usually resuspend in 200–250 μL of buffer which typically yields 40–60 ng/μL of DNA.

3.3 Anti-digoxigenin (or Anti-His-Tag) Antibody-Conjugated Quantum Dot

Conjugating quantum dots with antibody is performed per quantum dot manufacturer's protocol.

3.4 His6-EcoRI E111Q Mutant

DNA motion capture experiments require His6-EcoRI E111Q, a catalytically inactivated derivative of EcoRI for specific labeling of DNA with anti-His-tag antibody-conjugated quantum dots.

1. The starting plasmid pEQ111M was prepared by the laboratory of Paul Modrich.
2. Perform PCR for the EcoRI-E111Q region of the plasmid pEQ111M using primers EcoRI_NheI_F and EcoRI_EcoRI_R.
3. Digest the insert with NheI and EcoRI.
4. Ligate the insert into pET28b that is also cut by NheI and EcoRI. This generates pTG004.
5. Transform BL21 cell with pTG004, and grow cells in the presence of kanamycin.
6. Sonicate the BL21 cell pellet, obtained by centrifugation, in buffer R.
7. Apply the supernatant of the lysate to Ni-NTA resin for affinity chromatography, then, wash it with buffer R supplemented with 1 M NaCl. Wash it again with buffer R supplemented with 1 M NaCl and 50 mM imidazole.
8. Elute with buffer R supplemented with 250 mM imidazole that does not contain protease inhibitors.
9. Eluates were dialyzed against buffer S.
10. Add 1 volume of glycerol to 2 volumes of dialysate for storage and mix well.

3.5 Labeling Proteins with Fluorophores for Single-Molecule Imaging

Numerous strategies beyond the scope of this chapter exist to label proteins both specifically and nonspecifically [5–7]. Here we briefly describe two straightforward approaches to label a protein with a dye molecule.

1. Labeling of cysteine residues

   Surface-exposed Cysteine residues can be labeled with thiol-reactive dyes such as fluorophore-maleimide conjugates. We suggest following the manufacturer's instructions except with one caveat. Some protocols suggest that using up to 100× molar excess of TCEP (tris(2-carboxyethyl)phosphine) to protein concentration is fine. However, we have noticed that using high concentrations of TCEP can decrease protein labeling efficiency. We suggest using a TCEP concentration comparable to that of protein.

2. Labeling of primary amines

   NHS-Ester dyes can be readily conjugated to lysine residues or to the N-terminus of proteins. Follow the manufacturer's instructions for protein labeling.

3. Removing unreacted fluorescent dyes from the labeled protein reaction mix

Following the completion of the labeling reaction, unreacted dye must be removed so it does not contribute to background in imaging experiments.

Remove unreacted dyes following the fluorescent dye manufacturer's instruction or using a concentrator of an appropriate molecular weight cutoff that allows for removal of the free dye while retaining the protein (*see* **Note 13** for a caveat). Size-exclusion chromatography is also commonly employed to remove free dye.

4. Activity test for the fluorescently labeled protein

   It is essential to verify that the protein labeling has not affected the protein activity. For DNA-binding proteins that remodel DNA structure we recommend the procedures described in Subheading 3.9. Measure the DNA compaction rates by tracking the positions of a quantum dot labeled at the end of DNA, and confirm that they are identical or similar for both the labeled and unlabeled proteins.

3.6 Construction of Flow Cell for DNA Flow Stretching Experiments

This section describes how to construct a microfluidic flow cell to use in DNA flow stretching experiments. The bottom surface of the flow cell is a functionalized coverslip (described in Subheading 3.1) that captures DNA molecules. Here we describe the construction of a 2-channel flow cell but readers can readily adapt it depending on the number of channels they need.

1. Make four holes on a 20 mm × 25 mm × 1 mm quartz plate (Technical Glass Product) using a drill. These holes are where inlet and outlet polyethylene tubes (PE number: 60, wall thickness: 0.009″) will be inserted. The hole size should be very similar to PE60 tubing diameter to ensure that the tubing sits firmly in the hole.
2. Draw the flowcell layout as shown in Fig. 2a, and paste it on one side of double-sided tape cover (Grace Bio-labs 620001). The thickness of the double-sided tape (0.12 mm) will be the thickness (or height) of the flowcell. Each channel in the layout is 1.8 mm wide.
3. Using a razor blade or a scalpel, cut off flow cell channel regions (red and light green rectangles in Fig. 2a). Peel off the cover of the tape, and attach it on the quartz plate (*see* Fig. 2b). Make sure that the holes are located within the flow cell channel regions.
4. The PEGylated coverslip (24 mm × 60 mm) is unnecessarily bigger than the quartz plate (20 mm × 25 mm). Using a diamond scriber (VWR 52865-005), score the coverslip in the middle, and cut it by halves, generating two pieces of 24 mm × 30 mm coverslips (*see* **Note 14**).

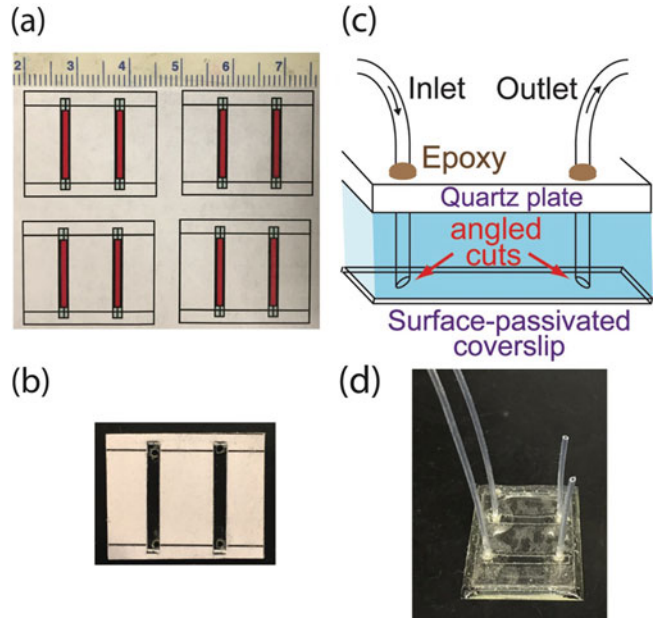


Fig. 2 (**a**) Flow cell layout, (**b**) Quartz plate with layout attached on it, (**c**) Schematic of flow cell. Note that inlet and outlet tubes were cut with ~30° angle, (**d**) Completed flow cell image

5. Peel off the remaining tape cover and put the surface-passivated coverslip on it. It is very common that air is trapped between the double-sided tape and coverslip. Using a flat-ended tweezer (Style: K35A, Electron Microscopy Sciences 78193-01 or equivalent) or the bottom of a microcentrifuge tube, eliminate the trapped air by applying pressure to the coverslip surface (*see* **Note 15**).
6. Prepare the inlet (7 cm long) and outlet (1–3 cm long) tubing. Cut one end of each tube at an ~30° angle (*see* Fig. 2c) to ensure smooth flow in the microfluidic channel area.
7. Apply epoxy around all four edges of the quartz plate at the junction between the quartz plate and the coverslip. Also, apply epoxy around the tubing insert sites (*see* Fig. 2d and **Note 16**).

3.7 TIRF Microscope Setup

3.7.1 Overall Scheme for a Microscope

1. Our home-built through-objective total internal reflection fluorescence (TIRF) microscope was built with an Olympus IX-71 microscope body frame (*see* **Note 17**). First, install 532 and 641 nm lasers on an optical table (*see* **Note 18**). Use three mirrors (M1, M2, and M3) to direct laser light toward desired directions as shown in Fig. 3a.
2. Install a long-pass dichroic mirror (LPD) that allows the laser beam with the longer wavelength (641 nm) to pass through while reflecting the other laser beam (532 nm) (Fig. 3a).

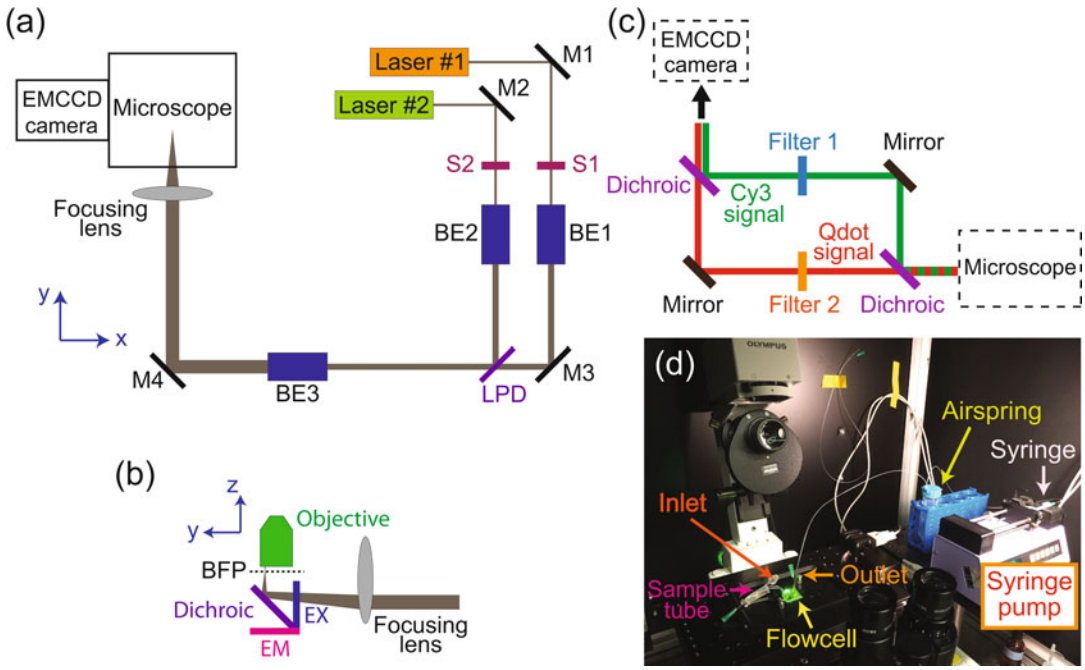


Fig. 3 (**a**) Schematic of our microscope setup. *M* mirror, *BE* beam expander, *LPD* long-pass dichroic mirror, *S* shutter, (**b**) TIRF angle (critical angle) realization. *BFP* back focal plane, *EX* excitation filter, *EM* emission filter. (**c**) Schematic of home-built custom dual-view for two-color imaging. (**d**) Actual microscope setup

3. Use another mirror (M4) to send the combined beam to the center of a microscope objective. The pointing of the laser beams can be controlled by adjusting the kinematic mirror mounts (M1–M4).
4. Build up beam expanders (BE1 and BE2) in front of each laser (*see* **Note 19**), generating expanded collimated beams.
5. Install another beam expander (BE3) to further expand the combined beam.
6. Place table-mounted shutters in the beam path (S1 and S2) to enable alternating laser excitation.
7. (*See* Fig. 3b) Install the focusing lens by mounting it on both horizontal (moving along the *y* axis) and vertical (moving along the *z* axis) translation stages.
8. Move the horizontal stage until the beam is focused on the back focal plane (BFP) of the microscope objective. When the lens is properly positioned the beam should be collimated upon passing through the objective lens. Adjust the vertical position of the focusing lens with the vertical stage to direct the beam onto the sample at the TIRF angle (critical angle).

9. (*See* Fig. 3c) In cases where two different fluorophores have to be imaged simultaneously, as in Subheading 3.9, an additional imaging module, known as a Dual-view, must be set up between the microscope and the EMCCD camera. Signals from two different emission wavelengths are optically separated by a dichroic mirror into two independent optical paths. The separated emission signals are then recombined with another dichroic mirror and directed onto separate halves of an EMCCD camera. Emission filters are inserted in each emission signal beam path (*see* **Note 20**).

3.7.2 Construction of an Air Spring to Minimize Flow Fluctuations

An air spring is connected between the outlet of the flow cell and the syringe pump, and dampens fast flow fluctuations. To make an air spring:

1. Fill a 50 mL conical tube with water so that only 2–5 mL air remains in the tube.
2. Make two holes in the 50 mL tube lid, whose sizes just fit PE60 tubing. Insert sufficiently long PE60 tubes into these holes so that one can connect to the outlet tubing of the flow cell and the other can connect to the syringe pump.
3. Seal the air spring by applying epoxy to the lid where the tubing is inserted, and the base of the lid of the 50 mL conical tube.
4. Connect the air spring to the flow cell and to the syringe pump.
5. In sum, the inlet tubing of the flow cell is inserted into a protein/DNA sample solution, and the outlet tubing is connected to the air spring. The other outlet of the air spring is connected to the syringe of the syringe pump (*see* Fig. 3d). When the syringe pump draws a syringe plunger, the protein/DNA sample solution is pulled through the flow cell.

3.8 Measuring Protein Diffusion on Individual DNAs

In this section, we describe how one can measure the diffusive properties of individual fluorescently labeled proteins on flow-stretched DNAs.

3.8.1 Experimental Procedure

1. Using a syringe, introduce 50–100 μL of 0.2–0.25 mg/mL streptavidin or neutravidin in a flow cell through the inlet tubing. Incubate for 5 min.
2. Drop microscope immersion oil on the objective and mount the flow cell. Connect the outlet tubing to a syringe pump through an air spring.
3. Wash the flow cell with EBB+BSA buffer or a buffer of your choice by quickly pulling the syringe plunger in the syringe pump. Prior to inserting the inlet tubing into the microcentrifuge tube containing the buffer, gently lift the air spring to produce a drop of solution from the inlet tube. Then rapidly

insert the inlet tube into the wash buffer and lower the air spring (*see* **Note 21**).

4. Dilute the biotinylated lambda DNA substrate using EBB+BSA buffer or a buffer of your choice, and flow in 150–200 μL at a rate of 0.03–0.05 mL/min. This step requires empirical optimization to achieve a density of tethered DNAs that allows for sufficient sampling while maintaining a reasonable distance between DNAs (*see* **Note 22**).
5. Flow in your sample buffer at a rate of 0.05 mL/min to remove any unbound DNAs and exchange buffers.
6. Flow in your protein sample (in sample buffer of your choice) at a rate of 0.05 mL/min and record a movie (*see* **Note 23**). If fluorophore photostability is a concern, supplement your buffer with an oxygen scavenging system (*see* **Note 24**).
7. Flow in ~10 nM Sytox Orange at a rate of 0.05 mL/min at the end of the experiment to directly visualize the tethered DNAs (*see* **Note 25**).

3.8.2 Data Analysis

1. Open up the raw image file with Fiji [8] or equivalent software, and identify regions of interest where you can observe fluorescent spots undergoing one-dimensional diffusion along DNA.
2. Track the fluorescent signal in a region of interest over time by fitting the raw signal image with two-dimensional Gaussian curve using our custom-written Matlab code (MathWorks) (based on a built-in function "lsqcurvefit") or a built-in function "GAUSS2DFIT" in IDL (Harris Geospatial Solutions). The center of the fitted Gaussian curve is approximately the position of the dye. This analysis will yield trajectories of dye position as a function of time.
3. Generate a net displacement histogram to see if the presence of flow biases protein diffusion in the flow direction. The net displacement is the position difference between initial fluorescence signal appearance and its final disappearance. If the protein moves via a Brownian diffusion mechanism with no flow bias, the net displacement histogram center should be located at zero.
4. If the net displacement histogram is not centered at zero, then the diffusion coefficient should be calculated after correcting for the flow bias as shown in Tafvizi et al. [9].
   (a) Concatenate all the position trajectories as if they were part of a single long trajectory. The end-to-end distance divided by total time is the average velocity of the concatenated trajectory that represents the flow bias.

(b) Subtract the effect of flow bias estimated by the average velocity of the concatenated trajectory from the individual raw trajectories.

5. Calculate the mean square displacement (MSD) for each corrected trajectory using:

$$\mathrm{MSD}\,(N,n) = \frac{\sum_{i-1}^{N-n}\left(y_{i+n} - y_i\right)^2}{N-n}$$

(a) y: position coordinate along the DNA.

(b) N: total number of data points for each trajectory from the appearance of the signal to its disappearance.

(c) n: from 1 to N.

6. Draw a graph of "MSD(N, n) versus $n\Delta t$" where Δt is EMCCD camera acquisition duration. Fit the graph to "MSD (N, n) = $2Dn\Delta t$". Here, D is a diffusion coefficient obtained from each individual trajectory (*see* **Note 26**).

3.9 Measuring DNA Structuring Activity by Bacterial Chromatin Proteins

Here, we describe an experimental scheme that enables simultaneous observation of spectrally distinct labeled proteins (e.g. Cy3-labeled protein) associating with an end labeled (e.g. quantum dot 705) lambda DNA (*see* **Note 27**). Also *see* **Note 28** for alternative imaging approaches.

3.9.1 Experimental Procedure

1. Add 0.3 μL anti-digoxigenin antibody-conjugated quantum dot 705 and 2.7 μL of bacteriophage lambda DNA with a biotin at one end and digoxigenin at the other end to 17.0 μL EBB buffer or a buffer of your choice. Allow 15–30 min for incubation at room temperature.
2. Apply 0.2–0.25 mg/mL streptavidin or neutravidin to the flow cell. After 5 min wash the flow cell. For more details *see* Subheading 3.8.1, **steps 1–3**.
3. Dilute the quantum-dot conjugated DNA by 20-fold using EBB+BSA buffer or a buffer of your choice (*see* **Note 29**).
4. Flow in your sample buffer at a rate of 0.05 mL/min in order to remove unbound quantum dot-conjugated DNAs and exchange buffers.
5. Turn off the flow. After 2 min, start recording a movie in the absence of any flow. In the middle of the movie when ~100 frames are recorded, turn on the buffer flow which will stretch

the DNAs. Keep recording the movie for additional 100–200 frames (*see* **Note 30**).

6. Flow in your protein sample at a rate of 0.05 mL/min and record a movie (*see* **Note 31**).

3.9.2 Data Analysis

1. Open up the raw image file with Fiji [8] or equivalent software, and select regions of interest (ROI) for both Cy3 and quantum dot channels (*see* **Note 32**).
2. Import the list of ROI into custom-written Matlab codes that perform Gaussian fitting on the signal and determine the positions of the quantum dot over time (*see* **Note 33**). Then, perform Gaussian fitting of the quantum dot signal to determine the positions of the quantum dot over time.
3. Correlate the distribution of Cy3-labeled proteins with the quantum dot positions.

3.10 DNA Motion Capture Assay

In this section, we describe how to carry out DNA motion capture experiments.

3.10.1 Experimental Procedure

1. Add 0.4 μL of anti-His-tag antibody-conjugated quantum dot 605 (stock concentration: usually ~1 μM) and 0.4 μL of 10-×-diluted His6-EcoRI E111Q mutant (final concentration: ~2 μM) to 16.7 μL of EBB2 buffer, and leave it at room temperature for 30 min.
2. Add 2.5 μL of DNA substrate that is tagged with Lambda-BL1Biotin oligo at its one end to it and incubate for 30 min at room temperature.
3. Flow in the DNA and quantum dot mixture into a microscope mounted flow cell as described in Subheading 3.8.1, **step 4**. Then, wash out the flow cell, and flow in the protein at a rate of 0.05 mL/min. Note the following differences.
   (a) The Dual-view part in the microscope setup is not needed.
   (b) Protein is not fluorescently labeled.
   (c) The buffer composition used in diluting the DNA and quantum dot mixture is EBB2 buffer.

There are five EcoRI-binding sites in the lambda genome. We typically observe three to five quantum dots on individual DNAs.

3.10.2 Data Analysis

1. Up to five quantum dots are labeled on a DNA. Track each quantum dot with two-dimensional Gaussian fitting. Alternatively, the five quantum dots can be simultaneously tracked by U-track software (Harvard University), following Jaqaman et al. [10].

4 Notes

1. This step can be omitted if the coverslip appears clean by visual inspection.
2. In this configuration, one side of either the first or the last coverslip faces the interior jar surface.
3. At this time, flip the first and the last coverslips and let the other side of each coverslip face the interior jar surface. It will ensure that both surfaces of the first and the last coverslips are cleaned during the cleaning procedures.
4. If you do not have enough time to complete the remaining functionalization steps, you may stop here, leaving the coverslips in MilliQ water filled jars for up to a few days.
5. Prior to the silanization step, it is crucial to remove any traces of water from the coverslips and the polypropylene jar since water rapidly induces hydrolysis of (3-Aminopropyl)triethoxysilane and thus prevents silanization of the coverslip surface.
6. After each use, the original stock solution bottle should be filled with inert gas such as argon or nitrogen and stored at 4 °C (for (3-Aminopropyl)triethoxysilane) or −20 °C (for PEG powders) for long-term storage.
7. The amounts were calculated assuming that 20 coverslips (24 mm × 60 mm) are PEGylated. With this scheme, about 4% of the PEG (polyethylene glycol) is biotinylated. Depending on the specific experiment, you may want to consider using more or less biotin-PEG. Reported values typically range between 0.5 and 4% biotin-PEG.
8. After coverslip surface passivation, test it with a protein or DNA of interest, since the success of the surface passivation is sample-specific. The surface passivation method described here is a good starting point for many single-molecule applications. However, if nonspecific binding is problematic, other surface passivation methods have been reported. For example, in addition to mPEG and biotin-PEG, bis-NHS-PEG can be co-mixed, and 0.05 M sulfo-DST (Soltek Ventures CL107) is applied [11]. Dichlorodimethylsilane (DDS)-Tween-20-based surface passivation method is another option to consider [12].
9. When adding or pipetting lambda DNA, cut the end of the pipet tip to make the entrance wider. Gently pipet the sample up and down to mix and never vortex the sample.
10. Excess oligos do not seem to interfere with tethering the lambda DNA but they can be removed if desired (*see* Subheading 3.2.2, **steps 6** and 7).

11. After eluting from the gel, DNA might stick to the dialysis tubing surface. The application of a reverse electric field can detach DNA from the tubing.
12. This step can be done for 1 h instead of overnight.
13. We noticed that, depending on what column we use, the protein activity could be altered. For example, in our single-molecule bacterial SMC study [13], we used BioRad's Micro Bio-Spin P-30 gel columns, since it did not alter the SMC protein activity.
14. We strongly recommend practicing this step with non-functionalized coverslips to minimize the risk of breaking functionalized coverslips.
15. Be careful not to apply pressure where the coverslip extends over the quartz plate as the coverslip will readily snap.
16. It will take ~5 min for the epoxy to cure.
17. TIRF microscopes are commercially available as well, but much more expensive.
18. If you intend to do two-color imaging for Cy3 and Cy5 fluorescent dyes, we suggest using 641 nm (laser #1) and 532 nm (laser #2) lasers. If Cy3 and quantum dot 705 need to be imaged, using only 532 nm laser is sufficient since it can excite both.
19. A beam expander is simply a telescope. It consists of either two convex lenses (Keplerian beam expander) or one convex and one concave lenses (Galilean beam expander). We used ACN127-025-A and AC254-200-A (Thorlabs) for initial beam expansion. The second beam expander is comprised of AC254-040-A and AC508-150-A (Thorlabs).
20. We used T640lpxr dichroic mirrors, ET585/65 and ET700/75M emission filters for Cy3 and quantum dot 705 imaging. There are also commercial Dual-view imaging modules.
21. Lifting up the air spring to produce a drop of solution from the inlet tube minimizes the risk of air bubbles being trapped in the inlet of the flow cell. To further minimize air bubbles we suggest degassing all buffers under vacuum for 30 min prior to use. Degassing can be done by leaving tubes with caps loosened under a vacuum chamber.
22. For the first experiment, a suggested guideline is to dilute the DNA by 100-fold. After flowing in 150–200 μL at a rate of 0.03–0.05 mL/min, allow the DNA containing solution to incubate in the flowcell for an additional 3 min without flow. Then, flow in 10 nM SYTOX Orange at a rate of 0.05 mL/min. The SYTOX Orange dye will stain flow-stretched DNAs enabling their visualization. Depending on the DNA density,

adjust the DNA concentration or flow conditions to achieve an optimal number of tethered DNAs. As a control experiment, slightly lift the air-spring which will quickly reverse the flow direction. The tethered DNAs should be stretched out along the reversed flow direction, indicating that the surface-passivated PEGylated coverslip is well passivated and that there is not significant nonspecific sticking of DNA to the surface.

23. In order to capture fast diffusion events, start with an EMCCD camera exposure time on the order of 50–100 ms. For Cy3 imaging, typical through objective excitation powers are approximately 0.1 mW of 532 nm laser light.
24. The well-known oxygen scavenging systems are based on the use of (1) glucose oxidase and catalase [14]; (2) PCA (protocatechuic acid), PCD (protocatechuate-3,4-dioxygenase), and trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) [15]; (3) a reducing and oxidizing system (ROXS) [16]; (4) pyranose oxidase and catalase [17]. The glucose oxidase and catalase systems used to be commonly employed in single-molecule experiments, but it leads to a significant drop in pH over time [17]. This drawback was overcome by the pyranose oxidase and catalase-based system. We routinely use the PCA-PCD-trolox system (10 mM, 50 nM, and 0.2–1.0 mM, respectively). Note that the pH of trolox stock should be increased to around 7–8, and, as a caveat, the use of excess Trolox in your imaging buffer may alter protein activity.
25. Direct visualization of DNA facilitates subsequent data analysis. For example, identifying where a labeled protein locates is straightforward.
26. Note that MSD(N, n) is expected to increase linearly with $n\Delta t$ for Brownian diffusion. Therefore, it is strongly suggested that one should only consider trajectories that pass a linearity test (e.g. Pearson correlation coefficient >0.9 obtained from CORREL function in Microsoft Excel).
27. This approach enables one to correlate protein binding with changes in DNA length (i.e. the activity of chromatin proteins). Since the emission signals from the labeled protein and DNA are spectrally well separated, they can be imaged on separate halves of an EMCCD camera using a dual-view apparatus.
28. If labeling protein leads to the loss or decrease of protein activity, use unlabeled protein, instead. In that case, the dual-view part in the microscope is unnecessary. By tracking the quantum dot labeled at the end of a DNA, quantitative information on DNA remodeling, including the compaction rate and the degree (fraction) of compaction, can still be obtained.

29. This step requires empirical optimization (*see* **Note 22**) to achieve the following goals: (1) the tethered DNA density that allows for enough sampling while keeping nearby DNAs well-separated, and (2) to find the optimal quantum dot-to-DNA incubation ratio for high DNA end labeling efficiency. For the first optimization experiment, after flowing in 150–200 μL of the quantum-dot conjugated DNA at a rate of 0.03–0.05 mL/min, allow an additional 3 min of incubation without flow. Wash out unbound quantum dots, unbound DNA or unbound quantum dot-conjugated DNA by flowing 200–400 μL of EBB+BSA buffer at a rate of 0.05 mL/min. At that time, record a movie. Then, flow 10 nM SYTOX Orange at a rate of 0.05 mL/min while recording another movie. The SYTOX Orange will visualize flow-stretched DNAs and check the surface-tethered DNA density. Comparing the two movies will indicate the quantum-dot labeling efficiency to DNA. Adjust the incubation conditions accordingly for your next experiments.
30. The average quantum dot position in the absence of the flow indicates the DNA-tether point while the difference between the quantum dot positions in the presence and absence of the flow is the DNA length stretched under flow.
31. We typically use 100 ms exposure time and collect on the order of a few hundred to a few thousand frames. The number of frames will need to be adjusted based on how fast your protein of interest structures DNA.
32. In the case of Fiji, this can be done by drawing a rectangle around a DNA compaction event and pressing "ctrl + t."
33. Two-dimensional Gaussian fitting is an option as mentioned in Subheading 3.8.2. In our algorithm, two-dimensional intensity values are projected onto the DNA long axis, and then one-dimensional Gaussian fitting was performed. The custom Matlab codes are available in the "supplementary information" section of our bacterial SMC paper (https://www.nature.com/articles/ncomms10200#supplementary-information) [13].

Acknowledgement

This work was supported by a National Science Foundation CAREER Award [MCB-1148818 to J.J.L].

References

1. Song D, Loparo JJ (2015) Building bridges within the bacterial chromosome. Trends Genet 31(3):164–173. https://doi.org/10.1016/j.tig.2015.01.003
2. Dame RT (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Mol Microbiol 56(4):858–870. https://doi.org/10.1111/j.1365-2958.2005.04598x
3. Price AC, Pilkiewicz KR, Graham TG et al (2015) DNA motion capture reveals the mechanical properties of DNA at the mesoscale. Biophys J 108(10):2532–2540. https://doi.org/10.1016/j.bpj.2015.04.022
4. Graham TG, Wang X, Song D et al (2014) ParB spreading requires DNA bridging. Genes Dev 28(11):1228–1238. https://doi.org/10.1101/gad.242206.114
5. Chen I, Dorr BM, Liu DR (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. Proc Natl Acad Sci U S A 108(28):11399–11404. https://doi.org/10.1073/pnas.1101046108
6. Los GV, Encell LP, McDougall MG et al (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 3(6):373–382. https://doi.org/10.1021/cb800025k
7. Shi X, Jung Y, Lin LJ et al (2012) Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. Nat Methods 9(5):499–503. https://doi.org/10.1038/nmeth.1954
8. Schindelin J, Arganda-Carreras I, Frise E et al (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9(7):676–682. https://doi.org/10.1038/nmeth.2019
9. Tafvizi A, Huang F, Leith JS et al (2008) Tumor suppressor p53 slides on DNA with low friction and high stability. Biophys J 95(1):1. https://doi.org/10.1529/biophysj.108.134122
10. Jaqaman K, Loerke D, Mettlen M et al (2008) Robust single-particle tracking in live-cell time-lapse sequences. Nat Methods 5(8):695–702. https://doi.org/10.1038/nmeth.1237
11. Blainey PC (2007) Single-molecule studies of protein-DNA interaction: diffusive search and sequence-dependent motors. Dissertation, Harvard University
12. Hua B, Han KY, Zhou R et al (2014) An improved surface passivation method for single-molecule studies. Nat Methods 11(12):1233–1236. https://doi.org/10.1038/nmeth.3143
13. Kim H, Loparo JJ (2016) Multistep assembly of DNA condensation clusters by SMC. Nat Commun 7:10200. https://doi.org/10.1038/ncomms10200
14. Yildiz A, Forkey JN, McKinney SA et al (2003) Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. Science 300(5628):2061–2065. https://doi.org/10.1126/science.1084398
15. Aitken CE, Marshall RA, Puglisi JD (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. Biophys J 94(5):1826–1835. https://doi.org/10.1529/biophysj.107.117689
16. Vogelsang J, Kasper R, Steinhauer C et al (2008) A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. Angew Chem Int Ed Engl 47(29):5465–5469. https://doi.org/10.1002/anie.200801518
17. Swoboda M, Henig J, Cheng HM et al (2012) Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. ACS Nano 6(7):6364–6369. https://doi.org/10.1021/nn301895c

Chapter 16

Unraveling the Biophysical Properties of Chromatin Proteins and DNA Using Acoustic Force Spectroscopy

Szu-Ning Lin, Liang Qin, Gijs J. L. Wuite, and Remus T. Dame

Abstract

Acoustic Force Spectroscopy (AFS) is a single-molecule micromanipulation technique that uses sound waves to exert force on surface-tethered DNA molecules in a microfluidic chamber. As large numbers of individual protein-DNA complexes are tracked in parallel, AFS provides insight into the individual properties of such complexes as well as their population averages. In this chapter, we describe in detail how to perform AFS experiments specifically on bare DNA, protein-DNA complexes, and how to extract their (effective) persistence length and contour length from force-extension relations.

Key words Acoustic force spectroscopy, Single-molecule manipulation, Protein-DNA interaction, DNA-binding protein, Bacterial chromatin protein

1 Introduction

Sound waves can be used to exert forces on objects; this concept is key to the application of controlled forces on surface-tethered microparticles using a method called acoustic force spectroscopy (AFS) [1, 14]. By applying force on the microparticle, force is exerted on the tether, DNA or a protein-DNA complex [1, 14]. The experimental layout for studying protein-DNA complexes is similar to that used for tethered particle motion (TPM) [2] and magnetic tweezers (MT) (*see* Chap. 17): one end of a DNA substrate is labeled with DIG to bind the anti-DIG on the sample carrier or flow chamber surface, while the other end of the DNA is labeled with biotin to bind on the silica or magnetic beads. In AFS, force is applied vertically to microspheres attached to surface-tethered DNA molecules as is the case for MT. The *x*-*y* motion of the bead is monitored and its diffraction pattern is used to determine its *z*-position from a look-up table (LUT) of radial profiles [1]. The

Szu-Ning Lin and Liang Qin are co-first authors.

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_16,

z-position of the bead (minus the bead radius) corresponds to the end-to-end distance of the DNA tether.

Acoustic pressure is generated by a vibrating piezo element attached to the bottom or top of the flow cell. The acoustic pressure transfers potential energy into the medium in the flow cell chamber and forms a standing wave. Particles, in this case, polystyrene or silica microspheres 1–5 μm in diameter (with a volume V), are forced to align at the nodes of the standing wave. By increasing the voltage (changing the amplitude of the wave), beads will experience a larger force (F) toward the wave node [3, 4]. The effective force applied on each bead is described by Eq. 1).

$$F = -V\nabla\left[\frac{1-\kappa^*}{4}\kappa_m P^2 - \frac{(\rho^*-1)}{1+2\rho^*}\rho_m v^2\right] \quad (1)$$

in which P is the acoustic pressure (energy gradient), v is the velocity of particles, and ρ^* ($=\rho_p/\rho_m$) and κ^* ($=\kappa_p/\kappa_m$) are the density ratio and compressibility ratio between the particle and the fluid, respectively [3]. The magnitude of the force applied to a bead is determined by the material and size of the bead, the medium inside the flow cell, and the vibration of the piezo. We routinely apply forces up to ~70 pN to achieve DNA overstretching and protein unfolding, with polystyrene microspheres, 4.5 μm in diameter.

Here, we describe the assembly of an acoustic force spectroscopy instrument around a commercial inverted microscope and explain how to perform experiments on DNA molecules and protein-DNA complexes. In addition, we demonstrate how structural and mechanochemical properties of protein-DNA complexes can be extracted from AFS data.

2 Materials

Prepare all solutions by using ultrapure water (prepared by purifying deionized water, to attain a sensitivity of 18 MΩ cm at 25 °C; MilliQ). Prepare solutions at room temperature (RT) and store at 4 °C.

2.1 Stock Solutions and Beads

2.1.1 Buffer for Analysis by Agarose Gel Electrophoresis (See Subheading 3.1)

TAE: 40 mM Tris–HCl (pH 7.6), 20 mM acetic acid and 1 mM EDTA.

2.1.2 Solutions for Reference Bead Preparation (See Subheading 3.2)

1. Buffer A: 100 mM $Na_2B_4O_7$, 150 mM NaCl, 0.05% w/v Pluronics (pH 8.3) (BASF), 30 μM Digoxigenin-NHS ester (Sigma-Aldrich).
2. Buffer B: 10 mM HEPES (pH 7.5).
3. Buffer C: 10 mM HEPES (pH 7.5) with 0.1% Pluronics.

2.1.3 Solutions for Cleaning Protocol (See Subheadings 3.4 and 3.5)

1. Bleach solution: 0.7 M NaClO.
2. Sodium thiosulfate solution: 10 mM $Na_2S_2O_3$.

2.1.4 Solutions for Passivation of the Flow Cell (See Subheading 3.6)

1. Phosphate-Buffered Saline (PBS, pH 7.4): 150 mM NaCl and 10 mM phosphate. 1 mM EDTA and 10 mM NaN_3 are added to prevent bacterial growth in the buffer.
2. Anti-digoxigenin solution: 200 μg/mL anti-DIG (Roche) in PBS.
3. Buffer D: 0.2% (w/v) BSA (Sigma-Aldrich) in PBS.
4. Buffer E: 0.5% (w/v) Pluronics (Sigma-Aldrich) in PBS.
5. Buffer F: 0.02% (w/v) Casein (Roche) and 0.02% (w/v) Pluronics in PBS.

2.1.5 Solutions for Passivation of the Flow Cell (See Subheadings 3.2 and 3.6)

1. 1.9 μm Streptavidin-coated bead (Kisker Biotech) in PBS.
2. 4.5 μm Streptavidin-coated bead (Kisker Biotech) in PBS.

2.2 DNA Substrates

DNA substrates for AFS experiments are generated via Polymerase Chain Reaction (PCR) using 5′ biotinylated and 5′ Digoxygenin-labeled primers (*see* **Note 1**). The length of DNA is designed to be in the range significantly shorter than the distance from the surface to the wave node. AFS is capable of measuring DNA substrates as short as 1 kbp and as long as 45.5 kbps [8]. Table 1 and Subheading 3.1, **step 2** summarize oligonucleotides used to generate our toolbox of DNA of different lengths (2000–8000 bp) and sequence content (32% and 50%) by PCR. All DNA substrates are stored at −20 °C after purification and concentration determination.

1. A DNA template contains the sequence of interest; plasmid pKYBI (8393 bp, New England Biolabs) and plasmid pRD227 are used as the templates for 8000 bp DNA substrate with 50% GC content and 2000 bp DNA substrate with 32% GC content, respectively.
2. 5′ biotin-labeled reverse primer (*see* Table 1).
3. 5′ Digoxygenin-labeled forward primer (*see* Table 1).
4. 100% DMSO (New England Biolabs).
5. Recombinant *Taq* DNA polymerase (5 U/μL) (Thermo Scientific).

Table 1
Primer sequences

Primer name	Sequence (5′–3′)	Modification
32% GC AFS General forward primer	GTGTGTGTGTGTGGTGTGTGGTGG ATACATATGCAACTTGAACGGCG TAAAAGAGG	5′ Digoxygenin
2000 bp 32% GC AFS reverse primer	GTGTGTGTGTGTGGTGTGTGGTGG TCCCTCACTAGTTTAGTACATGAACTG	5′ Biotin
50% GC AFS general forward primer	CTCTCTCTCTCTTCTCTCTTCTCT GAATTCGCGGCCGCCGTC	5′ Digoxygenin
2000 bp 50% GC AFS reverse primer	CTCTCTCTCTCTTCTCTCTTCTCT CAGTGGGAACGATGCCCTC	5′ Biotin
4000 bp 50% GC AFS reverse primer	CTCTCTCTCTCTTCTCTCTTCTCT CAGCGGTGGTTTGTTTGCCG	5′ Biotin
6000 bp 50% GC AFS reverse primer	CTCTCTCTCTCTTCTCTCTTCTCT CGATCCCCGGCAAAACAGC	5′ Biotin
8000 bp 50% GC AFS reverse primer	CTCTCTCTCTCTTCTCTCTTCTCT GGTACCAATGTTTTAATGGCGGATG	5′ Biotin

T = Modified T

6. Deoxyribose Nucleotide Triphosphate (dNTP mix) (Thermo Scientific, 2 mM).
7. *Taq* DNA polymerase reaction buffer (Thermo Scientific, 10×).
8. GenElute ™ PCR cleanup kit (Sigma-Aldrich).
9. Biorad T100 Thermocycler PCR or any other available PCR machine.
10. 1% agarose gel in 1× TAE.
11. Nanodrop® (Thermo Scientific).
12. GeneRuler DNA ladder (ThermoFisher Scientific).

