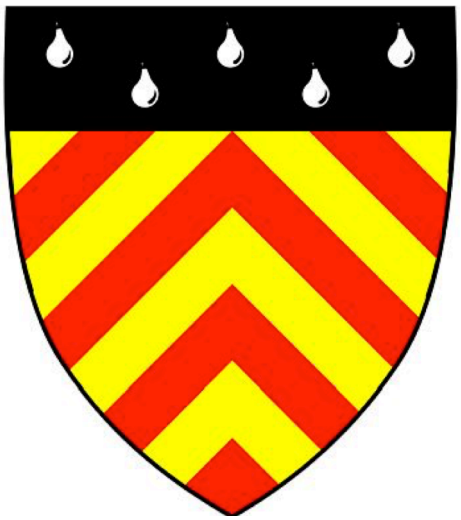
NUCLEOSOME-INDEPENDENT PERMANENTLY CONDENSED CHROMOSOME: INVESTIGATION OF NOVEL NUCLEAR ORGANISATION IN THE DINOFLAGELLATES

I Ian Hu

Clare Hall



September 2019

This dissertation is submitted for the degree of Doctor of Philosophy

Department of Biochemistry, University of Cambridge

Preface

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Degree Committee for the Faculty of Biology.

Abstract

Dinoflagellates hold vast diversities and are major contributors to overall marine primary production. Close relatives to the apicomplexan parasites and ciliates, the dinoflagellates, however, have a shockingly different nuclear biology. The dinoflagellates have massively inflated genomes (2-245 Gbps, up to 80X that of humans), permanently condensed liquid crystalline chromosomes throughout the life cycle, minimal histone proteins expression, and loss of nucleosome-mediated chromosome organisation. These changes co-occur with adoption of a novel nuclear protein termed DVNP (Dinoflagellate/Viral NucleoProtein). DVNP is highly positively charged, nucleus-located, and is found in only dinoflagellates and numerous marine large DNA viruses. I will address two issues in this thesis, 1) what are the properties of the protein DVNP, and can we infer if and how DVNP manages the permanently condensed chromosome state; and 2) how do other proteins accommodate this drastically different organisation of genetic material, more specifically how does transcription work in a cell with permanently condensed chromosomes? To elucidate the location of transcription, specific probe were designed and synthesised to locate RNA polymerase and newly synthesised nascent RNA in relation to the compacted chromosomes, and result showed that both are distributed at the periphery but not the inside of the chromosomes. On the other end, DVNPs are expressed and purified to perform biophysical and biochemical characterisation. In addition, using single-molecule optical tweezers microscopy, DVNP was observed to compact and change the mechanical properties of DNA. Most interestingly, the DVNP/DNA aggregates showed a propensity to travel along the DNA strand en masse. I then endeavoured to solve the NMR structure of DVNP. The results suggest that DVNP plays a central role in chromatin packaging in dinoflagellates. Finally, to marry the results obtained from both experiments, at the end of the thesis I propose a model of a novel nucleosome-independent chromatin management in dinoflagellates and how its transcription could proceed.

Acknowledgements

First and foremost, I'd like to thank my supervisor, Dr Ross Waller. Throughout my PhD Ross has constantly been supportive and enthusiastic, even with some of the unorthodox methods used in this thesis. He has nurtured me not only as an organised scientist, but also constantly taught me to ask the right questions. His perpetual love and enthusiasm for evolution and protistology is the basis of this thesis. He has taught me to write, as his expectations are the reason this thesis is in this current state of completion.

I would like to thank the comrades in the Waller and Carrington lab. Professor Mark Carrington for his often critical and helpful comments. Chapter Two of this thesis was the direct result of brainstorming sessions with Dr Konstantin Barylyuk, who also mentored me and provided perspectives throughout my PhD, as well as strengthening my knowledge in the fields of organic chemistry and chemical biology. Dr Eelco Tromer who introduced me into advanced evolutionary tools and helped with my deficiencies in molecular phylogeny. Dr Imen Lassadi and Dr Thomas Krueger for the countless discussions on mind-blowing dinoflagellate biology. My fellow PhD student, Carol Ke, for my very occasional abuse of seniority. And the rest of the labs for providing support, witty discussions, tolerance, and alcohol, without all of which this PhD would have crippled. Special mention alumni of the Waller lab, Dr Sebastian Gornik, whose initial confidence in and guidance for a young exchange student resulted in the eventual completion of this thesis.

I would like to thank the people in the Stott lab. Dr Katherine Stott coerced me into performing and learning more NMR experiments and analyses, and is personally responsible for the eventual generation of the NMR structure models presented in Chapter 7. Dr Mattew Watson, whose ample experiences in the lab and unlimited anecdotes made the bacterial expression processes less unbearable. Also Dr Andrew Travers and Professor Dame Jean Thomas, whose expertise in the fields and support proved to be invaluable to the generation of the models presented in this thesis.

I'd like to thank all my friends here in Cambridge. My Taiwanese friends and the football teams of Clare Hall and Trinity Hall had made my life here almost pleasant. Special thanks to the young and talented couple Dr Karol Fijalkowski and Ms Martyna Rzepecka who shared

love, values, and visions, and provided necessary writing potions at times in aid to the completion of this thesis.

Thanks to Dr Joanna Andrecka, not only for her support in the operation and data analyses for C-trap which was vital for the completion for Chapter 6, but also for her friendship, discussions, and anecdotes.

I'd like to thank my family, who have supported and provided for me and placed enormous amount of faith on me. I hope I will be worthy.

Lastly, I'd like to thank my beloved spouse, Ms Tingyin Liu, whose artistic nature has provided the essential balance for my life in Cambridge. Without her unconditional love, tolerance, and support, this thesis would not have been completed. I am forever in love and indebted.

Too many are not named in this acknowledgement. I have to thank the society, the karma, and the power of uncertainty and the universe for everyone and everything that had furthered this thesis.

Table of contents

List of figures

List of tables

Abbreviations

AUC	Analytical ultracentrifuge
bp	base pair
BSA	Bovine serum albumin
CD	Circular dichroism
Δε	Delta epsilon
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DVNP	Dinoflagellate/Viral NucleoProtein
EDC	1-ethyl-3-(-3-dimethylaminopropyl) carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
E. coli	*Escherichia coli*
FL	Full length
FPLC	Fast protein liquid chromatography
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HFF	Human Foreskin Fibroblast
IDP	Internally disordered protein
IFA	Immunofluorescence assay
IPTG	Isopropyl-β-D-thiogalactoside
ITC	Isothermal calorimetry
mRNA	Messenger RNA
MW	molecular weight
MS-spec	Mass spectroscopy
OD	Optical Density
PAGE	Polyacrylamide gel electrophoresis
RNAP	Ribonucleic acid polymerase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PSI-DNA	Polymer-and-salt induced condensed DNA
TEV	*Tobacco Etch Virus*
ThT	Thioflavin T
SDS	Sodium dodecyl sulfate
SOB	Super optimal broth
WT	Wild type

Chapter 1: A review on current knowledge of the dinoflagellates and the dinokaryon

1.1 The dinoflagellates and the "dinokaryon"

The dinoflagellates are a very diverse group of mostly single cellular marine and freshwater organisms. Most, if not all, dinoflagellates possess heterotrophic behaviour, while many still have functional chloroplasts for photosynthesis (Coats 1999; Jeong *et al.* 2010; Marie-Odile 2006). Dinoflagellates are major contributors to the total marine primary production, ranking only after diatoms in the eukaryotes (de Vargas *et al.* 2015; Malviya *et al.* 2016). The majority of dinoflagellates are single-celled motile organisms. In addition, several species form crucial complex endosymbiotic relationships with cnidarian animals, e.g. corals, sea anemones, and jellyfish (Pochon et al. 2006), and some are multinucleate crustacean parasites (Newman & Johnson, 1975; Appleton & Vickerman, 1998, 1996). Despite the apparent diversity, the phylum Dinoflagellata is, in fact, a monophyletic group that sits within the supergroup Alveolata, as described in the next paragraph. (Adl *et al.* 2005; Fensome *et al.* 1999; Leander & Keeling 2004; Saldarriaga *et al.* 2003, 2004; Zhang *et al.* 2007a).

The dinoflagellates and their relatives compose one of the largest and most diverse groups in eukaryotes (Guiry 2012). The dinoflagellates and their sister group, the perkinsids, are closely related to the apicomplexans and the ciliates, together forming the supergroup Alveolata (Fig. 1.1) (Gajadhar *et al.* 1991). The perkinsids are best known as bivalve parasites, although recently they have been reported to prey on fish and amphibians (Chambouvet *et al.* 2015; Isidoro-Ayza *et al.* 2017). The perkinsids share with the dinoflagellates the unusual post-transcriptional RNA modification of *trans*-splicing, where a small RNA sequence is transcribed separately, termed spliced leader, and spliced onto the 5' end of all pre-mRNA molecules. The apicomplexans are a group of animal intracellular parasites, responsible for diseases such as malaria and toxoplasmosis. The ciliates are common specimens for elementary biology classes for their abundance and ubiquitousness. They also bear an unusual chromosome management strategy. Two nuclei, a macronucleus and a micronucleus, exist in a cell. The macronucleus hosts hundreds of thousand copies of the chromosomes and is active during growth, yet divides by amitosis. Chromosome condensation and segregation never occur after replication, and the nucleus simply pinches in half, randomly distributing the contents into two daughter nuclei. This randomness eventually wears out the nuclear function, and the silent copy of the chromosomes hosted in the

micronucleus is used to regenerate a new copy of the macronucleus when necessary (Klobutcher 2001; Mochizuki 2010). All these examples demonstrate that the supergroup Alveolata indeed contains exceptional diversity.

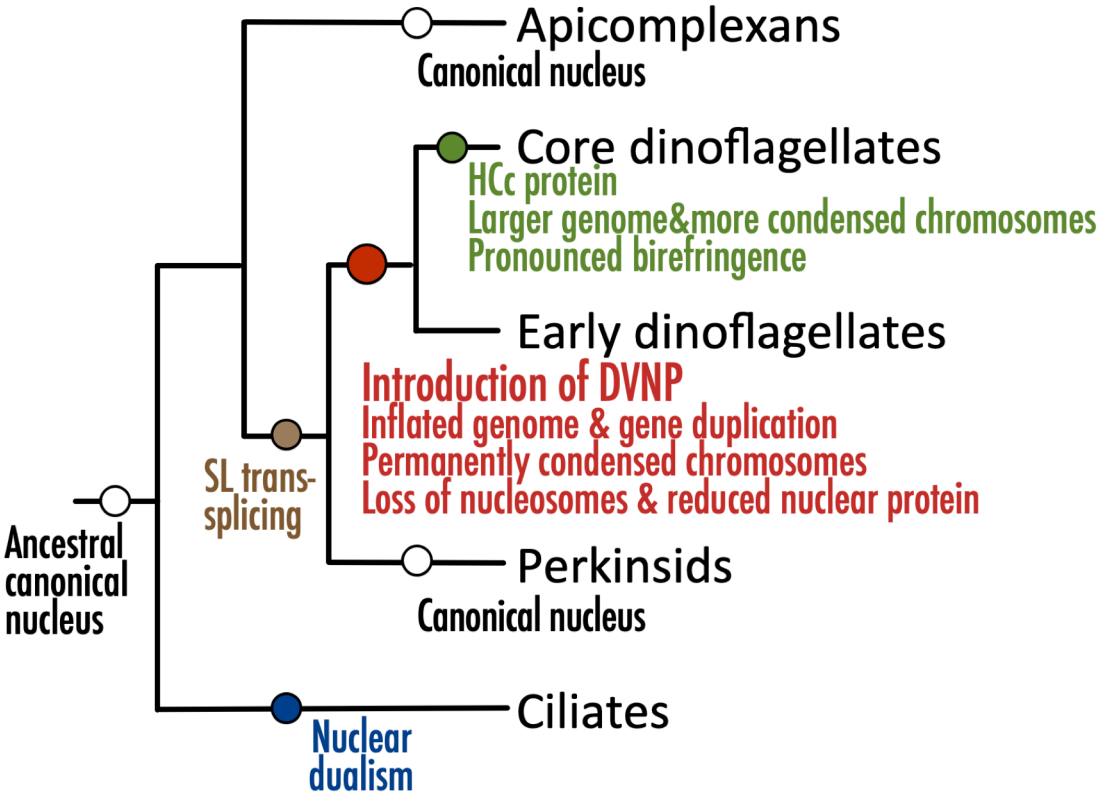


Fig. 1.1 | Simplified Evolutionary tree of Alveolata. HCc protein, a group of DNA-compacting protein homologous to baterial HU protein. SL, spliced leader RNA.

The dinoflagellates have a unique nuclear and chromosomal organisation only some of whose features are also found in other organisms. A "dinokaryon", the name given to the unusual dinoflagellate nucleus, comprises of several features. Firstly, the nuclear envelope does not break down during mitosis, a process termed closed mitosis, a feature shared with many eukaryotes (Adl *et al.* 2005). Secondly, it contains permanently condensed chromosomes throughout the life cycle but a permanent nucleolus with no condensed DNA within (Gao & Li 1986) (Fig. 1.2). The permanently condensed chromosomes are believed to be in a cholesteric liquid crystalline state, and are very often, but not necessarily, birefringent, causing the chromosomes to appear conspicuous under polarised light microscopy (Bouligand 1972; Cachon *et al.* 1989; Chow *et al.* 2010; Rill *et al.* 1989). In addition, the dinoflagellate genomes are massively inflated in size (approx. 1 - 245 Gbps (Lin 2006)) when compared with their close relatives (*Perkinsus marinus* 58 Mbps (Gornik *et al.* 2012),

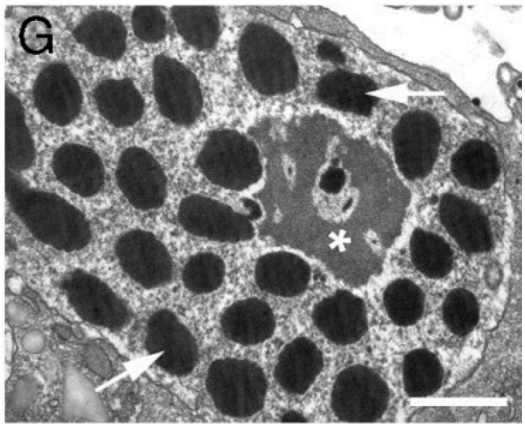


Fig. 1.2 | TEM image of a dinokaryon. Asterisk, uncondensed nucleolus. Arrow, condensed chromosome. Reproduced from Lukeš et al. 2009 with permission.

Toxoplasma gondii 80 Mbps (Kissinger *et al.* 2003; Sibley & Boothroyd 1992)). Along with the genome expansion occurred the extensive duplication of most, if not all, of the genes, demonstrated in EST and genome/transcriptome sequences (Bachvaroff *et al.* 2004; Bachvaroff & Place 2008; Beauchemin *et al.* 2012; Lin *et al.* 2010; McEwan *et al.* 2008). The dinoflagellates also have vastly diminished expression of histone proteins (Gornik *et al.* 2012; Marinov & Lynch 2015; Roy *et al.* 2018), and along with that loss of nucleosome-mediated DNA organisation (Bodansky *et al.* 1979; Gornik *et al.* 2012; Herzog & Soyer 1981; Rizzo & Burghardt 1980; Talbert & Henikoff 2012). Overall nuclear protein content is also observed to be much lower than canonical eukaryotes, having a protein to DNA ratio of 1:10 rather than the usual 1:1 by a combination of biochemical methods (Bhaud *et al.* 1999; Fukuda & Suzaki 2015; Gornik *et al.* 2012; Kato *et al.* 1997; Rizzo & Noodén 1973). Coincident with the loss of histones and all of the other unusual feats, dinoflagellates adopted a novel nuclear protein that is highly positively charged and small in size, characteristics similar to histones, termed Dinoflagellate/Viral NucleoProtein (DVNP) (Gornik et al. 2012). DVNP is also the most expressed gene in a dinoflagellate cell (Marinov & Lynch 2015). Some later derived dinoflagellates have then acquired another family of nuclear proteins the HCc family proteins from bacteria, and have overall even more expanded genome size and more condensed chromosomes (Chan *et al.* 2006; Chow *et al.* 2010; Vernet *et al.* 1990). All these most unusual features gave the dinoflagellate nucleus its own name of dinokaryon (Soyer 1971).

The origin of the dinokaryon and how it functions are still mysterious to biologists. Many attempts made to understand different aspects of the dinokaryon are described further in later chapters, however, currently, the community has only various experimental results as fragments of a puzzle, but have yet assembled it. There are no comprehensive and coherent explanations for how and why the dinoflagellates shifted toward this highly divergent nuclear structure. Because of their unique nuclear biology, dinoflagellates were once considered an intermediate in the evolution from prokaryotes to eukaryotes termed mesokaryotes (Dodge

1965), however evidence including genomic and transcriptomic data undisputedly support that the dinoflagellates are true eukaryotes that group within Alveolata despite their derived oddities (Adl *et al.* 2005; Hargraves 1998).

Based on the broad evolutionary history of the alveolates, the last common ancestor of dinoflagellates, apicomplexans, and ciliates very likely had a normal nucleus of canonical features (Adl *et al.* 2005; Lukeš *et al.* 2009). As demonstrated in Fig. 1.1, the apicomplexans and the perkinsids both have the canonical nucleosome-based DNA packaging system, although the *Perkinsus* species share one peculiar feature with the dinoflagellates, namely the 5' spliced leader *trans*-splicing, hence placing the emergence of the dinokaryon right after the divergence of the perkinsids. *Perkinsus* was once thought to be among the apicomplexans (Burreson *et al.* 1994), but subsequent molecular phylogenies group *Perkinsus* undoubtedly together with the dinoflagellates (Saldarriaga 2003; Saldarriaga *et al.* 2004). Evidently, when the dinoflagellates started diverging from the perkinsid lineage, substantial changes occurred. In addition, although the ciliates also have an equally bizarre means of chromosome management, there is no reason to believe that the two are related.

Coincidentally, there are currently no reliable methods for genetically modifying dinoflagellates, perhaps because of the highly divergent nuclear and chromosomal biology where a lot is still unknown. Studies reporting success of nuclear transformation, using seemingly conventional methodologies, have seen no follow-up publications, even by the original authors (Lohuis & Miller 1998; Ortiz-Matamoros *et al.* 2015a; Ortiz-Matamoros *et al.* 2015b). The joint effort of an international team has disproved the tractability of methods used in previously mentioned publications in *Symbiodinium microadriaticum* (Chen *et al.* 2019). The lack of a nuclear genetic modification system however limits significantly the methods available in the dinoflagellate research field and for the present study.

1.2 Optical/physical and biochemical properties of the dinokaryon

Dinokaryons are very conspicuous by electron microscopy (Fig. 1.3). The rod-shaped condensed chromosomes may display a toroidal twist, and the texture of the chromosomes may be a banding pattern or arch-like fibrils depending on the staining methods and the angle of sample sectioning (Bhaud *et al.* 2000; Cai *et al.* 1992; Livolant & Bouligand 1980; Oakley & Dodge 1979; Rill *et al.* 1989; Soyer 1977; Soyer & Haapala 1974). Under high magnification, unlike typical eukaryotes, there has been no observation of nucleosomes or 'bead-on-a-string' units in dinokaryons but only 6 nm smooth filaments, and it had been suggested several times that the dinoflagellates simply do not have nucleosomes (Bodansky *et al.* 1979; Herzog & Soyer 1981, 1983; Hinnebusch *et al.* 1980). In *Peridinium*, a dinoflagellate that also hosts an amitotic slave diatom nucleus in addition to the dinokaryon, in total extracted chromatin a smooth chromatin fibre can be observed as well as the bead-on-a-string structure (Fig. 1.3g) (Rizzo & Burghardt 1980; Tippit & Pickett-Heaps 1976).

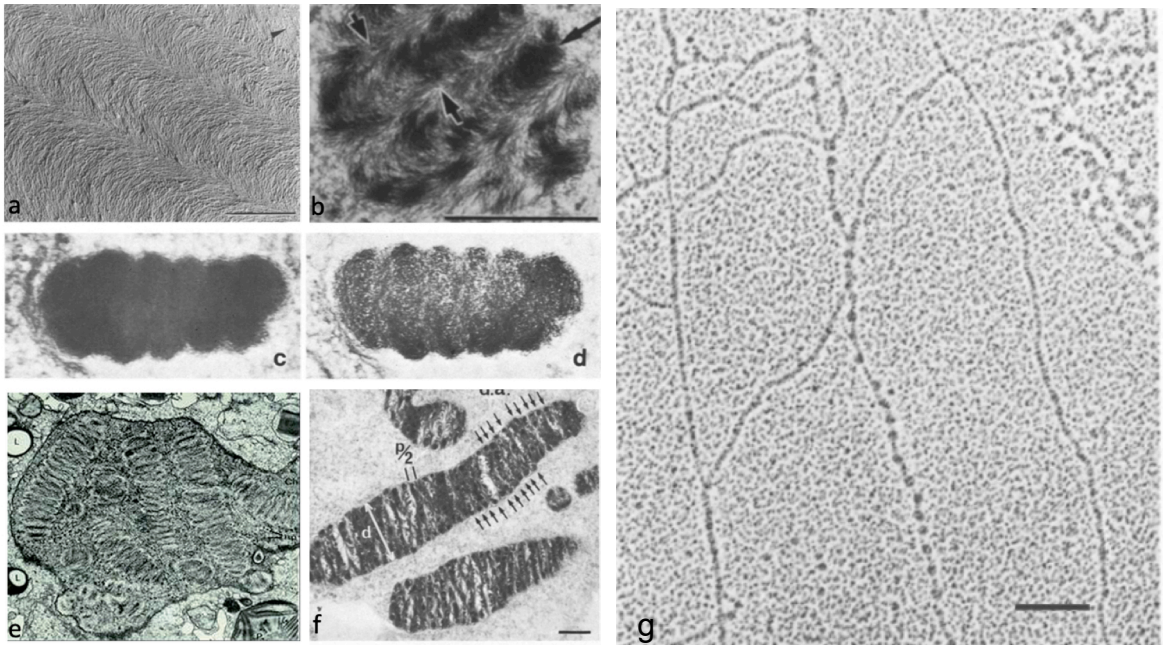


Fig. 1.3 | Dinokaryon under electron microscopy. a and b, arch-like fibrils of a dinokaryon, from Rill et al 1989 and Soyer & Haapala1974. Arrowhead in a, direction of shadow. Arrows in b, the arch-like patterns. c and d, toroidal twist of a dinokaryon chromosome, from Oakley & Dodge 1979; e, organization of the compacted chromosomes of *Prorocentrum*, from Soyer 1977; f, the banding pattern of the chromosomes, from Livolant & Bouligand 1980. Arrows, dense accumulations. d, diameter of the chromosome; a-f, bar = 1 μm. g, chromatin fibres of the binucleate dinoflagellate Peridinium. Note the presence of both smooth chromatin fibre and the bead-on-a-string structure. Bar, 0.1 μm. All images reused with permission.

It has been proposed that the DNA of the dinokaryon condensed chromosomes is organised into a liquid crystalline state based on the electron micrographs and other optical tests. It is suspected to be in a chiral "cholesteric" liquid crystal state, as represented in Fig 1.4 (Rill *et*

al. 1989; Strzelecka *et al.* 1988; Yee *et al.* 2012), the state being the apparent reason behind the signature arch-like fibrillar appearance observed from electron microscopy. It has long been known that DNA is able to self-assemble into cholesteric phases *in vitro* at very high concentration, supporting the possibility of this sort of DNA arrangement *in vivo* (Livolant & Bouligand 1978; Rill *et al.* 1989; Skuridin *et al.* 2016). The proposed cholesteric state is chiral and helical in nature, as the double-stranded DNA weaves through one plane before travelling to the next, offset by a small angle between planes. The helical pitch (P) of the chromosomes can be calculated by analysing the arch-like pattern in the EM images (Fig. 1.4). Many dinoflagellates' nuclei also have the property of birefringence, being positive under polarised light microscopy and circular dichroism (Cachon *et al.* 1989; Chow *et al.* 2010; Livolant & Maestre 1988; Rill *et al.* 1989). Birefringence is an optical property that refracts polarised light according to its angle, and in the cases of many ordered liquid crystals because of the chirality and helical nature of the molecular order. It is worth mentioning that species in the Syndiniales, the earliest branching dinoflagellates, have permanently condensed chromosomes, yet no birefringence has been observed (Chow *et al.* 2010). This result could suggest that although microscopically the chromosomes of the syndinian and core dinoflagellates look similar, the fundamental DNA organisation may be different, or that the still uncharacterised feature that gives a dinokaryon birefringence was either acquired or only became observable in the core dinoflagellates.

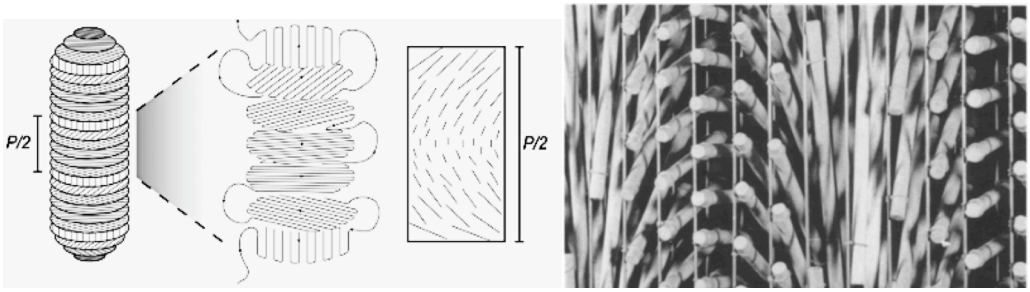


Fig. 1.4 | Representation of a cholesteric arrangement of liquid crystalline chromosome. Left, the proposed model redrawn in Chow et al, 2010; Right, wooden stick model from Rill et al 1989. Reused with permission. P/2, helical pitch of the cholesteric organisation.

The later branching species tend to have more expanded genome sizes than earlier ones, yet cell size and nucleus size do not seem to have increased proportionally (Chow *et al.* 2010; Dodge 1965; LaJeunesse *et al.* 2005; Spector 1984). Chow et al. measured the birefringence of six species of dinoflagellates with a Metropol polarising microscopy system, and presence of birefringence seemed to correlate with genome size (Chow *et al.* 2010). Birefringence requires the material to be chirally ordered at the molecular level, as it is the chirality of the

order that causes the optical phenomenon described above. It has thus been suspected that not just the genome size, but also the degree of DNA compaction, is what's causing the optical features, as the more tightly compacted and chirally ordered DNA would introduce more birefringence. To test this, Chow et al. measured the helical pitch by analysing electron micrographs of chromosomes from multiple dinoflagellate species and calculating the helical pitches. The study demonstrated that the strength of birefringence is positively correlated to helical pitch, which in turn is negatively correlated to DNA density; the tighter the DNA twist, the higher the DNA density, and the stronger the birefringence. It also demonstrated elegantly that nuclei expelled from cells of some non- or weakly birefringent dinoflagellate species have intrinsic birefringence properties. The study suggested that the other structures, e.g. the plasma membrane, also an ordered liquid crystal, may introduce negative interference to the optical properties, causing the loss or lack of birefringence at the cellular level. Chow et al. further proposed that since the highly compacted DNA is regulated by the comparatively much smaller amount of protein, the whole process of chromosome assembly must be driven by entropy, the process being highly regulated (Chow *et al.* 2010).

Biochemically, it has been repeatedly demonstrated that divalent cations, especially calcium and magnesium ions, are of great importance to the maintenance of dinokaryon chromosome structure, just as the two cations are now understood to play crucial roles in canonical eukaryotic chromatin organisation as well (Strick *et al.* 2001). X-ray spectrum analyses of electron microscopy of dinoflagellate nuclei sections confirm the high concentration of not only Ca^{2+} and Mg^{2+} ions, but also various elements in period IV of the periodic table, in the nucleus as well as the chromatin specifically. The ratio of divalent to phosphate ions within chromosomes was determined to be roughly 1:2, which correlates to one cation per two nucleotides, which can be further extrapolated as either one cation per base pair or one cation per two bases from two strands of double-strand DNA in a compacted state. Trypsin, EDTA, and EGTA were shown to have distinct impact on the integrity of the permanently condensed chromosome, either *in situ* or *ex vivo*, implying the involvement of both the divalent ions and a protein factor (Herzog *et al.* 1984; Herzog & Soyer 1983; Kearns & Sigee 1980; Rill *et al.* 1989; Rizzo 1979; Sun *et al.* 2012). A report that utilised secondary ion-probed mass spectrometry identified the co-localisation of secondary Mg^{+} and Ca^{+} ions and the permanently condensed chromosomes further confirmed the presence and importance of the divalent ions (Levi-Setti *et al.* 2008). In this study, the ratio of divalent ions to phosphate was determined to be 0.6 cations per base pair. The disagreement to earlier studies was discussed,

and these authors argued that this could be explained by the potential improper retention of the more soluble calcium and magnesium ions in the earlier studies. The rather irregular number of 0.6 cation per bp probably also implies that the DNA/divalent cation relationship is likely not a simple one, as previously proposed, yet the importance of the divalent cations was again attested. Interestingly, it has been demonstrated that divalent ions on a 2D plane as well as positive polyamines in 3D space have the activity to aggregate linear DNA strands (Bloomfield 1996, 1997; Cherstvy 2008; Koltover *et al.* 2000). This may well be one of the many factors that contribute to the final condensed state of the chromosome.

There are other potential factors responsible for the permanently condensed state of the chromosomes. A "structural RNA", of which the identity has yet to be discovered, was also mentioned in an earlier publication to hold the chromosomes in place, which can be dismantled with a high concentration of RNAse treatment. Dinoflagellate genomes are also known for containing an unusually high amount of modified bases: 12% to 70% of thymine is modified to 5-hydroxymethyluracil, with an hydroxyl addition to the methyl group on the pyrimidine ring, depending on species (Davies *et al.* 1988; Herzog *et al.* 1984; Rae & Steele 1978). To test whether dinoflagellate chromosomal DNA, its sequences, or the high level of modified bases are the crucial factor for the loss of nucleosomal DNA organisation, Herzog et al. (1984) were able to reconstruct the bead-on-a-string structure with purified dinoflagellate chromosomal DNA and corn histone proteins. Similarly, in a rather unorthodox study, Liu et al. (2000) first isolated dinoflagellate nuclear chromosomes then injected them into a frog egg. The chromosomes went through dramatic morphological changes and finally reached a dispersed state similar to that of chromatin in a regular eukaryotic fertilised egg. These results indeed suggested that the DNA sequences and the high level of modification do not exclude the nucleosomal type of DNA packaging, yet sheds little light on the exact mechanism of the dinokaryon chromosome compaction itself.

Several models for the 3-dimensional organisation of DNA have been proposed to try to explain the unusual structure of the dinoflagellate permanently condensed chromosomes. Most notably: i) the 'toroidal chromonema' model by Dodge (Oakley & Dodge 1979), ii) 'stacks of DNA discs' model by Bouligand (Bouligand & Norris 2001), iii) 'central core fibre' model by Spector (Spector *et al.* 1981), and iv) the 'peeled, cored, sliced pineapple' model by Rizzo (Levi-Setti *et al.* 2008) (Fig 1.5). It is possible that each of these experimentally supported models is correct at different levels of the grand scheme of chromosome management of the dinoflagellates, but the whole picture still requires more research input.

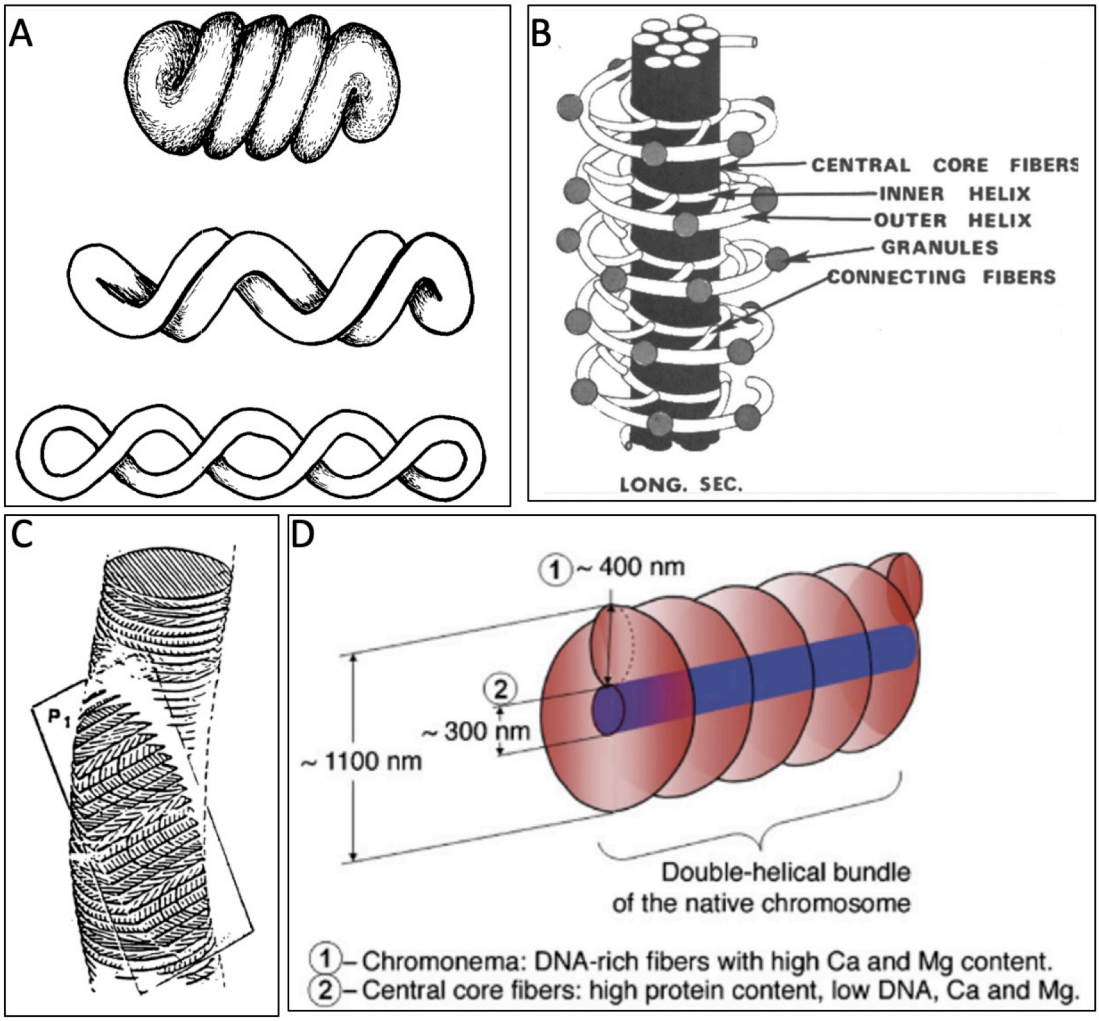


Fig. 1.5 | Current proposed models of the permanently condensed chromosomes. A, toroidal chromonema model (Oakley & Dodge 1979); B, centre core fibre model (Spector *et al.* 1981); C, stacks of DNA discs model (Bouligand & Norris 2001); D, 'peeled, cored, sliced pineapple' model (Levi-Setti *et al.* 2008). All images reused with permission.

1.3 Histone proteins of the dinoflagellates

In most canonical eukaryotes, histone proteins are an integral part of chromatin organisation: two copies of H3 and H4 form a tetramer, which is then joined by two H2A/H2B dimers forming an octamer. DNA wraps around for approximately 1.7 times into a structure called a nucleosome, the most basic packaging unit. Nucleosomes and the DNA form a "bead-on-a-string" structure, observable under electron microscopy. The nucleosomes then serve as basic units for more complicated and higher-order packaging (Annunziato 2008; Luger *et al.* 1997) (Fig. 1.6). DNA management and packaging are integral to the cell cycle, and canonical histone proteins are regulated by cyclins, whereas non-canonical histone variants exist in the cell in a mostly cell cycle-independent manner. Histone variants bear different functions to the canonical histones, e.g. DNA repair, transcription initiation, and termination (Ewen 2000; Hunt *et al.* 2013; Medina *et al.* 2012; Venkatesh & Workman 2015). All histone proteins, canonical or variant, possess a long N-terminal tail, on which multiple post-translation modifications, i.e. methylation, acetylation, phosphorylation and ubiquitination, form particular patterns called "histone codes" which are recognised by proteins for downstream processes such as activation of transcription or recruitment of other transcription factors (Jenuwein & Allis 2001).

Dinoflagellates seem to have abandoned the canonical histone packaging of chromatin (Rizzo 2003; Vernet *et al.* 1990). The canonical "bead-on-a-string" basic units were not present in electron micrographs, acid protein extraction showed a low ratio of basic nuclear protein to DNA, and histological methods show poor staining for basic nuclear proteins, as previously described in section 1.2. These results demonstrate discrepancy with nucleosome-mediated chromatin management (Herzog & Soyer 1981; Rizzo & Nooden 1972). In addition, in chromatin extract of *Peridinium* , a binucleate dinoflagellate which hosts a dinokaryon as well as a secondary eukaryotic nucleus, both smooth chromatin and bead-on-a-string were observed using the electron microscopy, demonstrating the absence of nucleosomes in the dinokaryon (Rizzo & Burghardt 1980; Tippit & Pickett-Heaps 1976). Furthermore, when dinoflagellate chromatin was digested with micrococcal nuclease, no nucleosome-mediated protection was observed; instead of the signature ladder-like protective pattern of the nucleosome, all that was observed is a smear ((Bodansky *et al.* 1979; Gornik *et al.* 2012; Herzog & Soyer 1981). The dinokaryon had been hence for a long time assumed to be completely histone-less.

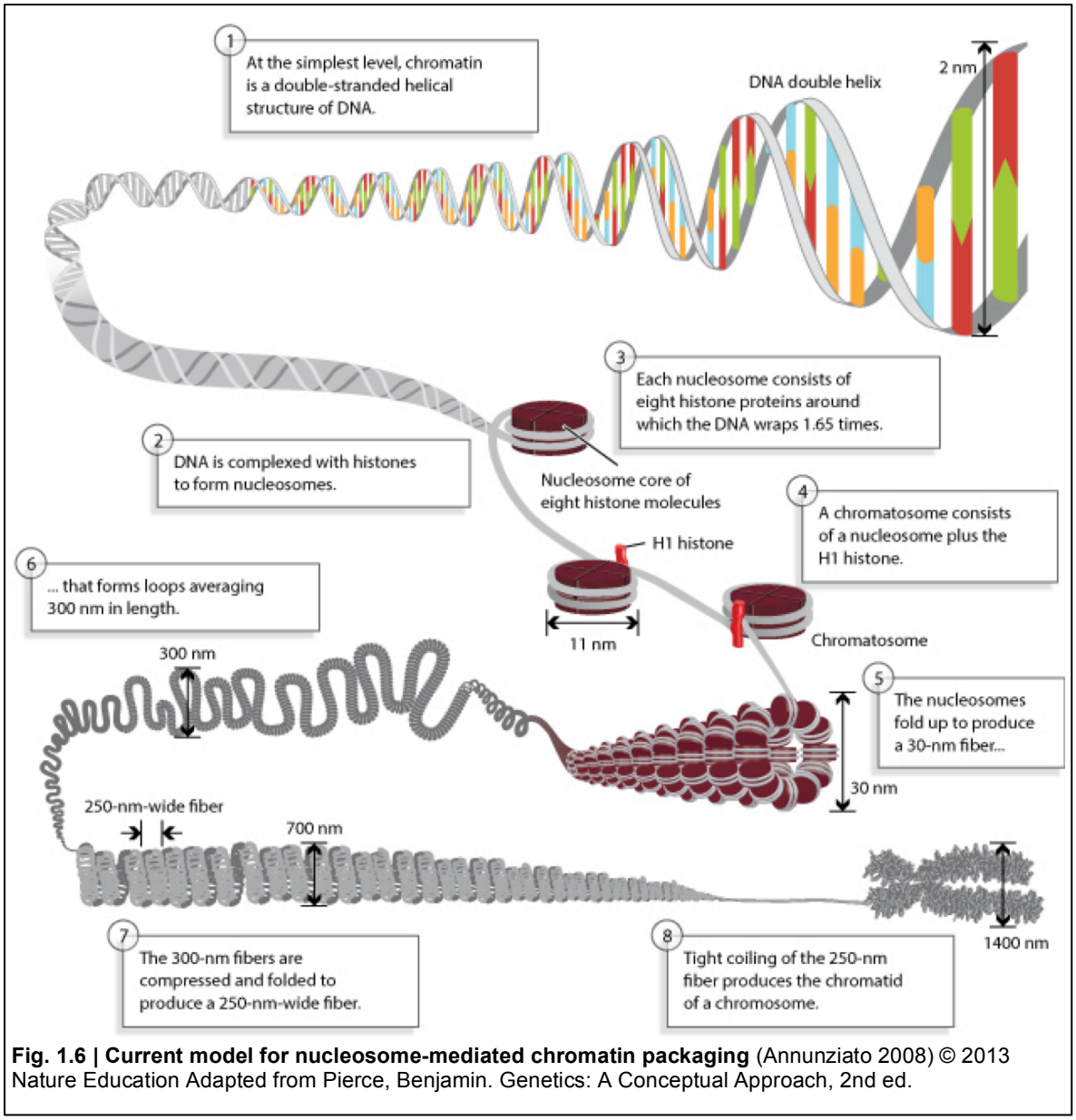


Fig. 1.6 | Current model for nucleosome-mediated chromatin packaging (Annunziato 2008) © 2013 Nature Education Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed.

Surprisingly, comparatively recent EST evidence and RNA-seq data demonstrated the existence of multiple canonical histone proteins and variants along with histone-modifying enzymes (Aranda *et al.* 2016; Bachvaroff & Place 2008; Baumgarten *et al.* 2013; Beauchemin *et al.* 2012; Lin *et al.* 2010; Roy & Morse 2012; Shoguchi *et al.* 2013). The RNA expression levels of histone proteins, however, seemed low compared to the most abundant nuclear protein DVNP and at the protein level, histones are still below detection limits of current instruments other than the highly sensitive mass spectroscopy (Gornik *et al.* 2012; Marinov & Lynch 2015; Rizzo 2003).

Marinov & Lynch (2015) performed a comprehensive study on dinoflagellate histone proteins based on all publicly available sequence information. The histone proteins in

dinoflagellates bear enormous variance between and even within each protein. Besides the usual trick of duplicating each gene multiple times, some of the histone proteins were appended with N- or C-terminal tails so long that the proteins reach up to 1000 amino acid residues, while others were almost a tenth in size (Fig. 1.7). Most of the common histone variants, e.g. H2A.Z, H2A.X, and CENP proteins, can also be found in dinoflagellates, albeit these bear equally significant divergence. The existence of these variants hints that the roles that these histone variants fulfil in the canonical eukaryotic nucleus may still exist in the highly derived dinokaryon, and are carried out by the now much more divergent variant counterparts. Histone tails and the conservation of their post-translational modification marks are also extremely divergent, and even frequently absent. To the original authors' surprise, histone modifiers, and especially members of the FACT complex, which plays crucial roles in transcription through chromatinised DNA templates (Orphanides *et al.* 1998; Reinberg & Sims 2006), are found in almost all dinoflagellates (Marinov & Lynch 2015). The authors argued that the presence of these proteins strongly supports the existence of nucleosomes and transcription through them, although no EM images of the bead-on-a-string structure have been observed or published (Marinov & Lynch 2015).

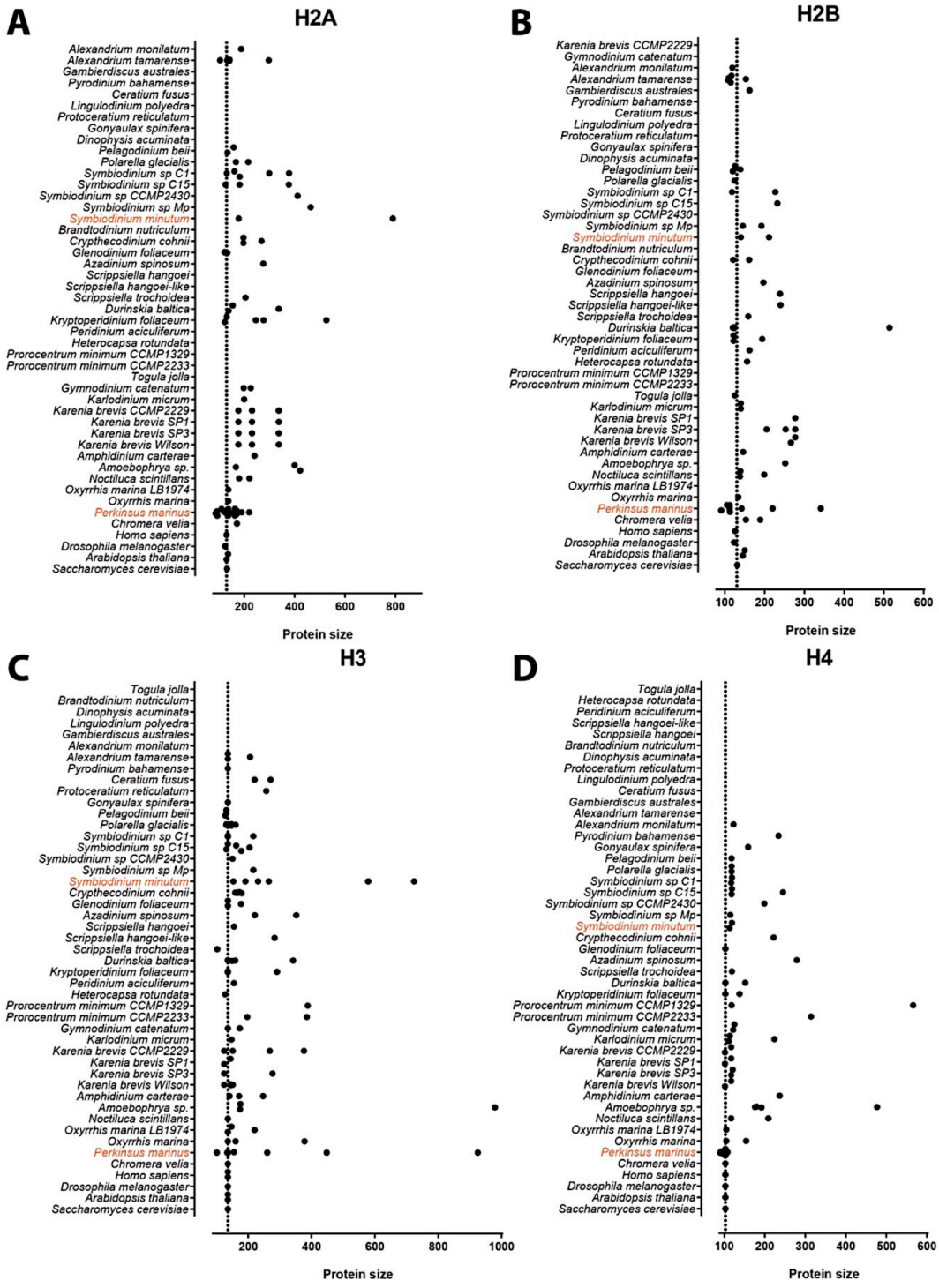


Fig. 1.7 | Size distribution of histone proteins in dinoflagellate species. Human, fruit fly, Arabidopsis, and yeast were presented at the bottom of each panel for comparison. Note the existence of multiple genes in each dinoflagellate species. Reproduced from Marinov & Lynch 2015 with permission.

1.4 Dinoflagellate/Viral NucleoProteins (DVNP)

DVNP is the most highly expressed transcript in a dinoflagellate cell at the RNA level (Marinov & Lynch 2015). At the protein level, DVNP seems to have replaced the histones. DVNP is unique in that the presence of this gene in the eukaryotes is limited to dinoflagellates (Gornik et al. 2012). It is a nuclear protein that exists in multiple highly similar copies in any dinoflagellate species. It is a small soluble protein (10-20 kDa) rich in positive amino acids, similar to histone proteins. As much as 30% of the entire sequence can be lysine and arginine, and pI can reach up to 11. DVNP binds to DNA with similar affinity to histone proteins, has at least several phosphorylation states, and the sequence is potentially prone to methylation and other post-translation modification (Gornik *et al.* 2012; Zhang *et al.* 2012). Based on sequence alignment and analysis, the protein can be separated into three regions; the middle highly conserved and hydrophobic core region, and the N- and C-terminal tails which are both highly enriched in positively charged amino acid residues with less sequence conservation (Gornik *et al.* 2012). One DVNP, *Hematodinium* DVNP.5, has been predicted to have seven potential phosphorylation sites, and at least three phosphorylated forms of total DVNP has been observed in *Hematodinium* nuclear extract (Gornik *et al.* 2012).

From the *in vivo* perspective, DVNP is known to co-localise with DNA in the dinoflagellate nucleus and its location correlates with DNA (Fig. 1.8). Although similar to histone in many aspects, DVNP is predicted to have a helix-turn-helix secondary structure that is different from the classic histone fold (helix-turn-helix-turn-helix) structure (Gornik *et al.* 2012; Luger

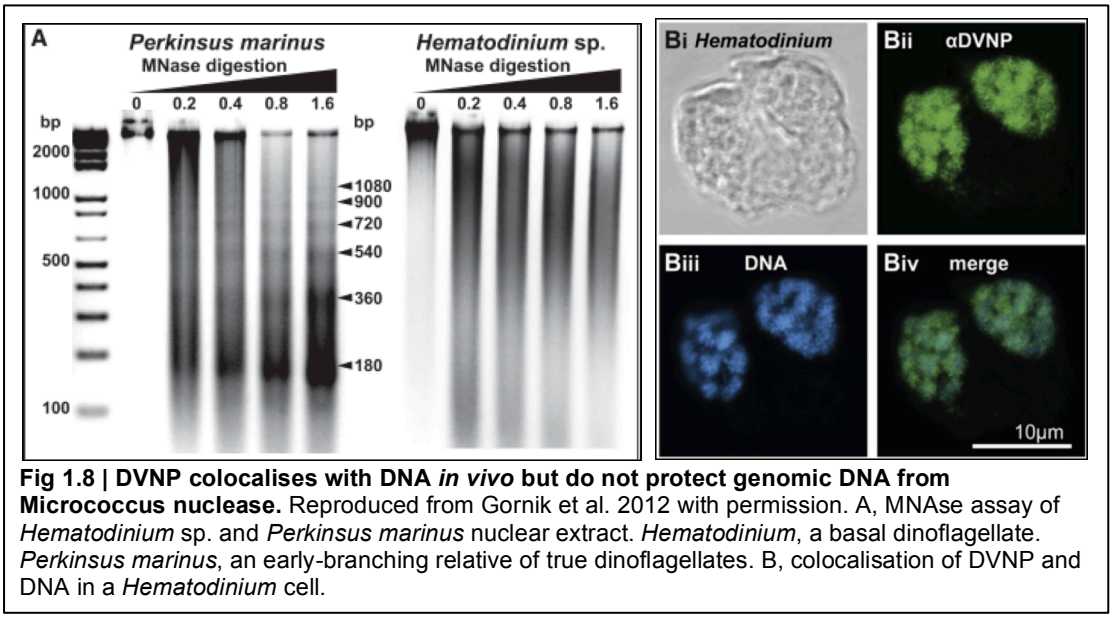


Fig 1.8 | DVNP colocalises with DNA *in vivo* but do not protect genomic DNA from Micrococcus nuclease. Reproduced from Gornik et al. 2012 with permission. A, MNAse assay of *Hematodinium* sp. and *Perkinsus marinus* nuclear extract. *Hematodinium*, a basal dinoflagellate. *Perkinsus marinus*, an early-branching relative of true dinoflagellates. B, colocalisation of DVNP and DNA in a *Hematodinium* cell.

et al. 1997). DVNP, unlike histones, does not seem to protect genomic DNA from micrococcus nuclease digestion by forming the nucleosome or similar unit structure (Fig. 1.8); when digested, instead of the signature ladder pattern, dinoflagellate chromosomal DNA is observed as a smear (Bodansky *et al.* 1979; Gornik *et al.* 2012; Herzog & Soyer 1981). Outside of dinoflagellates, DVNP homologues were only found in one family of marine viruses, Phycodnaviridae, across all living organisms and biological agents (Gornik *et al.* 2012). DVNP is conserved across early and late branching dinoflagellates and also conserved within Phycodnaviridae. The Phycodnaviridae DVNP, however, does not have the N-terminal lysine-rich extension found in dinoflagellates. DVNP may serve similar functions in both clades; the Phycodnaviruses are ancient large DNA viruses with enormous genomes (up to 500 kbps), and DVNP may be a means to manage its abundant genetic material (Dunigan et al. 2006; Van Etten et al. 2002; Gornik et al. 2019). As the family Phycodnaviridae is thought to be as old as the eukaryotes themselves, up to 2 million years by molecular clock, it is more likely that the DVNP gene was transferred by a virus to the earliest dinoflagellate ancestor and not otherwise, however with current knowledge it is difficult to verify this as no known dinoflagellate host to any members of the large DNA virus family has been identified (Van Etten et al. 1991; Van Etten et al. 2002; Dunigan et al. 2006; Gornik et al. 2012, Gornik et al. 2019).

There has been a recent attempt to recreate the nuclear biology of the very first ancestor of dinoflagellates by expressing dinoflagellate DVNP genes in baker's yeast and analysing the genomic and transcriptomic sequence data of the affected cells (Irwin *et al.* 2018). The results found that the protein exerts high toxicity towards cells and possibly replaces nucleosomes on DNA through high occupancy rather than direct competition, and also that cells expressing lower amount of histone proteins have better survivability, though with substandard statistics on many of the findings. Interestingly, the study found a general decrease in transcription activity, yet the lengths of transcripts were not affected, implying that only transcription initiation was affected and not elongation. Based on the similarities of high occupancy of DVNP and low histone expression between the yeast survivors and the dinoflagellates, the study seems to support the hypothesis that the gene DVNP may have flowed from a viral origin to an alveolate ancestor.

1.5 Transcription, splicing, and the transcription conundrum of dinokaryon

Very little was known about the transcription in dinoflagellates until quite recently. Scarce DNA and protein sequences were available, and the condensed chromosome organisation almost seems incompatible with the idea of transcription. However it was suggested that the primary underlying mechanism of RNA polymerase II is still conserved to an extent by an earlier classic study that used the fungal toxin α-amanitin, a potent RNA polymerase II inhibitor, to inhibit RNA synthesis in extracted *ex vivo* dinoflagellate nuclei (Rizzo 1979). Early studies utilising autoradiographic electron microscopy all showed RNA signals accumulating on the periphery, but not the inside, of the condensed chromosomes, supporting the hypothesis that transcription does not happen within the chromosomes, and only the portion of loose DNA looping out of the compacted chromosomes is active (Babillot 1970; Echeverría *et al.* 1993; Sigee 1983).

In canonical eukaryotes, RNA polymerase II (RNAPII) is a large protein complex that is the primary RNA polymerase responsible for the transcription of most mRNAs. The C-terminal tail of the largest subunit of RNAP contains 20-55 repeats of the heptapeptide sequence YSPTSPS, in which different residues can be phosphorylated at different stages –known as the "CTD code"– of the transcription process (Cramer *et al.* 2001; Meinhart & Cramer 2004). This repeat is highly conserved across the most commonly studied model organisms from human to yeast to green plants – though not so across all eukaryotes (Liu *et al.* 2010) – and is used as an antigen for generation of almost all commercial antibodies. Abundant transcription factors and helper proteins are required for successful transcription. It is known that not only DNA sequences but also certain histone codes are responsible for the recruitment of TFII transcription factors and upstream or downstream elements (Brickey & Greenleaf 1995). The transcription factors then recruit the polymerase, and after shifting to the correct phosphorylation state, the polymerase then starts the synthesis of nascent RNA.

Dinoflagellates seem to possess most of the proteins required for the transcription process (Marinov & Lynch 2015). However, several TFII protein families in dinoflagellates have fewer members than their canonical eukaryote counterparts, and several TFII proteins that are thought to be essential for transcription are not found in dinoflagellates, suggesting either loss of the functions of these proteins or acquisition of these functions by existing proteins. Furthermore, although the C-terminal heptapeptide repeat region does exist, the length and the amino acid composition of each repeat unit bears significant variance, and the most

conserved "YSPTSPS" sequence is rarely found (Roy & Morse 2012; Marinov & Lynch 2015; Hu & Waller, present work). The details are discussed further in Chapter 2.

The dinoflagellate seems incompatible with the current understanding of nuclear transcription. The current consensus on chromosome biology asserts that condensed heterochromatin deactivates specific DNA regions by compaction, reducing their availability to proteins, whereas open, uncompacted euchromatin is more accessible (Wegel & Shaw 2005). In this context, the whole dinoflagellate genome composed of permanently condensed chromosomes would seemingly be silent entirely, as none of the DNA sequences would be freely available for DNA and RNA polymerases as well as transcription factors to interact with. It has been proposed that under the liquid crystalline chromosomes model, while weaving through planes to planes of the condensed chromosome, exon DNA may loop out of the chromosome body (see Fig. 1.4), hence becoming the active sites for transcription where the protein could gain access and process by *trans*-splicing (Sigee 1983). However as more genomic and transcriptomic data of dinoflagellates become available, the seemingly ordinary organisation of introns and exons provides no evidence for the above scenario, which would require introns or intergenic regions kilobases in size to span the width of a chromosome (see Fig. 1.9). Other possible explanations to the transcription conundrum include either the polymerases, along with other proteins, readily travel into the inside of the chromosome, or instead that the chromosomes are in such fluid and dynamic state that DNA travels out of the chromosome to be accessed. Either way, this would suggest a completely different nature of the dinokaryotic chromosome to canonical eukaryotic heterochromatin.

Transcription commonly is spatially and temporally coupled with *cis*-splicing, i.e. the exons start being spliced together before the transcription of the gene is completed (Bentley 2014). Dinoflagellates are known to process their pre-mRNAs with *trans*-splicing to append a separately transcribed short "splice leader (SL)" sequence to the 5' end. The same process is observed in many other organisms across eukaryotes, including the trypanosome parasites, the nematodes, the cnidarians, and dinoflagellates' close relative the perkinsids (Douris *et al.* 2010; Zhang *et al.* 2007b, 2011). Dinoflagellates seem to also possess most of the proteins required for *cis*-splicing as most components serve in both processes (Roy & Morse 2012). Until emerging genomic and transcriptomic data suggested otherwise, dinoflagellates were once believed to be intron-less, for this is most compatible with the hypothesis of active DNA being on the periphery of chromosomes (Lidie *et al.* 2005). Recent genomic and

transcriptomic data for an early-branching dinoflagellate *Hematodinium* sp. indicates exceptionally large genes, as large as 100 kbs, with disproportionately small exons on average of two hundred bps, in between few dozens kbs of introns (Waller et al. unpublished) (Fig. 1.9). On the other end of the spectrum, the published draft of *Symbiodinium minutum*, major cnidarian symbiont, displays much more tightly packed genes curiously placed head to tail in the same direction, smaller introns of about 500 bp, but with the size and number of exons remaining relatively the same, all while maintaining a genome size of roughly 3 Gb (Shoguchi *et al.* 2013) (Fig. 1.9). It's currently unclear whether the introns are actually transcribed, or there is a possibility of *trans*-splicing at the exon level as well as the 5' spliced leader. *Trans*-splicing at exon level is found in the nuclear mRNA of *Giardia* (Kamikawa *et al.* 2011), and mitochondrial mRNA of *Diplonema* (Valach *et al.* 2014), a free-living Euglenozoa, and very likely mitochondrial mRNA of *Karlodinium*, *Alexandrium*, and *Symbiodinium*, all more derived dinoflagellate species (Jackson & Waller 2013). The presence of this phenomenon in other unrelated species warrants the possibility of nuclear exon level *trans*-splicing in dinoflagellates, especially when the mechanism is already present in the dinoflagellate mitochondrion. Nuclear *trans*-splicing would only have required the mechanism to be exported from mitochondrion to the nucleus. For comparison, the mitochondria and chloroplasts of dinoflagellates have both exported genes and pathways to the nucleus through horizontal gene transfer (Jackson & Waller 2013; Patron & Waller 2007; Teng *et al.* 2013).

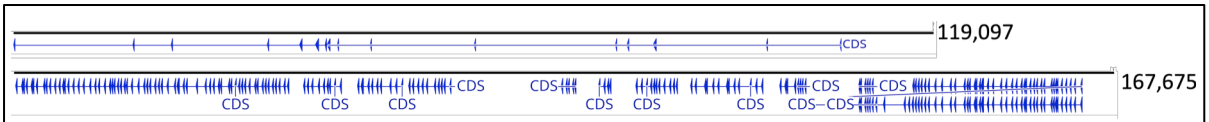


Fig. 1.9 | Genomic scaffolds of two representative dinoflagelates. Upper, *Hematodinium* (Waller, unpublished); lower, *Symbiodinium minutum* (Shoguchi et al. 2013). Sizes of contigs (bp) are marked on the right. Each arrowhead represents an exon, each occurrence of CDS represents one gene. The *Hematodinium* contig has one gene. *Symbiodinium* has twelve, all in the same direction.

1.6. Evolution of the dinokaryon and project overview

It is worth investigating what exactly happened and how it happened when the first dinoflagellates started diverging from its closest ancestor. In a seemingly short amount of evolutionary time, numerous events took place, namely: 1) the introduction of DVNP; 2) loss of histone proteins and nucleosomes; 3) reduced protein to DNA ratio in the nucleus, 4) inflation of genomes and duplication of genes; and 5) formation of permanently condensed chromosomes. Currently, there are no known intermediate species that has only part of the full suite of these characteristics, thus identifying them as sister groups to the current dinoflagellates that diverged during this transition. It is hence difficult to understand the progression of this dramatic change and whether there were causalities between the acquisitions of each of these features. In more derived dinoflagellate species, also referred to as "core dinoflagellates", several new DNA packaging proteins were introduced, including the HCc proteins and the Dinap proteins (Bhaud *et al.* 1999; Fukuda & Suzaki 2015), which coincides with more tightly packed genomic contents, and even smaller nuclear protein-to-DNA ratio (Chow *et al.* 2010; Fukuda & Suzaki 2015; Sala-Rovira *et al.* 1991). In addition, dinoflagellates also boast highly unusual mitochondria and chloroplast biology; the mitochondrial genome is fragmented into small linear double-stranded DNA (Jackson *et al.* 2012), whereas the chloroplast genome is largely reduced in content and exists as plasmid-like minicircles, each holding no more than 4 genes (Howe *et al.* 2008).

Perkinsus species form a sister group to the true dinoflagellates. They do not have the DVNP gene, neither do they possess most of the distinctive traits of the dinokaryon, e.g. reduced expression of histone and loss of nucleosomes, expanded genome, and reduced nuclear protein-to-DNA ratio. They do however share the spliced leader (SL) *trans*-splicing with the dinoflagellates, thus marking the starting point when the common ancestor of both groups started to diverge from canonical eukaryotes (Gornik et al. 2012). The shared ancestry makes the perkinsids excellent tools for comparison to understand the nuclear biology and evolution of the dinoflagellates. Coincidentally, the ciliates, close relative to both groups, also have a highly unusual chromosome management strategy, namely nuclear dualism, as described briefly in section 1.1. However, based on comparative studies, the ciliate nuclear duality is likely a derived feature gained after it branched off from the ancestor of apicomplexans and dinoflagellates, and there is no evidence supporting that the dinoflagellate and the ciliate DNA managements are related and no reason to suspect so.

DVNP seems to play a pivotal role in the grand scheme of the evolution of the dinokaryon. According to the evolution history, the ancestor of the dinoflagellates is most likely a free-living marine cell with chloroplast, and perhaps more importantly, had a canonical nucleus with a nucleosome-mediated DNA management system. It is then hard to imagine under what circumstances would the ancestor of the dinoflagellates undertake this drastic change to its nuclear biology without critical selection pressure. For relevance, Irwin's study in 2018 demonstrated that overexpression of DVNP, a toxic protein certainly not good for a common cell's well-being (Goh & Waller 2015; Gornik *et al.* 2012; Irwin *et al.* 2018), reduced occupancy of nucleosomes, and perhaps positively selected towards the individuals with lower histones expressions. This study is the only known case where causality between any of the highly unusual dinokaryon characteristics is established. The most likely scenario of what happened to the very first dinoflagellate is a large DNA virus with the gene DVNP, either as a toxic viral effector to attenuate the host, or as its own means of DNA management, infected an alveolate host (Gornik *et al.* 2019). The host, potentially with a lowered histone expression in the first place even before the event, somehow survived the infection, and its descendants domesticated DVNP by means of genome expansion and gene duplication, causing permanently condensed chromosomes and a completely rewired chromosome biology. One would imagine under this scenario, the first dinoflagellates with a mature dinokaryon and domesticated DVNP proteins were probably much fitter than the initially infected cells, which would have been out-competed during evolution resulting in the current lack of alveolate species with a transition state nucleus.

The present PhD study is centred on the theme of the permanently condensed chromosome. I will address this issue from two angles: 1) how do other nuclear proteins accommodate this drastically different organisation of genetic material, more specifically, how does transcription work in a cell with permanently condensed chromosomes? And 2) what are the properties of the protein DVNP, and can we infer if and how DVNP manages the state of the permanently condensed chromosome? For the first issue, I seek to elucidate the location of RNA polymerase and newly synthesised nascent RNA *in vivo* with innovative methods. To tackle the second issue, I analyse all the available information of DVNP across the tree of life *in silico* and try to understand the protein from an evolutionary perspective. I will also recombinantly express two DVNP proteins, from both a dinoflagellate and a viral source, and determine their biochemical and biophysical properties. Finally, to answer these questions in a continuous and coherent manner, I propose a model of how DVNP condenses and manages

DNA organisation, supported by all the experimental results provided in this study. These are discussed in separate chapters.

Chapter 2: Locating RNA polymerase II and transcription in dinoflagellates

2.1 Introduction

How transcription works in the case of the permanently condensed chromosomes of dinoflagellates has been a mystery. Current understanding of eukaryotic nuclear biology asserts that nucleosomal DNA can be compacted to reduce its accessibility to protein factors (Wegel & Shaw 2005). Permanently condensed chromosomes would then suggest no active DNA in the life cycle, clearly far from truth evident from the abundant RNA transcription and the fact that the cells are alive and active. This chapter seeks to detect the location of RNA polymerases and newly transcribed nascent RNA in the dinoflagellate cells as the first step to understand dinoflagellate nuclear biology.

The DNA in the permanently condensed chromosomes was proposed to be in a cholesteric liquid crystalline arrangement (see Fig. 1.4), where the DNA weaves back and forth, forming a 2D plane, and then travels to the next plane, of which the angle is offset by a few degrees (Rill *et al.* 1989; Strzelecka *et al.* 1988; Yee *et al.* 2012). The cholesteric liquid crystallinity model explains the conspicuous appearances of the dinokaryotic chromosomes in electron micrographs, however, does not solve the transcription conundrum described in the previous paragraph. It was then proposed that only the peripheral DNA of the chromosomes is accessible, the DNA linking plane to plane. Earlier studies using radioisotopic RNA nucleotides and autoradiographic electron microscopy all suggested that the active DNA is on the periphery of the chromosomes, supporting the hypothesis (Babillot 1970; Echeverría *et al.* 1993; Sigee 1983). The hypothesis however inevitably led to the proposition that the dinoflagellate has very large intergenic space that could span the width of the chromosomes and perhaps an intron-less gene structure (Lidie *et al.* 2005). This proposition turned out to be false; as large scale genomic and transcriptomic datasets became available, it seems that many dinoflagellates have fairly standard gene structures, intron/exon sizes, and intergenic spaces (Keeling *et al.* 2014; Mendez *et al.* 2015; Shoguchi *et al.* 2013). It has been shown that sample preparation processes, those for dinoflagellates included, may cause electron microscopy and electron micrographic autoradiography to be prone to artefacts (Kellenberger *et al.* 1992).

Hence, I decided to revisit some of the earlier experiments of elucidating the location of transcription and, in addition, the RNA polymerase, to gain more insight into the unique arrangement of DNA.

In canonical eukaryotes, RNA polymerase II (RNAPII) is a large protein complex and the primary RNA polymerase responsible for the transcription of most nuclear mRNAs. The C-terminal domain (CTD) of the largest subunit of RNAPII contains 20-55 repeats of the heptapeptide motif "YSPTSPS", in which different residues can be phosphorylated at different stages (known as the "CTD code") of the transcription process (Cramer et al. 2001; Meinhart & Cramer 2004). This repeat is conserved across many, but far from all, common eukaryotes, e.g. animals, fungi, and green plants (Liu *et al.* 2010), and is used as the antigen for almost all commercial antibodies against RNAPII. In dinoflagellates, most of the proteins required for the transcription process can be found in transcriptomic data (Roy & Morse 2013). However, though the C-terminal repeats region does exist, the heptad repeats are highly divergent from the consensus sequence YSPTSPS (Marinov & Lynch 2015). In addition, the distance between DNA strands in a condensed chromosome state can be as small as 30 angstroms (Koltover *et al.* 2000; Raspaud *et al.* 2005). I reasoned that antibody-based methods for localisation might have limitations due to the dimensions of an antibody molecule (IgG 12 nm, (Noll *et al.* 1982)) when compared with the condensed state of the chromosome, be it reduced penetration of the antibody. Hence a small molecule probe is preferred. An earlier study used the small molecule fungal toxin α-amanitin, a potent RNA polymerase II inhibitor, to inhibit RNA synthesis in isolated dinoflagellate nuclei *ex vivo,* suggesting the basic underlying mechanism of RNA polymerase II is still conserved (Rizzo 1979). I sought to utilise this reported specific affinity and use the fungal toxin as a probe for detecting RNAPII.

The amatoxin family consists of 5 members, α-, β-, γ- δ-, and ε-amanitin. All five are bicyclic peptides 8 amino acids in length biosynthesised in some of the toxic species in the *Amanita* genus of fungi. The amatoxins are very potent RNAPII inhibitors, interestingly acting on the hinge of the polymerase rather than the more commonly seen active site occupation (Bensaude 2011). The sole published use of a fluorescently modified α-amanitin as a probe for RNAP detection involved very

complicated chemical conjugation utilising the phenol group on the tryptophan side chain. The product demonstrated good sensitivity and signal to noise ratio, although with a much reduced affinity compared to the unmodified toxin (Wulf et al. 1980). Based on subsequently described X-ray crystallography results of the amanitin bound to the human RNAPII (Bushnell et al. 2002), I identified that the phenol moiety used for labelling in Wulf's study forms an important hydrogen bond with the polymerase. Labelling on this residue may cause steric hindrance, which could explain the reduced affinity of the labelled toxin. I reasoned that labelling another member of the amatoxin family, β-amanitin with a free carboxyl group that does not form interactions with the polymerase, should not hinder the binding to RNAPII thus provide equivalent sensitivity without compromising the binding affinity.

2.2 Results

2.2.1 Dinoflagellate RNAPII CTD heptapeptide motif is poorly conserved

Sequence alignment of RNAPII sequences of human, yeast, *Arabidopsis*, several dinoflagellates, and a representative of the closely related Apicomplexa, *Toxoplasma gondii*, demonstrates that RNAPII from all species are conserved. The conservation is, however, much lower in the CTD than in the rest of the protein (Fig. 2.1). The sequences of the CTD reveal that the archetype of the repeat heptapeptide motif "YSPTSPS" and the repetition are in fact present in all the species, however, in dinoflagellates, much higher variation in sequence and length is found in each repeat (Fig. 2.2).

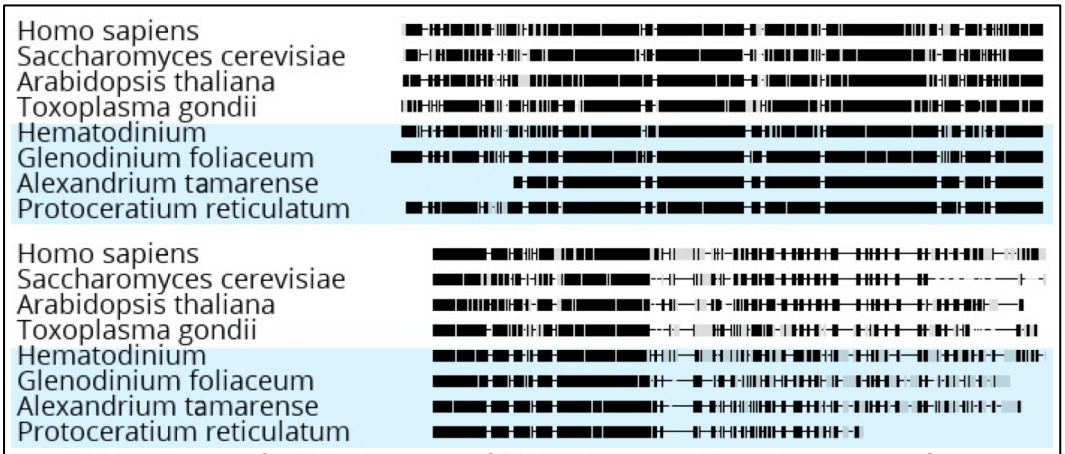


Fig. 2.1 | Illustration of global alignment of RNA polymerase II protein sequence from selected species across eukaryotes. Residues in higher agreement are represented as black blocks. Less conserved region are represented as grey blocks or space. Dinoflagellate sequences are in blue. Note the drop of conservation at the C-terminal end.