2.3 AFS Instrument

The AFS system is built around a commercially available inverted microscope combined with commercially available electronics and a commercially available AFS chip (*see* Fig. 1).

1. Microscope: Inverted microscope (Nikon, TE200) with condenser (Nikon, LWD lens), CFI Achromat 40× air objective (Nikon, NA = 0.65).
2. Illumination: Collimated LED (ThorLabs, 660 nm, 1200 mA) (*see* **Note 2**).
3. CMOS camera (Thorlabs, monochrome, pixel size 5.3 μm, 60 fps) connects to the computer.

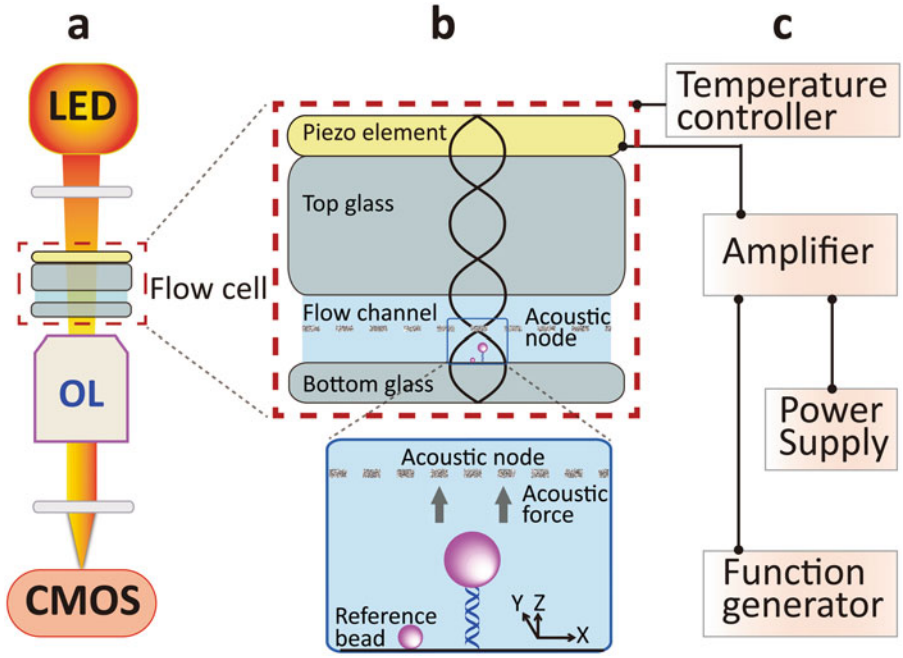


Fig. 1 Illustration of the principle of Acoustic Force Spectroscopy. (**a**) The acoustic force spectroscopy instrument consists of a flow cell, an inverted microscope visualizing targets with objective lens (OL), temperature controller connected to the AFS-Chip holder, a CMOS digital camera and 660 nm LED light source. (**b**) AFS flow cell consists of the piezo element and two glass slides with a fluidic channel in between. The acoustic wave generated by the piezo travels through the top glass to the bottom glass and the bottom glass as a reflector reflects the acoustic wave, producing the standing wave over the flow cell. The acoustic standing waves carry pressure profile, which generate acoustic forces. The tethered particles in the flow channel exposed to the acoustic force are driven in the direction of the acoustic pressure node. (**c**) The temperature controller independently connects to the AFS-Chip metal holder. Both power supply and function generator connect to the amplifier, which controls the piezo element

4. Stage: *Z*-axis piezo translation stage (PI, MCLS03200), driven by Nano-Drive controller system (MCL, Nano-Drive, MCLC03200) which connects to the computer.
5. Function generator (Keysight, 33220A).
6. Power supply (Voltcraft, VSP 1410).
7. Amplifier (Mini-Circuits, LZY-22+).
8. TMC Vibracontrol clean top isolation table.
9. Computer: "Advanced AFS workstation" (LUMICKS B.V., AFS-CPU).
10. AFS-chip (LUMICKS B.V., AFS-CH2) (*see* Fig. 2).
11. AFS-chip temperature control holder (*see* Fig. 3).
12. Custom pressure system. Syringes contain samples and buffers are connected to a gas pressure container (*see* Fig. 3).

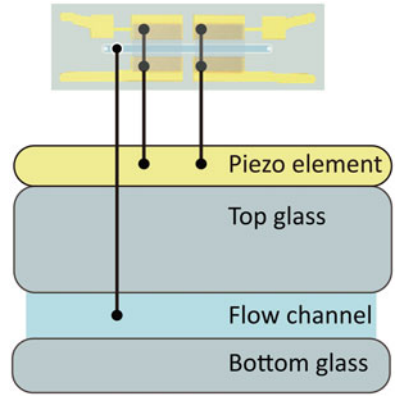


Fig. 2 Schematic representation of an AFS-chip. The AFS flow cell consists of the piezo element, top glass slide and bottom glass slide with a flow channel in between

2.4 Particle Tracking, Control, and Analysis

Three programs, a bead tracking program (*see* Subheading 3.6), a data analysis program (*see* Subheading 3.7.1), and an extensible worm-like-chain (eWLC) fitting program (*see* Subheading 3.7.2), are used during the measurement and data analysis process.

1. The tracking program is written in LabVIEW [1] and available online http://figshare.com/articles/AFS_software/1195874. A detailed manual is also provided with the software. It is used to:
   (a) Control output frequency and output power of vibration. Apply a linear force ramp mode, the voltage is ramped with a square-root function.
   (b) Create a template image of a bead (imaged via a LED with a camera) to track bead position using a template-matching algorithm.
   (c) Track the *x*, *y* movements of tethered beads and record a look-up-table (LUT) in *z* direction for each bead. Routinely, the *z*-stage is moved in 80 nm steps through the LUT range of 0–8000 nm. The *z*-distance range of the LUT has to be larger than the maximal extension of the DNA molecule.
   (d) Calibrate force-voltage relationship.
2. The data analysis program is also written in LabVIEW [1] and available online via the same link as indicated under Subheading 2.4, **item 1**). It is used to:
   (a) Perform real-time acquisition of three-dimensional bead position in the flow cell.
   (b) Determine anchor points of all the tracked beads.

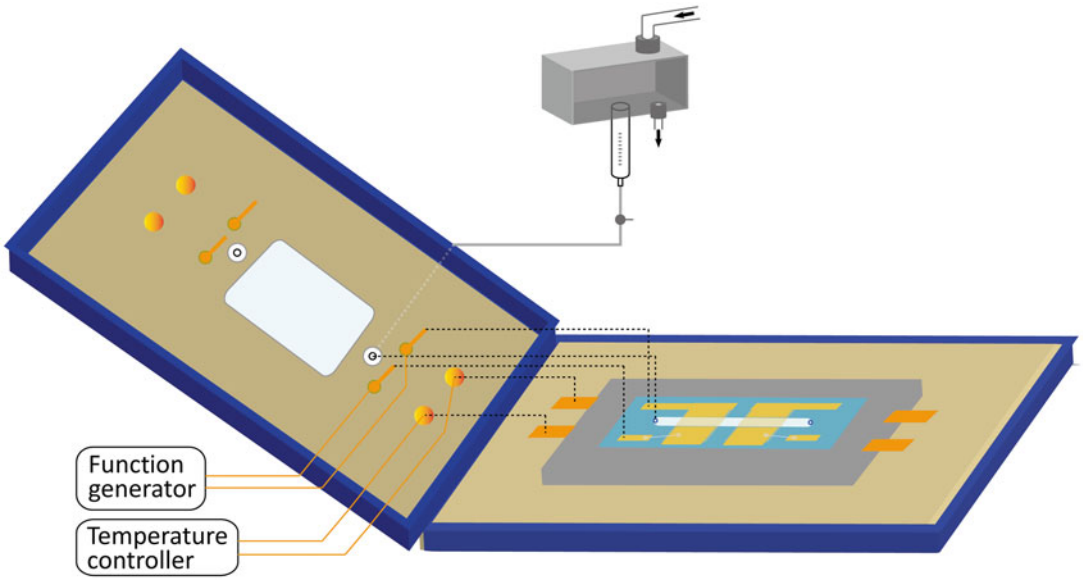


Fig. 3 Schematic representation of the temperature control holder. The AFS-chip fits into the temperature control metal holder (grey) and this metal holder fits into the flow cell holder (light brown and navy blue) that consists of lid and base. When the lid is closed, the metal points on the lid, connected to the function generator and temperature controller, will connect to their counterparts (metal regions) on the piezo and the metal holder (grey), and the holes on the lid, connected to the pressure system by a thin tube, will connect to their counterparts (holes) on the AFS-chip

(c) Correct for drift (*see* **Note 3**) based on positions of the surface attached reference beads.

(d) Generate force-extension (force-distance, FD) curves.

3. The WLC model fitting program is written in MatLab [5] and available online https://github.com/onnodb/FDFIT/tree/AFSFitting. It is used to:

(a) Fit FD curves exported from data analysis program to the eWLC model.

(b) Extract values of parameters reflecting the physical characteristics of bare DNA and protein-DNA complexes.

3 Methods

3.1 Generation of DNA Substrate Using PCR

1. DNA substrates for AFS experiments are generated by PCR. Carefully mix the reagents below in a PCR tube. Keep the enzymes in a −20 °C cold block and dNTP stocks on ice when taken outside the freezer.

Reagent	Quantity
dNTP mix (2 mM)	5 μL
Forward primer (10 pmol)	1 μL
Reverse primer (10 pmol)	1 μL
Taq Polymerase buffer (10×)	5 μL
DNA template (10 ng)	1 μL
Taq DNA Polymerase (5 U/μL)	0.2 μL
DMSO (100%)	1.25 μL
MilliQ	Add to 50 μL total volume

2. Use the program below to perform PCR (optimized for use in a Biorad T100 Thermocycler).

	Temperature	Duration	Cycle
Initialization	95	5 min	1
Denaturation Annealing Extension/elongation	95 65 72	30 s 30 s 4 min	35
Final elongation	75	10 min	1
Final hold	15	∞	1

3. Load 5 μL of each PCR product on a 1% agarose gel in TAE buffer, alongside a DNA marker for size estimation of the PCR

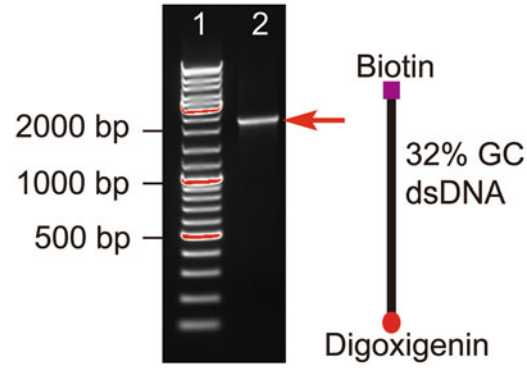


Fig. 4 Visualization of PCR product size by agarose gel electrophoresis. (1) 2 μL of the GeneRuler DNA marker. (2) 2 μL of the purified PCR product, and it is ready for use in Acoustic Force Spectroscopy experiments. The schematic representation of PCR-generated 5′ digoxigenin and 3′ biotin modified-DNA

product. Purify the PCR products with GeneElute PCR cleanup kit. *See* Fig. 4 for an example of the purified PCR products.

4. Use Nanodrop® to measure the concentration of PCR products. Store the purified DNA products in MilliQ at −20 °C.

3.2 Preparation of Reference Beads

1. Mix 8 μL of 0.5% (w/v) 1.9 μm polystyrene beads into 1 mL of buffer A.
2. To coat polystyrene beads with DIGs, incubate the bead solution with DIG-NHS at RT for 3 h with tumbling.
3. Centrifuge the bead solution at 2000 × *g* for 1 min, and discard the supernatant.
4. Remove the free DIG-NHS by washing the bead solution with 1 mL buffer C. Centrifuge at 2000 × *g* for 1 min and discard the supernatant. Repeat the wash twice.
5. Resuspend the beads in 1 mL buffer B. The beads can be used immediately or stored at 4 °C.

3.3 Preparation of Tether Beads

1. To exchange the storage buffer of the commercial beads with PBS, dilute 20 μL of the bead solution in 500 μL PBS.
2. Vortex the sample thoroughly.
3. Centrifuge at 2000 × *g* for 2 min and discard the supernatant, but avoid discarding beads. Resuspend the beads in 500 μL PBS.
4. Repeat **steps 2** and **3**, carefully discard ~450 μL of supernatant, leaving ~50 μL in tube.
5. Add 4 μL of the reference beads to the 50 μL solution from **step 4** and resuspend by vortexing.

3.4 Flow Cell and Tubing Cleaning

The tubing and the flow cell are used repeatedly in our system; we replace the syringe before each experiment.

1. Install the flow cell chip into its holder. The holder is connected to tubing (*see* Fig. 3).
2. Introduce 0.2 mL of bleach into the flow cell through the tubing and incubate for 10 min.
3. Rinse the tubing and the flow cell with MilliQ.
4. Introduce 0.2 mL $Na_2O_2N_3$ into the flow cell via the tubing and incubate for 10 min.
5. Flush 0.5 mL MilliQ into flow cell through the tubing.

3.5 Preparation of Flow Cell and Bead Tethers

To minimize waste of materials, in **steps 1**, **5**, and 7, the sample is manually pipetted into the flow cell without using the syringe. However, all the buffers for passivation are introduced through the syringe so the syringe and tubing are also passivized. All preparations are at RT.

1. Inject 20 μL of anti-digoxigenin solution into the flow chamber and incubate for 20 min.
2. To prevent air bubbles flowing into the flow chamber, introduce 0.5 mL of buffer D into the syringe and flush out the air present in the tubing before connecting the tubing to the holder. Place the chip in holder, leave drops at the two holes in the flow cell, and clamp the holder gently [6].
3. Inject 0.1 mL of buffer D and incubate for 30 min for the first time passivation. Discard the residual of buffer D in syringe.
4. Add 0.5 mL buffer E into the syringe, flush in 0.1 mL of buffer E, and incubate for 30 min for the second time passivation. Discard the residual of buffer E in the syringe.
5. Take out the chip to introduce 30 μL of DNA solution in the flow chamber and incubate for 20 min. Place the chip back into the holder and clamp the holder gently.
6. Add 0.5 mL buffer F into the syringe. Flush in 0.1 mL buffer F to remove free DNA.
7. Take out the chip to introduce 20 μL bead solution in the flow chamber. Place the chip back into the holder, clamp the holder gently, and incubate for 30 min.
8. Add 0.5 mL measuring buffer in the syringe.
9. The flow cell is ready for measurement.

3.6 Measurements

1. Switch on the illumination and bring the bottom-tethered beads into focus (*see* Fig. 2).
2. Start the bead tracking program in LabView.

3. Select the frequency for the piezo. The piezo frequency is given by the supplier (LUMICKS). Each piezo has a specific impedance, in other words, there is deviation of vibration frequency from chip to chip. More details have been described previously [9].
4. Remove untethered beads by flushing the flow cell with the measurement buffer at a flow rate of 0.2 μL/min until such beads are no longer observed in the region of interest (ROI).
5. Select the tethered beads in the ROI.

   Generate a LUT for each ROI, applying a constant force (≈10 pN) to the tethers to minimize bead motion. The "Create LUT" bottom in the program starts to move the sample stage (or objective) and record the ring patterns at different *z*-positions. We collect the LUT at 60 Hz with a camera exposure time in 16.6 ms, each stage moving step is 80 nm throughout a range of 0–8000 nm.
6. Start to record the tracking of the selected tethered beads.
7. Record the *x*-*y* motions of the tethered beads in the absence of force, the *x*-*y* motions are used for the determination of the single tethered beads in data analysis. The time required to sample all conformations depends on bead size; for beads with a diameter of 4.5 μm, a 10 min recording is sufficient.
8. To calibrate the force-voltage relationship of each tethered bead, apply a series of different forces on the tethered particles. To collect sufficient data for power spectra fitting (*see* Subheading 3.7.1), perform 2 min of *z*-position recording for each force.
9. Apply low force (~10 pN) and slow flowing rate (0.2 μL/min) while introducing protein solution. The flow in the flow chamber will result in a drag force on the tethered beads.
10. Apply a constant rate force ramp (120 ms between each force step) to generate FD curves of bare DNA and protein-DNA complexes.

3.7 Data Analysis

1. Load the data into the analysis program in LabView.

3.7.1 Generate FD Curves

2. Determine the single tethered beads and reference beads data by the root mean squared (RMS) displacement values and anisotropic ratio (*s*) from the *x*-*y* motions of each tethers.
   In Subheading 3.6, **step 7**, tethered beads randomly move around their anchor points without force applied. In AFS, RMS is used to quantify the degree of the tether motions in two dimensions over a period of time, *t*:

$$\mathrm{RMS} = \sqrt{\left\langle (x-\bar{x})^2 + (y-\bar{y})^2 \right\rangle_t} \tag{2}$$

in which $\bar{x}$ and $\bar{y}$ are the average positions of the tethered bead over time t.

Usually, not all tethered microspheres are attached to the surface via a single tether. Calculation of the anisotropic ratio (s) allows separation of single-tethered microspheres from stuck and multiple-tethered microspheres.

$$s = \frac{\lambda_{\text{major}}}{\lambda_{\text{minor}}} \tag{3}$$

in which λ_{major} and λ_{minor} represent the maximum and minimum values along the axes of the x-y-scatter plot respectively.

Single-tethered microspheres are expected to exhibit a perfectly symmetrical motion and to have an anisotropic ratio of 1. In our studies we use $s < 1.3$ as a threshold to discard multiple-tethered or poorly tracked particles. The particles that match the selection criterium are used for further analysis.

3. Remove the drift from the measured data by calculating the average drift of the stuck beads. Stuck beads are selected based on their x and y motions. In our studies we use RMS < 200/s ≈ 1 (*see* **Note 4**).
4. Determine the anchor points of the tethers by selecting the x-y motion trace without force applied during the measurement. Use "Anchor point" function in the program to determine the end-to-end length of the DNA molecule by Pythagoras calculation and the anchor point in data analysis program.
5. Calibrate the force-voltage relationship by selecting the voltage-time plot in the program where constant voltages were applied. Generate and fit power spectrum (*see* **Note 5**).
6. Generate FD curves by selecting the time period which force ramp was applied. Export FD curves in text file with "Export function" in the program for eWLC fitting (*see* Subheading 3.7.2).

3.7.2 Extensible Worm-Like Chain Model (eWLC) Model Fitting

The extensible worm-like-chain model in formula (4) describes the behavior of elastic polymers such as DNA and protein-DNA complexes [6].

$$\frac{z}{L_c} = 1 - \frac{1}{2}\left(\frac{k_B T}{F L_p}\right)^{\frac{1}{2}} + \frac{F}{K_0} \tag{4}$$

in which z is the extension and F is the external force, k_B is the Boltzmann constant, T is the absolute temperature, K_0 is the stretch modulus, L_p is the persistence length, and L_c is contour length. A typical value for the stretch modulus of double-stranded DNA is about 1000 pN [7].

1. Run the fitting program in MatLab software.
2. Import FD curves in MatLab program.
3. Select the data point of the FD curves which are taken below 30 pN.
4. Determine persistence length (L_p), contour length (L_c), and stretch modulus (K_0).

4 Analysis of Protein-DNA Complexes Using AFS

Architectural proteins bind to DNA via minor or major groove interactions and result in wrapping, bending, or bridging of the DNA. By applying force to protein-DNA complexes, the effect of proteins on DNA conformation and the binding behavior of these proteins can be investigated. Here, we discuss the effects of two types of DNA-binding proteins, HU and H-NS.

4.1 Force-Extension Curves of HU-DNA Complex

The HU protein compacts and stiffens DNA in a protein concentration-dependent manner. In our experimental system, DNA is compacted at concentrations below 400 nM. Above 400 nM, DNA is extended by the formation of a filament of HU proteins along DNA (*see* Fig. 5).

4.2 Force-Extension Curved of DNA and H-NS-DNA Complex

Histone-like nucleoid-structuring (H-NS) protein is a bacterial protein that plays a key role in chromosome organization and regulation. DNA and H-NS-DNA complex was studied by using AFS (*see* Fig. 6). The data show that DNA is stiffened by H-NS at 2000 nM, which is in agreement with previous studies [8, 9].

5 Notes

1. DNA substrates can also be prepared by other approaches, e.g., by filling in Digoxigenin/Biotin at two ends of cut plasmid or by ligating modified oligos/dsDNAs.
2. Diffraction ring patterns are required for accurate bead tracking. The light source needs to be monochromatic and aligned in parallel. To obtain collimated light, either a point source or an iris conjugated to your condenser is required. For monochromatic illumination a LED of defined wavelength or band-pass filter in the illumination path is suggested.
3. Movement of the machine or heat created by piezo vibration causes drift in the flow cell. A highly efficient piezo results in minimal heating of the system.
4. To correct drift signal, the program subtracts the displacements from x, y, z on the average traces of the selected stuck beads. The average displacement of the stuck beads is as a starting point, 0.

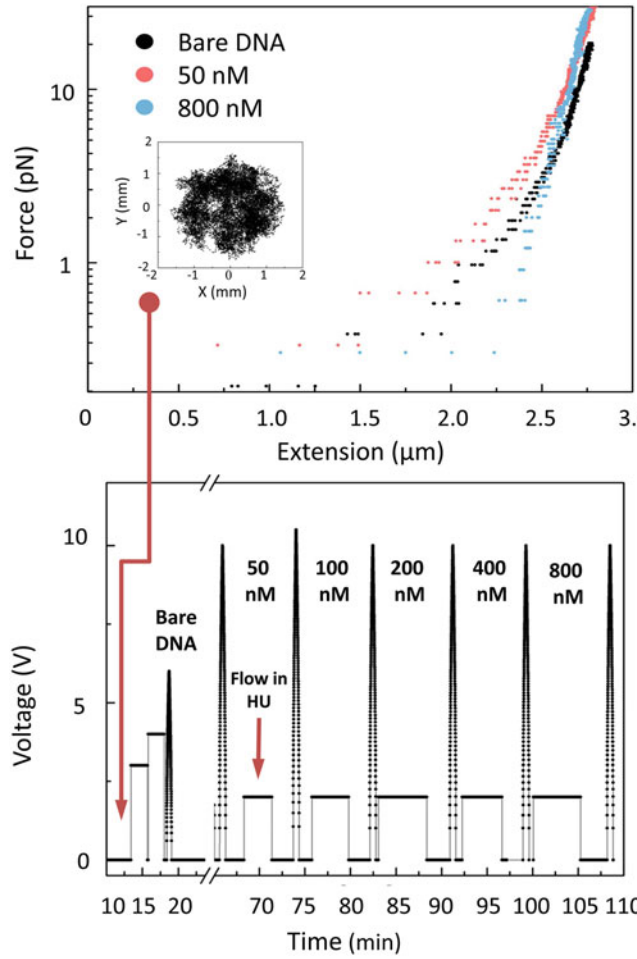


Fig. 5 Stretching curve of an HU-DNA complex and the flow chart of force application. The top figure shows force-extension curves of 8000 bp bare DNA (black), HU-DNA complex at low HU concentration (50 nM; red), and at high HU concentration (800 nM; blue). As an inset is shown the position over time of a DNA-tethered bead (anisotropic ratio, s, is 1.2). Bottom plot shows the force vs. time. A force of ~2 pN was applied while introducing proteins

5. The force power spectrum fitting is described in Norrelykke and Flyvbjerg [10]. Two important parameters that are obtained from the fitting are the frequency at the corner of the spectrum and the diffusion coefficient. The diffusion coefficient dependence can also be predicted from the bead size and the distance from the surface [11]. Nevertheless, the microsphere size varies in the same batch. Checking whether the theoretical value overlaps with the fitted value gives a reference of the error. This function is already included in the AFS data analysis program. Lorentzian formula and importance-weighted least squares fitting generate the force power spectra fitting [12, 13]:

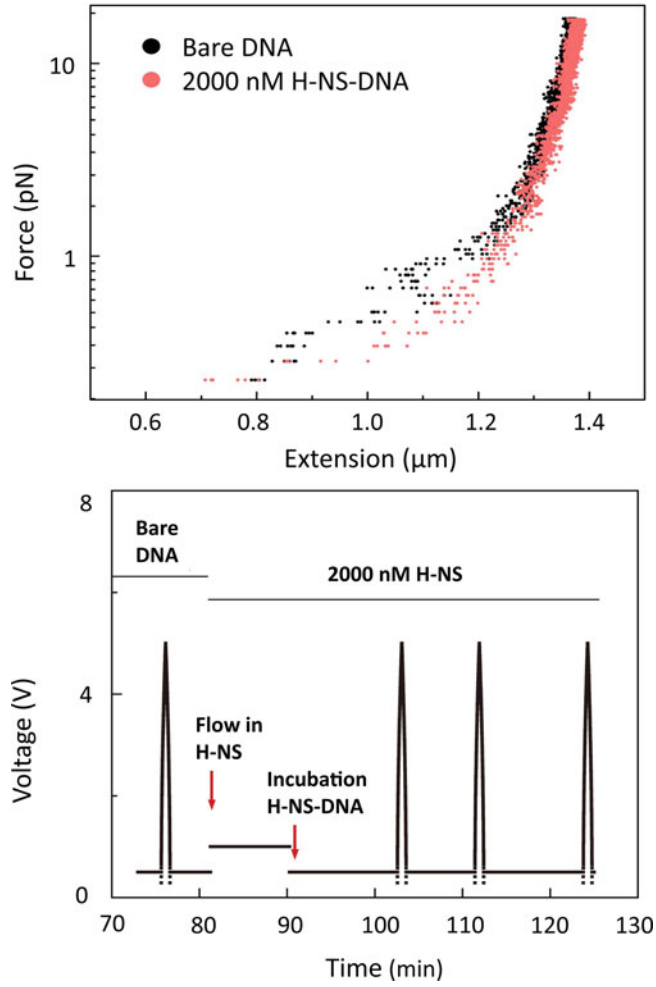


Fig. 6 Stretching curve of an H-NS-DNA complex and the flow chart of force application. Top plot shows force-extension curves of 4000 bp bare DNA (red) and H-NS-DNA complex (blue). Bottom plot shows the applied voltage vs. time. A low force of 0.8 pN was applied while introducing proteins and 0.2 pN was applied during incubation

$$\mathrm{P}(f) = \frac{D/(2\pi^2)}{f^2 + (k/(2\pi \times \gamma_{\mathrm{fax}}))^2}$$

where $D = k_{\mathrm{B}}T/\gamma_{\mathrm{fax}}$ is the microsphere diffusion constant, k_{B} is the Boltzmann constant, T is the temperature, γ_{fax} is the effective drag coefficient, f is the frequency, and k is the stiffness of a Hooke spring acting on the microsphere. $k = F/(L_{\mathrm{ext}} + R)$; in the expression of k, F is the force, L_{ext} is the measured extension of the DNA, and R is the radius of the microsphere.

Acknowledgments

The authors thank Andreas Biebricher for the development of surface passivation methods, Sandrine D'Haene, Seyda Acar, and Ramon van der Valk for DNA substrate design and technical assistance, LUMICKS B.V. for assistance in developing the AFS system. This work was supported by grants from FOM (R.T.D. and G.J.L.W.), NWO-VICI (R.T.D. and G.J.L.W.), HFSP (R.T.D.) and China Scholarship Council No. 201506880001 (L.Q.).

References

1. Sitters G, Kamsma D, Thalhammer G, Ritsch-Marte M, Peterman EJ, Wuite GJ (2015) Acoustic force spectroscopy. Nat Methods 12 (1):47–50. https://doi.org/10.1038/nmeth.3183
2. van der Valk RA, Laurens N, Dame RT (2017) Tethered particle motion analysis of the DNA binding properties of architectural proteins. Methods Mol Biol 1624:127–143
3. Settnes M, Bruus H (2012) Forces acting on a small particle in an acoustical field in a viscous fluid. Phys Rev E Stat Nonlinear Soft Matter Phys 85(1 Pt 2):016327. https://doi.org/10.1103/PhysRevE.85.016327
4. Gor'Kov L (1962) On the forces acting on a small particle in an acoustical field in an ideal fluid. Soviet Physics Doklady 6:773
5. Broekmans OD, King GA, Stephens GJ, Wuite GJ (2016) DNA twist stability changes with magnesium(2+) concentration. Phys Rev Lett 116(25):258102. https://doi.org/10.1103/PhysRevLett.116.258102
6. Odijk T (1995) Stiff chains and filaments under tension. Macromolecules 28(20):7016–7018. https://doi.org/10.1021/ma00124a044
7. Wang MD, Yin H, Landick R, Gelles J, Block SM (1997) Stretching DNA with optical tweezers. Biophys J 72(3):1335–1346. https://doi.org/10.1016/S0006-3495(97)78780-0
8. Lim CJ, Lee SY, Kenney LJ, Yan J (2012) Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. Sci Rep 2:509. https://doi.org/10.1038/srep00509
9. van der Valk RA, Vreede J, Qin L, Moolenaar GF, Hofmann A, Goosen N, Dame RT (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-bridging activity. elife 6. https://doi.org/10.7554/eLife.27369
10. Nørrelykke SF, Flyvbjerg H (2010) Power spectrum analysis with least-squares fitting: amplitude bias and its elimination, with application to optical tweezers and atomic force microscope cantilevers. Rev Sci Instrum 81 (7):075103
11. Schäffer E, Nørrelykke SF, Howard J (2007) Surface forces and drag coefficients of microspheres near a plane surface measured with optical tweezers. Langmuir 23(7):3654–3665
12. te Velthuis AJ, Kerssemakers JW, Lipfert J, Dekker NH (2010) Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. Biophys J 99 (4):1292–1302. https://doi.org/10.1016/j.bpj.2010.06.008
13. Berg-Sørensen K, Flyvbjerg H (2004) Power spectrum analysis for optical tweezers. Rev Sci Instrum 75(3):594–612. https://doi.org/10.1063/1.1645654
14. Kamsma D, Creyghton R, Sitters G, Wuite GJL, Peterman EJG (2016) Tuning the music: acoustic force spectroscopy (AFS) 2.0. Methods 105:26–33

Chapter 17

Unraveling DNA Organization with Single-Molecule Force Spectroscopy Using Magnetic Tweezers

Thomas B. Brouwer, Artur Kaczmarczyk, Chi Pham, and John van Noort

Abstract

Genomes carry the genetic blueprint of all living organisms. Their organization requires strong condensation as well as carefully regulated accessibility to specific genes for proper functioning of their hosts. The study of the structure and dynamics of the proteins that organize the genome has benefited tremendously from the development of single-molecule force spectroscopy techniques that allow for real-time, nanometer accuracy measurements of the compaction of DNA and manipulation with pico-Newton scale forces. Magnetic tweezers in particular have the unique ability to complement such force spectroscopy with the control over the linking number of the DNA molecule, which plays an important role when DNA organizing proteins form or release wraps, loops, and bends in DNA. Here, we describe all the necessary steps to prepare DNA substrates for magnetic tweezers experiments, assemble flow cells, tether DNA to magnetics bead inside flow cell, and manipulate and record the extension of such DNA tethers. Furthermore, we explain how mechanical parameters of nucleo-protein filaments can be extracted from the data.

Key words Magnetic tweezers, Single-molecule, Force spectroscopy, Rotational spectroscopy, DNA mechanics, DNA compaction, Bacterial chromatin, Eukaryotic chromatin

1 Introduction

1.1 Chromatin Organization by DNA-Binding Proteins

Genomes carry the genetic blueprint of all living organisms. Their organization requires strong condensation in a structure called chromatin, as well as carefully regulated accessibility to genes in chromatin for proper functioning of their hosts. The organization of chromatin is carried out by a large group of proteins that vary in abundance, structure, and function. Different kingdoms of life have evolved to implement different solutions for a compact, but dynamic organization of chromatin. Nonetheless, many of the involved proteins share common mechanisms to deform the DNA by inducing kinks, bends, wraps, and loops in the DNA trajectory. Examples of common DNA organizing proteins are histone proteins found in eukaryotes and in Archaea and smaller and more

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_17,

diverse proteins like HU, integration host factor IHF, and histone-like nucleoid-structuring protein H-NS in prokaryotes [1].

For a mechanistic understanding of the function of these proteins, required for a robust comprehension of chromatin-associated processes such as transcription, a detailed knowledge of the kinetics of DNA binding, as well as the mechanical consequences of their binding to DNA, is needed. In fact, there is a tight relationship between DNA conformation and its interaction with proteins: wrapping of DNA around proteins results in condensation of the DNA, both by bringing together the ends of the DNA and by inducing supercoiling in the surrounding DNA, which results in additional condensation by plectoneme formation. DNA supercoiling, induced by externally applied torque, will facilitate the binding of proteins that wrap or bend the DNA with similar chirality. Force and extension play similar related roles: some proteins induce stretching of DNA upon binding. Pre-stretching of DNA can facilitate binding or function of these proteins, making force an important parameter in the study of nucleo-protein filaments [2].

DNA forms the common thread in chromatin organization, and it is informative to gather and compare mechano-chemical experimental data on chromatin organization by different proteins and from different species, to gather a comprehensive picture of DNA organization. In terms of DNA mechanobiology, as well as from an evolutionary perspective, it is highly desirable to establish a physical understanding of chromatin organization throughout the kingdoms of life. In this chapter, we present protocols to set up single-molecule force spectroscopy experiments on DNA tethers interacting with a variety of chromatin proteins. Although some examples stem from studies of eukaryotic proteins, it is our experience that the methods for sample preparation, execution of the experiments, and data analysis are similar, and we provide protocols that can be used to study bacterial chromatin, as well as other nucleoprotein complexes.

1.2 Magnetic Tweezers

Magnetic Tweezers (MT) can probe the mechanical response of individual biomolecules to applied force and twist. The first MT experiments have been instrumental in understanding the mechanical behavior of bare DNA [3–5]. Since then, mechanical and structural properties of more complex biomolecules and DNA-protein complexes have been studied [6–11]. MT are especially suitable to this end since they require limited hardware investments, uniquely control both force and linking number, and have a relatively high throughput compared to optical tweezers – the most common alternative single-molecule force spectroscopy technique. Recent developments are an increase in maximum force to beyond the overstretching force for DNA [8, 12, 13], multiplexing capabilities, such that hundreds of molecules can be measured in parallel [14, 15] and freely orbiting magnetic tweezers (FOMT)

[16] or magnetic torque tweezers (MTT) [17], that allow for (partial) relief of twist or direct torque measurement. Another advance is the replacement of permanent magnets by electromagnets in electro-magnetic tweezers [5, 18]. Since there are no moving parts in electro-magnetic tweezers, the only factor limiting the speed of the experiments is the viscous drag of the magnetic beads themselves. Finally, combinations with (single-molecule) fluorescence have been reported [19–22]. These modern variations of the technique require similar sample preparation protocols, and can be seen as extensions of typical MT experiments.

A typical MT experiment involves a single DNA molecule tethered between a paramagnetic bead and a functionalized glass slide. The three-dimensional position of the bead is recorded in time by video microscopy and extracted by image processing. The bead can be manipulated by translation or rotation of a pair of magnets that are held above the tether. The force and torque that are exerted on the paramagnetic bead result in a restoring force and torque in the tethered molecule, which provides insight into the structure of the tether.

1.3 Measurement Schemes

MT can be operated in several measurement schemes. Traditionally, the magnet is fixed in a preset position during a measurement, resulting in a constant force during the experiment. For bare DNA, such a constant force measurement not only accurately reveals the extension of the tether, one can also directly calculate the force by quantification of the lateral movements of the bead [4]. Application of the equipartition theorem yields that the force must equal the product of thermal energy and extension, divided by the variance of the lateral fluctuations (*see* Subheading 3.7). Repeated measurements at different magnet positions yield a calibration of the force as a function of the magnet position [4, 23–26], which typically follows a mono- or bi-exponential decay [25]. The same measurement also provides an accurate map of the force-extension relation of DNA, which can be described by a Worm-like Chain (WLC) model. For transient protein-DNA interactions, a constant force experiment can directly reveal changes in extension of the tether upon flushing the protein solution in or out [27–29].

Having established a force calibration relation, one can revert to dynamic force spectroscopy, in which the extension of the tether is monitored during repositioning of the magnet [30]. Such a measurement scheme aids in revealing the force dependence of structures that may rupture or dissociate when increasing forces are applied. For instance, we have noticed that histone proteins readily dissociate when chromatin fibers are exposed to forces exceeding several pN [11].

Next to shifting the position of the magnets with respect to the tethered bead, the magnets can be rotated around the axis of the

tether. Since the torque on the bead is proportional to the strength of the magnetic field (*see* **Note 1**), and is typically much larger than the opposing torque provided by the DNA tether, the bead will immediately follow the rotation of the magnets. This can lead to accumulation of twist in the tether, if the twist is not relieved by swiveling around the bond between the tether and bead or surface, or by swiveling around a nick in the DNA. Constraining the rotation of the DNA ends requires more than one chemical bond between the DNA strand and the bead or surface, which can be achieved by introducing multiple affinity tags in the DNA ends. Such a DNA substrate allows for direct control of the linking number of the topological domain formed by the bead and the surface. Here, we describe the preparation of both torsionally constrained and torsionally free DNA substrates, allowing for a broad spectrum of applications.

Controlling the linking number of the DNA tether opens up a range of measurement schemes. At constant twist, force-extension measurements can be done in a similar fashion as with torsionally unconstrained DNA. Any twist-stretch coupling will be evident in the force-extension curve. Protein-induced (un-)wrapping will accumulate, which may affect the binding of subsequent proteins. A torsionally constrained tether can be pre-twisted before force spectroscopy experiments, to probe the chiral properties of the tether over a wide range of linking numbers [31]. The mechanical response of the tether to torque can also be measured directly by twisting the tether under constant force. In conclusion, a wide variety of single-molecule manipulation schemes are available with MT force spectroscopy.

1.4 Data Analysis

The mechanical properties of DNA provide a reference for data analysis of nucleo-protein filaments. The extension of DNA as a function of force can be described accurately by the WLC model [32]. At forces larger than 10 pN, the extensible WLC model provides a better fit [32, 33]. DNA extension as a function of twist requires a more elaborate analysis, using numerical methods, due to the coexistence of three conformations, i.e. twisted, melted, and plectonemic DNA [31]. However, for force and twist regimes where only two of these states prevail, analytical solutions exist [34]. In general, it is important to make sure that the force extension curves of bare DNA, under conditions relevant for studying chromatin, fit well to these models, prior to the study of more advanced structures, as protein-induced changes in extension can be difficult to interpret without proper reference.

Wrapping, bending, twisting, or deforming DNA otherwise will result in a change in extension of the tether that may depend on force and twist. In a force-extension curve, this may lead to a change in persistence length, contour length, stretch modulus, and/or twist persistence length. In Subheadings 3.10 and 3.11

we review some models that can be fitted to extract these parameters, as well as more dynamic methods to extract the kinetics of protein binding.

Below, we describe in detail how to make DNA substrates suitable for MT. Note that many variations of DNA substrates exist that the reader can adapt to specific needs. We report a detailed protocol for the assembly of DNA tethered beads in a flow cell, ready for measurements, and share some typical procedures for data analysis. These should provide a solid foundation to study chromatin at the single-molecule level with MT.

2 Materials

2.1 Stock Solutions

All aqueous buffers are prepared with ultrapure Milli-Q water and stored at 4 °C.

- Measurement buffer: 10 mM HEPES pH 7.5, 100 mM KCl, 10 mM NaN_3 (sodium azide), 0.2% Bovine Serum Albumin (BSA) (heat shock fraction, pH 7, ≥98%), 0.1% TWEEN 20 (*see* **Note 2**).
- Passivation Buffer: 10 mM sodium azide, 3.6% BSA and 0.1% TWEEN 20 (*see* **Note 2**).

2.2 Isolation of DNA Plasmids

- Desired DNA plasmid (*see* **Note 3**).
- XL1-Blue competent cells.
- Heat block, such as the Eppendorf ThermoMixer.
- Orbital shaker.
- Microcentrifuge, such as the Eppendorf 5424R.
- Lysogeny broth (LB) medium (*see* **Note 4**).
- LB agar plates (*see* **Note 4**).
- 500 mL Erlenmeyer flask.
- DNA plasmid isolation kit, such as the NucleoBondXtra Midi kit.

2.3 Digestion and Labeling of DNA

- Isolated DNA plasmid (*see* **Note 3**).
- Restriction enzymes (*see* **Note 3**).
- Reaction buffer for digestion.
- DNA purification kit, such as the Wizard SV Gel & PCR cleanup kit.
- Spectrophotometer for micro volume quantitation of nucleic acids, such as the BioDrop μLITE.
- Agarose for DNA/RNA-electrophoresis.
- Horizontal agarose gel electrophoresis system.

- dNTPs (100 mM).
- DNA Ladder.
- Ethidium Bromide (10 mg/mL).
- Klenow fragment.
- Reaction buffer for Klenow Fragment.
- Digoxigenin-labeled ddUTP (1 mM).
- Biotin-labeled ddUTP (1 mM).
- DNA plasmid as a template to construct DNA handles (*see* **Note 3**).
- Forward primer to construct DNA handles (*see* **Note 3**).
- Backward primer to construct DNA handles (*see* **Note 3**).
- T4 DNA Ligase.
- Reaction buffer for T4 DNA Ligase.
- DNA PCR kit, such as the FastStart Taq DNA Polymerase kit.
- Digoxigenin-labeled dUTP (1 mM).
- Biotin-labeled dUTP (1 mM).
- Sodium dodecyl sulfate (SDS) ≥98.5% (GC).
- PCR Thermal Cycler, such as the Bio-Rad T100.
- UV Transilluminator, such as the Bio-Rad UVT 2000.
- Imaging system for visualization of agarose electrophoresis gels, such as the ChemiDoc Imaging Systems.

2.4 Flow Cell Assembly

- High-precision crossover tweezers.
- Carbon steel scalpel.
- N_2 spray gun.
- 2-Propanol.
- Custom-built aluminum flow cell body.
- Custom-built Perspex mold.
- Embossing Tape (9 mm).
- Cover slip 24 × 40 mm (#1.5 thickness).
- Cover slip 24 × 60 mm (#1.5 thickness).
- PDMS kit, such as the Sylgard 184 Silicone Elastomer Kit.
- FEP Tubing (1/16″ outer diameter × 0.020″ inner diameter).
- Flexible electrical wire (of approximately 0.02″ OD).
- Fitting nut for 1/16″ outer diameter tubing.
- Ferrule for 1/16″ outer diameter tubing.
- Vacuum controller.
- 5 mL syringe.
- M4 bolts.

- Ultrasonic cleaning bath.
- Glass staining trough.
- 100 mL glass beaker.
- Microscope slide (75 × 26 × 1 mm).

2.5 Bead-Tether Assembly

- DNA substrate (torsionally constrained or torsionally free).
- Magnetic rack for bead separation.
- Streptavidin-coated paramagnetic beads (*see* **Note 5**).
- Nitrocellulose—1% in amyl acetate.
- Pentyl acetate (amyl acetate) puriss. p.a., ≥98.5% (GC).
- Anti-digoxigenin (AD).
- 1 mL syringe.
- 18G single-use needle.
- Flexible tubing (0.8 mm inner diameter/0.8 mm wall).
- Peristaltic pump (which can exert flow rates of approximately 200 μL/min).

2.6 Microscope

- Microscope objective (flat-field, NA = 1.3, 40×, oil immersion, field number 25).
- CMOS camera (25 Mpix, 8 bit, 30 fps, Camera Link—Full configuration).
- Infinity-corrected tube lens ($f = 200$ mm).
- Kinematic pitch/yaw adapter.
- LED-collimator ($\lambda = 645$ nm, 100 μW, 20 mA).
- Frame grabber for Camera Link—Full configuration.
- Multi-core PC (e.g., 10-core) with 32GB DDR3 memory.
- XYZ piezo stage (20 μm travel in Z [0.1 nm resolution]).
- Cube magnets (N50 magnetized, 5 mm).
- Stepper motor controller (6-axis, 1–256 micro-steps).
- Hollow-shaft stepper motor (2-phase, NEMA 8).
- Two translation stages to mount objective and magnets (20 mm travel [0.1 μm resolution]).
- XY manual positioning stage (20 mm travel in XY, 65 mm aperture).
- Ø1″ and Ø2″ broadband dielectric mirrors ($\lambda = 400$–750 nm).

3 Methods

3.1 Isolation of DNA Plasmids

1. Take 100 μL of XL1-Blue Competent Cells (at a concentration of approximately 3×10^8 cells/mL) from the −80 °C freezer and let thaw on ice (*see* **Note 6**). Add approximately 1.0 μL of the desired DNA plasmid (*see* **Note 7**) to the Eppendorf tube, gently mix, and incubate on ice for a minimum of 30 min.
2. Heat-shock the sample for 90–120 s at 42 °C by means of the thermomixer. Do not shake the sample or exceed the indicated time.
3. Place the tube back on ice for 60 s and add 900 μL of LB medium. Gently shake (240 rpm) and incubate the tube for 30–60 min at 37 °C.
4. Centrifuge the tube for 60 s at 20,000 × *g*, remove approximately 90% of the supernatant, and resuspend the sample.
5. Spread the sample on a LB agar plate containing the appropriate antibiotic (here: ampicillin) and incubate (upside down) at 37 °C overnight.
6. Pick up one colony and put it into a 500 mL Erlenmeyer flask containing 250 mL LB medium and the appropriate antibiotic (here: ampicillin). Incubate on the shaker (320 rpm) at 37 °C, overnight.
7. Isolate and purify the DNA plasmid DNA the NucleoBond®X-tra Midi kit (Macherey-Nagel) following the manufacturer's protocol.

3.2 Digestion and Labeling of Torsionally Free DNA Constructs

1. Digest the DNA plasmid by mixing the reagents listed in Table 1 in an Eppendorf tube and incubate at 37 °C, overnight (*see* **Notes 8** and **9**).
2. Inactivate BsaI by incubating for 20 min at 65 °C (*see* **Note 10**).
3. Purify DNA with Promega Wizard SV Gel & PCR cleanup kit following the manufacturer's protocol (*see* **Notes 11** and **12**). Determine the DNA concentration after purification by measuring A260 nm peak absorbance with the spectrophotometer and store a microliter of the reaction mixture for agarose gel electrophoresis.
4. Label one end of the digested DNA by mixing the reagents listed in Table 2 in an Eppendorf tube and incubate at 37 °C, for 2 h (*see* **Note 13**). The end concentration of (labeled) dNTPs should be 20 μM (*see* **Note 14**). Note that in this step the DNA is labeled with ddUTP rather than dUTP (*see* **Note 15**).

Table 1
Composition of digestion reaction mixture I, used for preparation of the torsionally free construct

pUC18 DNA plasmid (with 601-array) (2000 ng/μL)	20 μL	40 μg
NEBuffer 3.1 (10× concentrated)	80 μL	–
Milli-Q water	696 μL	–
BsaI (10 U/μL)	4 μL	40 units
Total volume	800 μL	

Table 2
Composition of labeling reaction mixture I, used for preparation of the torsionally free construct

BsaI-digested pUC18 DNA (with 601-array) (400 ng/μL)	50 μL	20 μg
Reaction buffer for Klenow Fragment (10× concentrated)	10 μL	–
Digoxigenin-11-ddUTP (1 mM)	2 μL	20 μM
dGTP (1 mM)	2 μL	20 μM
dCTP (1 mM)	2 μL	20 μM
Milli-Q water	24 μL	–
Klenow Fragment, LC (2 U/μL)	10 μL	20 units
Total volume	100 μL	

Table 3
Composition of DNA plasmid digestion reaction mixture II, used for preparation of the torsionally free construct

Single-labeled pUC18 DNA (with 601-array) (200 ng/μL)	50 μL	10 μg
NEBuffer 3.1 (10× concentrated)	20 μL	–
Milli-Q water	128 μL	–
BseYI (5 U/μL)	2 μL	10 units
Total volume	200 μL	

5. Purify DNA and store a microliter of reaction mixture for agarose gel electrophoresis.
6. Digest the DNA labeled in the previous steps with a single digoxigenin by mixing the reagents listed in Table 3 in an Eppendorf tube and incubate at 37 °C, overnight.
7. Inactivate BseYI by incubating for 20 min at 80 °C.
8. Purify DNA and store a microliter of reaction mixture for agarose gel electrophoresis.

Table 4
Composition of labeling reaction mixture II, used for preparation of the torsionally free construct

BseYI-digested single-labeled pUC18 DNA (with 601-array) (150 ng/μL)	50 μL	7.5 μg
Reaction buffer for Klenow Fragment (10× concentrated)	10 μL	–
Biotin-16-ddUTP (1 mM)	2 μL	20 μM
dCTP (1 mM)	2 μL	20 μM
Milli-Q water	32.25 μL	–
Klenow Fragment, LC (2 U/μL)	3.75 μL	7.5 units
Total volume	100 μL	

9. Label the second extremity of the DNA by mixing the reagents listed in Table 4 in an Eppendorf tube and incubate at 37 °C, for 2 h.
10. Purify DNA and store a microliter of reaction mixture for agarose gel electrophoresis.
11. Check the digestion reaction products by agarose gel electrophoresis using a 1% agarose gel. Prestain the gel with ethidium bromide by diluting the stock 10,000× in the agarose solution before it cures (*see* **Note 16**). Load the gel with the products obtained at different steps of the labeling procedure, flanked by two DNA ladders. Load approximately 50–100 ng of DNA per lane. Run the gel in TBE buffer in a horizontal agarose gel electrophoresis system for 1–2 h at 90 mV or until bands are adequately separated. Visualize the agarose electrophoresis gel in an imaging system suitable for ethidium bromide staining.

3.3 Digestion and Labeling of Torsionally Constrained DNA Constructs

1. Digest the DNA plasmid by mixing the reagents listed in Table 5 in an Eppendorf tube and incubate at 37 °C, overnight (*see* **Notes 8** and **9**). The two enzymes are used at the same time.
2. Inactivate restriction enzymes by incubating for 20 min at 80 °C.
3. Separate the digestion products by agarose gel electrophoresis using a 1% agarose gel. Prestain the gel with ethidium bromide by diluting the stock 10,000× in the agarose solution before it cures (*see* **Note 16**). Load the digested DNA in a large slot, obtained by using 1.0 mm comb and merging several lanes together (with tape), flanked by two DNA ladders. Before loading, add SDS to an end concentration of 0.5% (*see* **Note 17**).
4. Image the gel on the UV transilluminator to locate the bands of interest. Minimize the time on the UV transilluminator (*see* **Note 18**). Use a scalpel to cut the digested DNA from the gel, to separate the construct of interest from the DNA backbone.

Table 5
Composition of the reaction mixture for digestion of the DNA plasmid for the torsionally constrained construct

pUC18 DNA plasmid (with 601-array) (2000 ng/μL)	50 μL	100 μg
NEBuffer 3.1 (10× concentrated)	200 μL	–
Milli-Q water	1720 μL	–
BsaI (10 U/μL)	10 μL	100 units
BseYI (5 U/μL)	20 μL	100 units
Total volume	2000 μL	

Table 6
Composition of the PCR mixture for production of DNA handles for the torsionally constrained construct

Template pUC18 DNA plasmid (100 ng/μL)	1 μL	100 ng
FastStart Taq DNA Polymerase (5 U/μL)	0.5 μL	2.5 units
Forward primer to construct DNA handles (10 μM)	1 μL	100 nM
Backward primer to construct DNA handles (10 μM)	1 μL	100 nM
PCR Reaction Buffer + $MgCl_2$ (10× concentrated)	10 μL	
dNTP (10 mM)	1 μL	100 μM
Biotin-16-dUTP (1 mM) or Digoxigenin-11-dUTP (1 mM)	1 μL	10 μM
Milli-Q water	84.5 μL	
Total volume	100 μL	

5. Purify the digested DNA from the gel with the Promega Wizard SV Gel & PCR cleanup kit following the manufacturer's protocol (*see* **Note 19**).
6. Construct two sets of handles with multiple biotin or digoxigenin affinity tags by polymerase chain reaction (PCR) (*see* **Notes 20** and **21**). Mix the reagents listed in Table 6 in two separate Eppendorf tubes and place them in the thermal cycler. Use the PCR program depicted in Table 7 and store the tubes at 4 °C. Here, the DNA is labeled with dUTP rather than ddUTP (*see* **Notes 15** and **22**).
7. Digest the biotin PCR product with BsaI and the digoxigenin PCR product with BseYI by mixing the reagents listed in Table 8 in two different Eppendorf tubes and incubate at 37 °C, overnight.

8. Inactivate restriction enzymes by incubating for 20 min at 80 °C.
9. Purify digested PCR products with the Promega Wizard SV Gel & PCR cleanup kit following the manufacturer's protocol (*see* **Notes 12** and **23**).
10. Ligate the digested DNA plasmid with the digested PCR products containing the multiple biotin or digoxigenin affinity tags. The digested DNA and PCR products are mixed in an equimolar ratio. Mix the reagents listed in Table 9 in an

Table 7
The PCR program to produce the DNA handles. Steps 2–4 are cycled 30 times

Step	Temperature (°C)	Time (s)	Function
1	95	240	Melt the DNA
2	95	30	Melt the DNA
3	50	45	Anneal the primer to the ssDNA
4	72	60	Elongation by DNA polymerase
5	72	300	Fill in any protruding ends

Table 8
Composition of the digestion reaction mixture for generation of the DNA handles of the torsionally constrained construct

Multi-biotin PCR product or Multi-digoxigenin PCR product (200 ng/μL)	100 μL	20 μg
NEBuffer 3.1 (10× concentrated)	100 μL	–
Milli-Q water	798 μL or 796 μL	–
BsaI (10 U/μL) or BseYI (5 U/μL)	2 or 4 μL	20 units
Total volume	1000 μL	

Table 9
Composition of the ligation reaction mixture to produce the torsionally constrained construct

Digested pUC18 DNA (with 601-array) (100 ng/μL)	300 μL	30 μg
Digested multi-biotin PCR product (100 ng/μL)	50 μL	5 μg
Digested multi-digoxigenin PCR product (100 ng/μL)	50 μL	5 μg
T4 DNA Ligase Reaction Buffer (10× concentrated)	80 μL	–
Milli-Q water	315 μL	–
T4 DNA ligase (400 U/μL)	5 μL	2000 units
Total volume	800 μL	

Eppendorf tube and incubate at 4 °C, overnight. Before ligase is added, store a microliter of reaction mixture for agarose gel electrophoresis (*see* **Notes 24** and **25**).

11. Inactivate T4 ligase by incubating for 20 min at 65 °C.
12. Purify ligated DNA, elute at least two times (*see* **Note 26**). Store a microliter of reaction mixture for agarose gel electrophoresis.
13. Check the ligation reaction by agarose gel electrophoresis using a 1% agarose gel. Prestain the gel with ethidium bromide by diluting the stock 10,000× in the agarose solution before it cures (*see* **Note 16**). Load the gel with unligated and ligated reaction products, flanked by two DNA ladders. Load approximately 50–100 ng of DNA per lane. Run the gel in TBE buffer in a horizontal agarose gel electrophoresis system for 1–2 h at 90 mV or until bands are adequately separated. Visualize the agarose electrophoresis gel in an imaging system adjusted for ethidium bromide staining.

3.4 Flow Cell Assembly

1. Mix the two components of the PDMS in a ratio 10:1 (viscous monomer: non-viscous elastomer), approximately 2 mL per flow cell. Mix carefully and thoroughly (*see* **Note 27**). Degas the PDMS by keeping the mixture at 300 mBar for 2 h using the vacuum controller.
2. Clean the Plexiglas molds and aluminum flow cell body with 2-propanol and Milli-Q water, and dry in a stream of N_2. On the mold, place a strip of embossing tape which shapes the channel. Using a needle, puncture two holes through the tape, which are used to pass through two electrical wires. Place the 24 × 40 mm cover slip into the aluminum body. For the in- and outlet of the flow cell, insert two pieces of 3 cm flexible electrical wire into approximately 10 cm of FEP tubing. Assemble the fitting nut and ferrule onto the tubing. Mount the tubing onto the side of the flow cell. Guide the tubing through the side of the flow cell and secure the fitting nut tightly. Guide the electrical wire through the Plexiglas molds into the tubing. Fix the flow cell body and mold with the M4 bolts. The physical assembly of the flow cells is illustrated in Fig. 1.
3. Using a 5 mL syringe, slowly flow the degassed PDMS into the flow cell. Prevent bubbles as much as possible. Cure the flow cells at 65 °C for at least 5 h (up to overnight).
4. Remove electrical wires, creating connection channels in the PDMS to the in- and outlet of flow cell. Remove the Perspex mold. Use a scalpel to widen the channel in the flow cell to approximately 10 mm. Make sure the channel is connected to the in- and outlet of the flow cell.

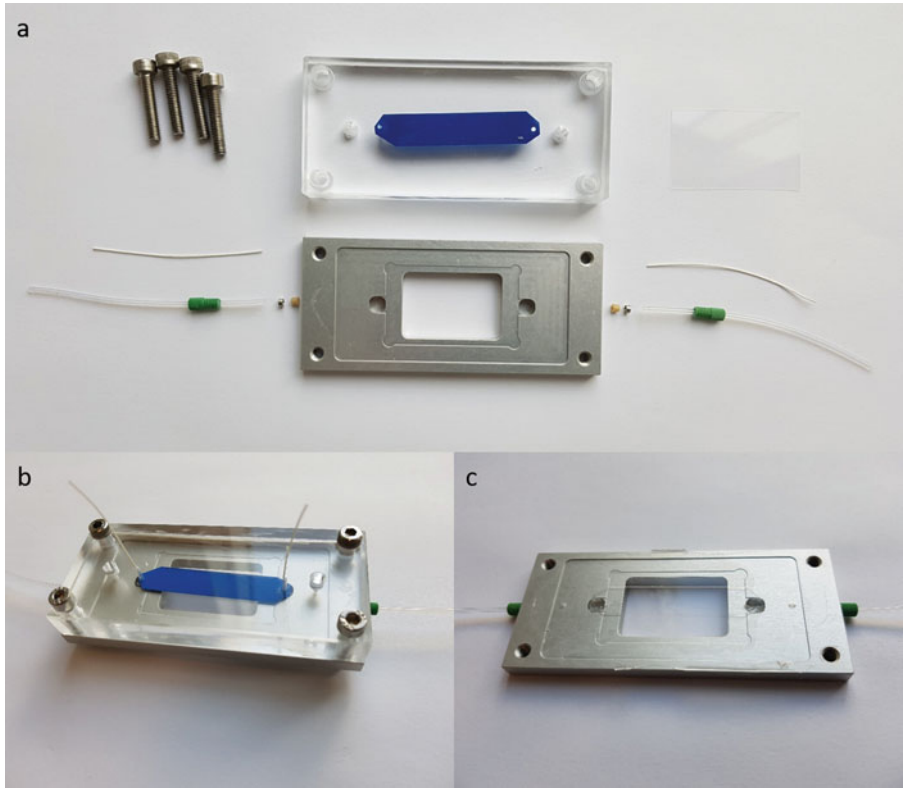


Fig. 1 The assembly of the custom-built flow cells is a three-step process. (**a**) Tubes and slides are assembled. The 24 × 40 mm glass slide is placed in the designated cavity of the aluminum body. Electric wire is inserted into the FEP tubing and guided upward through the mold. (**b**) After mounting the Perspex mold, the assembled flow cell is ready for casting PDMS. (**c**) Once the PDMS is cured, the electric wire and mold can be removed and the flow cell is covered by a 24 × 60 mm biochemically functionalized glass slide

5. Put 12 24 × 60 mm cover slips in a glass staining trough and sonicate in ultrasonic cleaning bath for 15 min in 2-propanol. Dry the cover slips in a stream of N_2 or allow the cover slips to dry in air.
6. Put 5 mL of 0.1% nitrocellulose in amyl acetate in a 100 mL glass beaker. Place a clean microscope object slide vertically in the solution. Using reversed tweezers, place one of the sonicated cover slips into the beaker directly next to the microscope slide. Capillary forces will fill the gap between the slides with nitrocellulose solution and will quickly coat the cover slip (*see* **Note 28**).
7. Once the gap between the cover slip and the object slide is filled, use reversed tweezers to carefully slide the cover slip from the object slide. Dry the cover slip in a stream of N_2. Place the cover slip with the nitrocellulose-coated side at the bottom of the flow cell, sealing the flow channel or store the coated cover slip for later use.

3.5 Bead-Tether Assembly

1. Connect the tube on one side of the flow cell to an empty syringe using a short length of flexible silicone tubing and a needle. Use suction to replace fluids. Carefully flush the flow cell with 1 mL of Milli-Q water, incubate for approximately 30 s.
2. Replace the Milli-Q water with 300 μL of 10 ng/μL anti-digoxigenin solution. Incubate at 4 °C for 2 h.
3. Passivate the flow cell by flowing in 1 mL of passivation buffer. Incubate at 4 °C overnight (*see* **Note 29**).
4. Flush the flow cell with 1 mL of measurement buffer.
5. Dilute 1 μg of DNA in 500 μL measurement buffer. Flush the flow cell with DNA and incubate at room temperature for 10 min (*see* **Note 30**).
6. Vortex and wash the 1 μg/μL paramagnetic beads in Milli-Q water by means of the magnetic rack and dilute 1 μL of paramagnetic beads in 500 μL measurement buffer. Flush the flow cell with beads and incubate at room temperature for 10 min (*see* **Note 31**).
7. Slowly flush out the excess beads from the flow cell with 500 μL of measurement buffer using a peristaltic pump at 500 μL/min (*see* **Note 32**).
8. The flow cell is now ready for measurements on bare DNA or incubation with DNA-binding proteins.

3.6 Initial Bead Selection and Height Calibration

1. Position the magnets at such a distance that force is negligible. In our setup, depicted in Fig. 2 (*see* **Note 33**), we position the magnets 10 mm above the sample, resulting in a force lower than 0.05 pN.
2. Place the flow cell containing the (pre-incubated) sample onto the microscope.
3. Put the beads in focus using the objective mounted onto the stepper motor. With the piezo stage, move the objective 10 μm above focus so distinct diffraction rings appear around the beads (*see* **Note 34**).
4. Select regions-of-interest (ROIs) of 150 × 150 pixels containing single, isolated beads (manually or by a bead-finding algorithm) (*see* **Note 35**).
5. Identify an immobile bead as a reference bead, and use this bead for phase calibration. Move the bead 20 μm in Z while cross-correlating the bead images with a set of computer-generated reference images. Calibrate the phase of the cross-correlation peak with bead height. When there are no immobile beads in the field of view, phase calibration can also be done at an intermediate force, i.e., 1 pN, to reduce the thermal fluctuations of the tethered bead, while keeping the tether intact.
6. Start tracking the beads (*see* **Notes 36** and **37**).

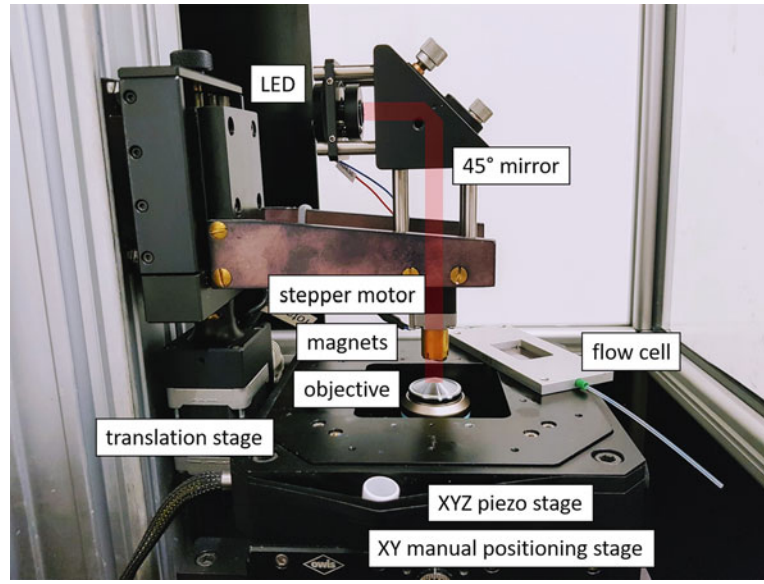


Fig. 2 The magnetic tweezers setup. A collimated LED is mounted in the kinematic pitch/yaw adapter and coupled into a hollow-shaft stepper motor by a Ø1″ mirror. The incoming light passes between a pair of magnets into the objective. The flow cell is mounted on the top of an XYZ piezo stage, above the objective (here, the flow cell is shifted to the side for clarity). The piezo stage is mounted on the top of an XY manual positioning stage, for coarse movements. The magnets and objective are mounted onto two translation stages. Below the objective, the light is coupled into the camera through an infinity-corrected tube lens by means of a Ø2″ mirror. The light path is depicted in red

3.7 Force Calibration

1. For force calibration, prepare a flow cell containing torsionally free DNA tethers. Preferably, the tethers are relatively long with respect to the magnetic bead size (*see* **Note 38**). Start the measurements with the magnets far away, exerting negligible force. Move the magnets quickly down to the position that needs to be calibrated, keep the magnets fixed for 120 s (*see* **Note 39**), and move the magnets upward. Append two stretches of negligible force before and after the experiment to provide a reference for drift correction. If possible, increase the sample rate f_s of the camera (*see* **Note 38**). Perform multiple calibration measurements at different magnet positions.
2. Load the data into data analysis software such as LabVIEW, Origin, Python, R, or Excel. Correct the drift by subtracting a term proportional with time from the extension prior to further analysis (*see* **Note 40**).
3. Exclude anomalous events, such as sticking (*see* **Note 41**), double bead attachment to a single tether (*see* **Note 42**), or double tether attachment to a single bead (*see* **Note 43**).

4. Evaluate the X-coordinates (the direction of the magnetic field) in the time-domain where the magnet was exerting a constant force. Compute the single-sided power spectral density (PSD) in units of $\mu m^2_{\mathrm{rms}}/\mathrm{Hz}$.
5. Exclude all spectral components below 1 Hz, to discard $1/f$ noise and to minimize the effect of drift.
6. Fit the PSD to a blur and aliasing corrected PSD function to obtain the cut-off frequency to calculate the force

$$\mathrm{PSD}_{\mathrm{corrected}}(f, f_{\mathrm{c}}) = \frac{\sigma^2}{f_{\mathrm{s}}} + \sum_{n=-1}^{1} \frac{k_{\mathrm{B}}T}{\gamma\pi^2(f_{\mathrm{c}}^2 + (f + nf_{\mathrm{s}})^2)} \left(\frac{\sin\left(W\pi(f + nf_{\mathrm{s}})\right)}{W\pi(f + nf_{\mathrm{s}})} \right)^2. \tag{1}$$

See **Note 44** for a derivation of Eq. 1 and a full description of the parameters.

7. Plot the Force as a Function of magnet position h, and fit the data with a bi-exponential decay [25]

$$F(h) = F_{\mathrm{max}}\left(\alpha \exp\left(\frac{-h}{h_1}\right) + (1-\alpha)\exp\left(\frac{-h}{h_2}\right)\right). \tag{2}$$

This relation is unique for each combination of magnets and the type of magnetic beads. For out setup and 2.8 μm beads, we obtained $F_{\mathrm{max}} = 85$ pN, $\alpha = 0.7$, $h_1 = 1.4$ mm and $h_2 = 0.8$ mm.

3.8 Force Spectroscopy Experiments

1. For a dynamic force spectroscopy experiment on a torsionally free substrate, move the magnets down, increasing the force, and subsequently upward, releasing the force. Move the magnets at a speed of 0.2 mm/s (*see* **Note 45**). Minimize the dwell time at the highest force to reduce the chance of a tether break. Several variations in trajectories are found in **Note 46**. Optionally, two stretches of low force can be appended before and after the experiment for drift correction.
2. For a dynamic force spectroscopy experiment on a torsionally constrained substrate, follow the same procedure as above. Optionally, pre-twist the torsionally constrained substrate before a dynamic force spectroscopy experiment: rotate the magnetic field while the stretching force upon the tether is kept low. Forces as low as 0.05 pN are sufficient for the bead to follow the rotation of the magnetic field. Rotate the magnets with a speed of 2 rotations per second (*see* **Note 47**).

3.9 Rotational Spectroscopy Experiments

1. For rotational spectroscopy on a torsionally constrained substrate, set the force by fixing the magnet height throughout the experiment. Rotate the magnets with a speed of 2 rotations per second. Apply positive twist as desired and relax the tether. Minimize the dwell time at the highest number of rotations (*see* **Note 48**). Apply an equal amount of negative twist at the same rotational speed, and relax the tether. Optionally, two stretches with fixed rotation can be appended before and after the experiment for drift correction.

3.10 Data Analysis of Force Spectroscopy Experiments

1. Load the data into data analysis software such as LabVIEW, Origin, Python, R, or Excel. Calculate force from magnet height using Eq. 2. Correct the drift by subtracting a term proportional with time from the extension, such that the parts of the extension trace that were appended to the start and the end have the same height (*see* **Note 40**). Alternatively, a (linear) time-dependent term can be included in the fitting model. Plot the force as a function of extension (*see* **Note 49**).
2. Exclude anomalous events, such as sticking (*see* **Note 41**), double bead attachment to a single tether (*see* **Note 42**), or double tether attachment to a single bead (*see* **Note 43**).
3. Fit the data.
   (a) For bare DNA, fit an extensible WLC model to the data:

$$z(F,t) = L_c\left(1 - \frac{1}{2}\frac{\sqrt{k_B T}}{F L_p} + \frac{F}{S}\right) + z_0 + dt \qquad (3)$$

   where z is the extension, L_c the contour length, L_p the persistence length, F the force, S the stretch modulus, k_B Boltzmann's constant, T the absolute temperature, and z_0 an offset in extension (*see* **Note 50**). Optionally, a drift, comprising amplitude d and time t, can be included.

   (b) For (bacterial) chromatin, fit appropriate models to the data. Generally, the effect of protein binding can be captured in a modified WLC model by adjusting fitting parameters (*see* **Note 51**). An example is shown in Fig. 3, reproduced from Van Noort et al. [6], where the effect of HU-binding on the mechanical parameters of a DNA tether is shown. Analysis of partial nucleoprotein filaments can be performed when the mechanical parameters of homogeneous, fully saturated nucleoprotein filaments have been established. Then, changes in extension can be described as a linear combination of the extensions of parts of bare DNA and parts of nucleoprotein filament

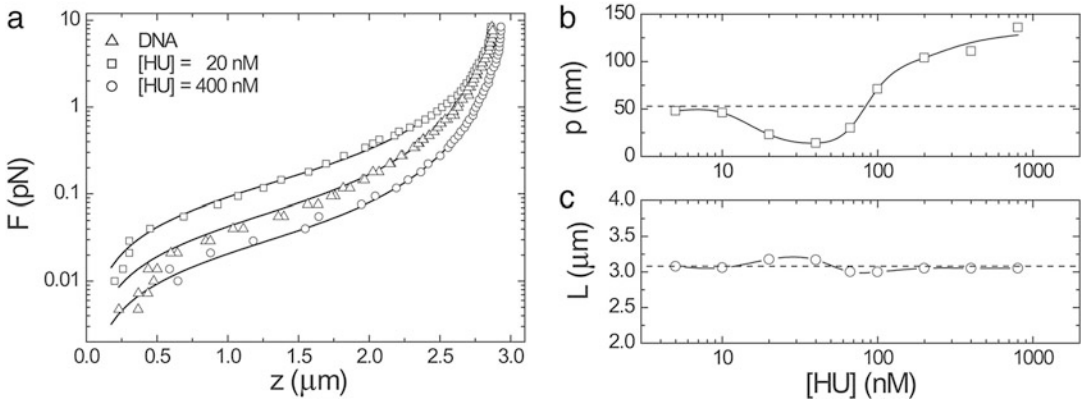


Fig. 3 The concentration of nucleoid-associated protein HU has a dual effect on the measured stiffness of a DNA tether. (**a**) The force-extension curves of DNA incubated with increasing concentration of HU. Solid lines represent fits to the WLC model (Eq. 3). The experiments reveal tether softening at nM HU concentration and tether stiffening at μM HU concentration. Experiments were performed in 60 mM of KCl and 20 mM HEPES (pH 7.9). (**b** and **c**) The persistence length and contour length as a function of HU concentration. The dashed lines represent bare DNA. Individual HU dimers introduce kinks in the DNA at nM HU concentration [55], resulting in a reduced persistence length. A rigid nucleoprotein filament is formed at high HU concentration. HU does not change the contour length. Reproduced from ref. 6 with permission from PNAS (Copyright (2004) National Academy of Sciences, U.S.A.)

[9, 35]. An example of a complex, dynamic chromatin structure is plotted in Fig. 4, showing force spectroscopy on a nucleosomal array reconstituted from tandem repeats of the *Widom* 601 sequence and eukaryotic histone proteins [36] (*see* **Note 52**).

3.11 Data Analysis of Rotational Spectroscopy Experiments

1. Load the data into data analysis software. Correct the drift following the same procedure as that of the force spectroscopy experiments. Correct the offset in Z by computing the first percentile of the extension as a measurement of the surface and subtract it from all Z coordinates. Calculate linking number density from magnet rotation (*see* **Note 53**). Plot the relative extension as a function of linking number density.
2. Exclude disruptive events, following the same procedure as that of the force spectroscopy experiments (*see* **Notes 54–57**).
3. Fit the data.
   (a) For bare DNA, to a three-state model [31], which describes the coexistence of twisted, melted, and plectonemic conformational states (*see* **Note 58**). An example is shown in Fig. 5.
   (b) For (bacterial) chromatin, fit an appropriate model to the data. The torsional response of chromatin fibers can be modeled as a linear combination of twisted and wrapped DNA [37] (*see* **Note 59**).

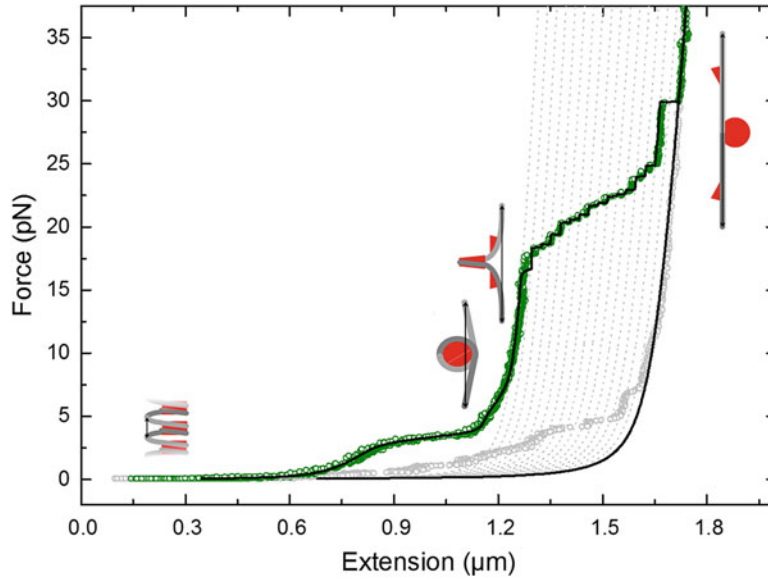


Fig. 4 Force spectroscopy on a eukaryotic chromatin fiber reveals different levels of compaction in 2 mM $MgCl_2$. The green curve is the stretch curve; the gray curve is the release curve. The black line describes a fit of the data to the statistical mechanics model [11]. The chromatin was reconstituted on an array of 15 Widom 601 nucleosome positioning sequences spaced by 50 base pairs [37]. The low-force regime shows a non-cooperative transition at 3.5 pN, typical for a solenoid chromatin fiber, where the interactions between the nucleosomes are broken and the outer turn of DNA unwraps from the histone core, forming a beads-on-a-string structure. At approximately 6 pN, another transition takes place where this structure is slightly extended. Above 10 pN stepwise unwrapping indicates release of the final wrap of DNA from the histone core. The dashed lines represent discrete 25 nm steps. The chromatin fiber follows the force-extension curve of bare DNA at forces exceeding 25 pN, and fits well with an extensible WLC model

4 Notes

1. Although the torque is proportional to the strength of the magnetic field, the force in MT is proportional to the gradient of the magnetic field and therefore rapidly reduces with distance.
2. In the measurement buffer, HEPES sets the pH and NaCl the ionic strength. TWEEN 20 is a detergent and prevents aggregation of proteins. Sodium azide is highly toxic and is added to prevent bacterial growth. When sodium azide is omitted, use sterile solutions or replace buffers frequently. BSA is used as a crowding agent and retains flow cell passivation. In the passivation buffer, a high concentration of BSA is used to block

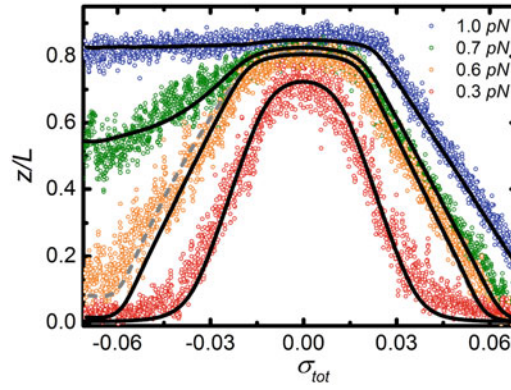


Fig. 5 The relative extension of a DNA tether as a function of linking number density shows an asymmetric response to twist at forces above 0.5 pN. At low force, plectonemes are formed for both negative and positive twists, reducing the measured extension. Negative twist induces DNA melting at higher force, which keeps the extension constant. The lines represent calculated extensions following a three-state model that distributes twisted, plectonemic, and melted regions in the DNA. Reproduced from ref. 32 with permission from Elsevier

nonspecific binding of DNA and beads and tethers to the surface.