2.2.2 Successful synthesis of the fluorescent conjugates with β-amanitin

Based on previously described X-ray crystallography results of the amanitin with human RNAPII (PDB ID:1K83) (Bushnell et al, 2002), the hydroxyl group on the hydroxytryptophan moiety used for labelling in Wulf's study directly forms hydrogen bond with a glutamate residue (Q822) of the polymerase (Fig. 2.3). Substituting the hydroxyl group with a bulky fluorophore may cause steric hindrance to the binding, a potential reason for the reported reduction of binding affinity in the original study (Wulf *et al.* 1980). The only difference between α- and β-amanitins, an asparagine versus an aspartate residue respectively, has no contact with the polymerase and points away from the binding pocket. I reasoned that the aspartate side chain on β-amanitin would be a better candidate for modifying with a fluorophore.

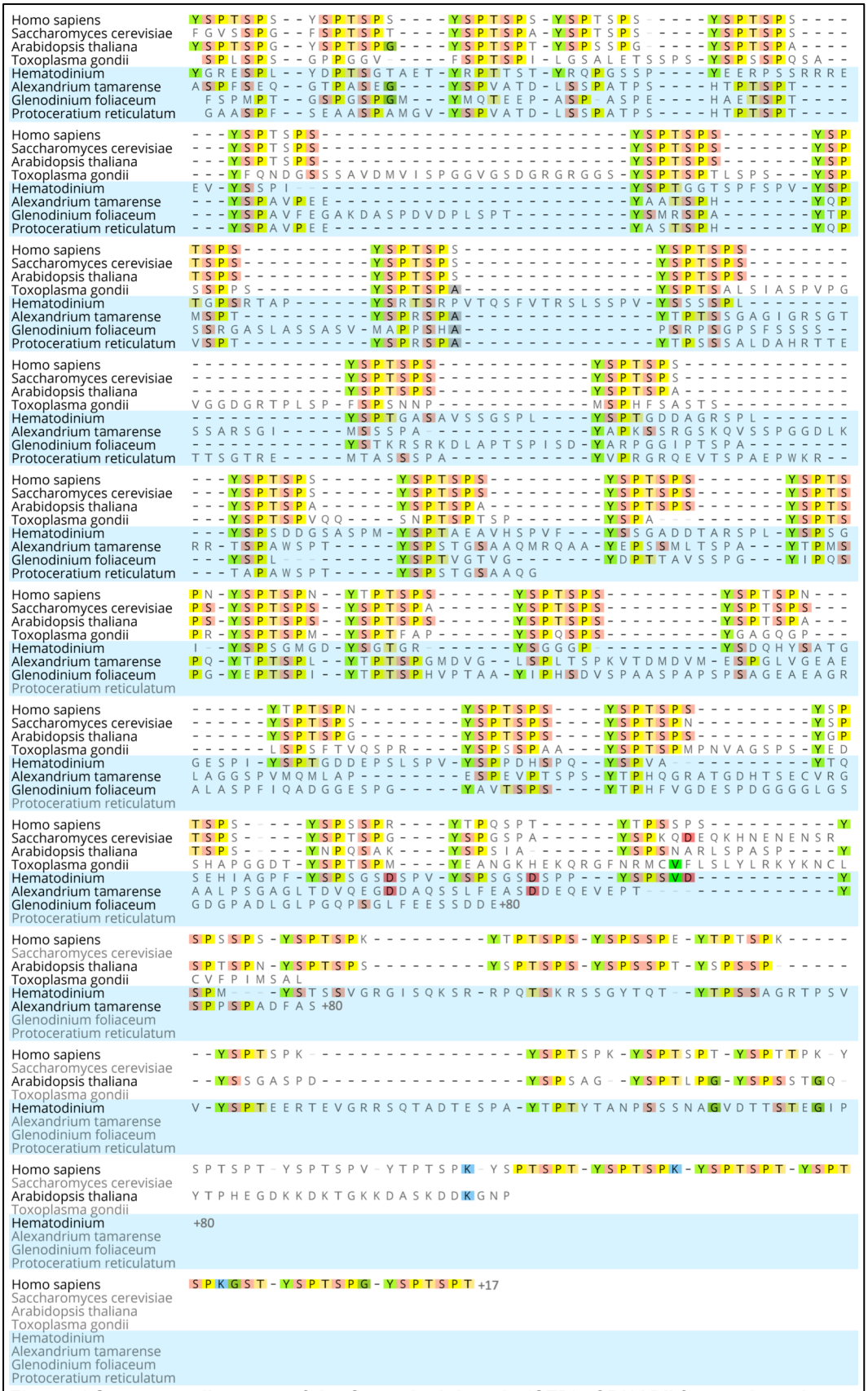


Fig. 2.2 | Sequence alignment of the C-terminal domain (CTD) of RNAPII from selected common eukaryotes and dinoflagellates. Each recognisable repeat of the sequences were aligned. Most common amino acid in each position is marked with colour. Extra sequences at the C-terminal end of *Homo sapiens*, *Hematodinium*, *Alexandrium tamarense*, and *Glenodinium foliaceum* that does not bear any resemblance to the heptapeptide are removed. Blue represents dinoflagellate sequences.

Two fluorophores were chosen for this study, Oregon Green®-488 (OG488) and Dyomics DY-654 for their spectral properties of high extinction coefficients and high stability (Fig. 2.4). Conjugation with OG488 was first tested in aqueous condition. 1-ethyl-3-(-3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) was used to activate the carboxyl group of β-amanitin, sulfo-*N*-hydroxysuccinimide (sulfo-NHS) to stabilise the activated carboxyl, and reacted with NHS-ester of OG488. This reaction resulted in a low reaction rate, and successful conjugate could not be detected by mass spectrometry (not shown). I sought to improve the reaction by performing the reaction in the organic phase (dimethylformamide, DMF) to minimise water competition to EDC, and accordingly the organic hydroxybenzotriazole was used in place of the water-soluble sulfo-NHS. At the same time, the concentrations of both reactants, the dye and the toxin, were raised to 10 mM to promote the reaction rate. After reverse phase HPLC fractionation, a total of three species of the potential conjugate in addition to the unreacted dye itself were purified. To identify these products, the fractions were analysed with mass spectrometry, the three species were confirmed to be the correct conjugate (m/z -1420.46), acetylated OG488 half-product (m/z -539.18), and self-reacted OG488 (m/z -975.29), respectively (Fig. 2.5). The correct product was then stored frozen.

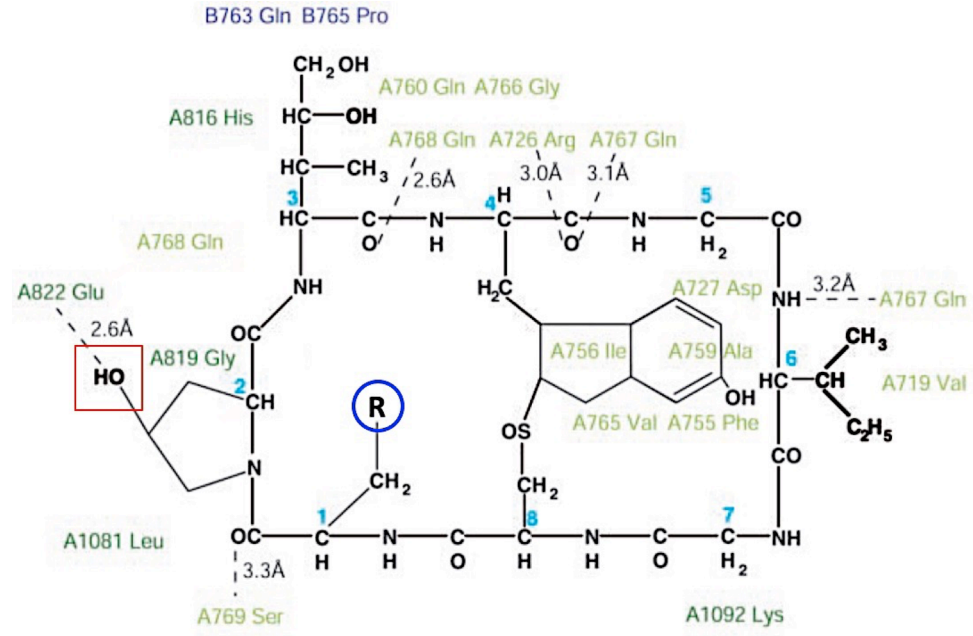


Fig 2.3 | Differences in the structure between α-amanitin and β-amanitin. Reproduced and modified from Bushnell et al. 2002. Copyright (2002) National Academy of Sciences. Red box, where Wulf's compound was modified. R, O=C-NH2 in α-amanitin, O=C-OH in β-amanitin. Human RNAPII residues within 4 Å when bound are shown in olive and blue. Hydrogen bonds are marked as dashed lines.

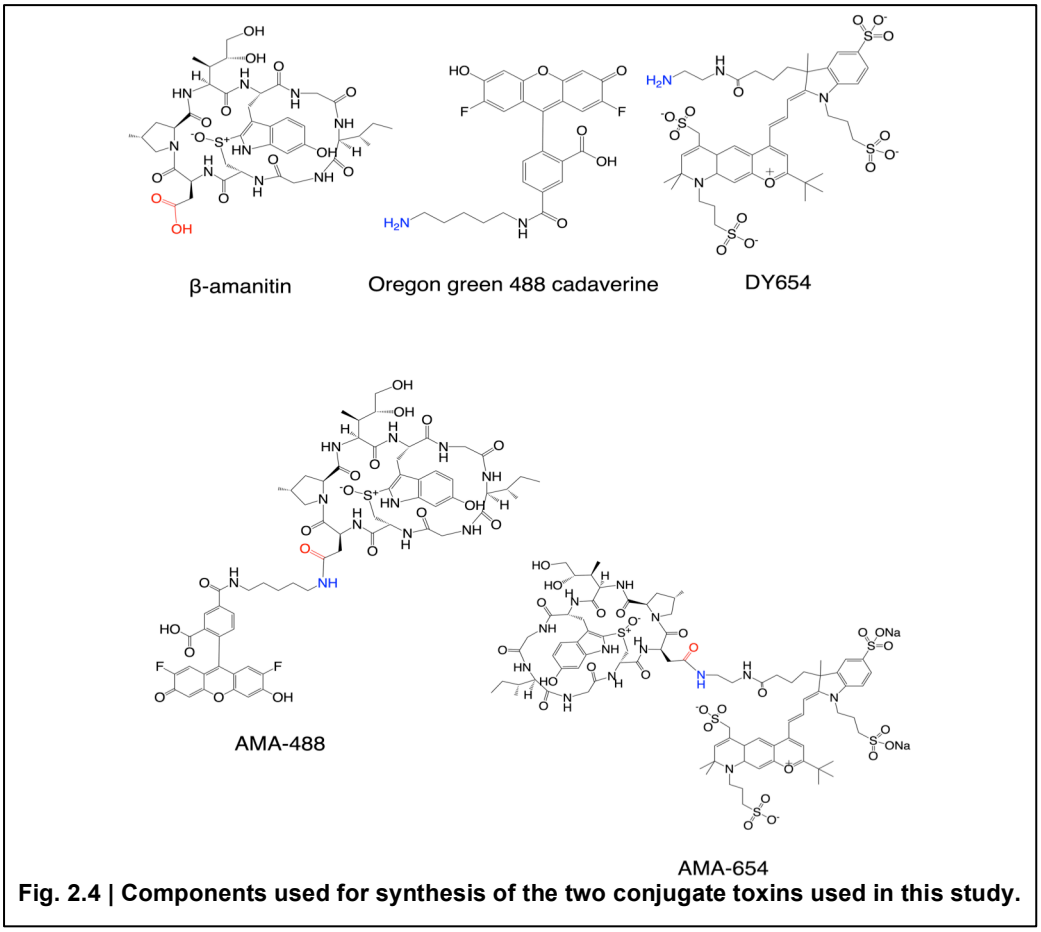


Fig. 2.4 | Components used for synthesis of the two conjugate toxins used in this study.

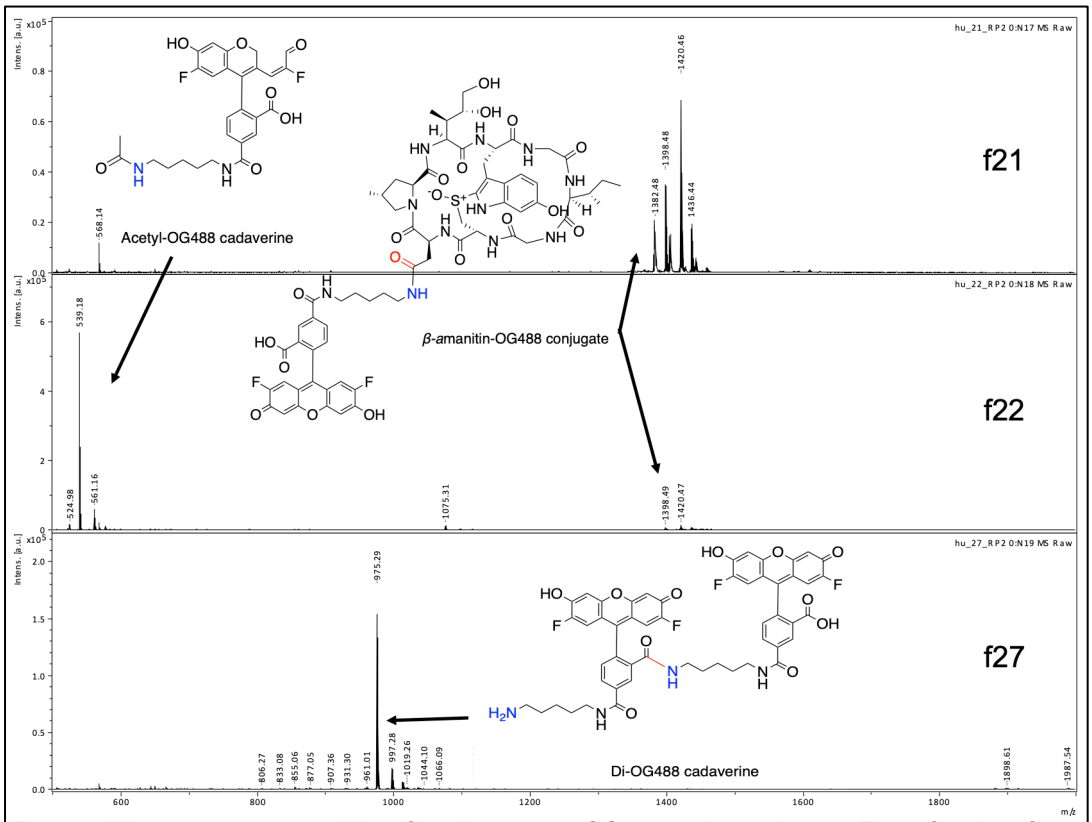


Fig. 2.5 | Mass spectrometry identifies the correct OG488-toxin conjugate. Three fractions from HPLC fractionation (f21, f22, f27) were analysed. Multiple peaks represent isotopic species.

The same scheme was applied to the far-red dye DY-654. In this case, the dye has no carboxyl group hence cannot form a self-reacted product. The reaction was fractionated by reverse-phase (RP)-HPLC, and the presence of correct conjugates confirmed again by mass spectroscopy (Fig. 2.6). The two dyes are termed AMA-488 and AMA-654, respectively.

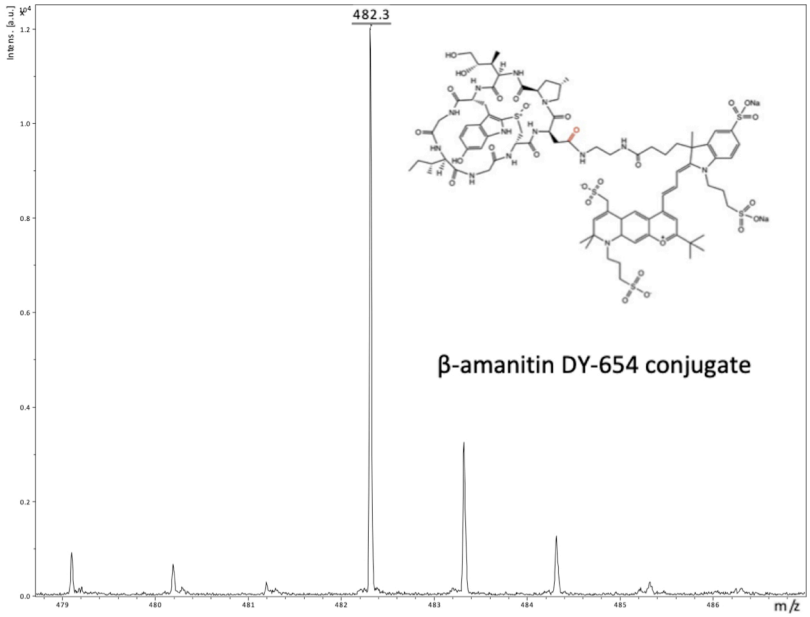


Fig. 2.6 | Mass spectrometry confirms identity of DY-654-toxin conjugate. Smaller peaks represent isotopic peaks.

2.2.3 The fluorescently conjugated amanitin molecules bind RNAPII

After fractionation by RP-HPLC, the fluorescent toxins were used to stain human foreskin fibroblast (HFF) cells to test their properties. Both 488 and 654 conjugates show a clear nuclear localisation pattern, with concentrated spots and patches where the DNA signal is correspondingly weaker, (Fig. 2.7A and B), resembling the spatial signature of immunofluorescence staining of RNAPII (Olivier-Van Stichelen & Hanover 2014).

To further validate the specificity of the conjugates, a fluorescence displacement assay was performed, where either labelled or unlabelled amanitin was preincubated with the sample firstly and then with the other to compete for the binding (Fig. 2.8). The result showed that fluorescently labelled amanitin molecule failed to be competed

off by, but also to compete off, unlabelled amanitin. This result suggested that both molecules had a very high affinity towards the same target molecule.

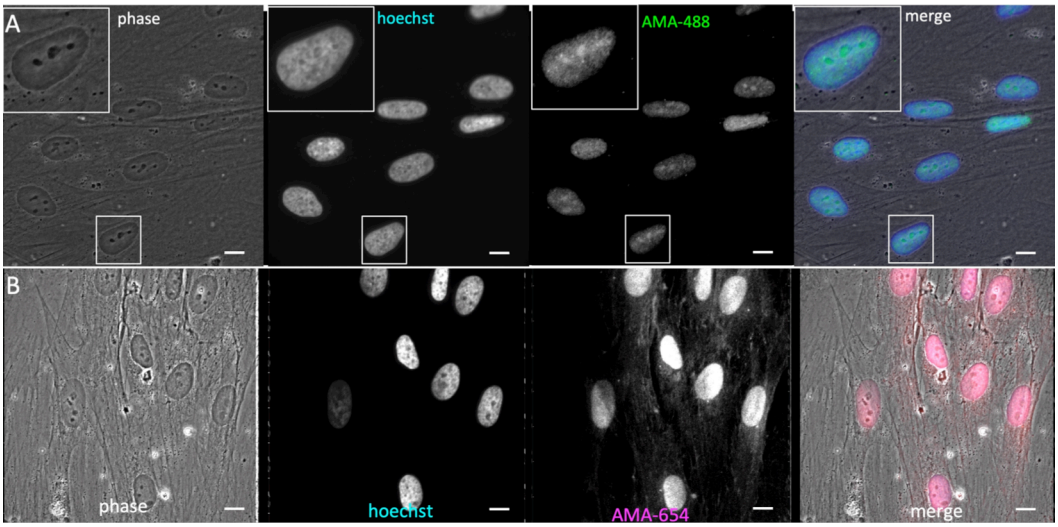


Fig 2.7 | Staining of fluorescently labelled β-amanitin on HFF cells. A, AMA-488. B, AMA-654. Scale bar represents 10 μm in all figures. Nucleus in white box is magnified on top left corner. AMA-488 and AMA-654, β-amanitin conjugates with OG488 and DY-654, respectively. Hoechst, DNA stain. Hoechst in merge images, cyan. AMA-488, green. AMA-654, magenta.

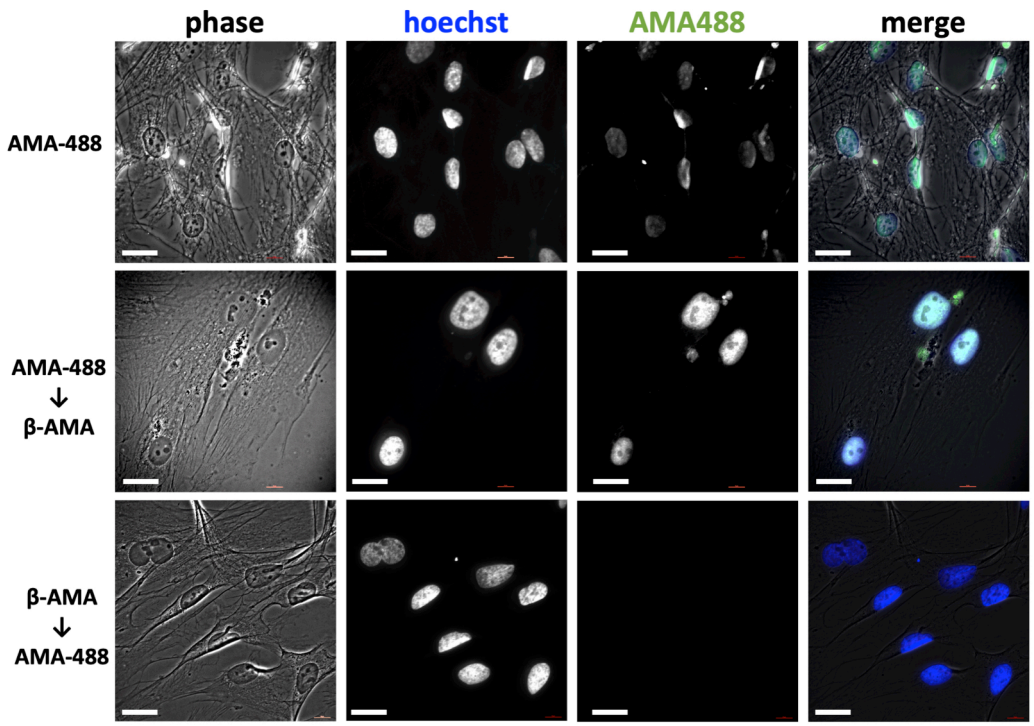


Fig. 2.8 | Fluorescence displacement assay of AMA-488 on HFF cells. Arrow represents the order of the application. Bar, 20 μm. Hoechst, DNA stain. AMA, unlabelled β-amanitin. Ac OG488, acetylated Oregon Green 488. AMA488, β-amanitin-Oregon Green 488 conjugate. Hoechst in merge images, blue. AMA488, green.

2.2.4 Fluorescent amanitin staining of *Hematodinium* dinoflagellate cells

After confirming the affinity of the conjugate, to understand the location of RNAPII in dinoflagellates, the stains were then applied to cells of the dinoflagellate *Hematodinium* sp. in culture (Fig. 2.9). Both 488 and 654 conjugates displayed clear nuclear localisation patterns. However, the AMA-488 must compete with a level of background autofluorescence from the cell in the same emission wavelength, and thus did not provide enough signal-to-noise ratio. Microscopic images taken with AMA-654 were free from this confounding background signal, and Z-series were taken for deconvolution and resolution improvement. The results seem to suggest a spatially negatively correlated distribution of RNAPII signal and the condensed chromosomes; at each optical section, the RNAPII signal surrounds but do not intrude the chromosomes (Fig. 2.9, 2.10). In addition, there seemed to be hotspots where more RNAPII signal accumulates, generating an overall scattering patchy pattern.

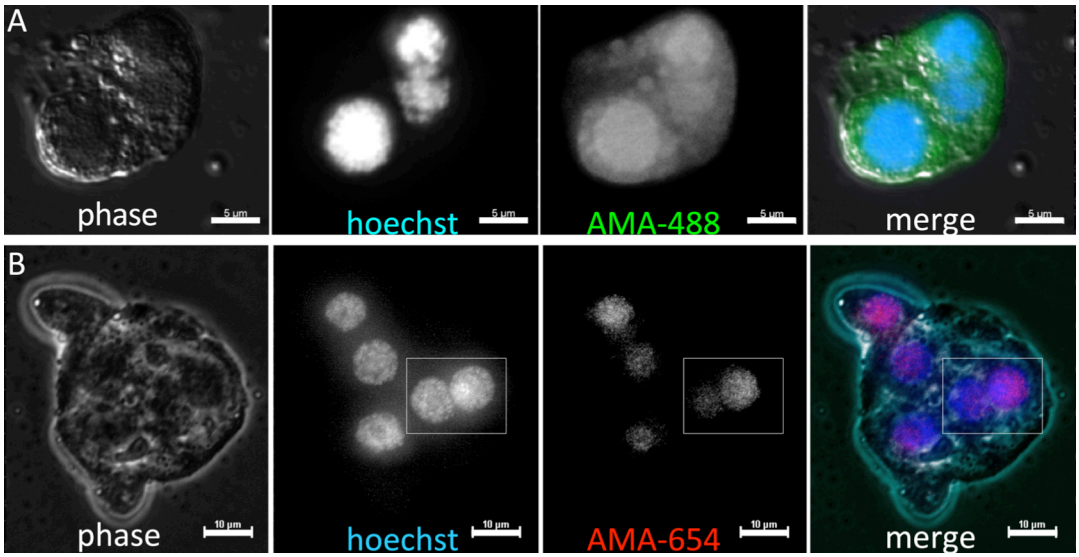


Fig. 2.9 | Staining of fluorescently labelled β-amanitin on *Hematodinium* cells. A, AMA-488. Scale bar, 5 μm. B, AMA-654. Image was deconvolved from a Z-stack series. Scale bar, 5 μm for A and 10 μm for B. Nuclei in white box are presented in Z-series in Fig. 2.9. Hoechst in merge images, cyan or blue. AMA-488, green. AMA-654, magenta.

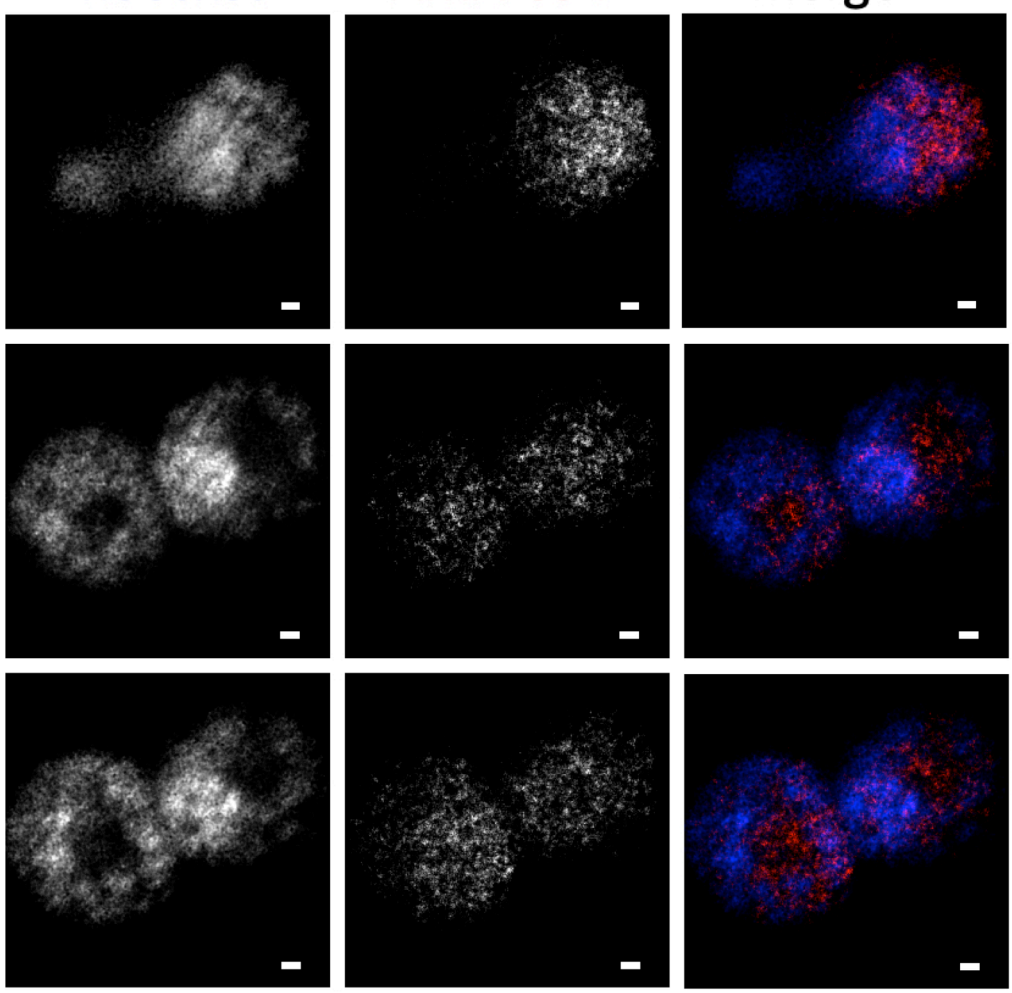


Fig. 2.10 | Three Z-series of a *Hematodinium* nuclei stained with AMA-654. Bar, 1 µm. Hoechst, DNA stain. AMA-654, β-amanitin DY-654 conjugate.

2.2.5 5-ethynyl uridine nascent RNA-labelling of *Hematodinium* dinoflagellate cells

An independent test to locate nascent RNA was performed in parallel to compare with the results I obtained by the modified amanitin. I designed an experiment to label nascent RNA with 5-ethynyl uridine (5-EU), a membrane-permeable uridine analogue which can later be chemically modified with "click" chemistry. Short labelling time was used to maximise information of the initial location of the RNA biosynthesis and its spatial relationships with the chromosomes. *Hematodinium* cells in culture medium were treated with 1 mM 5-EU for 5 minutes, then fixed with 4% paraformaldehyde, and the incorporated 5-EU labelled with the fluorophore AF 594. Microscopy was used to localise the nascent RNA signal (Fig. 2.11). The results displayed clear nuclear localisation, yet the signal extended slightly outside of the nuclei. The images were approaching the resolution of the microscope; however with z-series deconvolution, I was able to further dissect the relationships between the 5-EU and the condensed chromosomes. Within the nuclei, the 5-EU signal seemed to be spatially negatively correlated to that of the chromosomes, as demonstrated in Fig. 2.11.

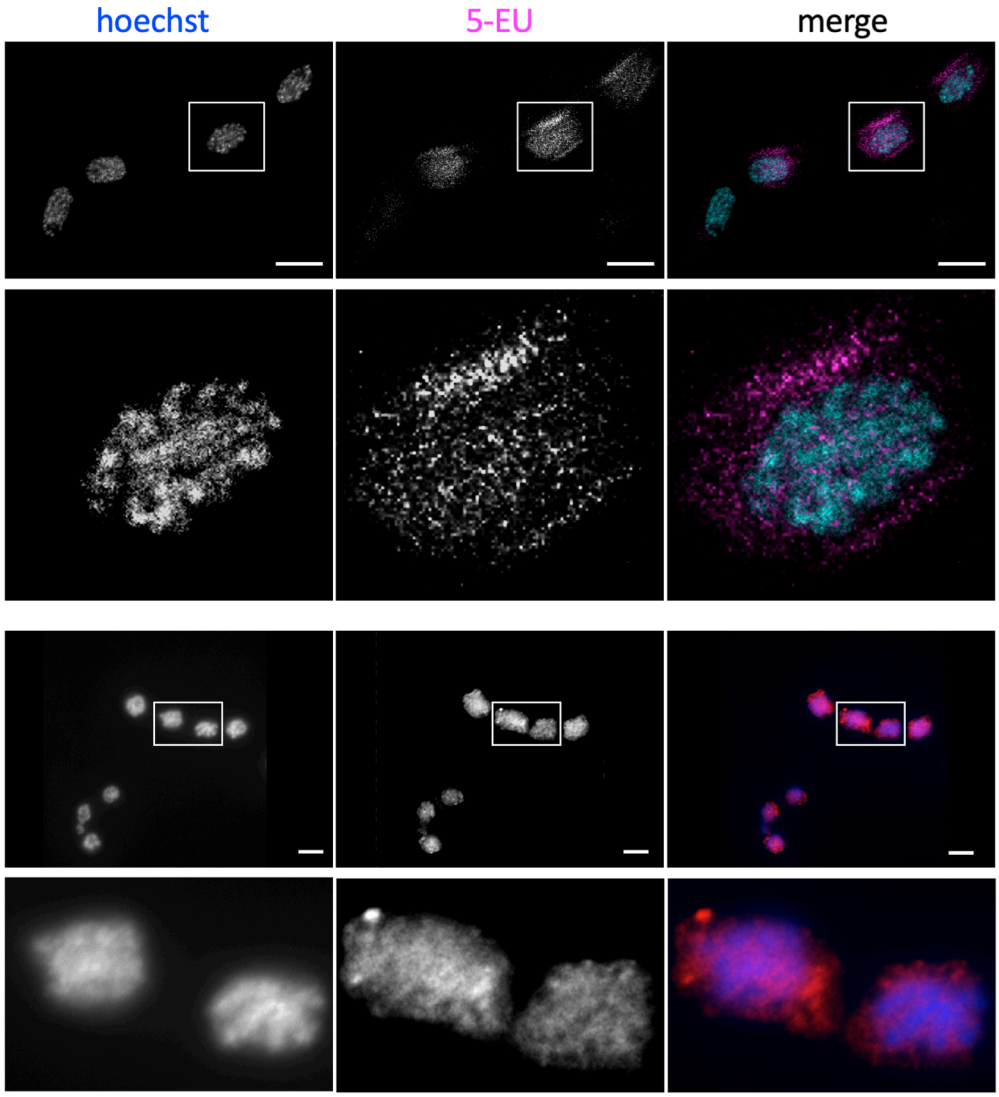


Fig 2.11 | Nascent RNA distribution in *Hematodinium* cells. The images were deconvolved from Z-series. Bar, 10 μm. Lower panels, blow up view of white boxes in the upper panels. Hoechst in merge, cyan or blue. 5-EU, magenta or red.

2.3 Discussion

Wulf's FITC-α-amanitin conjugation design utilised the phenol moiety on the tryptophan side chain (Wulf et al. 1980). The conjugation reaction was not clearly described in the publication, and any person without specific knowledge of organic chemistry will not be able to decipher the method. In this chapter, with the aid of a senior research fellow in the laboratory Dr Konstatin Barylyuk, I designed a simple one-pot conjugation of which all the four components are readily commercially available. The NHS conjugation reaction requires a total hands-on time of less than 1 hour. The mass spectroscopy results provided in this chapter confirmed the identity of the two designed molecules, AMA-488 and AMA-654. Amatoxins have extremely small Kd values around 10 pM and are very potent ligands to RNAPII (Cochet-Meilhac *et al.* 1974). The fluorescence displacement assays support that both fluorescent molecules as well as unlabelled β-amanitin have very tight binding characteristics that prevent the other to bind, supporting that the conjugate is indeed specifically interacting with RNAPII.

Dinoflagellates are not the only species that host permanently condensed chromosomes; a number of Euglenid and Cercomonad members also possess such a bizarre property (Hoppenrath & Leander 2006), as well as specific stages of animal cells, e.g. sperm cells (Derouchey *et al.* 2013; Mudrak *et al.* 2005; Talbert & Henikoff 2010). In these systems, the distance between two neighbouring molecules of DNA strands may be as close as 30 angstroms (Koltover *et al.* 2000; Raspaud *et al.* 2005), and currently it is still not known whether an antibody with a length of 120 angstrom (Noll *et al.* 1982) can be used for localisation in these systems. Under these scenarios, an antibody-based method for detection does not seem adequate. For these systems, small-molecule probes, such as the ones developed and described in this chapter, are likely more suitable. Furthermore, across eukaryotes, roughly half of all known RNAPII proteins do not bear a highly conserved heptapeptide repeat at the C-terminus (Fig. 2.12, Liu et al. 2010). The dearth of conservation would also render antibodies raised against this sequence unusable. Furthermore, many such taxa are currently not ready for genetic modification, and gene tagging is thus not an option (Waller *et al.* 2018). Besides their infamous toxicity to human and animals, amatoxins are known to inhibit RNA synthesis in at least the dinoflagellates, trypanosomatids, and ciliates, three independent lineages with equally divergent RNAPII C-terminus

(Bohatier 1981; Campbell *et al.* 2003; Rizzo & Noodén 1973). This fact alone suggests that the overall structure of the polymerase across eukaryotes is perhaps much more conserved than its sequences, and demonstrates the significance of the said structure. The conservation in structure is a reason to be optimistic that the fluorescently modified β-amanitin developed and described in this chapter, of which the probing mechanism is structure- rather than sequence-based, can be applied universally across eukaryotes, regardless of the composition of the CTD and the availability of a mature transformation system. In this chapter, I have demonstrated the correct and simple production of two small molecule (<2 kDa) probes designed for *in vitro* localisation of RNAPII suitable potentially across eukaryotes, an improvement from Wulf's study.

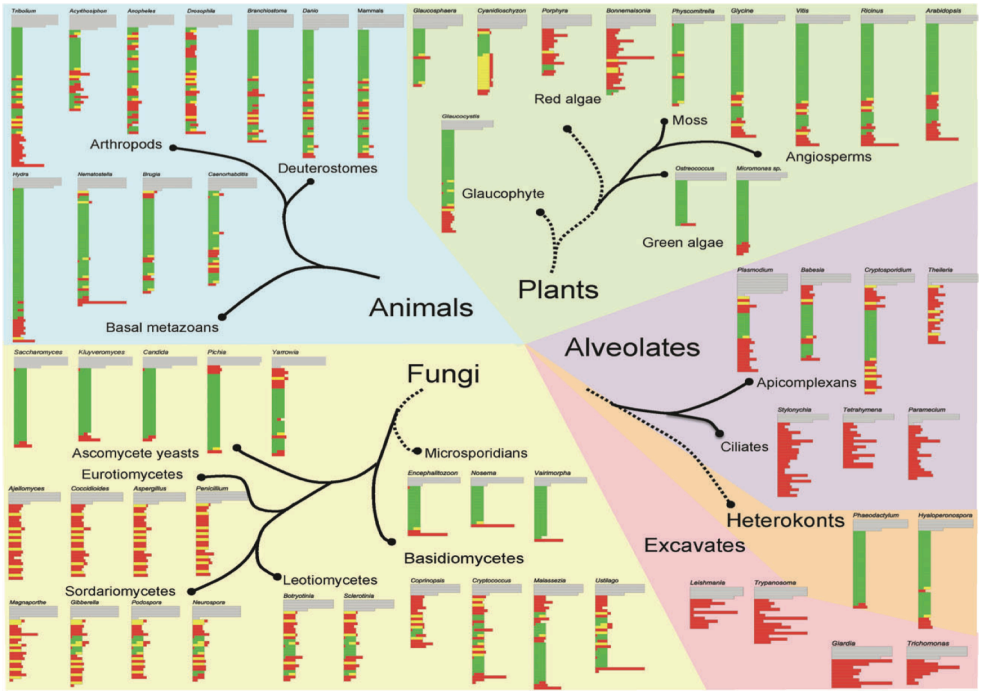


Fig. 2.12 | RNAPII CTD heptapeptide repeat conservation across eukaryotes. Reproduced from Liu et al. 2010 with permission. Green marks the presence of highly conserved repeats; yellow, medium; red, highly divergent.

This study is the first to try to locate RNA polymerase II in dinoflagellates to our knowledge, yet not the first to try to capture the location of nascent RNA. Earlier studies using H^3-labelled adenine and uridine for autoradiographic electron microscopy claimed comparable results that nascent signals accumulate at the periphery but not within the chromosomes (Fig. 2.13; Babillot 1970; Sigee 1983; Echeverría et al. 1993). The images were informative yet imprecise, with long labelling time, and image quality suffered greatly from the high variation of the sizes

of silver grains generated by using different developers. These results prompted the hypothesis that chromosomal loops on the periphery of the chromosomes are where active transcription occurs whereas the majority of the chromosomes are inaccessible and closed (Fig. 1.4; Rill et al. 1989). This system, however, requires intergenic regions and introns, if present, large enough to span the whole chromosomes, a feat that is not supported by the recently published genomic and transcriptomic data on dinoflagellates, as described in the introduction of this chapter. This inconsistency motivated me to revisit this long-standing issue with a modern method. An alternative hypothesis to the one described above is that the interior of liquid crystalline chromosomes is somehow still accessible to proteins and polymerases. Yet, results provided in this chapter are consistent with the earlier studies. For both amatoxin and 5-EU labelling, the clearly nuclear-localised signals seemed to be exclusive to that of the condensed chromosomes. I note that both experiments were near the resolution of the microscope, and even with z-series deconvolution, the stray light from above and below the focal plane could not be eliminated and caused the results to be less clear for interpretation. For the 5-EU experiment, signals can be seen around the chromosomes as well as outside the nuclei even, suggesting very fast RNA translocation towards the cytosol and justifies the short labelling time used in this study. I emphasise that the details in the images shown in this chapter, Figs. 2.10 and 2.11, are at the border of the resolution of a light microscope. In addition, although deconvolution was performed, the images may still be influenced by stray light from the planes above and below the focal plane. Quantification of co-localisation is desirable, however at this point these images are more suitable for qualitative characterisation. Further microscopy assays with higher resolution are required for quantitative analyses.

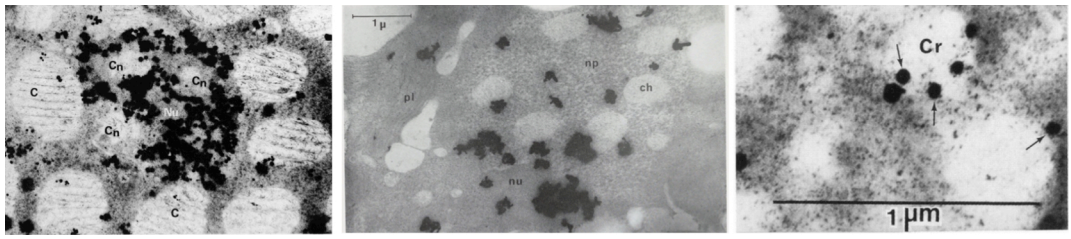


Fig. 2.13 | Earlier electron microscopic autoradiography of nascent RNA. Left, from Sigee 1983. Middle, Babillot 1970, reproduced from Bibliothèque nationale de France. Right, Echeverría et al. 1993. C, Ch, and Cr, condensed chromosomes. Black dots, radiographic signal. Cn, nucleolus. All images reused by permission.

As neither of the two hypotheses fit well with experimental data, the other logical explanation, which hasn't been explored before, is that the liquid crystalline chromosomes are in a condensed yet fluid state, and instead of moving into the interior of the chromosome, the polymerase stays anchored on the outside and draws out the template DNA as transcription proceeds, in a similar fashion to a sewing machine drawing new material in from one end and ejecting from the other (Fig. 2.14). This hypothesis is topologically similar to the transcription factories process; discrete sites in the canonical nucleus where abundant RNAPII molecules and transcription factors are concentrated and immobilised, and template DNA surrounding the factories go through the highly concentrated RNA polymerase molecules with increased transcription efficiency (Sutherland & Bickmore 2009). In both scenarios, the polymerase remains immobile, and it's the DNA template that does the movement.

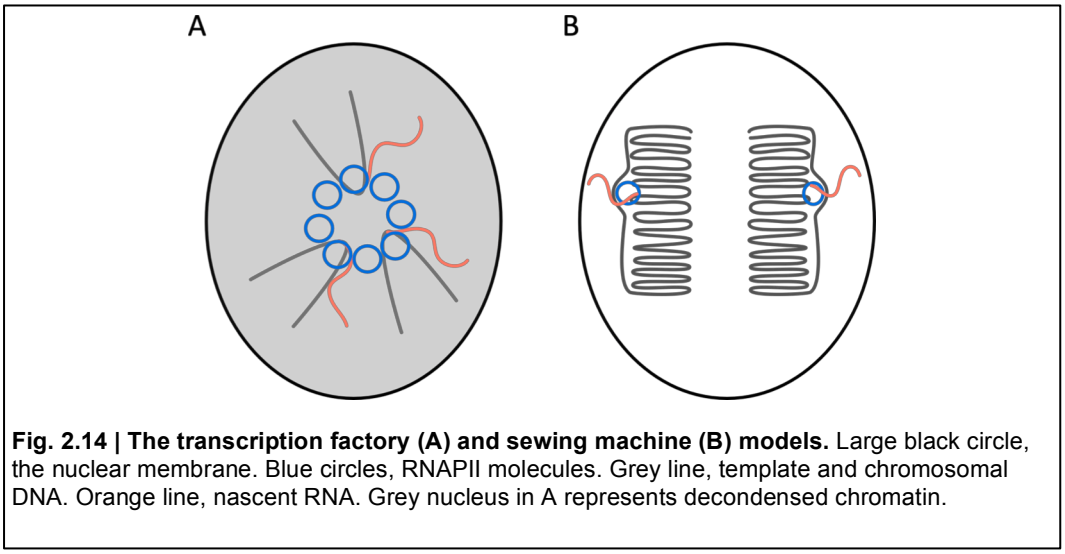


Fig. 2.14 | The transcription factory (A) and sewing machine (B) models. Large black circle, the nuclear membrane. Blue circles, RNAPII molecules. Grey line, template and chromosomal DNA. Orange line, nascent RNA. Grey nucleus in A represents decondensed chromatin.

In canonical eukaryotes, transcription factories are suggested to be associated with or even tethered to the nuclear matrix, a complex scaffold structure composed of cytoskeletal proteins, in order to remain stationary in relation to the DNA (Alfonso-Parra & Maggert 2010; Papantonis & Cook 2011; Rieder *et al.* 2012). The nuclear matrix in dinoflagellates has not been extensively explored, however, two studies provided electron micrographs of complex DNA-free web-like 3D networks inside the nucleus (Cai *et al.* 1992; Mínguez *et al.* 1994). The reproduced image, Fig. 2.15B of a DNAse-treated dinokaryon demonstrated the empty space with the silhouette of chromosomes, and curiously the DNA-free scaffold network itself seems to be reminiscent of both the amanitin and 5-EU staining signals. Furthermore, the

comparison between a canonical eukaryotic and a dinoflagellate nuclear matrix scaffold, as illustrated in Fig 2.15, demonstrated that the two are still highly similar. By making two assumptions that 1) RNAPII tethers to the nuclear matrix scaffold in a canonical nucleus, and 2) the same holds in a dinokaryon, I then make the deduction that the transition from a canonical nucleus to a dinokaryon actually did not require dramatic changes in the transcription process, despite the drastic differences in appearances of the two types of nuclei. Instead, the genome did compact into the permanently condensed chromosomes, but by preserving the fluidity of DNA in the condensed chromosomes, the dinokaryon was able to continue the use of the original transcription machinery, using the nuclear matrix scaffolds for the placement of the RNAPII molecules. Coincidentally, a mysterious "structural RNA" was proposed to have an important role in the integrity of the chromosomes, demonstrated by the chromosomes' RNAse sensitivity (Soyer-Gobillard & Herzog 1985). The identity of this RNA was never elucidated, but may be more easily put into context under the sewing machine hypothesis, albeit there is still no comprehensive explanation to the mechanistic details. This hypothesis certainly warrants more research input; a co-localisation study of the nuclear matrix scaffold with both the RNAPII and nascent RNA would be very informative, as well as live imaging of tagged RNAPII and nascent RNA, the latter currently not possible due to technical constraints.

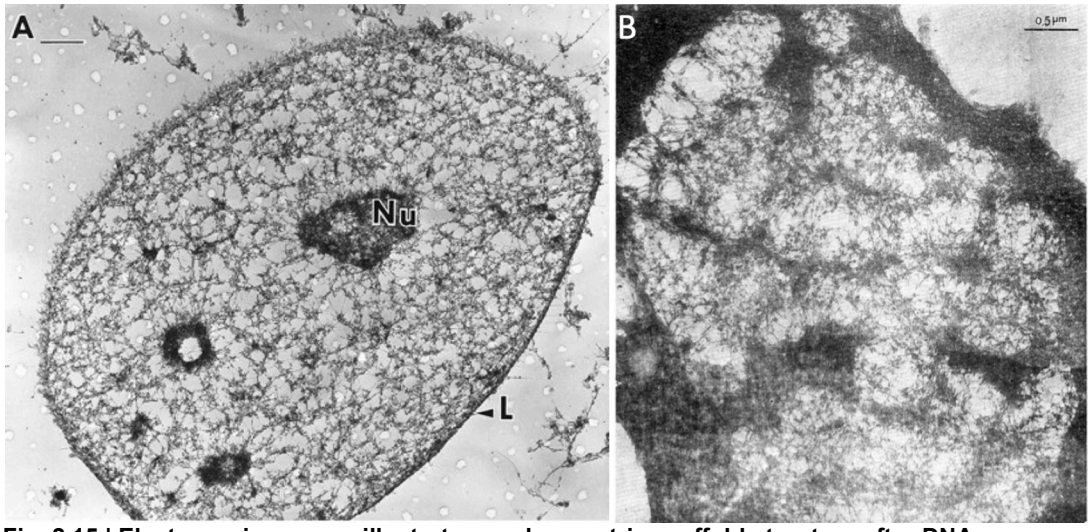


Fig. 2.15 | Electron microscopy illustrates nuclear matrix scaffold structure after DNAse digestion in a human and dinoflagellate nucleus. Left, human CaSki cell line reproduced from Nickerson et al. 2002. Nu, nucleolus. L, nuclear lamina. Copyright (2002) National Academy of Sciences. Right, *Crypthecodinium cohnii* dinoflagellate, reproduced from Cai et al.1992 with permission.

At the beginning of the chapter, I set out to ask the question, “How has the transcriptional activity been modified to cater to the drastically different chromosomal biology?” and it appears that the answer may be, not that much; the chromosomes may have changed a lot, but these changes may have been made to accommodate the already existent transcriptional machinery. The key then lies in the fluidity of DNA strands within the condensed chromosomes.

2.4 Methods

Sequence analyses

All sequence alignments, analyses, and visualisation were performed with Geneious® software suite (Kearse *et al.* 2012). Sequence alignments were performed using the muscle alignment algorithm plugin. Relevant sequences include: NCBI:KFH13943 (*Toxoplasma gondii* MAS RNAPII), NP_195305.2 (*Arabidopsis thaliana* RNAPII), NP_000928 (*Homo sapiens* RNAPII), EDV08410.1 (*Saccharomyces cerevisiae* RNAPII); MMETSP: CAMPEP_0168361026 (*Protoceratium reticulatum*), CAMPEP_0188424834 (*Glenodinium foliaceum*), CAMPEP_0186394260 (*Alexandrium tamarense*).

Synthesis of fluorescent dye β-amanitin conjugates

All reagents were dissolved in anhydrous DMF (Extra dry over molecular sieve, Acros) unless stated otherwise.10 mM of β-amanitin (abcam) dissolved in DMF (Sigma-Aldrich) was activated with freshly dissolved EDC (Sigma-Aldrich) to reach a final concentration of 15 mM, and stabilised with HOBt (Sigma-Aldrich) at the final concentration of 20 mM together at 4°C for 18 hours. Then a concentrated dye of equal molar amino-reactive DY-654 (Dyomics) or Oregon green® 488 (Thermofisher) was added to the mixture at 4°C and incubated for a further 8 hours. The samples were then purified with a Waters ACQUITY UHPLC system with a C18 reverse phase column. The final concentration of the conjugate was accessed by measuring absorbance at maximum absorbance wavelength with a nanodrop device and calculated with the extinction coefficient obtained from the vender.

Fluorescent amanitin staining assay and fluorescence displacement assay

HFF cells attached on poly-L-lysine coated glass coverslips were fixed with 4% paraformaldehyde for 10 minutes and washed. For *Hematodinium* sp., cells were concentrated by centrifugation at 300 g for 10 minutes, washed with Nephrops saline, and fixed with 1/3 volume of 16% paraformaldehyde solution to reach a final of 4% for 10 minutes. Then the cells were immobilised on poly-L-lysine coated glass slides. Both samples were then perforated with 0.1% Triton X-100 in PBS for 10 minutes and washed. Cells were blocked with 2% BSA in PBS for 30 minutes, and stained with 100 µm of fluorescent or native amanitin dissolved in 2% BSA. Cells were washed and counter-stained with Hoechst 33342 for 30 minutes. HFF cells on

coverslips were mounted with Fluorsave® (Merck Millipore) on glass slides and *Hematodinium* cells were mounted with #1.5 coverslips for microscopy.
Nephrops saline: NaCl 27.99 g/L, KCl 0.95 g/L. $CaCl_2$ 2.014 g/L, $MgSO_4$ 2.465 g/L, Na_2SO_4 0.554 g/L, HEPES 1.192 g/L, HCl to pH 7.8. Filter sterilised.
PBS: NaCl 8.0 g/L, KCl 0.2 g/L, Na_2HPO_4 1.44 g/L, KH_2PO_4 0.24 g/L.

5-Ethynyl Uridine staining assay

40 mL of late log phase *Hematodinium* cells were concentrated to 4 mL of its original medium by passing through a 50 mL syringe coupled to a 5 µm disk filter by gravity and back flushing, and incubated with 1 mM of 5-EU (Click Chemistry Tools) for 5 minutes. Cells were then fixed with 4% paraformaldehyde for 10 minutes, quenched with a drop of 1M Tris-HCl pH 7, washed, and immobilised on poly-L-lysine coated glass slides. Cells were perforated with 0.1% Triton X-100 for 10 minutes. Cells were then developed with 100 mM HEPES pH 7.5 (DEPC-treated) (Sigma-Aldrich), 10 µM AF594 azide (Click Chemistry Tools), 1 mM $CuSO_4$ (Sigma-Aldrich), and 100 mM freshly dissolved ascorbic acid (Sigma-Aldrich). Cells were finally counter-stained with Hoechst 33342 and mounted with Fluorsave and #1.5 coverslips for microscopy.

Microscopy

All images were taken with a Nikon Eclipse Ti system with a 100X oil objective, and Hamamatsu ORCA-Flash 4.0 C11440-22CU CMOS sensor. Phase contrast and fluorescence images were taken with the sensor. Fluorescence images were illuminated with CoolLED pE-4000 illumination system and standard single channel fluorescent cubes. Images were post-processed and deconvoluted with Nikon NIS-Element Advanced Research software suite.

Chapter 3: *In silico* analysis of DVNP proteins across dinoflagellates and marine viruses

3.1 Introduction

As illustrated in Chapter 1, the dinoflagellate nuclear protein DVNP has no homologs in any other eukaryotes, yet is found in several members of Phycodnaviridae, a group of marine large DNA viruses (Gornik et al. 2012). Its gain coincides with the early evolution of dinoflagellates and the highly derived characteristics of the dinokaryon, namely permanently condensed chromosomes, loss of nucleosomes, expansion of genomes, and lower protein:DNA ratio in the nucleus. Since the first publication of the discovery of such genes in viruses (Gornik et al. 2012), more large scale sequencing data, e.g. Tara Ocean expedition (Malviya et al. 2016) and the MMETSP project (Keeling et al. 2014), along with numerous metagenomic projects in several environments have become available. A more contemporary and comprehensive survey is hence possible, to provide more insight into both the distribution and evolution of the gene and the dinoflagellates.

3.2 Results

3.2.1 Conservation of DVNP across dinoflagellates and viruses, and structural information from conservation

To study the presence of this gene across the tree of life, during this PhD study, DVNP genes were collected from publicly available databases. As compared to earlier results by Gornik et al. at 2012, only a handful of new viral sequences emerged, most previously reported Phycodnaviruses. However, recent studies suggest that one particular species, *Organic Lake phycodnavirus* (ADX05902.1), has been reclassified to an early branching group of Mimiviridae, the OLPG group, later proposed as a subfamily by the name Mesomimivirinae (Gallot-Lavallée et al. 2017). Other instances of DVNP-hosting members of this group include *Phaeocystis globosa virus* (YP_008052414.1), *Chrysochromulina ericina virus* (YP_009173330.1), and

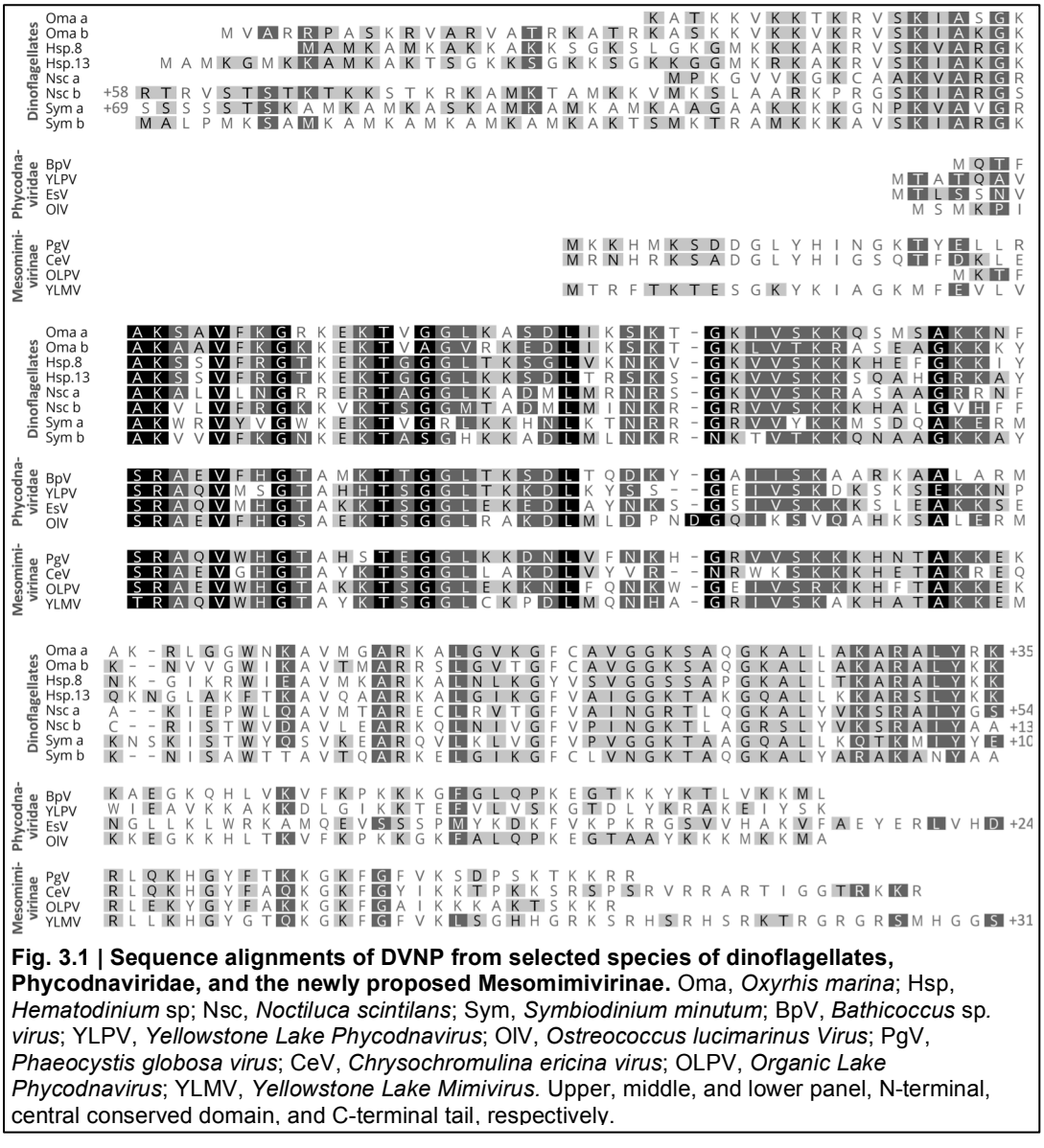


Fig. 3.1 | Sequence alignments of DVNP from selected species of dinoflagellates, Phycodnaviridae, and the newly proposed Mesomimivirinae. Oma, *Oxyrhis marina*; Hsp, *Hematodinium* sp; Nsc, *Noctiluca scintilans*; Sym, *Symbiodinium minutum*; BpV, *Bathicoccus* sp. *virus*; YLPV, *Yellowstone Lake Phycodnavirus*; OlV, *Ostreococcus lucimarinus Virus*; PgV, *Phaeocystis globosa virus*; CeV, *Chrysochromulina ericina virus*; OLPV, *Organic Lake Phycodnavirus*; YLMV, *Yellowstone Lake Mimivirus.* Upper, middle, and lower panel, N-terminal, central conserved domain, and C-terminal tail, respectively.

Yellowstone Lake Mimivirus (YP_009174133.1). A representative alignment of DVNP sequences of 4 species of dinoflagellates and each of the viral families is shown as Fig. 3.1.

DVNP protein sequences across dinoflagellates and both viral families (Fig. 3.1) are conserved in the core domain about 50 amino acids long, which contain many positively charged residues as well as hydrophobic residues. Dinoflagellate DVNPs have a highly positively charged N-terminal tail, which is present in neither of the virus families. After the highly conserved core domain, all DVNPs have a C-terminal tail of various lengths. The tail shows no sequence conservation between groups, although within each clade conservation can be seen. It also appears that the dinoflagellate C-terminal tail has an extra conserved stretch of about 30 amino acids long after the core domain. One clearly observable trait of the C-terminal tail across all clades is the widespread presence of positive charges, ie arginines and lysines, yet the locations of these charged residues do not appear to be conserved. To further understand the phylogenetic relationships between the genes, attempts were made to generate phylogenetic trees of all the sequences with several methods including maximum likelihood and maximum parsimony, however due to the shortness of the conserved region, phylogenetic trees with high confidence cannot be built, instead many branches are short with very low node scores. For an unrooted maximum likelihood tree with 50 bootstrap tests, although all viral sequences were grouped together, most basal nodes has a bootstrap value of less than 2 (Fig. 3.2), and overall the tree does not seem to be reliable.

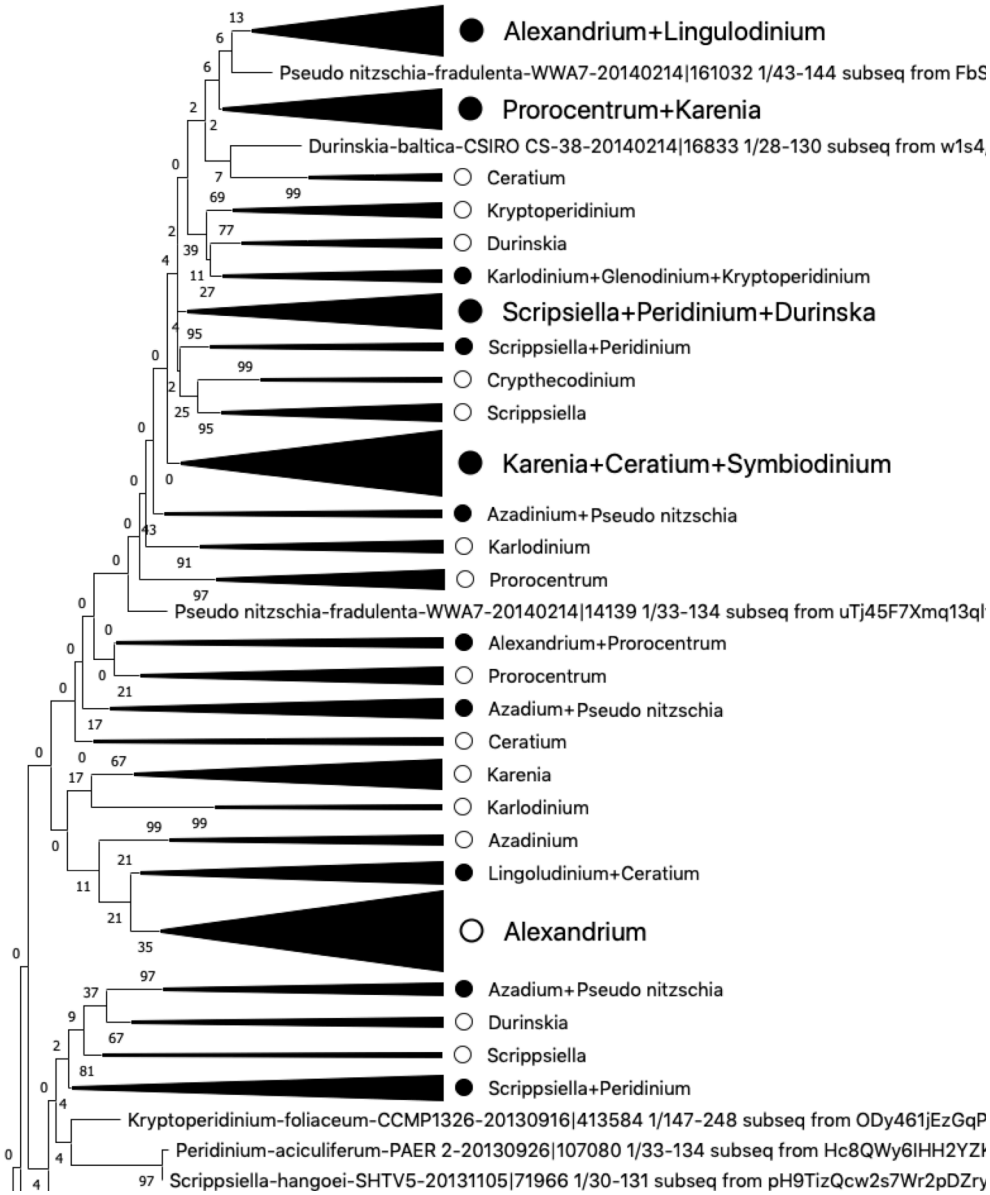


Fig. 3.2 | Maximum likelihood phylogenetic tree of dinoflagellate and viral DVNP protein sequences. Node values are bootstrap values. Monophyletic branches were collapsed and marked as blank circles. Several polyphyletic branches were collapsed for presentation, and are marked as filled circles.

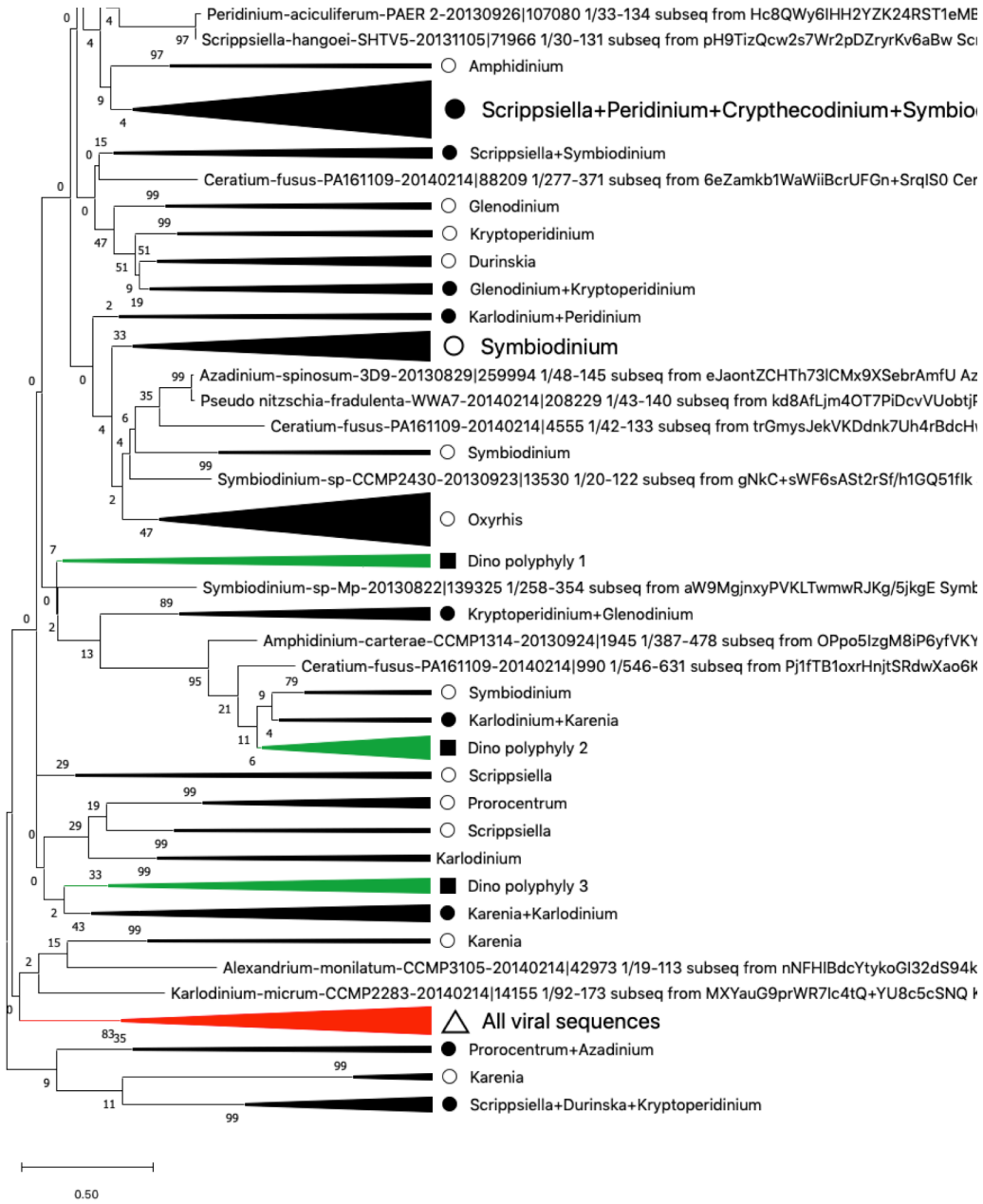


Fig. 3.2 (cont.) | Maximum likelihood phylogenetic tree of dinoflagellate and viral DVNP sequences. Node values are bootstrap values. Monophyletic branches were collapsed and marked as blank circles. Several polyphyletic branches were collapsed for presentation, and are marked as filled circles. All 12 viral sequences were grouped in one branch, which is marked with red branch colour and an empty triangle. A total of three highly complex dinoflagellate DVNP branches were also collapsed for better presentation effects. These are marked with green branch colour.

In order to understand sequence conservation within and between dinoflagellates and viruses, I employed Gremlin, an algorithm that simultaneously captures conservation and amino acid coevolution through hidden Markov model (HMM) multiple alignments. In August 2018 a pre-published large-scale metagenomic virus sampling dataset became available. In the database roughly 150 complete DVNP sequences of viral origin can be identified. In addition, dinoflagellate DVNP sequences were retrieved from all the publicly available databases, including MMETSP, and a total of 748 sequences were obtained. The two datasets were aligned and visualised with Jalview (raw sequences in Appendix 1), trimmed to comparable lengths, and then submitted to Gremlin separately (http://gremlin.bakerlab.org/) (Kamisetty et al. 2013). Fig. 3.3 and Fig. 3.4 shows logograms and gremlin contact maps of both datasets. Due to the nature of the multiple alignment algorithms, logograms presented here contain extra positional information. Conserved residues across sequences with various lengths in between are presented as logos and spaces of different widths. Large spaces between narrow logos indicate high variation in length between conserved amino acids. Comparison of both logograms shows a core domain with presence and positions of several neutral and hydrophobic residues highly conserved among and between both dinoflagellates and viruses. Further information can be extracted from the C-terminal tails of both logograms: there seems to be more conservation of existence of charges rather than the location of these charges, a phenomenon much more prominent in the viral sequences, evident by the spaces between and the width of these positively charged residues, whereas in the dinoflagellates the primary sequences are much more conserved. In addition, a phenylalanine is present in the C-terminal tail of both DVNPs, and an additional tryptophan and tyrosine in the

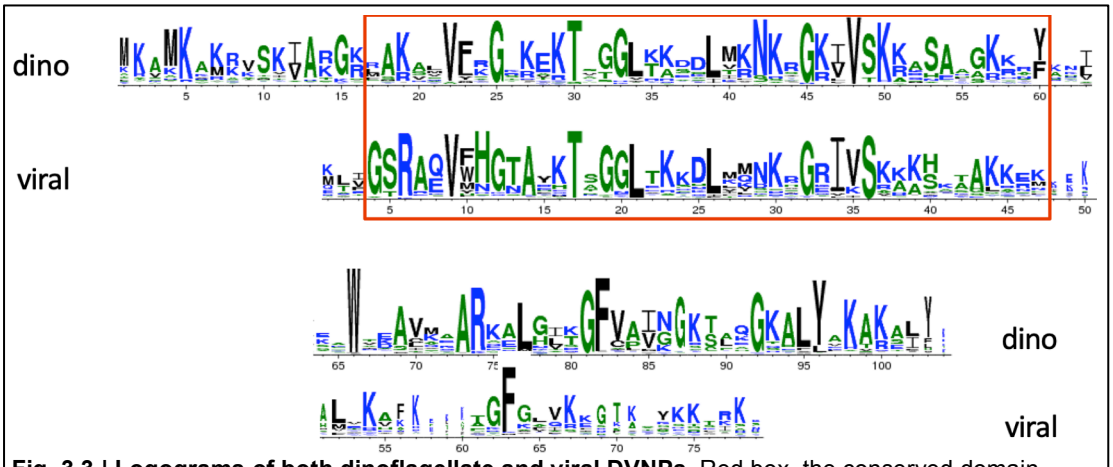


Fig. 3.3 | Logograms of both dinoflagellate and viral DVNPs. Red box, the conserved domain.

dinoflagellates. Note all three (Phe, Trp, Tyr) are flat and hydrophobic residues, suggesting a potential function of the protein that requires a flat or hydrophobic moiety.