3. For the DNA tethers, we generally use a pUC18-based DNA plasmid, carrying ampicillin resistance. We digest this plasmid with restriction enzymes BsaI and BseYI. For the handles of the torsionally constrained construct, we also use a pUC18 DNA plasmid as a template for PCR. The forward and backward primers were 5′ CTC CAA GCT GGG CTG TGT 3′ and 5′ GAT AAA TCT GGA GCC GGT GA 3′. Reconstituted chromatin fibers, such as the one shown in Fig. 4, were made using a plasmid containing 15 repeats of the *Widom* 601 positioning sequence spaced by 50 base pairs [36].
4. Make LB medium by diluting 10 g tryptone, 10 g NaCl and 5 g yeast extract in 1000 mL Milli-Q water. Make LB agar plates by diluting 10 g tryptone, 10 g NaCl, 5 g yeast extract, and 20 g agar in 1000 mL Milli-Q water.
5. We advise using Invitrogen Dynabeads™ M-270 Streptavidin or MyOne™ Streptavidin T1 paramagnetic beads. These beads have a relatively high iron content (MyOne: 26%/M-270: 14%) and monodispersed size distribution (CV $<$ 3%), which is important for reaching high forces and reliable bead tracking and force calibration.
6. When handling frozen competent cells, it is important to keep them cool. Frozen cells are very sensitive to temperature and will lose competence if not kept on ice.

7. Very low concentrations, down to picograms per microliter, are sufficient.
8. Use a reaction volume of at least 20–50 μL/μg DNA in a digestion.
9. Digestion of (hundreds of) micrograms of DNA plasmid requires overnight incubation. We frequently observe undigested DNA plasmid in agarose gel electrophoresis if we shorten the incubation time.
10. Inactivation temperatures vary for different restriction enzymes.
11. The Promega Wizard SV Gel & PCR cleanup kit uses high concentrations of chaotropic agents to perturb the hydrogen bond network of the solvent, enabling DNA to bind to the silica of the cleanup kit [38–41].
12. Multiple elutions with the spin column improve the yield. Furthermore, a longer incubation time increases the yield. For precious samples the increased yield outweighs the increased waiting time.
13. This step can be performed longer, up to overnight.
14. Keep the reaction volumes small to reduce cost.
15. The difference between dNTP and ddNTP is that the ddNTP is lacking an OH group at its 3′ carbon, which is required for DNA polymerization. It prevents incorporation of more than one affinity tag per DNA strand as the strand cannot be extended.
16. Ethidium bromide is an intercalating agent that fluoresces upon binding to DNA and illumination by a UV light source [42]. To visualize DNA in agarose gels it can be added to the agarose electrophoresis gel or to the running buffer.
17. BseYI can remain bound to DNA after digestion and alter the migration of DNA in agarose gels. To dissociate BseYI, add SDS to a final concentration of 0.5% or purify DNA before agarose gel electrophoresis, for instance with the Promega Wizard SV Gel & PCR cleanup kit.
18. Minimize UV exposure of the sample to prevent DNA damage. Accordingly, block UV exposure of the gel with a non-transparent material such as aluminum foil. Use flanking DNA ladders to estimate the position to cut the DNA fragment from the gel. After cutting, verify that the desired band is cut out by imaging on the UV transilluminator.
19. Dissolving the gel generally results in milliliters of DNA solution. To purify, load the spin column several times. Do not exceed the maximum amount of DNA that can be purified specified for the column.

20. DNA plasmid pUC18 was used as a template for the primers to construct DNA handles.
21. The PCR mixture contains 10% of biotin or digoxigenin labeled dUTP. The PCR product will thus have approximately 10–20 affinity tags incorporated.
22. Use labeled dUTPs instead of ddUTPs, since chain termination is undesired.
23. The small digestion products from the ends of the DNA (7 base pairs on the *BseYI*-side and 32 base pairs on the *BsaI*-side) are discarded by the Promega Wizard SV Gel & PCR cleanup kit, which has a cutoff at 100 base pairs.
24. The digested pUC18 DNA (with 601-array) is approximately seven times the length of the digested PCR products. To reach the approximate 1:1:1 molar ratio, use 1/7:1:1/7 weight ratio (multi-digoxigenin PCR product: digested pUC18 DNA (with 601-array): multi-biotin PCR product).
25. The unit definition of T4 DNA ligase can be quite ambiguous. In our protocol, we use excess amounts of T4 ligase for optimal yields.
26. Purifying a DNA construct with multiple affinity tags will strongly reduce the yield as the affinity tags increase DNA adhesion to the column. The yield can be improved by multiple elutions and longer incubation times (up to hours).
27. Mix thoroughly with, for instance, a flattened pipette tip to mix. Upon mixing, the PDMS starts to cure. Higher temperatures accelerate this process.
28. Nitrocellulose coating of the cover slip makes the surface hydrophobic, enhancing AD binding.
29. Incubate the flow cell with a high concentration of BSA to passivate the surface area. This step will reduce sticking of chromatin tethers during experiments.
30. DNA can be replaced by pre-incubated protein-DNA complexes, for instance, reconstituted or native chromatin.
31. During incubation of paramagnetic beads, keep the flow cell upright such that the beads move toward the coated cover slip by gravity.
32. During flushing, turn the flow cell upside down such that unbound beads will detach from the surface.
33. Our MT setup, depicted in Fig. 2, is home-built and equipped with custom control and tracking software written in LabVIEW (available upon request). A pair of magnets are mounted 0.8 mm apart, at the center axis of a hollow shaft stepper motor

that is mounted on a stepper motor-actuated translation stage. The magnets are oriented with their magnetization axis vertically, in opposite directions. For proper orientation, ensure that the magnets attract each other yet do not allow rotation around the axis connecting the centers of the magnets. The microscope objective is mounted on a stepper motor-actuated translation stage to allow for coarse focusing. Stepper motors are controlled by a 6-axis two-phase stepper motor driver. A collimated LED is mounted in a kinematic pitch/yaw adapter above the hollow shaft motor, and provides a homogeneous illumination of the flow cell. We use an inverted microscope layout, including a tube lens and monochrome camera, to record the out-of-focus images of the Lorentz-Mie scattering pattern of the tethered bead [43]. The flow cell is mounted on a XYZ piezo stage, for accurate control of the focus position. The XYZ piezo stage is mounted onto an XY positioning stage, for manual coarse translation of the flow cell. The microscope setup is surrounded by a box to minimize acoustic and thermal fluctuations and to protect the setup from dust.

34. Before each experiment, the focus should be adjusted so that we always measure in the same plane.
35. An automatic bead finder algorithm is currently implemented, which cross-correlates the field-of-view with a computer-generated reference image which resembles the beads. Beads are selected by setting a threshold on the cross-correlation amplitude.
36. During a typical measurement, a series of frames are acquired and processed in real time. The magnet position and rotation are recorded in synchrony by the same software that controls the frame grabber. Data collection is buffered in two parallel processes:
    (a) Image acquisition.

        The camera captures full frames at a speed of 30 fps. The assigned ROIs and frame numbers are extracted and stored in a buffer.

    (b) Image analysis.

        In parallel, the buffer is read out and analyzed. Each ROI is cross-correlated with a set of computer-generated reference images, using a FFT-based correlation algorithm. The maximum of the cross-correlation represents the center of the bead and is determined with sub-pixel accuracy. The phase of the cross-correlation at this location is also extracted, converted to a *Z*-coordinate and written into a

binary file. From the frame number and camera-frequency a time column is computed and stored in the same file. After each measurement, the binary file is transformed into an ASCII-file that allows for versatile data processing. This file contains a time column, a column with the motor positions and four columns per bead: three spatial coordinates and the amplitude of the cross-correlation.

37. Any appropriate bead tracking algorithm will suffice for this step. Use, for instance, the cross-correlation method [44], a center-of-mass calculation [45], directly fitting a Gaussian curve to the intensity profile [46], or the quadrant-interpolation method [47].

38. We used a 6955 base pairs DNA tether (≈2.4 μm) to calibrate 1.0 μm and 2.8 μm diameter magnetic beads and recorded images with a frame rate of 200 Hz. Longer tethers provide a better signal-to-noise ratio, therefore, it is not uncommon that Lambda-DNA is used for calibration, which is 48,502 base pairs (≈16.5 μm) in length. If appropriate camera corrections are applied (i.e., motion blur and aliasing, *see* **Note 44**), force calibration can be performed with shorter tethers.

39. The required measurement time τ_m for a desired statistical accuracy ε is described in Eqs. 4 and 5 [25]

$$\tau_m \approx \frac{12\pi^2 \eta R L_c}{F\varepsilon^2} \quad \text{for } F > 1\,\text{pN}, \tag{4}$$

$$\tau_m \approx \frac{8\pi^2 \eta R L_p L_c}{k_B T \varepsilon^2} \quad \text{for } F < 1\,\text{pN}, \tag{5}$$

where η is the viscosity of water, R the radius of the bead, L_c the DNA contour length, F the calibrated force, and L_p the DNA persistence length. Typically, the statistical accuracy ε is set to 0.05. We measure each magnet position for 120 s, which is well within limits.

40. A plot of extension as a function of time may reveal drift as a steady increase or decrease of extension over time. When beginning and ending a trajectory at low force (or no rotations), at which the Z-position of the bead is constrained by the bottom slide of the flow cell, compare the first percentile of the bead height of these two segments to quantify the drift. Calculate the first percentile rather than the average, since the surface poses a one-sided boundary to the bead motion. The first percentile will give a good estimation of the surface position, which is required for drift characterization. Typically, we measure less than 1 nm/s drift of the extension, yet this can accumulate to signification deviations of the measured force-extension curves.

41. When the bead is in close proximity to the surface, nonspecific interactions between the bead and the surface can cause the bead or part of the tether to stick to the flow cell bottom, immobilizing the bead. Small forces are generally sufficient to unstick the bead and resume the measurement as normal. As opposed to intra molecular rupturing or folding events, (un-) sticking events can be identified in the data as an abrupt step in the coordinates in three dimensions, rather than only in the Z-direction.
42. A double bead doubles the force exerted on the tether, therefore, transitions are measured at half the force at which they are expected (for instance, DNA overstretching measured at half the force). The measured persistence length is double of that measured with a single bead. Furthermore, the data is generally very noisy, since tracking becomes problematic with two beads in the same ROI.
43. A double tether halves the force exerted on the tether. Therefore, transitions are measured at double the force as they are expected and add up (for instance, double the amount of stepwise unwrapping events when stretching chromatin, measured at double the force). The measured persistence length is half that of a single tether. A plot of the X- versus Y-coordinates of the bead during drift characterization is asymmetric.
44. The magnetic force pulls the tethered bead in the Z-direction and is counteracted by a restoring force from the tether. Brownian motion causes the bead to move away from its central position. The potential energy $E_{\mathrm{p}}(x)$ in the direction of the field (X-direction) around the equilibrium position is described by:

$$E_{\mathrm{p}}(x) = \frac{1}{2} k_x \delta x^2, \tag{6}$$

where δx^2 is the variance of the bead fluctuations. The effective trap stiffness k_x follows

$$k_x = \frac{F}{z}, \tag{7}$$

where F is the stretching force and z the bead height. Equipartition theorem states that the energy per degree of freedom equals $\frac{1}{2} k_{\mathrm{B}} T$. The force therefore follows

$$F = \frac{k_{\mathrm{B}} TZ}{< \delta x^2 >}. \tag{8}$$

For an accurate experimental determination of the force however, it is necessary to look into the frequency dependence of

the thermal fluctuations of x. Start by computing the PSD in the direction of the field at constant force. Since the PSD follows a Lorentzian curve

$$\mathrm{PSD}_{\mathrm{thermal}}(f, f_c) = \frac{k_B T}{\gamma \pi^2 (f_c^2 + f^2)}, \tag{9}$$

it is fully defined by the lateral friction coefficient γ and the cut-off frequency f_c

$$f_c = k_x / 2\pi\gamma. \tag{10}$$

The friction coefficient depends on the bead radius R and the viscosity η (which increases with closer proximity to the surface, *see* **Note 60**):

$$\gamma = 6\pi\eta R, \tag{11}$$

leading together with Eq. 7 to

$$F = 12\,\pi^2 \eta\, R\, z\, f_c. \tag{12}$$

Thus fitting the cut-off frequency from the PSD yields the force for a given magnet position. However, the finite integration time of the camera averages out some of the bead's movement, an effect known as blurring. To account for blurring, a correction term C_{blur} must be applied [24–26, 48, 49]:

$$C_{\mathrm{blur}}(f) = \left(\frac{\sin(W\pi f)}{W\pi f} \right)^2, \tag{13}$$

where W is the frame integration time. The finite sampling frequency of the camera causes another artifact, known as aliasing. Aliasing can be described by folding back the parts of the spectrum that exceed the sampling frequency f_s

$$\mathrm{PSD}_{\mathrm{alias}}(f, f_c) = \sum\nolimits_n \mathrm{PSD}_{\mathrm{thermal}}(f + n f_s, f_c)\, C_{\mathrm{blur}}(f + n f_s). \tag{14}$$

It is usually sufficient to include only one aliasing term, i.e., $n = -1..1$ [24]. Finally, the PSD is offset by a tracking error of variance σ^2 that is independent of frequency [49]:

$$\mathrm{PSD}_{\mathrm{tracking}} = \frac{\sigma^2}{f_s}, \tag{15}$$

in which f_s is the sampling frequency. Overall, fitting

$$\mathrm{PSD}_{\mathrm{corrected}}(f, f_c) = \frac{\sigma^2}{f_s} + \sum_{n=-1}^{1} \frac{k_B T}{\gamma \pi^2 (f_c^2 + (f + n f_s)^2)} \left(\frac{\sin\left(W\pi(f + n f_s)\right)}{W\pi(f + n f_s)} \right)^2 \tag{16}$$

to the experimental spectrum of the lateral fluctuations will yield the cut-off frequency which is used to calculate the force.

45. The loading rate increases exponentially with magnet speed. For forces below 5 pN, we typically have a loading rate that ranges between 0 and 0.75 pN/s. High loading rates shift the rate of non-equilibrium events such as breaking of bonds between antibodies [50].
46. Several variations of magnet trajectories can be used:
    (a) Repeat loops to gain insight into the reversibility of DNA folding.
    (b) To approach a linear loading rate, break up the trajectory in several segments with decreasing magnet speeds.
    (c) Start with a short force ramp up to 0.5 pN to set free loosely stuck beads.
47. Torsionally constrained substrates build up torque while stretched, and consequently respond differently to force. These substrates can be pre-twisted before an experiment to study the role of supercoiling. For instance, DNA compacted by histone-like proteins has an intrinsically twisted structure.
48. Supercoiling reduces the bead height, which increases the probability of beads or tethers to stick to the surface. Minimizing the dwell time at the maximum number of rotations reduces sticking.
49. Plot force as a function of extension. Plotting it in this way is customary for optical tweezers and allows for easy comparison between methods. Because MT form a force clamp rather than a position clamp, this way of plotting may be counterintuitive.
50. For tethers that have a length that is in range with the size of the bead, off-center attachment of the DNA to the bead may lead to a large error in the measured extension. Superparamagnetic beads have a small, but finite permanent magnetic moment, giving them a preferred orientation when exposed to a magnetic field [51]. Since tethering happens in the absence of the magnetic field, application of force will turn the magnetic moment in the direction of the magnetic field, typically horizontally, while maximizing the height of the bead. When the DNA is attached at a location exactly at the circumference of the bead that has equal distance between the magnetic poles, it will rotate the bead, such that the attachment

point will be at the bottom of the bead. Any location remote from this circle will still rotate the bead around the horizontal axis along the poles, but cannot spin the bead along the horizontal axis perpendicular to the poles. Consequently, there will be an offset between the height of the bead and the extension of the tether, and the measured extension is underestimated. This artifact is specific for MT and can be corrected by shifting the data in the *Z*-direction, or by including this offset as a fitting parameter.

51. Specific modes of protein binding yield distinctive patterns:
    (a) Coating the DNA tether yields a stiffening of the DNA, which can be parameterized by an increased persistence length in the entropic regime at low force and/or an increased stretch modulus at high force. These two effects can only be differentiated when a large force range is probed. Examples are HU (at μM concentration) [6] and H-NS (at nM concentration) [9, 52, 53].
    (b) Bending induces a kink in the DNA trajectory, which is apparent as a reduced persistence length [54]. The magnitude of this reduction scales with the kink angle and the number of kinks, which reflect the structure and number of proteins bound to the tether. Examples of chromatin proteins that introduce a kink are HU (at nM concentration) [6, 55] and IHF (at nM concentration) [56].
    (c) Wrapping of DNA around proteins results in both a reduction of the persistence length, similar to bending and depending on the amount of DNA that is wrapped, and a reduction of the contour length. When exactly one wrap is formed, the persistence length is similar to that of bare DNA. The archetype of wrapping proteins is the eukaryotic histone octamer that wraps 1.7 turns of DNA [57–59]. Archaeal histones wrap smaller DNA lengths [60, 61].
    (d) Bridging involves protein-induced formation of contacts between distant part of the DNA tether and results in a large reduction of the contour length. The DNA that is captured in the loop does not contribute to the extension of the tether. An example of a bridging protein is H-NS, whose mode of binding depends on the concentration of monovalent and divalent ions [9, 52, 53].
    (e) Stabilization of single-stranded DNA. Melting parts of the dsDNA tether by force and or torque yields sections of single-strand DNA. This leads to a 1.6 times extension of the contour length, as well as a reduction of the persistence length down to a few nanometers in the absence of

protein covering the single-strand DNA [35]. Single-strand binding proteins like SSB prevent annealing of force induced, melted DNA, yielding force extension curves that display a large hysteresis [62].

52. Various force regimes reveal different chromatin conformations that can all be captured in a force-dependent linear combination of bare DNA and four different nucleosome structures [11]. Bacterial chromatin typically consists of proteins that bend, rather than wrap DNA, have a smaller footprint and are usually interspersed with different types proteins. Furthermore, they may have more varying sequence preferences. Nevertheless, a similar analysis, customized for different characteristics of individual proteins, can be used to retrieve a detailed understanding of chromatin folding based on force spectroscopy. When the chromatin conformation is not stable due to changes of protein concentration, force or salt, this may result in a change in extension. From such changes it is possible to extract binding and/or dissociation rates.
53. Calculate the linking number density, using $\sigma = \Delta Lk/Lk_0$ where Lk_0 is the linking number of DNA. Lk_0 equals the contour length in base pairs divided by the helical pitch of DNA, which is 10.4 base pairs. The change in linking number, ΔLk, equals the number of rotations of the magnets.
54. A double bead affects rotational spectroscopy measurements similar to force spectroscopy measurements. The tether is more likely to break since the force that is kept constant during the measurement is double of that expected.
55. Double tethers inhibit rotational spectroscopy measurements similar to force spectroscopy measurements. Double tethers can easily be identified when twisting the tether. During the first turn, the bead is pulled down by a large step, as the two molecules get braided. Subsequent twist lowers the bead for both negative and positive rotations, as opposed to single tethers, for which the extension is independent of negative twist at forces exceeding 1 pN.
56. The DNA tether can become nicked during sample preparation. A single nick is sufficient to release torque and consequently the tether cannot be used for rotational spectroscopy. These tethers can be discarded during an initial rotation experiment, or can serve as control for torsionally unconstraint tethers.
57. Plot all X-coordinates versus all Y-coordinates during rotation of the bead. The shape of this plot should be a circle. The radius of the circle equals the attachment offset from the center of the bead.

58. Since mechanical properties of twisted DNA are much more complex than those of torsionally free DNA, fitting the curves is challenging. The three-state model developed by Meng et al. [31] describes the coexistence of twisted, melted, and plectonemic conformational states. The data and the model feature symmetric buckling at low forces ($f < 0.6$ pN), where both positive and negative twist is absorbed by plectonemes. At higher forces ($f > 0.6$ pN) the extension-twist curve becomes asymmetric, and negative twist is absorbed by local melting of the DNA.
59. Quantitative analysis of the torsional stress in chromatin filaments is more involved. Bending proteins (such as HU of IHF) and wrapping proteins (such as HMf or eukaryotic histones) induce writhe in a torsionally constrained tether, while twist remains constant. The torsional stiffness of such fibers has not been measured, though quantification is highly desirable to test proposed models of twist-induced (de-)compaction of chromatin fibers and their role in transcription regulation. Eukaryotic tetrasomes have been shown to have a complex dynamic chirality [28, 29].
60. Beads experience an increasing viscosity as the distance to the surface approaches the bead diameter. [63]. The height dependence of the lateral friction coefficient γ is approximated by Faxén's law, as described in Eq. 17.

$$\gamma = \frac{\gamma_0}{1 - \frac{9R}{16h} + \frac{R^3}{8h^3} - \frac{45R^4}{256h^4} - \frac{R^5}{16h^3}} \tag{17}$$

where $\gamma_0 = 6\pi\eta R$ is the bulk friction coefficient, η the viscosity, R the radius of the bead, and h the distance of the bead center to the surface.

References

1. Luijsterburg MS, White MF, Van Driel R, Dame RT (2008) The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. Crit Rev Biochem Mol Biol 43:393–418
2. Bustamante C, Bryant Z, Smith SB (2003) Ten years of tension: single-molecule DNA mechanics. Nature 421:423–427
3. Smith S, Finzi L, Bustamante C (1992) Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. Science 258:1122–1126
4. Strick TR, Allemand JF, Bensimon D et al (1996) The elasticity of a single supercoiled DNA molecule. Science 271:1835–1837
5. Gosse C, Croquette V (2002) Magnetic tweezers: micromanipulation and force measurement at the molecular level. Biophys J 82:3314–3329
6. van Noort J, Verbrugge S, Goosen N et al (2004) Dual architectural roles of HU: formation of flexible hinges and rigid filaments. Proc Natl Acad Sci U S A 101:6969–6974
7. van Loenhout MTJ, van der Heijden T, Kanaar R et al (2009) Dynamics of RecA filaments on single-stranded DNA. Nucleic Acids Res 37:4089–4099
8. Lionnet T, Allemand JF, Revyakin A et al (2012) Single-molecule studies using magnetic traps. Cold Spring Harb Protoc 7:34–49

9. Lim CJ, Kenney LJ, Yan J (2014) Single-molecule studies on the mechanical interplay between DNA supercoiling and H-NS DNA architectural properties. Nucleic Acids Res 42:8369–8378
10. van der Valk RA, Vreede J, Crémazy F, Dame RT (2014) Genomic looping: a key principle of chromatin organization. J Mol Microbiol Biotechnol 24:344–359
11. Meng H, Andresen K, Van Noort J (2015) Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibers. Nucleic Acids Res 43:3578–3590
12. Yan J, Skoko D, Marko JF (2004) Near-field-magnetic-tweezer manipulation of single DNA molecules. Phys Rev E Stat Nonlin Soft Matter Phys 70(1 Pt 1):011905
13. Berghuis BA, Köber M, van Laar T, Dekker NH (2016) High-throughput, high-force probing of DNA-protein interactions with magnetic tweezers. Methods 105:90–98
14. Ribeck N, Saleh OA (2008) Multiplexed single-molecule measurements with magnetic tweezers. Rev Sci Instrum 79(9):094301
15. De Vlaminck I, Henighan T, Van Loenhout MTJ et al (2011) Highly parallel magnetic tweezers by targeted DNA tethering. Nano Lett 11:5489–5493
16. Lipfert J, Wiggin M, Kerssemakers JWJ et al (2011) Freely orbiting magnetic tweezers to directly monitor changes in the twist of nucleic acids. Nat Commun 2:439
17. Lipfert J, Kerssemakers JWJ, Jager T, Dekker NH (2010) Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. Nat Methods 7:977–980
18. Fisher JK, Cribb J, Desai KV et al (2006) Thin-foil magnetic force system for high-numerical-aperture microscopy. Rev Sci Instrum 77. https://doi.org/10.1063/1.2166509
19. Oliver PM, Park JS, Vezenov D (2011) Quantitative high-resolution sensing of DNA hybridization using magnetic tweezers with evanescent illumination. Nanoscale 3:581–591
20. Graham JS, Johnson RC, Marko JF (2011) Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA. Nucleic Acids Res 39:2249–2259
21. Graham JS, Johnson RC, Marko JF (2011) Counting proteins bound to a single DNA molecule. Biochem Biophys Res Commun 415:131–134
22. Long X, Parks JW, Bagshaw CR, Stone MD (2013) Mechanical unfolding of human telomere G-quadruplex DNA probed by integrated fluorescence and magnetic tweezers spectroscopy. Nucleic Acids Res 41:2746–2755
23. Vilfan ID, Lipfert J, Koster DA et al (2009) Magnetic tweezers for single-molecule experiments. In: Hinterdorfer P, Oijen A (eds) Single-molecule biophysics. Springer, New York, pp 371–395
24. te Velthuis AJW, Kerssemakers JWJ, Lipfert J, Dekker NH (2010) Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. Biophys J 99:1292–1302
25. Yu Z, Dulin D, Cnossen J et al (2014) A force calibration standard for magnetic tweezers. Rev Sci Instrum 85(12):123114
26. Daldrop P, Brutzer H, Huhle A et al (2015) Extending the range for force calibration in magnetic tweezers. Biophys J 108:2550–2561
27. Gupta P, Zlatanova J, Tomschik M (2009) Nucleosome assembly depends on the torsion in the DNA molecule: a magnetic tweezers study. Biophys J 97:3150–3157
28. Vlijm R, Lee M, Lipfert J et al (2015) Nucleosome assembly dynamics involve spontaneous fluctuations in the handedness of tetrasomes. Cell Rep 10:216–225
29. Vlijm R, Lee M, Ordu O et al (2015) Comparing the assembly and handedness dynamics of (H3.3-H4)2 tetrasomes to canonical tetrasomes. PLoS One 10(10):e0141267
30. Kruithof M, Chien F, de Jager M, van Noort J (2008) Subpiconewton dynamic force spectroscopy using magnetic tweezers. Biophys J 94:2343–2348
31. Meng H, Bosman J, Van Der Heijden T, Van Noort J (2014) Coexistence of twisted, plectonemic, and melted DNA in small topological domains. Biophys J 106:1174–1181
32. Marko JF, Siggia ED (1995) Stretching DNA. Macromolecules 28:8759–8770
33. Bustamante C, Smith SB, Liphardt J, Smith D (2000) Single-molecule studies of DNA mechanics. Curr Opin Struct Biol 10:279–285
34. Marko JF (2007) Torque and dynamics of linking number relaxation in stretched supercoiled DNA. Phys Rev E Stat Nonlin Soft Matter Phys 76(2 Pt 1):021926
35. Smith SB, Cui Y, Bustamante C (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. Science 271:795–798
36. Lowary P, Widom J (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol 276:19–42

37. Bancaud A, Wagner G, Conde e Silva N et al (2007) Nucleosome chiral transition under positive torsional stress in single chromatin fibers. Mol Cell 27:135–147
38. Chen CW, Thomas CA (1980) Recovery of DNA segments from agarose gels. Anal Biochem 101:339–341
39. Marko MA, Chipperfield R, Birnboim HC (1982) A procedure for the large-scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder. Anal Biochem 121:382–387
40. Boom R, Sol C, Salimans MMM et al (1990) Rapid and simple method for purification of nucleic acids. J Clin Microbiol 28:495–503
41. Salvi G, De Los Rios P, Vendruscolo M (2005) Effective interactions between chaotropic agents and proteins. Proteins Struct Funct Genet 61:492–499
42. LePecq JB, Paoletti C (1967) A fluorescent complex between ethidium bromide and nucleic acids. Physical-chemical characterization. J Mol Biol 27:87–106
43. Lee S-H, Roichman Y, Yi G-R et al (2007) Characterizing and tracking single colloidal particles with video holographic microscopy. Opt Express 15:18275
44. Gelles J, Schnapp BJ, Sheetz MP (1988) Tracking kinesin-driven movements with nanometre-scale precision. Nature 331:450–453
45. Lee GM, Ishihara A, Jacobson K (1991) Direct observation of brownian motion of lipids in a membrane. Proc Natl Acad Sci U S A 88:6274–6278
46. Anderson CM, Georgiou GN, Morrison IE et al (1992) Tracking of cell surface receptors by fluorescence digital imaging microscopy using a charge-coupled device camera. Low-density lipoprotein and influenza virus receptor mobility at 4 degrees C. J Cell Sci 101(Pt 2):415–425
47. van Loenhout MTJ, Kerssemakers JWJ, De Vlaminck I, Dekker C (2012) Non-bias-limited tracking of spherical particles, enabling nanometer resolution at low magnification. Biophys J 102:2362–2371
48. Wong WP, Halvorsen K (2006) The effect of integration time on fluctuation measurements: calibrating an optical trap in the presence of motion blur. Opt Express 14:12517–12531
49. van der Horst A, Forde NR (2010) Power spectral analysis for optical trap stiffness calibration from high-speed camera position detection with limited bandwidth. Opt Express 18:7670–7677
50. Sitters G, Kamsma D, Thalhammer G et al (2014) Acoustic force spectroscopy. Nat Methods 12:47–50
51. Klaue D, Seidel R (2009) Torsional stiffness of single superparamagnetic microspheres in an external magnetic field. Phys Rev Lett 102:1–4
52. Liu Y, Chen H, Kenney LJ, Yan J (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. Genes Dev 24:339–344
53. van der Valk RA, Vreede J, Qin L et al (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA bridging activity. elife 6:e27369
54. Kulić IM, Mohrbach H, Thaokar R, Schiessel H (2007) Equation of state of looped DNA. Phys Rev E Stat Nonlin Soft Matter Phys 75 (1 Pt 1):011913
55. Swinger KK, Rice PA (2004) IHF and HU: flexible architects of bent DNA. Curr Opin Struct Biol 14:28–35
56. Lin J, Chen H, Dröge P, Yan J (2012) Physical organization of DNA by multiple non-specific DNA-binding modes of integration host factor (IHF). PLoS One 7:e49885
57. Luger K, Mäder AW, Richmond RK et al (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389:251–260
58. Schalch T, Duda S, Sargent DF, Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature 436:138–141
59. Song F, Chen P, Sun D et al (2014) Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science 344:376–380
60. Henneman B, Dame TR (2015) Archaeal histones: dynamic and versatile genome architects. AIMS Microbiol 1:72–81
61. Mattiroli F, Bhattacharyya S, Dyer PN et al (2017) Structure of histone-based chromatin in Archaea. Science 357:609–612
62. Hatch K, Danilowicz C, Coljee V, Prentiss M (2008) Measurement of the salt-dependent stabilization of partially open DNA by Escherichia coli SSB protein. Nucleic Acids Res 36:294–299
63. Schäffet E, Nørrelykke SF, Howard J (2007) Surface forces and drag coefficients of microspheres near a plane surface measured with optical tweezers. Langmuir 23:3654–3665

Chapter 18

In Vitro Transcription Assay to Quantify Effects of H-NS Filaments on RNA Chain Elongation by RNA Polymerase

Beth A. Boudreau, Matthew V. Kotlajich, and Robert Landick

Abstract

While structuring of the bacterial nucleoid by nucleoid-associated proteins (NAPs) is critical for proper chromosomal organization and compaction, DNA-dependent RNA polymerase (RNAP) must frequently interact with and overcome the barriers these NAPs impose upon transcription. One particular NAP in *Escherichia coli* that influences transcription is the histone-like nucleoid structuring protein, H-NS, that binds to DNA and forms nucleoprotein filaments. To specifically investigate the effect that H-NS filaments have on RNAP elongation, we developed an in vitro transcription assay to assess transcript elongation by RNAP when transcribing DNA bound by an H-NS filament. In this method, initiation and elongation by RNAP are uncoupled by initiating transcription in the presence of three rNTPs to halt elongation just downstream of the promoter. Before elongation is restarted, an H-NS filament is formed so that elongation occurs on an H-NS nucleoprotein filament template. We also describe visualization and analysis of the transcription products from the nucleoprotein template which provides insight into how H-NS and RNAP interact. This method is a starting point to determine effects of NAPs on RNAP elongation in a variety of conditions.

Key words RNA polymerase, Transcription elongation, Pausing, Roadblocks, Nucleoid-associated proteins

1 Introduction

In bacteria, a dozen identified nucleoid-associated proteins (NAPs) bind throughout the chromosome to structure the nucleoid. The structuring of the nucleoid by NAPs is compatible with necessary DNA-dependent processes, such as DNA replication and transcription. How the replication or transcription machinery deals with potential interferences from NAPs is not well characterized. Some work has been done to characterize the interaction between NAPs and RNAP (reviewed in [1]), but many questions remain. Most NAPs are thought to regulate transcription initiation by binding to the promoter or nearby regions, where they either inhibit or assist RNAP binding at or near promoters [2]. However, NAPs could also affect elongation and termination by RNAP, and such effects

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_18,

could play key roles in regulating gene expression [3, 4]. One NAP in *E. coli*, the histone-like structuring protein (H-NS), has been found to contribute to gene silencing by decreasing initiation events at promoters, slowing RNAP elongation, and stimulating Rho-dependent termination [3–6]. To silence genes, H-NS forms nucleoprotein filaments by recognizing a high-affinity binding site and then oligomerizing along the DNA [7]. H-NS filaments adopt one of two distinct conformations depending upon a number of conditions: either a linear conformation, in which H-NS interacts with one segment of DNA, or a bridged conformation, in which H-NS interacts with two distal segments of DNA forming a DNA-protein-DNA bridge [8–10]. To characterize the interaction between H-NS and an elongating RNAP [4, 11, 12], we developed an in vitro transcription assay to measure the ability of RNAP to elongate through either a bridged or linear H-NS filament [3, 13]. This method can be easily adapted to study interactions between an elongating RNAP and any other NAP.

The protocol described here allows quantification of RNAP transcript elongation using a linear or bridged H-NS filament as a DNA template. The template contains a DNA fragment from the H-NS-silenced *bgl* operon of *E. coli* [14, 15]. The *E. coli bgl* operon encodes gene products responsible for utilization of β-glucosides, but is silenced in *E. coli* K-12 by an H-NS filament that nucleates upstream and downstream of the operon's relatively weak, cAMP-CAP-dependent promoter. The *bgl* operon also encodes an antisense promoter, from which transcript elongation is inhibited by H-NS in concert with the Rho termination factor. To enable formation of H-NS filaments, we created a chimeric template that replaced the *bgl* antisense promoter with the strong, constitutive λP_R promoter. Just downstream of the promoter, we placed a sequence that allows initiation with three rNTPs which halts transcription 26 nucleotides (nt) downstream of the start site to form stable, halted A26 elongation complexes (ECs) that contain a 26-nt RNA with an A at the 3′ end. H-NS filaments are then assembled on the downstream DNA before transcript elongation by RNAP is restarted by the addition of all four rNTPs. Samples are collected at time points throughout the course of transcription, and the RNA products are resolved using denaturing urea-PAGE. By preforming ECs before the addition of H-NS, and preventing subsequent re-initiation by the addition of the inhibitor rifampicin, transcription initiation is uncoupled from elongation to enable study of effects of H-NS filaments on the elongating RNAP.

This chapter describes each step of these procedures, including (1) DNA template preparation, (2) setup and execution of the in vitro transcription assay, (3) the electrophoretic mobility shift assay (EMSA), (4) urea-PAGE analysis of RNA products, and (5) generation of densitometry profiles to analyze RNA products (Fig. 1; *see* **Notes 1** and **2**).

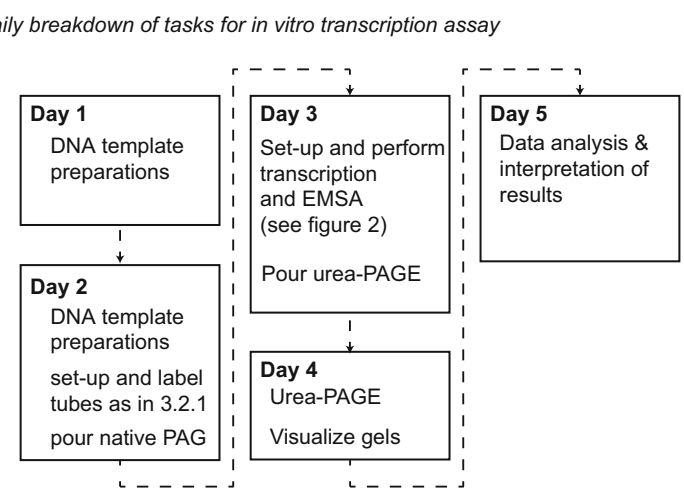


Fig. 1 Overall experimental design and organization broken down by day. The in vitro transcription and accompanying protocols can be completed over 5 days by dividing tasks as outlined here and discussed in **Note 1**. DNA template can be made in advance and used at later dates for multiple experiments. Gels should be poured no earlier than the day before they will be run to prevent drying out

2 Materials

Ultrapure water (MQH_2O) should be prepared by filtering deionized water to 18 MΩ cm at 25 °C and used where indicated.