I also employed Gremlin to try to understand potential structural elements of DVNPs and to compare DVNPs from dinoflagellates and viruses. A Gremlin contact map also bear structural information: as residues in contact in 3D space will often co-evolve together due to their cognate interaction, these co-evolved residues can be used to infer contacts in space, or even the 3D structure of the protein if enough high confidence contacts are present. In the contact maps (Fig. 3.4), predicted contacts are presented as a dot with the interacting amino acids as the axes. Both DVNPs have several high confidence contacts presented as large dark blue dots, suggesting the presence of certain stable structures, at least in the N-terminal half. Many contacts are close to the diagonal line, presenting local structural features, but not many long-range interactions are found. For the dinoflagellates' plot, the very N-terminal end is not very interactive, but the conserved core domain right after is seen to interact within itself much more than the rest of the protein, resulting in an interaction box (Fig. 3.4A, green box). In the centre of the plot, the straight lines eleven amino acids long parallel to the diagonal is strongly indicative of the presence of a helix (brown box). The C-terminal half, similarly to the N-terminal domain, shows prominent intra-domain contacts rather than long-range ones (pink box). For the viral plot, numbers of high confidence contacts are clearly less than their dinoflagellate counterpart. Although the overall trend is similar to the dinoflagellate contact map in that the N-terminal and C-terminal halves have much more intra-domain contacts, the plot carries much less interpretable information.

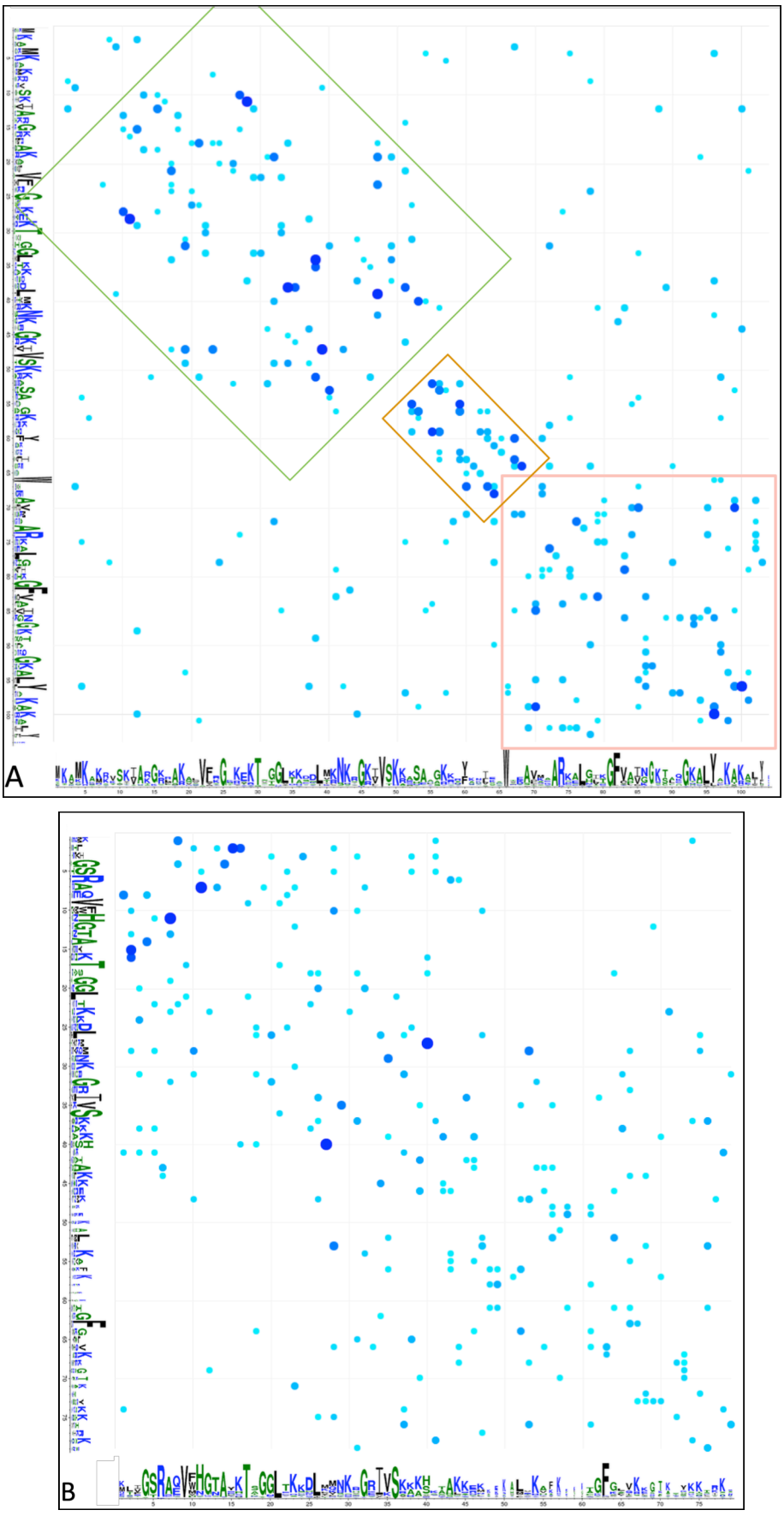


Fig. 3.4 | Gremlin co-evolution analyses contact map of DVNP. A, dinoflagellate sequences. B, viral sequences. Each dot represents a predicted contact, the darker the colour is the higher the confidence of the prediction. Boxes are explained in text above.

3.2.2 Charge, charge concentration, and charge distribution analysis

DVNP is highly positively charged with averaged pI value of 11. The negative charges on the phosphate backbone of DNA and the positive charges on histone proteins and their distribution are known to be important for their binding (Annunziato 2008; Fenley et al. 2010). I reasoned that the charges, their concentration, as well as distribution in DVNP, also a DNA-binding protein (Gornik et al. 2012), may have similarly important roles. In addition, charge and charge

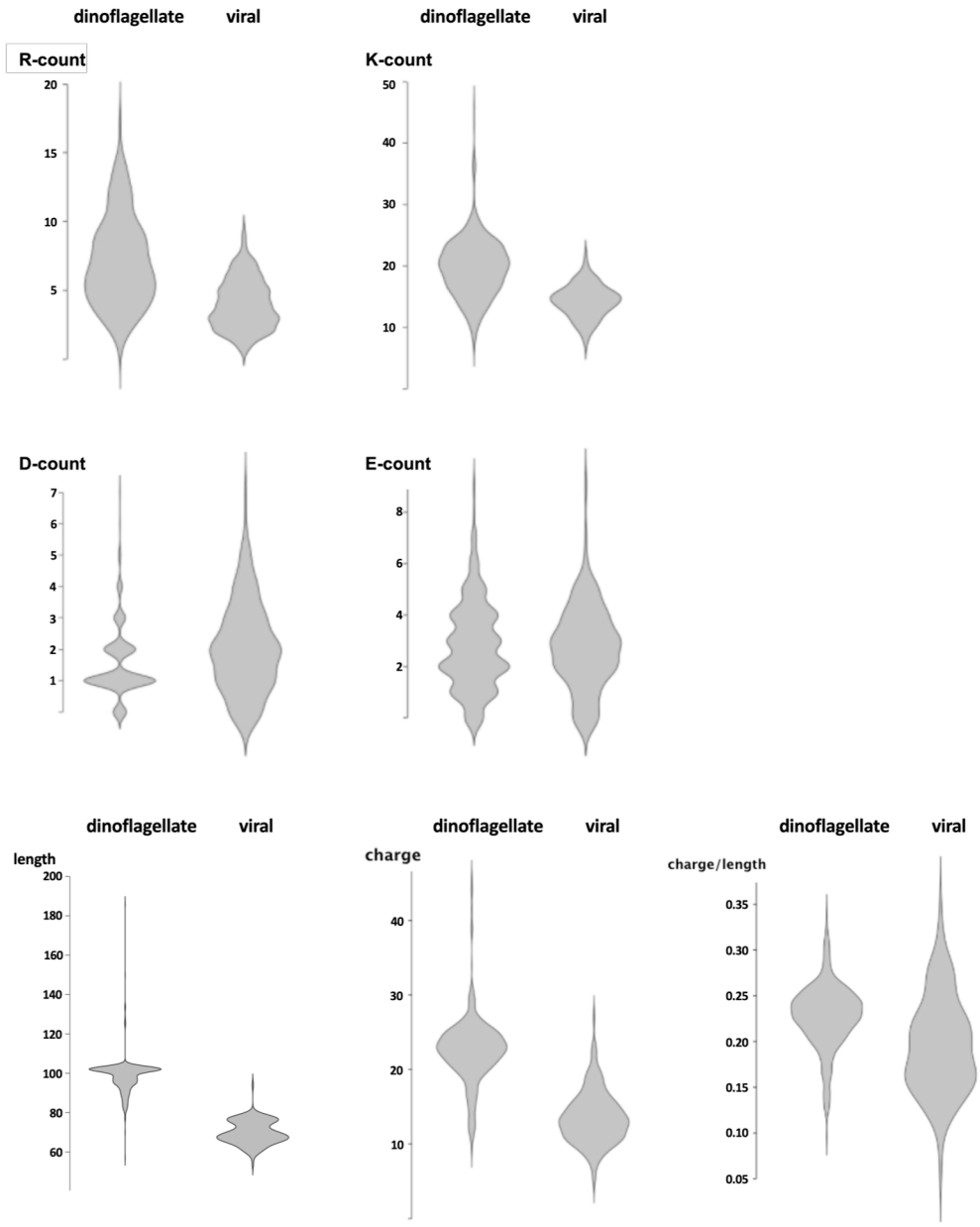


Fig. 3.5 | Charged amino acids, total length, total charge and charge/length distribution of dinoflagellate and viral DVNP proteins. R-count, K-count, D-count, E-count, number of lysine, arginine, aspartate, glutamate residues in the sequences.

distribution, both traits uncharacterised by primary sequences alone and were not captured in the conservation logograms, may also provide hints of how the gene DVNP has evolved. To understand the relationships of charge and the protein DVNP, I then performed charge analysis on all DVNP proteins globally (Fig. 3.5 & Table 3.1). All results presented here are comparisons of pooled dinoflagellate versus pooled viral DVNP sequences, with 748 and 161 sequences in each dataset. My results suggest that dinoflagellate DVNPs have higher variation in total protein length, as demonstrated by the long tails of the two relevant graphs in Fig. 3.4 and the standard deviation in Table 3.1. Dinoflagellates have on average higher counts of positively charged residues (lysines and arginines) and more individuals with higher counts, resulting in long tails in the violin plot and a higher skewness, i.e. less symmetrical distribution of values. For negatively charged residues (aspartates and glutamates) there are fewer differences between the two pools (Fig 3.5 & Table 3.1). For overall charge, the shape of distribution between the two seems to be similar although dinoflagellate sequences have a higher mean charge value, however the dinoflagellate pool again has a long tail skewing the dataset. In charge/length measurements, dinoflagellate DVNPs are overall marginally higher than their viral counterparts, however surprisingly the values are much more tightly clustered together with smaller standard deviation, and the tail causing the skewness in positively charged residues, sequence length, and net charge is not present (Fig. 3.5 & Table 3.1), indicating that the individuals with more charges are the longer ones as well.

	R count		K count		D count		E count		length		charge		charge/length	
	dino	viral	dino	viral	dino	viral	dino	viral	dino	viral	dino	viral	dino	viral
Mean	7.099	4.000	19.640	14.270	1.487	2.037	2.755	2.720	98.440	69.83	22.50	13.51	0.228	0.195
Std. Dev	3.243	1.803	4.769	2.711	1.037	1.418	1.665	1.582	7.943	5.805	4.411	3.479	0.035	0.053
Skew	0.537	0.460	0.684	-0.055	1.351	0.625	0.586	0.336	2.194	0.357	0.816	0.725	-0.457	0.335
Min.	0.00	1.00	7.00	7.00	0.00	0.00	0.00	0.00	57.00	54.00	9.00	5.000	0.096	0.053
Max.	18.00	9.00	46.00	22.00	7.00	7.00	9.00	9.00	186.00	94.00	46.00	27.00	0.341	0.350

Table 3.1 | Charge and charge distribution analyses and statistics of dinoflagellate and viral DVNP proteins. Skew, skewness. Min., minimum individual. Max., maximum individual.

In addition to overall charge, the placement of the charges is known to affect the overall conformation of a peptide (Das & Pappu 2013). To compare the charge

distribution on both dinoflagellate and viral DVNPs , I then performed a κ analysis. Das and Pappu constructed a parameter, κ ($0 \leq \kappa \leq 1$), as a numerical quantification of how well-mixed or homogenous oppositely charged amino residues in a particular peptide are placed (Das & Pappu 2013). Based on this parameter, structural aspects of the candidate peptide can be inferred, e.g. hairpin-like conformations where both ends with opposite charges attract each other (very high κ) or random coil ensemble IDPs (intrinsically disordered proteins) with very homogeneous mixing of both charges along the peptide (very low κ). When the κ value of dinoflagellate and viral DVNP protein sequences were calculated and plotted, dinoflagellate sequences have a much tighter spread of κ value, despite being larger in total number of sequences, and higher variation in length and charge (Fig. 3.6), similar to the charge/length analysis.

Das and Pappu further present a method to predict the structuredness of an IDP, based on the relation of fractions of the positively and negatively charged residues, ie (R+K) or (D+E) / sequence length, termed the Das-Pappu plot. While it is still unclear whether DVNP is an IDP or not, the plot Das and Pappu developed provides a useful tool to visualise the relationships between the positive and negative charges in each individual protein as well as overall charge/length, which can be inferred as the distance from a data point to the origin point. When viral DVNPs are plotted on Das-Pappu plot, a significant portion of proteins are loosely clustered within the borders of cyan region, where some outliers range into the center of the dark green as well as light green regions. With the dinoflagellate protein sequences, however, the majority tightly clusters in the grass green region and none of the data points wander into the dark green region as far as some of their viral counterparts do (Fig. 3.7), very much in accordance with the κ-value analysis and charged/length results.

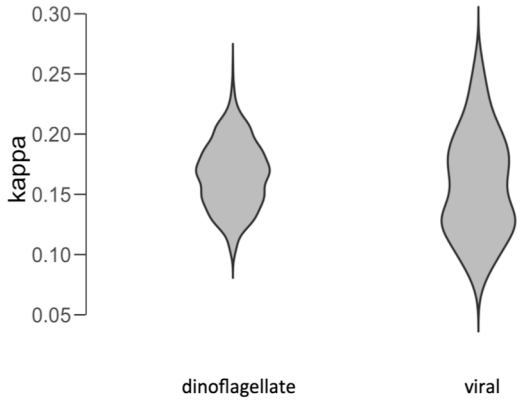


Fig. 3.6 | Distribution of κ-values of DVNP from dinoflagellate and viral sources.

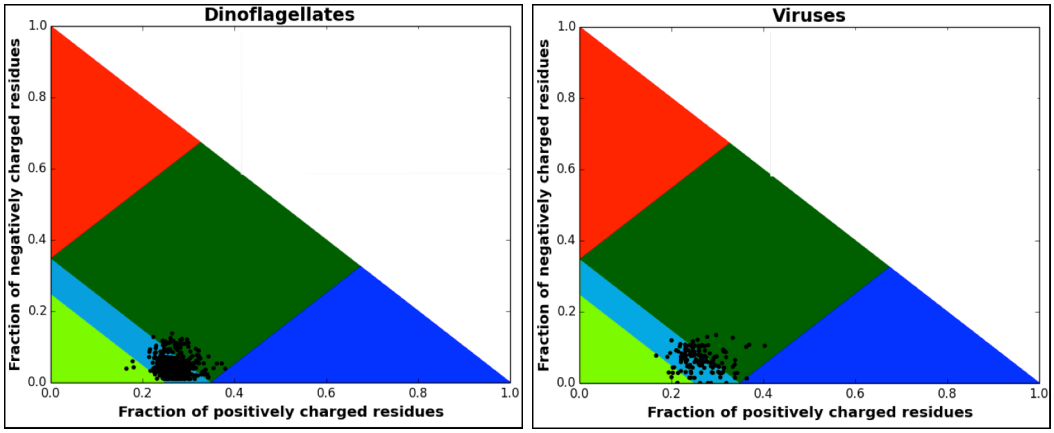


Fig. 3.7 | Das-Pappu plot of DVNPs from dinoflagellates and viruses. For an IDP, position in each colour is predicted to have the conformation of the following; Green, weak polyampholytes: globules and tadpoles. Cyan, collapsed or expanded depending on context. Olive, coils, hairpins & chimeras. Red and blue, negatively/positively charged strong polyelectrolyte: swollen coils.

3.3 Discussion

It appears that increasing amount of evidence are supporting that many large DNA viruses, especially Phycodnaviridae and Mimiviridae and its extended sister group Mesomimivirinae, are related (Fig. 3.8) (Gallot-Lavallée & Blanc 2017; Yutin et al. 2014; Zhang et al. 2015). The fact that the conserved gene DVNP is found in multiple separate clades in the large group suggests that it is likely that DVNP may have been an ancestral trait before the viruses radiated. However, there is the other possibility that the gene originated from Phycodnaviridae, but was horizontally transferred into a deep-branching OLPG member, possibly through an exchange event inside a co-infected host. In both cases, the most likely scenario of dinoflagellates acquiring this gene would be a failed or somehow ameliorated infection event between a DVNP-containing virus and an early dinoflagellate host. DVNP may have served as either a DNA management system for the large DNA virus, or a viral effector to weaken the host (Gornik et al. 2019). Either way, the host accommodated the gene, and adopted an entirely different DNA packaging method. The present study uses a large number of sequences from a metagenomic study targeting viruses. Although the source of some of the metagenomic sequences are not known, the implication is that the gene DVNP is probably more widely spread throughout the viruses than we supposed, and that the gene may have a much more ancient history in the viruses.

A robustly supported phylogenetic tree of DVNPs would provide much insight into the potential lateral gene transfer event when the gene DVNP was acquired by the dinoflagellates, and can also aid as a marker for the phylogeny of dinoflagellates, Mesomimivirinae viruses, and Phycodnaviruses. However, the conserved region of the protein is only approximately 50 amino acids long, and within the conserved region variance was too low to provide good support for any type of phylogenetic discrimination trialled in the present study. A maximum likelihood is shown (Fig. 3.2) in which all the known viral DVNP sequences cluster as a branch away from most dinoflagellate DVNP sequences, however as the bootstrap values were 0 in virtually every node close to the root. The fact that the viral sequences all branch away could possibly have been the result of all the viral sequences being shorter than their dinoflagellate counterparts. No realistic and confident conclusion could be drawn.

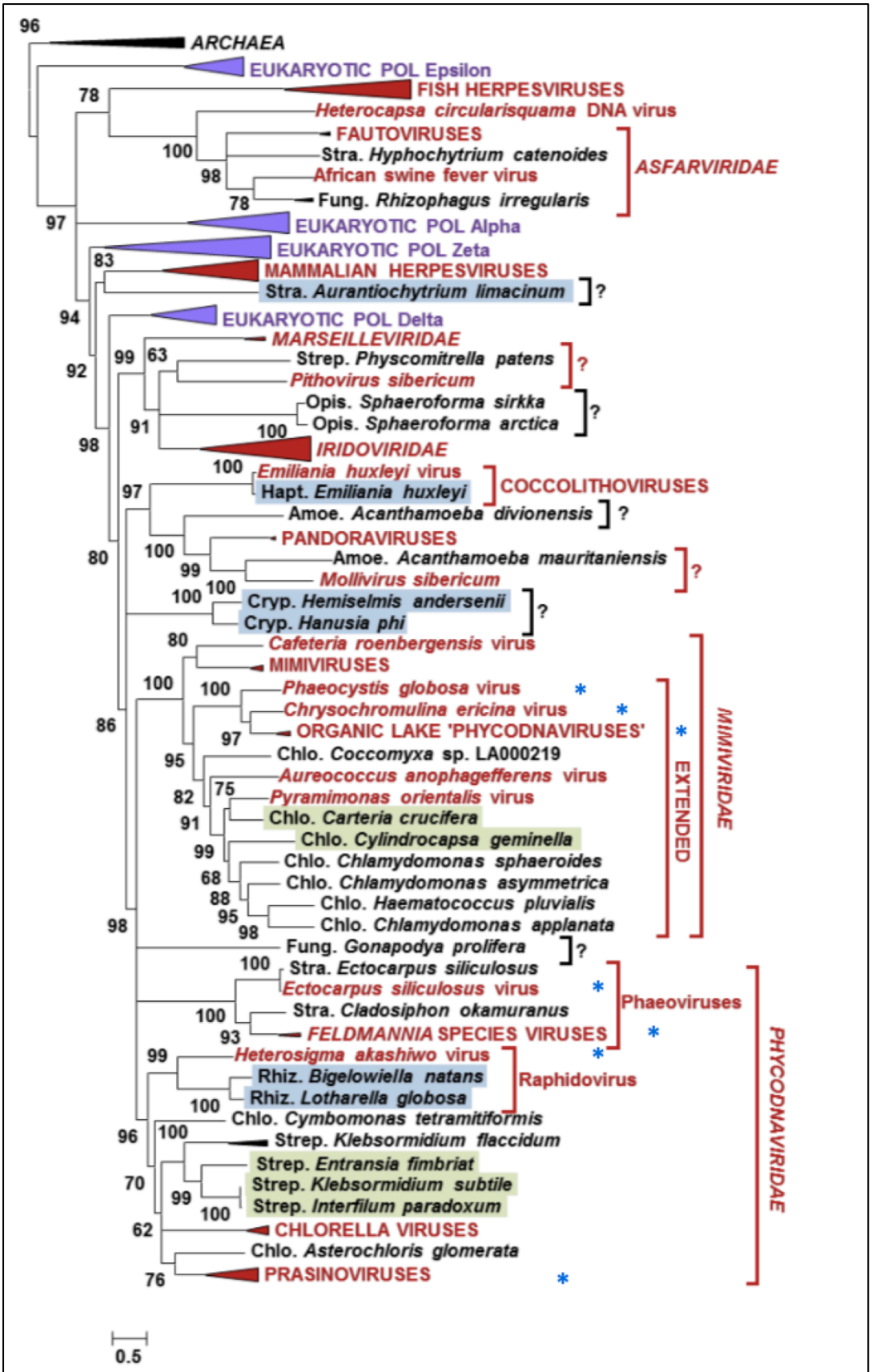


Fig. 3.8 | Maximum likelihood phylogenetic tree of DNA polymerase proteins. Reproduced from Gallot-Lavallée & Blanc 2017, (CC 4.0). Blue asterisks mark confirmed presence of DVNP. Numbers at the node presents bootstrap values.

By comparison, the viral DVNP logogram as well as contact map (Fig. 3.3 & 3.4) provided much lower conservation and interpretable information than their dinoflagellate counterpart, presumably because the lower number of sequences and the subsequent lower statistical power. However, I note that this cohort very likely contains members from both Phycodnaviridae and Mesomimivirinae. If the two pools have diverged and evolved different specific features, these diverged features would likely not have been seen in the combined analysis as the analysis reported on common features only. The relevant analysis would benefit from an even larger sample size and more sophisticated classification tools.

The presence of several conserved aromatic residues in the C-terminal tail may present insights into the protein. These include a GF motif that is in the centre of the C-terminal tail in both dinoflagellate and viral DVNP, and a tyrosine further down the C-terminal end, less conserved in viral sequences, in addition to another tryptophan in dinoflagellates. In the case of DNA-interacting proteins, a category which DVNP is indeed a member of, it has been demonstrated that aromatic residues may play important roles through cation-π or π-π stacking interactions with DNA (Sathyapriya & Vishveshwara 2004; Anjana et al. 2012; Wilson et al. 2014), suggesting possible protein-DNA interaction possibilities or even intercalation events.

The gremlin co-evolution analysis suffers a disadvantage: since the amino acid contacts prediction is based on statistical analysis of amino acid polymorphism, in several highly conserved amino acids, presumably the residues are so crucial that no mutations or variations are tolerated in the sequences, no correlation drawn and no prediction made. Examples here include the VFKG motif in the N-terminus, 48V in dinoflagellates, and the conserved phenylalanine described in the last paragraph. Nevertheless, the Gremlin analyses did indicate that there are indeed conserved structural elements, particularly within the core region of the protein, and conserved interactions within the termini of the protein. Overall, the dinoflagellate proteins appear more conserved in this 3D interpretation than the viral proteins, although this could be due to the heterogeneity of the viral samples, as mentioned above

The charge distribution results provide interesting implications. The dinoflagellate DVNPs have a much larger sample size (748 vs 161 viral proteins), a genetic

background that has the tendency to expand within the population (Bachvaroff & Place 2008), and much higher variation in size and in total charges. Despite these traits, they have surprisingly tightly clustered charge/length value and the distribution of charges (κ value), as illustrated in figures 3.3 and 3.4. Similarly, Fig. 3.7 demonstrates that dinoflagellates DVNPs are more homogeneous in characteristics of overall ratio of negative and positive charges as well as charge/length, despite the presence of many individuals with very long N- or C-terminal tails. This implies that the protein has been under a particular selection pressure to keep the overall charge distribution and charge/length constant, yet the length of the protein not. It is likely that this pressure is linked to its function *in vivo* in dinoflagellates, and the dinoflagellate proteins' properties appears to be more constrained than the viral DVNP proteins. This may also suggest that the function and roles of the protein in the vast viruses has diversified. In all known cases where the protein DVNP is expressed in eukaryote hosts, all hosts displayed significant deterioration of health (Gornik et al. 2012; Goh & Waller, 2015; Irwin et al. 2018). It is possible that the protein originally serves as a toxic viral effector to weaken the host for easier exploitation of the host transcription and/or translation machineries. Under this scenario, the viral proteins would receive relatively more freedom as long as the host is weakened yet not killed before the new virion particles are sufficiently synthesised and assembled. However, in the nucleus of a cell where protein-DNA dynamics are carefully balanced, it is conceivable that the properties of a structural protein have to be more tightly controlled. At any rate, further wet-bench experiments are required to test either of the possibilities raised here.

3.4 Methods

Statistics analyses

Amino acid counts, length counts, charge counts, and length/charge were performed with homemade Perl scripts with BioPerl module (Stajich et al. 2002). All statistics and graphs were performed with the freely available statistics software JASP (JASP Team 2019).

Charge distribution analyses

All analyses including κ-value and Das-Pappu plots for multiple DVNPs were performed with localCIDER v0.1.14 (Holehouse et al. 2017) and homemade Perl scripts packaging.

Sequences and co-evolution analyses

The viral dataset was made available by Dr Eelco Tromer's personal communication. Dinoflagellate sequences were mined from MMETSP database. All co-evolution analyses were performed with the Gremlin online server (http://gremlin.bakerlab.org) (Kamisetty et al. 2013) with default settings and HHblits search package.

Sequence visualisation and analyses

Sequence visualisation and alignments were performed with Geneious software v12 (Kearse et al. 2012) or Jalview2 (Waterhouse et al. 2009).

Evolutionary history analysis

Dinoflagellate and viral DVNP sequences with known sources were used. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model (Jones et al. 1992). The tree with the highest log likelihood (-29993.00) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Bootstrap values were displayed at the nodes. Evolutionary analyses were conducted in MEGA X (Kumar et al. 2018).

Chapter 4: Development of a scarless molecular cloning method –SLICchange

4.1 Introduction

To pursue structural studies of DVNP proteins and mutant forms of these, over-expression plasmids were required for bacterial transformation and induction of protein expression. These expression plasmids during optimisation required the inclusion of short sequences coding for affinity tags and protease recognition sequences. While several commercial options were available, e.g. direct gene synthesis or several methods described in the next paragraph, I identified that an effective and economical option had not been explored that allowed scarless addition of short sequences. Hence, I began a quest for an effective yet economic, scarless cloning system, of which the result is described in this chapter. Several new constructs had to be prepared in the course of this PhD study (further detailed in Chapter 5 section 2), and this method had been very effective for propelling the project forward.

Molecular cloning has always been an essential skill for molecular biologists. Since the discovery of HindII almost 50 years ago (Kelly & Smith 1970), we have utilised the activities of restriction endonucleases and DNA ligases to cut and paste DNA fragments. These restriction endonucleases recognise particular sequences and cut the DNA in two, but the existence or introduction of the recognition sites, or 'scars', in the target sequences is sometimes not practical. For example, extra sequences in an open reading frame might cause the protein to frameshift or the extra amino acids encoded by the recognition sites may have deleterious effects on the protein. More recently, scarless cloning and ligation-free cloning, both aiming to address this issue, have come into development, and techniques such as the Golden Gate Assembly, Sequence and Ligation-free Cloning (SLIC), Gibson Assembly, and Seamless Ligation Cloning Extract (SLiCE) have granted us increasing independence from classical restriction enzyme cloning (Engler *et al.* 2008; Gibson *et al.* 2009; Li & Elledge 2007; Zhang *et al.* 2014). However, these methods all have their drawbacks. Golden Gate Cloning relies on a new type of restriction enzyme that cuts outside its recognition site, thus giving more freedom on the restriction sequences. The strategy

is still dependent on the fact that the target DNA must not contain recognition sequences within. SLIC creates complementary single-stranded DNA on both the vector and insert DNA sequences with short homologies with the 5'- 3' nuclease activity with T4 DNA polymerase. This method removes the dependency on a recognition sequence, but the original protocol recommends using purified RecA recombinase, and partially relies on the repair mechanism in the competent cells to fill in the single-stranded gaps. SLiCE uses *E. coli* extract as material and demands careful monitoring of bacterial growth, equipment for cell rupture, and more labour. Gibson Assembly works in a similar fashion as SLIC with the chew back and annealing steps, but additionally has a repair step to fill-in and ligate the single-stranded gaps. The method, however, requires a mixture of three pricey enzymes –T5 exonuclease, Phusion polymerase, and *Taq* ligase– and albeit with high success rate and easier handling, comes with a much bigger price tag. In addition, the process requires step-wise temperature fluctuation for different enzymes to be active at different steps.

In another field for mutagenesis on plasmid DNA, QuikChange™ has been one of the most utilised technologies for site-directed mutagenesis (Kunkel 1985; Liu & Naismith 2008). The original QuikChange™ method uses PfuTurbo DNA polymerase along with perfectly complementary mutation-bearing primers (the QuikChange™ manual; Agilent Technologies). QuickChange™ works by the two new mutant primers binding to the original plasmids in each thermocycle and extension by the polymerase, and eventually, the original plasmids being fragmented by the restriction enzyme DpnI and the two newly synthesised mutant strands annealing together forming complete plasmids. As the newly synthesised strands always come from the original plasmids but not the already existing new strands, the accumulation of mutant DNA is linear, rather than the logarithmic amplification of a common PCR reaction. Due to the linear nature of the mutagenesis PCR, the amount of mutated product is typically low, making it not possible to verify the presence of the correct mutation-bearing DNA by gel electrophoresis, one of the most commonly used DNA quantification method. Users are required to blindly continue directly to the transformation step, without knowing the quality as well as the quantity of the mutated DNA product. In addition to the difficulty of verifying the presence of the

mutated DNA, QuikChange® is inherently incompatible with high-fidelity polymerases, most likely due to the fact that these polymerases would have higher affinities towards binding dimers of the fully complementary primers instead of the imperfect template-bound primers (Li *et al.* 2008; Wu *et al.* 2013; Xia *et al.* 2014). High-fidelity is of particular importance when a plasmid DNA several kilobases long is the target of the process. While partially-overlapping primers have seen success in QuikChange™ (Li *et al.* 2008; Liu & Naismith 2008; Qi & Scholthof 2008; Wu *et al.* 2013; Zheng *et al.* 2004), Xia et al. systematically demonstrated that mutagenesis PCR performed with imperfectly complementary primers and the high-fidelity polymerase Phusion is logarithmic in nature, using newly generated products in the earlier rounds of PCR as template (Xia *et al.* 2014). The authors recommended simply transforming *E. coli* competent cells with this linear DNA with both termini ending in a short homology, but also reported an unknown recombination mechanism in *E. coli* causing the linear template with homologous ends to re-circularise. An *E. coli* strain with induced expression of RecET recombination pathway was recommended to improve transformation efficiency (Xia *et al.* 2014; Xia & Xun 2017). The original Quikchange™ has the disadvantages of lack of checkpoints, high false-positive rate, and high error rate, due to the relatively low amount of the successful mutated DNA and the original template as well as the incompatibility to high-fidelity enzymes. A high-fidelity polymerase logarithmic PCR-compatible mutagenesis would mitigate all these shortfalls.

In this chapter, I made improvements to the original SLIC protocol, and combined the simplicity of SLIC and the power of logarithmic high-fidelity polymerase PCR, and created what I have called the SLICchange workflow for the reliable generation of plasmid mutation.

4.2 Results

4.2.1 The rationale of the SLICchange mutagenesis design

I incorporated the high-fidelity logarithmic mutagenesis PCR amplification with the SLIC (Sequence- and Ligation- Independent Cloning) method (Li & Elledge 2007) and created the workflow of SLICchange, which consists of two parts. The first part is generation of a linear DNA with identical sequences on both ends through logarithmic PCR, and the second is to circularise the DNA through the improved 'iSLIC' method, which will be introduced in following paragraphs (Fig. 4.1 and 4.2). Firstly, the PCR reaction is to generate a linear product with short homologies on both ends with imperfectly complementary primers, a method modified from Xia et al. 2014. Primer design is as follows: A primer is composed of 3 parts, the anchor region that is complementary to the plasmid, the mutation-bearing region, and optionally a tail region that is also complementary to the plasmid (Fig. 4.3). The criteria are as follows: the two primers should not be completely complementary; the anchor regions of both primers should have similar melting temperatures; the primers should have a

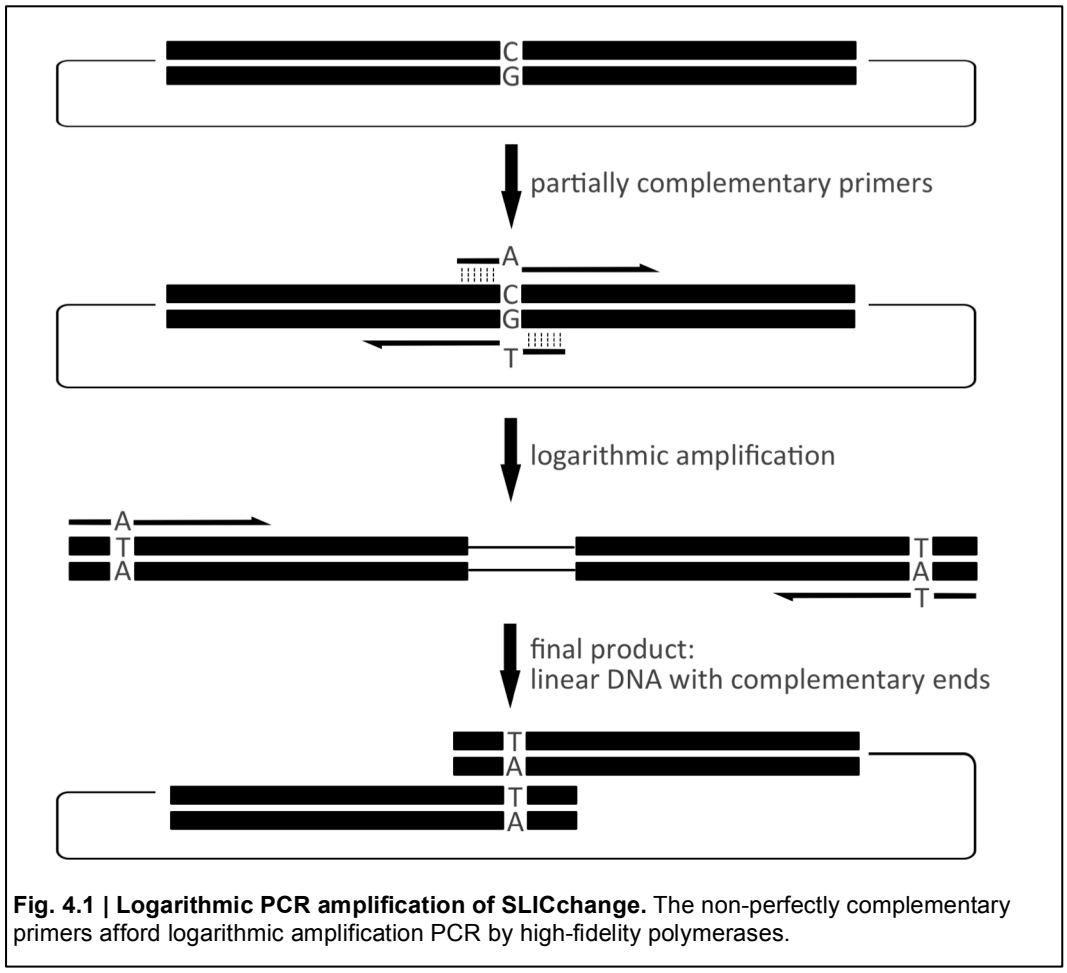


Fig. 4.1 | Logarithmic PCR amplification of SLICchange. The non-perfectly complementary primers afford logarithmic amplification PCR by high-fidelity polymerases.

5’ overlap of around 20 nucleotides, however, the overlap can also serve as the mutation and anchoring regions. A gradient PCR is performed with 3 different annealing temperatures, Tm and Tm ± 3°C, to systematically obtain the most successful product. The PCR reactions were electrophoresed on the gel and visualised, and the PCR reaction products with the highest product to template DNA ratio was purified and quantified. This step provides a checkpoint and pause point for the whole mutagenesis process.

After successful amplification of the PCR product with high-fidelity DNA polymerase and purification, the product is treated with T4 DNA polymerase to generate single-stranded homologies from the 3’ ends on both sides. After pausing the 3’ chewback and the annealing step, a filling-in step was performed by the same T4 DNA polymerase. This method only requires the T4 DNA polymerase (T4DNAP), a rigorously studied bacteriophage enzyme with known enzymatic characterisation and kinetics (Dale *et al.* 1985; Deen *et al.* 1983; Schwartz *et al.* 2015; Stocki *et al.* 1995).

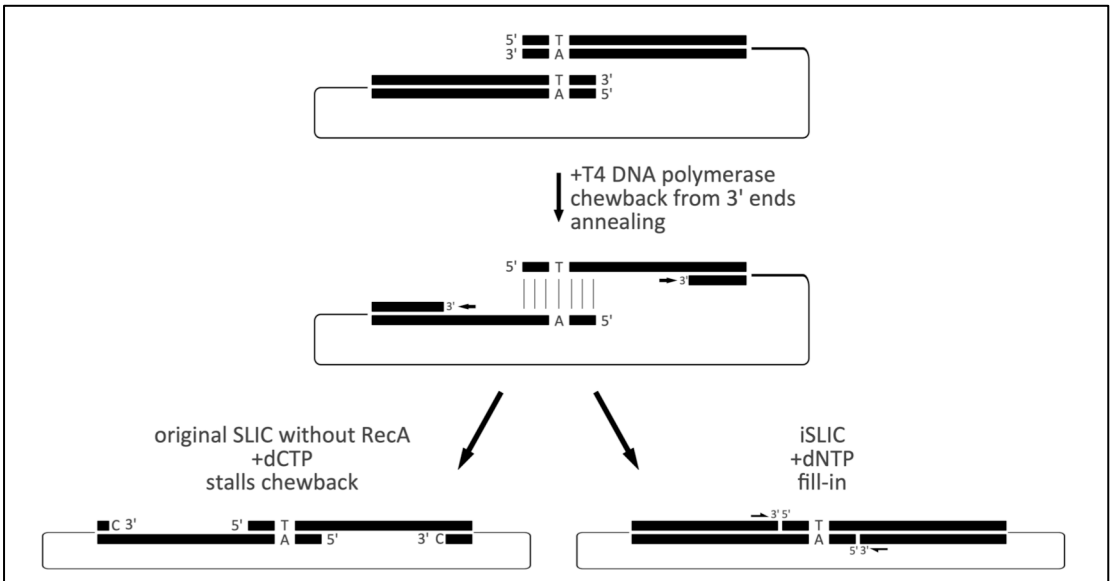


Fig. 4.2 | Comparison of the original SLIC and the proposed iSLIC.

The original SLIC method requires the linear DNA to be treated with T4DNAP with the absence of free nucleotides to exploit the 3’ to 5’ exonuclease activity and generate two complementary single-strand arms that can anneal to each other. dCTP was then added to the reaction to stall the chewback activity. Purified RecA was optionally applied to improve recombination of the overlapping ends of the plasmid

before transformation (Li & Elledge 2007) (Fig. 4.2). Since T4DNAP has no 5' to 3' exonuclease activity, nor does it have strand displacement activity (Hacker & Alberts 1994; Manosas *et al.* 2012), I reasoned that after initial formation of complementary linear overhangs that could self-anneal, by adding standard dNTP instead of only dCTP, T4DNAP would use its fill-in activity to restore the double-stranded region previously chewed back. This should create a plasmid requiring less repair in the bacterium and further improve transformation yield, the process termed iSLIC (Fig 4.2). In my hands, the improved and simplified SLICchange workflow provides satisfactory results. Here I present a simple high-fidelity polymerase-compatible mutagenesis/cloning method with clear and quantifiable results at each step.

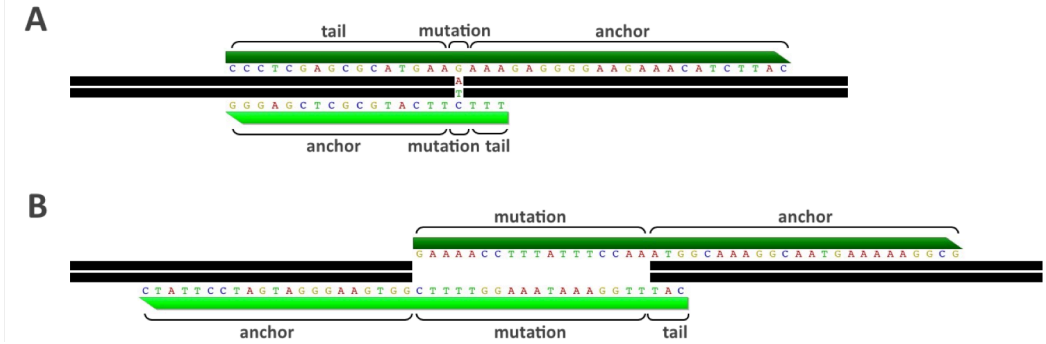


Fig. 4.3 | Primer design and sequences used in this study.

4.2.2 SLICchange site-directed mutagenesis with Phusion-based PCR

To test the working principle of the method, I performed SLICchange site-directed mutagenesis on one plasmid with the intention of generating a lysine silent mutation, from AAA to AAG. Primers were designed as described in section 4.2.1and Fig. 4.3. The anchoring regions of the primer set were calculated to have a melting temperature of 63°C. A gradient PCR was performed with annealing step ranged from 60°C to 66°C, with 3 different template concentrations (Fig. 4.4A). PCR reaction products with the greatest product to template DNA ratio was purified and quantitated (lane 7), and iSLIC reaction was performed, along with a negative control where T4DNAP was not added to induce chew-back and fill-in. The circularised DNAs were then directly used to heat-shock transform competent cells, which were subsequently plated on LB plates with appropriate antibiotics and incubated overnight. The negative control has 196 colonies, whereas the SLICchange products have 3 times more colony counts, 644 (Table 1). Colony PCR was performed, and ten colony PCR-positive clones were sequenced across the mutation-bearing region, and all were confirmed to be correct (Fig. 4.5).

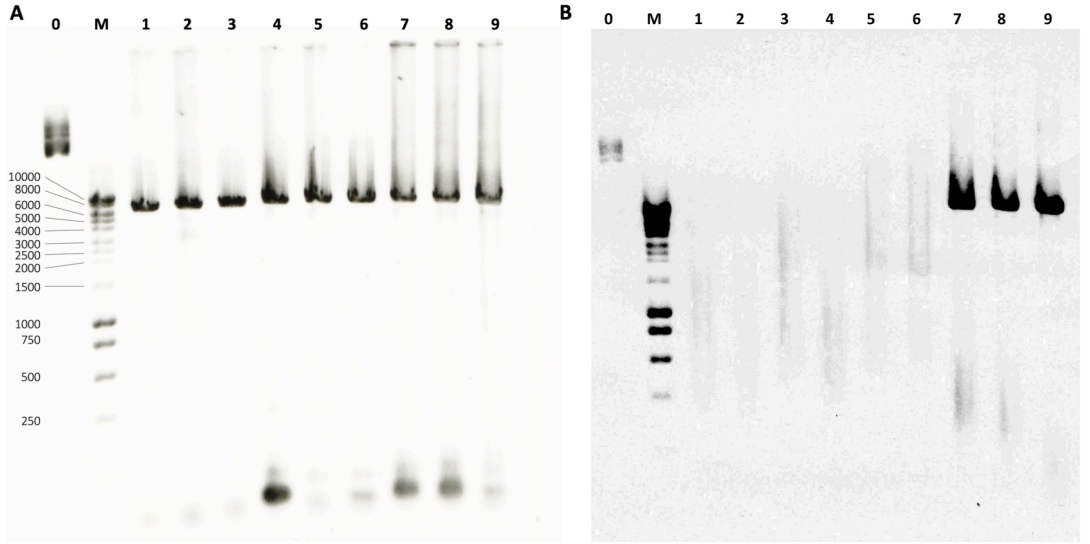


Fig. 4.4 | Gradient PCR for amplification of the mutant template for iSLIC reactions. A and B, the two experiments described in the study. Lanes 0, plasmid DNA PCR template; M, 1kb marker; lanes 1-3, predicted Tm -3°; 4-6, predicted Tm; 7-9, predicted Tm +3°. Lanes 1, 4, and 7, 1ng/100 µL template; 2, 5, 8, 10 ng/100 µL template; 3, 6, 9, 100ng/100 µL template. In both cases, expected size of DNA is ≈ 6.5 kb.

4.2.3 SLICchange sequence insertion with Phusion-based PCR

To test the capability of sequence insertion with SLICchange, I designed primers to insert the sequence that would translate Tobacco Etch Virus (TEV) protease recognition site, 18 nucleotides long in total (Fig. 4.3B) (Tropea *et al.* 2009). In this case, both primers introduce the insertion sequences to either end of the overlapping original region. Following PCR amplification (Fig. 4.4B), the iSLIC reaction was performed as described in the methods section as well as a negative control with no T4DNAP addition. I retrieved a ten-fold increase in colonies for the iSLIC process as compared to the negative control, 1323:125 (Table 4.1). Colony PCR was performed, and six positive clones were miniprepped and sequenced. Sequencing results confirmed correct sequence insertion for all tested plasmids (Fig. 4.5).

Colony counts	**SLIC+**	**SLIC-**
Exp 4.2.2 (substitution)	644	196
Exp 4.2.3 (insertion)	1323	125

Table 4.1 | Original colony counts of the two experiments in this study.

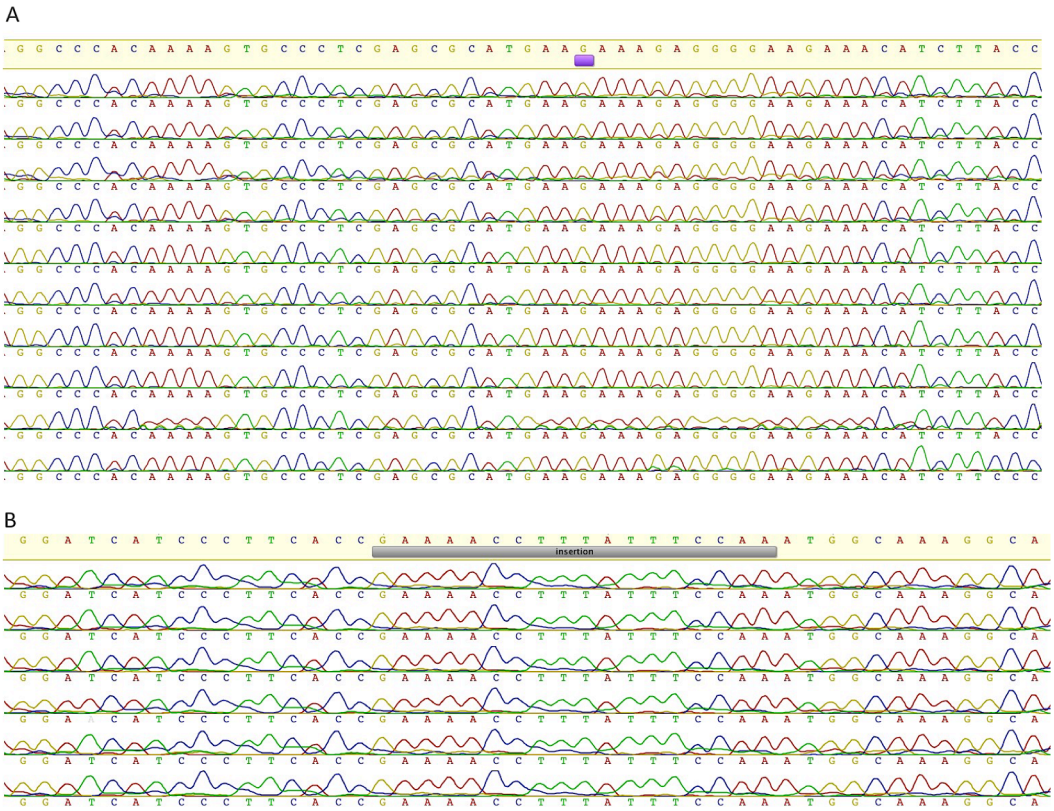


Fig. 4.5 | Sequencing chromatograms from individual colonies. A, 10 colonies from 4.2.1. Purple block represents the site of point mutation. B, 6 colonies from 4.2.2. Grey block represents the inserted sequence.

4.2.4 Comparison between similar scarless cloning methods

To further quantify the improvement of SLICchange versus previously described methods, the insertion mutagenesis presented in 4.3 was repeated 6 times using either the new iSLIC method devised in this study or one of two previously reported methods (Table 4.2). Using the same PCR product with short homologies at both ends, comparisons were made between: Xia's method of simply mixing the DNA components (negative control); the original SLIC no-RecA variation of Elledge using only dCTP; and the method developed in this study, iSLIC. In addition, to understand the effects of addition of nucleotides to overall transformation efficiency, an additional control was included of adding T4DNAP but omitting any nucleotides. Refrozen competent cells are known to have a drop in efficiency as much as 10-fold (Sambrook 2001). To test whether this new protocol would be compatible with most homemade competent cells with an averaged competency of less than 10^7 cfu/µg, refrozen cells were included in this trial as well.

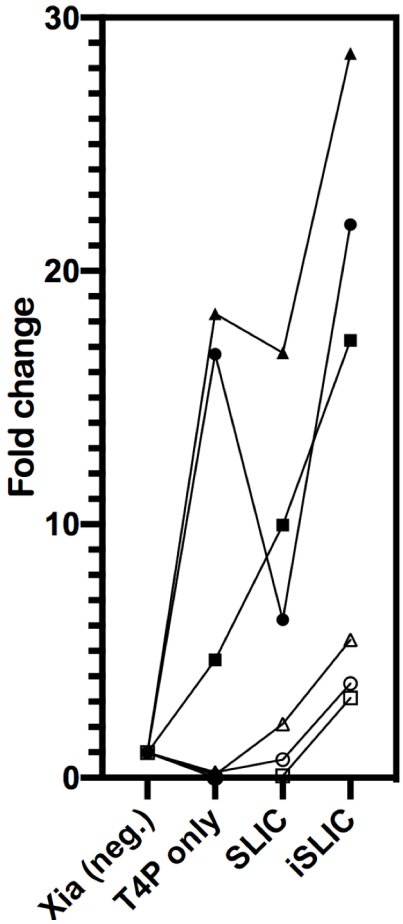


Fig. 4.6 | Fold changes of colony counts over 6 independent trials of the insertion experiment 4.2.3. Each symbol refers to an independent trial described in Table 4.2.

Condition/amount of DH5α comp. cells used	Xia (neg.)	T4P	SLIC	iSLIC	symbol
fresh cells, 200µl, 20' SLIC rxn, 100% plated	308	1432	3071	5317	■
fresh cells, 20µl, 20' SLIC rxn, 10% plated	35	585	218	764	•
fresh cells, 20µl, 20' SLIC rxn, 10% plated	22	403	369	629	▲
refrozen cells, 20µl, o/n SLIC rxn, 10% plated	13	0	1	41	□
refrozen cells, 20µl, 20' SLIC rxn, 10% plated	32	4	68	174	Δ
refrozen cells, 20µl, 20' SLIC rxn, 10% plated	14	3	10	52	O

Table 4.2 | Colony counts of experiment described in section 4.2.4 and their conditions

Linear DNA treated with each method were used to heat-shock transform competent *E. coli* cells, which were subsequently plated on LB plates with appropriate antibiotics. Colonies were counted for each plate (Table 4.2), and the fold changes calculated over the negative control in each trial (Xia), as presented in Fig 4.6. The result demonstrated that refrozen cells showed a wide variation of competency. Furthermore, the addition of dCTP in the original SLIC had mixed results on overall efficiency, though both SLIC and the no nucleotide groups fare better with fresh non-refrozen *E. coli* cells. With refrozen cells, neither condition performs better than mixing the cells with linear DNA (Xia method). However, iSLIC clearly outperformed every other method previously described in every condition tested.

4.3 Discussion

I have presented here a simple, quick, and economical method for generating site-directed mutagenesis or sequence insertion into a circular plasmid. Sequence deletion and substitution were performed with success and clearly outperformed other methods tested in this study. From PCR amplification of the original plasmid to the completion of new plasmid and sequence confirmation, the whole workflow takes only three days (Fig. 4.7).

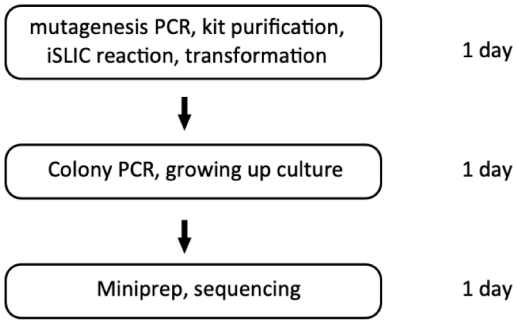


Fig. 4.7 | Proposed workflow timeline of one SLICchange reaction.

High-fidelity polymerases, as compared with other classic polymerases such as Taq and Pfu, bear the advantages of very high processivity and speed while maintaining low error rates. For amplification of long sequences such as a circular plasmid, enzymes such as Phusion® or Q5® polymerases are necessary to ensure no errors are made. However, due to their high fidelity nature, these are not compatible with the original QuikChange™ process (Li *et al.* 2008; Wu *et al.* 2013; Zheng *et al.* 2004). My SLICchange method, based on Xia et al.'s finding that logarithmic amplification is possible with high-fidelity polymerases with imperfectly complementary primer sets (Xia *et al.* 2014), circumvents the incompatibility. Unlike QuikChange™, SLICchange PCR can and is deliberately designed to amplify using products from earlier rounds as the template, i.e. logarithmic amplification. According to my results, the requirements of the primer sets are relatively simple, and total and anchor length of the primers are not of great concern, for in the experiments performed here a primer pair of uneven lengths was sufficient, though with a lower positive/negative colony count ratio (Fig. 4.4 & Table 4.1). The logarithmic nature of SLICchange PCR allows the user to scrutinise the quantity and quality of the PCR product by a simple agarose electrophoresis before the iSLIC step. A successful 50 µL Phusion reaction with a visible band at correct size by agarose electrophoresis typically quantitates more than 300 ng total after kit purification and is sufficient for multiple iSLIC reactions. This step provides a

pausing point should the cloning step does not work, something not offered by the original QuikChange™ method.

Unlike the original QuikChange™ method and Gibson Assembly, SLICchange requires no additional materials, kits, or ligases with the exception of T4 DNA polymerase only. SLICchange also provides improvement to the original SLIC process in that SLICchange utilises both the chew-back and fill-in activities of T4DNAP by using the standard dNTP mixture routinely used in PCR instead of dCTP nucleotide alone. In addition, SLICchange doesn't require specialised strains of *E. coli* as the *in vitro* annealing and repair steps removes the necessity for the recombination activity in host *E. coli* cells. I note that the negative control used for normalisation in section 4.4 is very similar to the process described in Xia et al.'s work (Xia *et al.* 2014), albeit this study employed a more common, *dnaA* strain of *E. coli* competent cells (DH5α in the present study; XL1-Blue MRF' in (Xia & Xun 2017)). SLICchange has markedly improved the transformation efficiency to the point where ultra-competent cells are no longer necessary for successful cloning, as 1% of a 1 x 10^8 competent cells plated gave up to several hundred colonies and even refrozen cells gave a sufficient amount of colonies to readily obtain successful transformants (Table 4.2). The higher efficiency, in combination with the lack of requirement for additional kits and enzymes, results in an overall significant reduction in labour and cost per each successful cloning. I hope this new method facilitates experiment design for fellow scientists and aids scientific research, as it had to the present doctoral study.

4.4 Methods

Amplification and purification

Gradient PCRs were performed as an initial test for the highest melting temperature and least amount of template possible to minimise colonies derived from the original plasmid. PCRs were routinely performed with Phusion® polymerase. Melting temperatures of primer sets were calculated with Tm Calculator (online resource; Thermofisher, Massachusetts, USA; (Allawi & SantaLucia 1997; Breslauer *et al.* 1986)), with the sequences of anchoring regions of both primers as targets. Three different PCR programs using the predicted value Tm and Tm ± 3°C as annealing temperatures were then performed. For each program, three template concentrations were employed: 1, 10, or 100 ng per 100 µL reaction volume. PCR reactions were then subjected to electrophoresis to assess PCR quality. PCR reaction products with the greatest product to template DNA ratio was purified and quantitated for iSLIC to minimise circular plasmid carry-over to the iSLIC step.

After PCR amplification and electrophoresis, PCR products of original template concentration > 10 ng / 100 µl were treated with the restriction enzyme DpnI directly in the PCR reaction buffer to remove original circular plasmid molecules (Activity/Performance Chart with Restriction Enzymes, New England Biolabs, Massachusetts, USA). The products were then purified with a PCR purification kit, and finally quantitated before the iSLIC reactions.

Improved SLIC (iSLIC) reaction.

50 ng of the purified PCR product was diluted in 10µL of 1X NEBuffer 2.1 (NEB). Then 0.4 µL of T4DNAP (ThermoFisher) is added to the reaction and incubated for 20 minutes at room temperature. The reaction is then placed on ice to suspend the chew-back activity, and 1 µL of 2 mM dNTP (Bioline Reagents, London, UK) is added to the reaction. The tube is briefly vortexed and incubated at room temperature for at least an additional 10 minutes for the fill-in reaction to complete. The reaction mixture is then used directly for bacterial transformation.

For controls, original SLIC was identical with iSLIC with the exception that 1 µL of 2 mM dCTP was used instead of dNTP. For negative control, T4DNAP was omitted, but dNTP was added at the same time as the iSLIC reaction.

Bacterial transformation

SLICchange is compatible with any commercially available or homemade chemically competent cells. Unless stated otherwise, 20 µL of Alpha-Select Silver Efficiency Competent cells (Bioline, 1 x 10^8 competency per 200 µL aliquot) was used per each SLICchange transformation, which typically gives transformation efficiency of 1 x 10^7 cfu/ug per reaction.

5 µL of each SLICchange reaction mixture or control reaction with no T4DNAP was added to 50 µL of competent cells. The cells were gently shaken and placed on ice for 40 minutes, heat-shocked at 42°C for 60 seconds, and placed on ice for 10 minutes (Green & Sambrook 2018). 0.5 mL of SOC medium (prepared in-house) is added to the cells and incubated at 37°C with shaking for 60 minutes. The cells are then centrifuged at 1000g for 5 minutes with a tabletop centrifuge, excess medium decanted, pellet resuspended in the remaining 100 µL of medium and plated on LB agar plate containing appropriate antibiotics. The plates were incubated in a 37°C incubator overnight. The next day, colonies on both plates were counted, and at least a total of 2 x (SLICchange / SLICchange - control) colonies picked from the iSLIC plate and sequenced.

SOC medium: NaCl 0.5 g/L, tryptone 20 g/L, yeast extract 5 g/L, autoclaved. 20 mL/L of filter sterilised 1M glucose is added after autoclave.

DNA sequence visualisation and manipulation

All sequence visualisation and manipulation works were performed with the Geneious v12 software, OSX (http://www.geneious.com) (Kearse *et al.* 2012). Artworks were performed with the software Pixelmator.

Chapter 5: Biophysical, biochemical, and histological characterisations of DVNP proteins

5.1 Introduction

The restricted distribution and *in silico* characterisation of the protein DVNP across the tree of life was described in Chapter 3. To directly test the properties of the DVNP protein, in this chapter I first sought to produce recombinant DVNP proteins from two sources for comparative purposes, DVNP.6 from *Hematodinium* sp., a dinoflagellate in laboratory culture, and from *Ostreococcus lucimarinus virus 5*, a Phycodnavirus, and test their biophysical and biochemical properties comparatively. In addition, I also explored the interaction of DVNP with DNA. The selected two proteins, *Hematodinium* DVNP.6 and Ostreoccoccal viral DVNP, represent two individuals from two separate clades, one from a dinoflagellate and one from a virus. According to the results obtained in Chapter 3, the two pools of proteins are different in several characteristics and had probably started to evolve to fulfil different biological roles. Comparing the biophysics and biochemical properties of the two DVNPs may provide a hint of what these roles are.

To characterise the biophysical properties of the *Hematodinium* and the Ostreococcal viral DVNPs, in this chapter I utilised a total of three biophysics methods, namely analytical ultracentrifugation (AUC), circular dichroism (CD), and isothermal calorimetry (ITC). AUC measures the steady-state sedimentation coefficient of particles in a liquid, a coefficient that is related to multiple factors, including the mass and shape of the particle and viscosity of the liquid phase. The sedimentation coefficients can then be modelled to obtain a molecular weight for each of the species present in the centrifugation run, and be used to distinguish between potential monomeric and multimeric species. The method requires long run times, usually at least overnight, and consequently, the inability to detect subtle changes or transient species (Howlett *et al.* 2006). CD measures the differential absorption of two differently circularly polarised lights in each wavelength, in steps of fixed spectral width, by the sample and reports a value called delta epsilon. A delta epsilon being non-zero is indicative of the existence of chiral or circularly ordered structures, whereas positive or negative delta epsilon reports the structure being right- or left-

handed. These structures include secondary and higher-ordered structures of proteins, DNA, or other biologically and optically active materials at different wavelengths. The CD spectra can later be deconvolved and modelled into secondary structure composition relatively accurately providing accurate mass are supplied for standardisation of the resulted spectra. ITC measures the changes in the form of heat of a particular reaction. If conditions were controlled accurately, ITC provides information including the free energy of the reaction, the stoichiometry of the reactants, and the disassociation constant (K_D) of the reaction. All three are commonly employed methods in biophysics study and complement one another for understanding the characteristics of a protein.

In addition, the chapter employs the biochemical test of nuclease protection assay. This assay tests for regular protective structures formed by the protein-DNA complex. The *Hematodinium* nuclear extract, when digested with Micrococcus nuclease (MNase), manifests a smear instead of a ladder. The ladder pattern is the signature of polynucleosomes, and the lack of ladder implies the absence of nucleosome-protected DNA in *Hematodinium* (Gornik *et al.* 2012). On the other hand, although DVNP is known to colocalise with DNA, the total nuclear protein profile is not clear, and it is not known what the other proteins that might also participate in DNA management are. This chapter seeks to provide a definite answer to whether the protein DVNP, albeit from dinoflagellate or viral sources, forms protective unit with double-stranded DNA by performing the *in vitro* MNase assay.

5.2 Results

5.2.1 Expression and purification of *Hematodinium* DVNP and *Ostreococcus* virus DVNP

To perform biochemical and biophysical tests on the two DVNP proteins, *Hematodinium* DVNP.6 (Genbank: AFY23230) and *Ostreococcus lucimarinus Virus 5* DVNP (YP_007674766), efforts were made to produce them in sufficient quantity. The sequences of the proteins are listed in Appendix 3. Several iterations of expression and purification schemes were used throughout the present study. For *Hematodinium* DVNP.6, the gene was initially cloned from mRNA into a pET100/d-TOPO vector with His6 tag at the N-terminus expressed in Rosetta2 *E. coli*, a strain that had been optimised for eukaryotic codon usage. This resulted in good soluble protein expression though with suboptimal purity, for a mixed but significant population of incompletely translated peptide contaminated the sample, suspected to be due to proteolysis, ribosome slippage, or incomplete translation (not shown). Later, a TEV protease cleavage site was introduced into the vector to allow generation of untagged full-length protein following nickel-NTA resin as the cleavage and purification method and TEV cleavage. This coding sequence was later subcloned into pET32a, with a fusion expression partner thioredoxin at the N-terminus to enhance expression level and bacterial tolerance to the protein. The gene was subsequently codon optimised by gene synthesis and switched to a more classical bacterial strain BL21(DE3), however, this resulted into protein accumulation in insoluble inclusion bodies requiring denaturing before purification and TEV cleavage. All the expression strategies described above underwent screens of expression

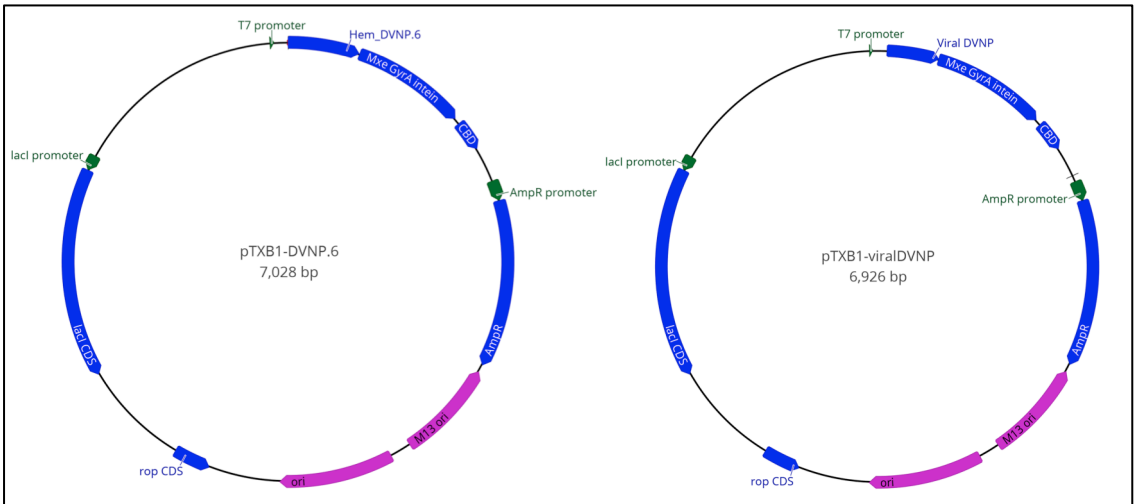


Fig. 5.1 | Plasmid maps for the two expression plasmids described in Section 5.2.1

temperature, expression time, and inducer (IPTG) concentration but without generating sufficient quantity or quality of expressed DVNP product (not shown). Eventually, the pTXB1 auto-cleavable intein system was adopted: the in-frame C-terminal chitin-binding domain fusion partner binds chitin resin when the protein is fully coded, ensuring correct translation of the N-terminal target protein. The chitin-binding domain also bears an enzymatic activity that moves the peptide bond at its N-terminus to the side chain of cysteine at the fusion point, causing it to be prone for a nucleophilic attack and subsequently cleavage. The attack was induced by a high concentration of reducing agent, most commonly used DTT, leaving the native full-length protein in the flow-through fraction. The protein coding sequence was cloned into the pTXB1 backbone using the same SLICchange technology described in Chapter 4. This scheme eventually afforded high yields of the protein with good purity (Fig. 5.2).

The *Ostreococcus* viral DVNP presented similar challenges for protein expression and purification. The gene was initially synthesised and placed in a pET3a backbone without codon optimisation, resulting in low level of expression. It was similarly subcloned into the plasmid backbone pET32a with a N-terminal thioredoxin fusion partner. The results were suboptimal as multiple smaller molecular weight species were present, similarly to the *Hematodinium* DVNP.6. Again, the constructs underwent screens of expression temperature, expression time, and IPTG concentration. Following the success of the intein fusion of DVNP.6, the codon-optimised version of viral DVNP was subsequently subcloned into the pTXB1 vector in an identical manner, resulting in successful expression and purification of the protein (Fig. 5.2).

Following the proof-of-concept successful expression and purification, I was able to scale up the bacterial expression to 3-5 litre batch size, with 2.5ml bed volume of chitin resin for gravity-fed affinity purification in a high salt condition (1M NaCl). The self-cleaved full-length proteins were then purified and fractionated through a 16/60 gel filtration column in the same high salt buffer with a GE Äkta Purifier chromatography system. Sample purity was confirmed by chromatogram as well as SDS-PAGE staining, and then the correct fractions were pooled, concentrated, and

dialysed into the desired buffer, 20 mM sodium phosphate pH7 and 100 mM KCl, and then concentrated again to the desired concentration. Protein samples were of high purity and suitable concentration for biophysics tests (Fig. 5.2). The proteins were then aliquoted and flash frozen with liquid nitrogen.

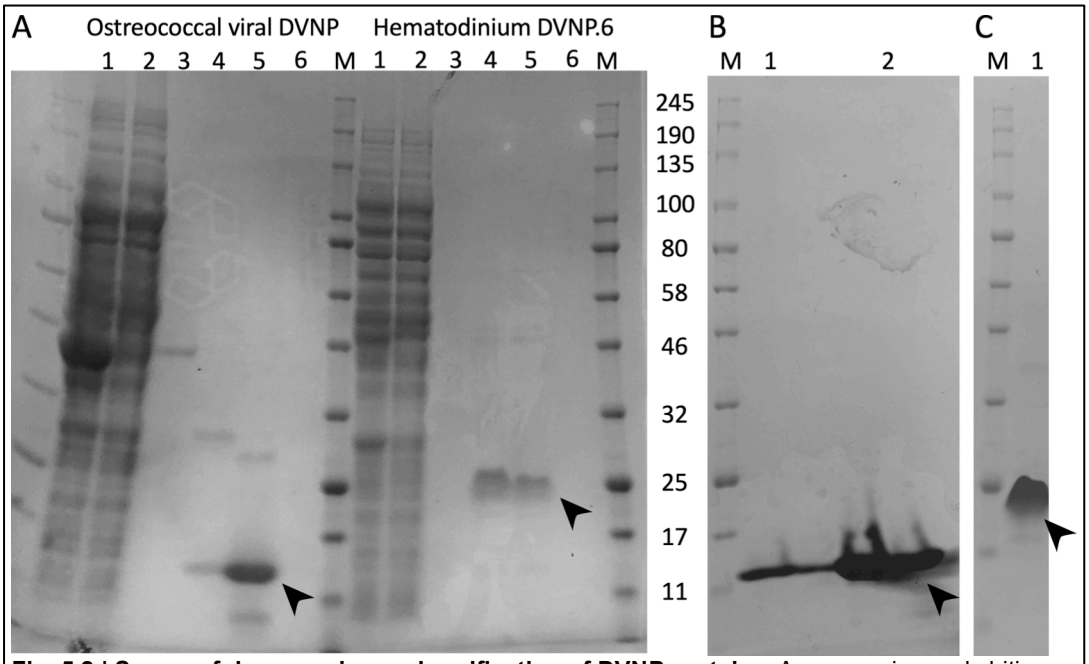


Fig. 5.2 | Successful expression and purification of DVNP proteins. A, expression and chitin column-purification results. Lane 1, total protein post induction. 2, flow-through of chitin column. 3, wash. 4, DTT pre-dlution. 5, eluate. 6, residual protein on beads. M, marker. Arrow head, location of target proteins. B, purified and concentrated viral DVNP. 1, flow-through of concentrator. 2, concentrate. C, Purified and concentrated Hematodinium DVNP.6.

During the expression of the proteins, *E. coli* cells were examined with the DNA stain DAPI with microscopy to determine the potential DNA compaction in bacterial cells. Changes in DNA compaction, however, were not observed in either of the DVNP-expressing *E. coli* (not shown).