2.1 DNA Preparation

1. pMK110, which is a plasmid containing the λP_R promoter, a C-less cassette (26-nt without a C) to halt RNAP downstream of the promoter, and a region of the *bgl* operon known to be bound by H-NS (*see* **Note 3**).
2. Primers 3071 (fwd) and 645 (rev) to amplify 1.5 kb template from pMK110.
   (a) Primer 3071: 5′-CGTTAAATCTATCACCGCAAGGG
   (b) Primer 645: 5′-CAGTTCCCTACTCTCGCATG
3. Taq DNA polymerase.
4. Taq reaction buffer (*see* **Note 4**).
5. 2.5 mM dNTP mix stocks (Fisher).
6. *Bgl* template PCR: Make desired volume of PCR mixture with final concentrations of 250 μM dNTPs, 250 nM forward primer 3071, 250 nM reverse primer 645, 0.4 ng/μL pMK110 (~2 μg in a 5-mL PCR), and 125 Units OneTaq DNA polymerase (*see* **Note 5**).
7. 15 mL conical tubes.
8. 100% ethanol.

9. 3 M NaOAc: Dissolve 40.8 g sodium acetate trihydrate in ~60 mL of water with stirring. Adjust pH to 5.2 with glacial acetic acid. Add MQH_2O to 100 mL. Filter-sterilize and store at room temperature.
10. 70% ethanol.
11. TE: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA.
12. 10× TBE: 890 mM Tris-base, 890 mM boric acid, 25 mM EDTA, pH 8.3. Add 180 g Tris-base, 55 g Boric Acid, and 9.3 g Na_2EDTA to 890 mL MQH_2O. Dissolve all components and adjust volume with MQH_2O to 1 L (*see* **Note 6**).
13. 0.5 M EDTA, pH 8.0: Add 93.05 g Na_2EDTA to ~400 mL MQH_2O. Adjust pH to 8.0 with NaOH pellets. Adjust volume to 500 mL. Autoclave (*see* **Note 7**).
14. 1× TBE with 4 mM EDTA: 200 mL of 10× TBE, 6 mL of 0.5 M EDTA, and MQH_2O to 2 L (*see* **Note 8**).
15. 1× TBE with 3.5 mM EDTA: 100 mL of 10× TBE, 2 mL 0.5 M EDTA, and MQH_2O to 1 L.
16. SeaKem LE Agarose (Lonza).
17. 10 mg ethidium bromide/mL MQH_2O.
18. A horizontal gel electrophoresis system with a gel container that is 9.5 cm wide × 10.8 cm long × 2.5 cm deep (*see* **Note 9**).
19. 80% glycerol.
20. 10× loading dye: Combine 1 mL 80% glycerol, 0.01 g of bromophenol blue (BpB), and 0.01 g of xylene cyanol (XC). Vortex to mix.
21. 3500 MWCO dialysis tubing (Spectra/Por #3, *see* **Note 10**).
22. TE-equilibrated phenol, pH 7.9 (*see* **Note 11**).
23. 24:1 chloroform:isoamyl alcohol (IAA).
24. DEPC-MQH_2O (nuclease-free water): stir 1 L MQH_2O with 1 mL diethyl pyrocarbonate (DEPC) at room temperature overnight. Autoclave to inactivate the DEPC before use.
25. 15 mM HEPES-KOH pH 8.0: Dilute 1 M HEPES-KOH, pH 8.0 (*see* Subheading 2.2.1) to 15 mM in DEPC-H_2O. Store wrapped in foil at 4 °C.
26. Nuclease-free low binding 1.5-mL microcentrifuge tubes.

2.2 In Vitro Transcription and RNA Purification

All buffers should be made with DEPC-treated MQH_2O and filtered with 0.2 μm syringe filters. Use of nuclease-free tips is recommended to avoid RNase contamination. In vitro transcription conditions (filament formation and elongation at 20 °C with 30 μM NTPs) are optimized for *E. coli* RNAP elongation through *E. coli* H-NS filaments, but can be modified for other experiments (*see* **Notes 2, 3, 13, 23,** and **53**).

2.2.1 In Vitro Transcription Reaction

1. RNase eliminator (e.g., Eliminase) (*see* **Note 12**).
2. *E. coli* RNA polymerase holoenzyme (*see* **Note 13**).
3. 100 mM individual rNTP (rATP, rUTP, rGTP, and rCTP) stocks in DEPC-treated MQH_2O stored at −80 °C in 40 μL aliquots until use (Promega).
4. 2 mM individual rNTP stocks (rATP, rUTP, rGTP, and rCTP): dilute 100 mM rNTP stocks in DEPC-treated MQH_2O. Store at −80 °C (*see* **Note 14**).
5. 20× initiating NTP mix: Make 10 μL by adding 1 μL of 2 mM rATP and rUTP and 0.25 μL rGTP for 200 μM ATP and UTP and 50 μM GTP as the 20× concentrations in DEPC-treated MQH_2O. CTP is excluded to allow formation of halted complexes at A26 and the GTP concentration is lower because [α-^{32}P]GTP will be added (*see* **Note 15**).
6. 10× chase NTP mix: add each rNTP (A, U, G, and C) to 300 μM (this is the 10× concentration) in DEPC-treated MQH_2O (*see* **Notes 16** and **17**).
7. [α-^{32}P]GTP (3000 Ci/mmol, 10 mCi/mL) (Perkin-Elmer) (*see* **Note 18**).
8. 2 mM ApU dinucleotide (Trilink or Ribomed) (*see* **Note 19**).
9. 25 mM DTT (*see* **Note 20**).
10. 1 M HEPES-KOH, pH 8.0: Add 59.57 g of HEPES to ~150 mL of DEPC-treated MQH_2O and stir to dissolve. Adjust pH to 8.0 with KOH pellets. Adjust volume to 250 mL with DEPC-treated MQH_2O. Filter and store solution in a dark glass bottle at 4 °C (*see* **Note 21**).
11. Acetylated BSA: Make 100 mL of BSA (5 mg/mL) in a sterile solution containing 0.1 M Na_2HPO_4 and 2 M sodium acetate at 0 °C. Dropwise, add 0.6 mL of acetic anhydride to the BSA solution. Incubate for 45 min and then dialyze extensively against cold and sterile MQH_2O. Aliquot (1 mL) and store at −20 °C [16].
12. 3× EMSA buffer: 120 mM HEPES-KOH, pH 8.0, 300 mM potassium glutamate, 24 mM magnesium aspartate, 0.066% Nonidet P-40, 0.3 mg acetylated BSA/mL, 15% glycerol (*see* **Notes 22** and **23**).
13. 1 mg rifampicin/mL: Add 4 μL of 50 mg rifampicin/mL to 196 μL water. Vortex to mix (*see* **Note 24**).
14. Recombinant RNasin (Promega).
15. Bacterial NAP of interest. Here we used purified H-NS as described in [3]. Working stocks of 20 μM were made by diluting H-NS stocks (~150 μM) in 3× EMSA Buffer (*see* **Note 25**).

16. 37 °C incubator.
17. 20 °C water bath.
18. 10× transcription elongation complex (ECs) mix: 1× EMSA buffer, 1× initiating NTP mix, 1 mM DTT, 150 μM ApU, and 100 nM purified DNA template (Subheading 3.1) leaving room to add [α-^{32}P]GTP and 150 μM RNAP holoenzyme when appropriate (*see* Table 2 and **Note 26**).
19. Elongation master mix: 0.55 U RNasin/μL, 0.1 mg rifampicin/mL, 0.5 mM DTT in DEPC-treated MQH_2O (*see* Table 4 and **Note 27**).
20. Elongation reaction mix: 1× EMSA buffer, desired amount of H-NS diluted in 3× EMSA buffer, elongation master mix, and room to add 10× ECs and 10× chase NTPs to 1× final concentrations (Table 3 and *see* **Notes 27** and **28**).

2.2.2 RNA Purification

1. 20 mg glycogen/mL stocks.
2. 0.5 M EDTA, pH 8.0.
3. EDTA/glycogen stop solution: 15 mM EDTA and 4 μL glycogen/100 μL final in DEPC-treated MQH_2O. Then aliquot 93 μL of the stop mix to 1.5-mL microcentrifuge tubes for each timed sample collection. Store at 4 °C until needed (*see* **Note 29**).
4. TE-equilibrated phenol, pH 7.9.
5. 100% ethanol and 70% ethanol solutions.
6. Ethanol precipitation aliquots: Aliquot 280 μL of 100% ethanol in 1.5 mL microcentrifuge tubes. Make enough tubes for all conditions and time points to be collected. Store at −20 °C until needed.
7. Nuclease-free, low-binding 0.6 and 1.5 mL microcentrifuge tubes.
8. 50 mL conical tube.
9. SpeedVac concentrator.
10. 2× urea stop dye: Combine 1 mL 10× TBE, 4.98 g urea, 0.5 mL 0.5 M EDTA, 2 mg XC, 2 mg BpB in 6 mL DEPC-treated MQH_2O. Heat at 65 °C for ~5 min to dissolve urea. Adjust water to 10 mL and store at room temperature.
11. Formamide stop dye: 95% formamide, 15 mM EDTA pH 8, 0.05% BpB, 0.05% XC. Combine 48 mL 99% formamide, 2 mL 0.5 M EDTA pH 8.0, 2 mg BpB and 2 mg XC. Mix and store 1.5 mL aliquots at −80 °C (*see* **Note 30**).

2.3 Electrophoretic Mobility Shift Assay

2.3.1 Gel Setup

1. Cold water circulating (4 °C) vertical electrophoresis system with appropriate gel plates and spacers (*see* **Note 31**).
2. Dichlorodimethylsilane.
3. 10% dichlorodimethysilane solution in chloroform: 1 mL dichlorodimethylsilane added to 9 mL of chloroform. Store in a fume hood in a dark bottle.
4. 0.5× TBE: Add 50 mL of 10× TBE to 950 mL MQH_2O. Mix to combine. Chill to 4 °C before use.
5. 1% agarose solution in 0.5× TBE: Add 0.5 g SeaKem LE agarose to 50 mL 0.5× TBE. Microwave to dissolve agarose so no particles remain in solution.
6. 100% glycerol.
7. 30% acrylamide:bisacrylamide 29:1 (*see* **Note 32**).
8. 3% native poly-acrylamide gel: Mix 7.5 mL 30% acrylamide (acrylamide:bisacrylamide 29:1), 3.75 mL 10× TBE, 2 mL 100% glycerol, MQH_2O to 75 mL. For 20 cm × 20 cm plates, a 75 mL solution will be enough to pour a gel. Apply mixture to a 0.2 μm filter (e.g., a 125 mL filter unit) to remove any particulate matter (*see* **Note 33**) and then apply to filter. Swirl to mix.
9. 10% Ammonium persulfate (APS) solution in water.
10. *N*,*N*,*N*′,*N*′-tetramethylethane-1,2-diamine (TEMED).
11. 10× loading dye (described above).
12. Whatman paper 3 mm Chr.
13. Gel-drying vacuum system (such the Welch Thomas Gelmaster Gel dryer vacuum system 142601).

2.3.2 Filament Formation on Non-EC DNA Controls

1. T4 polynucleotide Kinase (PNK) (NEB).
2. 1× PNK Buffer: 70 mM Tris–HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6 at 25 °C (NEB).
3. [γ-^{32}P]ATP (6000 Ci/mmol, 10 mCi/mL) (Perkin Elmer).
4. 10 μM rATP.
5. Purified PCR template (*see* Subheading 3.1).
6. 5′-end-labeled DNA: Kinase reaction with [γ-^{32}P]ATP using T4 PNK in 1× PNK buffer, 0.67 μM ATP, 20 μCi γ-^{32}P-ATP, 100 nM purified DNA template, and 5 units T4 PNK. Use DEPC-treated MQH_2O to bring reaction volume to 12.5 μL (*see* **Note 34**).
7. EMSA buffer, protein, rifampicin, nucleotides, DTT, and RNasin used for in vitro transcription (*see* Subheading 2.2).
8. DNA control master mix: 0.1× initiating NTP mix, 0.5 mM DTT, 15 μM ApU, 0.1 mg rifampicin/mL, 0.55 U RNasin,

and 10 nM labeled 5′-end-labeled DNA template. Add water to a total volume of 5 μL master mix per condition being tested. DNA control master mix conditions match elongation mix conditions to compare filament formation.

9. H-NS or NAP of interest.

2.4 Urea-PAGE

1. Two sets of glass plates, spacers, and comb to fit vertical electrophoresis system and hold an appropriate number of samples (*see* **Note 35**).
2. Vertical electrophoresis system with aluminum plates and flat back thermometer or adhesive thermometer strip on aluminum plates (*see* **Note 36**).
3. 2% dichlorodimethylsilane: combine 2 mL dichlorodimethylsilane and 98 mL chloroform.
4. 10× TBE.
5. 40% acrylamide:bisacrylamide 19:1 (*see* **Note 37**).
6. TMD-8 hydrogen and hydroxide form (Sigma-Aldrich, CAS 100915-97-7).
7. 6% urea-PAG: Make 100 mL solution with 0.5× TBE and 7 M urea. Combine 42.03 g solid urea pellets, 15 mL 40% acrylamide (acrylamide:bisacrylamide 19:1), and ~2 g of TMD-8 hydrogen and hydroxide form mixed bead resin with MQH_2O to ~80 mL (*see* **Note 38**). Dissolve urea in acrylamide with shaking at 37 °C for ~10 min, and then adjust volume to 95 mL with MQH_2O. Filter 5 mL of 10× TBE through a 0.2 μm filter (e.g., using a 125 mL filter unit) to remove any particulates that could have formed. Once TBE is filtered, add 95 mL of remaining gel solution to the remove TMD-8 and any additional particulates. Once gel solution is filtered, swirl to mix with TBE.
8. 12% urea-PAG: Make 100 mL of a 12% urea-PAG solution as described for the 6% urea-PAG in **item 7**, but with 30 mL of 40% acrylamide instead of 15 mL.
9. 10% ammonium persulfate (APS) solution in water.
10. *N*,*N*,*N*′,*N*′-tetramethylethane-1,2-diamine (TEMED).
11. Heat lamp or other light source to speed acrylamide polymerization process (*see* **Note 39**).
12. End-labeled pBR322 MspI digested marker (*see* **Note 40**) and 1:50 dilution of pBR322 labeled marker in 1× urea stop dye.
13. Power source capable of 1000 V.
14. Pre-exposed X-ray films or Whatman paper.
15. Phosphorimager screens or similar system to expose radioactive gels (*see* **Note 41**).
16. Amersham Typhoon scanner (GE Lifesciences).

2.5 Data Analysis

1. Computer.
2. ImageQuant software version 5.2.
3. Microsoft Excel.
4. Kaleidagraph version 4.5.0 (Synergy Software) (or similar comprehensive graphing software).

3 Methods

See Fig. 1 as a guide for the overview of when each step should be done. This method can be completed in 5 days where days 1 and 2 include DNA preparation (Subheading 3.1), days 2 and 3 involve in vitro transcription setup and execution (Subheading 3.2) alongside EMSA (Subheading 3.3), day 4 contains urea-PAGE (Subheading 3.4), and day 5 is reserved for data analysis (Subheading 3.5; *see* **Note 1**).

3.1 DNA Preparation

3.1.1 PCR Amplify and Concentrate DNA

1. Set up 5 mL of *bgl* template PCR mixture.
2. Mix reagents in a 15 mL conical tube. Aliquot 50 μL per PCR tube on ice (*see* **Note 42**).
3. Carry out PCR using a version of the protocol in Table 1 (*see* **Note 43**).
4. Once the PCR reaction is complete, pool the reaction and then aliquot 400 μL/1.5 mL nuclease-free, low-binding microcentrifuge tube.
5. Add 1/10 volume of 3 M sodium acetate, pH 5.2 and 2.5× volume of 100% ethanol to precipitate DNA.
6. Incubate for at least 1 h at −20 °C.
7. Spin at 20,000 × *g* for 20 min at 4 °C. Pour off the supernatant.

Table 1
Suggested thermocycling conditions for DNA template PCR

Step	Temperature and Time
1	95 °C for 2 min
2	94 °C for 30 s
3	54 °C for 30 s
4	72 °C for 2 min
5	Repeat steps 2–4 17 times
6	72 °C for 5 min
7	10 °C forever

8. Add 1 mL of 70% ethanol to wash excess salts and spin at 20,000 × *g* for 10 min at 4 °C.
9. Pour off the supernatant. Briefly spin again and pipet off remaining supernatant.
10. Air dry for 10–15 min until no ethanol remains (*see* **Note 44**).
11. Resuspend and combine all DNA pellets in ~500 μL of TE.

3.1.2 Gel Purify DNA

1. Put RNase-free gel box at 4 °C to chill the box for easier pouring (*see* **Note 45**).
2. Make 120 mL of 0.8% agarose solution in 1× TBE with 4 mM EDTA and microwave. Once cool enough to handle for ~10 s, add 300 μg of ethidium bromide/mL agarose (~5 μL) and slowly pour into gel box containing the comb (*see* **Note 46**) at room temperature. Let it solidify.
3. Fill gel electrophoresis system with 1× TBE with 4 mM EDTA to cover the gel.
4. To the 500 μL resuspended PCR product, add ~50 μL 10× loading dye and ~30 μL 80% glycerol to ensure sample sinks to the bottom of each well.
5. Load all the DNA onto one gel dividing over as many wells as necessary so each lane is about 80% full.
6. Run gel at 130 V until DNA is about 1/3 of the way into the gel (*see* **Note 47**).
7. Visualize band with long wave (365 nm) UV (*see* **Note 48**).
8. Use a clean, sterile scalpel blade to excise the agarose piece containing the desired DNA product.

3.1.3 DNA Electroelution

1. Rinse portion of 3500 MWCO dialysis tubing in DI water. Equilibrate in 1× TBE with 3.5 mM EDTA for about 1 min (*see* **Note 49**).
2. Clip the bottom of the dialysis tubing and add about 4 mL of 1× TBE with 3.5 mM EDTA. Add gel piece and remove as many air bubbles as possible before adding second clip.
3. Electroelute DNA by placing the dialysis tubing with the gel piece toward the positive end in enough buffer to just cover the gel piece (*see* Fig. 2).
4. Run at 130 V for about 30 min. Flip gel pieces over ½ way through to ensure the entire gel piece is covered by buffer and all DNA is eluted out of the gel piece. Ethidium bromide stained DNA should be observed collecting against the dialysis tubing toward the positive electrode.
5. Remove dialysis tubing from the buffer and unclip. Pipette 1× TBE (now containing DNA) out of dialysis tubing taking care

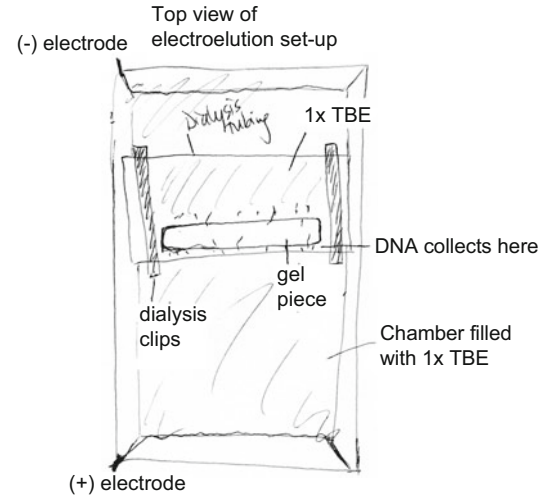


Fig. 2 Schematic of electroelution set-up viewed from the top. Gel fragment cut from gel purification (Subheading 3.1.2) should be placed inside dialysis tubing with ~4 mL of 1× TBE containing 3.5 mM EDTA. The gel should be positioned towards the positive electrode so the DNA will run out of the gel to the space between the gel piece and dialysis tubing. The chamber needs to contain enough 1× TBE with 3.5 mM EDTA to cover the top of the dialysis tubing. Because the tubing will float, halfway through the process, flip the dialysis tubing so the top (as viewed here) is now facing down

to rinse the gel piece and dialysis tubing with the 1× TBE + DNA solution as the DNA can stick to both surfaces.

6. Check with long wave UV light to be sure all DNA was collected.
7. Syringe-filter electroeluted TBE-containing DNA to remove any remaining gel pieces.

3.1.4 DNA Cleanup and Concentration

1. Divide electroeluted DNA into 400 μL aliquots in 1.5 mL microcentrifuge tubes.
2. Add 400 μL of TE-equilibrated phenol and vortex to emulsion (~25 s).
3. Spin at 10,000 × *g* for 7 min at 10 °C.
4. Transfer aqueous layer to 400 μL of 24:1 chloroform:IAA. Vortex briefly.
5. Spin at 10,000 × *g* for 7 min at room temperature.
6. Transfer aqueous layer to a 1.5 mL microcentrifuge tube (*see* **Note 50**).
7. Perform ethanol precipitation as in Subheading 3.1.1.
8. Resuspend in 15 mM HEPES, pH 8.0 to get between 0.5 and 1 μg template DNA/μL. Aliquot 50 μL into 0.6 mL tubes and

store at −20 °C. Remove a tube for use and store at 4 °C for 1–2 months to avoid additional freeze-thaw cycles.

9. Determine concentration by measuring A_{260}.

3.2 In Vitro Transcription

3.2.1 In Vitro Transcription Setup

The mixtures used here can be made in 0.6 mL nuclease-free tubes on ice *before* (*see* **Note 51**) adding radioactive material or indicated components (*see* **Note 52**). The work flow for in vitro transcription and EMSA performed side by side are described in Fig. 3.

1. Use the flow chart in Fig. 4 to determine the volume of each mixture to make based on the desired conditions to test and time points to collect. Example mixtures are described in Tables 2, 3, and 4.
2. Make EDTA/glycogen stop solutions and ethanol precipitation aliquots in 1.5 mL nuclease-free tubes for each time point to collect.
3. *Steps 3 and on are performed the day of the experiment.* Thaw proteins and buffer on ice and thaw rNTPs or NTP mixes, DNA, rifampicin, and DTT at room temperature. Once thawed, immediately place on ice.

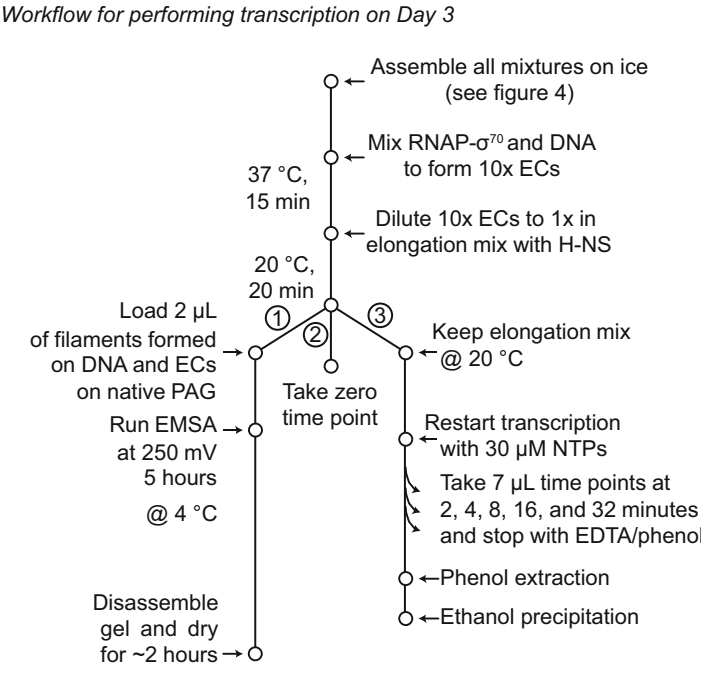


Fig. 3 The workflow for transcription and EMSA done side-by-side on day 3. When performing the in vitro transcription assay, the ECs are made first then H-NS filaments are formed. The EMSA gel is loaded and run while timed samples are collected during the transcription reaction. RNA products are cleaned up before disassembling the EMSA

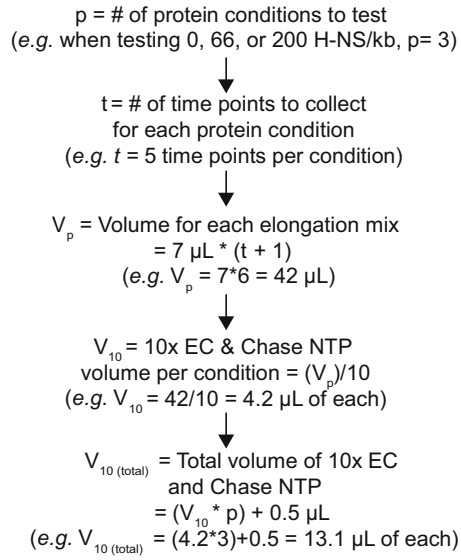


Fig. 4 Workflow to determine the volume of elongation mixes, 10× halted elongation mix, and 10× chase NTP mix required for each experiment. *See* Tables 2, 3 and 4 for volumes of each component in the mixtures

Table 2
Example of 10× halted EC reaction mix for 1× (9 µL) and 2.5× mixtures. The order of addition of components is indicated

Order to add component	Component	[Final]	1× amount (µL)	2.5× amount
[a]	3× EMSA Buffer	1×	3	7.50
[a]	20× initiating NTP mix	1×	0.45	1.125
[a]	25 mM DTT	1 mM	0.36	0.90
[a]	2 mM ApU	150 µM	0.675	1.69
Second to last	[α-^{32}P]GTP	20 µCi	1.1	2.75
Third to last	pMK110 PCR (1240 nM)	100 nM	0.725	1.81
LAST!	2 µM Holo RNAP	150 nM	0.675	1.69
[a]	DEPC-H_2O		2.01	5.04
Total Volume (µL)			9	22.5

[a]Add these at the same time on ice

4. Make both 20× initiating NTP mix and 10× chase NTP mix on ice (Figs. 3 and 4).
5. Make desired volume of 10× halted transcription elongation complex (ECs) mix without [α-^{32}P]GTP and RNAP holoenzyme (Figs. 3 and 4, Table 2).

Table 3
Example of elongation mix for forming bridged H-NS filaments on halted ECs and then subsequent transcription elongation reaction

Order to add components	Component	[Final]	1× amount (μL)	8× amount (μL)
[a]	3× EMSA Buffer	1×	1.7	13.6
[a]	20 μM H-NS	1 μM	0.3	2.4
Add after ECs formed	10× halted ECs mix	1×	0.6	4.8
Last (after filaments formed)	10× chase NTP mix	1×	0.6	4.8
[a]	Elongation master mix		2.8	22.4
Total volume (μL)			6	48

This table shows the amount needed to collect 1 timed sample and 8 timed samples. In this reaction, the total buffer volume of 2 μL comes from 3× EMSA buffer and 20 μM H-NS
[a]Add these at the same time on ice

Table 4
Example of elongation master mix for 1 reaction (timed sample) and for 50 reactions (timed samples)

Elongation master mix	[Final]	1× amount (μL)	50× amount (μL)
RNasin	40 U/μL	0.08	4
1 mg Rifampicin/mL	0.1 mg/mL	0.6	30
25 mM DTT	0.5 mM	0.12	6
DEPC-H_2O		2	100
Total volume (μL)		2.8	140

6. Make desired volumes of elongation master mix and elongation reaction mixtures on ice without adding 10× ECs or 10× chase NTPs (*see* Figs. 3 and 4, Tables 2, 3, and 4).
7. Make mixtures for EMSA samples (*see* Subheading 3.3) on ice before adding radioactive material. These two protocols are done side by side, but are described independently (*see* Figs. 1 and 3).

3.2.2 Performing In Vitro Transcription Reaction

1. Form 10× halted ECs by adding in the following order 20 μCi [α-^{32}P]GTP then 150 nM RNAP holoenzyme (final concentrations) to the 10× transcription elongation mix. Mix well by gentle pipetting and incubate at 37 °C for 15 min. Place on ice after incubation.
2. While ECs are forming, add 100 μL TE-equilibrated phenol to EDTA/glycogen containing tubes. Keep tubes at room temperature while collecting timed samples.

3. Dilute 10× ECs to 1× (10 nM ECs) in elongation mixes containing H-NS (*see* Table 3). Place tubes in 20 °C water bath (*see* **Note 53**). Incubate for 20 min to form H-NS filaments.
4. Form H-NS filaments on DNA without ECs at 20 °C alongside previous step (**step 3**) (*see* Subheading 3.3.2) as a control to ensure the presence of RNAP does not change H-NS filament formation.
5. Load 2 μL of each H-NS filament formed on DNA and ECs on native-PAG (*see* Subheading 3.3.3).
6. Take a time zero sample by removing 3 μL of each elongation mix and adding to 3 μL of 2× urea stop dye. Set aside.
7. Restart elongation by adding 1/10 volumes of 10× chase NTPs to restart the reaction (*see* Table 3). To begin time course at "0 s," set timer for 5 s, pipet up necessary amount of NTPs, begin timer, remove tube from water bath, add NTPs when timer goes off. Gently pipet up and down a few times to mix. Place the tube back in the water bath.
8. Collect timed samples by removing 7 μL of elongation mix ~4 s before desired time then adding to EDTA/glycogen/Phenol mix when timer reaches appropriate time to stop the reaction. Immediately vortex for ~5 s to completely stop the reaction. Set the tube aside. Allow for ~10–15 s between timed samples.
9. Continue collecting timed samples and then adding chase NTPs to other conditions until all reactions are completed.

3.2.3 RNA Cleanup

Removal of proteins from RNA is necessary to resolve gel electrophoresis bands cleanly as high levels of nucleic acid-binding proteins in the samples causes smearing of bands during urea-PAGE.

1. After all samples are collected, vortex the samples for 25–30 s to emulsify. Spin at 10,000 × *g* for 7 min at 10 °C to separate phases (*see* **Note 54**).
2. Take 90 μL of the top, aqueous phase being sure to avoid the interface or collecting any phenol. Combine the recovered aqueous phase with cold ethanol precipitation aliquots and vortex briefly to mix. Place tubes at −20 °C for at least 1 h to precipitate RNA products (ethanol precipitation).
3. During the precipitation, prepare urea-PAG described in Subheading 3.4.
4. Centrifuge tubes at 20,000 × *g* for 20 min at 4 °C. Pour off supernatants into a 50 mL conical tube to collect radioactive waste. Blot any drops of ethanol by inverting the tube briefly on a kimwipe, careful to avoid radioactive contamination.

5. Remove residual salt by adding 400 μL of 70% ethanol to each tube. *Do not vortex!* Spin at 8000 × *g* for 5 min at 4 °C. Pour off the supernatant.
6. Use a SpeedVac Concentrator to dry samples for 15 min (*see* **Note 55**).
7. Estimate the radioactivity in each pellet in counts per second using a Geiger counter. Resuspend in formamide stop dye so all have similar counts/μL. Loading 50–200 counts/s in each well is enough signal to scan the gel before the RNA bands separated by urea-PAGE diffuse.
8. Separate RNA products by urea-PAGE (*see* Subheading 3.4).

3.3 Electrophoretic Mobility Shift Assay

The EMSA run alongside the in vitro transcription assay is a good control to ensure H-NS binds to DNA to form filaments under the exact same conditions used for transcription. An additional control to form filaments on DNA that does not contain ECs as a side-by-side can confirm that RNAP does not interfere with filament formation is also described. Typically, the gel is poured on day 2 (day before transcription is performed), and then set-up to run alongside the in vitro transcription reaction. Since this step is described separately from transcription, see notes in Subheading 3.2 in vitro transcription and Fig. 3 to see when to perform these steps.

3.3.1 Pour and Set Up 3% Native Polyacrylamide Gel

1. Clean both gel plates well with Windex, water, then ethanol. Treat short glass plate with ~1 mL 10% dichlorodimethylsilane solution in a fume hood. Wipe excess off the plate with Kimwipe and let it dry for ~15 min (*see* **Note 56**). Clean both plates well with Windex, water, then ethanol as before.
2. Assemble 20 cm × 20 cm gel plate sandwich with 1.5 mm side spacers and clamp gel over spacer with binder clips. Insert the 25-well comb slightly offset at the top of the gel so there is room to pour in the gel solution (*see* Fig. 5).
3. Using a positive-displacement pipet, add ~4 mL of 1% agarose in 0.5× TBE to the bottom of the glass plates while the plate assembly is lying horizontally to create the bottom spacer (*see* **Note 57**). The agarose spacer should be about 1 cm tall. Allow to solidify (~5 min). This agarose is added to act as a plug to prevent acrylamide solution from leaking out of the bottom of the plate assembly during pouring and polymerization.
4. Prepare 3% acrylamide gel solution as in Subheading 2.3.1 just before pouring the gel.
5. Add 600 μL of 10% APS and mix. Add 80 μL of TEMED to catalyze the reaction and mix.
6. Hold the gel up at ~30° with respect to the bench then pour acrylamide solution into gel plates from the top using a 60 mL

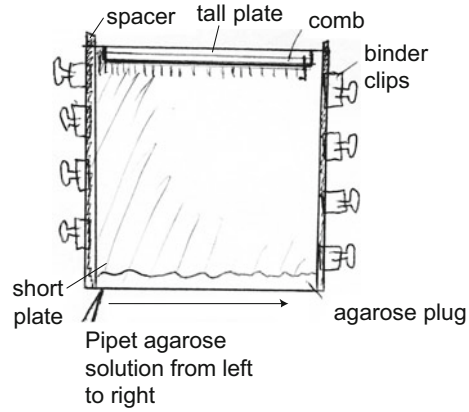


Fig. 5 Schematic of native-PA gel sandwich with assembled gel plates and agarose plug for native PAG. Spacers should be flush with the bottom and held in place with binder clips (not to scale). The agarose plug should be added from left to right, 1 mL at a time, after spacers are secured. The comb should be inserted at the top slightly offset to allow pouring of acrylamide solution between plates

syringe at a slow and steady pace. Insert a comb. Clamp gel plates just below the comb (*see* **Note 58**).

7. Place gel under heat lamp to polymerize (~1 h) (*see* **Note 59**).
8. The gel can be wrapped and stored overnight before proceeding to the next step. Wrap the comb with a damp paper towel and wrap the entire gel with cling wrap to keep moist. Do not let the gel dry out before running it.
9. On day 3, or when performing the transcription reaction, unwrap the gel and carefully remove the comb. Attach the gel to cooled vertical rig and add ~1 L of 0.5× TBE to both buffer chambers.
10. Clean out wells gently using a syringe and 18-gauge needle loaded with 0.5× TBE.
11. Prerun gel for at least 5 min at 250 mV before loading samples (*see* **Note 60**).

3.3.2 Form Filaments on DNA Not Containing ECs

1. Set up kinase reaction to make 5′-end-labeled DNA template and incubate at 37 °C for 30 min. Inactivate T4 PNK by incubating reaction at 65 °C for 15 min (*see* **Note 61**).
2. While setting up transcription mixes on day 3, make DNA control reactions for the number of filaments being tested. The volume of DNA control master mix needed is 5 μL for each NAP concentrations being tested plus one extra reaction. Total reaction volume including the master mix and NAP will be 7.5 μL for forming H-NS filaments on DNA without ECs.

3. Add 3× EMSA buffer and H-NS diluted in 3× EMSA buffer to separate 0.6 mL nuclease-free tube for each protein-DNA combination. Total volume of buffer and protein should be 2.5 μL. The final concentration of buffer components will be 1× after 5 μL of DNA control master mix is added.
4. When forming filaments for transcription experiment, add 5 μL of DNA master mix to each buffer and protein condition (total volume 7.5 μL).
5. Incubate DNA and H-NS at 20 °C for 20 min alongside elongation mixes containing H-NS and halted ECs (*see* Subheading 3.2.2).

3.3.3 Perform Native PAGE to Confirm Protein-DNA Interactions

1. After prerunning the gel, gently clean out the wells again by flushing with 0.5× TBE using a syringe and 18-gauge needle.
2. After forming filaments (**step 4** of Subheading 3.2.2, in vitro transcription and Subheading 3.3.2), load 2 μL of each DNA only sample and each transcription sample onto the 3% native gel.
3. Add 2 μL of 10× loading dye to an empty well or the well containing DNA only (*see* **Note 62**).
4. Run gel at 250 mV for ~5 h at 4 °C or until the XC band is at the bottom of the gel.
5. Disassemble the gel from gel rig and dispose of running buffer following radiation safety procedures.
6. Remove side spacers and agarose spacer from gel. Using a spatula/wedge, remove short plate from the gel. The gel should stick to the tall, non-silanized plate.
7. Place Whatman paper cut to the size of the gel on the top of the gel. Remove the tall plate and the gel should stick to Whatman paper.
8. Cover the gel with cling wrap and dry the gel on a gel drying vacuum system at 80 °C until completely dry (~2 h) (*see* **Note 63**).
9. Expose the dried gel to a phosphorimager screen or radiographic film and visualize. The length of the exposure depends on how radioactive the samples are, ranging from 15 min to a few hours.
10. The mobility of DNA species will decrease when RNAP, H-NS, or both are added to DNA (*see* Fig. 6).

3.4 Urea-PAGE

Gels should be cast on day 3 during the ethanol precipitation and run on day 4 after all timed samples have been collected and purified. For analyzing RNA products from pMK110, both 12% and 6% urea-PAGE were performed to resolve RNAs from 26–300

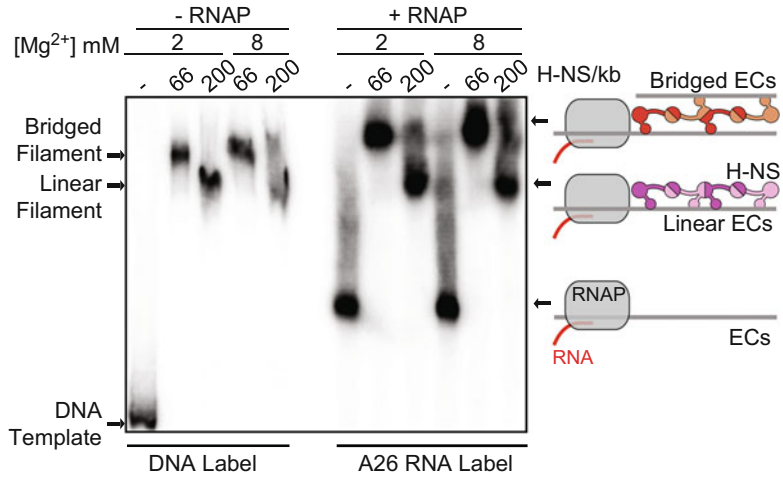


Fig. 6 Example of native PAGE run alongside transcription showing formation of H-NS filaments on DNA (left, DNA label) and on ECs (DNA containing RNAP, right, A26 RNA label). Species migrate slower when H-NS is bound to the DNA at either 66 or 200 H-NS/kb than without H-NS present. At 66 H-NS/kb, we found H-NS adopts a bridged filament conformation, whereas at 200 H-NS/kb, H-NS favors a linear filament [3]. Cartoons to the right of the gel show possible conformations of these filaments. One H-NS monomer is made up of two oligomerization domains (larger circles) and a DNA-binding domain (small circle) all connected by more flexible linkers (lines). The monomers interact to form two distinct conformations of filaments (linear on either DNA or ECs based on H-NS concentration described as linear and bridged). At the top, bridged H-NS filaments are made of red and tan H-NS monomers that interact with two segments of DNA (gray line). Linear filaments are shown as dark and light purple H-NS monomers interacting with one DNA segment. RNAP (gray rectangle) and RNA (red line) are shown at halted at the A26 site on the *bgl* template. This figure was originally published in Kotlajich, M.V., et al. eLife 2015;4:e04970

and 140–1482 nt, respectively. Two gel sandwiches should be made to accommodate both percentage gels. Electrophoresis on both gels occurs side by side.