5.2.2 DVNP binds DNA and forms µm-sized phase-separated coacervates

Prior to this study, DVNP was known to bind and colocalise with DNA; however its biochemical and biophysical properties were substantially uncharacterised. A simplified calculation of the DNA concentration in a dinoflagellate nucleus gives that 3.8 pg (Allen *et al.* 1975) / 0.524 picolitre (sphere volume with radius of 5 µm) = 7.25 mg/mL. A relative 1:10 protein:DNA ratio would yield a protein concentration of 0.725 mg/mL. To try to understand the DVNP's protective effects on DNA against nuclease digestion under this condition, an attempt was made where a highly concentrated (≈3 mg/ml) plasmid DNA sample (pET100./D-DVNP.6; Gornik *et al.* 2012, sequence shown in Appendix 4) was titrated with a highly concentrated (≈1 mg/ml) DVNP sample to reach a 10:1 DNA:protein ratio, without addition of salt except the phosphate and potassium chloride (100 mM) already present. Upon mixing, the sample quickly became milky macroscopically (not shown), and approximately spherical objects were formed from the solution that could be observed under the microscope. To test that the samples contained DNA, these were then stained with Pico488 (Lumiprobe), a multi-mode DNA-binding dye (Wang et al. 2017), and confirmed that the particles contained DNA (Fig. 5.3A, D). The object

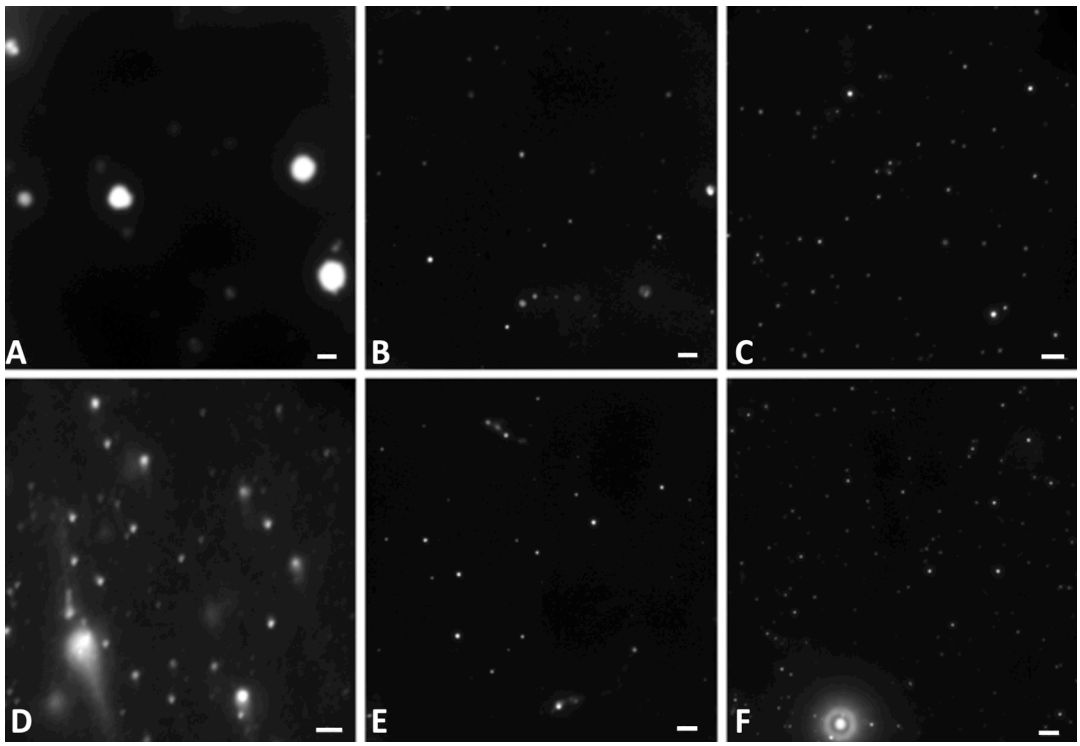


Fig. 5.3 | Protein:DNA ratio changes sizes of DVNP-DNA phase-separated coacervates. A-C, *Hematodinium* DVNP; D-F, Ostreococcal viral DVNP. A & D, protein:DNA = 1:10; B & E, 1:5; C & F, 1:1. Bar, 5 µm. Samples were stained with pico488 and visualised with fluorescence microscopy.

sizes were different for the two proteins as the *Hematodinium* DVNP constantly generated larger globules than its viral counterpart. To further explore the reason for this difference, DVNPs were titrated into independent DNA solutions to reach 1:5 and 1:1 DNA:protein ratio, and smaller particles were subsequently observed under the microscope for both proteins. The results suggested that the sizes of the coacervates negatively correlated to protein to DNA ratio (Fig. 5.3). The phenomenon was very clear between the 10:1 and 5:1 samples, but were much more subtle from the 5:1 to the 1:1 samples.

To verify that the objects consist of protein-DNA complexes, both the *Hematodinium* DVNP and Ostreococcal virus DVNP were chemically labelled with Alexa Fluor® 647. The objects showed homogeneous colocalisation of signals between the protein and dsDNA channels, confirming the protein-DNA complex identity (Fig. 5.4).

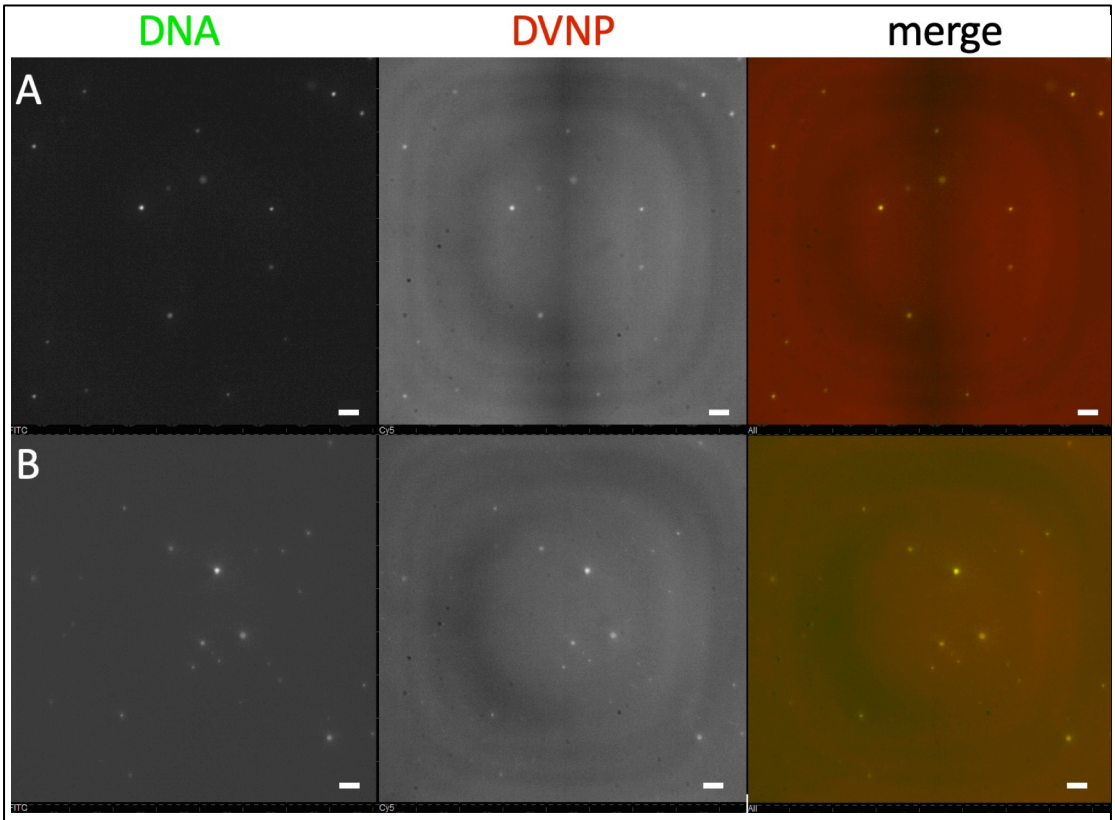


Fig 5.4 | Alexa Fluor-647 conjugated DVNP proteins form phase separated coacervates with double stranded plasmid DNA. Upper panel, *Hematodinium* DVNP, lower, Ostreococcal viral DVNP. Bar, 5 µm.

Rill et al. suggested that the chromosomes *in vivo* in dinoflagellates are in a highly fluid liquid crystalline state (Rill *et al.* 1989). To investigate the possibility of the *in*

vitro coacervates being liquid in nature, I employed photobleaching by high-power laser with the resolution of ≈100 nm. A subsection of the pico488®-stained putative coacervates was bleached with high power laser, and the recovery of the fluorescence was monitored (Fluorescence Recovery After Photobleaching, FRAP). Fluorescence started to recover within 200 ms after bleaching, and at 500 ms the recovery was complete (Fig. 5.5). This result supports that the coacervate is indeed in a very

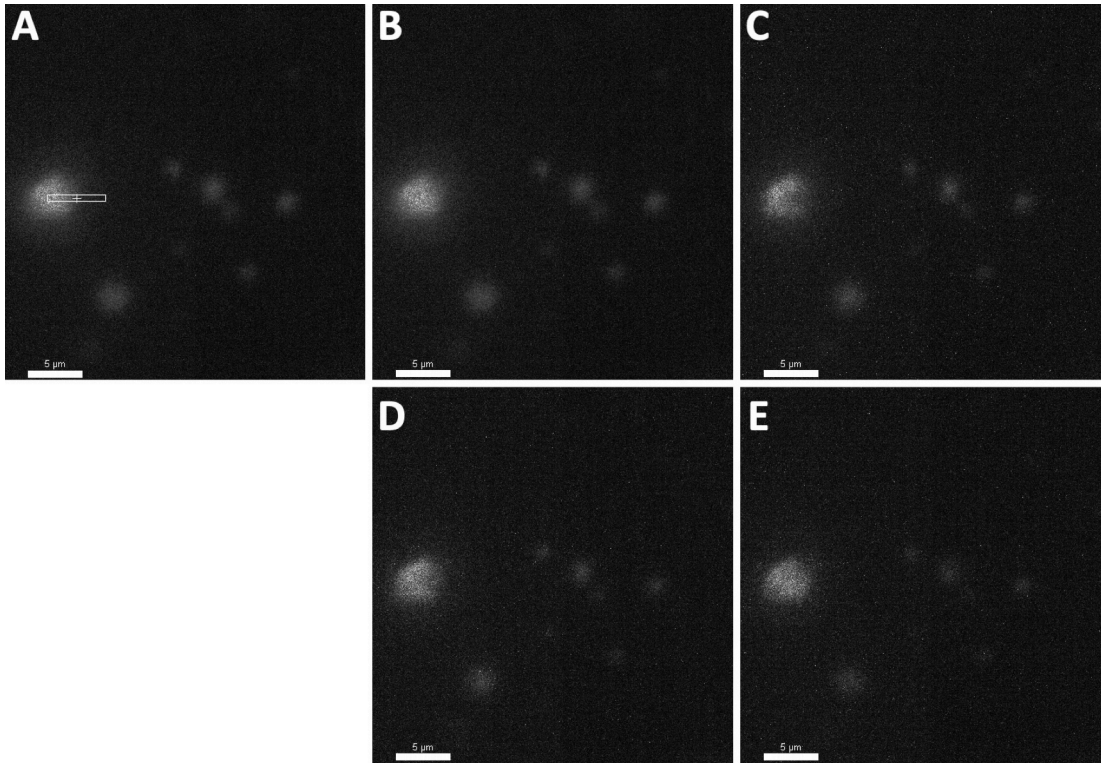


Fig. 5.5 | FRAP result of *Hematodinium* DVNP coacervate suggests high fluidity. A, presentation of the location for laser excitation. B, 100 ms prior to excitation. C, D, E, 100, 300, and 500 ms post excitation. Bar, 5 µm.

dynamic and fluid state.

5.2.3 Biochemical properties of DVNP-DNA complex

5.2.3.1 DVNP-bound DNA is not protected from Micrococcus nuclease digestion

Gornik's study (Gornik et at. 2012) presented that in a dinoflagellate nucleus, nucleosomal packaging and protection against nuclease digestion was not observed. Although the introduction of DVNP coincides with the loss of apparent nucleosomes, the lack of the ladder pattern by MNase digestion could still be the effect of the interaction within the whole nuclear proteome. To further understand the DNA-DVNP binding at the molecular level in a more controlled, *in vitro* environment, I performed a micrococcus nuclease digestion assay (MNase assay) in intermediate DVNP concentrations where phase-separation did not occur in solution. 10 µg of double-stranded plasmid DNA and 1 µg *Hematodinium* DVNP.6 or 0.7 µg Ostreococcal viral DVNP was mixed in 100 µl of 1X MNase buffer. Then the DNA was digested with an incremental amount of MNase at 12°C for 60 minutes. The DNA was then purified, precipitated, resuspended, and electrophoresed on 1% agarose gel (Fig. 5.6). The results demonstrate that with incremental MNase present, all three groups of DNA were digested progressively. Although a slight increase in the molecular range of the smears was observed when both DVNPs are applied, most apparent at DVNP.6+30 units of MNAse and viral DVNP+10 units of MNase, no sign of protective units or 'ladder' patterns of polynucleosome digestion was observed. By comparison, Mattirolli *et al.*'s work using 25 gel units of MNase per 100 µL of reaction volume on *in vitro* assembled nucleosomes assisted with chaperone proteins displays strong characteristics of a protective length of 150 bp (Mattirolli *et al.* 2017). However, the said positive control was not performed in the present study. It appears

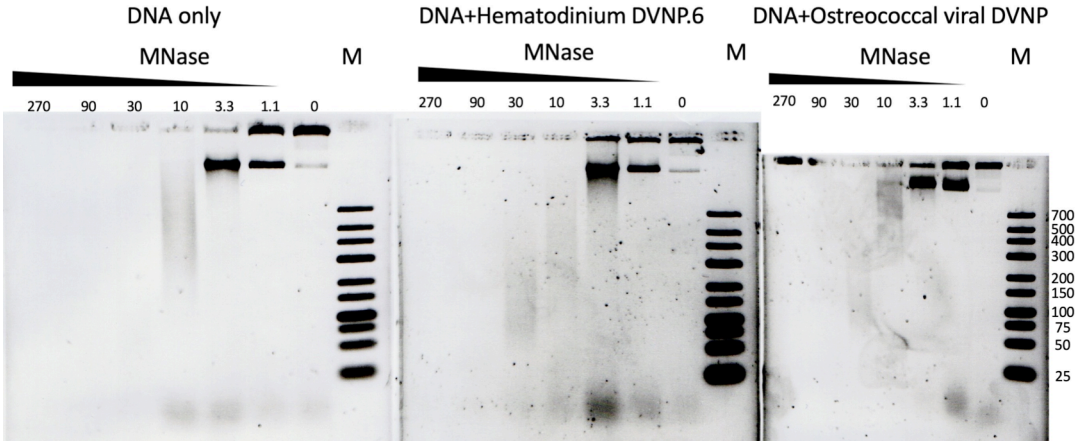


Fig. 5.6 | MNase digestion assay of DNA alone vs DNA with two DVNPs. Increment amount of MNase (0 to 270 gel units) were used to digest plasmid DNA.

that DVNP provides little to no protection to MNase. The results are in consistency to the results published by Gornik et al. (2012).

5.2.4 Characterising biophysical properties of DVNP proteins expressed heterologously

5.2.4.1 Analytical ultracentrifugation (AUC) analysis of DVNP proteins

To determine the steady-state molecular weights as well as oligomeric states, both *Hematodinium* DVNP.6 and Ostreococcal viral DVNP were subjected to AUC. The centrifugation was performed at 20°C at 50,000 g for 20 hours. The spectra collected were integrated and modelled by SedFit, and size distribution analyses were generated. Peaks in the analyses were calculated with a Bayesian approach, and the widths of peaks also represent the probabilities of the accurate sedimentation coefficient. Results of both the *Hematodinium* DVNP.6 and Ostreococcal viral DVNP are shown in Fig. 5.7. The small peaks close to 0S sedimentation coefficient are artefacts from the meniscus in the centrifugation chamber. The distribution plot of viral DVNP shows one species with high confidence, with a modelled molecular weight of 8.97 kDa (nominal 10.3 kDa). The *Hematodinium* DVNP sample shows a major species with high confidence with a modelled molecular weight of 13.2 kDa (DVNP.6 nominal 14.4 kDa). However, a new species of low amount and low confidence is present, with molecular weight determined as 62.5 kDa. The results suggest that for both DVNPs in this condition, they are strictly monomers. For *Hematodinium* DVNP.6, an additional species, very minor in volume, with modelled molecular weight of 62.5 kDa also exists, which is more likely a contaminant, but also possible a DVNP oligomer. Both proteins had modelled molecular weights smaller than their nominal value, suggesting that they are both slightly less compact or less structured than the standard smooth compact sphere model (Howlett *et al.* 2006).

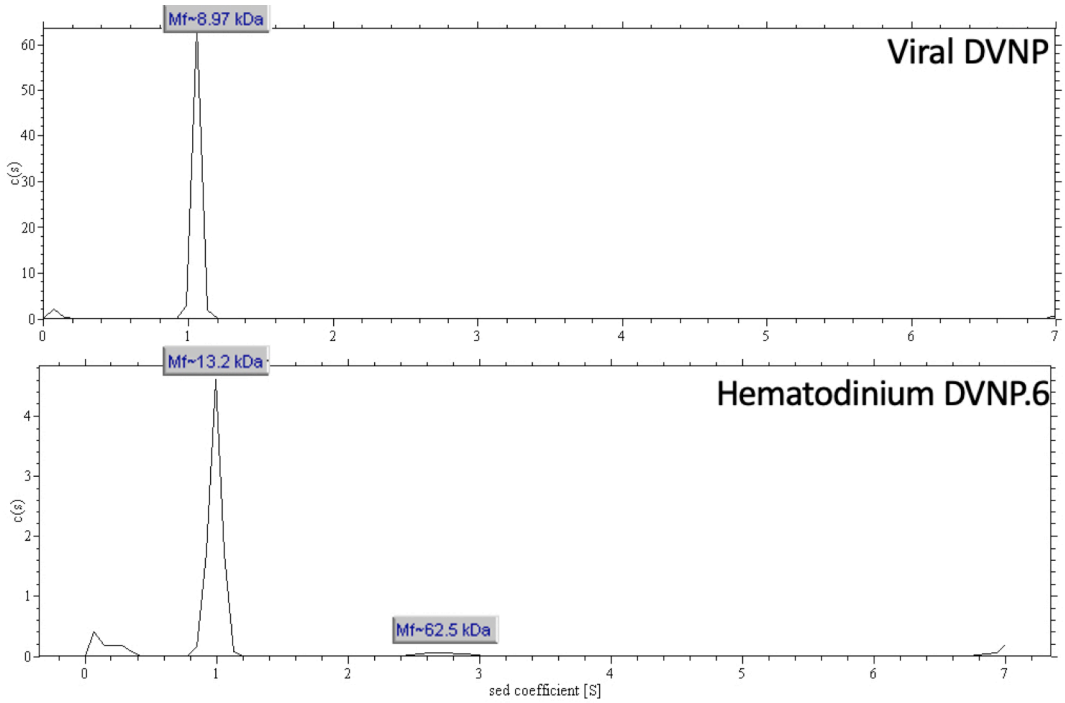


Fig. 5.7 | Size distribution analyses of Ostreococcal viral DVNP and Hematodinium DVNP.6 by AUC. lower, Ostreococcal viral DVNP. Upper, *Hematodinium* DVNP.6.

Based on the known DNA-binding activity (Gornik et al. 2012) and the prediction that DVNP usually occurs in the cell in association with DNA, I sought to mimic the DNA interaction for two reasons; 1) to test if in this context any protein-protein interactions or oligomerisations might occur, and 2) whether a conformation change is a part of the DNA binding process, which is detected by AUC as a substantial shift in modelled molecular weight. However, double-stranded DNA cannot be employed due to the coacervate formation phenomenon. To simulate the conditions of DVNP binding to DNA, but in the absence of DNA polymers, viral DVNP sample was measured with the inclusion of 200 mM of perchlorate anion. Perchlorate is a weak oxidiser with a similar atomic structure to a phosphate, albeit with a stronger dipole, and is commonly employed on DNA-interacting proteins to mimic properties of DNA (Tadeo *et al.* 2009; Wolsey 1973). The AUC results of Ostreococcal viral DVNP with 200 mM of perchlorate salt are shown in Fig. 5.8. Again, peaks closer to 0S are artefacts from modelling the meniscus. The control viral DVNP sample without addition of perchlorate reported a modelled molecular weight of 9.68 kDa, slightly larger than the previous run. This may be caused by the temperature fluctuation of the room hosting the machine causing an increase in the viscosity of the fluid. Present is another species with a molecular weight of 65.7 kDa, presumably a minor

contaminant with a peak so small that it is not visible. After the addition of the perchlorate ion, the presence of the contaminant persisted, and the majority of the protein species is still monomer with slightly higher modelled molecular weight (12.8 kDa). However, a second, distinct species was present when perchlorate was added, with a higher molecular weight (28.6 kDa) close to twice the monomer.

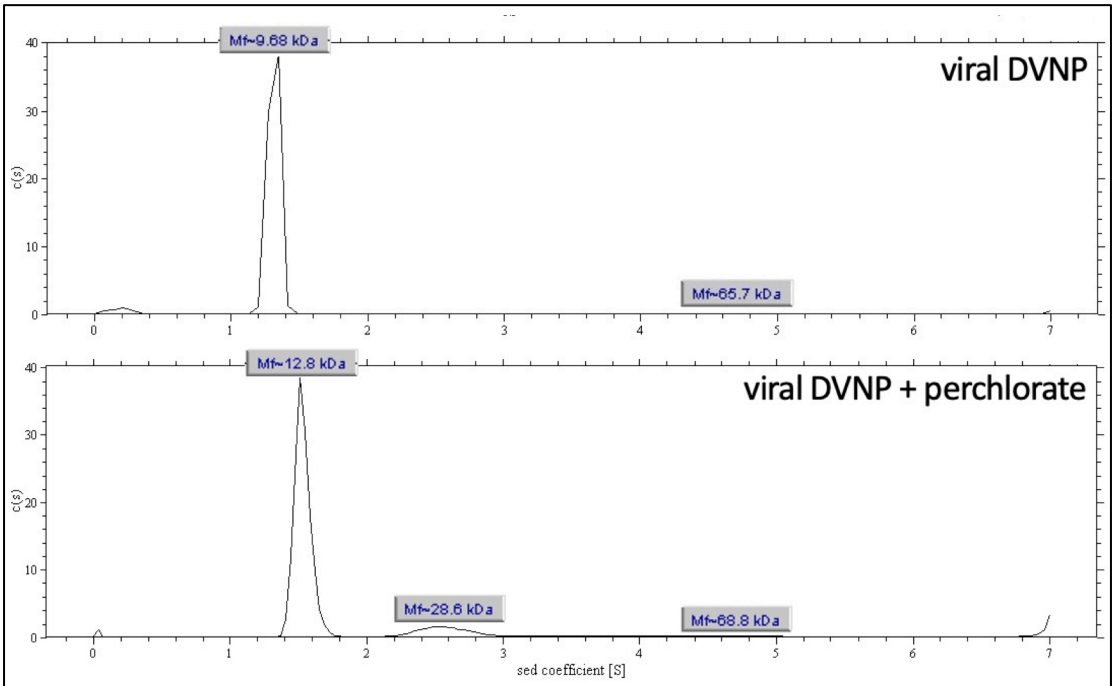


Fig. 5.8 | Addition of perchlorate ion induces a larger molecular weight species of DVNP detected by AUC analysis.

5.2.4.2 Circular dichroism (CD) analyses of DVNP proteins

To assess the possible contribution of protein secondary structures to the *Hematodinium* DVNP.6 and Ostreococcal viral DVNP, circular dichroism was employed. CD analysis of a His6-tagged *Hematodinium* DVNP.6 demonstrated significant structures (Fig. 5.9). The dip in the curve at 225 nm represents the signature of helices, and the drop at 210 nm and rise at closer to 190 nm represents a significant amount of coils, mostly negative ellipticity in this region, and other potential strands or helices. As circular dichroism signal is additive, the percentages of each secondary structures can be modelled by mixing reference spectra of different structures in a different ratio. The spectrum was removed of buffer signal, smoothed, and analysed by secondary structure prediction server Dichroweb. The result generated suggested that the spectrum is composed of 'unordered (35%), strands (33%), helix (18%) and turns (12%)'. The small difference between the experimental data and the reconstructed modelled data suggested that the model is credible.

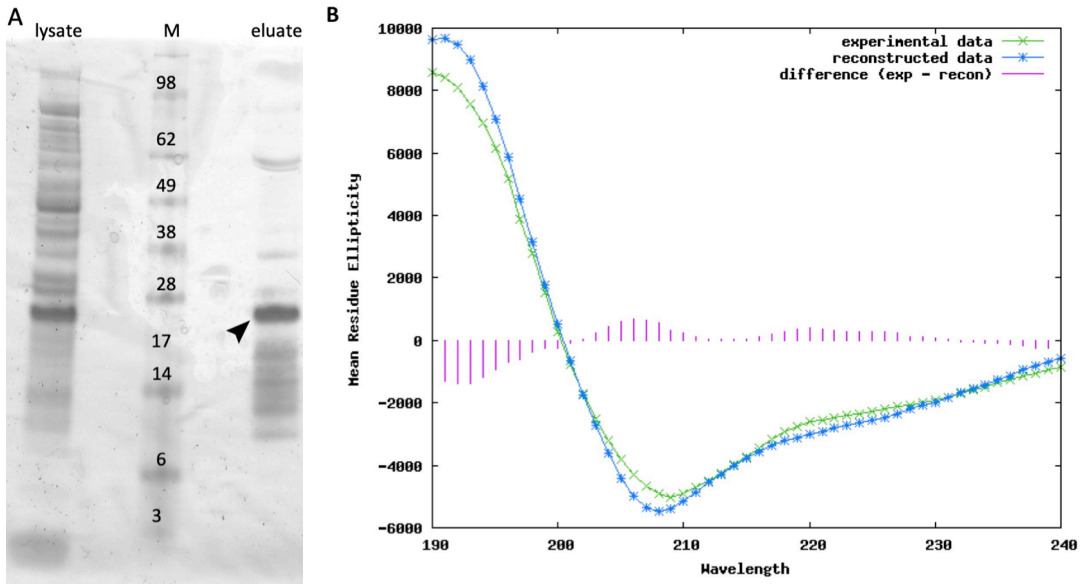


Fig. 5.9 | Circular dichroism results of His6-*Hematodinium* DVNP.6. A, coomassie blue stained SDS-PAGE of the His6-DVNP.6 sample during bacterial expression (lysate) and after purification (eluate). Arrowhead represents the target band of DVNP.6. B, CD spectrum results and modelled result of secondary structures. Modelled result (blue curve, reconstructed data) from dichroweb suggested the signal curve consisted of 'unordered (35%), strands (33%), helix (18%) and turns (12%).

The level of purity of the His6-tagged DVNP sample was previously considered to be suboptimal (described in section 5.1 and demonstrated in Fig. 5.9A) and also the effects of the His6 tag to the overall protein structure was considered difficult to discern. To address these concerns for the CD spectra shown in Fig 5.9, I performed

the CD analyses again with the purified full-length *Hematodinium* and viral DVNPs with no additional sequences, as described in section 5.2.1. The two samples were prepared in higher concentration and were expected to provide a better signal-to-noise ratio. In addition, the two samples were also measured in a thinner cuvette 20 µm in path length, designed to minimise the influence of the buffer components, especially in the far-UV region of less than 200 nm. Usage of thin cuvettes, however, requires even higher concentrations of the samples to provide sufficient signals. All the experimental signals were first removed of the signal from buffer alone and then smoothed (Fig. 5.10). The large dips at 201 nm represented a large coil fraction. The small dips at 225 nm signified the presence of helices, and the elevated reading in the near UV region from 189 to 201 nm implied strong definite secondary structures. When compared with the curve of Ostreococcal viral DVNP, the *Hematodinium* DVNP CD spectrum has a shallower dip at 225 nm, but is much steeper in the 201 nm dip. These differences suggest that by secondary composition ratio, *Hematodinium* DVNP.6 is composed of less helix component, but more coil, than the viral DVNP. By sequence comparison, *Hematodinium* DVNP has an N-terminal tail which does not occur in its viral counter-part and a longer C-terminal tail, both of which are highly positively charged, and may be responsible for the extra coil content in the CD spectrum. All traces resembled the previous experiment performed on His6-*Hematodinium* DVNP.6 (Fig. 5.9). However, the cleaner definition of the dip at 225 nm is the result of a higher signal-to-noise ratio benefited from higher concentration.

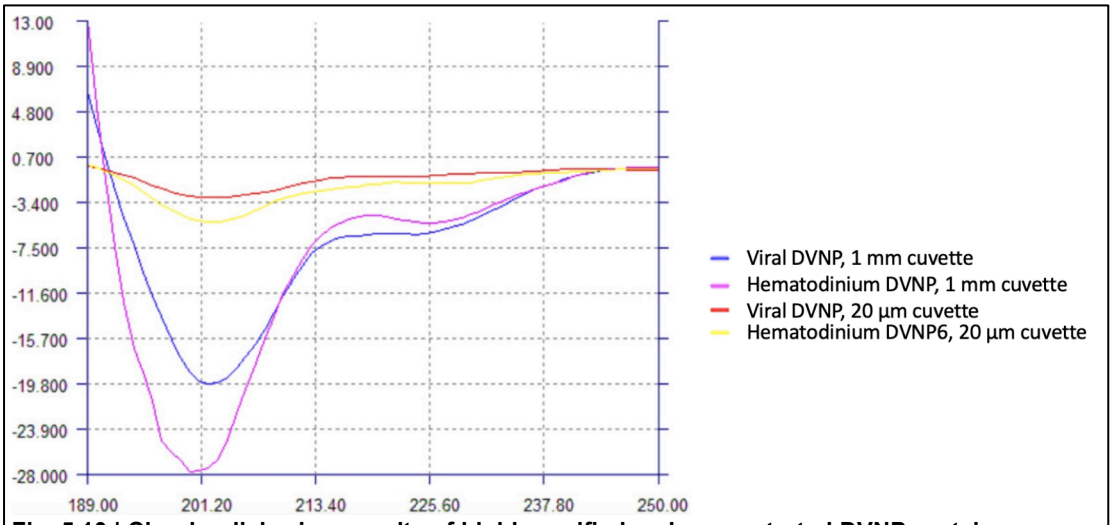


Fig. 5.10 | Circular dichroism results of highly purified and concentrated DVNP proteins expressed recombinantly. X axis, wavelength (nm). Y axis, delta epsilon.

The previous AUC experiment described above demonstrated that perchlorate addition resulted in potential viral DVNP dimerisation (Fig. 5.8), raising possibilities for the protein to multimerise. Additionally, macroscopic slow precipitation in the form of needle-like crystals were also observed with the addition of 200 mM of perchlorate salt (Fig. 5.11 left). To understand the secondary and potential tertiary structures in the condition where a novel, higher molecular weight species was found, CD was again employed. To capture potential higher order structure of viral DVNP, the measured spectrum was extended to 350 nm. Circular dichroism has a low tolerance for many salts, and to analyse a liquid phase sample with 200 mM of perchlorate salt would require the protein in an impossibly high concentration to overshadow the noise generated by the anion. However, the crystal precipitate could be retrieved, and held by the two sides of a 20 μm wide cuvette. It was assumed that the crystals were of high concentration of the viral DVNP protein and might provide informative signals. The signal trace of the buffer alone was arithmetically removed from the signal trace generated from the needle crystal (Fig. 5.11). The overall curve suffered from lower signal to noise ratio, judging from the roughness of the trace, however, qualitative exploration can be made though accurate quantitative analyses could not be reliably performed. The curve is reminiscent of both native DVNP proteins from 200 nm to 250 nm in that the trace is continuously below zero and the presence and scales of the two dips at 225 and 200 nm, supporting the preservation of secondary structures of viral DVNP in the crystal form. A general trend of decreasing delta epsilon, deviating away from zero, between 250 nm and 300 nm can also be

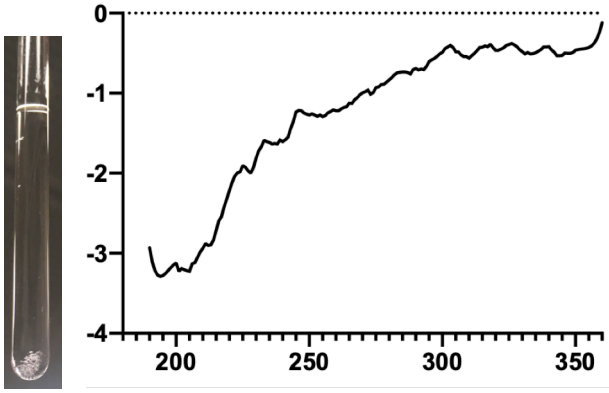


Fig. 5.11 | Circular dichroism result of perchlorate-precipitated viral DVNP crystal. Left, photo of perchlorate-precipitated viral DVNP crystal. Right, CD spectrum of the crystal. X axis, wavelength (nm). Y axis, delta epsilon. Note the extended range, up to 350 nm comparing to 250 nm in previous protein CD experiments.

observed. The dip has no correlation to secondary structures and implies the presence of tertiary structure (Kelly & Price 2000).

Finally, to understand whether the phase-separated coacervate composed of DVNP and DNA described in section 5.2.2 has any chiral order, an Ostreococcal viral DVNP-DNA coacervate sample was analysed by circular dichroism. Conditions were used that resulted in the formation of large coacervates micrometres in diameter, whereas macroscopically the solution becomes milky to the human eye (not shown). I reasoned that the proposed cholesteric liquid crystalline structure of dinoflagellate chromosomes might be the final result of this phase-separated coacervate phenomenon, and measured CD spectrum of the coacervates (Fig. 5.12). To capture the double-strand DNA signal, the wavelength range collected went as far as 360 nm, well out of the range of the commonly used 189 to 250 nm range for protein secondary structure characterisation (Lees *et al.* 2006). The coacervates were prepared freshly and confirmed increased turbidity right before measurements. Livolant and Mestre demonstrated with a CD microscope that the circular dichroism spectrum of a dinokaryon has the 270 nm DNA peak very much broadened and skewed towards longer wavelengths and even into the visible light region (Livolant & Maestre 1988). My result demonstrated that although there were some minor changes to the amplitudes of individual peaks, there were no pronounced peak deviation or formation of new peaks, suggesting the relative unchanged distribution of chiral orders in the sample as a whole. The changes in the far-UV region of 190 nm is likely the influence of the protein components. This result suggests that despite the drastic change in appearance, the protein and the DNA in the DVNP-DNA coacervates maintained their respective chiral orders, and seemingly did not form a regular chiral component.

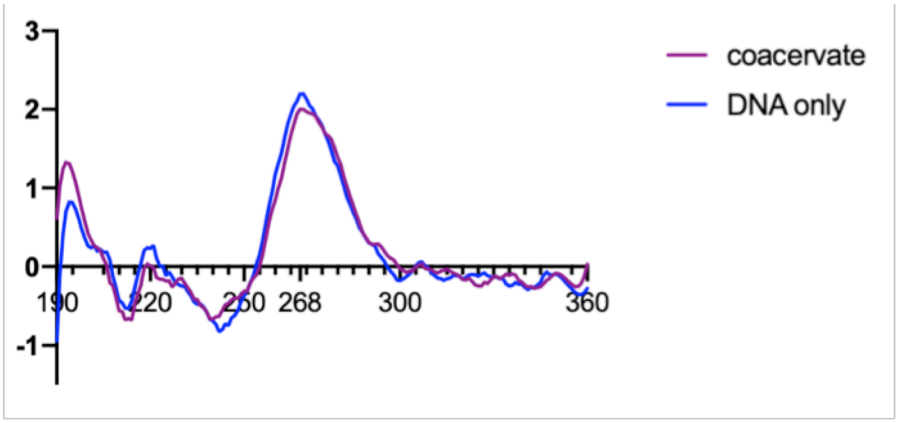


Fig. 5.12 | Circular dichroism result of viral DVNP-DNA coacervate. X axis, wavelength (nm). Y axis, delta epsilon.

5.2.4.3 Isothermal calorimetry (ITC) analyses of DVNP proteins

To elucidate the binding kinetics and stoichiometry of the DVNP-DNA binding reaction, both proteins were subjected to ITC with both 20-mer and 36-mer oligo dsDNA at 20°C. Oligomers 20 and 36 base pairs in size are routinely used in DNA-interacting proteins, making the results produce with both DVNPs more easily comparable with the known DNA-interacting proteins (Machha *et al.* 2013; Turner *et al.* 2018). 5 µM of both dsDNA were titrated with small doses of 85 µM of *Hematodinium* DVNP.6 or Ostreococcal viral DVNP at 2.5 minutes intervals. For both proteins, no thermodynamic differences were observed; the injections did not generate distinct peaks in any of the reactions, however small negative peaks which did not decrease in strength are present. These are probably due to slight differences in the buffer condition in the protein and the DNA. The results suggest the reactions are neither exothermic nor endothermic, but athermic (Fig. 5.13). For comparison purposes, an experiment with identical parameters of C-terminal tail of H1, another DNA binding protein, titrated by the same 36 oligomer DNA, was included in the graph with peaks decreasing in strength over 30 minutes and demonstrated an endothermic reaction that finally saturated at 25 minutes.

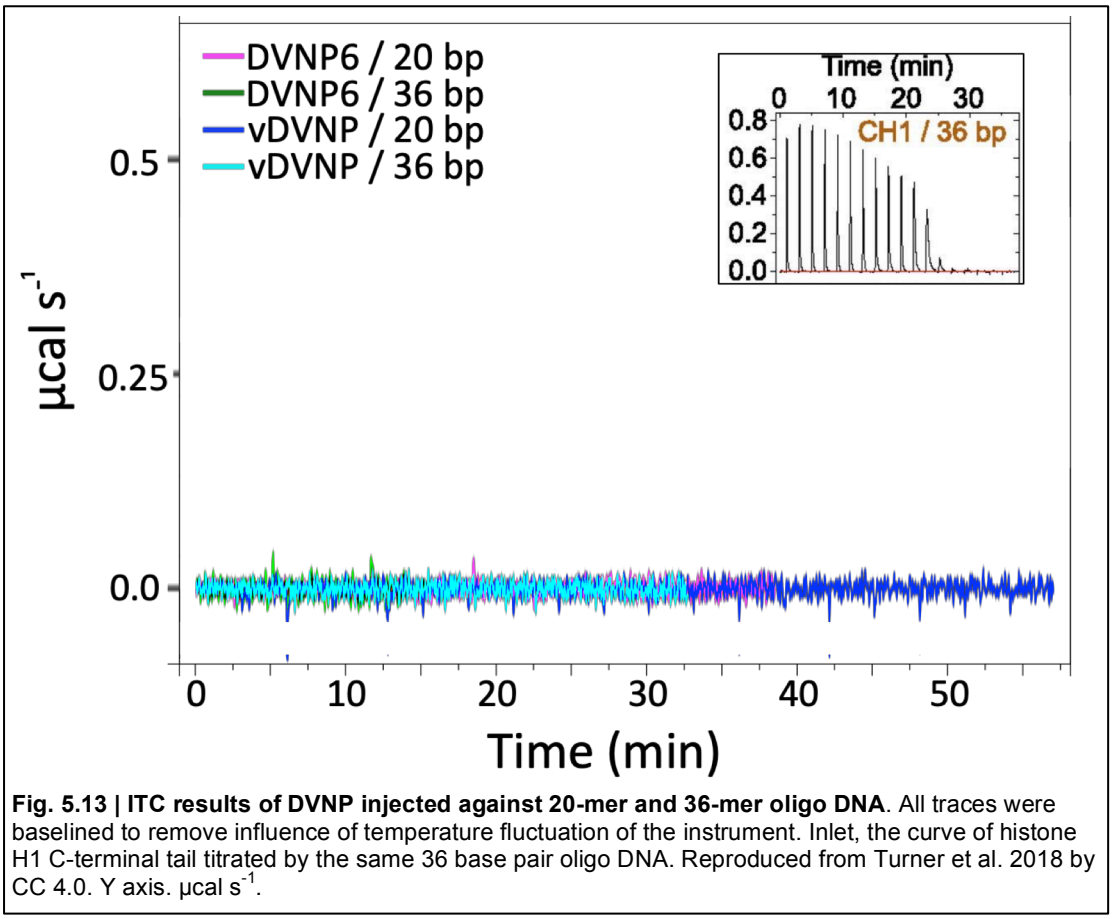


Fig. 5.13 | ITC results of DVNP injected against 20-mer and 36-mer oligo DNA. All traces were baselined to remove influence of temperature fluctuation of the instrument. Inlet, the curve of histone H1 C-terminal tail titrated by the same 36 base pair oligo DNA. Reproduced from Turner et al. 2018 by CC 4.0. Y axis. µcal s^{-1}.

5.2.5 Detecting presence of multimers by Thioflavin T staining and DVNP IFA *in vivo*

The absence of nucleosomal packaging of DNA and the presence of a novel abundant nuclear protein in dinoflagellates suggest a possible novel mechanism of DNA condensation. Results shown earlier in this chapter also provided a hint that as well as forming phase-separated coacervates with DNA, the novel nuclear protein DVNP may have the propensity to oligomerise in a DNA-rich environment. To test if proteinaceous amyloids could contribute to the organisation of dinoflagellate chromatin, I employed thioflavin T, the most commonly used method for amyloid fibre detection in diagnostics (Biancalana & Koide 2010), on *Hematodinium* cells.

Hematodinium cells were fixed with PFA and DVNP detected with αDVNP anti-sera. Samples were then stained with ThT as well as DAPI as a dsDNA marker. The chromosomes showed very specific and strong signals for DAPI, as most of the chromosomes can clearly be distinguished. The DVNP signal shows clear nuclear localisation in good accordance with Gornik et al. 2012 (Fig. 5.14). ThT staining also showed strong nuclear localisation, but lacks a clear definition of the borders of the signal. Additionally, some naked chromosomes, without nuclear and plasma membrane enclosure, can also be observed in the sample preparation, and these chromosomes show specific ThT fluorescence. Microscope images of these naked chromosomes provided higher resolution and more insight into the labelling patterns seen in these chromosomes, free from optical interference of the nuclear membrane and plasma membrane. In addition, liberating the chromosomes from a small confined 3D space into a 2D plane also resulted in less signal convolution on the Z-axis (Fig. 5.14). The signals from ThT and DVNP largely colocalised, but not perfectly, yet show distinct patterns from the DAPI signals. The DAPI signals seemed to be mostly on the outside of the chromosomes, with the inner region compartmentalised, whereas both ThT and DVNP signals resembled a central solid rod with the ends and the centre point stronger in intensity. There are noticeable differences between the ThT and DVNP channels suggesting the DVNP image is not mere bleed-through from the ThT channel.

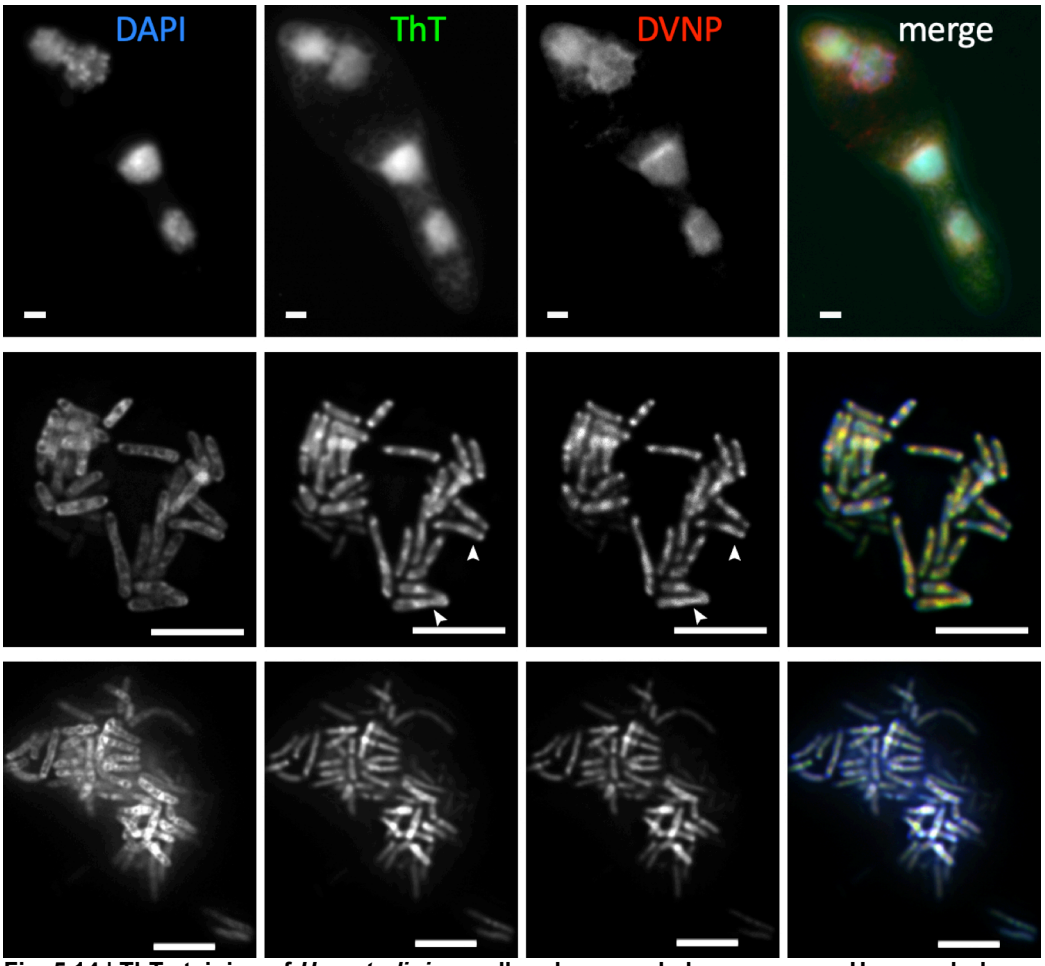


Fig. 5.14 | ThT staining of *Hematodinium* cell and exposed chromosomes. Upper, whole cell. Middle and lower, exposed chromosomes. Bar, 5 μm. Arrow heads mark differences between the ThT and DVNP channels.

5.2.6 DVNP mutant protein – N-terminal truncation

The analyses in Chapter 3 had revealed that the most variable part of DVNP across dinoflagellate and viral origins is the N-terminus tail. Viral DVNPs have a very short tail that is not highly charged, where dinoflagellate DVNPs tend to have a highly positively charged one. In all the biophysics tests and MNase assay performed, there seem to be no substantial differences between the two proteins used in this study, *Hematodinium* DVNP.6, and Ostreococcal viral DVNP. To understand the role of the N-terminal tail, I created a dinoflagellate DVNP.6 with the N-terminal extension removed (NΔ-DVNP.6).

```
>VIRAL_DVNP
-------------------------MSTKPIGSRAEVFHGTAEKTSGGLRAKDLMLDPNDGQIKSVQAHK
>NΔ-DVNP.6
---------------------------------MSSVFRGTKEKTSGGLKKSDLTRN-KSGKIVSKKASE
>DVNP.6
MAKAMKKAMKAKKSSAKKGKKSSKRHSKVAKGKRAKSSVFRGTKEKTSGGLKKSDLTRN-KSGKIVSKKASE


>VIRAL_DVNP
KSALERMKKEGKKHLTKVFKAKK---GKFALQPKEGTAAYKKKMKKMA------
>NΔ-DVNP.6
AAKKQFKKSGLSKFTEAVKKARKVLGIKGFQAVGGKSAKGQALLKKAREFYKK
>DVNP.6
AAKKQFKKSGLSKFTEAVKKARKVLGIKGFQAVGGKSAKGQALLKKAREFYKK
```

Fig. 5.15 | Sequence visualisation of viral DVNP, *Hematodinium* DVNP.6, and NΔ-DVNP.6

The NΔ-DVNP.6 mutant expression construct was modified from the pTXB1-DVNP.6 plasmid, using the iSLIC technique described in Chapter 4. The first 36 amino acids were removed, and instead the 36th amino acid, a lysine, was replaced with a methionine to facilitate translation initiation (Fig. 5.15). The truncated peptide corresponds to 6 amino acids into the viral DVNP by sequence alignment. The plasmid encoding the NΔ protein was introduced into BL21(DE3) cells. The mutant protein was successfully induced and purified using the same chitin column scheme described earlier in this chapter. However, during the dialysis process from the 1 M NaCl purification condition to 100 mM KCl, the condition used for most biophysics tests done in this study, the protein experienced a precipitation event, resulting in an amorphous aggregate (not shown). The precipitated state of NΔ-DVNP.6 prevented further biophysical characterisation but did suggest that the N-terminus provides some property to the stability of DVNP.6 in solution. The remnant protein in the solution was not sufficient for any biophysics measurements.

5.3 Discussion

5.3.1 DNA compaction and phase-separated coacervates

A remarkable feat of both of the DVNPs tested in this study is their capacity to compact DNA into micrometre-sized coacervates. Overcoming the negative charges on the backbone of the DNA is difficult, and most DNA-compacting proteins require either extra energy input or in the case of *in vitro* reconstitution of nucleosomes, require step-wise co-desalting of histone octamer and DNA from 2M KCl to physiological conditions (Dyer *et al.* 2003; Ganji *et al.* 2018; Hirano *et al.* 2008; Thomas & Kornberg 1978). Polymer-and-salt induced (PSI) DNA compaction is a type of DNA compaction where the charges of the phosphate backbones are neutralised, and the DNA concentration reaches a certain level and forces the DNA into an ordered compacted state. PSI-DNA does not require extra energy input, however, to neutralise the charges on DNA polymer, cations with valency larger than 4+ are required (Livolant & Maestre 1988; Marion *et al.* 2017). Biological relevant systems that may be of a similar approach to the PSI-DNA include the spermine- and spermidine-compacted chromosomes in sperm cells (Raspaud *et al.* 2005; Sikorav *et al.* 1994; Takahashi *et al.* 1997), and the highly positively charged N-terminal tail of histone H1 which is known to compact DNA and form phase-separated coacervates *in vitro* (Turner *et al.* 2018). However, for comparison with DVNP, the H1 tail itself lacks an internal structured domain and is an intrinsically disordered protein (IDP), while being able to compact DNA without the globular domain (Thomas & Stott 2012; Turner *et al.* 2018). Yet, considering the results presented in this chapter from circular dichroism, DVNP is probably not an IDP, making the exact mechanism of DNA compaction by DVNP likely different to PSI-DNA.

The biochemical analyses found that although DVNP binds and colocalises with DNA similarly to histone proteins, it behaves drastically differently. Both *in vivo* (Gornik *et al.* 2012) and *in vitro* (the present study) data demonstrated that DVNP does not protect DNA significantly from MNase digestion, suggesting a lack of a basic protein-DNA protective unit in a similar sense to the nucleosome. In addition, DVNP alone, without the addition of other components or step-wise desalting that is required for generating nucleosome *in vitro*, is enough to induce higher order of DNA compaction, a characteristic that histones or nucleosome do not have. By comparison, the bacterial protein Dps is known to be highly expressed during stationary phase, and is capable

of DNA-induced phase-separated nucleoid compaction. Jannissen *et al.* (2018) demonstrated that at high Dps concentration where DNA compaction was induced, the DNA-protein phase-separation specifically allows for accessibility of RNA polymerase, but not restriction endonuclease. The present study selected DVNP concentrations where DNA compaction is not induced and direct comparison cannot be made, however a follow-up experiment with similar parameter may be useful.

5.3.2 Biophysical properties of DVNPs

Since the discovery and characterisation of the protein DVNP (Gornik *et al.* 2012), this is the first study to produce the protein in good yield and quality for biophysics and biochemical approaches, albeit production was done in a prokaryote host. The self-cleavable chitin-binding tag method was optimised for eventual large-scale purification of DVNPs. In addition, the quality and especially the quantity of the protein preps afford the option to perform classic biophysics measurements, which prove to immensely informative when no genetic modification tools are available for the clade of dinoflagellates.

The AUC result demonstrated that both *Hematodinium* and viral DVNP proteins are monomeric in a low salt, DNA-free environment. However, it is important to note that AUC only detects steady-state condition since each run takes roughly 18 hours, and transient dimerisation/multimerisation or quick association/disassociation events are not detected. On the other hand, the addition of perchlorate ions indeed promoted the presence of the species of the 2.5S molecule with a modelled molecular weight of 28.6 kDa, whose measured sedimentation coefficient is close to two times of the monomer (1.3S with molecular weight 12.8 kDa), supporting the presence of dimers in these conditions. Perchlorate, one of the strongest chaotropic members of the reverse Hofmeister series salts, has been shown to promote hydrophobic interactions or protein polymer formation in solution (Okur *et al.* 2017; Sawyer & Puckridge 1973). It has a molecular structure and bond length that is similar to a phosphate ion, an integral part of DNA, which DVNP is known to interact with. It is hence reasonable to speculate that the effect of perchlorate may mimic the *in vivo* condition when DVNP interacts with DNA. In addition, the presence of perchlorate also has the effect of increasing the apparent molecular weight of viral DVNP monomers (Fig. 5.8). There may be several contributing factors for an apparent increased

sedimentation coefficient: 1) the bound perchlorate ions increasing the actual molecular weight of DVNP, 2) increased viscosity and inter-molecule repulsion by the addition of perchlorate salt, and 3) a potential conformational change of the monomer, 4) a shift in the dynamics in the monomer/dimer equilibrium if dimerisation indeed occurs. The sedimentation coefficient is inversely correlated to how similar the molecule is to a smooth perfect sphere, i.e. for a fixed molecular weight, the more deviant the object is from a perfect sphere, the smaller the sedimentation is (Brown & Schuck 2006). An increase of molecular weight would suggest the protein becoming more similar to a perfect sphere in shape. Since the molecular weight of bound perchlorate molecules does not sufficiently explain the difference in modelled molecular weights of the monomer Ostreococcal viral DVNP protein from the two AUC runs, option 3 and/or 4 most likely explain the effect here.

Circular dichroism results of both DVNPs provide an outline of the secondary structures of the protein monomer: a significant portion of helices, possibly some sheets, and a large component of unstructured coil. The CD spectrum of the perchlorate viral DVNP precipitate did not show significant changes in the protein secondary structure as manifested around the 190 - 250 nm range (Fig. 5.11), yet the changes in the 250 - 300 nm near-UV region suggest additional regular chiral or circular structure at another level. These changes might represent multimerisation or higher order of protein structure. The presence of these changes, and the potential existence of multimers, very much supports the postulation that the 2.5S molecule is indeed dimer of the protein DVNP.

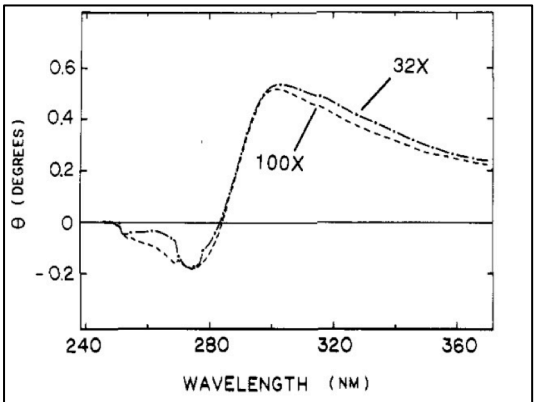


Fig. 5.16 | CD spectra for *Prorocentrum* nuclei from two lenses. Reprinted with permission from Livolant & Maestre 1988. Copyright (1988) American Chemical Society.

On the other hand, dinokaryons host permanently condensed chromosomes that have long been suspected to be cholesteric liquid crystalline (Fig. 1.4; Rill, 1989), which is chiral in nature. Results describing the phase-separated DVNP-DNA coacervates in section 5.2.2, especially the FRAP experiment, strongly supports that at concentrations that

resemble *in vivo* conditions, DVNP and DNA together form a liquid crystalline separated phase. This observation made DVNP a strong candidate as the major contributor to the packaging of the dinokaryotic cholesteric liquid crystalline chromosomes. These coacervates were, contrary to my expectations, were CD neutral, and did not generate any new structures that has CD signals, suggesting that packaging of the liquid crystal in these coacervates were not chiral in nature. In addition, a published study that utilised CD microscopy to measure the CD spectrum of the nuclei of *Prorocentrum micans*, a dinoflagellate, showed a positive DNA peak that is very much broadened towards the long wavelengths when compared with the spectrum of B-DNA (Fig. 5.12 & 5.16), due to the cholesteric arrangement of the condensed DNA. This is, however, not observed in the DVNP-DNA coacervates, again suggesting the paucity of cholesteric type of DNA compaction in the DVNP-DNA coacervates. Furthermore, the disagreement between the perchlorate precipitate and the coacervates CD experiments is unsettling: DVNP alone in combination with perchlorate formed regular chiral higher-order structures, whereas DVNP-DNA phase-separated coacervates did not show additional chiral order over the observed DNA signal. To maintain the desired size of the coacervates for CD, the DVNP to DNA ratio was roughly 1:10 (w/w). If the structure that was observed in the perchlorate precipitate sample were DVNP polymerisation indeed, more DVNP is probably required to generate such state, not just to compact DNA but to also have enough protein on the DNA to form protein-protein interaction. Alternatively, there may be other factors responsible for the dinokaryon cholesteric type of DNA compaction that has yet been identified.

The initial publication of DVNP measured that the stable DVNP-DNA interaction requires up to 0.8M of NaCl to dissociate, similar to that of a histone-DNA interaction (Gornik *et al.* 2012). The ITC results presented in this study, however, suggest that such interaction does not absorb nor release heat, but rather is thermodynamically neutral. This phenomenon can have several explanations, but most likely the total energy input and output, in this case, energy required to remove the hydration shell off the DNA and energy release upon protein-DNA binding, are identical and cancel each other out. This type of athermal protein interaction is not unprecedented in DNA-binding proteins, examples including the globular domain of histone H1 (Machha et al. 2013). Yet due to difficulty in publishing, the knowledge of these

athermal protein interactions are mostly kept within certain research areas and groups (Matthew Watson, personal communication). From an evolutionary point of view, this characteristic is conserved in two groups of very distant organisms where the protein DVNP exists, the dinoflagellates and the viruses, manifesting the importance of this athermic property, though the molecular mechanism of binding is still unknown.

Travers' computational modelling results suggest that the transition from 10-nm bead-on-a-string fibre to the next-level packaging structure 30-nm fibre can be induced with merely torsional stress on the DNA (Wu & Travers 2019). The 30-nm fibre has been argued to be an experimental artefact and its presence *in vivo* has been long debated (Maeshima *et al.* 2014). Wu & Travers' results provide substantial insights and bridge the gap between biology-related chromatin works and chemistry-related *in vitro* studies on liquid crystalline states of DNA. Torsional stress on DNA has been demonstrated to play essential roles in transcriptional regulation in both eukaryotes and prokaryotes (Meyer & Beslon 2014; Reymer *et al.* 2017; Safina *et al.* 2017; Teves & Henikoff 2014). The bacterial histone-like protein HU, important DNA-compacting protein and homologue to the dinoflagellate DNA-compacting protein HCc3 (Chan & Wong 2007; Sun et al. 2012), was reported to introduce negative supercoiling into dsDNA in the presence of topoisomerase I (Rouvière-Yaniv *et al.* 1979). It may be possible that the torsional stress may be one of the fundamentally conserved means to manage DNA in both prokaryotes and eukaryotes, albeit utilising different mechanisms. Following the same line of thought, perhaps DVNP-mediated DNA compaction also involves the control of torsional strength. Safina et al. in 2017 proposed that the FACT complex is a sensor of DNA torsion in eukaryotes, two years after Marinov & Lynch argued in their survey of dinoflagellate nuclear proteins that the presence of FACT complex members in dinoflagellates strongly implies the existence of nucleosomes (Marinov & Lynch 2015; Safina *et al.* 2017). It is thus possible that the underlying torsion-based system is retained, however not moderated by histone proteins and nucleosomes anymore. Instead DVNP-mediated chromosome management commandeered the system by modifying the binding partners of the FACT complex but retaining the torsion sensing mechanisms. Since torsion adjustments also require energy input, this perhaps also provides another angle to comprehend the athermal reactions that were observed in the ITC experiments. DNA torsion can act as an energy reservoir, balancing the energy input and output.

5.3.3 Thioflavin T staining of *Hematodinium*

There are currently three methods that are commonly used for diagnosis of protein amyloids (Biancalana & Koide 2010; Khurana *et al.* 2005). Electron microscopy provides precise results with the highest resolution, but is only used for *in vitro* reconstituted samples, as purifying such amyloids from an *in vivo* system is almost impossible. Congo Red/polarising microscopy used to be the most commonly used method; under the polarising microscope, amyloid positive fibres fluoresce with the colour of "apple green". The colour change in practice is, however, arbitrary and hard to quantify. Thioflavin T, which has been gaining popularity in past decades and took over Congo Red, fluoresces strongly in the green channel in a β-sheet-rich environment. ThT has the advantage of high fluorescence and ease of use and is compatible with most fluorescence microscopes with a FITC filter. Small numbers of false positive and false negative cases, however, have been reported (Cloe *et al.* 2011). Thioflavin T is known to bind with low specificity to proteins and DNA, however only fluoresces when a regular structure, e.g. β-sheet, stabilises the two ring components of the molecule, allowing efficient stabilisation of resonance (Biancalana & Koide 2010). The mechanism may be the reason for false positive and negative cases, for the requirement for being ThT positive is to be able to stabilise the two ring components of the molecule but not the presence of β-sheet.

The highly similar localisation of DVNP and thioflavin T staining *in vivo* supports that it is likely that DVNP may exist in the form of amyloid through multimerisation in the dinokaryon. The strong ThT fluorescence may not provide unequivocal evidence that DVNP *in vivo* indeed exists in the form of amyloid due to ThT's false positive occurrences. However, it does argue strongly that DVNP protein is in a highly ordered structure. This argument, in turn, supports the postulate that DVNP has a multimerisation activity that is induced with the interaction with dsDNA.

The staining results on the naked chromosomes and the chromosomes inside a nucleus did produce visibly different results; in the nuclei, it seems that the DVNP signals were around the DAPI signals, whereas in naked chromosomes, it seems the other way around (Fig. 5.14). Inside the nuclear environment, paraformaldehyde crosslinking may have caused restricted accessibility to the antisera. Alternatively,

this may be the result of an even higher order of packaging within the nuclei. On the other hand, on the exposed chromosomes, the disparity of the staining results observed between the nuclear stain and both DVNP and ThT provides exciting implications. The staining results of ThT, as well as DVNP on naked *Hematodinium* chromosomes, suggest that the chromosomes are DNA-rich on the outside, yet DVNP-rich in the centre longitudinally. Alternatively, the central part of the chromosomes with high DVNP content may be difficult for DAPI to penetrate after PFA fixation. The former, however, is reminiscent of the 'peeled, cored, sliced pineapple' model proposed by Levi-Setti et al. 2008 (Fig. 1.5). The present study is the first after the original publication, to the author's knowledge, to provide support to Levi-Setti's model with an alternative method. There has been, however, no known electron microscopy images that support the existence of such central core fibre that is rich in protein within the permanently condensed chromosomes of the dinoflagellates.

5.4 Methods

Protein expression and purification

All plasmids used in this chapter, pTXB1-DVNP.6, pTXB1-vDVNP, and pTXB1-NΔDVNP.6, were constructed in house with the scarless cloning method described in Chapter 4. Codon optimised DVNP.6 and vDVNP were synthesised by ThermoFisher, and subcloned into different vectors for this study. Codon optimised coding sequences are shown in Appendix 4. The plasmids were used to transform BL21(DE3) competent cells prepared in house (Sambrook & Russell 2006). Fresh colonies less than one week old were used to grow overnight inoculum in 5 mL of LB, and the next morning the inoculum were used to inoculate large volume cultures of LB 1:100 by volume, shaken at 180 rpm in 37°C. The cultures were monitored for optical density until reaches 0.5, when IPTG was added to the final concentration of 0.5 mM. Shaking speed was reduced to 90 rpm for another 3 hours.

After induction, the cultures were then centrifuged to pellet the cells, which were then washed twice in *E. coli* wash buffer. Cells were resuspended in 50 mL of chitin buffer and broken with Avestin Emulsiflex C5, with multiple passes and incremental pressure until 4000 psi was reached. The homogenate was then centrifuged at 8,000 g for 30 minutes to remove cell debris. The chitin column (2.5 ml bed volume per 3 litre culture) was washed beforehand with the chitin buffer. The supernatant of the centrifuged homogenate was passively passed through the chitin column 3 to 4 times by gravity. The loaded column was washed with 20X bed volume of the chitin buffer, and then 2X bed volume of the DTT-containing cleavage buffer was loaded on to the column before the column was sealed and left overnight at room temperature. The next morning, eluate from the column was collected, concentrated, and further purified with a GE Healthcare Äkta Purifier FPLC system with a Superdex HiLoad 16/60 pg size exclusion chromatography column using chitin buffer as liquid phase with flow rate set at 0.3 mL/min. Purified proteins were collected with a automated fraction collector, and relevant fractions were analysed and confirmed with a commercially available SDS-PAGE gel (TruPAGE, Sigma-Aldrich or NuPAGE, ThermoFisher) and Coomassie blue staining. The correct fractions were then pooled, concentrated, and dialysed into assay buffer. The dialysate was then quantitated, and further concentrated as required. The proteins were then aliqoted and flash frozen in liquid nitrogen.

LB medium: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl
E. coli wash buffer: 100 mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA
Chitin buffer: 1M NaCl, 20 mM sodium phosphate pH 7, 1 mM PMSF
Cleavage buffer: 1M NaCl, 20 mM sodium phosphate pH 7, 1 mM PMSF, 50 mM DTT
Assay buffer: 100 mM KCl, 20 mM sodium phosphate pH 7, 1 mM PMSF

Analytical Ultracentrifugation

AUC was performed with Beckman XL-I ultracentrifuge with AN-60 Ti rotor. All samples were exhaustively dialysed against the same buffer prior to the experiment. Exactly 400 µL of each sample at 1 ug/mL were loaded in the sample cells. Cells were loaded and locked into the rotor, which were placed inside the centrifuge with the interferometer installed, allowed to equilibrate to 20°C before the experiment was started at 50,000g for 18 hours. Data collected were analysed with the NIH software package SedFit (Brown & Schuck 2006).

Circular dichroism

Samples were analysed with Aviv Model 410 Circular Dichroism Spectrophotometer. Samples were placed in either a 1 mm or a 0.02 mm path length cuvette. Protein samples were analysed from 189 nm to 250 nm, whereas samples with DNA were analysed from 189 nm to 350 nm. Acquisition time for each wavelength is 3 seconds with dynamic dynode voltage setting. Each sample was measured 3 times and the results averaged and smoothed then removed of the signal trace of buffer alone arithmetically.

Isothermal calorimetry

Samples were analysed with a Malvern Microcal iTC200 system. For both DVNPs, the protein was placed in the cell at a concentration of 20 µM and injected with either 20 bp or 36 bp oligomer dsDNA at 20 mM. Raw data was visualised with Origin7 software suite.

Nuclease protection assay

10 µg of plasmid dsDNA (pET100/d-DVNP.6; Gornik *et al.* 2012; sequence shown in Appendix 4) is mixed with 1µg of DVNP.6 or 0.7 µg of vDVNP in a total of 100 µL

of 1x MNase buffer (NEB). The reactions were equilibrated to 12°C 30 minutes before MNase was added and incubated at 12°C for 60 minutes. The reactions were then phenol/chloroform purified and precipitated with isopropanol. The pellets were then solubilised in 10 µL of 10 mM Tris-HCl pH8, loaded onto a sodium borate 1% agarose gel and electrophoresised. The gel was stained with GelRed® after the run was completed.

Immunofluorescence assay and ThT staining

All steps were carried out at room temperature. *Hematodinium* cells were washed with Nephrops saline once and concentrated by centrifugation at 300g for 5 minutes. Cells were immobilised on poly-L-lysine coated slides, then fixed with 2% of paraformaldehyde and 0.1% of glutaraldehyde in PBS, and perforated with 0.1% of Triton X-100 in PBS. Cells were then washed twice with PBS and once with 10 mM of Tris-HCl pH 8 to reduce autofluorescence. Cells were then blocked with 2% BSA in PBS for 30 minutes, incubated with αDVNP sera 1:1000 in 2% BSA for 60 minutes, washed thrice with PBS, incubated in Alexa Fluor 594 α-rabbit secondary antibody 1:1000 for 60 minutes, and washed thrice with PBS. Cells were then stained with 20 µM of ThT dissolved in PBS for 60 minutes. Cells were washed thrice and stained with 0.1µg/ml DAPI for 30 minutes. Cells were washed twice with PBS followed by twice with water, briefly dried, and mounted in Fluorsave and #1.5 coverslips. The slides were then observed under the microscope using the same method described in Chapter 2.

Chapter 6 Interrogating DVNP-DNA interactions by single-molecule optical tweezers spectroscopy

6.1 Introduction

In the previous chapter, *Hematodinium* DVNP.6 and Ostreococcal viral DVNP were tested for their biochemical and biophysical characteristics in bulk. The methods applied, e.g. AUC, CD, and MNase assay require long running time and measure the steady-state condition, but often fail to report more subtle changes occurring on a smaller time scale. Since late 2018, the Department of Biochemistry had access to a Lumicks C-trap dual-trap single-molecule optical tweezers spectroscopy device. This single-molecule system allows us to measure molecular dynamics with high resolution.

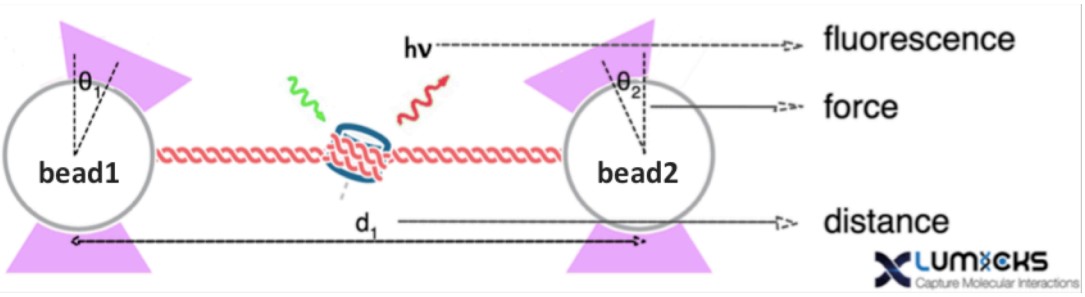


Fig. 6.1 | Working principle of Lumicks C-trap. Combination of tightly controlled laser traps and confocal microscopy affords the system simultaneous spectroscopic capability of distance, force, and fluorescence imaging. d_1, distance between the two beads. θ_1 and θ_2, degree of trap laser deflection.

The system employs two high-power infrared lasers to control two streptavidin-coated polystyrene beads in 3D space within a microfluidic glass chamber. A single piece of DNA, most often λ-phage DNA, biotinylated on both ends, is tethered between the two beads. The optical tweezers system is mounted within a confocal microscope with a Nikon 100x water lens, with 3 single-photon avalanche photodiode detectors for multi-channel fluorescence imaging. The laser traps move the beads and the tethered DNA between different channels of laminar flow in a microfluidic chamber and allow for controlled DNA-protein, dye, or ligand interactions whilst measuring the deflection of the laser to calculate the force the DNA is experiencing in real-time at high frequency (kHz) (Fig. 6.1). To further characterise how the protein DVNP, suspected responsible for the compacted state of permanently condensed chromosomes in the dinoflagellates, interacts with linear DNA with single-molecule precision, both *Hematodinium* DVNP.6 and Ostreococcal viral DVNP were tested

with the C-trap module for the ability of both to alter the characteristics of linear DNA, apply force to a linear DNA, and the behaviour of the protein/DNA interaction.

In an optical- or magnetic- tweezer device, DNA extension force properties provide a standard measure of the properties of a protein's interaction with the DNA (Baumann *et al.* 2000; Biebricher *et al.* 2015; Wang *et al.* 2017). A DNA extension curve is divided into three sections: 1) the entropic stretching regime, before the DNA is completely straightened; 2) the elastic regime, where DNA is fully extended, and more pressure is applied causing the DNA to start accumulating tension in a similar fashion to a rubber band; and 3) the overstretching regime, where the force is so strong that the two strands of the DNA start to melt and peel against each other until the DNA finally breaks (Fig. 6.2; Almaqwashi *et al.* 2016; Brower-Toland *et al.* 2002).

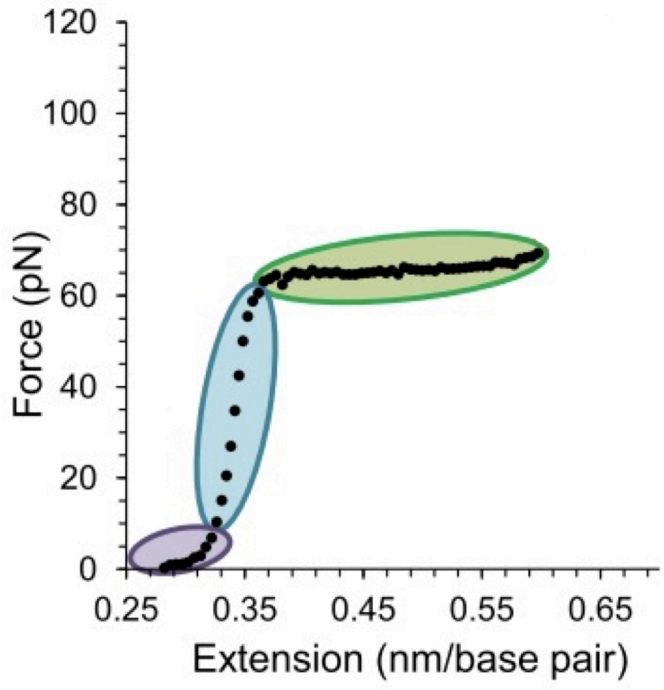


Fig. 6.2 | A typical dsDNA force-distance curve. Grey, the entropic stretching regime. Blue, the elastic region. Green, overstretching regime. Reproduced from Almaqwashi et al. 2016 with permission.

The C-trap instrument offers a high degree of freedom for recording data. High-frequency force measurements can be coupled to either bright field microscopy or one of the two fluorescence imaging tools, the confocal image/movie or the kymograph. The kymograph is a compilation of scans of a single line over time instead of a 2D area when compared with confocal microscopy, with high resolution in time at the expense of space.

6.2 Results

6.2.1 Measuring DVNP-mediated dsDNA-compaction with C-trap

To try to directly observe and quantify the DNA compaction events, both *Hematodinium* DVNP.6 and Ostreococcal viral DVNP were analysed with the C-trap optical tweezers. Firstly, to obtain a suitable protein concentration for the observation, a titration assay as performed with the viral DVNP by force clamp experiments. In a force clamp experiment, when a protein exerts tension force on the DNA, the tweezers system tries to maintain a constant force on the DNA by adjusting the distances between the beads. Consequently, DNA compaction could happen in a controlled manner. At 100 nM of viral DVNP, stochastic compaction events were observed within 30 minutes of waiting time with the time to sudden compaction variable. In contrast, at 200 nM, the delay before compaction commenced was reduced to seconds (not shown). Subsequently, all experiments were performed with 200 nM of protein. Following the identification of this threshold concentration, both *Hematodinium* and viral DVNPs were examined to monitor tension applied to linear λ-DNA. A single piece of relaxed λ-DNA was moved into the protein channel, and force was measured. Results showed that within 500 milliseconds after moving the trap into the protein channel, both *Hematodinium* DVNP.6 and Ostreococus virus DVNP generated tension force on the dsDNA of 5-7 pN and 0.8-1.8 pN respectively (Fig. 6.3).

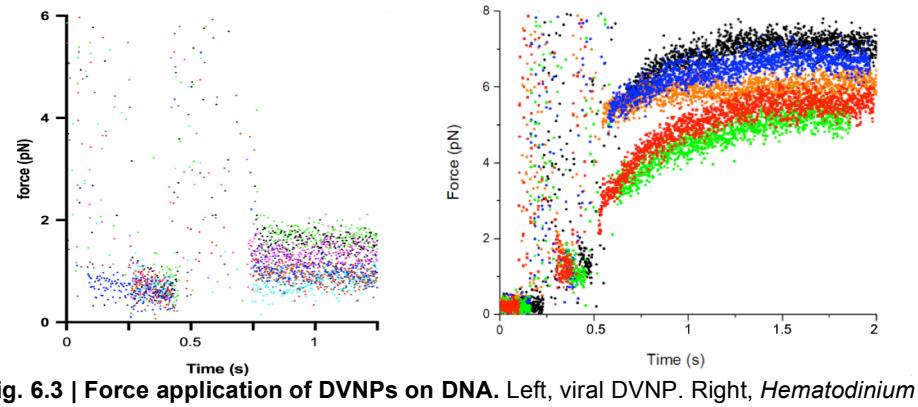


Fig. 6.3 | Force application of DVNPs on DNA. Left, viral DVNP. Right, *Hematodinium* DVNP.6. Each colour represents live force measurements from one DNA molecule. The high scatterings represent the trap moving between channels.

To test if the tension generated by DVNPs on DNA varies during the compaction event, force clamp experiments were performed. The template λ-DNA molecules

were extended to 16 µm, moved into the protein channel, and the force clamp was applied at 1 pN until the distance between the two beads reached 5 µM. A single molecule was measured three times for both *Hematodinium* DVNP.6 and Ostreococcal viral DVNP (Fig. 6.4). The results verified that both proteins compact DNA. However, the dynamics of the compaction events of the viral DVNP were more variable, as a compaction event took from 25 to 40 seconds, likely due to the force clamp being close to the tension applied by this DVNP. By comparison, *Hematodinium* DVNP compacted DNA much faster and more consistently in these experimental conditions, compacting the same length of DNA in 8 seconds, and the force traces almost completely overlapped. In addition, the *Hemadodinium* DVNP.6 placed much higher forces on the DNA during the compaction processes, roughly 6 pN to 1 pN exerted by viral DVNP. These results are in agreement with the previous force measurement results on relaxed DNA.

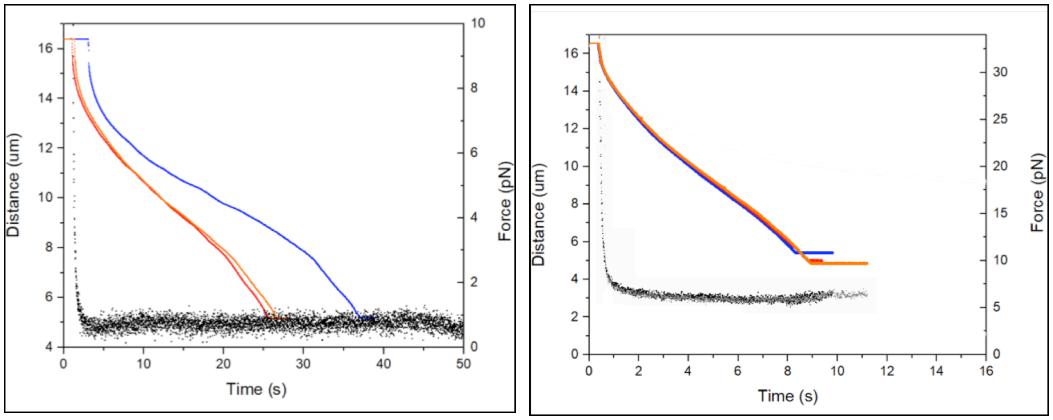


Fig. 6.4 | Force and distance measurements of force clamp experiments for viral and *Hematodinium* DVNP. Left, viral DVNP. Right, *Hematodinium* DVNP. For both figures, black dots are live measurements of force of three force clamps, and coloured curves are distances. Force clamps were initiated at 0 second, and manually terminated when distance reached 5 µm.

To directly observe DNA compaction by DVNP unconstrained by the tweezers system, two more experiments were performed. Firstly, a slingshot experiment was performed; after moving the λ-DNA extended to 16 µm into *Hematodinium* DVNP.6, one of the two laser traps was released. The freed bead, along with the DNA, shot towards the still trapped bead, and the two beads collided within 300 milliseconds (Fig. 6.5A). Secondly, a flow-stretch experiment was performed. A λ-DNA was only tethered on one side and straightened with active buffer flow and stained with Sytox Orange, a tension-sensitive intercalating DNA stain. *Hematodinium* DVNP.6 was flowed through the DNA, and three independent compaction events were recorded as kymographs (Fig. 6.5B). The kymographs demonstrated that after the initiation of

compaction, the entire linear DNA was compacted to the tether against the flow within from 1 second to 5 seconds. It was also observed that the compaction events seemed to have begun from the untethered free end of the DNA, judging by the higher fluorescence intensity of the free end when compared to the rest of the linear DNA in the kymograph. Both experiments demonstrated that *Hematodinium* DVNP.6 could compact DNA rapidly.

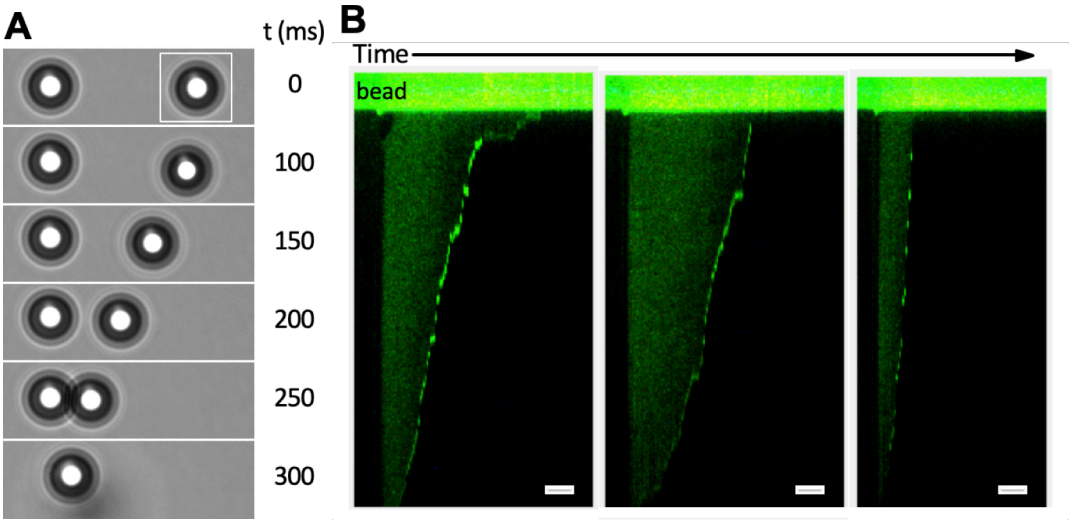


Fig. 6.5 | Observing DVNP.6-mediated DNA compaction in action. A, Slingshot experiment. B, flow-stretch experiment. Scale bar, 1 second. Details of both are in text.

6.2.2 DVNP modifies the characteristics of dsDNA

After characterising the compaction capability of the DVNP-DNA complex, to understand whether modification of the DNA itself was involved in the compaction or the compaction purely relied on protein-protein interaction, force-distance correlations were obtained. Force measurements were recorded while DVNP-induced compacted DNAs was extended in the protein channel. The concentration of 200 nM was used for both proteins. For *Hematodinium* DVNP.6, a three characteristics of the DVNP-DNA force-distance curve were observed: a highly irregular sawtooth pattern of force before the DNA was fully extended (5- 14 μm); during the elastic regime a reduced slope (15-17 μm); and a higher persistence force after entering melting regime (~17+ μm) (Fig. 6.6). The sawtooth pattern had a baseline force of about 7 pN. For viral DVNP, there are several differences. Firstly, the sawtooth pattern was also found, but at a much-diminished level; the reduced slope was not found during the elastic regime; the persistence force upon DNA melting was then slightly reduced, although not consistently.

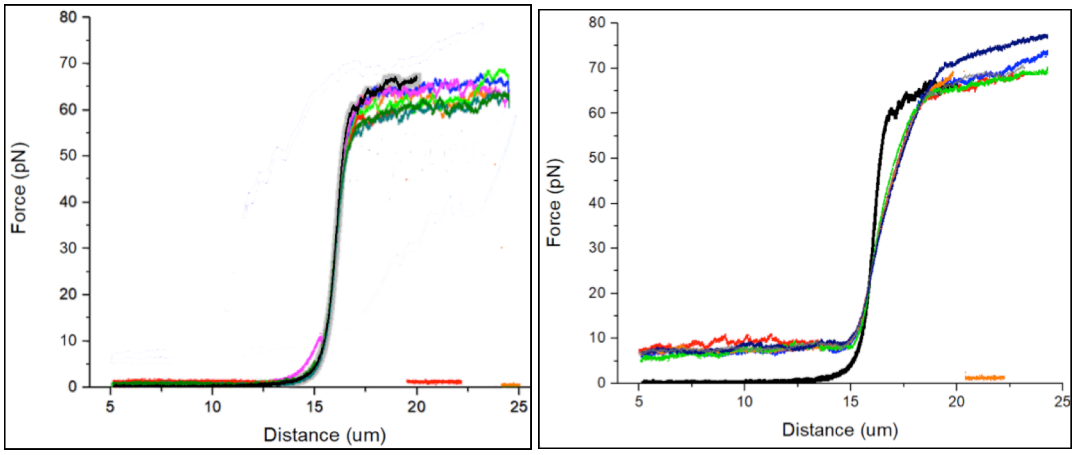


Fig. 6.6 | Force-distance curves of naked DNA and DNA-protein complex. Left, Ostreococcal viral DVNP. Right, *Hematodinium* DVNP.6. Each colour represents force-distance trace of one independent molecule. Black trace, dsDNA alone.

To test if the differences observed between the two DVNP proteins are due to actual functional differences or rather concentration-dependent differences, a much higher concentration (1 μM) of viral DVNP was tested (Fig. 6.7). The result demonstrated an elevated sawtooth pattern of force before the DNA was fully stretched, and slightly increased persistence force as well. The modification in the slope during the elastic regime that was present in the DVNP.6 curve, however, remained to be absent. The result suggested that *Hematodinium* DVNP.6 seems to be more efficient in its

performance. The result is also consistent with the previously demonstrated result that under the same concentration, Ostreococcal viral DVNP generated lower force on the DNA.

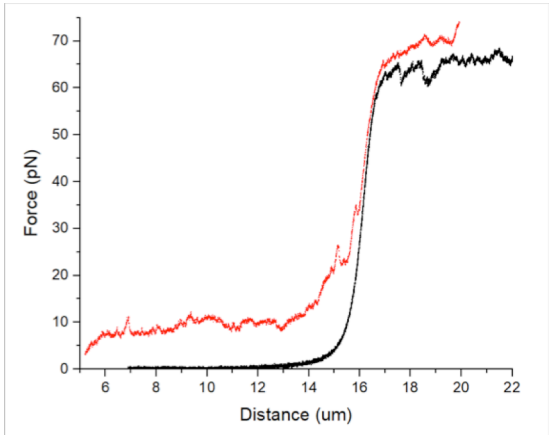


Fig. 6.7 | Force-distance curve of viral DVNP at 1 µM. Black trace, dsDNA alone.

I next asked if soluble DVNP in the buffer solution is exchanging with DNA-bound DVNP and necessary for the properties so far measured, or if stably bound DVNP drives the observed effects on DNA. To test this question, the DNA was moved into the protein channel and 'charged' with 200 nM of either Ostreococcal viral DVNP or *Hematodinium* DVNP.6 for 30 seconds. Afterwards, the DNA was moved back into the buffer channel, and force clamp was initiated with the target force set at 1 pN and terminated when the distance between the beads reached 5 µm, or the DNA did not compact within 30 seconds (Fig. 6.8). For the viral DVNP, it seemed that once moved out of the protein channel, the DNA failed to compact, although the same DNA molecule was able to compact when moved back into the protein channel. *Hematodinium* DVNP.6, on the other hand, was able to compact DNA several times once moved into the buffer-only channel, yet each subsequent time requiring a longer time to complete the compaction (Fig. 6.8). To correlate the extended time required for DNA compaction with the modified characteristics of DNA, force-extension experiments were performed after each compaction but stopped before the overextension phase to prevent DNA from breaking (Fig. 6.9). The results demonstrated that the DNA characteristic modifications seen on DNA by *Hematodinium* DVNP.6, namely the irregular sawtooth pattern and the reduced slope of the extension phase, lessened with each extension. The baseline of the sawtooth

pattern reduced from roughly 7 pN to 2 pN after 3 extensions, and the slope gradually also increased with each extension.

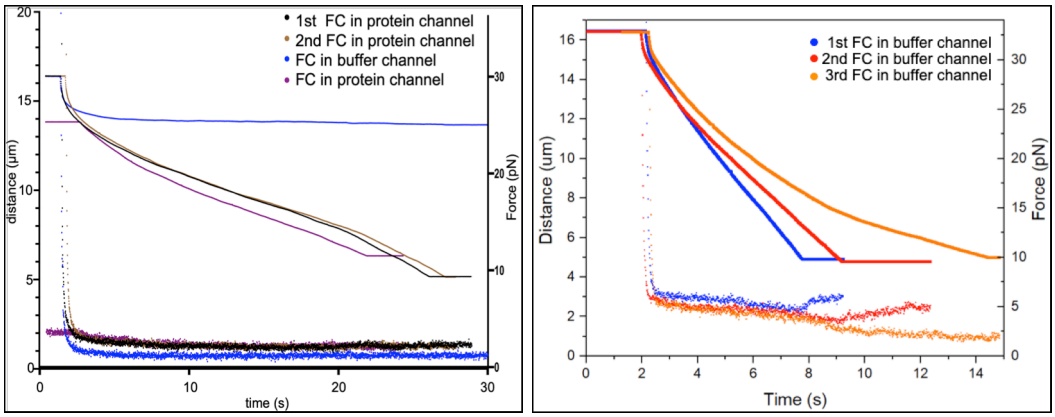


Fig. 6.8 | DVNP-DNA compaction without free DVNP proteins in solution. Left, Ostreococcal viral DVNP-charged λ-DNA failed to compact the DNA without external protein, but resumed compaction when moved back into protein-rich environment. Right, *Hematodinium* DVNP.6-charged λ-DNA could compact without external protein for several times, though slower each time. FC, force clamp.

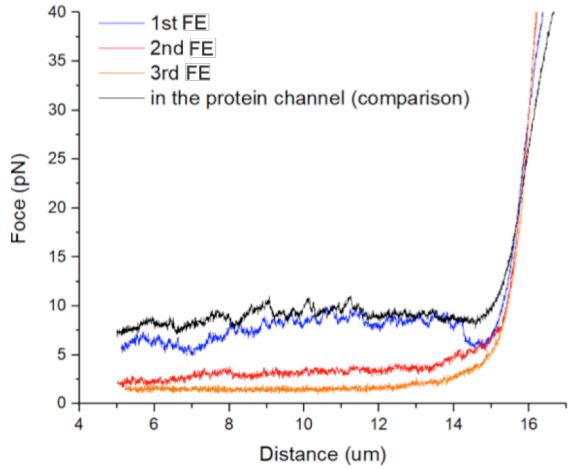


Fig. 6.9 | Force-extension curve of DNA compacted by *Hematodinium* DVNP.6 after each compaction event in Fig. 6.6. FE, force-extension.

6.2.3 Imaging the dynamics of DVNP-DNA interaction with C-trap

To directly observe the protein-DNA interaction and behaviours of both *Hematodinium* DVNP.6 and Ostreoccocal viral DVNP with imaging techniques, both DVNPs were labelled with a fluorophore that is compatible with the C-trap system. A far-red dye NHS-ester was the most suitable and straightforward to implement while not clash with the DNA dye Sytox Orange in spectral properties, and Alexa Fluor 647 was adopted. However, NHS-esters react with all primary amines, including the side chains of lysine, which both DVNPs possess in high percentages. To seek labelling conditions where the lysine side chains were not modified, the pH of the protein solutions were modified to 7, demoting the activity of side-chain amines and shifting the equilibrium towards the N-terminal amine. After the labelling reaction, to verify the activities of the labelled protein were similar to the unlabelled protein, the labelled *Hematodinium* DVNP.6 and Ostreococcal viral DVNP (red DVNPs hereafter) were tested with C-trap for their force responses. Labelled DVNP.6 was flowed through relaxed λ-phage DNA to observe the tension force response and the DNA imaged with a kymograph, similar to the experiment depicted in Fig. 6.3. The traps were moved into the protein channel at 9 seconds, demonstrated by the sudden intensification in fluorescence of the beads. A rapid force response was observed at the same time, of identical time scale and multitude as the unlabelled protein (Figs. 6.10 & 6.3). The response supported that the labelled protein behaved similarly as the native protein, and could be used for further experiments. I note the formation of one bright spot, or focus, on the λ-DNA at 25 seconds. This seemingly correlated with a very subtle response of a dip in force (Fig. 6.10).

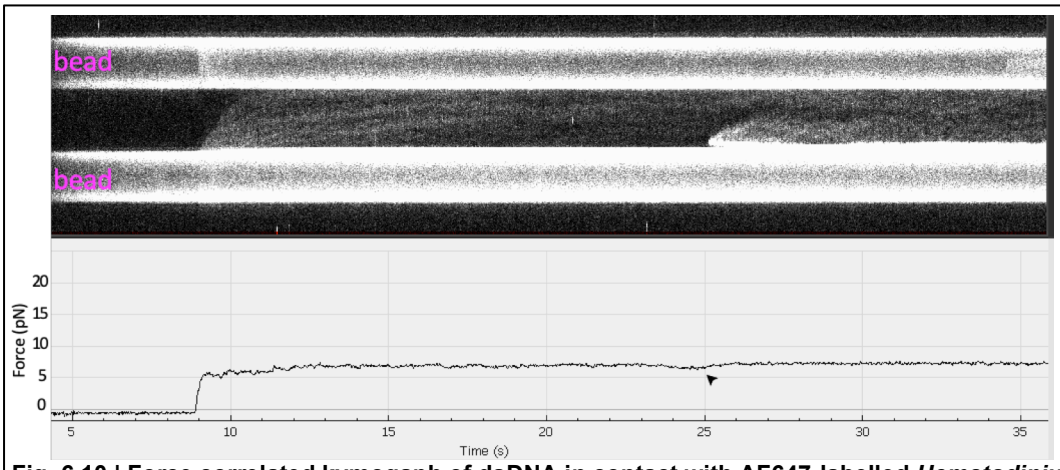


Fig. 6.10 | Force correlated kymogaph of dsDNA in contact with AF647-labelled *Hematodinium* DVNP.6. At 9 second, DVNP starts acting on DNA, applying a constant force. Note the additional foci formation at 25 seconds and the lack of correlation in force. Note the simultaneous formation of the dip in force marked by the arrowhead and the formation on one focus in the DNA.

Upon confirming that the labelled *Hematodinium* DVNP.6 demonstrated similar properties with respect to DNA interactions as the unlabelled DVNP.6, I then continued to perform co-imaging with dsDNA to observe the behaviour of DVNP proteins on dsDNA. The λ-DNA was moved into the protein channel, and the traps were imaged with the confocal microscope. In the first instance, the DNA, invisible in fluorescence channels, could be observed to become coated by the red protein (Fig. 6.11). Unexpectedly, spontaneous formation of bright globular foci could be observed by DVNP imaging. It was then observed that the foci were highly mobile, and the brighter of the foci seemed to have the agency to radiate smaller foci towards both sides (Fig. 6.11), processes as stochastic as foci formation itself. The mobility and direction of the motion of foci seemed to have no real correlation with one another; a focus could be completely immobile for minutes and then marched along the DNA for several micrometres in mere seconds.

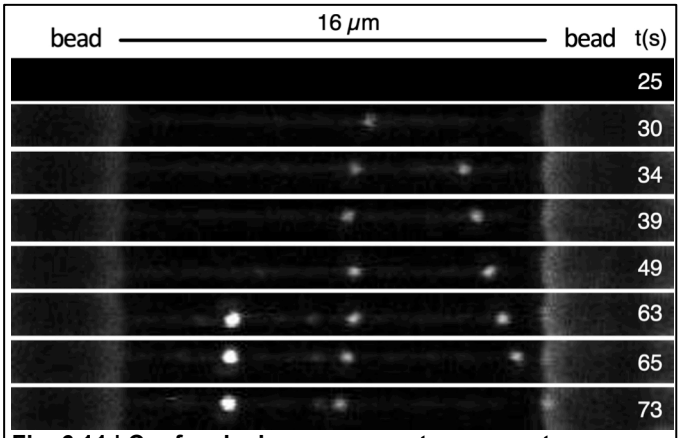


Fig. 6.11 | Confocal microscopy captures spontaneous formation and movements of bright foci by Hematodinium DVNP.6. Spheres on both ends are the polystyrene beads tethering the DNA.

To gain better resolution in time, I generated kymographs (Fig. 6.12). Again, the traps were moved into the protein channel after the kymograph capture had started. Remarkably, comet-like patterns representing the radiation from individual bright foci can be readily observed over time, as well as their mobility without any external stimuli. In Fig. 6.12A, the formation of comet patterns were dominant, and the formation of one comet seemed to have repelling or deflecting effects towards another. In Fig. 6.12B, instead, the foci were mostly immobile, and formation of comet patterns was not as prevalent, and the repelling effects were however not seen. The same experiment was also performed on the Alexa Fluor 647-labelled Ostreococcal viral DVNP (Fig. 6.12C). Bright foci were also observed, with discernable mobility though much less prominent than DVNP.6.

The brightness of the foci was also less than DVNP.6, and the comet pattern was not observed.

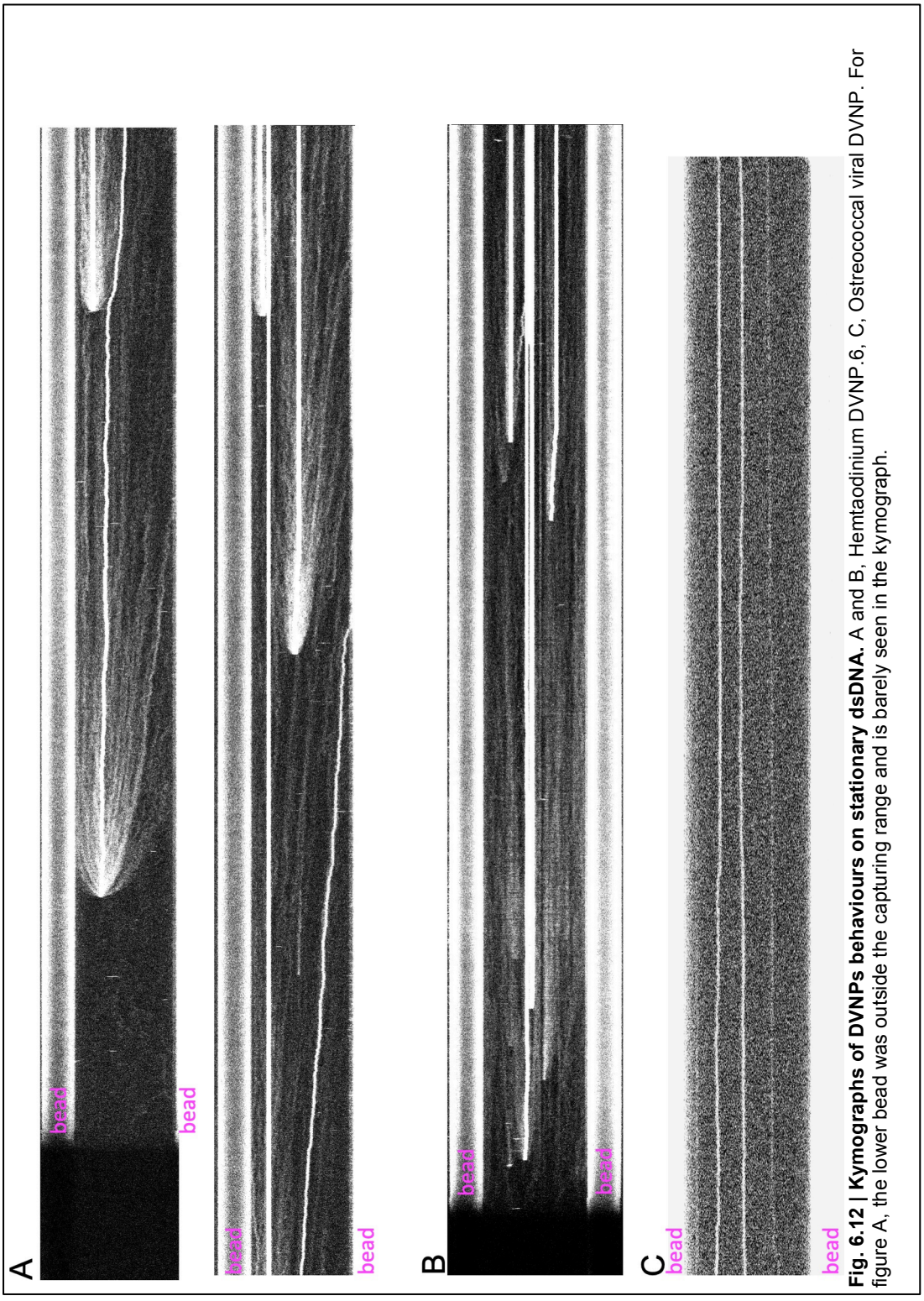


Fig. 6.12 | Kymographs of DVNPs behaviours on stationary dsDNA. A and B, Hemtaodinium DVNP.6, C, Ostreococcal viral DVNP. For figure A, the lower bead was outside the capturing range and is barely seen in the kymograph.

6.2.4 Spatial observations of compaction and decompaction across the linear DNA molecule

With the tool of labelled DVNPs, I was then able to ask more sophisticated questions regarding DVNP-DNA relationships. To elucidate whether the DNA compaction/decompaction events occur evenly throughout the whole DNA or there are local regions of tension points, I then performed DNA extension experiments with red DVNPs to correlate the labelled protein foci with the DNA extension (Fig. 6.13). During the progressive extension of DNA previously compacted by DVNPs, the bright red-DVNP foci served as markers on the DNA strand. For both proteins, it was observed that the distances between the foci did not increase with uniform ratios. In the case of *Hematodinium* DVNP.6, sharp displacement events in each focus occurred during the extension of the DNA. These steps did not concur in all foci and indicated that regions of local decompaction are occurring at different points in the DNA at different times (Fig. 6.13A). When force was plotted over time, these foci displacement events coincide with a short drop in force, evidence of this local decompaction and relaxation (Fig 6.13A inlet). For the Ostreococcal viral DVNP, a similar trend can be observed that sudden displacements of the foci seemed to correlate with sudden drops of DNA tension force (Fig. 6.13B). In the experiment demonstrated in Fig. 6.13B, one focus could be seen moving in opposite the direction to the extension of DNA and merging with another focus. Overall, the results presented here implied that each tooth on the force-extension curve represented a spatially local structural decompression event in the DNA as opposed to a global extension event.

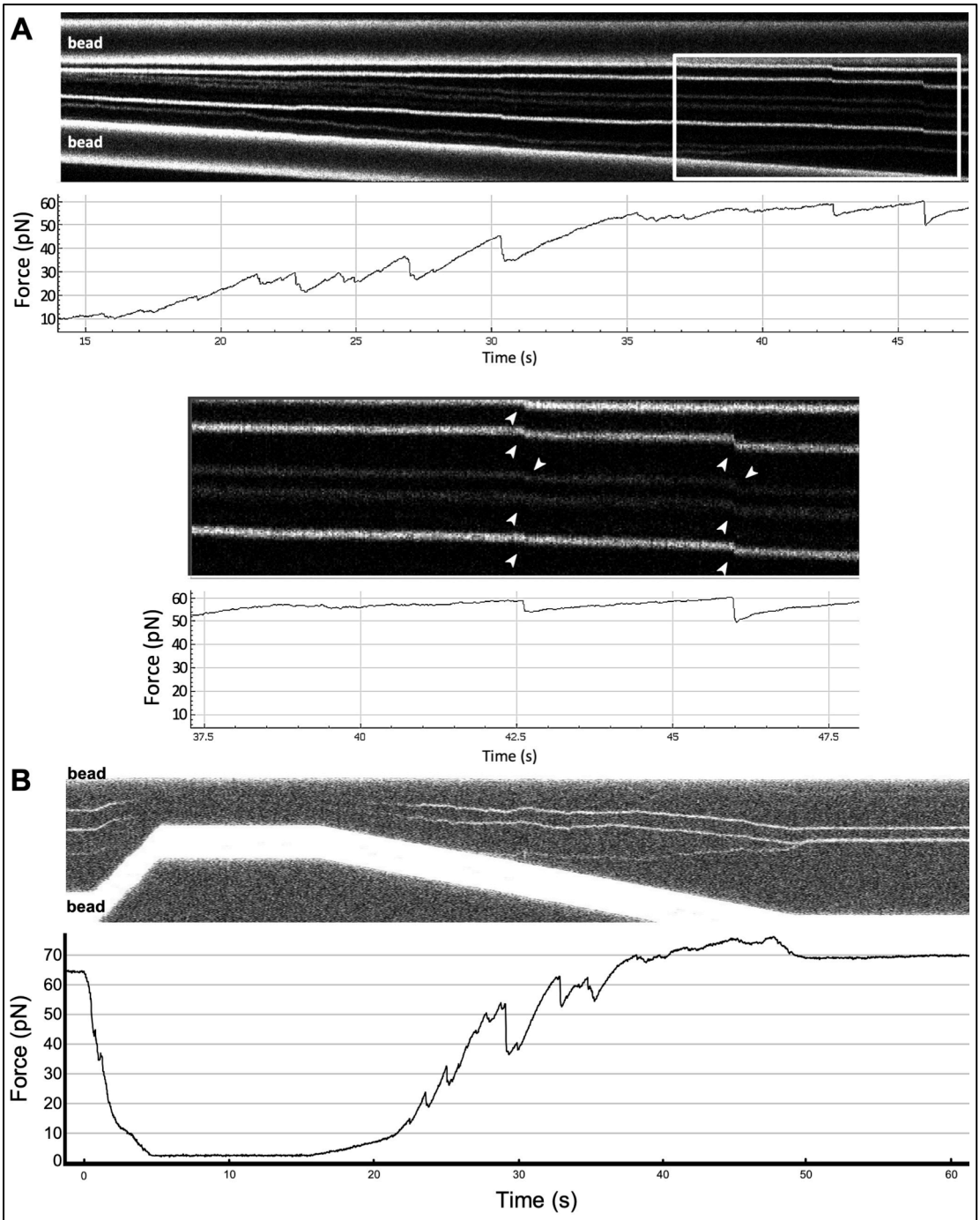


Fig. 6.13 | Correlating kymographs of Alexa Fluor-647 labelled DVNPs to force-extension curves. A, *Hematodinium* DVNP.6. Content in white box is zoomed in below. The sawtooth at 42.5 second can be traced to between the second and third foci in the kymograph, and the sawtooth at 46 second between first and second foci. B, Ostreococcal viral DVNP. Note the two foci merging at 50 seconds.

The tension-sensitive DNA dye Sytox Orange was then employed to assess the distribution of tension across the DNA and to correlate this with local regions of DVNP foci. Sytox is a tension-sensitive intercalating dye, which binds more strongly when the DNA is extended under tension (Biebricher *et al.* 2015; Flors *et al.* 2009). In the first instance, the red viral DVNP was co-imaged with Sytox Orange on a straightened λ-DNA (Fig. 6.14). The viral DVNP can be seen to form a large focus in the centre of the DNA. The Sytox dye showed even distribution across the DNA, with the exception of concentrated fluorescence at the location of the viral DVNP focus (Fig. 6.14). To capture higher resolution images with kymographs, new DNA molecules were captured and placed into the protein channel. However, the formations of the foci were stochastic and unpredictable, and DNA extension and relaxation were used to induce foci formation. Two such experiments were shown in Fig 6.15 where DNA extensions were performed. In both experiments, the λ-DNA was relaxed and moved into the protein channel, where an extension was performed afterwards. In 6.15A, one focus could be seen forming during the extension process, yet became highly mobile from the end of the extension process to the ensuing relative stasis period of the DNA. In 6.15B, one focus could be seen immediately after the DNA was moved into the protein channel, the event marked by the sudden increase in the intensity of the two beads in both channels. The Sytox channel, however, demonstrated elevated fluorescence at the focus as well even when the DNA is not under tension or extended. The focus in Fig. 6.15B disappeared after the extension has started, as did the Sytox fluorescence site. At full extension, the DNA was highly fluorescent in the Sytox channel; however, the staining was not homogeneous but rather patchily distributed, but no DVNP foci could be correlated to the Sytox staining pattern.

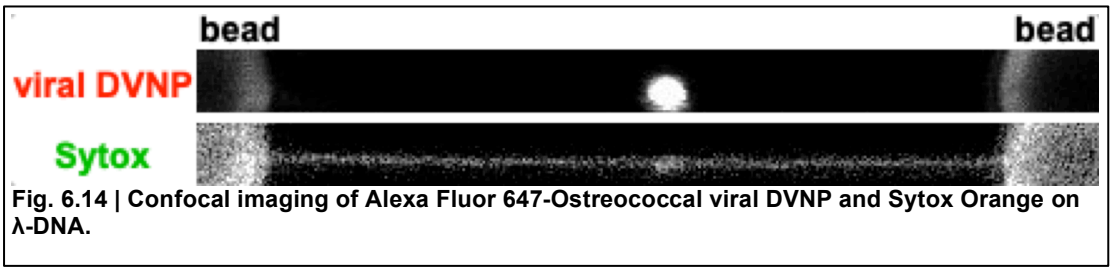


Fig. 6.14 | Confocal imaging of Alexa Fluor 647-Ostreococcal viral DVNP and Sytox Orange on λ-DNA.

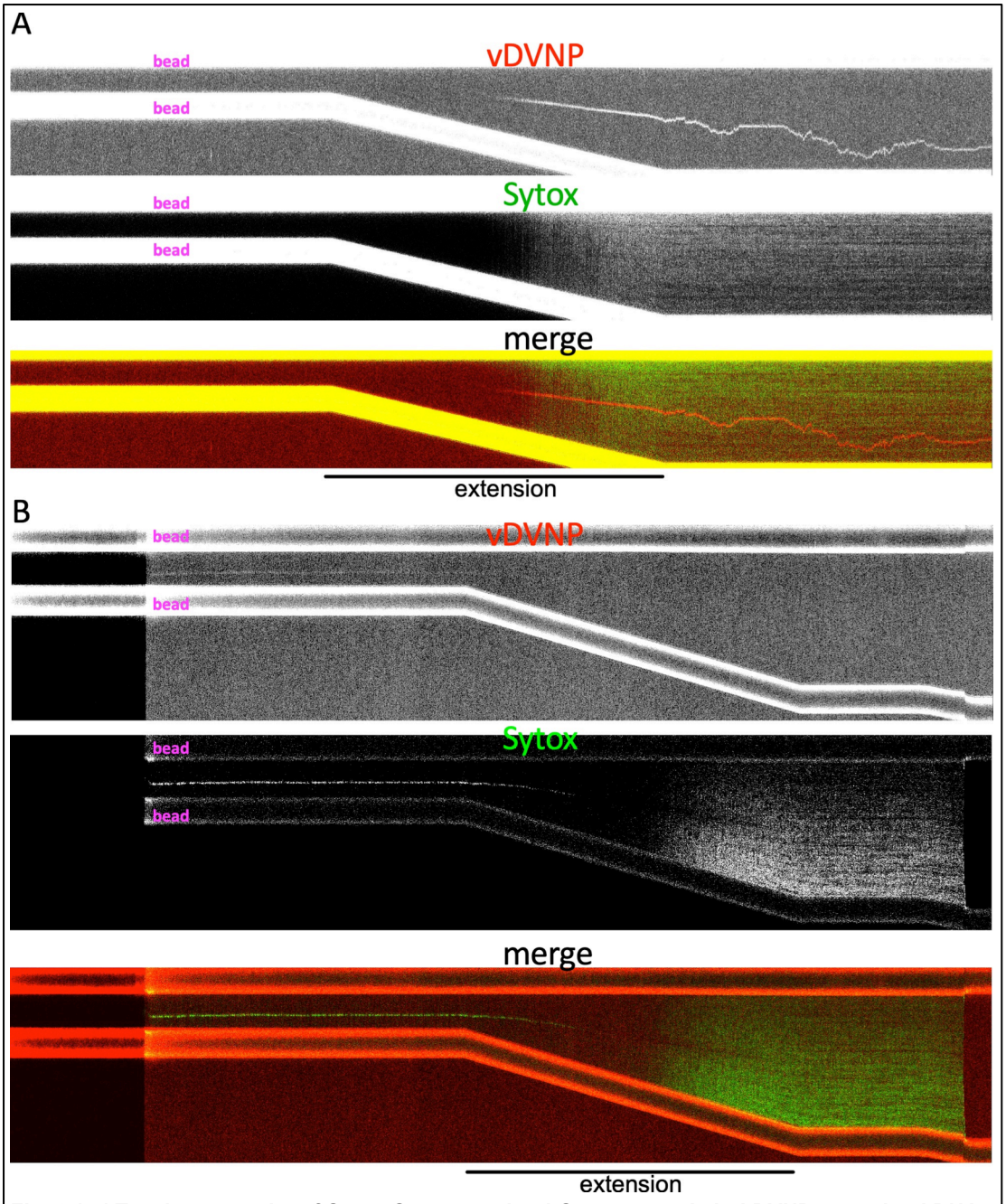


Fig. 6.15 | Two kymographs of Sytox Orange and red Ostreococcal viral DVNP co-stained DNA.

Hematodinium DVNP.6 was also co-imaged with Sytox Orange. Under identical conditions, DVNP.6 more readily formed multiple foci. For DVNP.6, two cycles of force clamp compaction and extension were performed for each molecule (Fig. 6.16). In the co-staining experiments, though with reliable foci formation, the previously observed comet pattern was not found. The bright DVNP foci signal, however, correlated well with the Sytox-negative parts of the DNA, and the two signals seemed to be negatively correlated (Fig. 6.9A). However, the phenomenon that was observed

on the viral DVNP, i.e. colocalisation of DVNP foci and Sytox signal under low tension, seemed to be also present, as marked by the arrowhead in Fig. 6.16B.

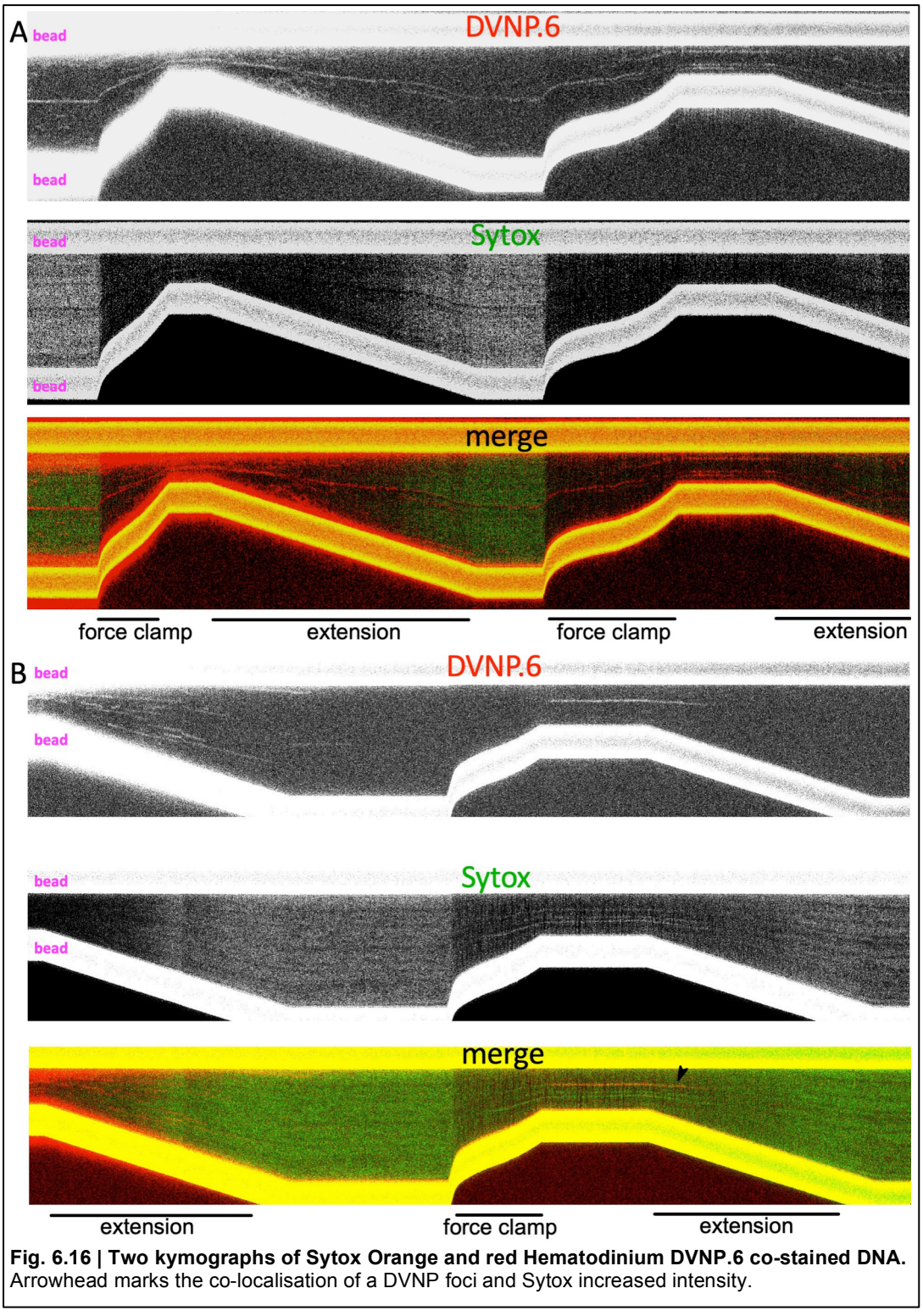


Fig. 6.16 | Two kymographs of Sytox Orange and red Hematodinium DVNP.6 co-stained DNA. Arrowhead marks the co-localisation of a DVNP foci and Sytox increased intensity.

6.3 Discussion

The results provided in this chapter demonstrated that the two DVNPs are both able to compact linear DNA. The DNA compaction activity could be clearly observed by the force clamp experiments described in Fig. 6.4 and the experiments in Fig. 6.5. In the force clamp experiment, to compensate for the force applied on DNA by the proteins, the system reduced the distance between the two traps. These data showed that force was maintained at a constant level throughout the whole process for both proteins, although *Hematodinium* DVNP.6 maintained a higher force of 5-7 pN whereas Ostreococcal viral DVNP barely above 1 pN. This result is rather counter-intuitive, as decreased length between traps should bring the relaxation of DNA and reduced amount of force on DNA. This is, however, not the case. This is might be caused by new protein-protein interaction establishing as the DVNP-coated dsDNA began to compact, creating more interfaces for interaction. The new interactions, in turn, could create the tension force faster than the trap could release by decreasing the distance between the two beads.

A sawtooth pattern is commonly observed in force-extension curves of other DNA packaging proteins measured by single-molecule spectroscopy. Multi-nucleosome associated DNA, for example, generates a sawtooth pattern in force-distance plots of about 20 pN in height (Fig. 6.17) (Brower-Toland *et al.* 2002). By comparison, DVNP-packaged DNA has very irregular sawteeth heights of less than 10 pN. These force responses seem chaotic and unpredictable when compared to a polynucleosome. Under identical concentrations, Ostreococcal viral DVNP produced much less prominent sawteeth than the dinoflagellate one, but viral DVNP of higher concentration generated an irregular sawtooth pattern very similar to a dinoflagellate curve with a similar baseline tension force (Figs. 6.6 & 6.7), implying that the two proteins are functionally similar in this respect and the differences were perhaps caused by the efficacy of the proteins' activities. Both DVNPs were able to recompact an extended DNA with free protein in the environment, and in the case of

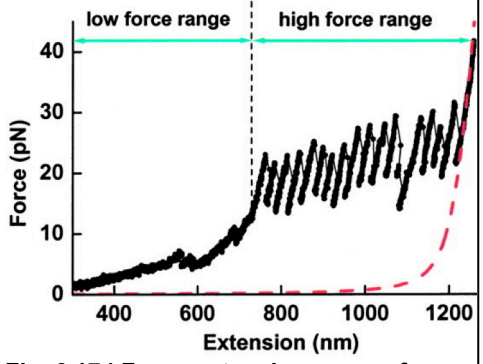


Fig. 6.17 | Force-extension curve of a polynucleosome. Adapted from Brower-Toland et al. 2002. Copyright (2002) National Academy of Sciences.

Hematodinium DVNP.6, a loaded DNA was able to recompact several times in the absence of free protein in the solution. This result also seems to support that *Hematodinium* DVNP.6 require less concentration to compact DNA. Nucleosome, however, cannot be reformed by simply allowing more exposure time of histone proteins with relaxed DNA. Using both fluorescently labelled DVNP proteins, it was possible to not only observe the DNA compaction and extension in real-time but also visualise the DNA compaction and sudden decompression events to regions of lengths observable by microscopy, and furthermore tracing the foci of DVNP proteins along with extension and compaction events. The ease of reestablishment of the interaction in 100 mM of KCl that was disrupted by the extension suggests that the interaction is not a complex large conformation change or higher-order DNA packaging event as the case of the nucleosome, but rather a simple, most likely hydrophobic and/or salt bridge-mediated protein-protein interaction, as depicted in diagram Fig. 6.18.

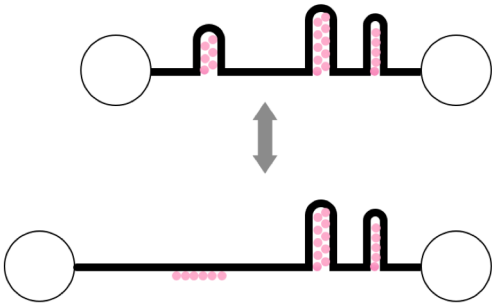


Fig. 6.18 | Proposed scenario for DVNP-mediated self-reestablishable DNA compaction observed by C-trap. White large circles, optical trapped beads. Black line, dsDNA. Pink solid circles, DVNP.

In a force-distance curve of a single-molecule dsDNA extension experiment, the slope of the elastic region of the curve represents the elasticity of the DNA molecule. *Hematodinium* DVNP.6 reduced this slope, implying the reduction of required force to extend the DNA. DVNP also increased the maximum length of DNA before melting, and therefore also the force required for this melting event. This type of alteration to the mechanical properties of DNA has been associated

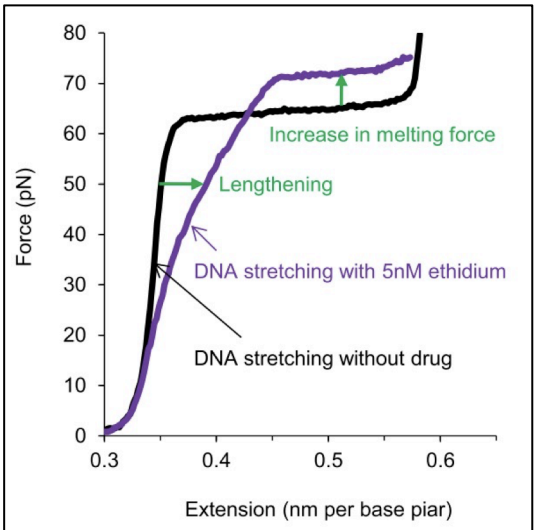


Fig. 6.19 | dsDNA elongation and stabilisation due to intercalation by ethidium on a force-extension curve. Reproduced from Almaqwashi et al. 2016 with permission.

with DNA intercalators, as intercalated object between the otherwise tightly packed stacks of DNA extends its length, decreasing the difficulty to further reach the fully extended state (Fig. 6.19; Almaqwashi et al. 2016; Wang et al. 2017).

In the first instance, it is hard to imagine that a DNA-binding protein larger than 10 kDa can be a DNA intercalator. However, there have been numerous reports of intercalation of amino acid side chains to DNAs causing DNA to kink, mostly DNA-binding transcription factors, including TATA-box binding protein (Kuznetsov *et al.* 2006; Sandmann & Sticht 2018). The involved residues are small hydrophobics (Val and Leu) and flat aromatics. The conserved aromatic residues in the C-ter half of dinoflagellate DVNPs, namely a GF motif, a Trp, and a Tyr (Fig. 3.2), could very well serve as the intercalating residues, and the conservation of them provides a clue for their possible importance. Note that in the conservation plot (Fig. 3.2), the length and composition of the peptide between the conserved core and these aromatics are relatively unconserved, and only the overall charge and the presence of the aromatics are. In addition, the mutual exclusive staining result of fluorescent DVNP and Sytox Orange in the single-molecule C-trap experiments further supports that the two may potentially compete for the same binding sites (Fig. 6.16A). Sytox Orange is a tension-sensitive dye, where DNA under tension decreases the dissociation constant of the dye to DNA, but the association constant maintains relatively constant (Biebricher *et al.* 2015). This result, however, can be interpreted in three possible ways: 1) DVNPs occupy the intercalation binding sites, possibly through the conserved aromatic side chains; 2) the DVNP-DNA complex forms a tightly packed cluster of which the steric hindrance suppresses the interaction of dye intercalation; or the 3) DVNP-DNA complex relaxes the DNA, resulting in local regions of less Sytox-binding (Biebricher *et al.* 2015). Nonetheless, the DVNP intercalator scenario fits better with the modified properties of DNA demonstrated by the reduced slope in the force-extension curve (Fig. 6.6), as well as the lack of comet patterns in the kymographs of red *Hematodinium* DVNP.6 and the intercalating dye Sytox co-staining experiment.

On the other hand, Ostreococcal viral DVNP also compacted DNA, yet did not seem to cause a reduced slope in the force-extension curve. Moreover, instead of mutual exclusion with Sytox Orange, which was observed with DVNP.6, viral DVNP seemed

to colocalise with Sytox when the DNA is relaxed. These imply that although the potential DNA intercalation activity is absent in viral DVNP, DVNP-DNA complexes or foci are under high tension to stabilise the Sytox binding, perhaps by generating curvature in the linear DNA and hence stretching it. Alternatively, the large focus may be a large DNA-DVNP complex with more DNA, and although the kinetics for the dye did not change, the higher amount of DNA brought more dye into the same focus. The results might also imply that the activity in *Hematodinium* DVNP.6 that causes the reduced slope in the force-extension curve, potentially intercalation, was acquired later after the viral and dinoflagellate DVNPs branched apart, assuming all dinoflagellate DVNPs have similar intercalation-like properties while their viral counterparts do not. Where exactly did the acquisition occur cannot be determined at the moment: a properly supported phylogenetic tree of all known dinoflagellate and viral DVNPs would provide some hint to this; however as demonstrated in Chapter 3 this is currently not possible. A larger sample pool of DVNPs from multiple dinoflagellates and viral lineages is also necessary to test the possibilities that either *Hematodinium* or *Ostreococcus* virus gain or lost the property independently.

For comparison, the bacterial DNA-compacting protein HU is a non-specific B-DNA minor groove binder (Serban et al. 2003). The results from the present study, that the two DVNPs may rely on intercalation to interact with double-stranded DNA, does not preclude the possibility of the proteins also acting with the major or minor grooves of DNA. During the course of the present study, an effort was made to understand the potential binding modes by co-imaging with DNA dyes with known binding properties and both Hematodinium and viral DVNP. This would also have allowed us to further verify the results obtained from the co-imaging with the intercalator dye Sytox Orange. However the C-trap apparatus that was available to the department was not compatible with the minor groove binder DAPI and Hoechst, and there are in effect no known major groove binder fluorescent DNA dye. I have later identified a classic histological chromogenic dye, Methyl Green, as a potential major groove binder that fluoresces in the far-red channel, however at this point the department has lost access to the correlative optical tweezers apparatus.

The phenomena of the observed foci fast and random mobility and splitting and merging (Figs. 6.12 & 6.13) are astounding. After all, it is difficult to comprehend

why a DNA-compacting protein (Chapter 5.2.1) is striding along the DNA, forming foci, and breaking apart, and even harder as to how it is done unprovoked and without any other environmental stimuli besides phosphate buffer and simple potassium chloride salt. Before I continue the discussion, let me summarise the dinokaryon conundrum and the results demonstrated in the previous chapters: permanently condensed chromosomes, bound and packaged presumably by the most abundant protein in the nucleus DVNP, with transcription happening only at the peripheries of the chromosomes, yet with seemingly ordinary genome structure. Perhaps the motile nature of DVNP actually is the very property that allows such an enigmatic system to work. In Chapter 2 I proposed the sewing machine transcription model that involves the RNA polymerase molecules anchoring on the outside of the chromosomes while DNA template travels from the chromosomes outwards, through the polymerase, and back into the chromosome, a process topologically similar to transcription factory. A relatively motile relationship between the DNA itself and the DNA packaging protein would fill in the gaps in the scenario described above: the protein provides an immobile scaffold for the DNA to travel on without the need to completely remodel the protein-DNA and protein-protein structure as transcription continues. For a canonical eukaryote, the elongation of transcription requires nucleosome displacement by the FACT complex before the arrival of the PolII large subunit and reforming the nucleosomes after transcription (Venkatesh & Workman 2015; Workman 2006). A dinokaryon that hosts no known nucleosomes could have replaced this process with merely maintaining the DNA fluidity, albeit in a permanently compacted liquid crystalline state.

The molecular mechanism of the sliding motion of DVNP is at the moment uncertain. There are, however, other DNA-sliding proteins that have structural basis, including the trimer DNA clamp PCNA (Fig. 6.20) (Clore 2011; De March *et al.* 2017; Iwahara *et al.* 2004; Omichinski *et al.* 1993). This protein has been shown to interact with the phosphate backbone atoms with the lysine and arginine side-chains, and proposed to slide by a ‘cogwheel’ motion where all the positive side chains interact with a neighbouring phosphate atom iteratively and stochastically. DVNP has a C-terminal tail (as well as N-terminal tail for the dinoflagellate ones) that is highly concentrated with lysines and arginines, and it is possible that DVNP uses a similar approach, forming salt bridges or polar interactions with the phosphate backbones. A stochastic

salt-bridges-mediated approach of making contacts with DNA backbone by DVNP also accounts for the mobility of the foci observed in the C-trap experiments.

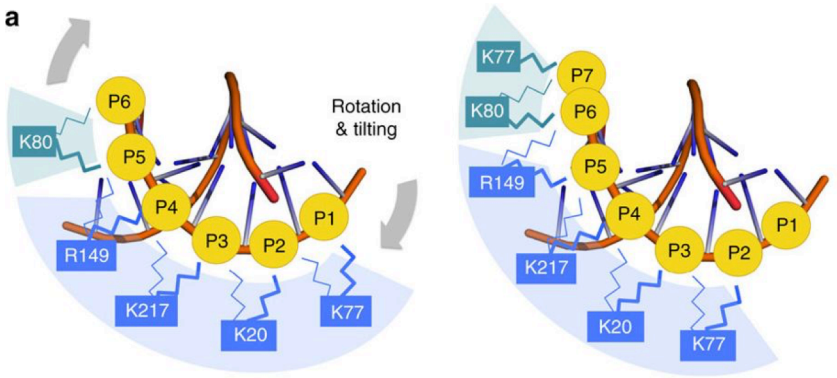


Fig. 6.20 | 'Cogwheel' DNA-sliding model of PCNA. The positive side chains stochastically form new bonds individually with the phosphate atom before or after the originally interactive phosphate, and when all residues make contact with the adjacent phosphate, the protein moves forward by 1 bp. Reproduced from de March et al. 2017 (CC 4.0).

The bacterial HU protein seems to share some commonality. Although known as a non-specific minor groove binder, the binding of HU to DNA involves securing a phosphate backbone of DNA between Gly46 and Lys83, and a proline is used to intercalate the DNA and secure the binding (Swinger & Rice 2004). Mutagenesis suggested importance of a series of arginines and lysines, also similar to the highly positively charged DVNP (Bhowmick et al. 2014). In addition, HU has been shown to have a dimerisation activity (White et al. 1999; Swinger 2003), although the mobility of the protein on DNA has not been observed yet. The interaction mechanism of HU and dsDNA also supports that DVNP may be adopting such a similarly unusual mechanism.

6.4 Methods

Labelling DVNP with Alexa Fluor 647

500 µL of 100 µM of either *Hematodinium* DVNP.6 or Ostreococcal viral DVNP in 20 mM sodium phosphate pH7 and 100 mM potassium chloride prepared as described in Chapter 5 was used for the labelling reactions. Alexa Fluor 647 NHS-ester was purchase from Thermofisher. The dye was dissolved in dry DMF at 1 mM. Then the dye was mixed with protein samples at dye:protein ratio of 1:4, briefly mixed, wrapped in foil, and left at 4°C overnight. The reaction mixture was then dialysed against 20 mM sodium phosphate and 100 mM potassium chloride for 48 hours, with 3 changes of buffer per day. The mixtures was quantified by measuring absorbance at 205 mM with a Nanodrop One spectrophotometer and flash frozen in liquid nitrogen and stored in -80° until required.

C-trap® dual trap optical tweezers single-molecule spectroscopy

Laser traps were firstly used to trap streptavidin coated polystyrene beads. Afterwards the traps were moved through a DNA channel to tether DNA between the two beads, and a force-extension curve was then performed to ensure only one DNA molecule was tethered. The traps were then moved into sample channel, which can be loaded with either labelled or unlabelled protein with the combination of Sytox Orange when required. All images and data were captured by proprietary software Bluelake designed and developed by Lumicks. Images were output in tiff format, and force information was recorded in hierarchical hdf5 format. The information was then replotted by either OriginLab analysis software or by Graphpad Prism after data export with HDFView freeware.

Technical support received:

C-trap was operated by or under supervision of Lumicks product specialist Joanna Andrecka, as well as screen captures of the Bluelake software of the C-trap system or OriginLab plot outputs.

Chapter 7: Understanding the structural properties of DVNPs by NMR spectroscopy

7.1 Introduction

To elucidate the molecular mechanism of the activities of DVNPs observed in chapters 5 and 6, insight into the structures of DVNP proteins was sought. Due to its small and hydrophilic nature, the predicted probability of successfully forming a crystal suitable for X-ray crystallography was low, and both *Hematodinium* DVNP.6 and Ostreococcal viral DVNP were predicted to be 'recalcitrant' to crystallisation (Overton *et al.* 2008; Overton & Barton 2006). I, therefore, sought to gain more insight into the structural properties of DVNP by nuclear magnetic resonance (NMR) spectroscopy.

When compared with X-ray diffraction, NMR has the advantage of analysing the proteins closer to their native states: salt conditions suitable for a particular protein have a high chance to be compatible with NMR, but this generally is not the case with protein crystallisation. Also when designed and adjusted carefully, NMR can capture potential inter- and intra-molecular dynamics as well as properties such as secondary structure boundaries within a protein (Roberts 1993), whereas X-ray crystallography captures only a snapshot of what might be a dynamic protein. This aspect is especially important when the subject is an Intrinsically Disordered Protein (IDP) or has abundant dynamics in its activity.

In this chapter, I seek to express large amounts of ^{15}N and/or ^{13}C labelled *Hematodinium* DVNP.6 and Ostreoccocal viral DVNP, and employed NMR methodologies to gain insight into the structures of these novel nucleoproteins.

7.2 Results

7.2.1 Expression and purification of isotopically labelled DVNP proteins suitable for NMR spectroscopy analysis

To produce labelled DVNP protein suitable for NMR analysis, two minimal media expression schemes were tested. Firstly, DVNP-expressing bacteria were grown and induced in M9 minimal medium. Secondly, DVNP-expressing bacteria were grown in the rich medium LB, and after O.D. reached 0.5, the cells were transferred to the minimal M9 medium for protein expression (Fig. 7.1). Eventually, it was determined that the optimal conditions were sole growth in M9 minimal medium and 37°C induction for 5 hours due to the much less non-target bacterial proteins in this route (Fig. 7.1). Choice of carbon source was glucose solely, nitrogen ammonium chloride and ^{15}N labelled celtone supplement to boost the otherwise slow growth rate. The purification strategy was identical to that described in Chapter 5, utilising a self-cleavable chitin-binding fusion partner and chitin beads column by gravity-fed affinity chromatography followed by gel filtration.

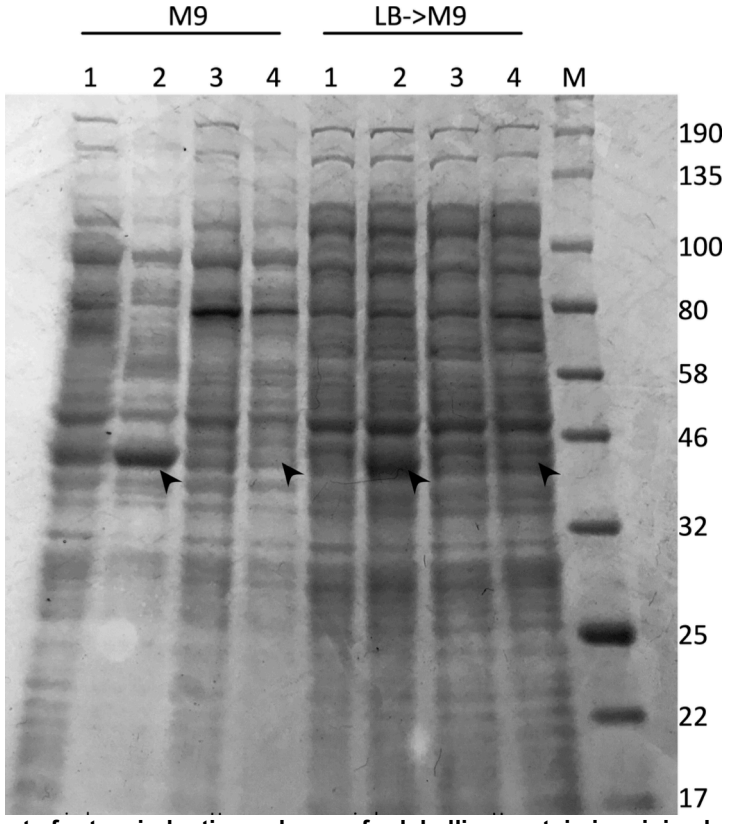


Fig. 7.1 | Tests for two induction schemes for labelling protein in minimal media. M9, growth and induction in identical M9 medium. LB->M9, growth in LB, induction in M9 media. Both were induced at 37°C for 5 hours. Loading in each lane was standardised with absorbance at 600 nm. Lane 1, viral DVNP, pre-induction. 2, viral DVNP, post induction. 3, DVNP.6, pre-induction. 4, DVNP.6, post-induction. Arrows, location of expected overexpressed protein. M, marker. Unit, kDa.

^{15}N labelled *Hematodinium* DVNP.6 and Ostreococcal virus DVNP were both expressed in 3L batches and purified. Both proteins were expressed and purified utilising the same methodologies described in Chapter 5, and SDS-PAGE confirmed that both samples were adequate for NMR analyses (Fig. 5.2). The expression tests shown in Fig. 7.1 suggested that for both expression schemes, overexpression of viral DVNP is much better than *Hematodinium* DVNP.6 judging by the intensity of the target band.

Based on the better resolved ^{15}N-heteronuclear single-quantum coherence (to be described in the next section) and hence a better chance leading to fruitful structural insights, the viral DVNP was first pursued for further ^{13}C labelling. The expression was performed in 3L-size expression batches, and high purity of highly concentrated ^{13}C-labelled viral DVNP was acquired through gel filtration chromatography and protein concentrator (Fig. 7.2).

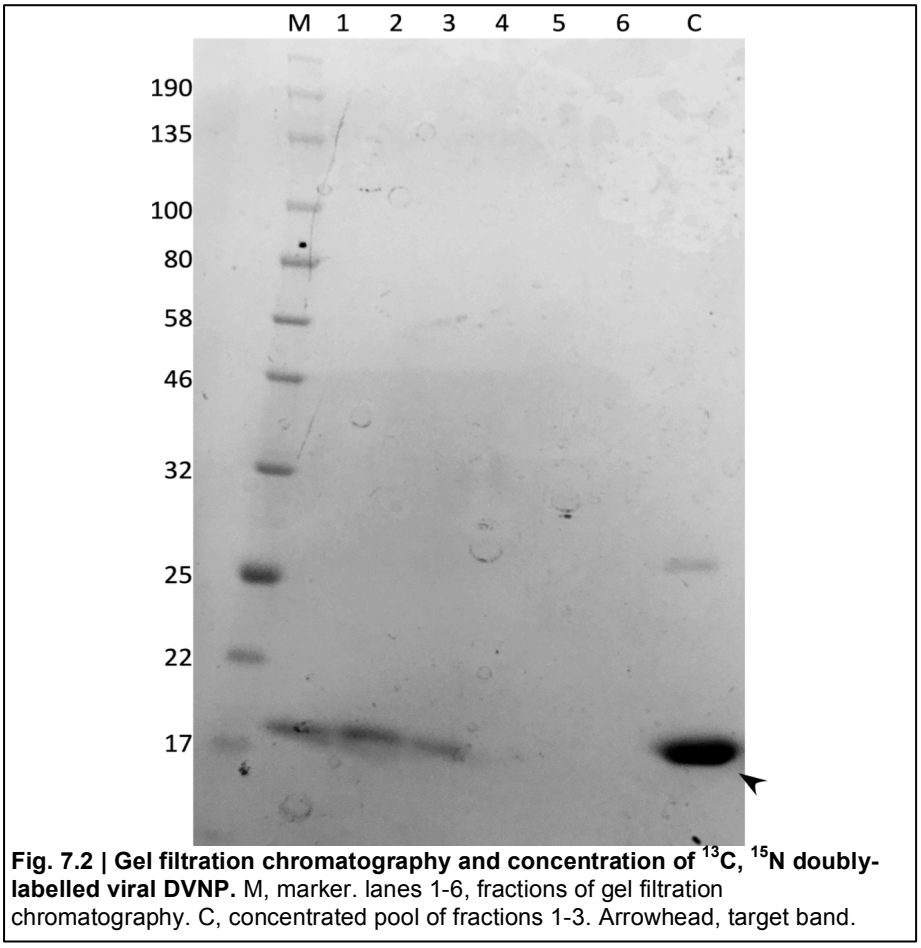


Fig. 7.2 | Gel filtration chromatography and concentration of ^{13}C, ^{15}N doubly-labelled viral DVNP. M, marker. lanes 1-6, fractions of gel filtration chromatography. C, concentrated pool of fractions 1-3. Arrowhead, target band.

7.2.2 The general molecular dynamics and structural states of DVNPs

To understand the structural states of DVNPs, purified ^{15}N-labelled *Hematodinium* DVNP.6 and viral DVNP were subjected to ^{15}N-heteronuclear single-quantum coherence (HSQC), a method that measures the immediate chemical environment of each H-N bond. Both *Hematodinium* DVNP.6 and viral DVNP samples were recorded at 278, 288, and 289°K. In addition, viral DVNP was also recorded at 273°K. The best-resolved spectra (278°K) were later used for data analyses (Fig. 7.3). For the viral DVNP, roughly half of all the observable peaks were centred in the middle of the spectrum with respect to both the X-axis (^{15}N) and Y-axis (^{1}H). The other half of peaks, however, displayed wide spreads over both axes. On the other hand, the *Hematodinium* DVNP.6 spectrum showed much less peak dispersion by HSQC, regardless of the temperatures (Fig 7.3). Besides the doublet peaks on the upper right-hand side of the HSQC, around 7.0~7.5 ppm ^{15}N, 107~112 ppm ^{1}H, which represent the amino side-chains of Arg, Asp, and Glu, the rest of the peaks representing the backbone amide bonds showed very little dispersion. These results suggested that for viral DVNP, roughly half of the viral protein shows a high level of disorder, displaying low chemical shift dispersion and narrow line widths in both the ^{15}N and ^{1}H dimension, whereas the other half showed signs of stable structures, manifested by the distinctness in the immediate chemical environments of the H-N bonds hence the dispersion of the peaks. By comparison, the *Hematodinium* DVNP.6 showed little signs of structures, judging by the lack of peak dispersion.

In order to capture DVNP-DNA interactions by NMR, a short double-strand DNA oligo of 20 bp was titrated into the viral DVNP sample (sequence is shown in Appendix 5). At equal charge ratio of both viral DVNP and DNA, the highly concentrated protein quickly formed non-soluble coacervates and precipitated to the bottom of the NMR tube and failed to generate any signal by the spectroscope (not shown). Consequently, more DNA was added to the sample, and at equal molar ratio (1:2 protein to DNA charge ratio), some of the coacervates seemed to have redissolved back into the solution, determined by the reduction of turbidity by the human eye (not shown). A ^{15}N-HSQC spectrum was subsequently recorded. The overall spectrum is reminiscent of the monomer viral DVNP protein; however most distinct and dispersed peaks had disappeared, and the existing peaks in the centre of the plot had become broader with a much lower signal to noise ratio (Fig 7.4). The

disappearance of the structured peaks suggests that these residues are perhaps in a larger structure that is not responsive within the time frame of the NMR signal collection. The broadened peaks in the disordered region suggest that each of the corresponding residues may be in multiple states, e.g. multiple binding conditions, and HSQC had captured the average of the conditions.

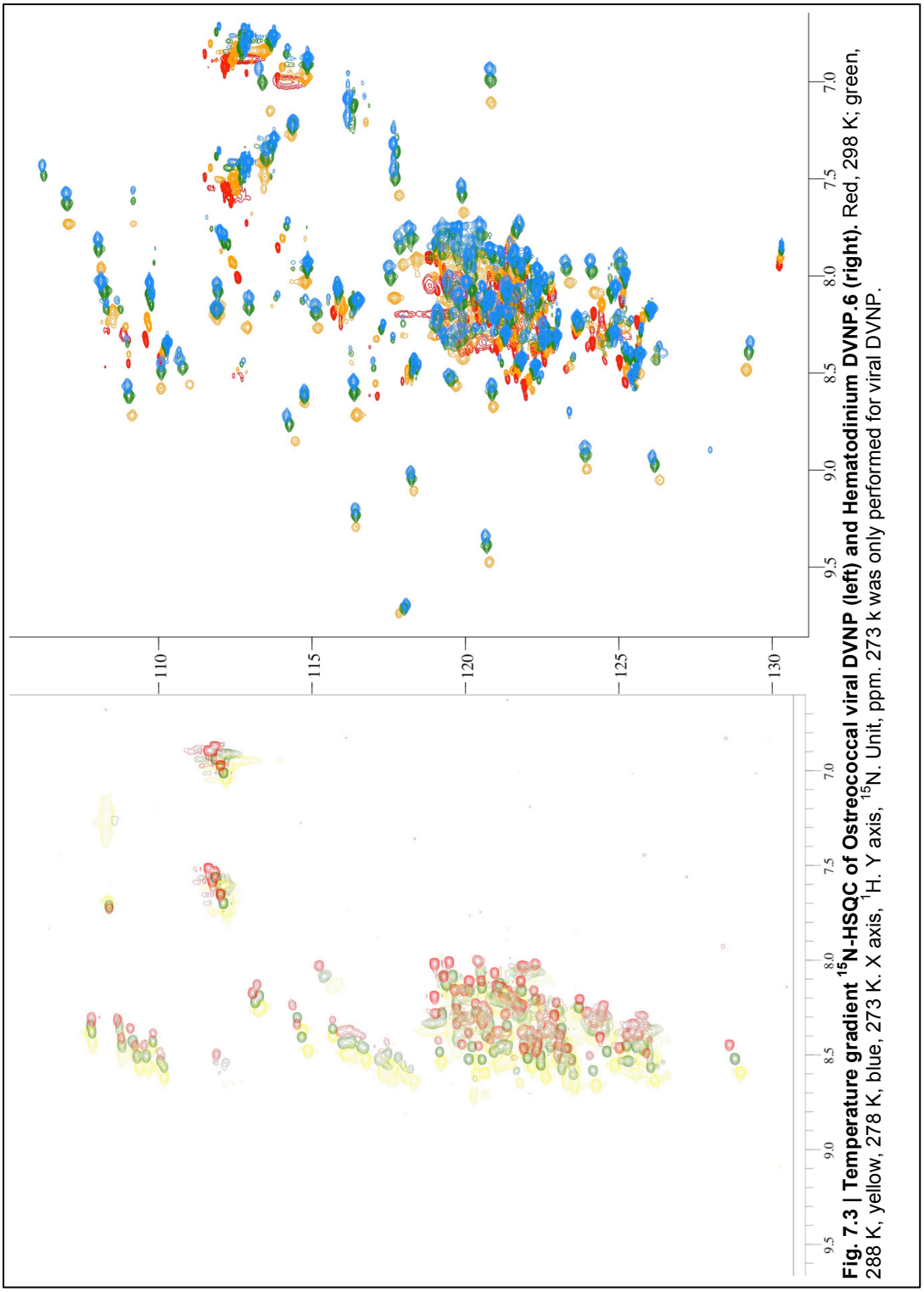


Fig. 7.3 | Temperature gradient ^{15}N-HSQC of Ostreococcal viral DVNP (left) and Hematodinium DVNP.6 (right). Red, 298 K; green, 288 K, yellow, 278 K, blue, 273 K. X axis, ^{1}H. Y axis, ^{15}N. Unit, ppm. 273 k was only performed for viral DVNP.