1. Clean short glass plate (31.12 cm × 41.28 cm) thoroughly with Windex, water, then ethanol.
2. Silanize short plate by coating with ~2 mL of 2% dichlorodimethylsilane and let it dry in fume hood for about 10 min (*see* **Note 64**).
3. Clean and dry both short and tall glass plates with Windex, water, then ethanol.
4. Assemble urea-PA gel plate sandwich. Place cleaned spacers along edges of plates and use binder clips to secure plates and spacers together. Insert clean comb at top of plates (*see* Fig. 7 and **Note 65**).

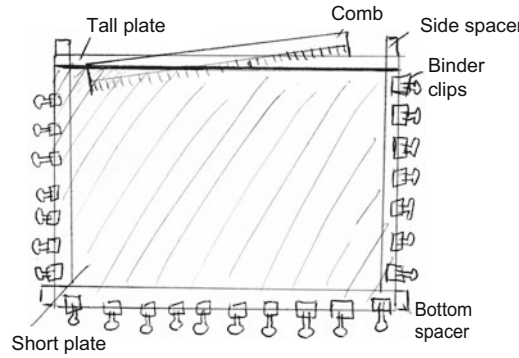


Fig. 7 Schematic of gel sandwich made for urea-PAG casting. The assembly is done horizontally where the short plate is assembled on the top of the tall plate with the side and bottom spacers in between. Binder clips (not to scale) are placed around the edges of the sandwich to secure spacers. The comb is placed offset in the top of the sandwich to facilitate pouring. Agarose can be added between the bottom and side spacers if leaks occur during pouring

5. Make 12% and 6% urea-PAG solutions just before pouring the gel.
6. Add 200 μL 10% APS and the gel solution at room temperature and mix. Then, add 100 μL TEMED and mix.
7. Using a 60 mL syringe, pour gel solution at a slow and steady rate between glass plates held at ~30° angle above the bench with the comb side up to allow gel solution to flow to the bottom.
8. Gently insert the comb the rest of the way without introducing air bubbles and secure with clamp over the comb.
9. Gel will polymerize in ~45 min under heat lamp. Once polymerized, the gel can be stored overnight by wrapping the top of the gel (over the comb) with damp paper towels and cling wrap to prevent the gel from drying out (*see* **Note 66**).
10. To set up the gels to run (day 4), carefully remove the comb and bottom spacer. Use a razor blade to remove any excess acrylamide that might have polymerized around the edges and the wells. Set up the gel in vertical gel rig. Place foam spacers at the top of the short spacer above side spacers to prevent buffer leaking out of the top. Attach an aluminum plate on the outer face of the gel with binder clips to diffuse heat across the gel so lanes run evenly. The aluminum plate should also contain an adhesive thermometer to monitor temperature during the run (*see* **Note 67**).
11. Add 0.5× TBE to top and bottom buffer reservoirs (~2 L total will fill this gel system described here) remove air bubbles and unpolymerized acrylamide within the wells using an 18-gauge

needle and syringe. Air bubbles trapped where the bottom spacer was can be removed using a bent (~110°) needle attached to the syringe.

12. Prerun gel at 50 W for ~1 h or until gel temperature reaches ~45 °C.
13. Heat samples and 1:50 dilution of labeled marker at 95 °C for 3 min.
14. Spin at 20,000 × *g* for 1 min to collect the sample in the bottom of the tube.
15. Quickly load 1 μL of each sample or marker onto gels (*see* **Note 68**). For best resolution, the sample should form a single band at the bottom of each well. If the sample is too diffuse in the well, the RNA products will not be resolved. Early timed samples (and time zero sample), which have short RNA products, should be loaded on the 12% gel whereas later timed samples with longer RNA products should be loaded on the 6% gel. Some of the samples can be loaded on both gels for better resolution of the RNA products (*see* **Note 69**).
16. Run 12% gel until XC band is ~¾ of the way down the gel to visualize RNA products from 26 to ~300 nt.
17. Run the 6% gel so that both dye bands run off, then load ~3 μL of 2× urea stop dye in 3–4 wells and run until BpB band is ~1/2 way down the gel to resolve RNA products from ~140 nt to the end of the template (1482 nt).
18. Remove buffer from reservoirs and dispose of in proper containers as the bottom buffer typically contains radioactivity. Remove the gel from the vertical gel electrophoresis system and remove the side spacers. Cool the gel by running cool water along the short plate.
19. With short plate facing up, remove the top plate by pushing up with a spatula wedged between the plates. The gel should stick to the tall, non-silanized plate. Place a pre-exposed X-ray film (*see* **Note 70**) on the top of the gel and slowly remove the gel from the top plate taking care not to rip the gel. The gel should stick to the film. Cover the gel with cling wrap, ensure no liquid can contact phosphorimager screen by taping the wrap down on the back of the film.
20. Expose to phosphorimager screens from 2 h to overnight depending on radioactivity of samples.
21. Visualize RNA products with phosphorimager scanner (e.g., Amersham Typhoon; GE Lifesciences) (*see* Fig. 8). RNA products will get longer with time. Locations of RNAP pausing at will appear as bands that last for differing persistence depending on pause characteristics.

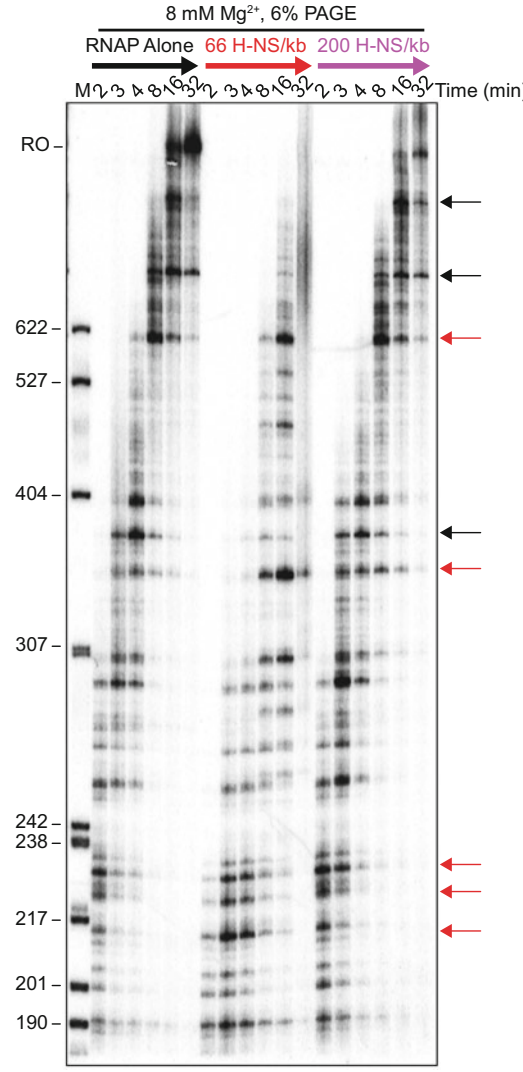


Fig. 8 Example of 6% urea-PAGE result showing RNA products generated in the presence of none, 66, or 200 H-NS/kb at 2, 3, 4, 8, 16, and 32 min and labeled pBR322 MspI digested marker (M) run alongside. Bands indicate RNA products generated at different time points showing where RNAP was on the DNA template at that time point. Black arrows indicate sequences where RNAP pauses along the template. Some pauses (red arrow) are enhanced (i.e., appear for longer amounts of time) by H-NS. This figure was originally published in Kotlajich, M.V., et al. eLife 2015;4:e04970

3.5 Data Analysis

Analysis of transcription results can be easier if the transcription products are plotted as a function of nucleotide length and relative signal level, and compared across conditions or time points (Fig. 9) rather than only visualized on a gel. These pseudo-densitometry plots then show the distribution of RNA products by length instead of gel position as in a true densitometry plot. The conversion of

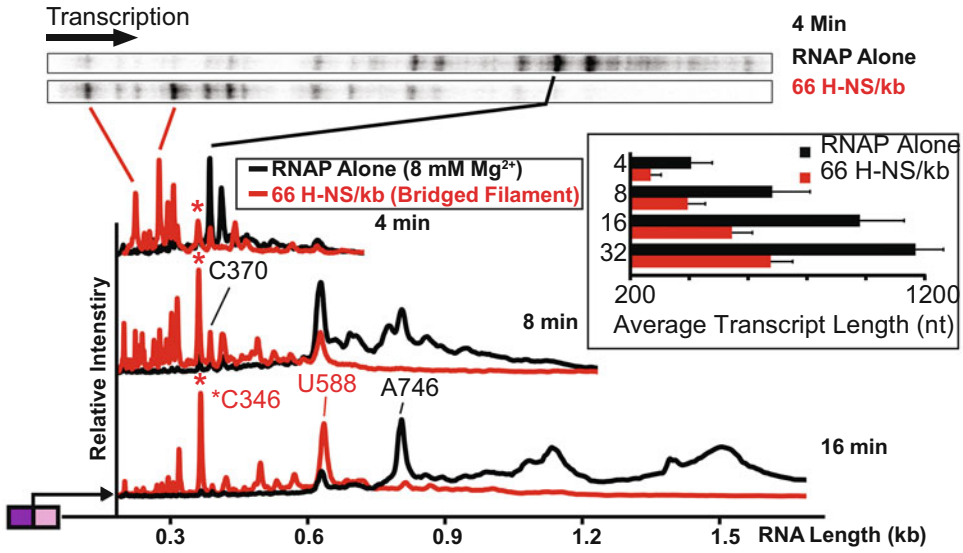


Fig. 9 Example of pseudo-densitometry profiles created in Microsoft Excel from line analysis in ImageQuant software. Here, pseudo-densitometry profiles of RNAP alone and 66 H-NS/kb are compared at different time points. ATL (average transcript lengths, nt) are inserted and show an average of at least three independent experiments. Pauses at various positions (C346, C370, U588, and A746) are highlighted; red text indicates H-NS stimulated pauses and black indicates no H-NS enhancement of pauses. Enhanced pauses appear over more time points than non-enhanced pauses. This figure was originally published in Kotlajich, M.V., et al. eLife 2015;4:e04970

each gel lane (sample) to such a plot using the Excel spreadsheet program is described here. In this analysis, the DNA marker lengths are assumed to be similar to RNA lengths of the transcripts and used to convert the pixels to a nucleotide length. The plots show the major pauses and the overall transcription pattern for each sample directly comparable across conditions. The average transcript length (ATL) of each sample can also be calculated from these data. A comparison of ATL can help to distinguish overall differences between conditions, but is less useful than comparing changes in RNAP behavior at each template position.

3.5.1 Conversion of Pixels to Nucleotides

1. Use ImageQuant software to draw lines over the length of the image of each gel lane of interest. The line should be thick enough to encompass most of the density of signal in each lane, and should start and end at similar transcript lengths for each lane. It is most straightforward to analyze the results when lines are drawn from bottom to top so the lines length dimension will be output as a smaller value for shorter transcripts (*see* **Notes 71** and **72**).
2. Use the same line characteristics for each lane, including the marker lane, and align each line to the bottom of the gel so each line starts at the same transcript length.

3. Create a graph of each line in ImageQuant by highlighting all lines and selecting "create graph" from the analysis menu. Highlight all graphs, then copy and paste in Excel. Data columns should appear in the excel spreadsheet listing a pixel numbers and a corresponding rfu(z) (intensity) value for each pixel number in the lane.
4. In Excel, create a graph of marker intensity vs. pixel number which should show a peak pattern similar to the bands observed on the gel. Identify the pixel location for each peak by hovering the mouse over the peak and confirming the center of the band with the highest rfu(z) value. Each pixel value can be matched to the known nucleotide size for each marker band.
5. In Kaleidagraph, create a scatter plot of nucleotides vs. pixels using the identified pixel value for each nucleotide length in the marker lane.
6. In Kaleidagraph, fit these data to a polynomial function using the "Curve Fit" function.
   (a) For a 6% gel, most graphs can be fit with a 5- or 6-factor polynomial function. For a 12% gel, a combination of a 4-factor (fit to nucleotides <147) and a 6-factor (fit to all values) will fit the pixel vs. nucleotide graph well.
   (b) Equation 1. Example of a 4-factor polynomial function

$$\text{Nucleotide} = \left(\text{M4} \times \text{pixel}^4\right) + \left(\text{M3} \times \text{pixel}^3\right) + \left(\text{M2} \times \text{pixel}^2\right) + (\text{M1} \times \text{pixel}) + \text{M0}$$

   where M0–M4 are coefficients defined by the polynomial curve fit in Kaleidagraph.

7. In Excel, convert each pixel on the gel to a nucleotide using the function determined in Kaleidagraph and shown above. Depending on the location, some pixels will be converted to fractional nucleotide lengths (*see* Fig. 10). When combining pixel-to-nucleotide conversions from multiple fits, use the nucleotide value they converge on as the value to switch between a 4-factor and 6-factor.

3.5.2 Making Pseudo-Densitometry Plots

1. Before graphs are made, the intensity for each lane should be normalized so each lane has the same total intensity. Normalization is accomplished by first calculating the sum of the rfu(z) in each lane and then identifying the largest sum (Fig. 11 "total rfu(z)" row and column AA). For each lane, divide the largest sum by the sum of that lane (Fig. 11, formula bar). This generates a multiplication factor that each rfu(z) value in that lane should be multiplied by to equalize the total rfu(z) in each lane.

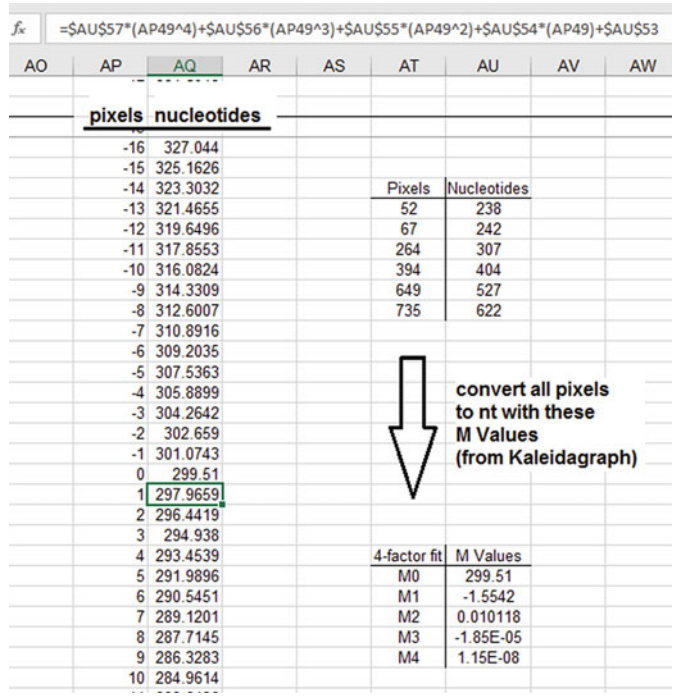
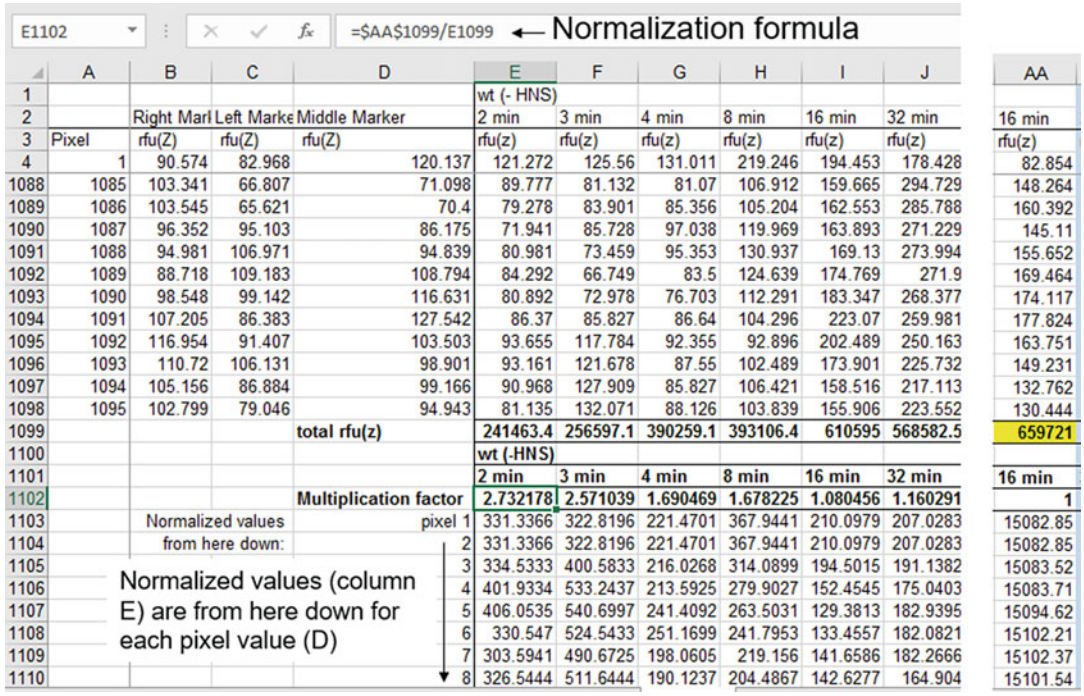


Fig. 10 A screenshot from an Excel workbook showing the formula used to convert pixels (column AP) to nucleotides (AQ). The formula uses the coefficient values (M values) from Kaleidagraph (column AU) generated from a curve-fit of the pixels vs. nucleotide plot (using pixel and nucleotide values in columns AT and AU. The conversion is indicated by large arrow). The pixel-to-nucleotide formula is shown in the top formula bar (fx) for the highlighted cell

2. Now that each pixel has a related length value in nucleotides, graphs for each lane can be made. Plot the normalized rfu(z) values vs. length in nucleotides for one time point and all the conditions on the same graph (*see* **Note 73**).
3. Plots can be modified by changing the x-axis range if any peaks do not line up between conditions for a certain time point because the gel ran with a "smile" or showed a wavy pattern (*see* **Notes 74** and **75**).

3.5.3 Calculating Average Transcript Lengths (ATLs)

ATLs can be calculated to reveal the average progression of RNAP at each time point or in each condition and thus enable a quantitative comparison of RNAP behavior between samples. For example, the ATL should increase with time as RNAP progresses along the template in the RNAP alone condition (Fig. 9). If many pauses are stimulated, the ATL will be decreased relative to minimal or no pause stimulation (Fig. 9, 66 H-NS/kb condition). An example equation used to calculate ATL is



	A	B	C	D	E	F	G	H	I	J	AA
1					wt (- HNS)						
2		Right Marl	Left Marke	Middle Marker	2 min	3 min	4 min	8 min	16 min	32 min	16 min
3	Pixel	rfu(Z)	rfu(Z)	rfu(Z)	rfu(z)	rfu(z)	rfu(z)	rfu(z)	rfu(z)	rfu(z)	rfu(z)
4	1	90.574	82.968	120.137	121.272	125.56	131.011	219.246	194.453	178.428	82.854
1088	1085	103.341	66.807	71.098	89.777	81.132	81.07	106.912	159.665	294.729	148.264
1089	1086	103.545	65.621	70.4	79.278	83.901	85.356	105.204	162.553	285.788	160.392
1090	1087	96.352	95.103	86.175	71.941	85.728	97.038	119.969	163.893	271.229	145.11
1091	1088	94.981	106.971	94.839	80.981	73.459	95.353	130.937	169.13	273.994	155.652
1092	1089	88.718	109.183	108.794	84.292	66.749	83.5	124.639	174.769	271.9	169.464
1093	1090	98.548	99.142	116.631	80.892	72.978	76.703	112.291	183.347	268.377	174.117
1094	1091	107.205	86.383	127.542	86.37	85.827	86.64	104.296	223.07	259.981	177.824
1095	1092	116.954	91.407	103.503	93.655	117.784	92.355	92.896	202.489	250.163	163.751
1096	1093	110.72	106.131	98.901	93.161	121.678	87.55	102.489	173.901	225.732	149.231
1097	1094	105.156	86.884	99.166	90.968	127.909	85.827	106.421	158.516	217.113	132.762
1098	1095	102.799	79.046	94.943	81.135	132.071	88.126	103.839	155.906	223.552	130.444
1099				**total rfu(z)**	**241463.4**	**256597.1**	**390259.1**	**393106.4**	**610595**	**568582.5**	**659721**
1100					wt (-HNS)						
1101					2 min	3 min	4 min	8 min	16 min	32 min	16 min
1102				**Multiplication factor**	**2.732178**	**2.571039**	**1.690469**	**1.678225**	**1.080456**	**1.160291**	**1**
1103		Normalized values		pixel 1	331.3366	322.8196	221.4701	367.9441	210.0979	207.0283	15082.85
1104		from here down:		2	331.3366	322.8196	221.4701	367.9441	210.0979	207.0283	15082.85
1105				3	334.5333	400.5833	216.0268	314.0899	194.5015	191.1382	15083.52
1106				4	401.9334	533.2437	213.5925	279.9027	152.4545	175.0403	15083.71
1107				5	406.0535	540.6997	241.4092	263.5031	129.3813	182.9395	15094.62
1108				6	330.547	524.5433	251.1699	241.7953	133.4557	182.0821	15102.21
1109				7	303.5941	490.6725	198.0605	219.156	141.6586	182.2666	15102.37
1110				8	326.5444	511.6444	190.1237	204.4867	142.6277	164.904	15101.54

Fig. 11 Example normalization formula from a screenshot of an Excel workbook showing the total rfu(*z*) value for seven time points (row 1099, total rfu(*z*)) with the maximum total signal highlighted in yellow (row 1099, column AA). The multiplication factor was determined using the formula in the formula (fx) bar at the top (for column E, WT − H-NS condition at 2 min). The rfu(*z*) values for each pixel were multiplied by the multiplication factor to give the normalized rfu(*z*) in the rows 1102+ below the multiplication factor. For example, the rfu(*z*) at pixel 1 for WT − H-NS at 2 min was 121.27 before normalization and then 331.336 after normalization with the 2.73 multiplication factor

$$\mathrm{ATL} = \frac{\Sigma(\mathrm{rfu}(z) \times \text{transcript length})}{\Sigma(\mathrm{rfu}(z))}$$

where rfu(z) is the intensity of each position minus the background signal (*see* **Note 76**) and the transcript length is derived from the pixel to nucleotide conversion described above. For each sum, all points within the desired range of the lane should be included (e.g., from 150 to 800 nt for the 4 min time point).

4 Notes

1. A typical timeline for the method is given in Fig. 1 and described here which covers 5 days from start to finish. Days 1 and 2: DNA preparation will take 1.5 days plus 2–3 h to set up all reagents for in vitro transcription and pour native PAG. Day 3: 1–2 h to set up the experiment, then perform transcription and run EMSA alongside, about a ~7 h day in total. The urea-PAG should be cast on day 3 and not day 2 so that it does

not dry out before use. Day 4: Run urea-PAGE and visualize gels. Day 5: Data analysis. In vitro transcription and analysis by EMSA need to occur on the same day so that RNA products can be interpreted alongside conformation of EC and filament formation. Due to the time it takes to perform urea-PAGE (~5 h), electrophoresis and analysis are completed on the 2 days following in vitro transcription and the EMSA, but could be performed on the same day if required.

2. Modifications to this protocol, including the use of a different DNA template, buffer conditions, NTP concentrations, time points, and temperature could be applied to determine the impact of H-NS filaments under different conditions or to investigate how other NAPs impact elongation by RNAP.
3. The full sequence of pMK110 and an aliquot of the plasmid can be obtained from Addgene (ID #99535). The template described here was designed for *E. coli* RNAP and H-NS filaments, so it contains the appropriate promoter and a long region of AT-rich DNA on which H-NS filaments are known to form. Features of the DNA template can be modified (e.g., promoter, downstream DNA sequence, or length) to suit the properties of the roadblocking protein being studied or the RNAP being used.
4. Any Taq polymerase can be used as long as it has high enough fidelity to make long DNA products. We have found OneTaq DNA polymerase from NEB to be well suited for this amplification. Use the supplied OneTaq standard reaction buffer (20 mM Tris–HCl, 22 mM NH_4Cl, 22 mM KCl, 1.8 mM $MgCl_2$, 0.06% IGEPAL® CA-630, 0.05% Tween® 20, pH 8.9 at 25 °C) for PCR.
5. Be sure to check for the synthesis of the correct PCR product with a small-scale (~50 μL) reaction before setting up large-scale PCR. The annealing temp or buffer components may need to change for different reactions. A 5 mL PCR of *bgl* template used here will yield enough DNA for ~12 transcription experiments including DNA controls. The PCR can be scaled up or down if using a different template or different amount of DNA is required.
6. The pH will be at 8.3 after addition of listed reagents and should not be adjusted or the desired ion balance will be disrupted!
7. EDTA will go into solution around pH 8.0.
8. The additional EDTA will chelate any free metal ions which can damage DNA during storage and prevent interpretable in vitro transcription products.

9. This size gel electrophoresis system can be custom made or a comparable system can be purchased from Bio-rad or other company (e.g., product #170445 from Bio-rad which casts a 15 × 7 cm gel).
10. Dialysis tubing of a different molecular weight cutoff may be required depending on the size of the template used in the experiment. The MWCO must be at least one-third the size of the template (e.g., for a 1.5 kb template with a molecular weight of 965 kDa, dialysis tubing of at least 300 kDa must be used) to avoid loss of sample.
11. We purchase water-saturated phenol supplied with alkaline buffer (Tris) to adjust the pH to 7.9.
12. Maintaining an RNase-free environment is key to success of in vitro transcription, particularly when transcribing long DNA templates. RNA degradation will obscure the results. RNase activity will be evident by the appearance of short RNA products at times after longer RNA products have already been synthesized. Typically, treatment of bench surface, agarose gel running boxes, and pipets with an RNase remover product (e.g., Eliminase from Thermo-Fisher) works well to remove RNases from the working area and supplies. Use nuclease-free filter tips to prevent any RNase contamination on pipettes from causing RNA degradation.
13. This protocol was developed using core RNAP and σ^{70} purified as described in [17–19]. Purified *E. coli* RNAP holoenzyme obtained commercially (e.g., NEB) or purified by other methods may work also, but we have not compared other enzymes. If using commercial sources of RNAP, be sure the enzyme is concentrated enough to form 100 nM ECs. Additionally, this protocol could be modified to use RNAP from other bacterial species. Appropriate changes to the DNA template (e.g., to the promoter sequence) would need to be made to match the RNAP chosen for experiments.
14. Fresh dilutions of NTPs should be made after 2 months or if any contamination is suspected.
15. Small (2 μL) aliquots of 2 mM rNTPs can be made for use in initiation mixes.
16. Larger (50 μL) aliquots of 2 mM rNTPs can be used for chase NTPs.
17. 10× Chase NTPs should be made the day of or the day before the experiment. Aliquots of 10× chase NTP mixes can also be stored at −80 °C. The NTP concentration can be modified. 30 μM NTPs is a low substrate concentration, causing slower transcription elongation, allowing for visualization of maximal pause stimulation by a roadblocking protein. 1 mM NTPs can be used if more physiological conditions are desired, but time

points need to be taken more quickly (e.g., *10, 20, 40, 60, 120, and 180 s*).

18. Since radioactive isotopes are used, safety protocols and regulations must be followed when performing this experiment.
19. ApU can be stored in 25 mM stock solutions at −80 °C and then diluted to 2 mM ApU for working stocks. ApU is a dinucleotide to help initiate transcription.
20. 25 mM DTT can be made by diluting a 1 M stock of DTT in MQH_2O. DTT should be stored at −20 °C.
21. Storing the HEPES buffer in the dark at 4 °C will prevent formation of free radicals which can damage nucleic acids and protein. Degradation of HEPES is indicated when the solution turns a yellow color. Dispose of any yellow solution and make fresh before using in transcription experiments.
22. Add solid potassium glutamate and magnesium aspartate to HEPES buffer. Heat (~65 °C) to dissolve reagents as it can be difficult to get magnesium aspartate into solution. After these components are dissolved, add NP-40, BSA, and glycerol. Filter and store aliquots of the solution at −80 °C to prevent formation of free radicals.
23. This buffer was optimized to study H-NS family proteins, but can be adapted to look at other conditions for H-NS or another bacterial nucleoid-associated protein. Changes in pH, buffering ions, salt concentration, magnesium levels, glycerol, temperature, etc. can be made. We recommend trying new conditions alongside the described EMSA buffer to compare ability of RNAP to initiate and elongate and H-NS to form filaments as a positive control.
24. Make a 50 mg rifampicin/mL stock by dissolving 50 mg rifampicin in DMSO or dimethylformamide. Store this stock solution wrapped in foil at −20 °C. Use proper safety procedures when handling organic solvents. The 50 mg/mL stock will last for ~6 months wrapped in foil at −20 °C and the 1 mg/mL stock will last about 1 month at −20 °C.
25. Working solutions of proteins should be in 3× EMSA buffer so that when added to the elongation mix, the final buffer concentration is 1× EMSA buffer. For H-NS, this was accomplished by diluting the freezer stocks of H-NS (~150 μM) to 20 μM in 3× EMSA buffer. In other cases, the stock solution of protein might be at or near the desired concentration for a working stock (e.g., ~20 μM). In these situations, the protein storage buffer should be exchanged for 3× EMSA buffer by dialysis. The acetylated BSA could be removed from 3× EMSA buffer and added separately in the reaction to avoid having the protein of interest stored with BSA.

26. 10× ECs should be made the day of the experiment on ice. We recommend adding DNA last to avoid losing DNA if a mistake is made when adding other reagents. *Do not add [α-^{32}P]GTP or RNAP holoenzyme until indicated!*
27. Make the elongation master mix and elongation reaction mixture the day of the experiment on ice. *Do not add 10× ECs or 10× chase NTPs until the appropriate time during the experiment!*
28. The final EMSA buffer concentration in the elongation mix should be 1× after including volumes of both the buffer and diluted H-NS added.
29. EDTA will chelate the Mg^{2+} to stop RNAP activity. The phenol (added later) will separate proteins from the nucleic acids. The glycogen is added as a carrier for the ethanol precipitation.
30. Formamide can degrade to formic acid and ammonia, which can degrade nucleic acids, so be sure to use deionized formamide. Formamide stop dye should be stored in aliquots at −80 °C. Working stocks can be kept at −20 °C, but if the solution is no longer frozen, discard because the formamide has degraded and will damage RNA.
31. A vertical electrophoresis system with an option to attach a circulating water bath is available from a variety of vendors such as ThermoFisher (product #FB-VE20-1). This 20 × 20 cm gel system provides enough space to load all samples for transcription conditions and DNA controls. This gel setup can be run on the bench. Alternatively, the native PAGE can be run in a gel rig in a cold room or dairy case.
32. We use a prepared solution from Bio-rad, catalog #161-0156.
33. A 3% native PAG was necessary to visualize H-NS filaments formed on 1.5 kb DNA because of the large size of the H-NS: DNA complex. If performing an experiment with smaller DNA templates or proteins that bind to fewer or discrete sites, a higher percentage gel can be used.
34. Adjust volume as needed for number of samples being tested. Each sample requires 0.75 μL of 100 nM DNA.
35. We use custom gel plates and vertical electrophoresis system with plates that are 33.02 cm × 41.28 cm and 31.12 cm × 41.28 cm for the tall and short plates, respectively. Spacers and comb are 0.05 cm thick and contain either 54 or 80 wells that easily hold 1–2 μL of sample. Plates of this size hold ~80 mL of gel solution.
36. A commercial gel running system (e.g., Owl vertical electrophoresis system, ThermoFisher) could be used. Aluminum plates are used to even distribute heat across the gel. Addition of an adhesive thermometer strip allows for monitoring of temperature during electrophoresis.

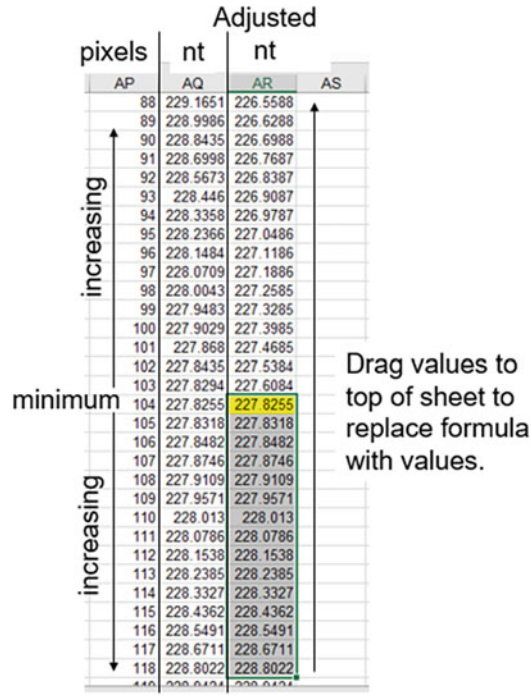
37. We use a prepared solution from Bio-rad, catalog #161-0144.
38. TMD-8 hydrogen and hydroxide form is added to deionize the urea, as many urea products sold contain ions that would disrupt the ion balance required for resolving bands on urea-PAGE. TMD-8 should not contact 10× TBE as the ion balance will be ruined. Filter 10× TBE separately from remaining acrylamide gel solution as described.
39. The heat lamp is useful for speeding up the polymerization process. A 60 W yellow bulb should generate enough heat to speed up polymerization. If no lamp is used, polymerization will take longer, but the gel will form correctly.
40. A typical protocol for making labeled marker is as follows. Treat pBR322 MspI digest ladder with calf intestinal phosphatase (CIP) from NEB. Phenol extract and then ethanol precipitate to remove CIP. Phosphorylate pBR322 with [γ-^{32}P]ATP with a 25 μL PNK reaction containing 1× PNK buffer, 5 μL [γ-^{32}P]ATP, 0.67 μM ATP, 20 ng CIP-treated pBR322 ladder/μL, and 1 μL T4 PNK. Incubate for 30 min at 37 °C. Inactivate PNK by incubating at 65 °C for 15 min. Dilute in 1× urea stop dye before loading on a gel. Ladder can be stored at −20 °C.
41. If signal on gel is low and a longer exposure is required, the urea-PAG can be dried to prevent diffusion of the bands within the gel. As with the native-PAG, use a gel-drying vacuum system and dry the gel for ~1 h at 80 °C.
42. Aliquoting 50 μL/tube works well for a 96-well type setup. A larger volume could be used if the thermocycler allows a larger volume.
43. The key to generating good in vitro transcription DNA template from PCR is to limit the number of cycles. Fewer cycles reduce the chance for misincorporation of nucleotides or mispriming of the primers and a higher chance of getting the desired product. This reaction should include a larger amount of template DNA than typically used in PCR to compensate for the reduced number of cycles.
44. A SpeedVac Concentrator could be used in this step if desired instead of air drying. *See* Subheading 3.2.2 for the description of SpeedVac use.
45. To make an agarose gel running box RNase-free, clean with DIH_2O and 70% ethanol. Then coat surface with Eliminase and let it dry. Use DEPC-treated MQH_2O to rinse out gel running box before use.
46. A gel running box with a gel container that is 9.5 cm wide × 10.8 cm long × 2.5 cm deep will hold ~120 mL of 0.8% agarose. A 9-well comb will generate wells that will each hold ~60 μL. The concentrated PCR product from the 5-mL

PCR will fit on this size gel. A larger gel or multiple gels may need to be run to accommodate larger scale PCRs.

47. The ethidium bromide-stained band will be visible by eye because there is so much DNA present in the gel.
48. Short wave UV will damage the DNA and should never be used if a high-quality DNA template is to be produced.
49. Dialysis tubing should be treated with solutions of EDTA and sodium bicarbonate to remove trace heavy metals and sulfides, which can damage DNA or RNA, as necessary. A general protocol is to boil dialysis tubing in MQH_2O containing 5 mM EDTA and 5 mM $NaHCO_3$ for about 15 min, then soak in MQH_2O for 15 min. Boil dialysis tubing again in MQH_2O containing 5 mM EDTA for ~15 min followed by soaking in MQH_2O for 15 min. Fully washed dialysis tubing can be stored indefinitely in 20% ethanol in MQH_2O.
50. Save some sample to run on an agarose gel to estimate the amount of DNA. This is also a good opportunity to confirm that the DNA product from the PCR and other steps is the desired product.
51. Setup of mixes can be done at any bench. Once radioactive nucleotides are to be used, move to a designated area and work behind a plexiglass shield to protect yourself from beta-decay of the ^{32}P. Samples can be quite radioactive so it is imperative to keep samples behind a shield and check your gloves frequently for contamination (especially after handling phenol samples).
52. There are many components to this assay so it is imperative to keep all the tubes organized. Adding an open tube rack to the ice bucket can help with positioning of tubes if they cannot stay organized within the ice bucket. A simple and clear labeling strategy, such as color coding, is also important for the organization.
53. The temperature for this experiment (20 °C) was optimized for H-NS filament formation and RNAP elongation to occur at the same time. The temperature at which H-NS (or other NAPs) are incubated with the DNA can be changed depending on conditions at which binding is observed. For example, elongation in the presence of H-NS filaments formed at 37 °C was also tested [3]. Depending on temperature, timed samples might need to be adjusted as RNAP will elongate faster at higher temperatures.
54. Chilling the centrifuge before this spin will greatly increase the separation of phases. If the top layer is still cloudy, repeat spin.
55. Samples can also be air dried if no SpeedVac is available. Take care to remove as much remaining ethanol with a pipet before air drying to ensure removal of all ethanol in a timely manner.