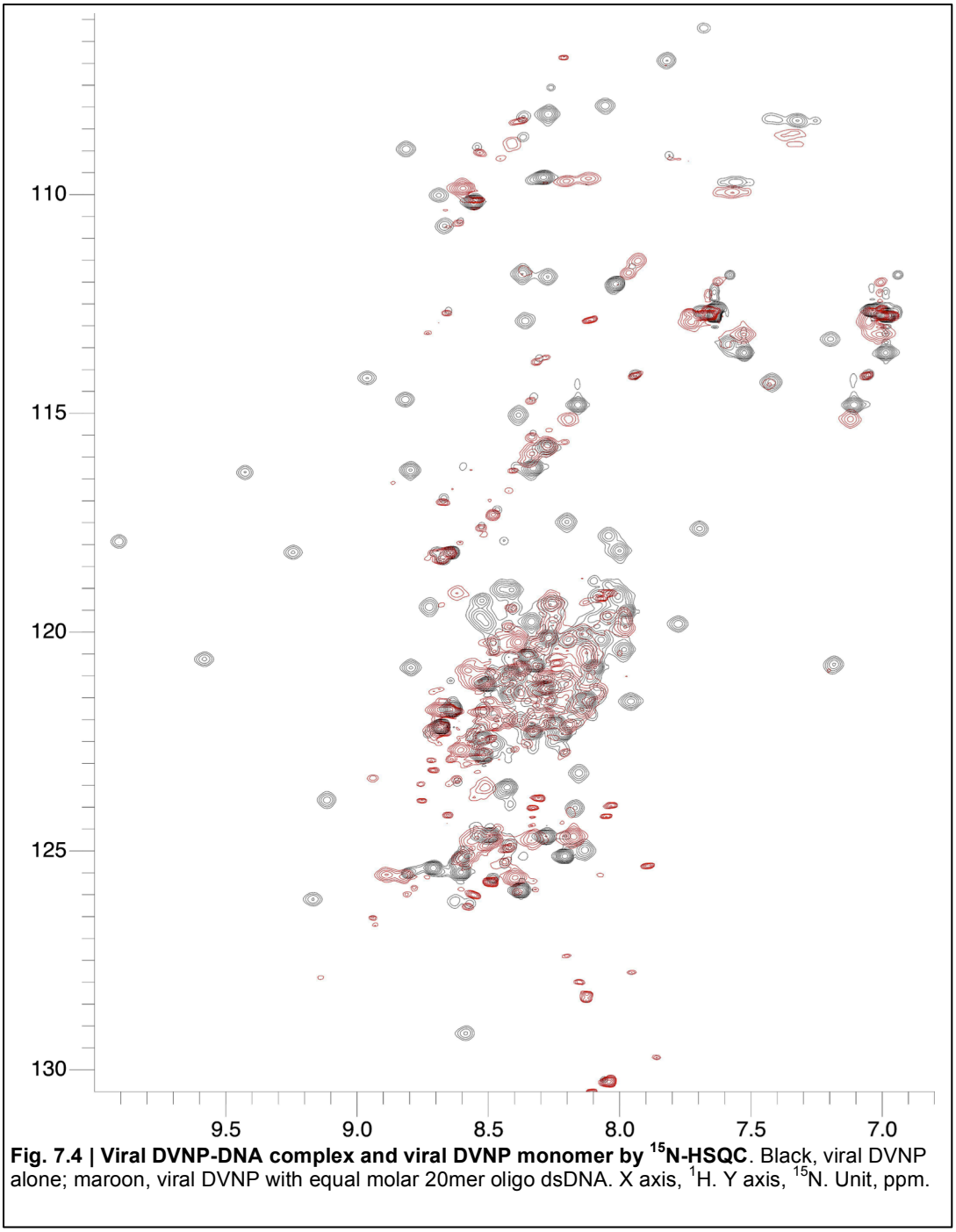


Fig. 7.4 | Viral DVNP-DNA complex and viral DVNP monomer by ^{15}N-HSQC. Black, viral DVNP alone; maroon, viral DVNP with equal molar 20mer oligo dsDNA. X axis, ^{1}H. Y axis, ^{15}N. Unit, ppm.

7.2.3 The secondary structure of Ostreococcal viral DVNP

To generate secondary structural information of DVNP, the viral DVNP was selected to be labelled with both ^{13}C and ^{15}N for further 3-dimensional NMR experiments. Due to the results from previous chapters that the two DVNP proteins share considerable similarities, it was anticipated that the structural knowledge learnt from viral DVNP would be representative the *Hematodinium* DVNP as well. In addition, the viral DVNP presented a ^{15}N-HSQC with much more potential structural information available. After purification, the doubly labelled protein was used to produce several 3D through-bond spectra by NMR spectroscopy in order to assign the Hα and Hβ of the protein, including HNCA, HNCACB, HNCOCA, HNCOCACB, HNCO, and HNN. Each residue in the protein in each spectrum correlates to the residue preceding and/or succeeding it, and all residues were identified by comparing all the spectra, or 'walking' the carbon backbone (Fig. 7.5). During the multi-dimension carbon walking process, the peaks of the protein in the ^{15}N-HSQC map of Ostreococcal viral DVNP were assigned (Fig. 7.6).

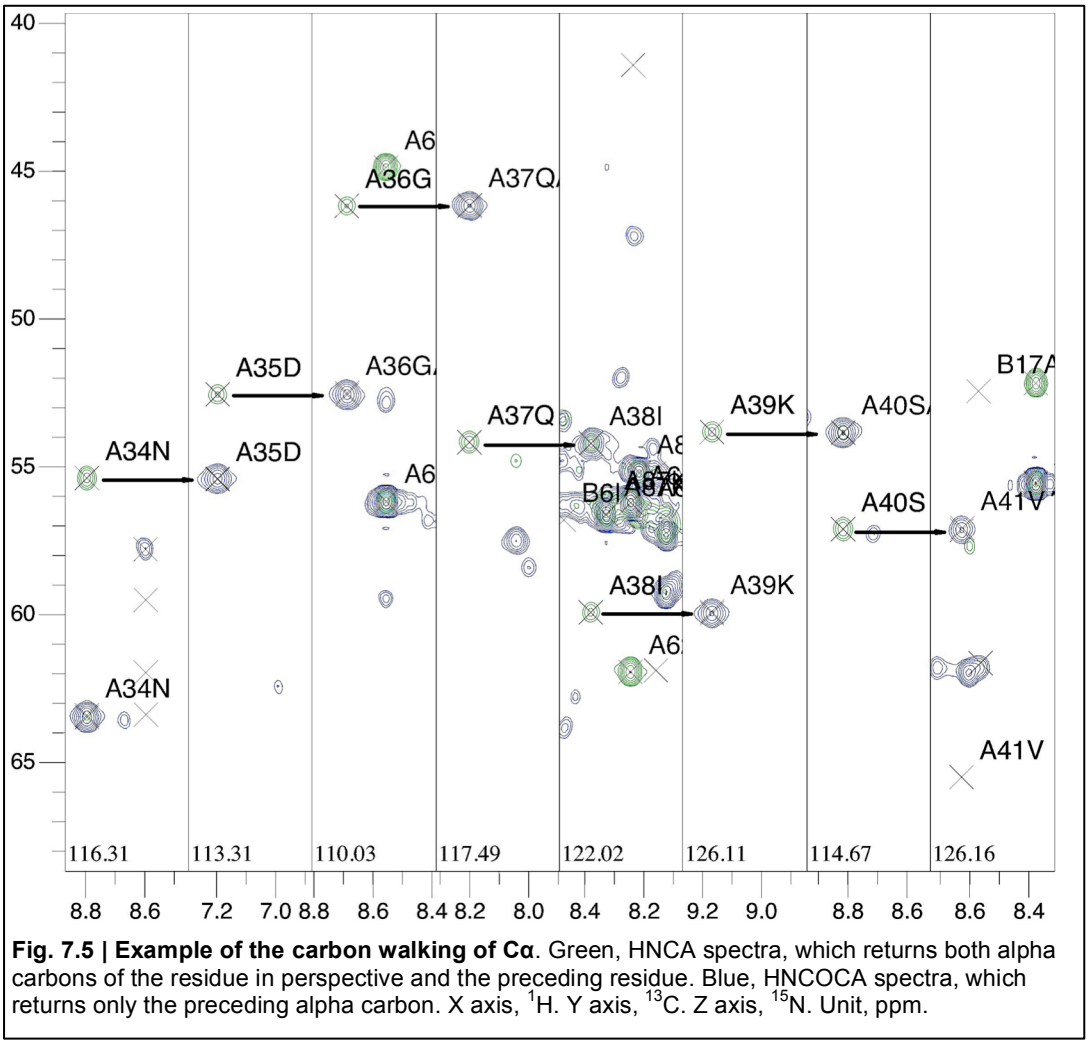


Fig. 7.5 | Example of the carbon walking of Cα. Green, HNCA spectra, which returns both alpha carbons of the residue in perspective and the preceding residue. Blue, HNCOCA spectra, which returns only the preceding alpha carbon. X axis, ^{1}H. Y axis, ^{13}C. Z axis, ^{15}N. Unit, ppm.

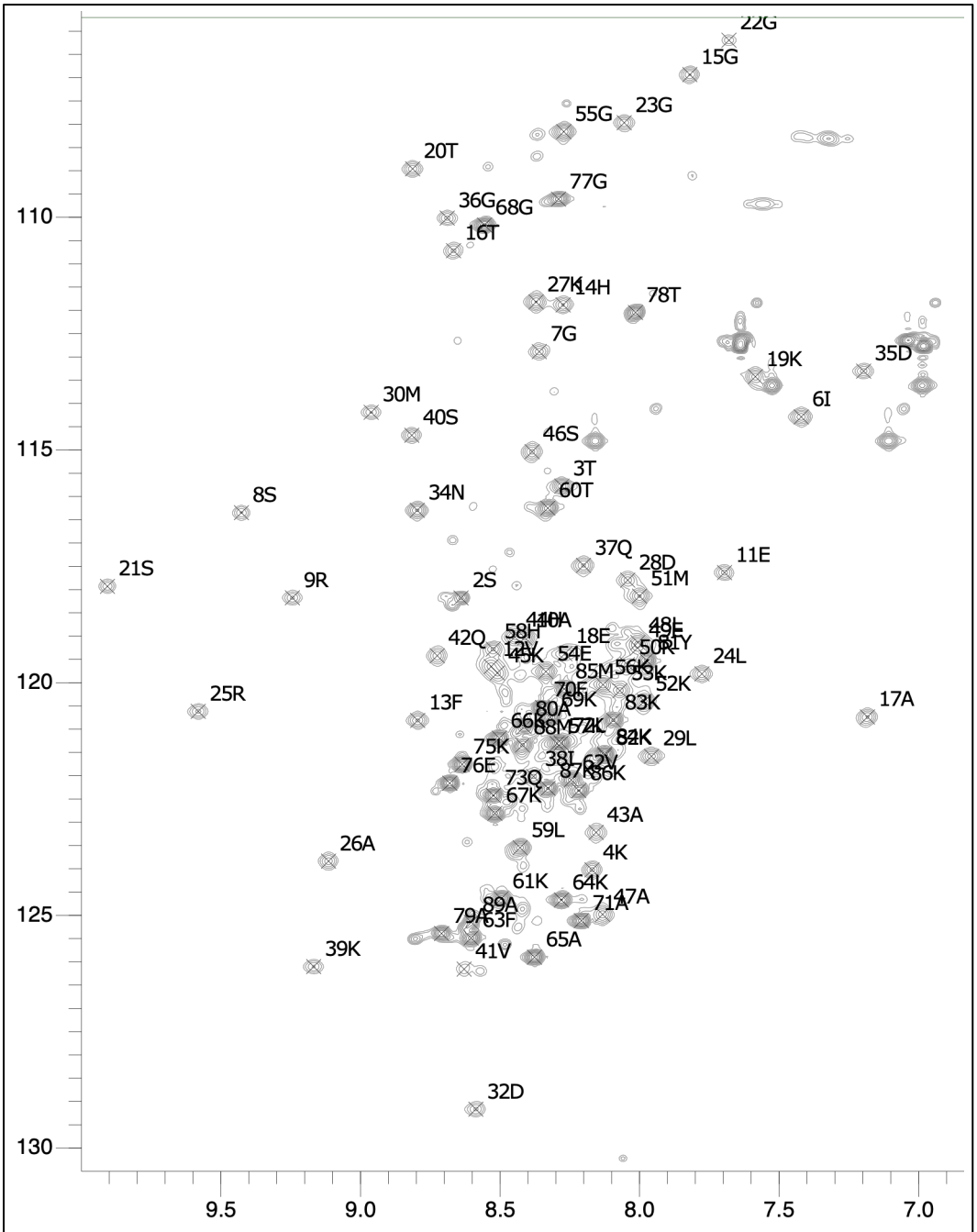


Fig. 7.6 | Annotated 15N-HSQC spectrum of Ostreococcal viral DVNP. Unassigned peaks on the upper right corner are amide side-chains. Minor peaks represents another folding form, which will be discussed later in the chapter. X axis, ^{1}H. Y axis, ^{15}N. Unit, ppm.

Once residue spectra had been identified, information regarding the secondary structure of the viral DVNP was derived by determining the dihedral angles φ and ψ. The Hα and Hβ angles were then calculated by two algorithms, DANGLE and TALOS, and the secondary structure was generated (Fig. 7.7 and Table 7.1). DANGLE utilises *de novo* generation of possible dihedral angles for each residue from *ad hoc* NMR information as well as consideration of the previous and following amino acids with a Bayesian approach (Cheung *et al.* 2010). TALOS works by a data mining approach with an extensive database of protein with known structures (Cornilescu *et al.* 1999). The two calculation algorithms produced very similar results; the overall secondary structure displays a small helix at residues 9-13, two small extended sheets between residues 25 and 40, and a large helix from 42 to 55. At several residues where DANGLE did not have enough confidence to predict a secondary structure including 7G, 18E, 36G, 54E 55G, and 68G, TALOS invariably gave a prediction of a coil. Notably, 80A and 81Y were predicted by DANGLE as coil whereas TALOS predicted the two residues as a small helix with low confidence. Prior to NMR-based modelling of the secondary structure of DVNP, the *in silico* method EMBOSS tool garnier (Garnier *et al.* 1978) was used to test for any predicted secondary structures. When the outputs of these predictions were compared, the results displayed a much different scenario (Fig. 7.7). As the two evidence-based methods, DANGLE and TALOS provided similar results, it is likely that the tool garnier is not adequate for secondary structure prediction of the protein DVNP.

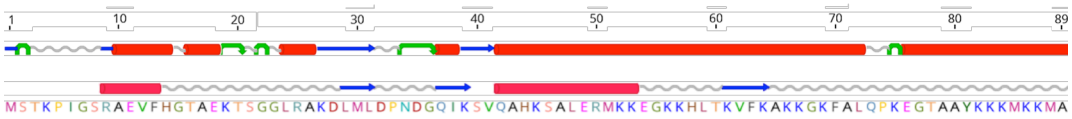


Fig. 7.7 | Illustration of secondary structure predicted by EMBOSS garnier versus calculated results of DANGLE. Upper, predicted by garnier. Lower, calculated results from DANGLE. Red, helix; blue, extended; grey, coil; green, turn.

Heteronuclear Overhauser effect (hetNOE) spectra were recorded to verify the predicted secondary structure independently (Fig. 7.8). Heteronuclear NOE reports the tumbling rate ratio of each residue versus the whole protein by measuring the signal intensities of the protein at different magnetisation pulse regimes. Its result is, therefore, sensitive to secondary structures that can limit residue freedom of movement. The hetNOE of viral DVNP indicated a relatively stable structure starting

from the third residue from N-terminus of the protein; then a sudden drop of structuredness just after residue 55. Finally, a small stretch of residues towards the C-ter end, roughly residues 75-85, indicated some structure. The hetNOE result is mostly in agreement with the secondary structure calculation results; the sudden drop of tumbling ratio at residues 15 and 55 both correspond to the terminations of calculated secondary structures (Table 7.1). The structure at the C-terminal end predicted by hetNOE was not captured in DANGLE, however, exists as a low confidence small helix in TALOS. Interestingly, both secondary structure calculators identify a small stretch of extended sheet at around residue 62, yet this is not supported by heteronuclear NOE.

viral DVNP		DANGLE								TALOS	
No	seq	ϕExp	ϕUpper	ϕLower	ψExp	ψUpper	ψLower	Ω	SS	CONF.	SS
1	M									0	C
2	S									0.94	C
3	T	-92.67	-40	-150	127.45	179.99	80	180	C	0.93	C
4	K	-71.91	-50	-150	146.02	179.99	100	180	C	0.95	C
5	P	-60.36	-40	-90	140.74	170	120	180	C	0.88	C
6	I	-135.9	-90	-180	156.58	-170	110	180	C	0.54	C
7	G	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.35	C
8	S	-87.66	-50	-150	164.98	-170	130	180	C	0.83	C
9	R	-55.01	-40	-80	-35.03	-10	-60	180	H	0.66	H
10	A	-64.07	-40	-90	-39.53	-20	-70	180	H	0.96	H
11	E	-64.72	-40	-90	-43.4	-20	-70	180	H	0.98	H
12	V	-64.58	-40	-90	-42.84	-10	-70	180	H	0.98	H
13	F	-59.97	-40	-90	-35.21	-10	-60	180	H	0.79	H
14	H	-85.18	-60	-110	-5.87	30	-40	180	C	0.19	C
15	G	75.89	110	50	22.95	50	-30	180	C	0.89	C
16	T	-71.77	-40	-110	145.52	179.99	90	180	C	0.88	C
17	A	-135.41	-90	-180	163.37	179.99	130	180	C	0.75	C
18	E	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.84	C
19	K	-135.32	-70	-180	163.26	-170	130	180	C	0.83	C
20	T	-123.55	-80	-170	135.4	179.99	100	180	E	0.86	C
21	S	-58.14	-30	-100	129.97	170	100	180	C	0.89	C
22	G	-85.87	-60	-120	-0.52	30	-30	180	C	0.92	C
23	G	94.34	120	60	-3.13	50	-30	180	C	0.9	C
24	C	-93.58	-50	-120	135.02	160	100	180	C	0.83	C
25	R	-75.69	-50	-130	155.74	179.99	130	180	C	0.78	C
26	A	-55.11	-40	-80	-34.98	-10	-60	180	H	0.47	H
27	K	-95.49	-60	-140	-3.99	40	-60	180	C	0.48	H
28	D	-115.05	-40	-180	140.77	-150	70	180	C	0.41	C
29	C	-150.74	-90	-180	155.69	179.99	120	180	E	0.57	E
30	M	-135.26	-90	-180	152.83	179.99	120	180	E	0.81	E
31	C	-115.91	-70	-150	123.7	160	90	180	E	0.15	E
32	D	-75.41	-40	-130	116.27	179.99	70	180	C	0.86	C
33	P	-63.69	-40	-100	-17.16	10	-50	180	C	0.9	C
34	N	-74.3	-40	-120	-25.08	20	-60	180	C	0.91	C
35	D	-79.31	-40	-160	144.15	179.99	90	180	C	0.96	C
36	G	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.95	C
37	Q	-66.98	-40	-180	147.16	-170	100	180	C	0.69	C
38	I	-84.26	-50	-130	118.08	160	100	180	C	0.12	E
39	K	-132.8	-100	-160	152.4	179.99	110	180	E	0.38	E
40	S	-96.22	-50	-160	126.51	170	90	180	C	0.4	E
41	V	-61.3	-30	-100	-22.11	10	-60	180	C	0.66	H
42	Q	-64.85	-40	-90	-25.11	0	-70	180	H	0.85	H
43	A	-68.04	-40	-90	-37.34	-10	-70	180	H	0.87	H
44	H	-58.31	-40	-90	-43.55	-20	-70	180	H	0.88	H
45	K	-64.39	-40	-90	-34.66	0	-70	180	H	0.89	H

Table 7.1 | Data output of DANGLE and TALOS secondary structure calculation algorithms. φ, ψ, and Ω, three dihedral angles of an amino acid. SS, secondary structure. H, helix; E, extended; C, coil. CONF., confidence.

viral DVNP		DANGLE								TALOS	
No	seq	ϕExp	ϕUpper	ϕLower	ψExp	ψUpper	ψLower	Ω	SS	CONF.	SS
46	S	-64	-40	-90	-42.14	-20	-70	180	H	0.94	H
47	A	-63.99	-40	-90	-35.29	-10	-70	180	H	0.97	H
48	C	-65.48	-40	-90	-27.61	10	-60	180	H	0.96	H
49	E	-63.26	-40	-90	-31.97	0	-70	180	H	0.94	H
50	R	-66.62	-40	-100	-29.51	10	-60	180	H	0.92	H
51	M	-70.05	-30	-120	-22.76	30	-70	180	H	0.82	H
52	K	-74.67	-40	-120	-22.93	30	-50	180	H	0.49	H
53	K	-67.19	-40	-120	144.12	179.99	90	180	C	0.13	C
54	E	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.76	C
55	G	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.9	C
56	K	-75.43	-40	-170	149.18	179.99	90	180	C	0.93	C
57	K	-152.56	-40	-180	163.9	-170	120	180	C	0.89	C
58	H	-65.67	-40	-120	146.57	179.99	110	180	C	0.66	C
59	C	-69.48	-40	-120	135.75	179.99	90	180	C	0.78	C
60	T	-103.06	-50	-150	122.3	179.99	80	180	C	0.77	C
61	K	-106.74	-60	-170	134.68	179.99	90	180	E	0.4	C
62	V	-107.82	-50	-160	123.39	170	90	180	E	0.47	E
63	F	-101.88	-60	-150	117.26	160	90	180	E	0.32	E
64	K	-99.1	-50	-160	116.31	170	90	180	E	0.43	C
65	A	-65.04	-40	-110	145.77	179.99	90	180	C	0.79	C
66	K	-65.12	-40	-110	136.72	179.99	100	180	C	0.88	C
67	K	-101.3	-50	-160	138.96	179.99	90	180	C	0.88	C
68	G	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.91	C
69	K	-78.62	-40	-180	142.34	179.99	90	180	C	0.84	C
70	F	-103.28	-40	-180	141.01	-170	80	180	E	0.57	C
71	A	-64.98	-40	-90	145.37	179.99	110	180	C	0.43	C
72	C	-67.5	-40	-150	136.51	170	90	180	C	0.71	C
73	Q	-69.38	-30	-120	146.02	-170	100	180	C	0.91	C
74	P	-65.33	-40	-90	154.74	170	120	180	C	0.91	C
75	K	-106.05	-60	-150	119.92	170	90	180	C	0.94	C
76	E	-90.78	-40	-180	144.74	-170	90	180	C	0.95	C
77	G	-152.24	-40	-180	173.29	-160	100	180	C	0.96	C
78	T	-107.64	-40	-180	147.42	-170	90	180	C	0.9	C
79	A	-104.84	-60	-140	4.93	40	-40	180	C	0.53	C
80	A	-112.84	-40	-180	147.39	-170	90	180	C	0.19	H
81	Y	-99.15	-60	-130	1.96	30	-40	180	C	0.2	H
82	K	-97.36	-50	-130	1.62	40	-40	180	C	0.39	C
83	K	-70.05	-40	-120	-16.26	30	-50	180	C	0.7	C
84	K	-75.12	-40	-130	-15.28	30	-60	180	C	0.76	C
85	M	-87.87	-50	-130	-6.1	30	-60	180	C	0.87	C
86	K	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.92	C
87	K	-65.01	-40	-110	145.08	179.99	110	180	C	0.87	C
88	M									0.84	C
89	A									0	C

Table 7.1 cont. | Data output of DANGLE and TALOS secondary structure calculation algorithms. φ, ψ, and Ω, three dihedral angles of an amino acid. SS, secondary structure. H, helix; E, extended; C, coil. Yellow box marks the only disagreement between the two algorithms. CONF., confidence.

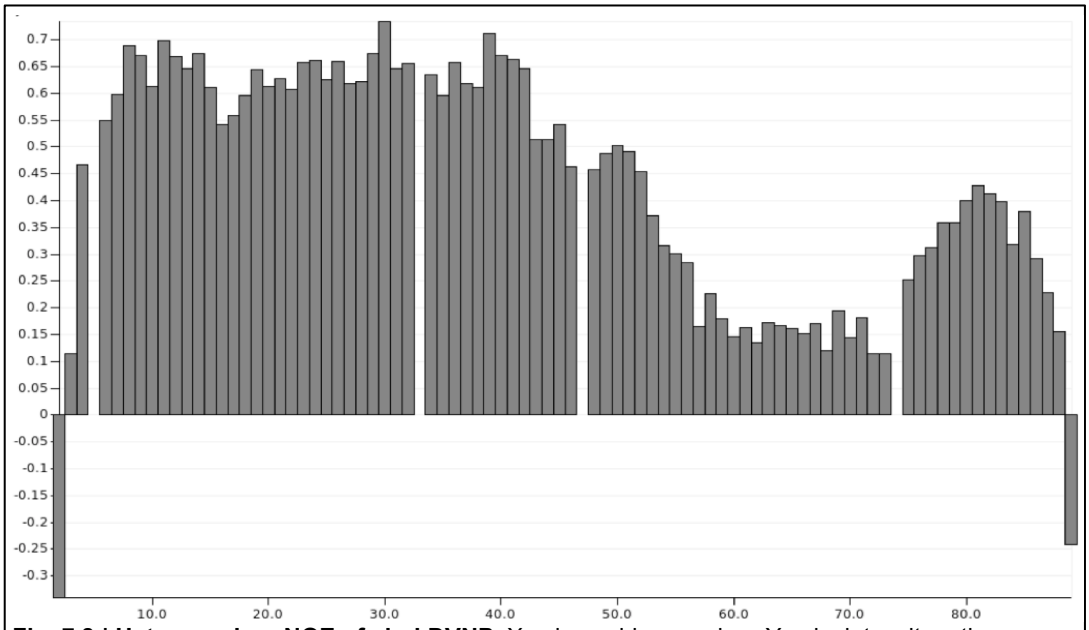


Fig. 7.8 | Heteronuclear NOE of viral DVNP. X axis, residue number. Y axis, intensity ratio.

7.2.4 Solving the 3D structure of Ostreococcal viral DVNP

For a complete understanding of the structure of DVNP and its possible mechanism of DNA interaction and compaction, a tertiary, 3D structural model was sought. Several further 3D spectra were then recorded. ^{13}C-HSQCs were recorded to identify carbon-hydrogen bonds, similar to a ^{15}N-HSQC. ^{13}C-TOCSY (TOtal Correlated SpectroscopY) and ^{15}N, ^{13}C-HccoNH through-bond spectra were recorded to correlate side-chain C-H bonds to the other hydrogen atoms on the same side-chains and also to the backbone carbon and nitrogen atoms. Finally, for the two through-space spectra, ^{15}N-NOESY (Nuclear Overhauser Effect SpectroscopY) was recorded to discover interactions associated with each Hα and ^{13}C-NOESY for interactions with hydrogen atoms on side-chains. The individual atoms and pair of interactions were identified and assigned empirically by correlating peaks in different spectra with shared axes, with the prior knowledge of known characteristics of each type of amino acids. I note that some amino acids in the C-terminal region, i.e. the abundant lysine residues, are unresolved and overlapping in many of the spectra (for example see Fig 7.9), and may result in imprecise peak placement. After all the assignments were completed, the locations of all peaks in the two NOESY spectra were identified to create two restraint lists.

In the NOESY experiments, either a nitrogen or carbon atom is pulsed, the pulse transfers to the immediate hydrogen, then transfers to another hydrogen atom that is in close proximity in 3D space, and finally reports one frequency of the final acceptor hydrogen atom. Since many hydrogen atoms can share the same frequency, each restraint on the lists, i.e. each peak in the NOESY spectra represents the interaction of one specific NH or CH bond to one or more real contacts in 3D space out of many possible candidates. The two lists, in combination with the sequence of the protein, the dihedral angles, and the secondary structure calculation results from TALOS, were used to model the structure with the iterative 3D structure-modelling program ARIA (Linge *et al.* 2003). ARIA performs iterative *in silico* folding and refolding of the protein and tries to fulfil all the through-space restraints. After each iteration, ARIA returned twenty most energetically favourable potential structures and a list of violations, which were NOESY restraints that none of the generated structures can fulfil and conflict with other restraints. These violations were then empirically examined, and signals determined as false or noise were removed from the restraint

lists. The Gremlin contact maps produced by coevolution analyses described in Chapter 3 were also used in conjunction to improve the restraint lists where high confidence contacts were predicted. The improved restraint lists were then resubmitted to ARIA for another iteration, where the calculation process was repeated and more weight was placed on potential through-space interactions with higher confidence. The procedure repeated until no more violations were produced, and high-confidence structure models were produced.

During the process of identification of all the peaks in the spectra as well as the violations, it was discovered that there were two different folding states of DVNP, of which the details are described in the next section. For some residues, the switching between the two folding states created exchange peaks, which behaved similarly to magnetisation transfer between different atoms. As these peaks are not real through-space interactions but only represent the switching between states, they interfered with the structure calculation adversely, hence were manually identified and removed from the NOESY restraint lists.

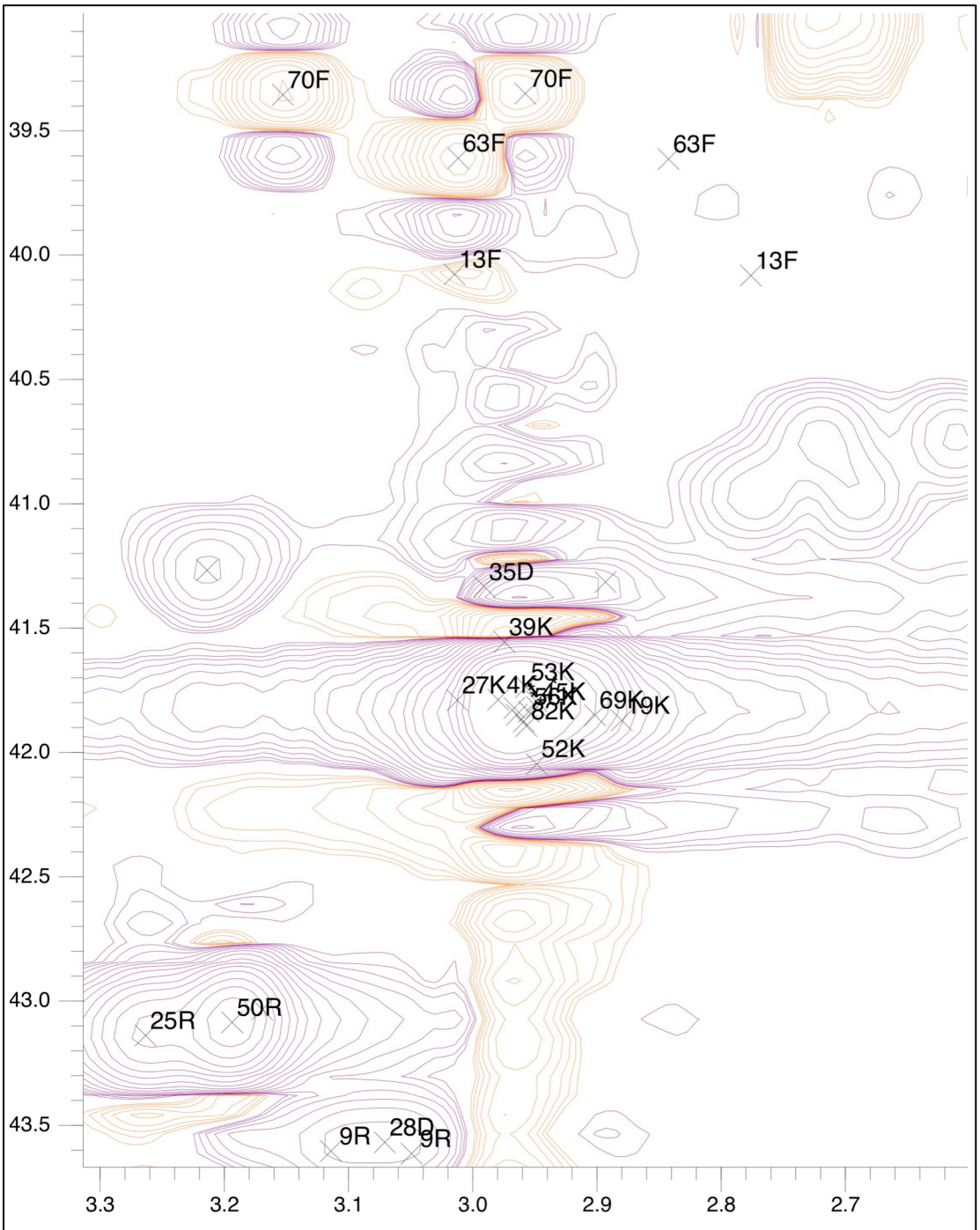


Fig. 7.9 | Example lysine peak overlapping in ^{13}C-HSQC. The peak in the centre represents the overlapping peaks of about 30 lysine Cε-Hε. Each X marks a peak representing one set of C-H bond. Purple, positive contour. Orange, negative contour. X axis, ^{1}H. Y axis, ^{13}C. Unit, ppm.

Here I present the first 3D NMR structure models on the protein DVNP (Fig. 7.10). Seven energetically most stable structures were aligned and superimposed. According to the tightly clustered ensemble of the structures, the first half of the protein, from the N-terminus to the end of a large helix (1-55) is folded tightly around a hydrophobic core, generally planar in shape. The N-terminus immediately falls into a small local structure of a tightly packed turn and a small helix (blue to cyan). The structure is followed by an antiparallel β-sheet in direct contact with the small helix (blue-green to bright green). A curved helix 11 residues long (bright green to mustard) then lies immediately adjacent to the sheet. The other half of the protein, after extending from the end of the helix along the antiparallel β-sheet, becomes much more disordered, and seemingly wraps and shields one side of the core domain. The C-terminal part of the protein by comparison to the N-terminal half, is much less converging and much more diverse, as the N-terminus structures of the seven models align almost perfectly (Fig. 7.11).

To investigate potential DNA-binding regions of DVNP in the context of this structural information, charge distribution was overlaid with the model structure (Fig. 7.12). The charge distribution on the protein is non-uniform. The positive charges are scattered and few in the N-terminal structure but predominantly occupy the entirety C-terminal tail. The negative charges are few and sit mostly at the edges of the planar structure, including the end of the antiparallel sheet and the top and lower side of the very N-terminal structure. In addition, surface rendered charge distribution profiles were generated of the whole protein and the N-terminal half (1-55) (Figs. 7.13 & 7.14). The results demonstrated that the model of the full protein is essentially an irregular globule with high positive surface charge across the whole surface. The surface without the C-terminal tail, however, reveals positive charges at the edges of the structure and a groove in the N-terminal structure where the beginning of the C-terminal tail (56-64) fits (Fig. 7.14). Some hydrophobic patches are also observed to be underneath the C-terminal tail.

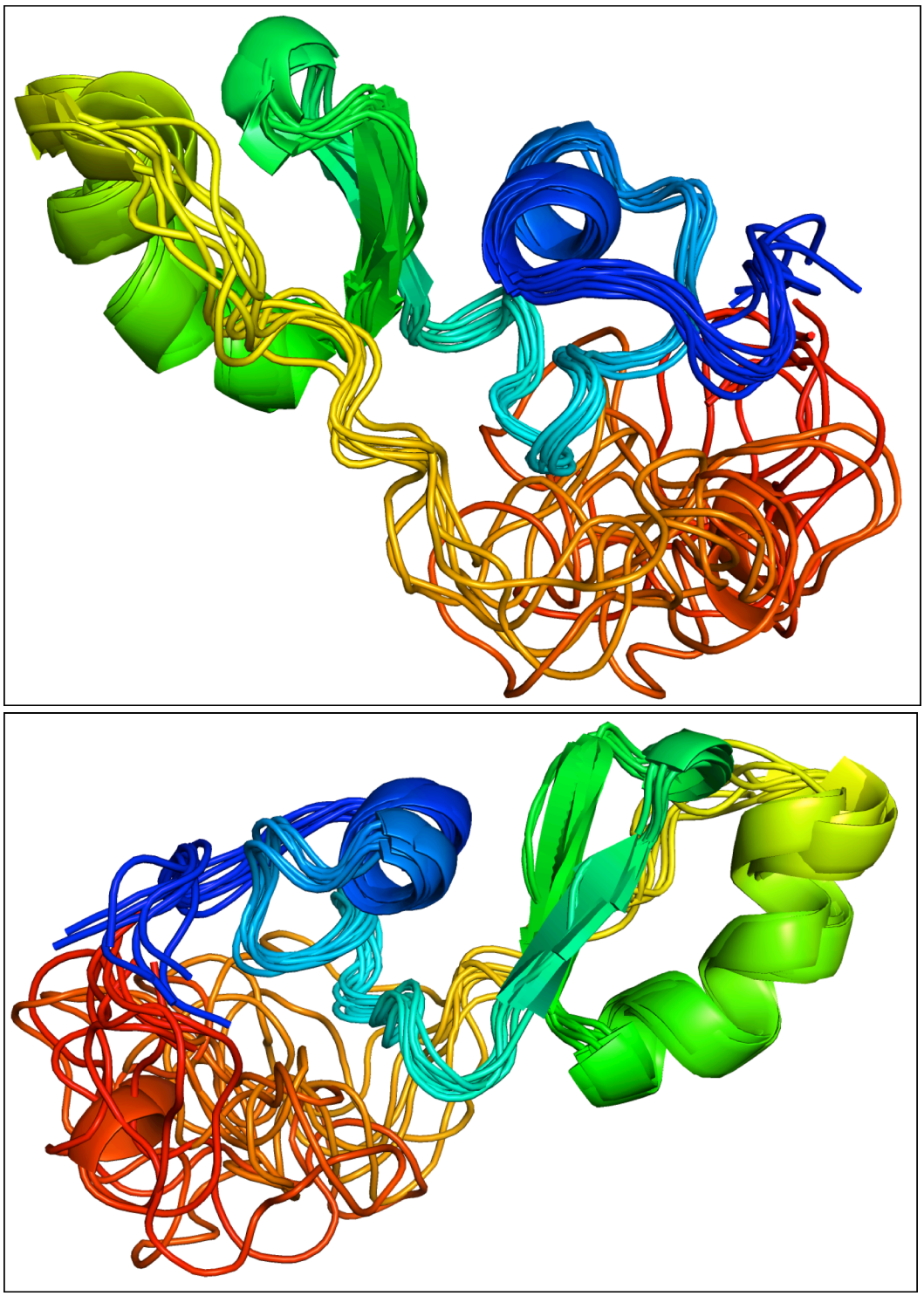


Fig. 7.10 | Ensemble of seven lowest free-energy models of viral DVNP that conform to restraints by ARIA. Blue end, N-terminus; red end, C-terminus.

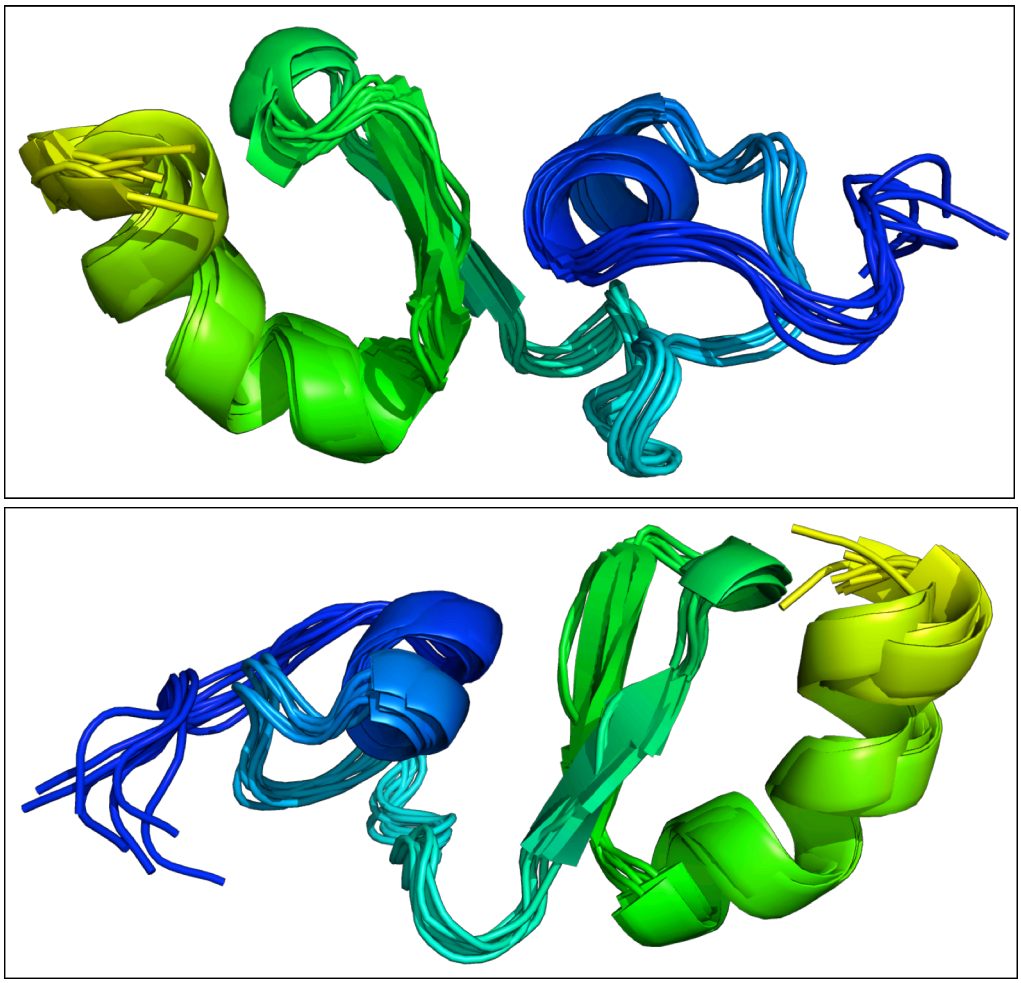


Fig. 7.11 | The N-terminal structured region (1-55) of the ensemble of seven lowest energy models of viral DVNP. Blue, N-terminus; yellow, C-terminal end of the structures.

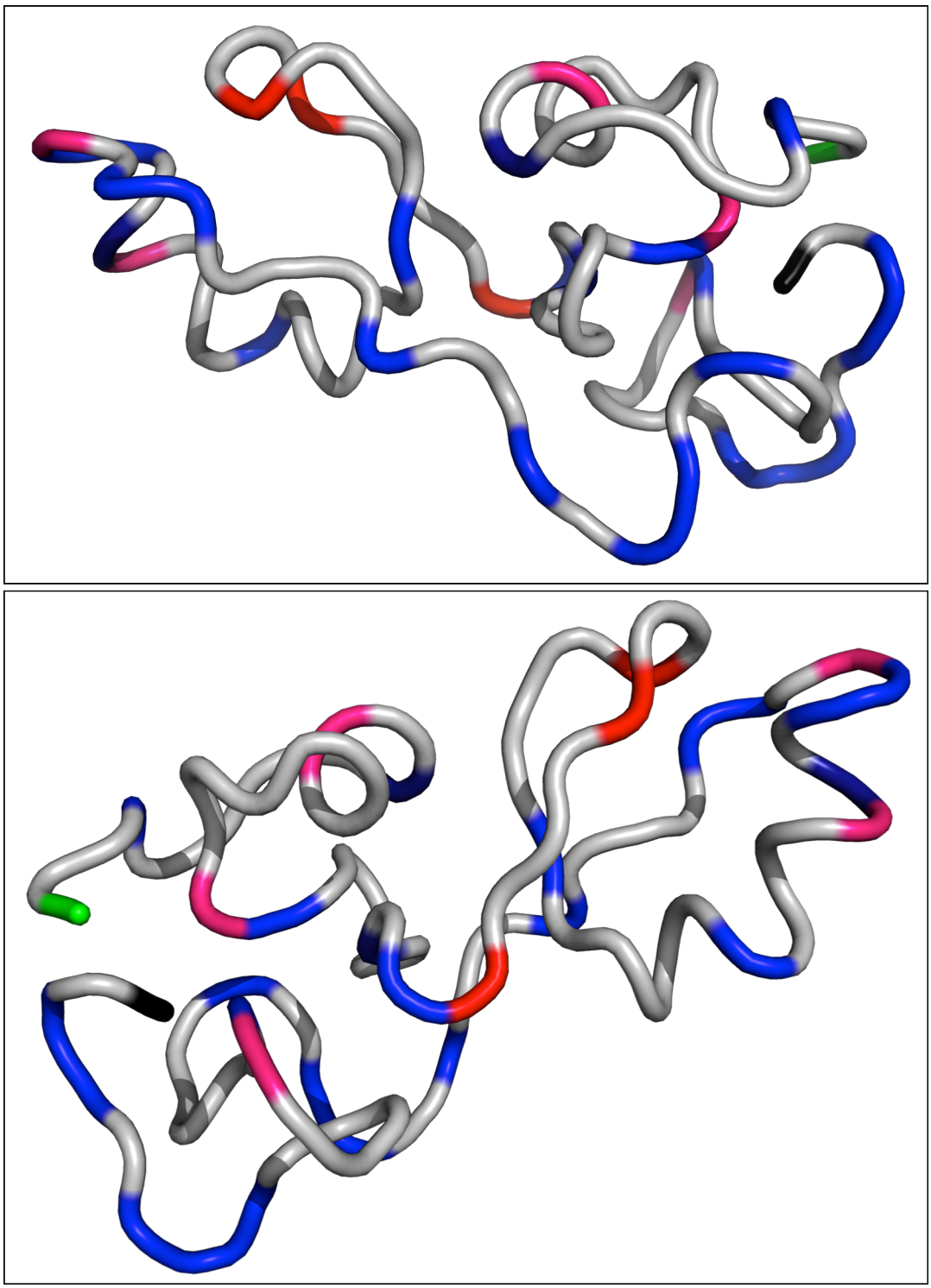


Fig. 7.12 | Distribution of charged amino acids within viral DVNP. For clarity only one model is shown. Blue and purple, lysines and arginines. Red and pink, aspartates and glutamates. Green, N-terminal end. Black, C-terminal end. The model is placed in the same orientation as fig. 7.9 and 7.10.

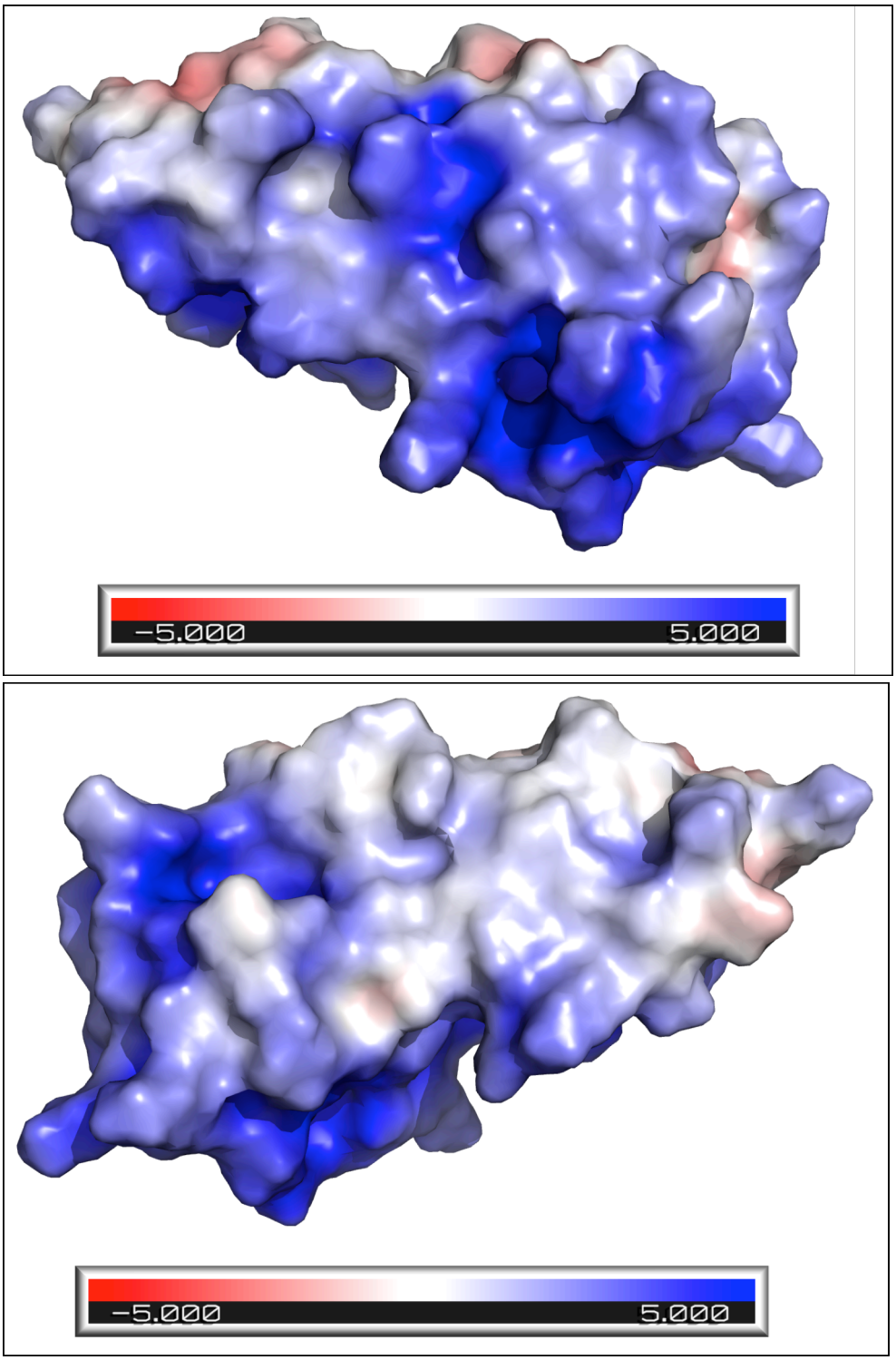


Fig. 7.13 | Surface charge render of viral DVNP. The same model as Fig. 7.12 was used. Blue and red, positive and negative charges. The model is placed in the same orientation as previous figures.

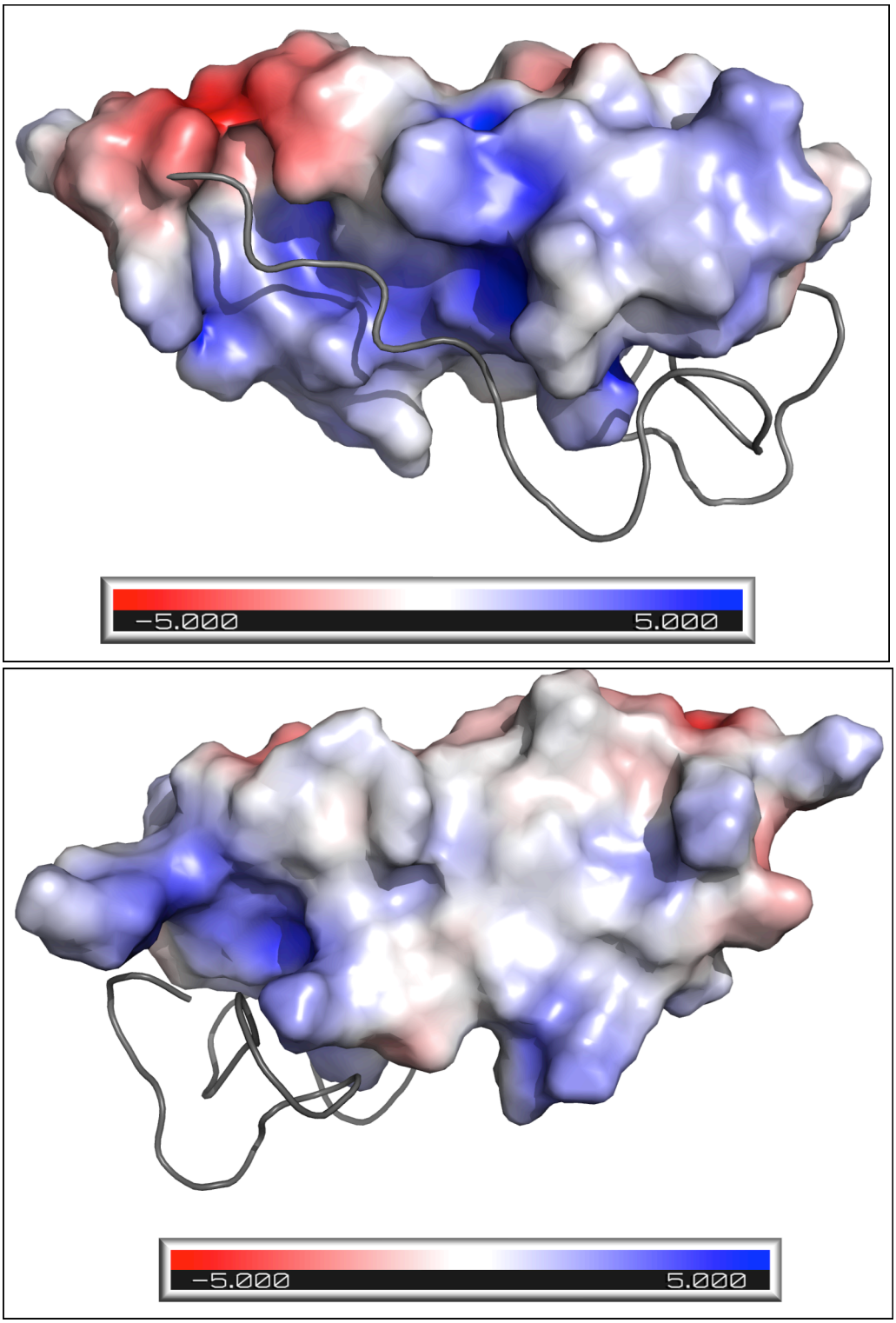


Fig. 7.14 | Surface charge rendering of the N-terminal half of viral DVNP. The same model as Fig. 7.12 and 7.13 was used. Blue and red, positive and negative charges. The model is placed in the same direction as previous figures.

7.2.5 An alternative folded form of Ostreococcal viral DVNP

During the sequential analyses of all the spectra, it was noted that there was another minor population of peptides besides the *Ostreococcus* virus DVNP in question, possibly even interacting with each other, evident by the additional number of peaks in both the ^{15}N and ^{13}C HSQC and also additional unexplained side-chain peaks in the ^{13}C-TOCSY spectra. Initially this was suspected to be a contaminant protein, but subsequently, it became apparent that there was, in fact, another folding form of the viral DVNP, and the peaks for several residues in this form captured by NMR (Fig. 7.15). This alternative folding form was recognised because of the discovery of additional peaks in the TOCSY through-bond experiment which identifies C-H pairs on the same side-chain. Since this was a through-bond experiment, and the number of C-H bonds on the side-chain of an amino acid was established, additional peaks cannot be 3D spatial interaction peaks and had to be because of conformational exchange. The alternative folding form was termed B form, and its peaks B peaks.

To remove the influence of B peaks on the structural calculation by ARIA, efforts were made to identify as many B peaks as possible in all the spectra. B form had much less signal strength overall, and the distribution was heterogeneous along the length of the protein. Most of the amino acids in an alternative form cannot be detected in any spectra. However, there are a handful of amino acid residues that display strong and stable signals that were eventually detected and identified. An example can be found in Fig. 7.16, a through-bond TOCSY experiment. Under normal circumstances for an alanine, the magnetisation received by Hα is transferred through the C-C bond and received by Hβ and vice versa, and two peaks are observed on the same horizontal line when either Hα or Hβ is in focus. Here for Ala17, however, an additional peak was found in both scenarios. Additionally, when the additional peak is in focus, energy transfers to both of the A peaks were also observed. Additional examples include several alanines and the only two isoleucines in the protein, and all show evident signals on the ^{13}C-HSQC (Fig. 7.17). For Hβ, Hγ, and Hδ of the isoleucines, two sets of C-H pairs can be observed for each with distinct chemical shifts (Fig. 7.17 cont.), all of which were traced from TOCSY spectra similarly to Ala17 described above.

Fig. 7.18 displays the assignment chart of the Ostreococcal viral DVNP protein in both A and B forms. The A form shows almost complete assignment throughout the whole protein, with the exception of prolines, which naturally do not have signals in many of the experiments due to the lack of Hα, and several clusters of under-resolved lysines from Hγ to Hε (Fig. 7.9). The assignment condition for the B form, on the other hand, is far from complete, since most of the residues do not produce clear and reliable signals in the B form. The incompleteness of assignment of the B form prevented further structural analyses on this folding form.

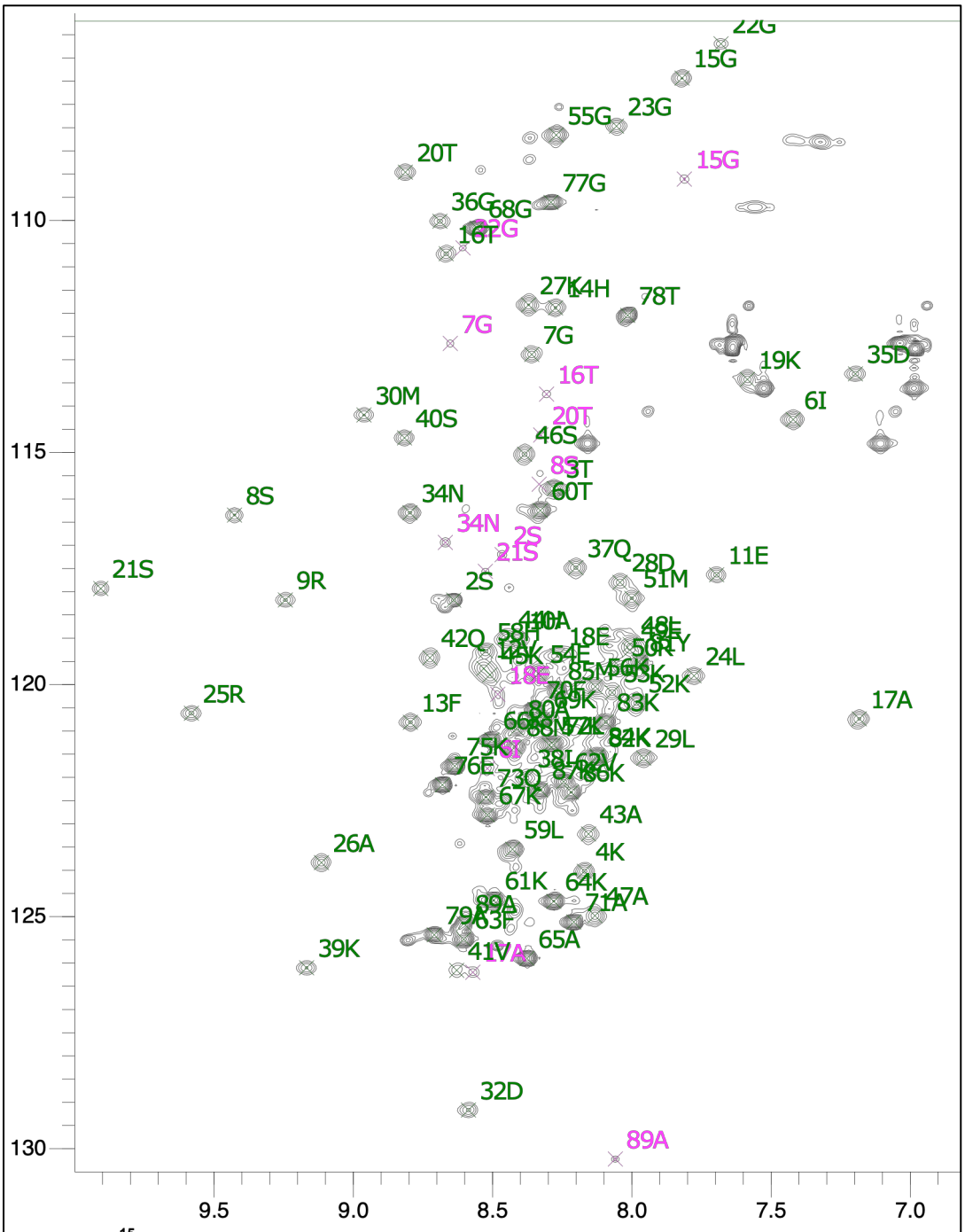


Fig. 7.15 | ^{15}N-HSQC peak distribution of the two folding forms of viral DVNP. Green, the structured 'A' form. Magenta, the alternative 'B' form. X axis, ^{1}H. Y axis, ^{15}N. Unit, ppm.

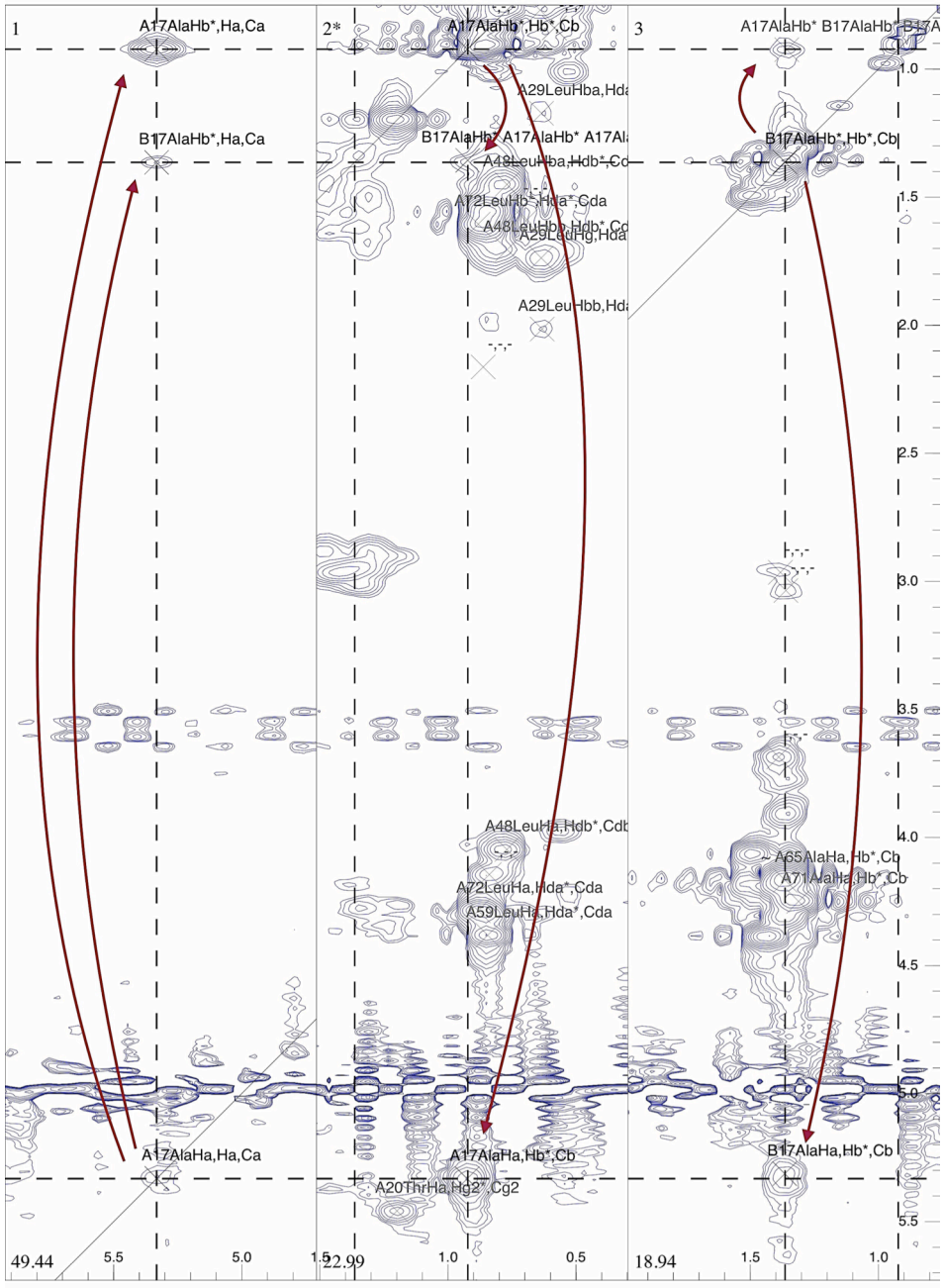


Fig. 7.16 | Example of inter-folding states dynamics of viral DVNP in ^{13}C-TOCSY through-bond side-chain correlation experiment. Panels 1, 2, and 3 represents the interactions from 'A' state Hα to Hβ in both 'A' and 'B' state, 'A' state Hβ to 'B' state Hβ and 'A' state Hα, and 'B' state Hβ to 'A' state Hα and Hβ, respectively. Dimensions x, y, z: ^{1}H, ^{1}H, ^{13}C. Unit, ppm. Arrows represent the direction of magnetisation transfer.

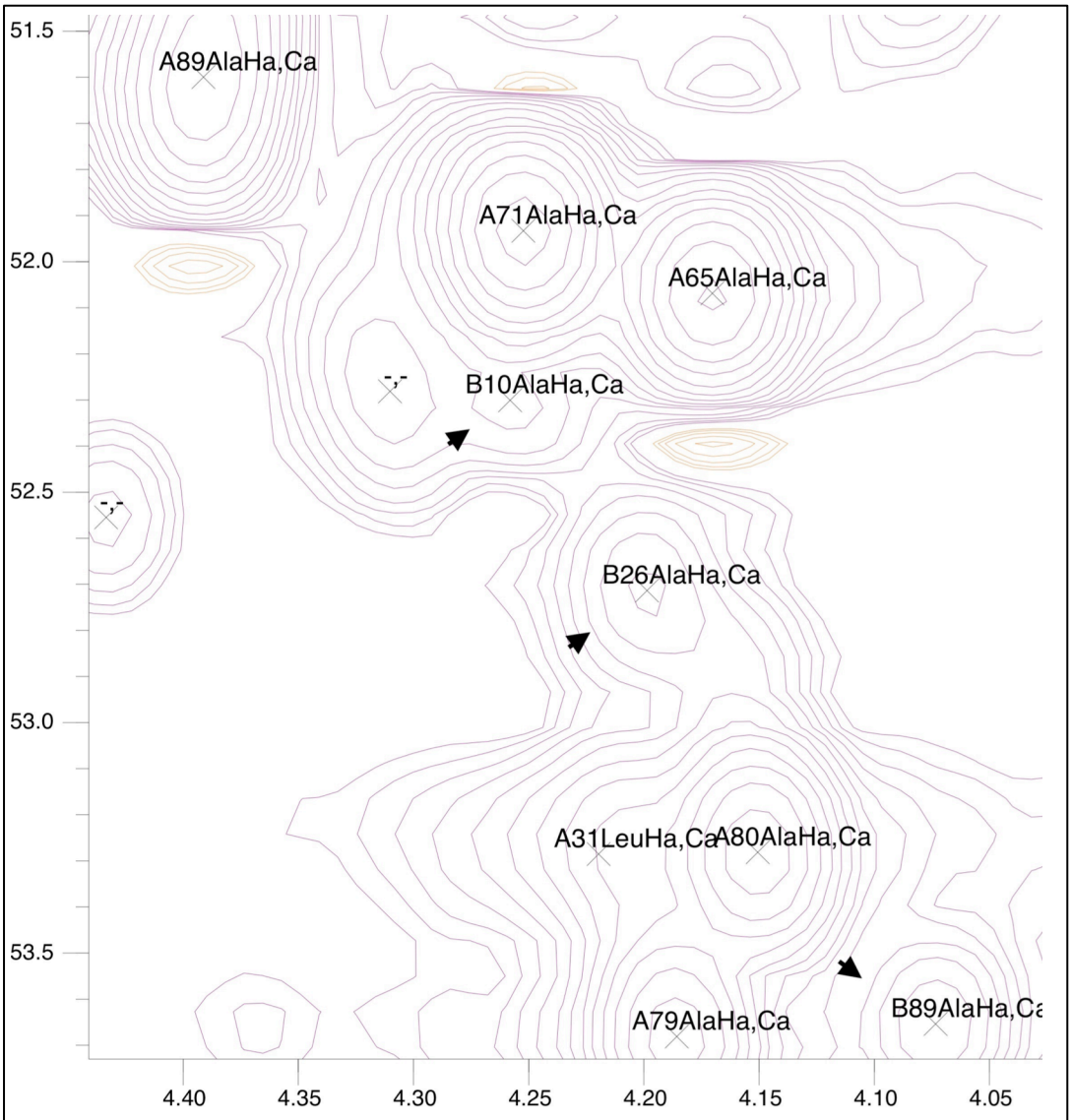


Fig. 7.17 | Examples of the existence of 'B' form peaks of viral DVNP in the ^{13}C-HSQC. In this region where alanine Cα-Hα pairs converge, several peaks from the B form of viral DVNP can be observed, including Ala 10, Ala 26, and Ala 89. X axis, ^{1}H. Y axis, ^{13}C. Unit, ppm. Purple, positive contour; orange, negative. Arrows, B peaks.

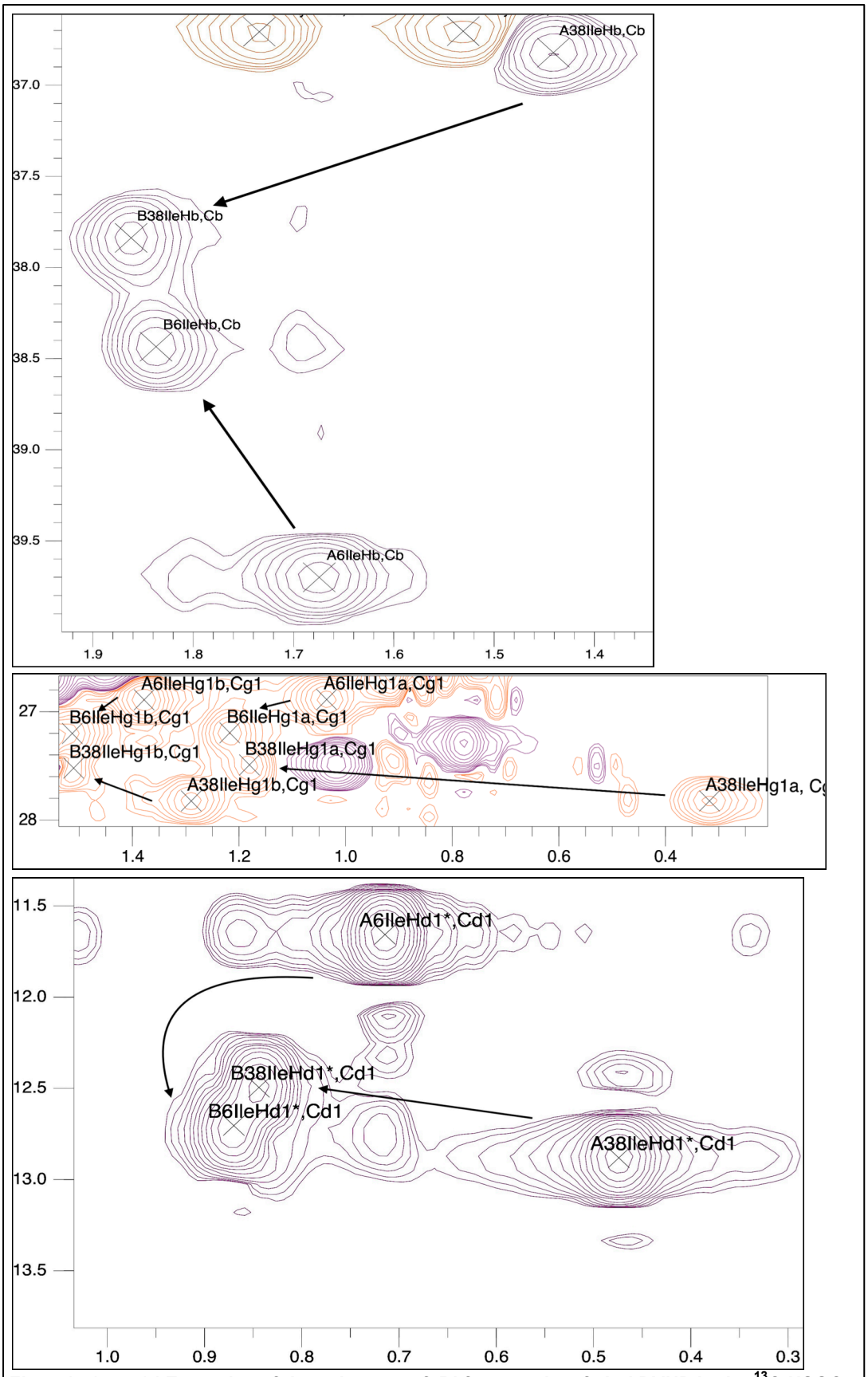


Fig. 7.17 (cont.) | Examples of the existence of 'B' form peaks of viral DVNP in the ^{13}C-HSQC. The Cβ-Hβ, Cγ-Hγ and Cδ-Hδ pairs of Ile 6 and Ile 38 all displayed additional distinct characteristics different from the A form. X axis, ^{1}H. Y axis, ^{13}C. Unit, ppm. Purple, positive; orange, negative. Arrows, movement from A to B peak.