56. It is imperative to treat glass plates to prevent the gel from sticking to the plates. 10% dichlorodimethylsilane in chloroform works well to prevent the 3% native gel from sticking to plates. Additionally, clean plates prevent bubble formation during pouring and ensure easier removal of the gel from the plates after running.
57. A typical 1000 μL pipette can also be used to add the agarose plug to the bottom of the gel sandwich; however, the hot agarose can damage pipette components. The use of a positive-displacement pipette (Rainin) can prevent damage to pipettes and is recommended.
58. Air bubbles frequently appear between the glass plates and the native gel solution after pouring. Typically, these do not interfere with a successful run of the gel, but can prevent good well formation. Cleaned comb and plates should prevent formation of bubbles and additional clamps can ensure good contact between the gel and plates by pushing excess air out.
59. A heat lamp is not necessary, but is recommended to speed up the polymerization process. If no heat lamp is used, additional APS and TEMED can be added to decrease polymerization time.
60. Prerunning enables crisp band formation by removing any imbalance of ions. The wells should also be cleaned of particulates or unpolymerized acrylamide in the well by gently squirting buffer into each well with a syringe and 18-gauge needle.
61. Do not use higher temperatures to inactivate T4 PNK because the DNA will be denatured.
62. Dye may disrupt protein-DNA interactions.
63. Drying this gel is necessary since liquid from the gel will damage the phosphorimager screen. The thickness of the gel can also blur the radioactive signal. Drying time depends on thickness of the gel and gel dryer used. The time described here is for the gel and gel dryer system used in this method.
64. As with treating small plates for EMSA, take care when using dichlorodimethylsilane. The 2% dilution in chloroform is sufficient for treatment of large plates for urea-PAGE.
65. To prevent the acrylamide solution from leaking out of the gel, a good fit between spacers needs to be achieved. A small amount of 1% agarose can be added to the spacer junction to improve the seal between spacers.
66. Gels can be stored wrapped with a paper towel and cling wrap for 48 h if the gel stays wet. Dispose of the gel if it dries out.
67. The aluminum plate should not contact the buffer to avoid contacting the electrical circuit.

68. Heating samples is important for the long RNAs to be completely denatured. Work quickly when loading the gels to keep the gel from cooling too much, which will decrease the resolution of bands on the gel due to RNA secondary structure forming.
69. The gel percentages should be determined based on the DNA template being used in the experiment. For the *bgl* template described here, this two-gel system worked well to visualize RNA products from 26 to 1482 nt long.
70. If pre-exposed X-ray films are not available, Whatman filter paper can be used as support for the gel. In this case, drying the gel before exposure would be recommended.
71. This analysis technique was developed using ImageQuant v5.2 where lines can be drawn manually. This analysis can likely be done in newer versions of ImageQuant as well, though different commands may be needed to create lines on each lane and to ensure that the small pixel numbers correspond to short RNA lengths.
72. Be sure that no signal is saturated on the gel. All signals from RNA lengths must fall in the linear scale of signal to be meaningful.
73. If more than three conditions are being compared, it is useful to stack each condition along the *y*-axis. This can be done by adding a constant value (e.g., +15,000) to each rfu(z) value for the conditions that should appear higher on the pseudo-densitometry plot (*see* Fig. 11, normalized data in column AA).
74. In the best-case scenario, all lanes will run evenly on the gel meaning that all bands line up at about the same pixel value on the gel. In reality, the samples may not run completely evenly because the electrophoresis may be slower near the edges (i.e., smile) or be uneven (i.e., be wavy). To realign lanes so that peaks that represent the same RNA product line up, the pixel to nucleotide conversion can include converting "negative" pixel values to nucleotides. These additional nucleotide values can be used to shift the *x*-axis for different lanes to match peaks on the pseudo-densitometry profiles (*see* Subheading 3.5.2 and Fig. 10).
75. Another adjustment to the nucleotide conversion is done because the curve fit may result in nucleotide values reaching a minimum before the bottom of the gel giving smaller pixel values increasing nucleotide values rather than decreasing (Fig. 10). To make the smaller pixels smaller nucleotide values, paste the values from the curve fit derived as in Fig. 10 in the next column and use the drag and fill feature in excel to fill cells from the 15 values above the minimum to the 1 pixel row



AP	AQ	AR
88	229.1651	226.5588
89	228.9986	226.6288
90	228.8435	226.6988
91	228.6998	226.7687
92	228.5673	226.8387
93	228.446	226.9087
94	228.3358	226.9787
95	228.2366	227.0486
96	228.1484	227.1186
97	228.0709	227.1886
98	228.0043	227.2585
99	227.9483	227.3285
100	227.9029	227.3985
101	227.868	227.4685
102	227.8435	227.5384
103	227.8294	227.6084
104	227.8255	227.8255
105	227.8318	227.8318
106	227.8482	227.8482
107	227.8746	227.8746
108	227.9109	227.9109
109	227.9571	227.9571
110	228.013	228.013
111	228.0786	228.0786
112	228.1538	228.1538
113	228.2385	228.2385
114	228.3327	228.3327
115	228.4362	228.4362
116	228.5491	228.5491
117	228.6711	228.6711
118	228.8022	228.8022

Fig. 12 Example of adjusting nucleotide (nt) values from Excel spreadsheet screen shot. In some curve fits, the pixel-to-nt conversion will calculate a minimum nt values at a position other than pixel 1, which corresponds to the bottom of the gel. In the same curve fit provided in Fig. 10, this minimum is at pixel 104. To correct, the same formula-derived values of the nt from the curve fit were pasted from column AQ to column AR. Nt values corresponding to pixels 104 to 118 (i.e., the ~15 nt values at and next to the minimum that increase with increasing pixels) were highlighted. Use Excel's fill handle (the small box in the bottom-right of the highlighted cells) to drag these cells to row 1 at the very top of the sheet. This will adjust the nt values from pixel 1 to pixel 104, setting the corrected nt minimum at pixel 1. At this size range, fractional nt are often present

(Fig. 12, arrow). In this way, nucleotide values will only be increasing as pixel values increase.

76. Background subtraction can usually be done by creating a linear function that dictates how the background changes with gel position. This approach determines the background to be subtracted for each pixel. Occasionally, background may need to be defined by a nonlinear function.

References

1. Dillon SC, Dorman CJ (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nat Rev Microbiol 8(3):185–195. https://doi.org/10.1038/nrmicro2261

2. Browning DF, Grainger DC, Busby SJ (2010) Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. Curr Opin Microbiol 13(6):773–780. https://doi.org/10.1016/j.mib.2010.09.013

3. Kotlajich MV, Hron DR, Boudreau BA, Sun Z, Lyubchenko YL, Landick R (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. eLife 4:e04970. https://doi.org/10.7554/eLife.04970
4. Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R (2012) Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. Genes Dev 26 (23):2621–2633. https://doi.org/10.1101/gad.196741.112
5. Lim CJ, Lee SY, Kenney LJ, Yan J (2012) Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. Sci Rep 2:509. https://doi.org/10.1038/srep00509
6. Rimsky S, Zuber F, Buckle M, Buc H (2001) A molecular mechanism for the repression of transcription by the H-NS protein. Mol Microbiol 42(5):1311–1323. https://doi.org/10.1046/j.1365-2958.2001.02706.x
7. Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, Gualerzi CO, Mavathur R, Muskhelishvili G, Pon CL, Rimsky S, Stella S, Babu MM, Travers A (2007) High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. Nucleic Acids Res 35 (18):6330–6337. https://doi.org/10.1093/nar/gkm712
8. Liu Y, Chen H, Kenney LJ, Yan J (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. Genes Dev 24(4):339–344. https://doi.org/10.1101/gad.1883510
9. Arold ST, Leonard PG, Parkinson GN, Ladbury JE (2010) H-NS forms a superhelical protein scaffold for DNA condensation. Proc Natl Acad Sci U S A. 107(36):15728–15732. https://doi.org/10.1073/pnas.1006966107
10. van der Valk RA, Vreede J, Qin L, Moolenaar GF, Hofmann A, Goosen N, Dame RT (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-bridging activity. eLife 6. https://doi.org/10.7554/eLife.27369
11. Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, Hinton JC (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. PLoS Pathog 2(8):e81. https://doi.org/10.1371/journal.ppat.0020081
12. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, Fang FC (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. Science 313(5784):236–238. https://doi.org/10.1126/science.1128794
13. Haft RJF, Keating DH, Schwaegler T, Schwalbach MS, Vinokur J, Tremaine M, Peters JM, Kotlajich MV, Pohlmann EL, Ong IM, Grass JA, Kiley PJ, Landick R (2014) Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. Proc Natl Acad Sci U S A 111 (25):E2576–E2585. https://doi.org/10.1073/pnas.1401853111
14. Dole S, Nagarajavel V, Schnetz K (2004) The histone-like nucleoid structuring protein H-NS represses the *Escherichia coli* bgl operon downstream of the promoter. Mol Microbiol 52(2):589–600. https://doi.org/10.1111/j.1365-2958.2004.04001.x
15. Nagarajavel V, Madhusudan S, Dole S, Rahmouni AR, Schnetz K (2007) Repression by binding of H-NS within the transcription unit. J Biol Chem 282(32):23622–23630. https://doi.org/10.1074/jbc.M702753200
16. Gonzalez N, Wiggs J, Chamberlin MJ (1977) A simple procedure for resolution of *Escherichia coli* RNA polymerase holoenzyme from core polymerase. Arch Biochem Biophys 182:404–408
17. Nayak D, Voss M, Windgassen T, Mooney RA, Landick R (2013) Cys-pair reporters detect a constrained trigger loop in a paused RNA polymerase. Mol Cell 50(6):882–893. https://doi.org/10.1016/j.molcel.2013.05.015
18. Gribskov M, Burgess RR (1983) Overexpression and purification of the sigma subunit of *Escherichia coli* RNA polymerase. Gene 26:109–118
19. Svetlov V, Artsimovitch I (2015) Purification of bacterial RNA polymerase: tools and protocols. In: Artsimovitch I, Santangelo TJ (eds) Bacterial transcriptional control: methods and protocols. Springer, New York, NY, pp 13–29. https://doi.org/10.1007/978-1-4939-2392-2_2

Part III

In Silico Approaches

Chapter 19

Deciphering 3D Organization of Chromosomes Using Hi-C Data

Andreas Hofmann and Dieter W. Heermann

Abstract

In order to interpret data from Hi-C studies genome-wide contact probability maps need to be translated into models of functional 3D genome organization. Here, we first present an overview of computational methods to analyze contact probability maps in terms of features such as the level and shape of compartmentalization. Next, we describe approaches to modeling 3D genome organization based on Hi-C data.

Key words Hi-C, Modeling, Polymer, Simulation

1 Introduction

Following the generation of contact probability maps based on Hi-C sequencing data (*see* Chapters 1 and 2), analysis of these maps is required to extract models of 3D organization and to establish structure-function relationships. It is important to keep in mind that Hi-C contact probability maps are matrices containing the contact probabilities of all pairs of loci at a given resolution and are conventionally visualized as heat maps. We are abbreviating contact probability maps as contact maps throughout this chapter.

There are two main challenges associated with the interpretation of Hi-C contact maps. First, the information contained in contact maps reflects an ensemble average of genome conformation of a large heterogeneous population of cells. Second, contact maps contain contact probabilities reflecting the probability that any pair of genomic loci co-localizes and hence characterize the neighborhood for each genomic locus. This is a major difference with imaging methods that can measure the spatial position of genomic loci in the context of the folded genome inside the cell.

There are two distinct approaches in developing spatial models of genome organization based on Hi-C contact maps. The first approach consists of hypothesis-driven modeling, which includes

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_19,

genome folding principles, such as DNA looping, as physical principles in the framework of polymer simulations. The properties of the ensemble of simulated polymer conformations are next compared to Hi-C data, often by comparison of experimental and simulated contact maps and defined structural features (such as domains) therein. The second approach uses the contact map as input to establish 3D structural models best fitting the experimental data.

Here, we first discuss how Hi-C contact maps can be analyzed in terms of prominent structural features and can be correlated with other types of genome-wide data. Next, we review approaches to the 3D modeling of chromosomes. We conclude with a discussion of the application of such modeling approaches on bacterial genomes (Fig. 1).

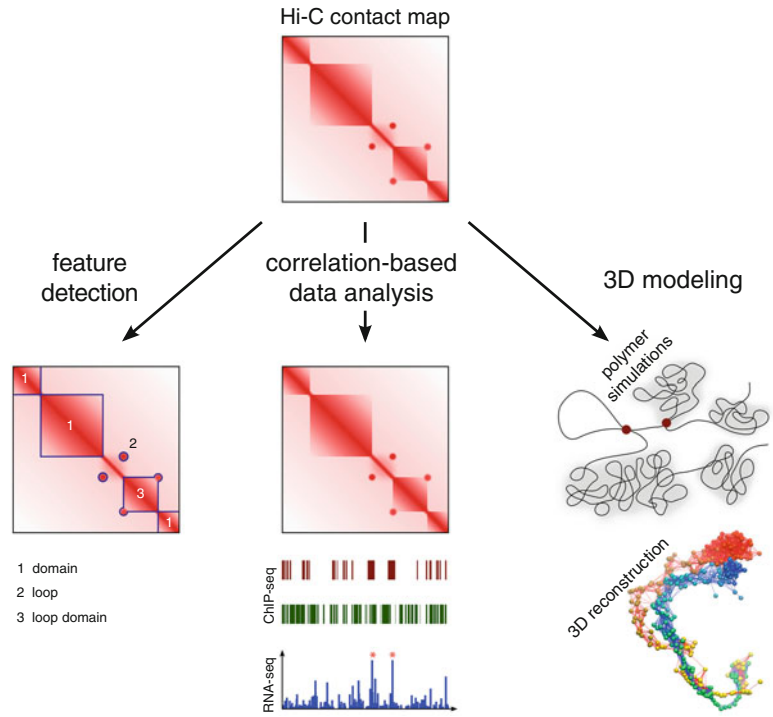


Fig. 1 Schematic overview of approaches to analyzing and modeling of Hi-C contact maps. *Left*: Feature detection methods analyze the contact map in therms of prominent features, such as domains, loops and loop domains. *Middle*: Contact maps can be analyzed alongside with other genome-wide data, such as the expression level of genes (RNA-seq) or data on proteins being attached to the genome (ChIP-seq). *Right*: Polymer simulations and 3D reconstruction methods (exemplary 3D representation adapted from [5]) aim for developing 3D models of the spatial organization of chromosomes based on Hi-C data

2 Analyzing Hi-C Contact Maps

This section describes (1) the issue of comparing Hi-C contact maps, (2) methods for detecting certain features abundantly emerging in contact maps across different species, such as compartments, domains, loops, and more complex structures, and (3) the analysis of other genomic data alongside with Hi-C data.

2.1 Comparison of Contact Maps

How can Hi-C contact maps of the same genome, but acquired under different conditions, be compared? Although qualitative differences between two contact maps can be readily detected by visual inspection, it is not straightforward to quantify differences. The underlying mathematical challenge consists in quantifying the similarity, or rather the dissimilarity, of two matrices. However, well-known similarity measures, such as the cosine similarity S_C or the Jaccard index, take vectors as input and give as output a number between 0 and 1, where 0 indicates absolute dissimilarity and 1 absolute similarity. The same is true for statistical measures of correlation, such as the Pearson correlation coefficient ρ, that is commonly used for comparing Hi-C contact matrices and yield values between $+1$ and -1, where 1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. The computation of these measures requires decomposition of the two-dimensional contact matrices into one-dimensional vectors row-by-row.

In order to illustrate this problem, two contact maps of the *Caulobacter crescentus* chromosome [1] are shown in Fig. 2. In this example, the qualitative difference is the missing compartmentalization in the vicinity of the main diagonal of the contact map of

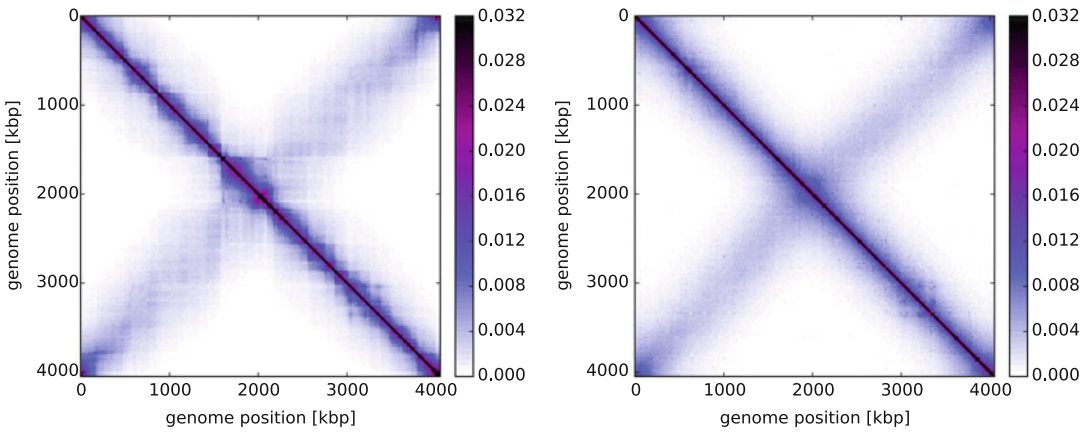


Fig. 2 Hi-C contact probability maps of the *C. crescentus* genome (left: wild-type; right: Rifampicin treated). The color code is as follows: the darker the color, the higher the contact probability. Although there is a clear qualitative difference between the two maps, quantifidcation of this difference is not straightforward

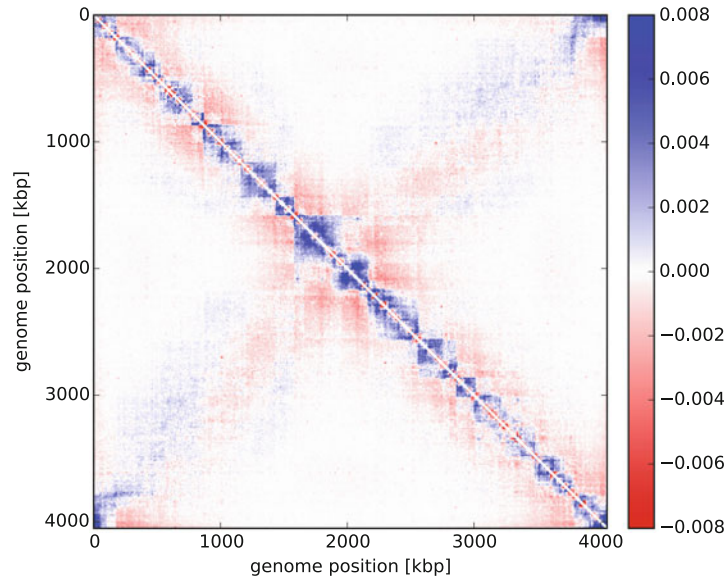


Fig. 3 Subtracting the two Hi-C contact maps of the *C. crescentus* genome. The Rifampicin treated *C. crescentus* contact matrix has been subtracted from the wild-type one and the resulting matrix is illustrated using a heat map. A comparable contact probability in both matrices is depicted in white, enhanced contact probability in the wild-type matrix in comparison with the Rifampicin treated one in blue and the contrary case in red

Rifampicin-treated *C. crescentus*. But how to quantify this difference between the two maps? Both the Pearson correlation coefficient ($\rho = 0.97$) and the cosine similarity ($S_C = 0.98$) indicate high similarity between the two maps. This is due to the inherent emphasis of these measures on comparing global features in the matrices, such as the diagonal dominance and the presence of a secondary diagonal. By element-wise subtraction of one matrix from the other and depicting the resulting matrix as a heat map (*see* Fig. 3), it is possible to quantitatively display the differences between two contact matrices. The missing compartmentalization along the main diagonal in the Rifampicin-treated *C. crescentus* contact matrix can be recognized by the blue domains along the diagonal in this heat map.

2.2 Feature Detection

Eukaryotic and bacterial chromosomes have been found to be organized in compartments. In mammalian genomes domains comprise multiple length scales such as “A and B compartments” on the scale of Mbp [2] and “Topologically Associating Domains” (TADs) on length scales ranging from 10 kbp up to 1 Mbp [3, 4]. “Chromosomal Interaction Domains” (CIDs) in bacteria occur on the same length scale as TADs and are considered equivalent [1, 5–7]. More complex structures of eukaryotic genomes that have been found using Hi-C experiments include loops and loop

domains [8]. The observation of domains at different length scales highlights the need to be able to quantitatively characterize domains. On the basis of the detected domain structure, it is possible to compare different contact maps.

There are various different methodological approaches to identifying the domain structure in Hi-C contact maps. They can be divided into two different classes depending on whether or not domains or domain boundaries are detected algorithmically. A first attempt at identification of TAD boundaries was presented by Dixon et al. [3] and is motivated by the observation that TADs are demarcated by regions that are biased in their interaction probability; the upstream domain boundary is preferentially interacting downstream while the downstream boundary is preferentially interacting upstream. This method is based on a two-step strategy. First, the 2D contact information is translated into the directionality index encoding the ratio of downstream and upstream interactions. Next, Dixon et al. argued that the directionality index can be considered an observation of a hidden directionality bias that can be determined using a hidden Markov model and which allows the segmentation into domains. As an alternative, downstream interactions can also be directly compared to upstream interactions in order to derive whether the strength of interactions is significantly stronger in one direction compared to the other. Domain boundaries correspond to positions where this preferred direction of interactions abruptly changes [1]. Lévy-Leduc et al. [9] developed a 2D model that fits a block diagonal matrix to observed contacts using maximum likelihood. In this model blocks correspond to domains. Chen et al. [10] presented a method for identifying TADs based on the interpretation of the Hi-C matrix as a weighted graph whose vertices are genomic loci and whose edge weights are contact probabilities of pairs of loci. As TADs are regions within the Hi-C matrix characterized by high internal contact probability, their identification can be translated to the problem of segmenting the graph into components with strong intra-connections and weak inter-connections. This graph partitioning is realized using spectral decomposition. Further methods use combinatorial optimization to find an optimal TAD hierarchy [11, 12].

All these methods for domain detection assume that the domains are distinct, contiguous blocks of increased contact probability. In order to detect both domains and more complex structures including loops or loop domains [8] without bias, we developed a probabilistic graphical model that makes no a priori assumptions on the domain structure [13]. Within this approach, the Hi-C contact matrix is analyzed using an Ising-like probabilistic graphical model whose spin coupling constant is proportional to each lattice point (entry in the contact matrix). This approach is also relying on the graph theoretic interpretation of Hi-C matrices and does not yield domain boundaries, but a contour separating

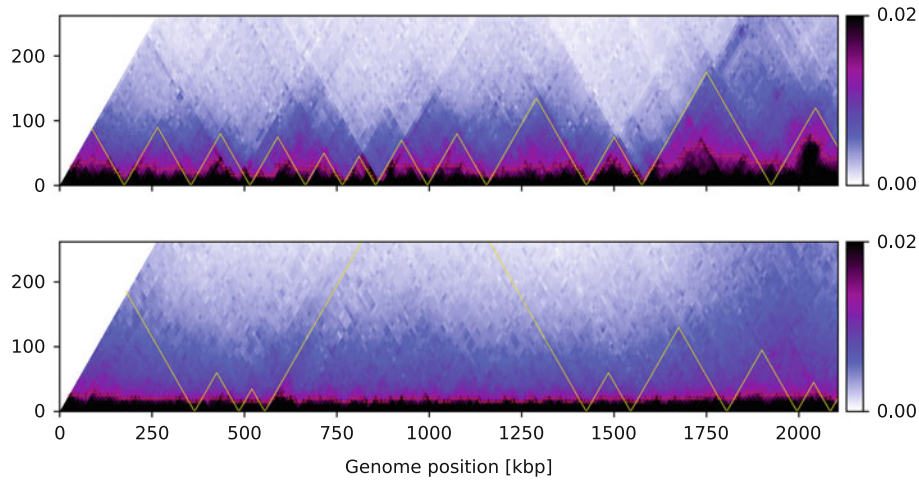


Fig. 4 Comparison of the results of the approach using the directionality index and our probabilistic graphical model. To this end, excerpts of the 45°, anti-clockwise rotated Hi-C contact maps of the *C. crescentus* chromosome shown in Fig. 2 are used. Contrary to the directionality index approach our method does not yield domain boundaries or rather a domain structure (yellow), but a contour (red) separating contact probabilities of a certain strength from the background. The results of our method clearly show the missing compartmentalization along the diagonal in the contact map of the Rifampicin-treated *C. crescentus* chromosome

contact probabilities of an adjustable strength from the background. This iso-strength contour allows identification and characterization of compartments irrespective of whether or not there are contiguous domains in the form of squares.

In order to illustrate and visually compare the results of our approach and that based on the directionality index, we used the Hi-C contact maps of *C. crescentus* shown in Fig. 2. The contiguous domain structure computed on the basis of the directionality index as well as the iso-strength contour yielded by our method permit characterization of the compartmentalization of the wild-type *C. crescentus* chromosome (*see* Fig. 4). The iso-strength contour is nearly flat for the contact map of the Rifampicin-treated *C. crescentus* chromosome using the same model parameters. This indicates that following rifampicin perturbation this chromosome is not compartmentalized; instead the captured interactions between genomic loci decrease uniformly as their genomic distance increases. The absence of a contiguous domain structure cannot be captured by methods like those based on the directionality index.

2.3 Correlation-Based Data Analysis

Hi-C contact information can be analyzed alongside with other genomic data such as the expression level of genes or data on proteins being attached to the genome. This has the advantage of being able to gain insights into the correlation between the two types of information instead of only analyzing each separately. Using such a combined analysis it was found that the mouse

genome is organized into domains of coordinately regulated enhancers and promoters that coincide with TADs [14]. First, ChIP-seq data of RNA polymerase II indicative of active promoters and H3K4me1 as a mark for enhancers were compared across different tissues and cell types. As these signals were concordantly enriched within clusters in the genome, much like TADs in contact maps of mammalian genomes, the authors also compared both types of domains and determined that they indeed overlap. Another example of such correlation-based data analysis of ChIP-seq-binding profiles and Hi-C data showed that the proteins CTCF and cohesin associate with loops that have been detected within contact maps [8]. The previously discussed domain boundaries in the contact map of the wild-type *C. crescentus* chromosome have been found to correlate with the position of highly expressed genes identified through DNA microarray analysis experiments.

We mentioned only a few exemplary studies that found interesting features of Hi-C contact maps to be correlated with other genomic data. These correlations, though not implying causality, are interesting for further specialized studies and hypothesis-driven modeling approaches since they hint at possible mechanisms underlying 3D genome organization.

3 3D Modeling

Hi-C experiments yield information that can be interpreted using computational models of chromosome organization. There are two key strategies for building such models. The first data-driven strategy, referred to as 3D reconstruction, uses the contact probabilities as summarized in the contact map to determine an optimal structural model of the data. The second strategy aims at establishing general principles of folding for organization of chromosomes using physical principles in the framework of polymer simulations. Contrary to the first strategy, Hi-C data is not used as an input for these polymer models, but rather for validation. Here, we review several methods employing either of the two strategies. For a more complete overview of 3D reconstruction methods, we refer to the review of Serra et al. [15].

3.1 3D Reconstruction

The goal of 3D reconstruction algorithms is to use the contact map as input to recapitulate the underlying 3D structure of a genome. In this approach 3C-based data is used to obtain spatial restraints for modeling the genome; 3D reconstruction is also known as restraint-based structure modeling. The basic concept for the reconstruction is simple: the closer two genomic loci are in 3D space, the higher the probability is that they interact. In technical terms, the assumption is that the Euclidean distance between two loci is inversely proportional to their contact probability. Following

this basic notion, there are two strategies for translating the contact probabilities within the contact map into a set of 3D coordinates of loci representing the genome. In the first, optimization-based, approach the total difference between pairwise distances in the hypothesized set of 3D coordinates is minimized and the corresponding distances are inferred from the observed contact probabilities. In the second, model-based, strategy, the observed contact probabilities are assumed to follow a probability distribution from which 3D structures can be inferred.

Irrespective of the underlying strategy, these methods output either a single consensus 3D structure or an ensemble of 3D structures. Both consensus and ensemble methods have advantages and disadvantages. Ensemble methods are biologically more plausible, because they reflect the fact that Hi-C data is obtained from an ensemble of conformations. However, the analysis of an inferred ensemble of 3D structures is not straightforward. One option is the characterization of the ensemble average [16], another option is to select a few structures that are representative of the diversity of the ensemble [17]. Consensus methods, in contrast, generate a single structure, which can be thought of as a visualization of the contact map and is easy to analyze. Computationally, ensemble methods are more demanding than consensus methods, because they need to sample from a very large dimensional space of candidate 3D structures.

3.1.1 Optimization-Based Methods

"ShRec3D" [18] is a method that seeks for analytically reconstructing a consensus 3D structure. It builds upon the fact that the contact matrix can be interpreted as the adjacency matrix of a weighted graph whereby the problem is reformulated in terms of embedding a graph into Euclidean space. This problem, in turn, is well known in the literature and can be solved using classical Multidimensional Scaling (MDS) [19]. Given a set of distances between the vertexes of a graph, this method returns an Euclidean set of coordinates. Therefore, the definition of distances between the vertexes of the graph representing the contact matrix is crucial within this framework. The authors chose the shortest path distance for this purpose, but did not show how other distance definitions, such as the resistance distance or connectivity-based distances, perform compared to that choice. "ChromSDE" [20] is a numerical method that jointly optimizes the 3D structure and a parameter that maps contact frequencies to spatial distances. The main difference with ShRec3D is the translation of contact frequencies to spatial distances by numerical optimization. Both methods reconstruct a consensus 3D structure. In contrast, Kalhor et al. have proposed an optimization framework that generates an ensemble of structures [16]. The idea behind this approach is to convert contact probabilities into a set of contact restraints for the

3D structures in the ensemble. However, any given contact is enforced with its contact probability, hence only in a fraction of the inferred structures in the ensemble.

3.1.2 Probabilistic Modeling Methods

Different from the optimization-based approaches, probabilistic modeling methods assign an uncertainty to the spatial distances between genomic loci. The observed contact frequency of two loci is typically assumed to follow a Poisson distribution [21, 22]. This accounts for the fact that 3C-based experiments detect contact frequencies among restriction fragments and, hence, count data. This approach is valid for non-genome-wide input data. However, these methods are not valid for Hi-C input data as these consist of contact probabilities rather than contact frequencies among genomic loci. The Markov chain Monte Carlo (MCMC)-based method "MCMC5C" [17] is an exception in this respect since it assumes a Gaussian distribution for the input contact data; therefore it can model both Hi-C contact probabilities and other 3C-based contact frequencies. In this approach DNA is modeled as a chain of beads representing the 3D structure, which is iteratively changed using random moves that can be either accepted or rejected depending on whether the new 3D structure is more probable given the data. After a sufficient number of iterations, this MCMC scheme samples 3D structures that fit the experimental contact data. By running many of those simulations in parallel, a large ensemble of structures is generated. Hu et al. proposed a probabilistic method called "BACH" [21] that models the contact data using a Poisson distribution. Contrary to MCMC5C, Monte Carlo methods are used in order to gradually refine an initial structure conformation and generate a consensus 3D structure. "PASTIS" [22] also models the contact data using a Poisson distribution. It uses maximum likelihood estimation of the model parameters for reconstructing the 3D structure with the highest likelihood given the observed contact data.

3.2 Polymer Simulations

Polymer models incorporating known or hypothesized physical or biological principles can also be used to model chromosomes. Contrary to 3D reconstruction methods, such models do not infer conformations using Hi-C data, but rather use such data to test whether generated ensembles of 3D structures agree with contact maps or key features thereof.

A polymer can be described using various properties like its average end-to-end distance or its radius of gyration. Here, we focus on two biologically relevant quantities that can be compared to experiments, such as Hi-C or microscopy imaging: (1) the contact probability between two loci as a function of their genomic distance which can be deduced from the Hi-C data; (2) the mean squared distance (MSD) of two loci as a function of their genomic distance, a quantity that can be measured, for example, by

fluorescence in situ hybridization (FiSH) experiments. Both quantities are averaged over the conformational ensemble in any polymer model and over a population of cells in a Hi-C experiment. Simple polymer models include the random coil and the self-avoiding chain. The random coil is the simplest model; it is characterized by non-interacting monomers. Self-avoiding chains exhibit excluded volume interactions leading to an increased effective volume compared to the random coil. The worm-like chain or Kratky-Porod model introduces an intrinsic stiffness by associating a bending of the chain with an energy cost. Hence, it can describe semi-flexible polymers, such as double-stranded DNA. The fractal globule model [23] describes a compact polymer state that emerges during polymer condensation as a result of topological constraints. It has been reported to agree with the initial Hi-C data of the human genome [2] since it shows the same scaling behavior of the contact probability as a function of the genomic distance at a scale of ~ 1–10 Mbp [24]. However, it does not explain findings from FiSH experiments that display a leveling-off in the MSD for genomic separations above 10 Mbp [25]. In the dynamic loop (DL) model [26] the chromosomal fiber is represented as a self-avoiding chain allowed to form probabilistic intra-polymer cross-links between non-adjacent monomers. As a consequence, loops of different size are formed. The main model parameter is the looping probability, a measure for the probability that a loop is formed between two non-adjacent monomers. The DL model explains both the scaling behavior of the contact probability and the leveling-off of the MSD on the basis of the dynamic formation and dissolution of loops. The "strings and binders switch" (SBS) model [27], which assumes diffusible factors (binders) being responsible for loop formation by linking two monomers of the polymer, is a special case of the DL model that implicitly incorporates the properties of such binders in the looping probability parameter. The polymer fiber is also modeled as a self-avoiding chain and the binding molecules are represented by Brownian particles with a certain concentration. The loop extrusion model [28] proposes that loop-extruding factors form increasingly larger loops, which are stalled by boundary elements, such as bound proteins at domain boundaries. Different from the DL and the SBS model [29, 30], it also explains the formation of domains [31, 32].

4 Summary

Hi-C contact maps either can be analyzed in terms of feature detection or can be used as a starting point for developing 3D models of the spatial organization of chromosomes. In the specific case of feature detection, it would be useful to have methods that

allow discrimination of different possible patterns instead of detecting only a specific one. Existing methods, such as directionality index-based approaches, have shown the existence of domains in eukaryotic as well as bacterial chromosomes [1, 5–7], but fail to identify loops or loop domains. There are two different ways of modeling 3D organization of chromosomes: 3D reconstruction and polymer modeling. While the first generates the most likely 3D structure given the contact data amenable to visual inspection, polymer modeling supports clarifying hypotheses of chromosomal organization deduced from contact maps. The Hi-C studies of the two bacterial chromosomes *C. crescentus* [1] and *B. subtilis* [5] exemplify the difference between hypothesis-driven modeling and 3D reconstruction. Using ShRec3D, a consensus 3D structure for the *B. subtilis* chromosome is reconstructed and then used to illustrate the point that the chromosome folds into a helicoidal shape and is organized into domains. In contrast, Le et al. deduced from their gathered Hi-C and microarray data the hypothesis that the *C. crescentus* chromosome consists of domains comprised of supercoiled DNA plectonemes and boundaries being transcription-induced. Based on this hypothesis, they performed simulations of a bottle-brush polymer with linear boundary elements and compared the simulated with the experimental contact maps.

Acknowledgment

We would like to thank Remus T. Dame, Frédéric Crémazy and Fatema Zahra Rashid for the stimulating and fruitful discussions. This work was supported by a grant from the International Human Frontier Science Program Organization (RGP0014/2014).

References

1. Le TBK, Imakaev MV, Mirny LA, Laub MT (2013) High-resolution mapping of the spatial Organization of a Bacterial Chromosome. Science 342(6159):731–734. https://doi.org/10.1126/science.1242059
2. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326 (5950):289–293. https://doi.org/10.1126/science.1181369
3. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398):376–380. https://doi.org/10.1038/nature11082
4. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation Centre. Nature 485 (7398):381–385. https://doi.org/10.1038/nature11049
5. Marbouty M, Le Gall A, Cattoni DI, Cournac A, Koh A, Fiche J-B, Mozziconacci J, Murray H, Koszul R, Nollmann M (2015) Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by hi-C and

super-resolution imaging. Mol Cell 59 (4):588–602. https://doi.org/10.1016/j.molcel.2015.07.020

6. Wang X, Le TBK, Lajoie BR, Dekker J, Laub MT, Rudner DZ (2015) Condensin promotes the juxtaposition of DNA flanking its loading site in Bacillus subtilis. Genes Dev 29 (15):1661–1675. https://doi.org/10.1101/gad.265876.115
7. Trussart M, Yus E, Martinez S, Baù D, Tahara YO, Pengo T, Widjaja M, Kretschmer S, Swoger J, Djordjevic S, Turnbull L, Whitchurch C, Miyata M, Marti-Renom MA, Lluch-Senar M, Serrano L (2017) Defined chromosome structure in the genome-reduced bacterium mycoplasma pneumoniae. Nat Commun 8:14665. https://doi.org/10.1038/ncomms14665
8. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at Kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021
9. Lévy-Leduc C, Delattre M, Mary-Huard T, Robin S (2014) Two-dimensional segmentation for analyzing hi-C data. Bioinformatics 30(17):i386–i392. https://doi.org/10.1093/bioinformatics/btu443
10. Chen J, Hero AO, Rajapakse I (2016) Spectral identification of topological domains. Bioinformatics 32(14):2151–2158. https://doi.org/10.1093/bioinformatics/btw221
11. Filippova D, Patro R, Duggal G, Kingsford C (2014) Identification of alternative topological domains in chromatin. Algorithms Mol Biol 9:14. https://doi.org/10.1186/1748-7188-9-14
12. Weinreb C, Raphael BJ (2016) Identification of hierarchical chromatin domains. Bioinformatics 32(11):1601–1609. https://doi.org/10.1093/bioinformatics/btv485
13. Heermann DW, Hofmann A, Weber E (2017) Domain boundary detection in Hi-C maps: a probabilistic graphical model approach. ArXiv170303656 Q-Bio
14. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B (2012) A map of the cis-regulatory sequences in the mouse genome. Nature 488(7409):116–120. https://doi.org/10.1038/nature11243
15. Serra F, Di Stefano M, Spill YG, Cuartero Y, Goodstadt M, Baù D, Marti-Renom MA (2015) Restraint-based three-dimensional modeling of genomes and genomic domains. FEBS Lett 589(20):2987–2995. https://doi.org/10.1016/j.febslet.2015.05.012
16. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L (2012) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat Biotechnol 30(1):90–98. https://doi.org/10.1038/nbt.2057
17. Rousseau M, Fraser J, Ferraiuolo MA, Dostie J, Blanchette M (2011) Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. BMC Bioinformatics 12 (1):414. https://doi.org/10.1186/1471-2105-12-414
18. Lesne A, Riposo J, Roger P, Cournac A, Mozziconacci J (2014) 3D genome reconstruction from chromosomal contacts. Nat Methods 11 (11):1141–1143. https://doi.org/10.1038/nmeth.3104
19. Young FW (2013) Multidimensional scaling: history, theory, and applications. Psychology Press
20. Zhang Z, Li G, Toh K-C, Sung W-K (2013) 3D chromosome modeling with semi-definite programming and hi-C data. J Comput Biol 20 (11):831–846. https://doi.org/10.1089/cmb.2013.0076
21. Hu M, Deng K, Qin Z, Dixon J, Selvaraj S, Fang J, Ren B, Liu JS (2013) Bayesian inference of spatial organizations of chromosomes. PLoS Comput Biol 9(1):e1002893. https://doi.org/10.1371/journal.pcbi.1002893
22. Varoquaux N, Ay F, Noble WS, Vert J-P (2014) A statistical approach for inferring the 3D structure of the genome. Bioinformatics 30 (12):i26–i33. https://doi.org/10.1093/bioinformatics/btu268
23. Grosberg AY, Nechaev SK, Shakhnovich EI (1988) The role of topological constraints in the kinetics of collapse of macromolecules. J Phys 49(12):2095–2100. https://doi.org/10.1051/jphys:0198800490120209500
24. Mirny LA (2011) The fractal globule as a model of chromatin architecture in the cell. Chromosom Res 19(1):37–51. https://doi.org/10.1007/s10577-010-9177-0
25. Mateos-Langerak J, Bohn M, de LW, Giromus O, Manders EMM, Verschure PJ, Indemans MHG, Gierman HJ, Heermann DW, van DR, Goetze S (2009) Spatially confined folding of chromatin in the interphase nucleus. Proc Natl Acad Sci 106 (10):3812–3817. https://doi.org/10.1073/pnas.0809501106
26. Bohn M, Heermann DW (2010) Diffusion-driven looping provides a consistent framework

for chromatin organization. PLoS One 5(8): e12218. https://doi.org/10.1371/journal.pone.0012218

27. Barbieri M, Chotalia M, Fraser J, Lavitas L-M, Dostie J, Pombo A, Nicodemi M (2012) Complexity of chromatin folding is captured by the strings and binders switch model. Proc Natl Acad Sci 109(40):16173–16178. https://doi.org/10.1073/pnas.1204799109
28. Alipour E, Marko JF (2012) Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res 40 (22):11202–11212. https://doi.org/10.1093/nar/gks925
29. Hofmann A, Heermann DW (2015) The role of loops on the order of eukaryotes and prokaryotes. FEBS Lett 589(20):2958–2965. https://doi.org/10.1016/j.febslet.2015.04.021
30. Nicodemi M, Pombo A (2014) Models of chromosome structure. Curr Opin Cell Biol 28:90–95. https://doi.org/10.1016/j.ceb.2014.04.004
31. Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci 112(47): E6456–E6465. https://doi.org/10.1073/pnas.1518552112
32. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA (2016) Formation of chromosomal domains by loop extrusion. Cell Rep 15(9):2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085

Chapter 20

Molecular Dynamics Simulation of a Feather-Boa Model of a Bacterial Chromosome

Debasish Chaudhuri and Bela M. Mulder

Abstract

The chromosome of a bacterium consists of a mega-base pair long circular DNA, which self-organizes within the micron-sized bacterial cell volume, compacting itself by three orders of magnitude. Unlike in eukaryotes, it lacks a nuclear membrane, and freely floats in the cytosol confined by the cell membrane. It is believed that strong confinement, cross-linking by associated proteins, and molecular crowding all contribute to determine chromosome size and morphology. Modeling the chromosome simply as a circular polymer decorated with closed side-loops in a cylindrical confining volume, has been shown to already recapture some of the salient properties observed experimentally. Here, we describe how a computer simulation can be set up to study structure and dynamics of bacterial chromosomes using this model.

Key words Chromosome modeling, Bottle brush polymer, Confinement, Crowders, Langevin heat bath

1 Introduction

A typical *E. coli* cell has a diameter of about 0.8 μm and lengths between 2 and 4 μm. The *E. coli* chromosome consists of a 4.7 Mbp of circular DNA which results in a length of 1.6 mm, and thousands of crosslinking proteins that belong to either of two classes: (1) nucleoid-associated proteins (NAPs), or (2) ATP-dependent active structural modification of chromosome (SMC) proteins. The circular DNA is supercoiled due to a globally maintained, enzymatically controlled, under-twist. This supercoiling brings distant parts of the DNA spatially close to each other. The combined effect of supercoiling, protein-mediated cross-linking, and electrostatic zippering can cause formation of loops in the chromatin [1, 2]. This process may be further enhanced by osmotic pressure due to smaller molecules in the cytosol generating further compaction, as well as cell wall confinement. The chromosome self organizes into a compact structure called nucleoid, which "floats" in the cytosol in the absence of any nuclear membrane-like confinement.

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837,
https://doi.org/10.1007/978-1-4939-8675-0_20,

The dynamics of loop formation and their distribution along the DNA chain is expected to be complex, involving poly-dispersity in loop sizes and topological entanglements. Indeed, electron microscopy of lysed *E. coli* bacteria showed an approximate bell-shaped distribution of loop sizes, with a mean and maximum close to 10–12 kbp [3]. The local contact map, which measures spatial contacts of different parts on DNA contour, has been explored through chromosome conformation capture experiments [4].

Live-cell imaging revealed that the bacterial nucleoid in rod-shaped bacteria like *E. coli* [5] and *B. subtilis* [6] has a helical shape with ∼1 μm pitch, such that 3–4 turns are seen in wild-type bacteria. Thus a question arises as to what extent such configurations can indeed emerge as equilibrium configurations of real polymers with short-range interactions and, hence, provide clues to potential purely physical mechanisms. Using a minimal structured polymer model of chromosome and rod-shaped confinement, we showed previously that entropic effects alone are sufficient to yield such helical shapes [7]. This polymer-based modeling suggests that the helicoidal shape in *E. coli* and *B. subtilis* is directly connected to the shape of the confinement provided by bacterial cells.

The effect of confinement on the shape of self-avoiding polymers has long been appreciated theoretically [8, 9]. Scaling arguments suggest that as the confining dimension D becomes smaller than the radius of gyration of a polymer Rg, the polymer rather than behaving as a single "blob" of size Rg breaks up into connected blobs having size D, indicating a crossover to a lower dimensional behavior. The amount of compaction required can be appreciated by considering that a 1.6 mm long linear chain (here dsDNA with a typical persistence length of 50 nm) has a radius of gyration of 33 μm, which is much larger than the diameter of a micron-sized cell. The last decade has seen increasing interest in understanding the configuration and dynamics of the *E. coli* chromosome [5] and its implications for chromosome segregation prior to division [10–12]. Experiments in which the *E. coli* chromosome was mechanically compacted revealed its soft nature, and also changes in entropic force due to molecular crowding were shown to control its size [13]. The entropic force caused by the restrictions that confinement imposes on polymer conformation has been suggested as a possible explanation of the ubiquitous helical morphology of chromosomes in rod-shaped bacteria [7], and as a possible mechanism for chromosome segregation [11].

Here, we describe the feather-boa model of a self-avoiding polymer decorated with a cloud of loops, to model the complex local structure of the chromosome [7]. The same model has previously been considered by Reiss et al. [14], in the absence of confinement. This type of polymer model is similar to so-called "bottle-brush" polymers, extensively studied by Binder and coworkers [15–17]. The latter work has shown that such polymers

exhibit a local resistance to bending due to the entropic repulsion between the side chains. The closed side-loops in the feather-boa model lead to enhanced effective repulsion compared to the linear side-chains of the bottle-brush. This stiffening, combined with intra-chain packing effects within the cylindrical confinement, leads to the spontaneous formation of helical configurations [7]. Taking into account the crowded environment of the cytosol is expected to lead to depletion effects, with a resultant osmotic pressure further stabilizing the helix.

2 Materials

2.1 Model Specification

In the feather-boa model of the bacterial chromosome we consider a polymer composed of a ring-like backbone chain of length $l_b = n_b\sigma$, with n_b beads and bond length σ. To this backbone, side loops of length $l_s = n_s\sigma$ are joined at equal spacing of $n_g\sigma$. The main-chain and the side-loops are physically the same. This structured polymer is assumed to be made of self-avoiding harmonic chains. In this model, consecutive beads interact through a harmonic potential.

$$V_b = (k_s/2)(\mathbf{d}_i - \sigma\mathbf{u}_i)^2 \tag{1}$$

where the bond vector is $\mathbf{d}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$, $\mathbf{r}_i$ is the position of ith bead, σ the equilibrium bond-length, and $\mathbf{u}_i = \mathbf{d}_i/|\mathbf{d}_i|$ is the local tangent vector to the chain. Non-bonded beads repel each other through the Weeks-Chandler-Andersen (WCA) potential [18].

$$V_{WCA} = 4\varepsilon\left(\sigma/r_{ij}\right)^{12} - \left(\sigma/r_{ij}\right)^{6} + 1/4, \tag{2}$$

If, the inter-monomer separation $r_{ij} < 2^{1/6}\sigma$, else $V_{WCA} = 0$. Here ε and σ set the energy and length scale of the system respectively. We use $k_s = 100\varepsilon$. The interaction of all beads with the confining walls is modeled through,

$$V_{wall} = 2\pi\varepsilon\left[(2/5)\left(\sigma/r_{iw}\right)^{10} - (\sigma/r_{iw})^{4} + 3/5\right] \tag{3}$$

if the distance of the ith monomer from a wall $r_{iw} < \sigma$ and $V_{wall} = 0$ otherwise. Using this model, we perform Molecular Dynamics simulations in the presence of a Langevin heat bath (*see* **Note 3**).

The cytosolic environment composed of much smaller molecules, apart from providing a *heat bath*, is expected to generate an entropic depletion effect, further compressing the polymer. To capture this effect, one can model the cytosol as an *ideal gas* of crowders that do not interact between themselves, but repel chromosome segments [19]. The repulsion can be modeled via the WCA potential of Eq. 2.

2.2 Implementation

We implemented parallel program simulations using MPI on a Linux cluster of 2.6 GHz Intel Xeon E5 series processors. The simulations were performed using ESPResSO molecular dynamics package that uses c/c++ routines, in the presence of Langevin thermostat. Depending on the number of particles and size of the simulation box, peak performance was achieved when we used between 20 and 30 cores of processors. The typical runtime varies from 20 to 30 h. For example, a simulation of a 40,000 beads chain in a cylinder of length 150σ and diameter 25σ takes 20 h runtime on a 24 core machine.

3 Methods

This system can be simulated employing a velocity-Verlet molecular dynamics algorithm in the presence of a Langevin thermostat fixing the temperature at $k_BT = 1$ which sets the energy scale. The simulation may be performed by either using a self-written program or as implemented in a molecular dynamics package like ESPResSo [20].

3.1 Molecular Dynamics with Langevin Heat Bath

This is a discretized scheme of integrating underdamped Langevin equations of motion. The equation of motion of a meso-scale i-th particle of unit mass is given by

$$d^2\mathbf{r}_i/dt^2 = \mathbf{f}(r_i) - \gamma \mathbf{v}_i + \eta_i, \tag{4}$$

where $-\gamma\mathbf{v}_i$ denotes the viscous dissipative force due to a free flowing fluid and $\boldsymbol{\eta}_i(t)$ is the associated Gaussian distributed stochastic force due to incessant molecular motion of fluid molecules. This model assumes a length scale and time scale separation between the solvent and solute molecules, with the solute being bigger and slower. Here $\mathbf{r}_i$ (t) is the position vector of the i-th particle, and $\mathbf{f}(\mathbf{r}_i)$ denotes all possible non-stochastic forces acting on a solvent particle, due to other solvents and confinement. The fluctuation dissipation theorem requires the following relations between the deterministic and stochastic components of force due to the solvent, $\langle\eta_\alpha(t)\rangle = 0$ and $\langle\eta_\alpha(t)\eta_\beta(t')\rangle = 2\ \gamma\ \delta_{\alpha,\beta}\ k_BT\ \delta(t - t')$ where indices α, β may denote components of a vector, or particle index.

The integration scheme is based on the velocity-Verlet algorithm for performing molecular dynamics (MD). This is a symplectic scheme that conserves time-reversal symmetry allowing a Hamiltonian system to follow a shadow Hamiltonian conserving the overall phase space volume.

The velocity-Verlet scheme for MD simulation to integrate Newtonian dynamics $d^2\mathbf{r}_i/dt^2 = \mathbf{f}(r_i)$ is given by the following two steps:

$$\mathbf{r}_i(\mathrm{t}+\delta\mathrm{t}) = \mathbf{r}_i(\mathrm{t}) + \mathbf{v}_i(\mathrm{t}).\delta t + \frac{\delta t^2}{2\mathrm{m}}\mathbf{f}_i(\mathrm{t})$$

$$\mathbf{v}_i(\mathrm{t}+\delta t) = \mathbf{v}_i(\mathrm{t}) + \frac{\delta t}{2m}(\mathbf{f}_\mathrm{i}(\mathrm{t}) + \mathbf{f}_i(\mathrm{t}+\delta t)). \tag{5}$$

To integrate the dynamics in the presence of a Langevin heat bath, one needs to replace the Newtonian force $\mathbf{f}_i$ by $\{\mathbf{f}_i(\mathrm{t}) - \gamma\mathbf{v}_i(\mathrm{t}) + \eta_i(\mathrm{t})\}$ incorporating the force due to heat bath $\{-\gamma\mathbf{v}_i(\mathrm{t}) + \eta_i(\mathrm{t})\}$. Thus the above steps of velocity-Verlet change to

$$\begin{aligned}
&\mathbf{r}_i(\mathrm{t}+\delta\mathrm{t}) = \mathbf{r}_i(\mathrm{t}) + \mathbf{v}_i(\mathrm{t}).\delta\mathrm{t} + \frac{\delta\mathrm{t}^2}{2m}\Big(\mathbf{f}_i(\mathrm{t}) - \gamma\mathbf{v}_i(\mathrm{t}) + \eta_i(\mathrm{t})\Big)\\
&\mathbf{v}_i(\mathrm{t}+\delta\mathrm{t}) = \mathbf{v}_i(\mathrm{t}) + \frac{\delta\mathrm{t}}{2m}\{\mathbf{f}_i(\mathrm{t}) - \gamma\mathbf{v}_i(\mathrm{t}) + \eta_i(\mathrm{t}) + \mathbf{f}_i(\mathrm{t}+\delta\mathrm{t}) - \gamma\mathbf{v}_i(\mathrm{t}+\delta\mathrm{t}) + \eta_i(\mathrm{t}+\delta\mathrm{t})\}.\\
&= \left(1 - \frac{\gamma\delta\mathrm{t}}{2m}\right)\mathbf{v}_i(\mathrm{t}) + \frac{\delta\mathrm{t}}{2m}\{\mathbf{f}_i(\mathrm{t}) + \eta_i(\mathrm{t}) + \mathrm{f}_i(\mathrm{t}+\delta\mathrm{t}) + \eta_i(\mathrm{t}+\delta\mathrm{t})\} - \frac{\gamma\delta\mathrm{t}}{2m}\mathbf{v}_i(\mathrm{t}+\delta\mathrm{t}).
\end{aligned} \tag{6}$$

Taking all the $\mathbf{v}_i(\mathrm{t}+\delta\mathrm{t})$ terms together, the velocity update can be written as

$$\begin{aligned}
\mathbf{v}_i(\mathrm{t}+\delta t) = &\frac{2m-\gamma\delta t}{2m+\gamma\delta t}\mathbf{v}_i(\mathrm{t})\\
&+ \frac{\delta t}{2m+\gamma\delta t}[\mathbf{f}_\mathrm{i}(\mathrm{t}) + \eta_i(\mathrm{t}) + \mathbf{f}_i(\mathrm{t}+\delta t) + \eta_i(\mathrm{t}+\delta t)]
\end{aligned} \tag{7}$$

Thus the simulation algorithm is as follows.

1. Knowing $\mathbf{r}_i(\mathrm{t})$ one can calculate the interaction potential that gives forces $\mathbf{f}_i(\mathrm{t})$ felt by the ith particle. Store $\mathbf{F}_i(\mathrm{t}) = \mathbf{f}_i(\mathrm{t}) + \eta_i(\mathrm{t})$.
2. Calculate $\mathbf{r}_i(\mathrm{t}+\delta\mathrm{t})$, using $\mathbf{f}_i(\mathrm{t})$ and $\mathbf{v}_i(\mathrm{t})$:

$$\mathbf{r}_i(\mathrm{t}+\delta t) = \mathbf{r}_i(\mathrm{t}) + \mathbf{v}_i(\mathrm{t}).\delta t + \frac{\delta t^2}{2m}(\mathbf{F}_i(\mathrm{t}) - \gamma\mathbf{v}_i(\mathrm{t})). \tag{8}$$

3. Half Update Velocity:

$$\mathbf{v}_\mathbf{i}^*(\mathrm{t}+\delta t) = \frac{2m-\gamma\delta t}{2m+\gamma\delta t}\mathbf{v}_\mathrm{i}(\mathrm{t}) + \frac{\delta t}{2m+\gamma\delta t}\mathbf{F}_\mathrm{i}(\mathrm{t}). \tag{9}$$

4. Calculate new force $\mathbf{f}_i(\mathrm{t}+\delta\mathrm{t})$ depending on the new position $\mathbf{r}_i(\mathrm{t}+\delta\mathrm{t})$. Store $\mathbf{F}_i(\mathrm{t}+\delta\mathrm{t}) = \mathbf{f}_i(\mathrm{t}+\delta\mathrm{t}) + \eta_i(\mathrm{t}+\delta\mathrm{t})$. Notice that, in calculating forces for each i at $\mathrm{t}+\delta\mathrm{t}$, one needs $\mathbf{r}_i(\mathrm{t})$ for all i.
5. Full update velocity using this new force:

$$\mathbf{v}_i(\mathrm{t}+\delta t) = \mathbf{v}_\mathbf{i}^*(\mathrm{t}+\delta t) + \frac{\delta t}{2m+\gamma\delta t}\mathbf{F}_\mathrm{i}(\mathrm{t}+\delta t). \tag{10}$$

The total force on each polymer bead can be derived from the bead-spring potential connecting consecutive monomers, the WCA repulsion capturing self-avoidance, and the repulsion from confining walls. For example, the harmonic interaction between nearest neighbors is

$$V_\mathrm{s} = \frac{\mathrm{k_s}}{2}\sum_{i=1}^{N-1}[\boldsymbol{d}_i - \sigma\boldsymbol{u}_i]^2, \tag{11}$$

where σ is the equilibrium bond-length. The force acting on each bead, due to the harmonic interaction with nearest neighbors is

$$\mathbf{f}_\mathbf{i}^\mathrm{s} = -\mathrm{k_s}(\mathbf{d}_\mathrm{i} - \sigma\mathbf{u}_i) + \mathrm{k_s}(\mathbf{d}_{i-1} - \sigma\mathbf{u}_{i-1}). \tag{12}$$

For a N-bead polymer, the first term survives if $i < \mathrm{N}$ and the second term survives when $i > 1$. Similarly, all other forces may be determined from the corresponding interaction potentials indicated in the previous section.

The stochastic forces can be simulated using the following arguments. Integration of the correlation of the stochastic force.

$$\int^{\delta t} \mathrm{d}t \int^{\delta t} dt' \langle \eta_\alpha(t)\eta_\beta(t')\rangle = 2\gamma\delta_{\alpha,\beta}k_B T\delta t$$

The noise $\eta_\alpha(\mathrm{t})$ obeys Gaussian statistics. Thus one may use the Box-Muller algorithm to generate a Gaussian random number θ_α with unit deviate [21] such that $\langle\theta_\alpha\theta_\beta\rangle = \delta_{\alpha,\beta}$ to obtain the stochastic force at a given time step, $\eta_\alpha = (2\ \gamma\mathrm{k_B T}/\delta\mathrm{t})^{1/2}\ \theta_\alpha$.

3.2 Program Implementation Using ESPResSo

In simulating the feather-boa polymer in the ESPResSo package, the first step is to create a polymer configuration with a backbone attached to equally spaced side-loops using elementary geometry (*see* Fig. 1). For an open backbone the easiest initial condition is to use a linear configuration. To model the circular chromosome of bacteria one may choose a circular initial configuration of the backbone. Note that the side loops are free to rotate around the main chain, and physically the beads of the backbone and side loops are all the same. Only to distinguish side loops from the backbone we used two different colors in Fig. 1. In performing the integration we used $\delta\mathrm{t} = 0.01$.

To implement the simulation, carry out the following steps.

1. Set the initial configuration of the feather-boa chain, and initialize velocities obeying the Maxwell-Boltzmann distribution with $\mathrm{k_B T} = 1$, using a Gaussian random number generator.

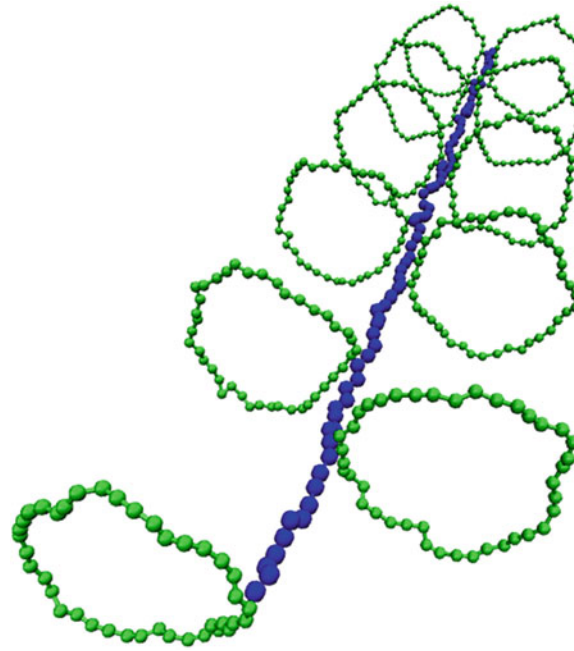


Fig. 1 (Color online) A segment of the feather-boa polymer is shown. The backbone is denoted by blue and the side loops by green beads. Physically all the beads are identical. Consecutive beads are connected by a spring of equilibrium length σ, and non-bonded beads repel each other by the WCA potential. The loops are free to rotate around the backbone

2. Use a large cylindrical confining geometry that is big enough to encapsulate the initial configuration of the chain. Use cylindrical sidewalls and planar (or sphero-cylindrical) caps at the ends of the cylinder. The repulsive potential from the walls ensures that the chain will stay within the confinement.
3. Equilibrate the system. Ensure configurational equilibration by following eigenvalues of the radius of gyration tensor, and components of the end-to-end separation vector. At equilibrium these values will saturate and fluctuate around the equilibrium mean.
4. Slowly reduce the confining volume toward the target geometry of a much smaller cylinder. The target has to be set such that the diameter and length of the cylinder are much smaller than the radii of gyration of the feather-boa chain in bulk (*see* **Note 7**). The volume reduction has to be done in small steps with a scaling factor s_f, infinitesimally smaller than 1, taking care that the positions of all the monomers are also scaled by this factor, i.e., $\mathbf{r}_i \rightarrow s_f \mathbf{r}_i$, $L \rightarrow s_f L$ and $D \rightarrow s_f D$.

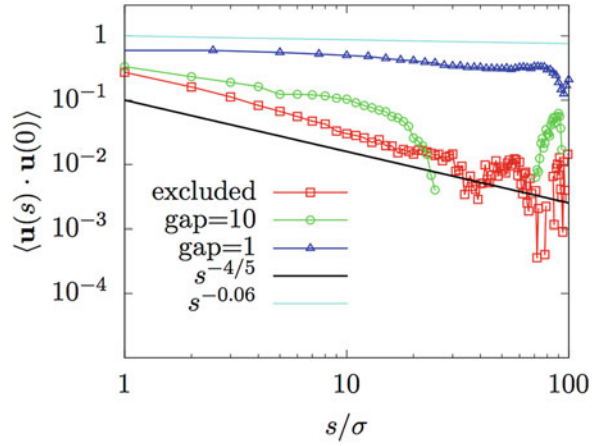


Fig. 2 (Color online) In the absence of confinement: The tangent-tangent correlation $\langle \mathbf{u}(0).\mathbf{u}(s) \rangle$ for excluded volume chain, chain associated with side loops of size $l_s = 21\ \sigma$ with every 10-th bead on backbone (gap = 10), and that with every bead of backbone (gap = 1). As expected, the excluded volume polymer shows a power-law decay $s^{-4/5}$ consistent with growth of end to end distance $R(s) = \langle r(s)^2 \rangle \sim s^b$ with $b = 3/5$. With increasing density of side-loops the correlation decays more slowly, finally saturating to $s^{-\alpha}$ with $\alpha = 0.06$ (*see* **Note 12**)

5. After rescaling, re-equilibrate the system by returning to **step 3**. **Steps 3** and **4** have to be repeated until the target volume and shape of the confinement have been obtained.
6. Once the target volume and shape of the confinement is achieved, equilibrate the system once more by performing **step 3**, before starting to accumulate data.

3.3 Model Verification: Emergent Stiffness

The side-loops in the feather-boa polymer lead to an emergent bending stiffness that can be quantified using the tangent-tangent correlation function. This is easily seen from a comparison of a feather-boa chain with a simple self-avoiding polymer. Fig. 2 shows that the simulation results of self-avoiding chains indeed obey the Flory prediction $\langle \mathbf{u}(s).\mathbf{u}(0) \rangle \sim s^{-\alpha}$ with $\alpha = 4/5$. However, in the presence of side-loops the value of exponent α decreases with increasing side-loop density, reflecting the increase in bending rigidity (*see* **Note 12**).

4 Notes

1. Note that choice of the time step $\delta t = 0.01$ used in our simulations is related to the fastest modes in the system, i.e., controlled by the spring constant $k_s = 100 k_B T$ used in our simulations. The harmonic mode sets a frequency $\sim k_s^{1/2}$ and

thus if one decides to use a larger k_s, δt has to be reduced as $k_s^{-1/2}$ accordingly.

2. The value of k_s is chosen such that the bond length fluctuations remain less than 5%.
3. The reason for modeling chromosomal environment by a Langevin thermostat is as follows. The bacterial cell is composed of highly dense and poly-dispersed molecular components, containing genetic material like the nucleoid, plasmids, and cytoskeletal elements like actin and tubulin homologues, as well as molecular machines like RNA polymerase and ribosomes. As is shown in recent experiments, reduction of metabolic activity brings the cell close to a glassy state [22]. Thus, it is important to note that the system we are considering remains in a fluid-like state due to strong metabolic activity. This allows us to model the environment of the chromosome, to first approximation, as a freely flowing fluid with its fluctuations captured by an effective temperature.
4. In these simulations, special care needs to be taken so that no particle is left outside the confinement, during the initialization, or the scaling to reduce the initial confining volume. Moreover, at the latter step two particles must not come too close to each other or to the wall, as that may lead to diverging repulsive force. This is the reason why the scaling factor s_f needs to be taken as close to unity as possible. In our experience reduction of volume by using a scale factor $s_f = 0.99$ along with relaxing the configurations by equilibrating over $\tau_r = 10^4$ MD steps helped in avoiding any possible overlap and bond breaking.
5. It is advisable to measure the actual relaxation time for a given degree of confinement before deciding τ_r and setting up long runs toward data accumulation and analysis.
6. After each rescaling, one may first use force capping over several steps before starting equilibrium relaxation. In force capping steps, simulation is run with the repulsive interaction implemented with a preset maximum capping value to remove possible overlaps that might have generated during rescaling. The cap is lifted slowly as the overlap gets removed. In the ESPResSo package [20] this is implemented by using *ljforcecap*.
7. The chromosome length is 10^3 times longer than the typical bacterial cell. The effect of confinement is pronounced in this limit. For example, in ref. 7 a feather-boa chain of length 8400 σ was considered which was confined to a cylinder of length $L = 50.75\sigma$ and diameter $D = 29.5\sigma$.
8. Consider a feather-boa chain with backbone of length $l_b = 200\sigma$ and side-loops of length $l_s = 40\sigma$ attached at each backbone

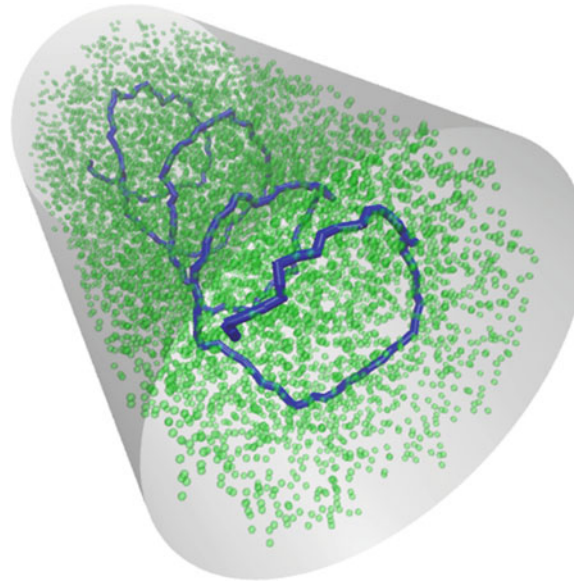


Fig. 3 (Color online) Equilibrium structure for a linear backbone chain of length $l_b = 200\sigma$ (thick blue line) to which side-loops of length $l_s = 40\sigma$ are attached at each backbone monomer. The polymer is confined within a cylinder of length $L = 50.75\sigma$ and diameter $D = 29.5\sigma$. The side-loop monomers are shown as transparent green beads

monomer, confined within a cylinder defined by length $L = 50.75\sigma$ and diameter $D = 29.5\sigma$. We equilibrated this system over $\tau_{eq} = 10^6$ steps after reaching this target volume, before collecting 10^3 data points separated by 10^3 time steps. The backbone chain displays a marked helical ordering with $\approx 12.5\ \sigma$ of helical pitch measured along the long axis of the cell (Fig. 3).

9. The degree of helicity can be quantified by considering the tangent-tangent correlation function $\langle \mathbf{u}(s) \cdot \mathbf{u}(0) \rangle$, where a monomer position along the contour is given by $s = i\sigma$ with $i = 0,1,...,200$. Its Fourier transform yields a structure function $S(q)$ with a peak at a dimensionless wavenumber $q_{max} = l_b/\lambda$, where λ denotes the pitch of the helix measured along the backbone chain.
10. As demonstrated in Ref. [7], the helical morphology in cylindrical confinement is stabilized by two effects: (a) emergent bending stiffness of the backbone due to side loops, (b) stacking of the loops under strong confinement. The first effect is an inherent property of the feather-boa polymer.
11. To quantify the emergent bending stiffness one may study the tangent-tangent correlation of a free feather-boa chain in the absence of confinement. As an example, we consider such a chain of backbone length $l_b. = 100\sigma$, associated with side-

loops of length $l_s = 20\sigma$ grafted on (a) every 10-th backbone monomer and (b) every backbone monomer. We compare the results with (c) a self- avoiding polymer that has the same $l_b = 100\sigma$ chain but in the absence of any side loop. The Flory scaling of end-to-end distance for a self-avoiding chain $\langle r(s)^2\rangle \sim s^{2b}$ implies a power-law decay of the tangent-tangent correlation $\langle \mathbf{u}(s).\mathbf{u}(0)\rangle \sim s^{-\alpha}$ with $\alpha = 2 - 2b$. Using $b = 3/5$ for three-dimensional real chains, one obtains a decay exponent $\alpha = 4/5$. Fig. 3 shows that the simulation results of self-avoiding chains indeed obey this prediction. With increasing density of side-loops, the value of exponent α decreases, finally showing $\alpha = 0.06$ for the feather-boa chain with every monomer of the backbone connected to a side-loop. This reflects a larger bending rigidity for backbones with increasing density of side-loops. The small value of α corresponds to exponent $b \approx 1$ signifying a rigid rod-like scaling in end-to-end separation $\langle r(s)^2\rangle \sim s^2$ (Fig. 2).

12. The mechanism leading to this enhanced rigidity is the following. The self-avoiding side-loops repel each other such that a locally bent conformation of the polymer becomes energetically less favorable. As a result the backbone stiffness increases with side-loop size ls, and the linear density of side loops along the backbone. The power-law nature of the correlation precludes any interpretation in terms of the worm-like chain model of polymer [23–26], where the correlation function decays as an exponential with the intrinsic persistence length. Nevertheless, it is intuitively clear that the side-loop packing causes a free-energy cost associated with local bending deformations. For example, the end-to-end distance $R \approx 413\ \sigma$ obtained for backbone length $l_b = 500\sigma$, can be reproduced by a worm-like chain with persistence length $l_p \approx 391\ \sigma$ [27], where l_p is comparable to the chain length l_b, characterizing a significantly stiff chain.
13. The chromosomal DNA of bacteria such as *E. coli* is a circular DNA. Thus, it is more appropriate to consider a circular feather-boa chain to model properties of the *E. coli* chromosome. As we have shown in ref. 7, within cylindrical confinement the backbone loop gets organized into two parallel helices running along the long axis of the cylinder, a prediction that was later verified in experiments on *E. coli* chromosome [5].
14. The feather-boa model successfully captures the emergent helical organization of bacterial chromosomes in cylindrical cells. This helicity is robust with respect to small variations of length and width of the confining cylinder. The thickening of the backbone due to side-loops enhances bending rigidity, and their stacking under confinement leads to the helical

shape. Arguably, a change in confining geometry would lead to a different shape. For example, if the confinement is spherical like that in several coccoid species, with a big enough diameter such that the backbone may open up completely, our model predicts a ring-like morphology of the chromosome, with the ring being free to rotate. Changes in chromosomal shape as a function of controlled but small variations in *E. coli* cell shapes have already been reported [5]. Another important contributor to chromosomal morphology is the cytosol of bacteria [13]. Osmotic pressure exerted by the cytosol may hold the chromosome floating inside cellular confinement. Some recent studies focused on entropic segregation of chromosome and ribosomes within bacterial cell, with chromosome localizing in a more central position of the cell while ribosomes concentrate near the cylindrical wall and the endcaps [28].

Acknowledgments

D.C. thanks SERB, India for financial support under Grant No. EMR/2016/001454. The work of B.M.M. is part of the research programme of the Netherlands Organisation for Scientific Research (NWO) and was performed at the research institute AMOLF. We thank Pinaki Swain for providing Fig. 2.

References

1. Zimmerman SB (2006) Shape and compaction of Escherichia coli nucleoids. J Struct Biol 156:255–261
2. van der Valk RA, Vreede J, Cremazy F, Dame RT (2014) Genomic looping: a key principle of chromatin organization. J Mol Microbiol Biotechnol 24:344–359
3. Postow L, Hardy CD, Arsuaga J, Cozzarelli NR (2004) Topological domain structure of the Escherichia coli chromosome. Genes Dev 18:1766
4. Marbouty M, Cournac A, Flot JF, Marie-Nelly-H, Mozziconacci J, Koszul R (2014) Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. elife 3: e03318
5. Fisher JK, Bourniquel A, Witz G, Weiner B, Prentiss M, Kleckner N (2013) Four-dimensional imaging of E. coli nucleoid organization and dynamics in living cells. Cell 153:882–895
6. Berlatzky IA, Rouvinski A, Ben-Yehuda S (2008) Spatial organization of a replicating bacterial chromosome. Proc Natl Acad Sci 105:14136–14140
7. Chaudhuri D, Mulder BM (2012) Spontaneous helicity of a polymer with side loops confined to a cylinder. Phys Rev Lett 108:268305
8. Chaudhuri D, Mulder BM (2011) Size and shape of excluded volume polymers confined between parallel plates. Phys Rev E 83:031803
9. de Gennes P-G (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca
10. Bates D, Kleckner N (2005) Chromosome and replisome dynamics in E. coli: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. Cell 121:899–911
11. Jun S, Mulder B (2006) Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome. Proc Natl Acad Sci 103:12388–12393

12. Wiggins P, Cheveralls K, Martin J, Lintner R, Kondev J (2010) Strong intranucleoid interactions organize the Escherichia coli chromosome into a nucleoid filament. Proc Natl Acad Sci 107:4991–4995
13. Pelletier J, Halvorsen K, Ha B, Paparcone R, Sandler SJ, Woldringh CL, Wong WP, Jun S (2012) Physical manipulation of the Escherichia coli chromosome reveals its soft nature. Proc Natl Acad Sci 109:E2649–E2656
14. Reiss P, Fritsche M, Heermann DW (2011) Looped star polymers show conformational transition from spherical to flat toroidal shapes. Phys Rev E 84:051910
15. Wang R, Virnau P, Binder K (2010) Conformational properties of polymer mushrooms under spherical and cylindrical confinement. Macromol Theory Sim 19:258–268
16. Hsu H-P, Paul W, Binder K (2010) Standard definitions of persistence length do not describe the local "intrinsic" stiffness of real polymer chains. Macromolecules 43:3094–3102
17. Theodorakis PE, Hsu H-P, Paul W, Binder K (2011) Computer simulation of bottle-brush polymers with flexible backbone: good solvent versus theta solvent conditions. J Chem Phys 135:164903
18. Weeks JD, Chandler D, Andersen HC (1971) Role of repulsive forces in determining the equilibrium structure of simple liquids. J Chem Phys 54:5237
19. Dijkstra M, van Roij R, Evans R (1998) Phase Behavior And Structure Of Binary Hard-Sphere Mixtures. Phys Rev Lett 81:2268–2271
20. Limbach H-J, Arnold A, Mann BA, Holm C (2006) ESPResSo—an extensible simulation package for research on soft matter systems. Comput Phys Commun 174:704–727
21. Flannery BP, Teukolsky S, Press WH, Vetterling WT (2007) Numerical recipes in C: the art of scientific computing. Cambridge University Press, New York
22. Parry B, Surovtsev I, Cabeen M, O'Hern C, Dufresne E, Jacobs-Wagner C (2014) The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. Cell 156:183
23. Doi M, Edwards S (1992) The theory of polymer dynamics. Clarendon, Oxford
24. Saito N, Takahashi K, Yunoki Y (1967) The statistical mechanical theory of stiff chains. J Phys Soc Jpn 22:219–226
25. Dhar A, Chaudhuri D (2002) Triple minima in the free energy of semiflexible polymers. Phys Rev Lett 89:65502
26. Chaudhuri D (2007) Semiflexible polymers: dependence on ensemble and boundary orientations. Phys Rev E 75:21803
27. Hermans J, Ullman R (1952) The statistics of stiff chains, with applications to light scattering. Physica 18:951–971
28. Mondal J, Bratton BP, Li Y, Yethiraj A, Weisshaar JC (2011) Entropy-based mechanism of ribosome-nucleoid segregation in E. coli cells. Biophys J 100:2605–2613

INDEX

Remus T. Dame (ed.), *Bacterial Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 1837, https://doi.org/10.1007/978-1-4939-8675-0,

G

H

I

L

M

N

O

P

Q

R

S

T

V

W