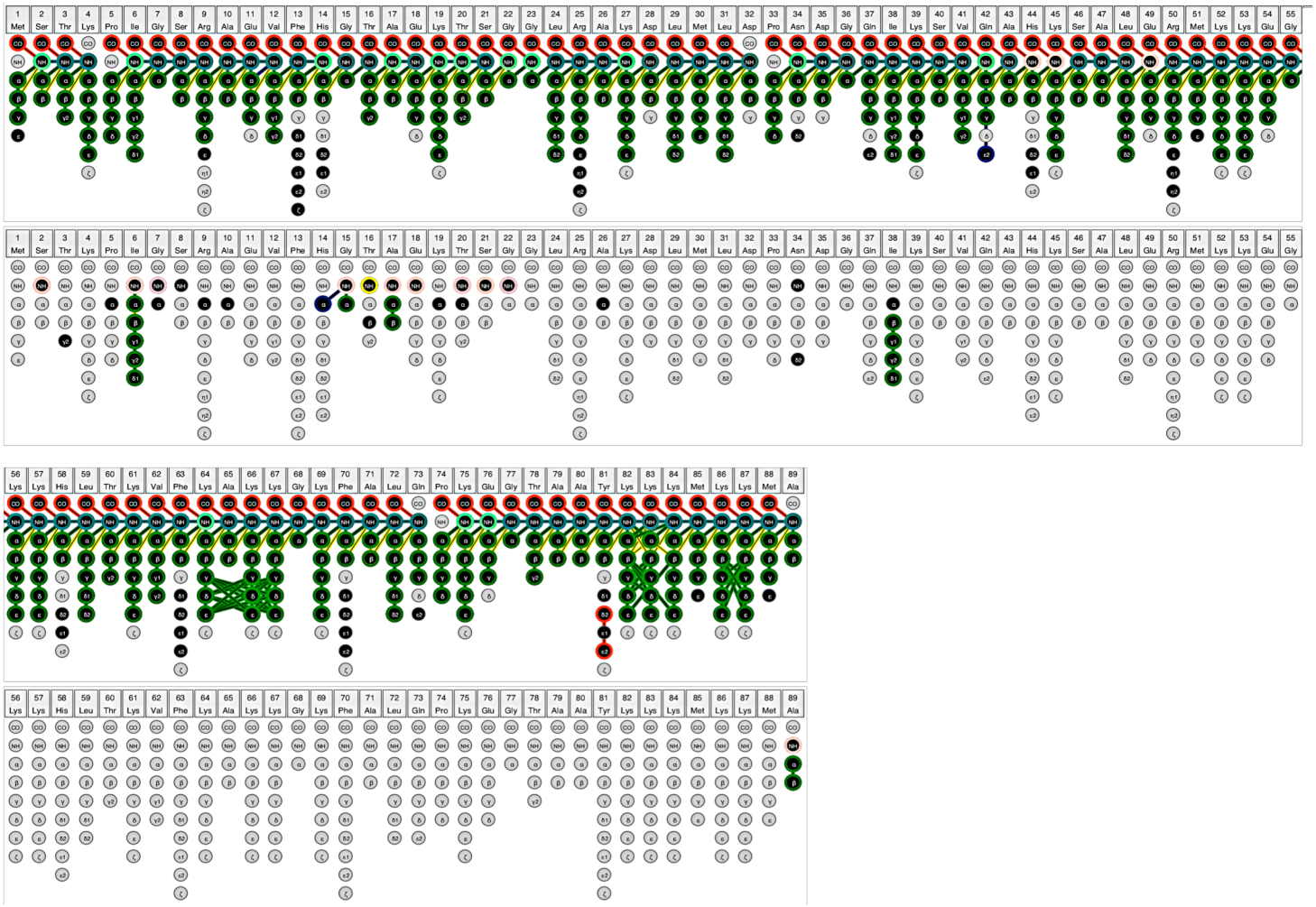


Fig. 7.18 | Assignment graph of both the A-form and B-form of Ostreococcal viral DVNP. Each filled circle represents an annotated oxygen, nitrogen, or carbon atom. Coloured lines and circles represent annotated hydrogen atoms and known associations in different spectra. The three green clusters in the C-terminal half of A form is the result of lysine overlapping.

7.2.6 Capturing DVNP interaction with perchlorate by NMR

To try to ameliorate the alternative folded form interfering with the protein structure calculation described in section 7.2.5, a number of salts were tested in a bid to push the protein into one conformation over the other, most notably phosphate and perchlorate (see also section 5.2.5). I reasoned that since in the *in vivo* condition, the protein DVNP very likely interacts with the phosphate backbone of DNA, using phosphate ions could be a proxy of DNA interactions *in vivo*. Phosphate concentrations up to 500 mM were tested, but no significant changes were observed in spectra, suggesting no overall conformational changes of DVNP (not shown). An alternative salt of 200 mM sodium perchlorate was also tested. Macroscopically, precipitation started to form: fibrils or crystals with the shape of eyelashes started to slowly collect at the bottom of the NMR tube over the time scale of a 2D NMR spectrum acquisition, approximately two hours (Fig. 7.19). Instead of pushing the equilibrium of the folding forms towards one over the other per earlier expectation, the perchlorate salt seemed to have caused significant shift across much of the ^{15}N-HSQC spectrum (Fig. 7.20), suggesting a change of local environment of many N-H bonds. To quantify the effect of perchlorate addition, the shifts of the peaks were plotted by residue numbers (Fig. 7.21 & 7.22). The result demonstrates that the C-terminal half of the protein, albeit still highly disordered judging by ^{15}N-HSQC, are the most heavily influenced by the addition of perchlorate.

Fig. 7.19 | Perchlorate-induced viral DVNP precipitation in a NMR tube.

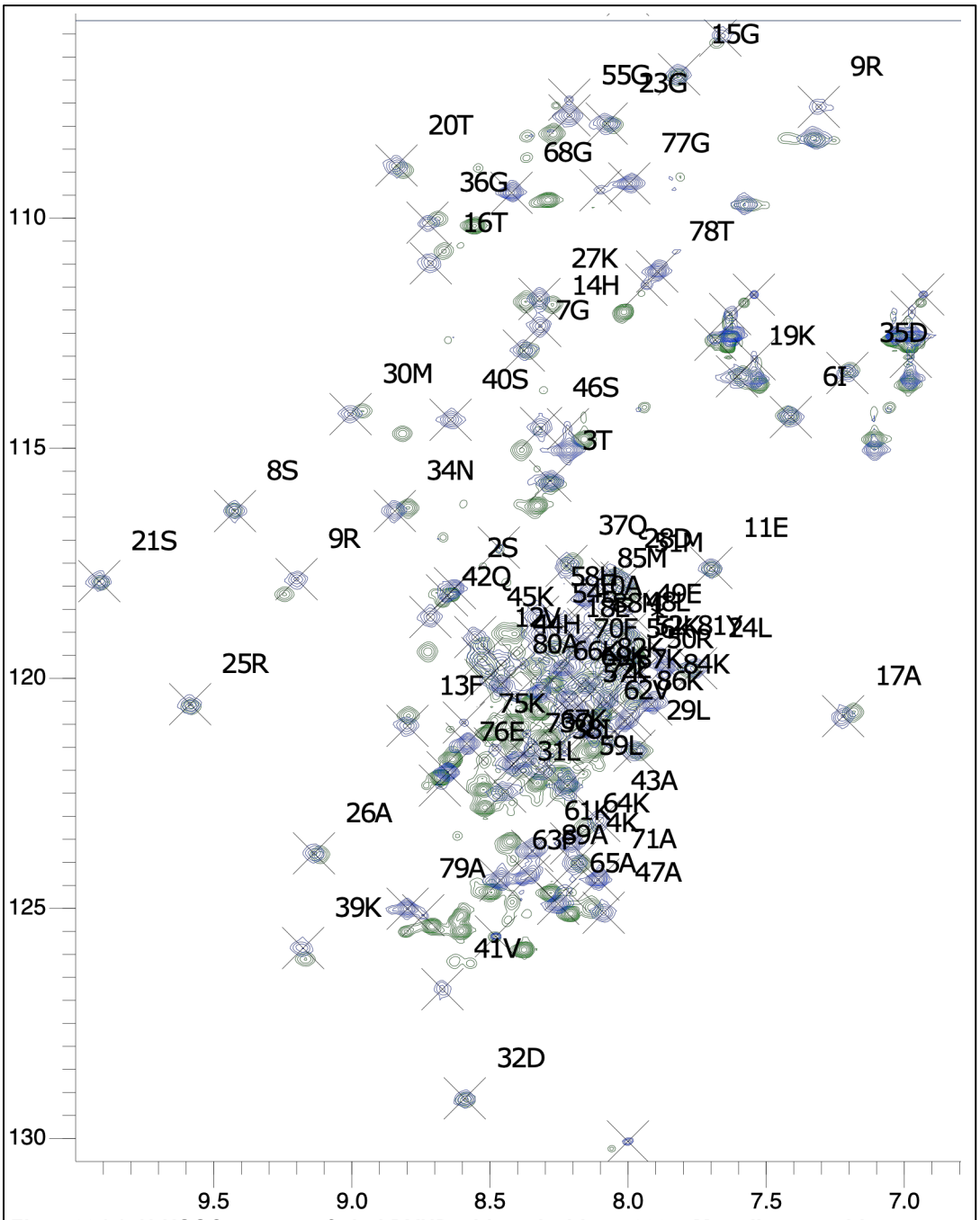


Fig. 7.20 | 15N-HSQC spectra of viral DVNP with and without 200 mM sodium perchlorate. Green and blue, before and after addition of perchlorate. A-form backbone peaks are assigned to the reference unsalted spectrum. Only the backbone peaks are shown.

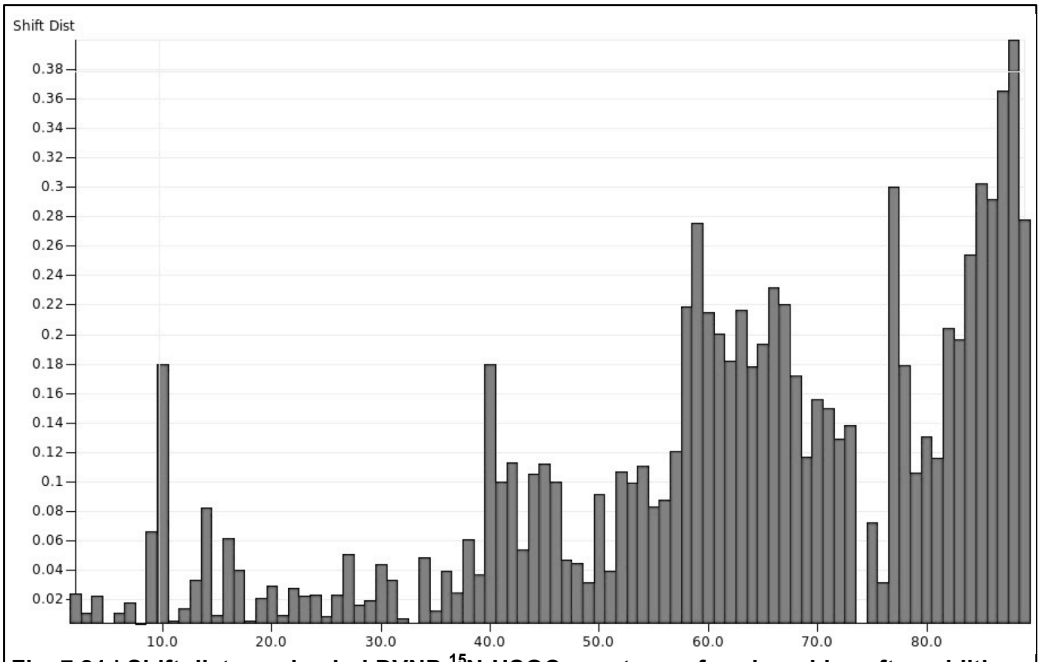


Fig. 7.21 | Shift distance in viral DVNP ^{15}N-HSQC spectrum of each residue after addition of perchlorate. X axis, residue number. Y axis, shift distance. Unit, ppm.

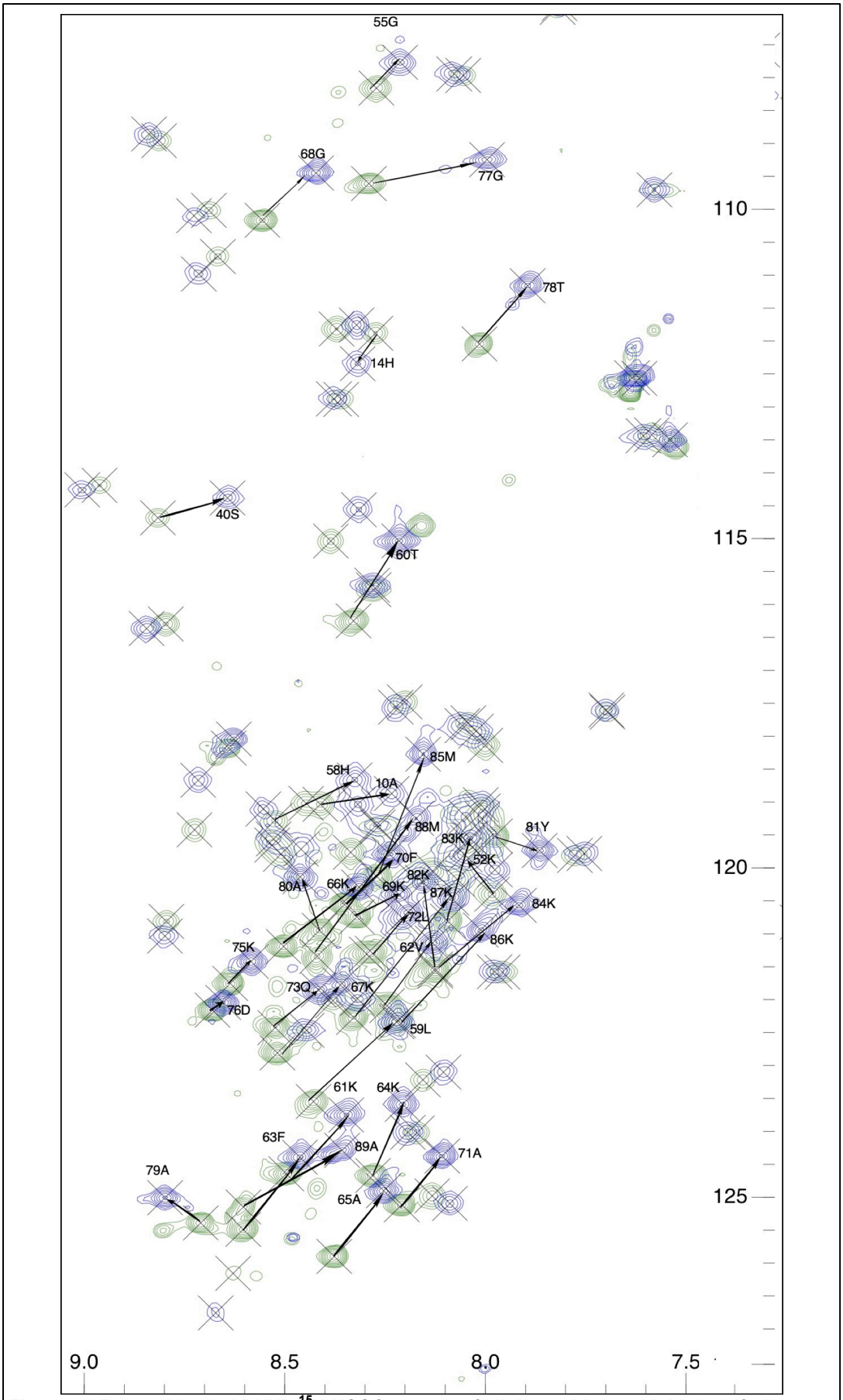


Fig. 7.22 | Zoomed in viral DVNP ^{15}N-HSQC spectra of all the residues with high shift distance after addition of perchlorate. Green and blue, before and after addition of perchlorate, respectively.

7.3 Discussion

The present study marks the first successful structural determination of the protein DVNP. Key to these studies was the optimisation of expression and purification, and several obstacles had to be overcome to achieve a suitable protein sample. At the beginning of the study, very heterogeneous mixtures with C-terminal tails of varying lengths were produced, apparently through translation slippage and/or premature translation termination. These problems were overcome with the strategy of a self-cleavable affinity domain placed at the C-terminus, the only scarless C-terminal affinity tag to my knowledge. The heterogeneity of the earlier protein preparations would have generated great noises by NMR spectroscopy and impeded the assignment, rendering the whole task impossible to complete. In the present PhD study, only the *Ostreococcus lucimarinus virus 5* DVNP was labelled with ^{13}C, due to the better chance for resolving a structure determined by the better dispersion of peaks in the initial ^{15}N-HSQC spectrum.

The structure models of Ostreococcal viral DVNP eventually obtained from the calculation show excellent resolution and agreement in the N-terminal half (Fig. 7.10 - 7.12). In the assembly of the 7 most energetically favourable structure models, the N-terminal halves almost completely overlap with no disagreement. The C-terminal halves of the models are more divergent by comparison, yet the common theme of the models is the lack of any structural feature. There are two possibilities why this is, the apparent first being that the C-terminal half is genuinely unstructured, and the calculation reflected that; the second possibility is that the signal overlapping, especially of the abundant and presumably less structured lysine residues that concentrate in the C-terminal half, prevented the software from correctly identifying the real contacts from noise. The heteronuclear NOE data (Fig. 7.6) largely support that the protein starts to lose its structure after residue 55, consistent with the first explanation of a disordered domain. However, the small stretch of modelled helix right at the end of C-terminal tail around Tyr81 that was identified by TALOS and supported by heteronuclear NOE is present in 2 of the 7 most energetically favourable models. Tyr81 is one of the most conserved amino acids in the whole sequence in DVNP based on sequence analysis (chapter 3), an aromatic that is both present in the dinoflagellate and viral clades. Similarly, a small peptide around Val62 that was

marked as extended by both algorithms is also not represented clearly on the structural models derived from ARIA. It is reasonable to think that the two residues, 81Y and 62V, may have essential functions and participates in structure formation, but at the moment the structural information on the residues is still confounded by the lack of resolved signal around them. More structural insights may require further manual deconvolution of the signals from the original spectra and additional acquisition of spectra.

The high level of structure of DVNP was unexpected. All relevant sequences of DVNP were analysed by the Metaprdos protein disorder predictor (online tool; Ishida & Kinoshita 2007), and all were predicted to have a high tendency to be disordered at least in the N-terminal half (Fig. 7.19). Furthermore, based on the size and hydropathy of the protein, one hypothesis was that the protein would compact DNA in a similar

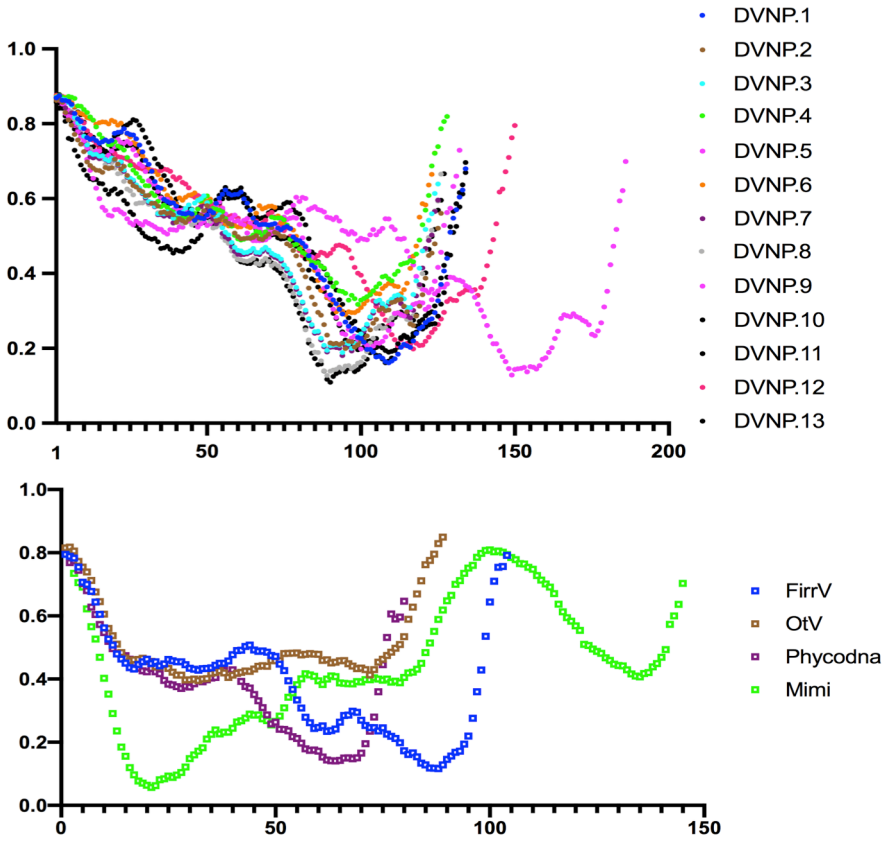


Fig. 7.23 | Metaprdos protein disorder prediction results for DVNP proteins. X axis, residue number. Y axis, probability of disorder. Upper, 13 DVNP proteins found in *Hematodinium* sp. Lower, four DVNP proteins of viral origin found in publicly available database. FirrV, Fieldmannia irregularis virus; OtV, Ostreococcus tauri Virus; Phycodna, Organic Lake Phycodnavirus; Mimi, Yellowstone Lake Mimivirus. Consistent values lower than 0.5 imply stable structure.

fashion of PSI-DNA, neutralising the backbone phosphate charges and allowing the DNA to fold upon itself (Leforestier & Livolant 1993; Livolant & Maestre 1988; Raspaud *et al.* 2005; Sikorav *et al.* 1994). Yet, the NMR results provided strong evidence that there is a globular core structural element in the stable structure of viral DVNP; the high spread of peaks in the ^{15}N-HSQC, heteronuclear NOE, and the 3D model calculated from NOESY restraints all independently support that the N-terminal core region has a stable globular structure, contrary to the Metaprdos prediction.

The alternative folding form of Ostreococcal viral DVNP provides interesting challenges and implications as well. The distribution of the few peaks identified to belong to B form mostly locate in the N-terminal half of the protein, hinting that the alternative form may involve in the structural aspects of the N-terminal half. This is also consistent with the placement of B form peaks on the ^{15}N-HSQC (Fig. 7.15). All the B form peaks seem to be concentrated towards the centre on the ^{1}H axis, suggesting the tendency of disorder. Due to the placement of the residues, it is less likely that the A and B forms share other amino acids in their original state. It is more likely that the whole of B form represents a less structured form of Ostreococcal viral DVNP.

The mechanism for the DVNP-DNA compaction activity remains to be determined. The presence of a hydrophobic core implies that the PSI-DNA compaction might not explain DVNP's activity adequately and that a more complex mechanism might be in place. The PSI-DNA type of simple charge neutralisation of the phosphate backbones does not seem to integrate well with the complex globular core domain. Nonetheless, I note the planar nature of the folded domain, its surface hydrophobicity buried underneath the C-terminal tail, and the existence of the antiparallel beta-sheet before the major helix 41-55 (Fig. 7.14). These factors raise the possibility of the protein being able to dimerise or multimerise in a manner similar to amyloid formation, with the buried surface as the interface. The Thioflavin T staining results presented in Chapter 5 (Fig. 5.14) also support the existence of multimers or amyloid formation. These data imply that the multimerisation activity may be intertwined with the DNA compaction phenomenon. However, none of the spectra recorded actually captured intermolecular interactions; the protein appeared to be in only monomeric form in low

salt as well as in high concentrations of phosphate or perchlorate –albeit with significant changes in intramolecular dynamics evident by the ^{15}N-HSQC. Additionally, although dsDNA oligo and ^{15}N, ^{13}C-doubly labelled viral DVNP do form interactions, the intense precipitation inhibited the generation of any useful signal. Titration of shorter lengths of dsDNA oligos with DVNP protein may be required to inhibit protein multimerisation so the NMR spectroscopy may truly capture the dynamics of DNA-DVNP interaction unconfounded.

The discrepancy between the results obtained from the *Hematodinium* DVNP.6 and Ostreococcal viral DVNP was disconcerting at face value. The two proteins shared regions of high conservation in their sequences, including the core domain that was modelled as a globular structure in the viral protein. Furthermore, the biochemical and single-molecule spectroscopy data also reflected very similar protein properties. Our first comparative ^{15}N-HSQC spectra, however, suggest otherwise; a significant portion of the viral DVNP is highly structured, whereas the *Hematodinium* DVNP showed much less sign of a stable structure. Nevertheless, I note that in Chapter 3, the Gremlin co-evolution analysis revealed that the dinoflagellate DVNP very likely has a helix element between the highly conserved core sequences and the beginning of the variable C-terminal tail (Fig. 3.3). The same trend could not be found in the viral sequences presumably due to the paucity of the number of sequences and the resulting lack of statistical power. However, this helix is evidently present in every one of the NMR models at the correct position (Fig. 7.24). This result strongly supports that DVNP from the viruses and the dinoflagellates indeed share similar structures, abundant features, properties and possibly, functions.

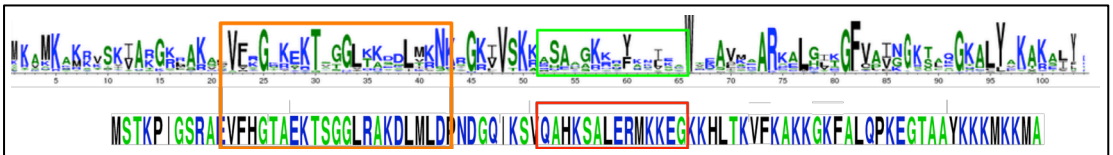


Fig. 7.24 | Sequence alignment of dinoflagellate DVNP logoplot and *Ostreococcus lucimarinus virus 5* DVNP. Upper, dinoflagellate logogram. Lower, viral DVNP sequence. Orange box, conserved N-terminal globular region. Green, predicted location of helix element by Gremlin. Red box, location of helix element by NMR spectroscopy.

As for the reason for the apparent lack of structure by ^{15}N-HSQC of the *Hematodinium* DVNP.6, the existence of an alternative folding state for viral DVNP strongly implies that DVNP.6 is fast switching between two states as well. Since the

spectra collections were performed at millisecond scale, if such equilibrium is not slow enough for both states to be captured separately by the spectrometer, the final spectrum would be the average of both folding states weighted by the time a protein molecule spends in each state (Roberts 1993). The apparent lack of structure of the *Hematodinium* protein may be the result of this phenomenon. Although there is still the obvious explanation that the dinoflagellate DVNP simply lacks a rigid structured region, this seems unlikely based on firstly the similarities of the data collected through biophysical, biochemical, and single-molecule spectroscopy means, and secondly the coincidence of the predicted major helix component in the dinoflagellate co-evolution prediction and the helix element that is presented in the NMR structure of the viral DVNP (Fig. 7.24).

Addition of both double-stranded DNA and perchlorate ions had profound effects on DVNP. Oligo DNA formed insoluble coacervates with viral DVNP initially, but when DNA was titrated past charge neutralisation point, the coacervates started to redissolve back into the solution. This result is consistent with the hypothesis that DVNP-DNA binding is based on charge neutralisation, and overcharging the phase-separated coacervates increases the intra-coacervate repellence and redissolves the molecule (Cummings & Obermeyer 2018; Turner *et al.* 2018). The redissolved DVNP-DNA complex, however, lost many signals in the most structured regions according to the structural models and the existing peaks in the C-terminal tail had become much broader. Peak broadening is most likely caused by slow exchanges between multiple species, causing the weighted average of the two or more states to be incompletely separated peaks, generating a broadened appearance (Roberts 1993). In this case, the broadened peaks would most likely be the average of the residues in the monomer form and multiple DNA-interactive forms. For the missing peaks of the structural parts, it is most likely that the molecular tumbling rate was reduced so much that the peptide was not responsive to the magnetisation in the millisecond time scale of a ^{15}N-HSQC experiment. This phenomenon is common in aggregation-prone proteins or proteins that oligomerise or polymerise (Karamanos *et al.* 2015), consistent with the hypothesis raised in Chapters 5 and 6 that protein oligomerisation is involved in DVNP-induced DNA compaction.

For the perchlorate experiment, although the initial goal for addition was to mimic DNA binding, perchlorate addition resulted in protein aggregation and precipitation (see Fig. 7.19). Since DNA-DVNP precipitate did not generate signals in ^{15}N-NSQC (not shown) and neither did the structural part of soluble viral DVNP-DNA complex (Fig. 7.4), presumably, the perchlorate-induced DVNP protein oligomers also were not able to generate much signals in the ^{15}N-HSQC experiments (Karamanos et al. 2015). This would suggest that the spectra presented in Figs. 7.20 and 7.22 were mainly generated by DVNP monomers in association with perchlorate ions, free from the influences of oligomerisation. The two figures, as well as the shift distance plot Fig. 7.21, clearly demonstrate that the majority of the residues affected by perchlorate addition were within the C-terminal tail. However, the effect caused by the perchlorate addition did not seem to bring structure or order to this part of the protein, as none of the residues left the unstructured central region in the ^{15}N-HSQC spectrum but merely caused a 'musical chairs' of most of the peaks (Fig. 7.22). The peaks migration most likely reflects the remodelling of the immediate chemical environments of the C-terminus residues by interaction with the perchlorate ions. The peak shifting, however, also demonstrated that the N-terminal peaks are not affected much by the addition of perchlorate salt, suggesting that the integrity of the N-terminal structure was not influenced by the addition of perchlorate ions. Following the same line of thought, if we accept the presumption that perchlorate is an appropriate substitute for dsDNA in *in vitro* experiments, the same implication can be applied to the dsDNA titration experiment where the structural peaks of viral DVNP disappeared: that the disappearance of the structural peaks would not have involved structural changes in the N-terminal structure. This deduction again supports the hypotheses that there was indeed a multimerisation event and that the multimerisation was the main reason for the disappearances of the peaks.

7.4 Methods

^{15}N and/or ^{13}C-labelled protein expression and purification

E. coli strain BL21(DE3) hosting either plasmids pTXB1-vDVNP or DVNP.6 was grown in LB overnight. The next morning, the overnight culture was used to inoculate 100 times volume of M9 minimal media with $^{15}NH_4Cl$ and or ^{13}C-glucose supplemented. The culture was shaken at 180 RPM at 37° until $O.D._{600}$ of 0.5. 0.5 mM of IPTG was then used to induce protein expression, and the culture was shaken at 90 RPM for 5 hours. The cells were then concentrated and proteins purified and processed by methods described in Chapter 5.

M9 minimal medium (Rédei 2008):

Na_2HPO_4 40 mM, KH_2PO_4 22 mM, NaCl 12 mM, autoclaved.

NH_4Cl 9 mM, $MgSO_4$ 1 mM, $CaCl_2$ 100 μM, glucose 0.2% (w/v), filter sterilised and added after the saline was autoclaved.

NMR spectroscopy

Spectra were taken with a Bruker AvanceIII AV600 system with 5 mm QCI CryoProbe for samples with 100 mM potassium chloride and Bruker Avance DRX500 (Ultrashield) for samples in 500 mM phosphate and 200 mM perchlorate. Both instruments were operated by the facility manager Dr Katherine Stott. All the analyses were performed with the software suite CCPNMR (Vranken *et al.* 2005). Structure calculation program ARIA (Linge *et al.* 2003) was performed on the departmental server 'Hydra' hosted in the Department of Biochemistry, University of Cambridge (hydra.bioc.private.cam.ac.uk).

Protein sequence visualisation and in silico prediction

Sequence visualisation were performed with Geneious software v12 (Kearse *et al.* 2012). *In silico* secondary structure prediction was performed with Geneious with the EMBOSS garnier plug-in (Garnier *et al.* 1978).

Artwork

Artworks were processed with software suite Pixelmator. Protein structure models were visualised with the PyMOL molecular graphics system, Version 1.2r3pre, Schrödinger, LLC.

Chapter 8: Conclusion: a proposed model for a nucleosome-independent permanently condensed chromosome arrangement and its transcription

The dinokaryon is uniquely distinctive from the nuclei of all other eukaryotes, but a clear understanding of it has been elusive. The discrepancy between the vast differences of a dinokaryon and a canonical nucleus, and yet its evident lineage of being derived from such a state, has confounded scientists for decades. In this PhD study, I sought to solve some of the mysteries of the permanently condensed chromosomes by trying to answer two questions: 1) where does transcription occur in this peculiar set up of permanently condense chromosomes, and 2) what does DVNP do, a protein that seemed to have had very important roles in the shaping of the dinokaryon and exists in only the dinoflagellates and select marine large DNA viruses?

Chapter 2 focused on the transcription conundrum of the dinokaryon permanently condensed chromosomes. A living dinoflagellate cell, albeit with permanently condensed chromosomes, still has genes expressed, exons spliced together, and proteins made. Where are the active genes in relation to the chromosomes then? The results described in this study agree with the earlier literatures, that the RNAPII and nascent RNA signals do not penetrate into the chromosomes, but instead wrap around the chromosomes in a patchy manner. To explain how the exons embedded in the chromosomes are accessible still to the transcription machinery, I reasoned that the DNA must travel outside the chromosomes to the periphery where the RNAPII molecules are. I posited the 'sewing machine' model. In this model, the transcription machinery is tethered to the periphery of the chromosomes and from within the chromosomes draw out template DNA, which, after being transcribed returns to the condensed body of the chromosomes (Fig. 2.14). The hypothesis is analogous to the 'transcription factory' concept in canonical eukaryotic cells, where discrete locations in the nucleus are concentrated in RNAPII molecules and transcription factors in order to increase transcription efficiency. The factories are tethered, possibly to an intra-nuclear proteinaceous scaffold termed nuclear matrix (Alfonso-Parra & Maggert 2010; Papantonis & Cook 2011; Rieder *et al.* 2012), and template DNA does all the

relative travelling of the two. I noted the similarities between the nuclear matrix of a canonical eukaryote and a dinokaryon and proposed that the two processes are not only analogous but also homologous. In both processes, the RNAPII molecules are tethered to the intra-nuclear scaffold. Despite the condensed representation of the dinokaryotic chromosomes, the DNA does all the moving, just as its counterpart in a canonical nucleus. This hypothesis, however, requires a critical component that had not been described previously yet, which is the fluidity of DNA within the condensed chromosomes.

In canonical eukaryotes, accessibility of DNA has been the criterion discriminating inactive, condensed heterochromatin from transcriptionally active euchromatin. In the present thesis, characterisation of DVNP was performed, the nuclear protein of viral origin whose introduction into dinoflagellates coincides with many of the oddities in their nuclear biology. This study finds that a) DVNP localises to the chromosomes *in vivo* yet on exposed chromosomes does not perfectly colocalise with DNA stain; b) DVNP has the activity to compact DNA alone; and c) DVNP-compacted DNA is apparently in a fluid state. In addition, when DNA was immobilised between two beads in the C-trap experiments to study DVNP's behaviour on DNA, it was discovered that both *Hematodinium* DVNP.6 and Ostreococcal virus DVNP have the capacity to move along DNA at the same time as forming aggregates. In combination with the previous knowledge that DVNP is the most highly expressed gene in a dinoflagellate cell (Marinov & Lynch 2015), these findings provide hints to a new type of chromosome, where DVNP protein provides the structural scaffold, and the DNA rests on the DVNP scaffold yet remains fluid and accessible to RNAPII molecules that are outside of the chromosomes. The proposed new type of DNA arrangement is fundamentally different to the nucleosome-mediated DNA packaging, and is supported by both *in vitro* and *in vivo* results that DVNP-mediated DNA compaction does not form protective structural packaging unit as the nucleosome. It is very possible that this putative DNA packaging does not involve DVNP protein forming a central core of the chromosome with DNA on the outside, since none of the published electron microscopy images showed any hint of this kind of structure. Instead, this organisation could happen at the most basic level similar to the nucleosome and the 30 nm-fibre, and the chromosomes would use this organisation to build up to more complicated high-order structures. However, I note the proposed

chromosome structure by Levi-Setti et al. 2008 (Fig. 1.5), which involves a protein-rich central core fibre surrounded by DNA-rich chromonema. It is possible, though unlikely, that the TEM images failed to differentiate DNA from protein in the sections of dinokaryons. TEM with immunogold labelling against DVNP and deeper sections over dinokaryons across multiple species of dinoflagellates would provide more data to test the hypothesis raised here.

The molecular mechanism of DVNP-induced DNA packaging is also suggested by this study where the potential of DVNP forming oligomers in association with double-stranded DNA is shown. This evidence includes the amyloid sensitive ThT-staining of *Hematodinium* chromosomes, coacervate formation by the high concentration of both DVNP and DNA, and the observation of protein-rich foci by single-molecule spectroscopy. Additionally, the use of perchlorate, an anion commonly used to characterise DNA-binding proteins for its molecular similarity to a phosphate ion, induces the formation of a novel species with molecular weight approximately twice that of a DVNP monomer. These results are strongly suggestive of the hypothesis that DVNP-DVNP interactions might be the structural basis of the DNA arrangement described above. In other eukaryotic or bacterial DNA compacting proteins, multimerisation seems to be a common theme. HCc3 in the later branching dinoflagellates and the bacterial HU and H-NS proteins were all shown to have dimerisation or multimerisation activities (Chan & Wong 2007; White et al. 1999; Arold et al. 2010). In the case of H-NS, Arold et al. (2010) proposed that the protein forms a superhelical scaffold for DNA condensation. In addition, Dps induced-DNA phase separation has shown to not inhibit the process of transcription elongation, whereas its effect on transcription initiation was not tested (Janissen *et al.* 2018). Together these previous studies support the hypothesis proposed in the present study, that it is possible that by dimerisation or multimerisation, the nuclear protein DVNP forms a protein scaffold for DNA to travel on.

The proposed DNA packaging however would generate obstacles for the canonical transcription system to initiate. Conceivably, a protein-mediated condensed state of DNA would be less penetrable or accessible for transcription factors or other components required for the initiation of transcription, although the fluidity of DNA under this hypothesis resolves the difficulty for transcription elongation. It has been

suspected that the dinoflagellates may take on a polycistronic type of transcription, partly due to the fact that most dinoflagellate genes lack identifiable gene elements and promoters and no TATA-box binding protein found in the transcriptome (Beauchemin et al. 2012; Roy & Morse 2014). Dinoflagellates also have a vastly diminished suite of TFII members, with no TFIIA, TFIIB, TFIIE, and TFIIF present (Roy & Morse 2014). In addition, the existence of spliced leader RNA trans-splicing in many species are coupled to the polycistronic transcription, including the nematodes, the cnidarians, and the blood parasite the trypanosomes (Tautz 2009). The existence of SL in the dinoflagellates is supportive of this possibility, and while Beauchemin *et al.* (2012) argue that the disparity of expression levels in closely positioned genes in dinoflagellates suggests otherwise, the trypanosomes are known to have virtually uncoupled expression levels of genes in the same polycistron that is solely controlled at the post-transcriptional level (Schwede et al. 2012). I caution that even though there are noticeable differences in expression patterns in the RNA of a dinoflagellate in different conditions, e.g. circadian clock (Walz et al. 1983), without the ability to knock out certain elements it is impossible to dissect how much influences the transcription processes had, and it is entirely possible that all the observable differences at the steady-state RNA level were regulated at the post-transcriptional level. It is still possible, even more so considering the findings of this thesis, that the dinoflagellates are at least partly polycistronic. Speculatively, assuming a continuous transcription *in vivo* where all genes were transcribed equally yet regulated either post-transcription or controlled by simply adjusting the copy number of the genes in the genome, initiation of transcription is then less concerning for the compacted system.

It has long been understood that dinokaryons are birefringent in nature, suggesting the cholesteric liquid crystallinity of DNA within the permanently condensed chromosomes (Rill et al. 1989). As opposed to a true solid crystal, liquid crystallinity describes a liquid that is ordered. This phenomenon is also consistent with the proposed hypothesis of DNA fluidity. There is no simple solution to test this hypothesis, although Hi-C chromosomes conformation capture may be able to provide meaningful insight.

The present study also seeks to understand the molecular and structural basis of DVNP-DNA binding and compaction. The data generated in this thesis demonstrate that DVNP is highly structured in the N-terminal half, and that this structured component might be involved in the potential oligomerisation in the DVNP-induced DNA compaction event. The C-terminal tail likely has no fixed structure, and is the most probable candidate for the actual DNA binding through charge neutralisation of the phosphate backbone of DNA by the abundant positively charged residues in this part of the protein. The C-terminal tail of histone H1 also employs this type of protein-DNA interaction (Turner *et al.* 2018), and it compacts DNA through PSI-DNA type of DNA compaction, which shares similarities by circular dichroism with a dinokaryon (Livolant & Maestre 1988). This type of protein-DNA interaction has been shown to exist and supports the DNA mobility required for the sewing machine model on a structural basis (Clore 2011; De March *et al.* 2017). In both studies, the proteins had been observed to be sliding bidirectionally on the double-stranded DNA. In the case of De March et al. 2017, PCNA utilises multiple lysine and arginine residues to form a series of interactions with the phosphate backbone of the DNA, and the sliding of PCNA is achieved by all the positive side chains stochastically forming interactions with the phosphate atoms adjacent to the initially bound nucleotides. DVNP may utilise a similar mechanism for the observed mobility. In addition, the single-molecule spectroscopy results presented in Chapter 6 also demonstrated that *Hematodinium* DVNP.6 modified the properties of DNA. These results suggested that DNA intercalation events may take place in the DVNP-DNA binding inside a dinokaryon, or through a currently unknown mechanism lengthen DNA as well as reducing the resistance of DNA to tension.

This thesis also provides evidence that DVNP of the *Ostreococcus* virus switches between two conformations. Based on the similarity of behaviours of DVNPs from both a dinoflagellate (*Hematodinium* DVNP.6) and the virus in many of the assays described in this thesis, it is possible that the switching behaviour applies to the dinoflagellate DVNPs as well and could be a universal property of this molecule. Based on the NMR structures described in chapter 7, the most possible molecular scenario of the conformation change would be a switch between: A) the tightly packed ‘closed’ form where the C-terminal tail shields the N-terminal structure as well as holding the different elements of the structure together, and B) the ‘open’

form, where the C-terminal tail is not associated or loosely associated with the N-terminal structured region.

Based on the above results, I here posit a molecular model of DVNP-mediated DNA compaction (Fig. 8.1). DVNP's behaviour as monomer in solution switches between the two conformations, allowing for the free positively charged C-terminal tail to bind the phosphate backbone of dsDNA. When bound, the N-terminal half of the protein, still in the open conformation, becomes prone to forming dimers or oligomers with other DNA-bound DVNP monomers. When an oligomer is formed by monomers bound to different pieces of DNA or indeed different parts of the same piece of DNA, compaction is thus generated

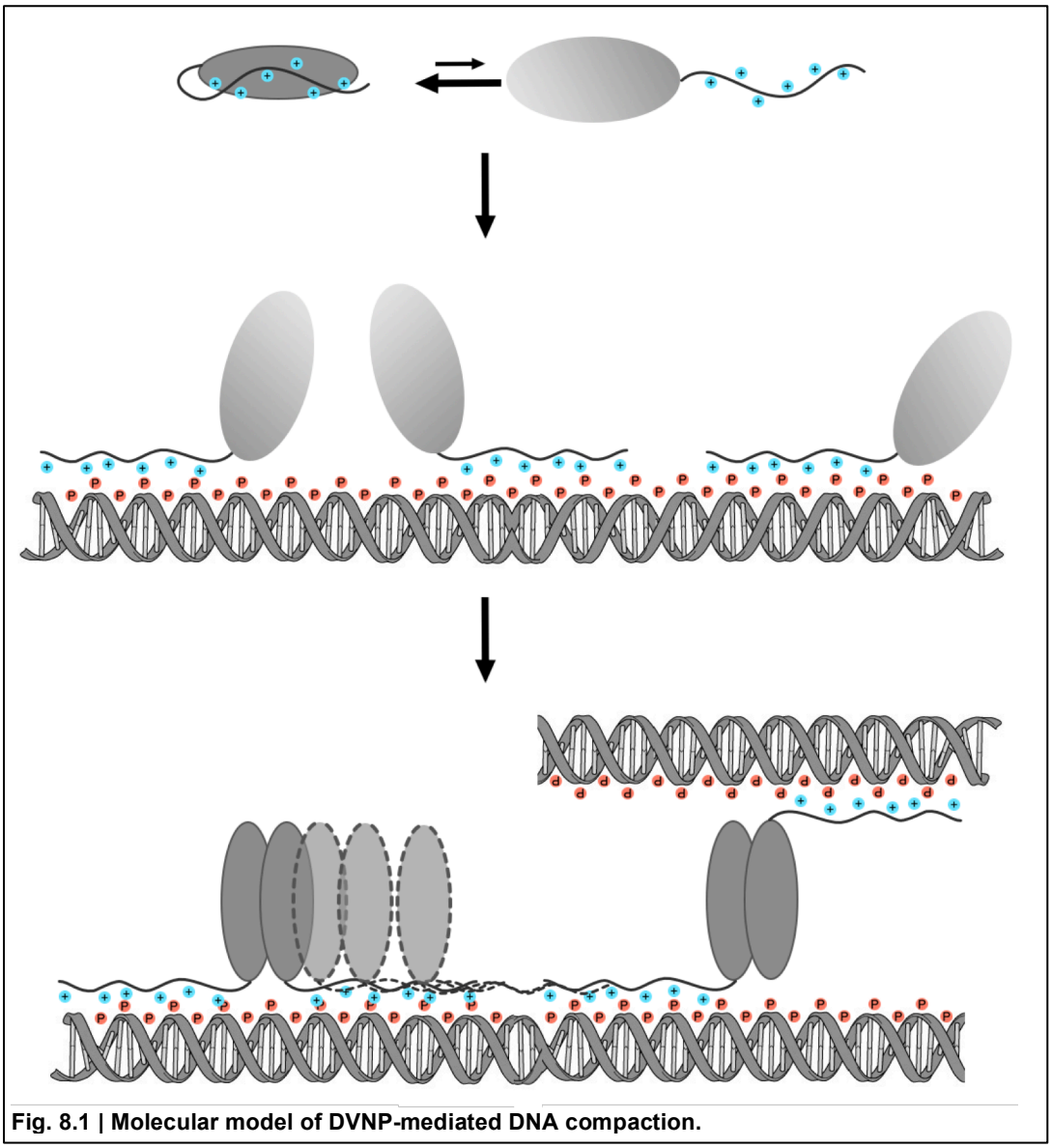


Fig. 8.1 | Molecular model of DVNP-mediated DNA compaction.

The charge and charge distribution analyses in Chapter 3 presented that comparatively speaking variation in dinoflagellate DVNP protein length is much higher than the viral DVNP proteins, and the same trend holds true for total charges. However, dinoflagellate DVNPs have a greater tendency to maintain a fixed ratio of charge versus length, the homogeneity of charge distribution, as well as the ratio of total positive versus negative charges. This tendency implies a selective pressure for the dinokaryon to have more control of the traits of charge/length, charge distribution, and total positive/negative charges over its DVNP (Table 8.1). The viral DVNPs may face less pressure than the dinoflagellate DVNPs. Whether the viral DVNP is a DNA management protein for the viral genome or as an effector to attenuate the host, the viral DVNP could afford more freedom as long as viral DNA is decompacted within the host and the replication inside the host cell is completed. Yet, for dinoflagellates, the properties of the putative DNA-packaging protein DVNP may be crucial for many processes within the nucleus. In the previous paragraphs, I posited the DVNP-DNA binding model in which the positive charges of the protein form interactions with the phosphate backbones of the DNA (Fig. 8.1). The affinity of the C-terminal tail to the phosphate backbone may be highly dependent on the charge distribution as well as charge strength. Under this scenario, DVNP proteins with the placement of charges showing unconstrained change would have cost the dinoflagellate cell its finely tuned relationships between the transcription machinery and the chromosomes. The hypothesis of dinoflagellate DVNPs using charge distribution to regulate the DNA compaction is consistent with independent experimental results presented throughout this thesis. However, more experiments are required to verify the hypothesis.

	dino vs viral
Variation in length	>
Variation in charge	>
Variation in charge/length	<
Variation in charge distribution	<
Variation in pos/neg charges	<

Table 8.1 | Comparisons of dinoflagellate versus viral DVNP sequences

The very first dinoflagellate expressing the gene DVNP was likely infected by a DVNP-hosting large DNA virus. The virus likely used DVNP for its own DNA packaging, but DVNP may also have a dual role as a potential virulence factor (Gornik *et al.* 2019). According to experiments expressing DVNP in other eukaryotes presented by Gornik et al. 2012, Goh & Waller 2015, and Irwin et al. 2018, this would

have had dramatic health consequences for this infected cell, possibly including histone displacement from the host's chromatin. I note that in the aforementioned three publications, no clear observations of dramatic changes in chromatin structure were observed in the experimental models (*Toxoplasma gondii* and baker's yeast) expressing DVNP heterologously. Presumably, the very first DVNP-expressing dinoflagellate cell also did not experience dramatic chromatin restructuring. Regardless of the source of the genome expansion, I note that genome expansion could have been a beneficial response to the DVNP expression in early dinoflagellate. Irwin et al. had shown that over-expression of DVNP in a yeast cell reduced transcription initiation by occupancy, presumably by hindering transcription factors and RNAP molecules locating and binding the promoters. With an expanded genome and duplicated genes, the overall chance for a vital gene to be found by RNAP and hence transcribed would have become larger, although the chances of each individual copy of the gene being transcribed would have remained low. However, it is equally possible that the expansion of the genome occurred independently before the introduction of DVNP. An expanded genome with duplicated genes prior to the challenge of DVNP would have yielded the same results, and afforded the cell to better tolerate the protein DVNP, while making up the lost fitness for having a larger genome than necessary (Bentkowski *et al.* 2015; Lefébure *et al.* 2017).

Results presented in this thesis suggest that DVNP only induces phase-separated DNA compaction above a certain DNA concentration threshold. If DVNP were indeed the main factor for the appearance of the condensed chromosomes of current dinoflagellates, I argue that the condensation likely would have been a derived feature that occurred later than the introduction of DVNP and the initiation of genome expansion, after the genome expansion brought the DNA concentration over the threshold. For comparison, Livolant demonstrated that *in vitro* formation of cholesteric PSI-DNA compaction and polycation-mediated DNA compaction, with similar CD spectra with the dinokaryotic chromosomes (Livolant & Maestre 1988), both required high concentration of DNA (Livolant 1986). In addition, the euglenids also host permanently condensed chromosomes. The studies on euglenid nuclear genomes are few (Bicudo & Menezes 2016), and cytophotometric methods placed the genome sizes from 3 Gb to 24 Gb (Ebenezer *et al.* 2017; Zakryś 1988). The only sequence-based study on the genome of a euglenid, *Euglena gracilis*, estimated its

genome size at 500 Mb (Ebenezer *et al.* 2019). The canonical nuclei-bearing kinetoplastids, extracellular parasitic and free-living relatives of the euglenids, have genome sizes less than 40 Mb (Berriman *et al.* 2005; Ivens *et al.* 2005; Jackson *et al.* 2008; Real *et al.* 2013). Additionally, bacterial over-expression of HCc3, another DNA-compacting protein that was acquired by the late-branching dinoflagellates, has shown to compact *E. coli* chromatin (Sun *et al.* 2013). On the other hand, DNA compaction was not observed in DVNP-expressing *E. coli* (Chapter 5). These results support that the chromosome compaction in the dinokaryon, or more broadly DNA phase-separation, may require a certain threshold of DNA concentration to be reached. As the dinoflagellate cell walked down the one-way street of evolution, the protein DVNP, initially a viral DNA packaging protein and/or virulence factor, gradually co-evolved and eventually became tamed over the genome expansion process, where the cell reached a balance between the compacted state of DNA and accessibility of the transcription machinery. The two hypotheses proposed here, namely that genome expansion or the introduction of DVNP might have helped to establish the other, and that chromosome compaction was the consequence of the presence of DVNP and expanded genome size, are coherent with the data presented in this thesis. Both are highly speculative and require more sophisticated methodologies to test.

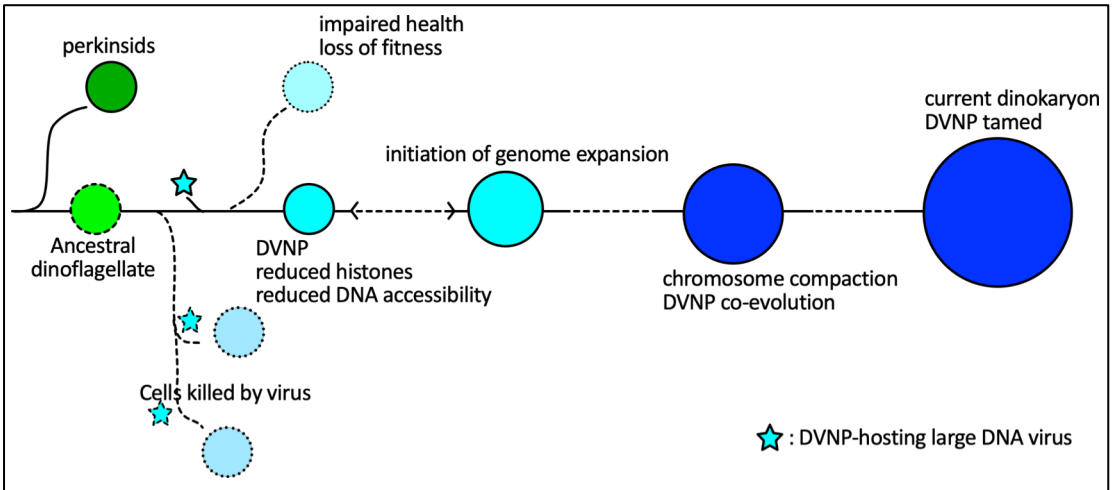


Fig. 8.2 | Proposed projection of evolution history of the dinoflagellates. Sizes of the spheres represent genome sizes. Cyan, presence of DVNP. Dark blue, condensed chromosome state.

I note that this thesis, though broad by methodologies used, utilises simplified elements in each *in vitro* experiments. None of the recombinantly purified DVNP experiments took into account of the potential post-translational modification, and none of the double-stranded DNA includes modified bases, such as 5-hydroxymethyluracil replacement of thymine, or epigenetic markers such as methylation or acetylation. The bacterial HU protein is known to modulate its DNA compaction by acetylation on its abundant lysine residues (Ghosh et al. 2016), and it is very likely that the post-translational marker on DVNP may help to regulate the DNA-packaging processes. Yet, I believe the methodologies used in this thesis are sufficient for qualitative studies, partly based on the correct folding of natively-purified DVNP proteins, which fold differently when compared to a refolded DVNP protein (Goh, 2015). In addition, the results that were obtained *in vitro* conform with available *in vivo* data, including the NMR protein structure model and the Gremlin contact map structure prediction, as well as the fluidity of the DVNP proteins observed on C-trap and the sewing machine model deduced from imaging techniques. It is my sincere hope that this thesis serves as a starting point for more quantitative studies in the future that will eventually unravel the mystery of the dinokaryon.

To summarise, this thesis has provided evidence that the protein DVNP, of which the occurrence seemed to have coincided with genome expansion and chromosome condensation in dinoflagellate evolution history, is sufficient to drive DNA into a highly compacted state through phase separation. The phase-separated DVNP-DNA complex is highly dynamic and fluid, and relative motion of DVNP protein and DNA is possible. In combination with the localisation results of RNA polymerase II and nascent RNA, these results provide the possibility of a new type of DNA management system. This putative system conforms to some of the known enigmatic and seemingly contradictory properties of the dinokaryon: namely, 1) the liquid crystalline condensed chromosomes, 2) very large genes that probably inevitably span the width of the chromosomes, and 3) transcription mostly occurs on the periphery of the chromosomes. By hypothesising that RNA polymerase outside the chromosomes draws the fluid DNA from within the chromosome cores, this model resolves the conflicts between the known properties. It is, however, still unclear how the initial events of the dinokaryon formation process occurred, and whether chromosome compaction is linked to the increased genome size. These questions may require

further knowledge of the roles of viral DVNPs to answer. This work has demonstrated that there are many similarities to the DVNP proteins from dinoflagellates and viruses. Yet, there are notable differences in protein size and charge distribution properties as well as the presence of a charged N-terminal tail. Understanding the role of DVNP for viruses might help us predict the early effects of viral DVNP in an early dinoflagellate, and how the viral DVNP-DNA relationship transformed into a stable chromatin management system. This work has provided the first broad insight into the molecular properties of DVNP, and how these might contribute to a radical new chromatin system for eukaryotes. It now paves the way for these further questions to be addressed of the molecular mechanisms, and evolutionary processes, that allowed dinoflagellates to make such changes but remain major players in global food webs and geochemical systems.

Bibliography

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., McCourt, R.M., Mendoza, L., Moestrup, O., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A. V, Spiegel, F.W. & Taylor, M.F.J.R. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology* **52**, 399.

Alfonso-Parra, C. & Maggert, K.A. (2010). Drosophila SAF-B Links the Nuclear Matrix, Chromosomes, and Transcriptional Activity. *PLOS ONE* **5**, e10248.

Allawi, H.T. & SantaLucia, J. (1997). Thermodynamics and NMR of Internal G·T Mismatches in DNA. *Biochemistry* **36**, 10581–10594.

Allen, J.R., Roberts, T., Loeblich, A. & Klotz, L. (1975). Characterization of the DNA from the dinoflagellate crypthecodinium cohnii and implications for nuclear organization. *Cell* **6**, 161–169.

Almaqwashi, A.A., Paramanathan, T., Rouzina, I. & Williams, M.C. (2016). Mechanisms of small molecule-DNA interactions probed by single-molecule force spectroscopy. *Nucleic Acids Research* **44**, 3971–3988.

Anjana, R., Vaishnavi, M.K., Sherlin, D., Kumar, S.P., Naveen, K., Kanth, P.S. & Sekar, K. (2012). Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes : a study based on orientation and distance. *Bioinformation* **8**, 1220–1224.

Annunziato, A. (2008). DNA Packaging: Nucleosomes and Chromatin. *Nature Education* **1**, 26.

Aranda, M., Li, Y., Liew, Y.J., Baumgarten, S., Simakov, O., Wilson, M.C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V.B., Ryu, T., Ravasi, T., Bayer, T., Micklem, G., Kim, H., Bhak, J., LaJeunesse, T.C. & Voolstra, C.R. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports* **6**, 1–15.

Arold, S.T., Leonard, P.G., Parkinson, G.N. & Ladbury, J.E. (2010). H-NS forms a superhelical protein scaffold for DNA condensation. Proceedings of the National Academy of Sciences 107, 15728–15732.

Babillot, C. (1970). Study of incorporation of uridine-3H in the nucleus in the Amphidinium carteri, Dinoflagellata. *Comptes rendus hebdomadaires des seances de l'Academie des*

sciences. Serie D: Sciences naturelles **271**, 828–831.

Bachvaroff, T.R., Concepcion, G.T., Rogers, C.R., Herman, E.M. & Delwiche, C.F. (2004). Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* **155**, 65–78.

Bachvaroff, T.R. & Place, A.R. (2008). From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate Amphidinium carterae. *PLoS One* **3**, e2929.

Baumann, C.G., Bloomfield, V.A., Smith, S.B., Bustamante, C., Wang, M.D. & Block, S.M. (2000). Stretching of single collapsed DNA molecules. *Biophysical Journal* **78**, 1965–1978.

Baumgarten, S., Bayer, T., Aranda, M., Liew, Y., Carr, A., Micklem, G. & Voolstra, C. (2013). Integrating microRNA and mRNA expression profiling in Symbiodinium microadriaticum, a dinoflagellate symbiont of reef-building corals. *BMC genomics* **14**, 704.

Beauchemin, M., Roy, S., Daoust, P., Steve, D.-B., Bertomeu, T., Letourneau, L., Lang, F.B. & Morse, D. (2012). Dinoflagellate tandem array gene transcripts are highly conserved and not polycistronic. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15793–15798.

Bensaude, O. (2011). Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity? *Transcription* **2**, 103–108.

Bentkowski, P., Van Oosterhout, C. & Mock, T. (2015). A Model of Genome Size Evolution for Prokaryotes in Stable and Fluctuating Environments. *Genome Biology and Evolution* **7**, 2344–2351.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. *Nature reviews. Genetics* **15**, 163–175.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., Bohme, U., Hannick, L., Aslett, M.A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U.C.M., Arrowsmith, C., Atkin, R.J., Barron, A.J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T.-J., Churcher, C., Clark, L.N., Corton, C.H., Cronin, A., Davies, R.M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M.C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B.R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A.X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P.J., Moule, S., Martin, D.M.A.,

Morgan, G.W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C.S., Peterson, J., Quail, M.A., Rabbinowitsch, E., Rajandream, M.-A., Reitter, C., Salzberg, S.L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A.J., Tallon, L., Turner, C.M.R., Tait, A., Tivey, A.R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M.D., Embley, T.M., Gull, K., Ullu, E., Barry, J.D., Fairlamb, A.H., Opperdoes, F., Barrell, B.G., Donelson, J.E., Hall, N., Fraser, C.M., Melville, S.E. & El-Sayed, N.M. (2005). The genome of the African trypanosome Trypanosoma brucei. *Science* **309**, 416–422.

Bhowmick, T., Ghosh, S., Dixit, K., Ganesan, V., Ramagopal, U.A., Dey, D., Sarma, S.P., Ramakumar, S. & Nagaraja, V. (2014). Targeting Mycobacterium tuberculosis nucleoid-associated protein HU with structure-based inhibitors. Nature Communications 5, 4124.

Bhaud, Y., Géraud, M.L., Ausseil, J., Soyer-Gobillard, M.-O. & Moreau, H. (1999). Cyclic expression of a nuclear protein in a dinoflagellate. *The Journal of eukaryotic microbiology* **46**, 259–267.

Bhaud, Y., Guillebault, D., Lennon, J., Defacque, H., Soyer-Gobillard, M.-O. & Moreau, H. (2000). Morphology and behaviour of dinoflagellate chromosomes during the cell cycle and mitosis. *Journal of cell science* **113**, 1231–1239.

Biancalana, M. & Koide, S. (2010). Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochimica et biophysica acta* **1804**, 1405–1412.

Bicudo, C.E.M. & Menezes, M. (2016). Phylogeny and classification of euglenophyceae: A brief review. *Frontiers in Ecology and Evolution* **4**, 1–15.

Biebricher, A.S., Heller, I., Roijmans, R.F.H., Hoekstra, T.P., Peterman, E.J.G. & Wuite, G.J.L. (2015). The impact of DNA intercalators on DNA and DNA-processing enzymes elucidated through force-dependent binding kinetics. *Nature Communications* **6**, 1–12.

Bloomfield, V.A. (1996). DNA condensation. *Current opinion in structural biology* **6**, 334–341.

Bloomfield, V.A. (1997). DNA Condensation by Multivalent Cations. *Biopolymers* **44**, 269–282.

Bodansky, S., Mintz, L.B. & Holmes, D.S. (1979). The mesokaryote Gyrodinium cohnii lacks nucleosomes. *Biochemical and Biophysical Research Communications* **88**, 1329–1336.

Bohatier, J. (1981). Synthèses macromoléculaires au cours de la morphogenèse de régénération chez le cilié Condylostoma magnum: Action de differents inhibiteurs I. Synthèses d'Acides nucleiques. *Archiv für Protistenkunde* **124**, 144–172.

Bouligand, Y. (1972). Twisted fibrous arrangements in biological materials and cholesteric mesophases. *Tissue Cell* **4**, 189–217.

Bouligand, Y. & Norris, V. (2001). Chromosome separation and segregation in dinoflagellates andbacteria may depend on liquid crystalline states. *Biochimie* 187–192.

Breslauer, K.J., Frank, R., Blöcker, H. & Marky, L.A. (1986). Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3746–3750.

Brickey, W.J. & Greenleaf, A.L. (1995). Functional Studies of the Carboxy-Terminal Repeat Domain of Drosophila RNA Polymerase II in Vivo. *Genetics* **140**, 599–613.

Brower-Toland, B.D., Smith, C.L., Yeh, R.C., Lis, J.T., Peterson, C.L. & Wang, M.D. (2002). Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proceedings of the National Academy of Sciences* **99**, 1960 LP – 1965.

Brown, P.H. & Schuck, P. (2006). Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophysical journal* **90**, 4651–4661.

Burreson, E.M., Ragone Calvo, L.M., La Peyre, J.F., Counts, F. & Paynter, K.T. (1994). Acute osmotic tolerance of cultured cells of the oyster pathogen Perkinsus marinus (Apicomplexa:Perkinsida). *Comparative biochemistry and physiology. Part A, Physiology* **109**, 575–82.

Cachon, J., Sato, H., Cachon, M. & Sato, Y. (1989). Analysis by polarizing microscopy of chromosomal structure among dinoflagellates and its phylogenetic involvement. *Biology of the Cell* **65**, 51–60.

Cai, S., Zeng, C., Li, J. & Zhai, Z. (1992). Identification of the nuclear matrix and chromosome scaffold in dinoflagellate Crypthecodinium cohnii. *Cell Research* **2**, 165–181.

Campbell, D.A., Thomas, S. & Sturm, N.R. (2003). Transcription in kinetoplastid protozoa: why be normal? *Microbes and Infection* **5**, 1231–1240.

Chambouvet, A., Gower, D.J., Jirků, M., Yabsley, M.J., Davis, A.K., Leonard, G., Maguire, F., Doherty-Bone, T.M., Bittencourt-Silva, G.B., Wilkinson, M. & Richards, T.A. (2015). Cryptic infection of a broad taxonomic and geographic diversity of tadpoles by Perkinsea protists. *Proceedings of the National Academy of Sciences* **112**, E4743 LP-E4751.

Chan, Y.-H. & Wong, J.T.Y. (2007). Concentration-dependent organization of DNA by the dinoflagellate histone-like protein HCc3. Nucleic Acids Research 35, 2573–2583.

Chan, Y., CM Kwok, A., SH Tsang, J. & TY Wong, J. (2006). Alveolata histone-like proteins have different evolutionary origins. *Journal of Evolutionary Biology* **19**, 1717–1721.

Chen, J.E., Barbrook, A.C., Cui, G., Howe, C.J. & Aranda, M. (2019). The genetic intractability of Symbiodinium microadriaticum to standard algal transformation methods. PLoS ONE 14, 1–19.

Cherstvy, A.G. (2008). DNA Cholesteric Phases : The Role of DNA Molecular Chirality and DNA - DNA Electrostatic Interactions. *Journal of Physical Chemistry B* **112**, 12585–12595.

Cheung, M.-S., Maguire, M.L., Stevens, T.J. & Broadhurst, R.W. (2010). DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of magnetic resonance* **202**, 223–233.

Chow, M.H., Yan, K.T., Bennett, M.J. & Wong, J.T. (2010). Birefringence and DNA condensation of liquid crystalline chromosomes. *Eukaryotic Cell* **9**, 1577–1587.

Cloe, A.L., Orgel, J.P.R.O., Sachleben, J.R., Tycko, R. & Meredith, S.C. (2011). The Japanese mutant Aβ (ΔE22-Aβ(1-39)) forms fibrils instantaneously, with low-thioflavin T fluorescence: seeding of wild-type Aβ(1-40) into atypical fibrils by ΔE22-Aβ(1-39). *Biochemistry* **50**, 2026–2039.

Clore, M.G. (2011). Exploring translocation of proteins on DNA by NMR. *Journal of Biomolecular NMR* **51**, 209–219.

Coats, D.W. (1999). Parasitic life styles of marine dinoflagellates. *J Euk Microbiol* **46**, 402–409.

Cochet-Meilhac, M., Nuret, P., Courvalin, J.C., Chambon, P. & Mddecine, F. De. (1974). Determination of the Cellular Number of RNA Polymerase B Molecules. *Biochimica et biophysica acta* **353**, 185–192.

Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* **13**, 289–302.

Cramer, P., Bushnell, D.A. & Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876.

Cummings, C.S. & Obermeyer, A.C. (2018). Phase Separation Behavior of Supercharged Proteins and Polyelectrolytes. *Biochemistry* **57**, 314–323.

Dale, R.M.K., McClure, B.A. & Houchins, J.P. (1985). A rapid single-stranded cloning

strategy for producing a sequential series of overlapping clones for use in DNA sequencing: Application to sequencing the corn mitochondrial 18S rDNA. *Plasmid* **13**, 31–40.

Das, R.K. & Pappu, R. V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences* **110**, 13392 LP – 13397.

Davies, W., Jakobson, K.S. & Nordby, Ø. (1988). Characterization of DNA from the Dinoflagellate Woloszynskia bostoniensis. *The Journal of Protozoology* **35**, 418–422.

Deen, K.C., Landers, T.A. & Berninger, M. (1983). Use of T4 DNA polymerase replacement synthesis for specific labeling of plasmid-cloned inserts. *Analytical Biochemistry* **135**, 456–465.

Derouchey, J., Hoover, B. & Rau, D.C. (2013). A comparison of DNA compaction by arginine and lysine peptides: A physical basis for arginine rich protamines. *Biochemistry* **52**, 3000–3009.

Dodge, J.D. (1965). Chromosome structure in the dinoflagellates and the problem of the mesokaryotic cell. *Excerpta Medica International Congress Series* **91**, 264–265.

Douris, V., Telford, M.J. & Averof, M. (2010). Evidence for multiple independent origins of trans-splicing in Metazoa. *Molecular biology and evolution* **27**, 684–93.

Dunigan, D.D., Fitzgerald, L.A. & Van Etten, J.L. (2006). Phycodnaviruses: a peek at genetic diversity. *Virus research* **117**, 119–132.

Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M. & Luger, K.B.T.-M. in E. (2003). Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. In *Chromatin and Chromatin Remodeling Enzymes, Part A*: 23–44. Academic Press.

Ebenezer, T.E., Carrington, M., Lebert, M., Kelly, S. & Field, M.C. (2017). Euglena gracilis Genome and Transcriptome: Organelles, Nuclear Genome Assembly Strategies and Initial Features. In *Euglena: Biochemistry, Cell and Molecular Biology*: 125–140. Schwartzbach, S.D. & Shigeoka, S. (Eds.). Cham: Springer International Publishing.

Ebenezer, T.E., Zoltner, M., Burrell, A., Nenarokova, A., Novák Vanclová, A.M.G., Prasad, B., Soukal, P., Santana-Molina, C., O'Neill, E., Nankissoor, N.N., Vadakedath, N., Daiker, V., Obado, S., Silva-Pereira, S., Jackson, A.P., Devos, D.P., Lukeš, J., Lebert, M., Vaughan, S., Hampl, V., Carrington, M., Ginger, M.L., Dacks, J.B., Kelly, S. & Field, M.C. (2019). Transcriptome, proteome and draft genome of Euglena gracilis. *BMC Biology* **17**, 1–23.

Echeverría, O.M., Jiménez-García, L.F., Gonzalez-Cerón, G., Elizundia, J.M. & Vázquez-Nin, G.H. (1993). Cytochemical and autoradiographic study of the nucleus of a symbiotic dinoflagellate. *Caryologia* **46**, 261–274.

Engler, C., Kandzia, R. & Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PloS one* **3**, e3647.

Van Etten, J.L., Graves, M. V, Muller, D.G., Boland, W. & Delaroque, N. (2002). Phycodnaviridae--large DNA algal viruses. *Archives of virology* **147**, 1479–1516.

Van Etten, J.L., Lane, L.C. & Meints, R.H. (1991). Viruses and viruslike particles of eukaryotic algae. *Microbiological reviews* **55**, 586–620.

Ewen, M.E. (2000). Where the cell cycle and histones meet. *Genes & development* **14**, 2265–2270.

Fenley, A.T., Adams, D.A. & Onufriev, A. V. (2010). Charge state of the globular histone core controls stability of the nucleosome. *Biophysical journal* **99**, 1577–1585.

Fensome, R.A., Saldarriaga, J.F. & Taylor, F. "Max" J.R. (1999). Dinoflagellate phylogeny revisited: reconciling morphological and molecular based phylogenies. *Grana* **38**, 66–80.

Flors, C., Ravarani, C.N.J. & Dryden, D. (2009). Super‐Resolution Imaging of DNA Labelled with Intercalating Dyes. *Chemphyschem* 2201–2204.

Fukuda, Y. & Suzaki, T. (2015). Unusual features of dinokaryon, the enigmatic nucleus of dinoflagellates. In *Marine Protists: Diversity and Dynamics*: 23–45.

Gajadhar, A.A., Marquardt, W.C., Hall, R., Gunderson, J., EV, A.-C. & Sogin, M.L. (1991). Ribosomal RNA sequences of Sarcocystis muris, Theileria annulata and Crypthecodinium cohnii reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Molecular and Biochemical Parasitology* **45**, 147–154.

Gallot-Lavallée, L. & Blanc, G. (2017). A glimpse of nucleo-cytoplasmic large DNA virus biodiversity through the eukaryotic genomics window. *Viruses* **9**, E17.

Gallot-Lavallée, L., Blanc, G. & Claverie, J.-M. (2017). Comparative Genomics of Chrysochromulina Ericina Virus and Other Microalga-Infecting Large DNA Viruses Highlights Their Intricate Evolutionary Relationship with the Established Mimiviridae Family. *Journal of virology* **91**, 1–16.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H. & Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Science* 102–105.

Gao, X.P. & Li, J.Y. (1986). Nuclear division in the marine dinoflagellate Oxyrrhis marina. *Journal of cell science* **85**, 161–175.

Garnier, J., Osguthorpe, D.J. & Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of molecular biology* **120**, 97–120.

Ghosh, S., Padmanabhan, B., Anand, C. & Nagaraja, V. (2016). Lysine acetylation of the Mycobacterium tuberculosis HU protein modulates its DNA binding and genome organization. Molecular microbiology 100, 577–588.

Gibson, D.G., Young, L., Chuang, R.Y. & Venter, J.C. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–347.

Goh, Y.H. & Waller, R.F. (2015). *An Investigation of Dinoflagellate Nuclear Reorganisation Following Replacement of Histones with a Novel Nucleoprotein, Master's thesis*. Univeristy of Cambridge.

Gornik, S.G., Ford, K.L., Mulhern, T.D., Bacic, A., McFadden, G.I. & Waller, R.F. (2012). Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Current Biology* **22**, 2303–2312.

Gornik, S.G., Hu, I., Lassadi, I. & Waller, R.F. (2019). The Biochemistry and Evolution of the Dinoflagellate Nucleus. *Microorganisms* **7**, 245.

Grainger, D.C. (2016). Structure and function of bacterial H-NS protein. Biochemical Society Transactions 44, 1561–1569.

Green, M.R. & Sambrook, J. (2018). The Hanahan Method for Preparation and Transformation of Competent Escherichia coli: High-Efficiency Transformation. *Cold Spring Harbor protocols* **2018**.

Guiry, M.D. (2012). How many species of algae are there? *Journal of Phycology* **48**, 1057–1063.

Hacker, K.J. & Alberts, B.M. (1994). The rapid dissociation of the T4 DNA polymerase holoenzyme when stopped by a DNA hairpin helix. A model for polymerase release following the termination of each Okazaki fragment. *The Journal of biological chemistry* **269**, 24221–24228.

Hargraves, P.E. (1998). Identifying Marine Phytoplankton. *Eos, Transactions American Geophysical Union* **79**, 99–99.

Herzog, M., Boletzky, S. & Soyer, M.-O. (1984). Ultrastructural and biochemical nuclear aspects of Eukaryote classification: Independent evolution of the dinoflagellates as a sister group of the actual Eukaryotes? *Origins of Life* **13**, 205–215.

Herzog, M. & Soyer, M.-O. (1981). Distinctive features of dinoflagellate chromatin. Absence of nucleosomes in a primitive species Prorocentrum micans E. *European journal of cell*

biology **23**, 295–302.

Herzog, M. & Soyer, M.-O. (1983). The native structure of dinoflagellate chromosomes and their stabilization by Ca2+ and Mg2+ cations. *European journal of cell biology* **30**, 33–41.

Hinnebusch, A., Klotz, L., Immergut, E. & Loeblich, A. (1980). Deoxyribonucleic acid sequence organization in the genome of the dinoflagellate Crypthecodinium cohnii. *Biochemistry* **19**, 1744–1755.

Hirano, Y., Takahashi, H., Kumeta, M., Hizume, K., Hirai, Y., Otsuka, S., Yoshimura, S.H. & Takeyasu, K. (2008). Nuclear architecture and chromatin dynamics revealed by atomic force microscopy in combination with biochemistry and cell biology. *Pflügers Archiv : European journal of physiology* **456**, 139–53.

Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O.G. & Pappu, R. V. (2017). CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophysical journal* **112**, 16–21.

Hoppenrath, M. & Leander, B.S. (2006). Dinoflagellate, Euglenid, or Cercomonad? The ultrastructure and molecular phylogenetic position of Protaspis grandis n. sp. *The Journal of eukaryotic microbiology* **53**, 327–342.

Howe, C.J., Nisbet, R.E.R. & Barbrook, A.C. (2008). The remarkable chloroplast genome of dinoflagellates. Journal of Experimental Botany 59, 1035–1045.

Howe, C., Ho, F., Nenninger, A., Raleiras, P. & Stensjö, K. (2018). Differential biochemical properties of three canonical Dps proteins from the cyanobacterium Nostoc punctiforme suggest distinct cellular functions. Journal of Biological Chemistry 293, 16635–16646.

Howlett, G.J., Minton, A.P. & Rivas, G. (2006). Analytical ultracentrifugation for the study of protein association and assembly. *Current Opinion in Chemical Biology* **10**, 430–436.

Hunt, C.R., Ramnarain, D., Horikoshi, N., Iyengar, P., Pandita, R.K., Shay, J.W. & Pandita, T.K. (2013). Histone Modifications and DNA Double-Strand Break Repair after Exposure to Ionizing Radiations. *Radiation research* **179**, 383–392.

Irwin, N.A.T., Martin, B.J.E., Young, B.P., Browne, M.J.G., Flaus, A., Loewen, C.J.R., Keeling, P.J. & Howe, L.J. (2018). Viral proteins as a potential driver of histone depletion in dinoflagellates. *Nature Communications* **9**, 1535.

Ishida, T. & Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic acids research* **35**, W460-4.

Isidoro-Ayza, M., Lorch, J.M., Grear, D.A., Winzeler, M., Calhoun, D.L. & Barichivich, W.J. (2017). Pathogenic lineage of Perkinsea associated with mass mortality of frogs

across the United States. *Scientific Reports* **7**, 10288.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M., Adlem, E., Aert, R. & others. (2005). The Genome of the Kinetoplastid Parasite, Leishmania major. *Science* **309**, 436.

Iwahara, J., Schwieters, C.D. & Clore, M.G. (2004). Ensemble Approach for NMR Structure Refinement against 1H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule. *Journal of the American Chemical Society* **126**, 5879–5896.

Jackson, A.P., Quail, M.A. & Berriman, M. (2008). Insights into the genome sequence of a free-living Kinetoplastid: Bodo saltans (Kinetoplastida: Euglenozoa). *BMC genomics* **9**, 594.

Jackson, C.J., Gornik, S.G. & Waller, R.F. (2012). The Mitochondrial Genome and Transcriptome of the Basal Dinoflagellate Hematodinium sp.: Character Evolution within the Highly Derived Mitochondrial Genomes of Dinoflagellates. Genome Biology and Evolution 4, 59–72.

Jackson, C.J. & Waller, R.F. (2013). A Widespread and Unusual RNA Trans-Splicing Type in Dinoflagellate Mitochondria. *PloS One* **8**, e56777.

JASP Team. (2019). JASP (Version 0.10.2) [Computer software].

Janissen, R., Arens, M.M.A., Vtyurina, N.N., Rivai, Z., Sunday, N.D., Eslami-Mossallam, B., Gritsenko, A.A., Laan, L., de Ridder, D., Artsimovitch, I., Dekker, N.H., Abbondanzieri, E.A. & Meyer, A.S. (2018). Global DNA Compaction in Stationary-Phase Bacteria Does Not Affect Transcription. Cell 174, 1188-1199.e14.

Jenuwein, T. & Allis, C.D. (2001). Translating the histone code. *Science* **293**, 1074–1080.

Jeong, H.J., du Yoo, Y., Kim, J.S., Seong, K.A., Kang, N.S. & Kim, T.H. (2010). Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. *Ocean Science Journal* **45**, 65–91.

Jones, D.T., Taylor, W.R., & Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences 8: 275-282.

Kamikawa, R., Inagaki, Y., Tokoro, M., Roger, A.J. & Hashimoto, T. (2011). Report Split Introns in the Genome of Giardia intestinalis Are Excised by Spliceosome-Mediated trans -Splicing. *Current Biology* **21**, 311–315.

Kamisetty, H., Ovchinnikov, S. & Baker, D. (2013). Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences* **110**, 15674 LP – 15679.

Karamanos, T.K., Kalverda, A.P., Thompson, G.S. & Radford, S.E. (2015). Mechanisms of amyloid formation revealed by solution NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **88–89**, 86–104.

Kato, K.H., Moriyama, A., Huitorel, P. & Cosson, J. (1997). Isolation of the major basic nuclear protein and its localization on chromosomes of the dinoflagellate, Oxyrrhis marina. *Biology of the Cell* 43–52.

Kearns, L. & Sigee, D. (1980). The occurrence of period IV elements in dinoflagellate chromatin: an X-ray microanalytical study. *Journal of cell science* **46**, 113–127.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A. & Worden, A.Z. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLOS Biology* **12**, e1001889.

Kellenberger, E., Johansen, R., Maeder, M., Bohrmann, B., Stauffer, E. & Villiger, W. (1992). Artefacts and morphological changes during chemical fixation. *Journal of Microscopy* **168**, 181–201.

Kelly, S. & Price, N. (2000). The Use of Circular Dichroism in the Investigation of Protein Structure and Function. *Current Protein & Peptide Science* **1**, 349–384.

Kelly, T.J.J. & Smith, H.O. (1970). A restriction enzyme from Hemophilus influenzae. II. *Journal of molecular biology* **51**, 393–409.

Khurana, R., Coleman, C., Ionescu-Zanetti, C., Carter, S.A., Krishna, V., Grover, R.K., Roy, R. & Singh, S. (2005). Mechanism of thioflavin T binding to amyloid fibrils. *Journal of Structural Biology* **151**, 229–238.

Kissinger, J.C., Gajria, B., Li, L., Paulsen, I.T. & Roos, D.S. (2003). ToxoDB: Accessing the Toxoplasma gondii genome. *Nucleic Acids Research* **31**, 234–236.

Klobutcher, L.A. (2001). Macronucleus. In *Encyclopedia of Genetics*: 1135–1136. Brenner, S. & Miller, J.H.B.T.-E. of G. (Eds.). New York: Academic Press.

Koltover, I., Wagner, K. & Safinya, C.R. (2000). DNA condensation in two dimensions. *Proceedings of the National Academy of Sciences* **97**, 14046–14051.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549.

Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proceedings of the National Academy of Sciences* **82**, 488–492.

Kuznetsov, S. V, Sugimura, S. & Vivas, P. (2006). Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proceedings of the National Academy of Sciences* **103**, 18515–18520.

LaJeunesse, T.C., Lambert, G. & Andersen, R.A. (2005). Symbiodinium (Pyrrhophyta) genome sizes (DNA content) are smallest among dinoflagellates. *Journal of Phycology* **41**, 880–886.

Leander, B.S. & Keeling, P.J. (2004). Early evolutionary history of dinoflagellates and apicomplexans (Alveolata) as inferred from hsp90 and actin phylogenies. *Journal of Phycology* **40**, 341–350.

Lees, J.G., Miles, A.J., Wien, F. & Wallace, B.A. (2006). A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* **22**, 1955–1962.

Lefébure, T., Morvan, C., Malard, F., François, C., Konecny-Dupré, L., Guéguen, L., Weiss-Gayet, M., Seguin-Orlando, A., Ermini, L., Sarkissian, C. Der, Pierre Charrier, N., Eme, D., Mermillod-Blondin, F., Duret, L., Vieira, C., Orlando, L. & Douady, C.J. (2017). Less effective selection leads to larger genomes. *Genome Research* **27**, 1016–1028.

Leforestier, A. & Livolant, F. (1993). Supramolecular ordering of DNA in the cholesteric liquid crystalline phase: an ultrastructural study. *Biophysical journal* **65**, 56–72.

Le Treut, G., Képès, F. & Orland, H. (2016). Phase Behavior of DNA in the Presence of DNA-Binding Proteins. Biophysical Journal 110, 51–62.

Levi-Setti, R., Gavrilov, K.L. & Rizzo, P.J. (2008). Divalent cation distribution in dinoflagellate chromosomes imaged by high-resolution ion probe mass spectrometry. *European journal of cell biology* **87**, 963–976.

Li, J., Li, C., Xiao, W., Yuan, D., Wan, G. & Ma, L. (2008). Site-directed mutagenesis by combination of homologous recombination and DpnI digestion of the plasmid template in Escherichia coli. *Analytical biochemistry* **373**, 389–391.

Li, M.Z. & Elledge, S.J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature methods* **4**, 251–256.

Lidie, K.B., Ryan, J.C., Barbier, M. & Van Dolah, F.M. (2005). Gene expression in Florida red tide dinoflagellate Karenia brevis: analysis of an expressed sequence tag library and development of DNA microarray. *Marine biotechnology* **7**, 481–493.

Lin, S. (2006). the Smallest Dinoflagellate Genome Is Yet To Be Found: a Comment on Lajeunesse Et Al. "Symbiodinium (Pyrrhophyta) Genome Sizes (DNA Content) Are Smallest Among Dinoflagellates." *Journal of Phycology* **42**, 746–748.

Lin, S., Zhang, H., Zhuang, Y., Tran, B. & Gill, J. (2010). Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. *Proceedings of the National Academy of Sciences* **107**, 20033–8.

Linge, J.P., Habeck, M., Rieping, W. & Nilges, M. (2003). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**, 315–316.

Liu, H. & Naismith, J.H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC biotechnology* **8**, 91.

Liu, P., Kenney, J.M., Stiller, J.W. & Greenleaf, A.L. (2010). Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Molecular Biology and Evolution* **27**, 2628–2641.

Liu, X.L., Shen, Y., Chen, E.J. & Zhai, Z.H. (2000). Nuclear assembly of purified Crythecodinium cohnii chromosomes in cell-free extracts of Xenopus laevis eggs. *Cell research* **10**, 127–137.

Livolant, F. (1986). Cholesteric liquid crystalline phases given by three helical biological polymers : DNA, PBLG and xanthan. A comparative analysis of their textures. *Journal de Physique* **47**, 1605–1616.

Livolant, F. & Bouligand, Y. (1978). New observations on the twisted arrangement of dinoflagellate chromosomes. *Chromosoma* **68**, 21–44.

Livolant, F. & Bouligand, Y. (1980). Double helical arrangement of spread dinoflagellate chromosomes. *Chromosoma* **80**, 97–118.

Livolant, F. & Maestre, M.F. (1988). Circular Dichroism Microscopy of Compact Forms of DNA and Chromatin in Vivo and in Vitro: Cholesteric Liquid-Crystalline Phases of DNA and Single Dinoflagellate Nuclei. *Biochemistry* **27**, 3056–3068.

Lohuis, M.R. & Miller, D.J. (1998). Genetic transformation of dinoflagellates (Amphidinium and Symbiodinium): expression of GUS in microalgae using heterologous promoter constructs. *Transformation* **13**, 427–435.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260.

Lukeš, J., Leander, B.S. & Keeling, P.J. (2009). Cascades of convergent evolution: The corresponding evolutionary histories of euglenozoans and dinoflagellates. *Proceedings of the National Academy of Sciences* **106**, 9963 LP – 9970.

Machha, V.R., Waddle, J.R., Turner, A.L., Wellman, S., Le, V.H. & Lewis, E.A. (2013). Calorimetric studies of the interactions of linker histone H10 and its carboxyl (H10–C) and globular (H10–G) domains with calf-thymus DNA. *Biophysical Chemistry* **184**, 22–28.

Maeshima, K., Imai, R., Tamura, S. & Nozaki, T. (2014). Chromatin as dynamic 10-nm fibers. Chromosoma 123, 225–237.

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A. & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences* **113**, E1516–E1525.

Manosas, M., Spiering, M.M., Ding, F., Bensimon, D., Allemand, J.-F., Benkovic, S.J. & Croquette, V. (2012). Mechanism of strand displacement synthesis by DNA replicative polymerases. *Nucleic Acids Research* **40**, 6174–6186.

De March, M., Merino, N., Barrera-Vilarmau, S., Crehuet, R., Onesti, S., Blanco, F.J. & De Biasio, A. (2017). Structural basis of human PCNA sliding on DNA. *Nature Communications* **8**, 1–7.

Marie-Odile, S.-G. (2006). Edouard Chatton (1883-1947) and the dinoflagellate protists: concepts and models. *International Microbiology* **9**, 173–177.

Marinov, G.K. & Lynch, M. (2015). Diversity and Divergence of Dinoflagellate Histone Proteins. *G3: Genes|Genomes|Genetics* **6**, 397–422.

Marion, S., San Martín, C. & Šiber, A. (2017). Role of Condensing Particles in Polymer

Confinement: A Model for Virus-Packed "Minichromosomes." *Biophysical Journal* **113**, 1643–1653.

Mattirolli, F., Gu, Y. & Luger, K. (2017). Measuring Nucleosome Assembly Activity in vitro with the Nucleosome Assembly and Quantification (NAQ) Assay. Bio-Protocol 25, 1032–1057.

McEwan, M., Humayun, R., Slamovits, C.H. & Keeling, P.J. (2008). Nuclear genome sequence survey of the Dinoflagellate Heterocapsa triquetra. *The Journal of eukaryotic microbiology* **55**, 530–5.

Medina, R., Ghule, P.N., Cruzat, F., Barutcu, A.R., Montecino, M., Stein, J.L., van Wijnen, A.J. & Stein, G.S. (2012). Epigenetic control of cell cycle-dependent histone gene expression is a principal component of the abbreviated pluripotent cell cycle. *Molecular and cellular biology* **32**, 3860–3871.

Meinhart, A. & Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**, 223–226.

Mendez, G.S., Delwiche, C.F., Apt, K.E. & Lippmeier, J.C. (2015). Dinoflagellate gene structure and intron splice sites in a genomic tandem array. *Journal of Eukaryotic Microbiology* **62**, 679–687.

Meyer, S. & Beslon, G. (2014). Torsion-Mediated Interaction between Adjacent Genes. *PLOS Computational Biology* **10**, e1003785.

Mínguez, A., Franca, S. & Espina, D. (1994). Dinoflagellates have a eukaryotic nuclear matrix with lamin-like proteins and topoisomerase II. *Journal of cell science* **107**, 2861–2873.

Mochizuki, K. (2010). DNA rearrangements directed by non-coding RNAs in ciliates. *Wiley interdisciplinary reviews. RNA* **1**, 376–387.

Mudrak, O., Tomilin, N. & Zalensky, A. (2005). Chromosome architecture in the decondensing human sperm nucleus. *Journal of Cell Science* **118**, 4541–4550.

Noll, F., Lutsch, G. & Bielka, H. (1982). Structure of IgG and IgY molecules in ribosome-Antibody complexes as studied by electron microscopy. *Immunology Letters* **4**, 117–123.

Oakley, B.R. & Dodge, J.D. (1979). Evidence for a double-helically coiled toroidal chromonema in the dinoflagellate chromosome. *Chromosoma* **70**, 277–291.

Okur, H.I., Hladílková, J., Rembert, K.B., Cho, Y., Heyda, J., Dzubiella, J., Cremer, P.S. & Jungwirth, P. (2017). Beyond the Hofmeister Series: Ion-Specific Effects on Proteins and Their Biological Functions. *The Journal of Physical Chemistry B* **121**, 1997–2014.

Olivier-Van Stichelen, S. & Hanover, J.A. (2014). X-inactivation normalizes O-GlcNAc transferase levels and generates an O-GlcNAc-depleted Barr body. *Frontiers in genetics* **5**, 256.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. & Gronenborn, A.M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **261**, 438–446.

Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S. & Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**, 105–116.

Ortiz-Matamoros, M.F., Islas-Flores, T., Voigt, B., Menzel, D., Baluška, F. & Villanueva, M.A. (2015). Heterologous DNA Uptake in Cultured Symbiodinium spp. Aided by Agrobacterium tumefaciens. *PLOS ONE* **10**, e0132693.

Ortiz-matamoros, M.F., Villanueva, M.A. & Islas-flores, T. (2015). Transient transformation of cultured photosynthetic dinoflagellates (ymbiodinium spp.) with plant-targeted vectors. *Ciencias Marinas* **41**, 21–32.

Overton, I.M. & Barton, G.J. (2006). A normalised scale for structural genomics target ranking: the OB-Score. *FEBS letters* **580**, 4005–4009.

Overton, I.M., Padovani, G., Girolami, M.A. & Barton, G.J. (2008). ParCrys: a Parzen window density estimation approach to protein crystallization propensity prediction. *Bioinformatics* **24**, 901–907.

Papantonis, A. & Cook, P.R. (2011). Fixing the model for transcription: the DNA moves, not the polymerase. *Transcription* **2**, 41–44.

Patron, N.J. & Waller, R.F. (2007). Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *BioEssays* **29**, 1048–58.

Qi, D. & Scholthof, K.-B.G. (2008). A one-step PCR-based method for rapid and efficient site-directed fragment deletion, insertion, and substitution mutagenesis. *Journal of Virological Methods* **149**, 85–90.

Rae, P. & Steele, R. (1978). Modified bases in the DNAs of unicellular eukaryotes: an examination of distributions and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates. *Biosystems* **10**, 37–53.

Raspaud, E., Durand, D. & Livolant, F. (2005). Interhelical spacing in liquid crystalline spermine and spermidine-DNA precipitates. *Biophysical journal* **88**, 392–403.

Real, F., Vidal, R.O., Carazzolle, M.F., Mondego, J.M.C., Costa, G.G.L., Herai, R.H., Würtele, M., de Carvalho, L.M., Carmona e Ferreira, R., Mortara, R.A., Barbiéri, C.L., Mieczkowski, P., da Silveira, J.F., Briones, M.R. da S., Pereira, G.A.G. & Bahia, D.

(2013). The genome sequence of Leishmania (Leishmania) amazonensis: functional annotation and extended analysis of gene models. *DNA research : an international journal for rapid publication of reports on genes and genomes* **20**, 567–581.

Rédei, G.P. (Ed.). (2008). M9 Bacterial Minimal Medium BT - Encyclopedia of Genetics, Genomics, Proteomics and Informatics. In : 1135. Dordrecht: Springer Netherlands.

Reinberg, D. & Sims, R.J. 3rd. (2006). de FACTo nucleosome dynamics. *The Journal of biological chemistry* **281**, 23297–23301.

Reymer, A., Zakrzewska, K. & Lavery, R. (2017). Sequence-dependent response of DNA to torsional stress: a potential biological regulation mechanism. *Nucleic Acids Research* **46**, 1684–1694.

Rieder, D., Trajanoski, Z. & McNally, J.G. (2012). Transcription factories. *Frontiers in genetics* **3**, 221.

Rill, R.L., Livolant, F., Aldrich, H.C. & Davidson, M.W. (1989). Electron microscopy of liquid crystalline DNA: direct evidence for cholesteric-like organization of DNA in dinoflagellate chromosomes. *Chromosoma* **98**, 280–286.

Rizzo, P. & Noodén, L. (1973). Isolation and Chemical Composition of Dinoflagellate Nuclei. *Journal of Eukaryotic Microbiology* **20**, 666–672.

Rizzo, P.J. (1979). RNA synthesis in isolated nuclei of the dinoflagellate Crypthecodinium cohnii. *Journal of Protozoology* **26**, 290–294.

Rizzo, P.J. (2003). Those amazing dinoflagellate chromosomes. *Cell research* **13**, 215–7.

Rizzo, P.J. & Burghardt, R.C. (1980). Chromatin structure in the unicellular algae Olisthodiscus luteus, Crypthecodinium cohnii and Peridiniun balticum. *Chromosoma* **76**, 91–99.

Rizzo, P.J. & Nooden, L.D. (1972). Chromosomal proteins in the dinoflagellate alga Gyrodinium cohnii. *Science* **176**, 796–797.

Roberts, G.C.K. (Ed.). (1993). *NMR of Macromolecules. A Practical Approach*. Oxford University Press.

Rouviere-Yaniv, J. & Kjeldgaard, N.O. (1979). Native Escherichia coli HU protein is a heterotypic dimer. FEBS letters 106, 297–300.

Roy, S., Jagus, R. & Morse, D. (2018). Translation and Translational Control in Dinoflagellates. *Microorganisms* **6**, 30.

Roy, S. & Morse, D. (2012). A full suite of histone and histone modifying genes are transcribed in the dinoflagellate Lingulodinium. *PLoS ONE* **7**, e34340.

Roy, S. & Morse, D. (2013). Transcription and Maturation of mRNA in Dinoflagellates.

Microorganisms **1**, 71–99.

Safina, A., Cheney, P., Pal, M., Brodsky, L., Ivanov, A., Kirsanov, K., Lesovaya, E., Naberezhnov, D., Nesher, E., Koman, I., Wang, D., Wang, J., Yakubovskaya, M., Winkler, D. & Gurova, K. (2017). FACT is a sensor of DNA torsional stress in eukaryotic cells. *Nucleic Acids Research* **45**, 1925–1945.

Sala-Rovira, M., Geraud, M.L., Caput, D., Jacques, F., Soyer-Gobillard, M.-O., Vernet, G. & Herzog, M. (1991). Molecular cloning and immunolocalization of two variants of the major basic nuclear protein HCc from the histone-less eukaryote Crypthecodinium cohnii (Pyrrhophyta). *Chromosoma* **100**, 510–8.

Saldarriaga, J.F. (2003). Multiple protein phylogenies show that Oxyrrhis marina and Perkinsus marinus are early branches of the dinoflagellate lineage. *International Journal of Systematic and Evolutionary Microbiology* **53**, 355–365.

Saldarriaga, J.F., "Max" Taylor, F.J.R., Cavalier-Smith, T., Menden-Deuer, S. & Keeling, P.J. (2004). Molecular data and the evolutionary history of dinoflagellates. *European Journal of Protistology* **40**, 85–111.

Saldarriaga, J.F., McEwan, M.L., Fast, N.M., Taylor, F.J.R. & Keeling, P.J. (2003). Multiple protein phylogenies show that Oxyrrhis marina and Perkinsus marinus are early branches of the dinoflagellate lineage. *International Journal of Systematic and Evolutionary Microbiology* **53**, 355–365.

Sambrook, J. (2001). *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press.

Sambrook, J. & Russell, D.W. (2006). The Inoue Method for Preparation and Transformation of Competent E. Coli: "Ultra-Competent" Cells. *Cold Spring Harbor Protocols* **2006**.

Sandmann, A. & Sticht, H. (2018). Probing the role of intercalating protein sidechains for kink formation in DNA. *PLoS ONE* **13**, 1–21.

Sathyapriya, R. & Vishveshwara, S. (2004). Interaction of DNA with clusters of amino acids in proteins. *Nucleic Acids Research* **32**, 4109–4118.

Sawyer, W.H. & Puckridge, J. (1973). The Dissociation of Proteins by chaotropic salts. *The Journal of biological chemistry* **243**, 8429–8433.

Schwartz, M., Sarusi, A., Deitch, R.T., Tal, M., Raz, D., Sung, M.H., Kaplan, T. & Hakim, O. (2015). Comparative analysis of T4 DNA ligases and DNA polymerases used in chromosome conformation capture assays. *BioTechniques* **58**, 195–199.

Schwede, A., Kramer, S. & Carrington, M. (2012). How do trypanosomes change gene expression in response to the environment? Protoplasma 249, 223–238.

Serban, D., Arcineigas, S.F., Vorgias, C.E. & Thomas Jr., G.J. (2003). Structure and dynamics of the DNA-binding protein HU of B. stearothermophilus investigated by Raman and ultraviolet-resonance Raman spectroscopy. Protein Science 12, 861–870.

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H., Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., Toyoda, A., Kuroki, Y., Fujiyama, A., Medina, M., Coffroth, M., Bhattacharya, D. & Satoh, N. (2013). Draft Assembly of the Symbiodinium minutum Nuclear Genome Reveals Dinoflagellate Gene Structure. *Current Biology* **23**, 1399–1408.

Sibley, L.D. & Boothroyd, J.C. (1992). Construction of a molecular karyotype for Toxoplasma gondii. *Molecular and biochemical parasitology* **51**, 291–300.

Sigee, D.C. (1983). Structural DNA and genetically active DNA in dinoflagellate chromosomes. *BioSystems* **16**, 203–210.

Sikorav, J.L., Pelta, J. & Livolant, F. (1994). A liquid crystalline phase in spermidine-condensed DNA. *Biophysical journal* **67**, 1387–1392.

Skuridin, S.G., Vereshchagin, F. V., Salyanov, V.I., Chulkov, D.P., Kompanets, O.N. & Yevdokimov, Y.M. (2016). Ordering of double-stranded DNA molecules in a cholesteric liquid-crystalline phase and in dispersion particles of this phase. *Molecular Biology* **50**, 783–790.

Soyer-Gobillard, M.-O. & Herzog, M. (1985). The native structure of dinoflagellate chromosomes: involvement of structural RNA. *European journal of cell biology*.

Soyer, M.-O. (1971). Structure du noyau des Blastodinium Dinoflagellés parasites. *Chromosoma* **33**, 70–114.

Soyer, M.-O. (1977). A modification of the Karnovsky fixation for better preservation of dinoflagellate nuclear structures. *Biologie Cellulaire*, Biology of the cell 297–300.

Soyer, M.-O. & Haapala, O.K. (1974). Electron microscopy of RNA in Dinoflagellate chromosomes. *Histochemistry* **42**, 239–246.

Spector, D.L. (1984). Chapter 4 - Dinoflagellate Nuclei. In *Cell Biology*: 107–147. San Diego: Academic Press.

Spector, D.L., Vasconcelos, A.C. & Triemer, R.E. (1981). DNA duplication and chromosome structure in the dinoflagellates. *Protoplasma* **105**, 185–194.

Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G.R., Korf, I., Lapp, H., Lehväslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E.,

Wilkinson, M.D. & Birney, E. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome research* **12**, 1611–1618.

Stocki, S. a, Nonay, R.L. & Reha-Krantz, L.J. (1995). Dynamics of bacteriophage T4 DNA polymerase function: identification of amino acid residues that affect switching between polymerase and 3' --> 5' exonuclease activities. *Journal of molecular biology* **254**, 15–28.

Stojkova, P., Spidlova, P. & Stulik, J. (2019). Nucleoid-associated protein HU: A lilliputian in gene regulation of bacterial virulence. Frontiers in Cellular and Infection Microbiology 9, 1–7.

Strick, R., Strissel, P.L., Gavrilov, K. & R, L.-S. (2001). Cation-chromatin binding as shown by ion microscopy is essential for the structural integrity of chromosomes. *Journal of cell biology* **155**, 899–910.

Strzelecka, T.E., Davidson, M.W. & Rill, R.L. (1988). Multiple liquid crystal phases of DNA at high concentrations. *Nature* **331**, 457–460.

Sun, S., Wong, J.T.Y., Dong, F. & Liu, M. (2012). Histone-like protein HCcp3-induced liquid crystalline DNA condensation. Chemistry Letters **41,** 874-876.

Sun, S., Liu, M., Dong, F., Fan, S. & Yao, Y. (2013). A histone-like protein induces plasmid DNA to form liquid crystals in vitro and gene compaction in vivo. *International Journal of Molecular Sciences* **14**, 23842–23857.

Sun, S., Wong, J.T.Y., Liu, M. & Dong, F. (2012). Counterion-Mediated Decompaction of Liquid Crystalline Chromosomes. *DNA and Cell Biology* **31**, 1657–1664.

Sutherland, H. & Bickmore, W. (2009). Transcription factories: gene expression in unions? *Nature reviews Genetics* **10**, 457–66.

Swinger, K.K., Lemberg, K.M., Zhang, Y. & Rice, P.A. (2003). Flexible DNA bending in HU–DNA cocrystal structures. The EMBO Journal 22, 3749–3760.

Swinger, K.K. & Rice, P.A. (2004). IHF and HU: flexible architects of bent DNA. Current opinion in structural biology 14, 28–35.

Tadeo, X., López-Méndez, B., Castaño, D., Trigueros, T. & Millet, O. (2009). Protein Stabilization and the Hofmeister Effect: The Role of Hydrophobic Solvation. *Biophysical Journal* **97**, 2595–2603.

Takahashi, M., Yoshikawa, K., Vasilevskaya, V. V. & Khokhlov, A.R. (1997). Discrete Coil−Globule Transition of Single Duplex DNAs Induced by Polyamines. *The Journal of Physical Chemistry B* **101**, 9396–9401.

Talbert, P.B. & Henikoff, S. (2010). Histone variants—ancient wrap artists of the epigenome.

Nature reviews Molecular cell biology **11**, 264–75.

Talbert, P.B. & Henikoff, S. (2012). Chromatin: packaging without nucleosomes. *Current Biology* **22**, R1040.

Tautz, D. (2008). Polycistronic peptide coding genes in eukaryotes—how widespread are they? Briefings in Functional Genomics 8, 68–74.

Teng, C.Y., Dang, Y., Danne, J.C., Waller, R.F. & Green, B.R. (2013). Mitochondrial genes of dinoflagellates are transcribed by a nuclear-encoded single-subunit RNA polymerase. *PloS one* **8**, e65387.

Teves, S.S. & Henikoff, S. (2014). DNA torsion as a feedback mediator of transcription and chromatin dynamics. *Nucleus* **5**, 211–218.

Thomas, J.O. & Kornberg, R.D. (1978). The study of histone--histone associations by chemical cross-linking. *Methods in cell biology* **18**, 429–440.

Thomas, J.O. & Stott, K. (2012). H1 and HMGB1: modulators of chromatin structure. *Biochemical Society Transactions* **40**, 341 LP – 346.

Tippit, D.H. & Pickett-Heaps, J.D. (1976). Apparent amitosis in the binucleate dinoflagellate Peridinium balticum. *Journal of cell science* **21**, 273–289.

Tropea, J.E., Cherry, S. & Waugh, D.S. (2009). Chapter 19: Expression and purification of soluble His 6-tagged TEV protease. In *Methods in Molecular Biology: High Throughput Protein Expression and Purification*. Humana Press.

Turner, A.L., Watson, M., Wilkins, O.G., Cato, L., Travers, A., Thomas, J.O. & Stott, K. (2018). Highly disordered histone H1−DNA model complexes and their condensates. *Proceedings of the National Academy of Sciences* **115**, 11964 LP – 11969.

Valach, M., Moreira, S., Kiethega, G.N. & Burger, G. (2014). Trans-splicing and RNA editing of LSU rRNA in Diplonema mitochondria. *Nucleic acids research* **42**, 2660–2672.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horak, A., Jaillon, O., Lima-Mendez, G., Luke, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E., Boss, E., Follows, M., Karp-Boss, L., Krzic, U., Reynaud, E.G., Sardet, C., Sullivan, M.B. &

Velayoudon, D. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605–1261605.

Venkatesh, S. & Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nature reviews Molecular cell biology* **16**, 178–189.

Vernet, G., Sala-Rovira, M., Maeder, M., Jacques, F. & Herzog, M. (1990). Basic nuclear proteins of the histone-less eukaryote Crypthecodinium cohnii (Pyrrhophyta): two-dimensional electrophoresis and DNA-binding properties. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1048**, 281–289.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. & Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696.

Waller, R.F., Cleves, P.A., Rubio-brotons, M., Woods, A., Bender, S.J., Edgcomb, V., Gann, E.R., Jones, A.C., Teytelman, L., Von, P., Wilhelm, S.W. & Collier, J.L. (2018). Strength in numbers : Collaborative science for new experimental model systems. *PLOS Biology* 1–10.

Walz, B., Walz, A. & Sweeney, B.M. (1983). A circadian rhythm in RNA in the dinoflagellate, Gonyaulax polyedra. Journal of comparative physiology.

Wang, Y., Schellenberg, H., Walhorn, V., Toensing, K. & Anselmetti, D. (2017). Binding mechanism of PicoGreen to DNA characterized by magnetic tweezers and fluorescence spectroscopy. *European Biophysics Journal* **46**, 561–566.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. & Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191.

Wegel, E. & Shaw, P. (2005). Gene activation and deactivation related changes in the three-dimensional structure of chromatin. *Chromosoma* **114**, 331–337.

White, S.W., Wilson, K.S., Appelt, K. & Tanaka, I. (1999). The high-resolution structure of DNA-binding protein HU from {\it Bacillus stearothermophilus}. Acta Crystallographica Section D 55, 801–809.

Wilson, K.A., Kellie, J.L. & Wetmore, S.D. (2014). DNA-protein π-interactions in nature: Abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. *Nucleic Acids Research* **42**, 6726–6741.

Wolsey, W.C. (1973). Perchlorate salts, their uses and alternatives. *Journal of Chemical Education* **50**, A335.

Workman, J.L. (2006). Nucleosome displacement in transcription. *Genes and Development*.

Wu, C. & Travers, A. (2019). Modelling and DNA topology of compact 2-start and 1-start chromatin fibres. *Nucleic Acids Research* 1–23.

Wu, D., Guo, X., Lu, J., Sun, X., Li, F., Chen, Y. & Xiao, D. (2013). A rapid and efficient one-step site-directed deletion, insertion, and substitution mutagenesis protocol. *Analytical biochemistry* **434**, 254–258.

Wulf, E., Bautz, F.A., Faulstich, H. & Wieland, T. (1980). Distribution of fluorescent α-amanitin (FAMA) during mitosis in cultured rat kangaroo (PtK1) cells. *Experimental cell research* **130**, 2–7.

Xia, Y., Chu, W., Qi, Q. & Xun, L. (2014). New insights into the QuikChange™ process guide the use of Phusion DNA polymerase for site-directed mutagenesis. *Nucleic Acids Research* 1–9.

Xia, Y. & Xun, L. (2017). Revised mechanism and improved efficiency of the quikchange site-directed mutagenesis method. In *Methods in Molecular Biology*: 367–374.

Yee, B., Sagulenko, E., Morgan, G.P., Webb, R.I. & Fuerst, J.A. (2012). Electron tomography of the nucleoid of Gemmata obscuriglobus reveals complex liquid crystalline cholesteric structure. *Frontiers in Microbiology* **3**, 1–8.

Yutin, N., Colson, P., Raoult, D. & Koonin, E. V. (2013). Mimiviridae: Clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. *Virology Journal* **10**, 1–13.

Yutin, N., Wolf, Y.I. & Koonin, E. V. (2014). Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* **466–467**, 38–52.

Zakryś, B. (1988). The nuclear DNA level as a potential taxonomic character in Euglena Ehr. (Euglenophyceae). *Archiv für Hydrobiologie* **49**, 483–504.

Zhang, H., Bhattacharya, D. & Lin, S. (2007a). A Three-Gene Dinoflagellate Phylogeny Suggests Monophyly of Prorocentrales and a Basal Position for Amphidinium and Heterocapsa. *Journal of Molecular Evolution* **65**, 463–474.

Zhang, H., Dungan, C.F. & Lin, S. (2011). Introns, alternative splicing, spliced leader trans-splicing and differential expression of pcna and cyclin in Perkinsus marinus. *Protist* **162**, 154–167.

Zhang, H., Hou, Y., Miranda, L., Campbell, D.A., Sturm, N.R., Gaasterland, T. & Lin, S. (2007b). Spliced leader RNA trans-splicing in dinoflagellates. *Proceedings of the National Academy of Sciences* **104**, 4618–4623.

Zhang, W., Zhou, J., Liu, T., Yu, Y., Pan, Y., Yan, S. & Wang, Y. (2015). Four novel algal

virus genomes discovered from Yellowstone Lake metagenomes. *Scientific Reports* **5**, 1–13.

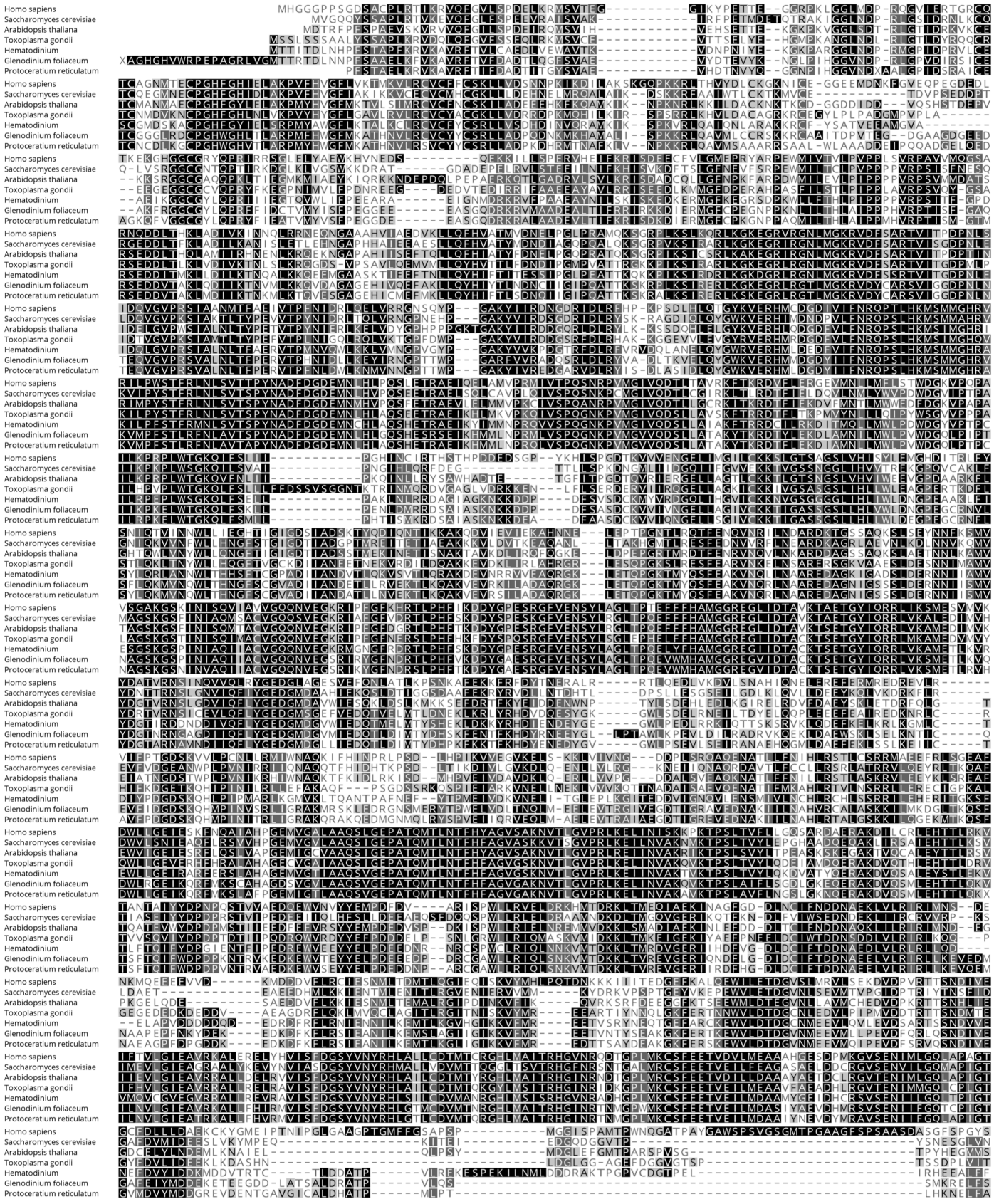
Zhang, X., Wen, H. & Shi, X. (2012). Lysine methylation: Beyond histones. *Acta Biochimica et Biophysica Sinica* **44**, 14–27.

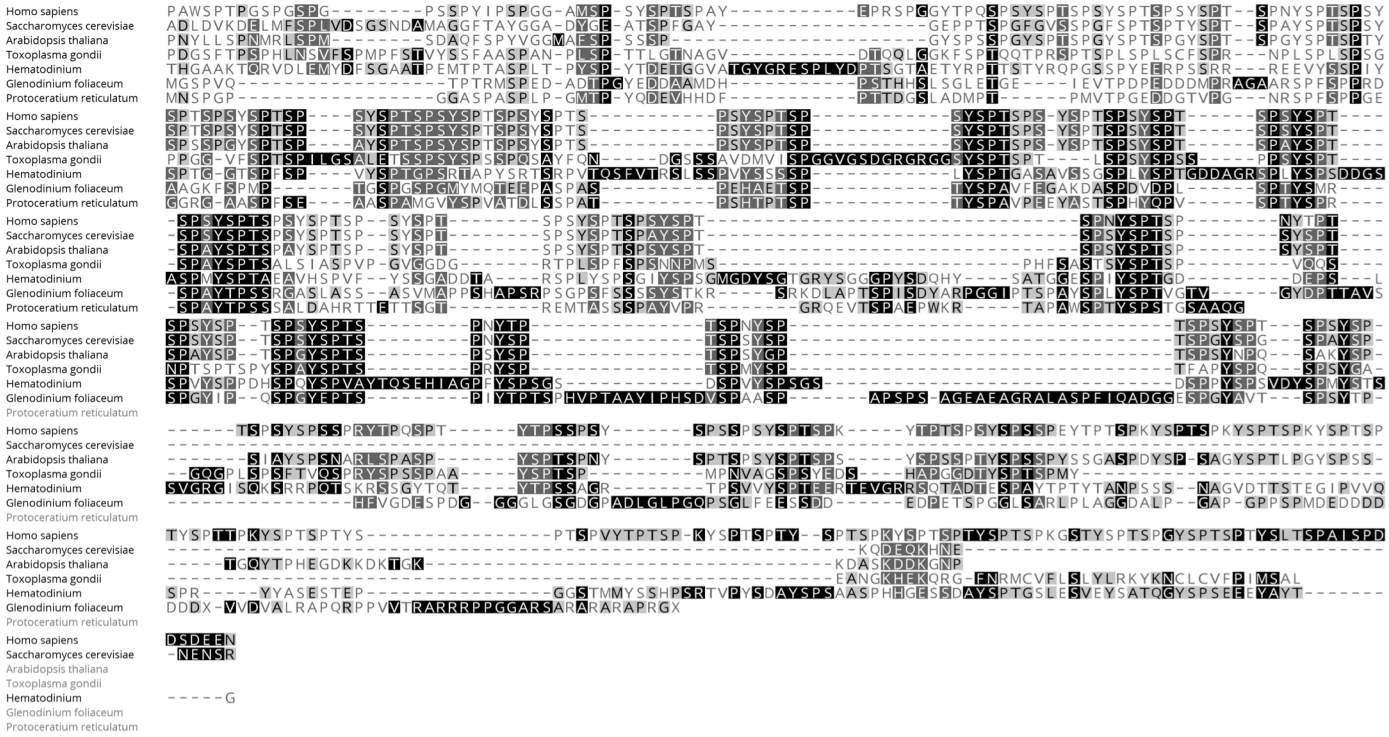
Zhang, Y., Werling, U. & Edelmann, W. (2014). Seamless Ligation Cloning Extract (SLiCE) cloning method. *Methods in molecular biology* **1116**, 235–244.

Zheng, L., Baumann, U. & Reymond, J.-L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research* **32**, e115–e115.

Appendices

Appendix 1: Sequences of RNA polymerase II from 7 species, *Homo sapiens, Saccharomyces cerevisiae, Arabidopsis Thaliana, Toxoplasma gondii, Hematodinium sp., Glenodinium foliaceum, and Protoceratium reticulatum.*




Homo sapiens
Saccharomyces cerevisiae
Arabidopsis thaliana
Toxoplasma gondii
Hematodinium
Glenodinium foliaceum
Protoceratium reticulatum

Appendix 2: Multiple alignment of dinoflagellate DVNP sequences

>Azadinium-spinosum-3D9-20130829|17390_1/32-135 [subseq from] wE3qSTiHAwaONzs5nipTsrSUatw Azadinium spinosum 3D9
AGAMKGKRVSKIATGKRARAAVFSGRKEKTASGMTKSKLMKNSRGKIVSKARAAVAKKNFAKLGLWTSAVTA
ARKALNIRGFVAINGKSAEGKAIYAKAKSLYK
>MMETSP0661-20131031|7467_1/151-253 [subseq from] ZcG9a7GXqzNknK64EYQWj5AAtik Alexandrium margalefi, Strain AMGDE01CS-322
-MKVKAARVSDIARGKRARQMVWEGRKRKTATGMTKDRLMKNPRGKIVAKRRSAAIRSRYARLGIWTEAVMA
ARKALGVVGFVALNGKTVHGKAVYVKARAIYT
>Kryptoperidinium-foliaceum-CCMP1326-20130916|222598_1/17-120 [subseq from] t2arFKVtvKjb2ewM7QONONkWZNg Kryptoperidinium foliaceum CCMP1326
MRSKRAGKSKKIARGRVAKAMVLRGSREKTSGGLKSGDLMKNKRGKVVSKRQSAQGRRAYSLIEGWTDSVME
ARRALHISGFVAINGKTLQGKALYVKAKALYT
>Glenodinium-foliaceum-CCAP1116_3-20130913|263558_1/31-133 [subseq from] ICq7npAlhLdQyQt+PDsgsKoeRwE Kryptoperidinium foliaceum
NRGPKSKKGRGLASGRLAKAMVLRGSRVKTTGGLKSEDLMRNKRGRVVSKRKSANGRRAYAQIEAWTDSVVE
ARRALHISGFVAINGKTLQGKALYVKARSLY-
>Glenodinium-foliaceum-CCAP1116_3-20130913|40125_1/55-157 [subseq from] oYHTBo6fqsO1+8ZO49sLlvQXEMI Kryptoperidinium foliaceum
KRGAKGKGKSKIARGRLAKAMVLRGSKVKTSGGLKSEDLMRNKRGKIVSKRKSANGSRAYKFIEGWMDSVVE
ARRALHISGFVAINGKTLQGKALYVKARSLF-
>Karlodinium-micrum-CCMP2283-20140214|36473_1/41-143 [subseq from] BJxTYMkngg1LsjsgahYkb26g0N4 Karlodinium micrum, Strain CCMP2283
QNKPKKSRSSKVATGRFAKAMVLRGSREKTSGGLRKEDLMRNKRGKVVSKRASAHGKRAFKRVEAWIDSVMQ
AREALHTRGFVAINGKTLHGKALYHKAKTLH-
>Karlodinium-micrum-CCMP2283-20140214|199510_1/10-110 [subseq from] dqt7Lm8B6cQ1WTiu+2MnsSK9DNw Karlodinium micrum, Strain CCMP2283
--TKAVSKKSIIARGSRAKAMVLRGSKARTSGGLKAEDLMKNKRGKVVSKKASAHGMRAFKRIEGWVDVVME
ARAALHSRGFVAINGKTLQGKALYFKAKQLY-
>Alexandrium-monilatum-CCMP3105-20140214|47100_1/18-121 [subseq from] 8VOx6155l9DVICtrqDCyHh1iZRg Alexandrium monilatum, Strain CCMP3105
AKATKAKRVSKVARGRFAKALVLRGRRERTVGGVRKDGLMRNKRGRIVSKRASACAMKKYCYLEDWTQALMQ
ARQALHCKGFVAVNGKSLHGKALYLKAKAIHQ
>MMETSP0126-20121128|16784_1/20-122 [subseq from] A/25v8s8bktRLR2fFbFkXWA2oaU Strombidinopsis acuminata
GKAMKAKRVSKIAKGKLAKAVVLRGSKEKTSGGLTKDMLFKNKRGKVVSKKASAVAKKRFSTFKKWTESIVS
ARKALNVTGFVAVNGKTAQGKAIYAKAKAMY-
>ht00362-ht/1-93 [subseq from] zJPoagQVG9kSQ1nao6nziOByC9c Heterocapsa triquetra
----------KIAKGKFAKAVVLRGSKEKTSGGLTKDQLFKNKRGKVVSKKASAVAKKRFSSMKKWTESIQS
ARKALNVSGFVAINGKTSQGKALYAKAKALY-
>ht01357-ht/9-111 [subseq from] HQsDjewf6iucXQP3AWKoc9e6Ee4 Heterocapsa triquetra
MKAMKAKKVSKVAKGKFAKAVVMRGSKAKTSGGLTKDQLFKNKRGKIVSKKASAAGKKRFAAIKGWFDAVKS
ARKALNVSGFVAINGKTSQGKALYAKAKAMY-
>MMETSP0448-20130528|5947_1/1-97 [subseq from] ZsfKDMmQbK4dUvlC7+Ub6HoRJcM Heterocapsa triquetra
------KKVSKVAKGKFAKAVVMRGSKAKTSGGLTKDQLFKNKRGKIVSKKASAAGKKRFAAIKGWFEAVKS
ARKALNVSGFVAINGKTSQGKALYAKAKAMY-
>ht01122-ht/2-101 [subseq from] 9KV8kyP46OWihPdr7zCabqPGkcc Heterocapsa triquetra
---MKAKKVSKVAKGKFAKAVVLRGSKAKTSGGLTKDQLFKNKRGKVVSKKASAAGKKRFAAIKGWSQAVQS
ARKALNVSGFVAINGKTSQGKALYAKAKALY-
>MMETSP0126-20121128|27282_1/18-112 [subseq from] 4eSd9kbzfuZq9nUYSlbcB0Mlkgw Strombidinopsis acuminata
MKAMKAKKVSKVAKGKLAKAIVLRGSKAKTSGGLTKDQLFKNKRGKVVSKKASAAGKKRFAAIKGWSQAVQS
ARKALNVSGFVAINGKTSQGKAL---------
>MMETSP1462-20131121|155014_1/18-121 [subseq from] FX3cE1mB9HhGaJaPTqwKhnpBkBQ Brandtodinium nutriculum, Strain RCC3387
ARAMKSKRPSKIAKGRMAKVMVFRGSKEKTASGLKADALMKNKRNRVVSKRRSAHGKRQFKNVESWVKSVTE
ARSALHLSGFVAINGKTLQGKALYVKARALWK
>MMETSP1462-20131121|65783_1/46-148 [subseq from] PyZ4eTF9QR8hIpD84gmWU5/SykE Brandtodinium nutriculum, Strain RCC3387
ARAMKAKRESKIARGRMAKAMVLRGSKEKTVGGLRADGLMRNKANKVVSKRASAHGKRQFKNIESWLESLTE
ARSALHVSGFVAVNGKTLQGKALYVKAKALW-
>MMETSP1462-20131121|11579_1/50-152 [subseq from] lEAYYXywV9KCYo30l4gK6M9rEO8 Brandtodinium nutriculum, Strain RCC3387
AGAMKAKRASKVAKGRMAKAMVLRGTKEKTVGGLKAEALMRNRRNKVVSKRASAHGKRQFKNVEAWVESVVE
ARTALHVSGFVAINGKSLQGKALYVKAKALW-
>MMETSP1462-20131121|155014_1/174-275 [subseq from] FX3cE1mB9HhGaJaPTqwKhnpBkBQ Brandtodinium nutriculum, Strain RCC3387
AGGSKTKRASKIARGRMAKAMVLRGSKEKTVGGLKADALMRNKRNKVVSKRASAHGKRQFKNIESWVDSVVE
ARAALHVSGFVAVNGKTLQGKALYVKAKAL--
>MMETSP1462-20131121|36613_1/42-143 [subseq from] 9sJhvSP0rK9ZuB6WcSV+U6Ht+So Brandtodinium nutriculum, Strain RCC3387
AGVKRKQRASKIARGRMAKVMVLRGSKEKTVGGLKADALMKNKRNKVVSKRASAHGKRQFKNIESWVTSVVE
ARAALHVSGFVAVNGKTLQGKALYVKAKAL--
>Peridinium-aciculiferum-PAER_2-20130926|168390_1/49-150 [subseq from] jPlaqYy+7PGgFN2DMoyjruw6vQ8 Peridinium aciculiferum PAER_2
GGAMKSKRASKIARGRMAKALVLRGSKEKTVGGLTKDALMRNKRGKIVSKRSSAAGKRSYKQIEAWVDSVVE
ARKALHISGFVAINGKTLQGKALYVKSKAL--

>Scrippsiella-hangoei_like-SHHI_4-20140214|254654_1/16-117 [subseq from] jDY5UblhbSdtJWuwMYrrApj9XcY Scrippsiella hangoei-like, Strain SHHI-4
GGAMKSKRTSKIARGRMAKALVLRGSKEKTVGGLTKDALMRNKRGKIVSKRSSAAGKRSYKQIEAWVDSVME
ARKALHISGFVAINGKTLQGKALYVKSKAL--
>Scrippsiella-hangoei-SHTV5-20131105|36646_1/49-150 [subseq from] H6Z4saHm3GyNtkauZJgvijB3v0c Scrippsiella hangoei SHTV5
GGAMKSKRTSKIALGRMAKALVLRGSKEKTVGGLTKDALMRNKRGKIVSKRSSAAGKRSYKQIEAWVDSVME
ARKALHISGFVAINGKTLHGKALYVKSKAL--
>Scrippsiella-hangoei-SHTV5-20131105|127602_1/201-302 [subseq from] pcVGb8U15YXQC/g8LvhGNCAIKWM Scrippsiella hangoei SHTV5
MKVMKVKRVSKIAKGRMAKALVFRGSKEKTVGGLKQDALMKNKRGKIVSKRGSAAAKRKYTKIEEWVECLVE
ARKALHITGFVAVNGRTLQGKALYVKTKAL--
>Scrippsiella-hangoei_like-SHHI_4-20140214|11055_1/59-160 [subseq from] xUtFmNR3pPOx3XGFa+HTHEsw51A Scrippsiella hangoei-like, Strain SHHI-4
MKVMKVKRVSKIAKGRMAKALVFRGSKEKTVGGLKQDALMKNKRGKIVSKRGAAAGKRKYTKIEEWVECLVE
ARKALHITGFVAVNGRTLQGKALYVKTKAL--
>Peridinium-aciculiferum-PAER_2-20130926|112890_1/55-156 [subseq from] pfGij4jFo0UvwtAcJgRoxIWtJOE Peridinium aciculiferum PAER_2
MKVMKVKRVSKIAKGRMAKALVFRGSKEKTVGGLKQDALMKNKRGKIVSKRGAAAGKRKYSKIEEWVECLVE
ARKALHITGFVAVNGRTLQGKALYVKTKAL--
>MMETSP1098-20130426|7337_1/19-120 [subseq from] Vm+VguSFR4CWOxd3+VgvFviyFzI Spumella elongata, Strain CCAP 955
MKVMKVKRVSKIAKGRMAKALVFRGSKEKTVGGLKQDALMTNKRGKIVSKRGSASGKRKYTKIEEWVECLVE
ARKALHITGFVAVNGRTLQGKALYVKTKAL--
>MMETSP1098-20130426|7102_1/45-146 [subseq from] Yp18FR/+0gNEk2OADM+PuDrNZOc Spumella elongata, Strain CCAP 955
GGAMKSKRTSKIALGRMAKALVLRGSKEKTVGGLTKDALMRNKRGKIVSKRGSASGKRKYTKIEEWVECLVE
ARKALHITGFVAVNGRTLQGKALYVKTKAL--
>Scrippsiella-trochoidea-CCMP3099-20130930|8520_1/39-140 [subseq from] LnqJTgs4vL3d8HSQK0yXvXTc/7s Scrippsiella trochoidea CCMP3099
APAMKAVRASKVARGRMAKAQVFKGRKEKTSGGLKQDSLMQNKRGKIVSKRLSALGKRRYANVEAWVDSVME
ARRALHISGFVAINGKTLQGKALYVKTRAL--
>Scrippsiella-hangoei_like-SHHI_4-20140214|27363_1/14-115 [subseq from] lYbECBG+po8aPVb7N8lXkVRBgiM Scrippsiella hangoei-like, Strain SHHI-4
MKAMKAKRVSKIAKGRMAKALVLRGSKEKTVGGLRKDSLMKNARGRVVSKRASANGKRRYKNIETWIESVLE
ARKALHVSGFVAINGRTLQGKALYVKSKAI--
>Scrippsiella-hangoei-SHTV5-20131105|127814_1/14-115 [subseq from] H7dcPkj2SLYtlSfVE8DaBUXxIgw Scrippsiella hangoei SHTV5
MKAMKAKRVSKIAKGRMAKALVLRGSKEKTVGGLRKDSLMKNARGRVVSKRASANGKRRYKNIETWIESVLE
ARKALHVSGFVAINGKTLQGKALYVKSKAI--
>Peridinium-aciculiferum-PAER_2-20130926|40376_1/14-115 [subseq from] y7cRYkf1iTwbl9HYCC9GcY0Xc0s Peridinium aciculiferum PAER_2
MKAMKAKRVSKIAKGRMAKALVLRGSKEKTVGGLRKDSLMKNARGRVVSKRASANGKRRYKNIETWIESVLE
ARKALHVSGFVAINGKTLQGKALYVKTKAI--
>Scrippsiella-hangoei_like-SHHI_4-20140214|25550_1/14-115 [subseq from] e1Zywwo6aUS8qp1jADRtsR4OGVM Scrippsiella hangoei-like, Strain SHHI-4
MKAMKAKRVSKIAKGRMAKALVLRGSKEKTVGGLRKDSLMKNARGRVVSKRASANGKRRYKNIETWIESVLA
ARKALHVSGFVAINGKTLQGKALYVKSKAI--
>Scrippsiella-hangoei-SHTV5-20131105|13755_1/14-115 [subseq from] gEmfOUJYIOSMoL+BS2chzvawsSM Scrippsiella hangoei SHTV5
MKAMKAKRVSKIAKGRMAKALVLRGSKEKTVGGLLKDSLVKNARGRVVSKRASANGKRRYKNIETWIESVLA
ARKALHVSGFVAINGKTLQGKALYVKSKAI--
>Scrippsiella-hangoei-SHTV5-20131105|5254_1/51-152 [subseq from] jyzj3k1F88KdddKHqGDBNiWVmbs Scrippsiella hangoei SHTV5
MKATSAKRVSKIAKGRMAKALVLRGSKEKTVGGLRKDSLMKNARGRVVSKRASAHGKRRYKNIETWVESVLE
ARKALHVSGFVAINGRTLQGKALYVKSKAI--
>Scrippsiella-hangoei-SHTV5-20131105|127814_1/297-398 [subseq from] H7dcPkj2SLYtlSfVE8DaBUXxIgw Scrippsiella hangoei SHTV5
TKAMKVKRASKIAKGRMAKALVLRGSKEKTVGGLRKDSLMKNARGRVVSKRASANGKRRYKNIETWVESVLA
ARKALHVSGFVAINRRTLQGKALYVKSKAI--
>Scrippsiella-trochoidea-CCMP3099-20130930|11141_1/23-124 [subseq from] uwP2nYYy4P3Oz4WtUCFCfKZF97g Scrippsiella trochoidea CCMP3099
ATAMKAKRATKVAKGRMAKALVFRGSKEKTTGGLKQESLMKNKRGKIVSKRQSAIGKRRYNNVEAWVDSVME
ARKALHISGFVAINGKTLQGKALYVKTKSL--
>Scrippsiella-trochoidea-CCMP3099-20130930|27374_1/10-111 [subseq from] xJA1bO9rXTouwhPq9NQFJ0oK8VQ Scrippsiella trochoidea CCMP3099
AAPMKAKRATKVAKGRMAKAQVFTGRKEKTVGGLKQDSLMKNKRGKIVSKRQSAQGKRKYSNVETWVDSVME
ARRALHISGFVAINGKTLQGKALYVKTRAL--
>Scrippsiella-trochoidea-CCMP3099-20130930|396511_1/165-266 [subseq from] 6ma0xsRcq7y6U4gllt0squbfVaU Scrippsiella trochoidea CCMP3099
ATPMKAKRATKVAKGRMAKAQVFRGRKEKTVGGLKQESLMKNKRGKIVSKRQSALGKRKFSNVEAWVDSVME
ARQALHISGFVAINGKTLQGKALYVKTRAL--
>Scrippsiella-trochoidea-CCMP3099-20130930|11921_1/21-122 [subseq from] NX62NZlFjf3clystrAyEQHtHMaU Scrippsiella trochoidea CCMP3099
AAPMKAKRATKVAKGRMAKAQVFRGRKEKTAGGLKQESLMKNKRGKIVSKRQSALGKRKFSNVEAWVDSVME
ARQALHISGFVAINGKTLQGKALYVKTRAL--
>Scrippsiella-trochoidea-CCMP3099-20130930|398462_1/238-339 [subseq from] kMhNb2enynhULlxDwGP4VYQWYmc Scrippsiella trochoidea CCMP3099
MKAMKTKRASKVARGRMAKAMVFRGSKEKTVGGLKQESLMKNKRGKIVSKRQSAQGKRRYANVEGWVDSVMA
ARKALHISGFVAINGKTLQGKALYVKSKAL--

>Scrippsiella-hangoei-SHTV5-20131105|63059_1/41-142 [subseq from] ymQrFDdxgs6PSoc1/0OIxaZ2POI Scrippsiella hangoei SHTV5
MKAMKAKRVSKVAKGRMAKAMVFRGSKEQTVGGLKQEALMKNKRGKIVSKRASAAGKRRFRNIETWLDSVME
ARKALHVSGFVAINGKTLQGKALYVKSKAI--
>Scrippsiella-hangoei_like-SHHI_4-20140214|35053_1/45-146 [subseq from] 6pJ57H6q/RGaGFHbGr1oieh08vk Scrippsiella hangoei-like, Strain SHHI-4
LKAMKAKRVSKVAKGRMAKAMVFRGSKEQTVGGLKQEALMKNKRGKIVSKRASAAGKRRFRNIESWLDSVME
ARKALHVSGFVAINGKTLQGKALYVKSKAI--
>Peridinium-aciculiferum-PAER_2-20130926|179373_1/30-121 [subseq from] TQKrkRnRSxipPruaKOJOJJiBPGs Peridinium aciculiferum PAER_2
MKAMKAKRVTKVAKGRMAKAMVFRGSKEKTVGGLKQEALMKNKRGKVVSKRASANGKRRFRNIETWLDSVME
ARKALHVSGFVAINGKTLQG----------
>Peridinium-aciculiferum-PAER_2-20130926|113902_1/50-151 [subseq from] 5LJF0JP9DlpPADnLRnZLUYFpRas Peridinium aciculiferum PAER_2
MKSIKSKRVPKVAKGRMAKAMVFRGSREKTVGGLKQEALMKNKRGKVVSKRASANGKRRFRNIELWLESVME
ARTALHFSGFVAINGKTLQGRALYVKSKAL--
>Durinskia-baltica-CSIRO_CS-38-20140214|10169_1/19-120 [subseq from] PmXNKiXcty0njsbuofGXJwq5yPI Durinskia baltica, Strain CSIRO CS-38
KKVMKAKRVSKVARGRMARYMVLKGSREKTVGGLTQEALMKNKRGKVVSKRASANGKMRFKNIEGWLDSVME
ARKALHVSGFVAINGKTLQGKALYVKSKAL--
>MMETSP1462-20131121|54632_1/33-134 [subseq from] hkRZpzbN27ZkFEbNYb7VIAh0L7k Brandtodinium nutriculum, Strain RCC3387
RGAMKAKRVTKVARGRLAKAMVFRGSKEKTVGGVKKEGLMKNKRDKVVSKRQSAAGTRAYRHIEAWVDSVVE
ARKALHISGFVAINGKSLQGKALYVKTKAI--
>Karenia-brevis-SP3-20130916|255427_1/205-306 [subseq from] CGn0NGRkG8o8ZVam3Jxi9e0axHk Karenia brevis SP3
KKVKKAKRVSQIAKGRMARAMVLRGSKEKTVGGLKKDMLMRNKRGKVVSKRAAAHGRRVYRNVEDWVEAVME
AREALHARGFVAINGKTLQGKALYVKAKAL--
>Karenia-brevis-CCMP2229-20130916|271706_1/4-105 [subseq from] nfYuHaYzRwqQ9pBK1CNfOCsOleE Karenia brevis CCMP2229
MKVMKAKRVSKIAQGKMAKAMVLRGSKEKTVGGLKKDMLMRNKRGKVVSKRAAAHGRRVYRNVEDWVEAVME
AREALHARGFVAINGKTLQGKALYVKAKAL--
>Karenia-brevis-Wilson-20130916|58292_1/3-104 [subseq from] WHFI1crbV78C1ORiEQ28w/dOQ/Y Karenia brevis Wilson
MKVMKAKRVSKIAQGRMAKAMVLRGSKEKTVGGLKKDSLMKNKRGKVVSKRAAAHGRRAYRNVEDWIEAVME
AREALHARGFVAINGRTLQGKALYVKAKAL--
>Karenia-brevis-CCMP2229-20130916|271706_1/174-272 [subseq from] nfYuHaYzRwqQ9pBK1CNfOCsOleE Karenia brevis CCMP2229
---VKAKRVSKIAQGKMAKAMVLRGSKEKTVGGLKKDSLMKNKRGKVVSKRAAAHGRRAYRNVEDWIEAVME
AREALHARGFVAINGRTLQGKALYVKAKAL--
>Karenia-brevis-Wilson-20130916|379606_1/197-295 [subseq from] 1driopnj9aMppW6IGo0FTGCmnlA Karenia brevis Wilson
---XKAKRVSKIAQGRMAKAMVLRGSKEKTVGGLKKDSLMKNKRGKVVSKRAAAHGRRAYRNVEDWIEAVME
AREALHARGFVAINGRTLQGKALYVKAKAL--
>Karenia-brevis-SP1-20130916|292963_1/158-247 [subseq from] DAggHOptNBGFQLSC9TpNz38t2VA Karenia brevis SP1
----------ASQSHXKKLVLRGSKEKTVGGLKKDSLMKNKRGKVVSKRAAAHGRRAYRNVEDWIEAVME
AREALHARGFVAINGRTLQGKALYVKAKAL--
>Karenia-brevis-Wilson-20130916|38660_1/4-105 [subseq from] taaVtfuxpVrJgu+qTtbufDz9pms Karenia brevis Wilson
MKVMKKKQASKIAKGKMAKVMVLRGSKEKTRGGLTKDSVMRNKRGKVVSKKASAHGRRVYRNIEDWTEALME
AREALHARGFVAVNGKTLQGKALYVKAKAI--
>Karenia-brevis-SP1-20130916|288311_1/4-105 [subseq from] ChZXbf/9ib36EJNTQybNV02NBns Karenia brevis SP1
MKVMKAKRTSKVAQGKFAKVLVFRGSKEKTRGGLTKDSVMRNKRGKIVSKKASAQSRRNYRHIESWTEAVME
AREALHARGFVAINGKTLQGKALYVKAKAL--
>Karenia-brevis-CCMP2229-20130916|47017_1/45-146 [subseq from] Iy/CD5UWMU56n3iSCe+XP5+aHA0 Karenia brevis CCMP2229
TKVMKAKKVSKIAKGKLAKALVFRGSKEKTRGGLTKDSVMRNKRGKIVSKKASAQGRRNYRNIEDWTEAVME
AREALHARGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-Wilson-20130916|378019_1/140-241 [subseq from] fr9v7yZYzY1dJjCtuEJCorm+bk8 Karenia brevis Wilson
TKVMKAKKVSKIAKGKLAKALVFRGSKEKTRGGLTKDSVMRNKRGKIVSKKASAQGRRNYRNIEDWTEAVME
AREALHARGFVAINGKTLQGKALYVKAKAL--
>Karenia-brevis-Wilson-20130916|38938_1/14-115 [subseq from] 6hfXhqMtTxF084objjcxMucY5dw Karenia brevis Wilson
MKGMKAKKVSKIAQGKLAKALVFRGSKEKTRGGLTKDSVMRNKRGKIVSKKASAQGRRNYRNIEDWTEAVME
AREALHARGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-SP1-20130916|288311_1/189-289 [subseq from] ChZXbf/9ib36EJNTQybNV02NBns Karenia brevis SP1
MKGMKAKKVSKIAQGKLAKALVFRGSKEKTRGGLTKDSVMRNKRGKIVSKKASAQGRRNYRNIEDWTEAVME
AREALHARGFVAINGKSLQGKALYVKAKL---
>Karenia-brevis-SP3-20130916|252702_1/14-114 [subseq from] 0bUX8i0mNx3iJDE/hUnQuTSZybs Karenia brevis SP3
MKGMKAKKVSKIAQGKLAKALVFRGSKEKTRGGLTKDSVMRNKRGKIVSKKASAQSRRNYRNIENWTEAVME
AREALHARGFVAINGKSLQGKALYGKLKL---
>Durinskia-baltica-CSIRO_CS-38-20140214|21350_1/42-143 [subseq from] 2zWSigNMXPlPHA0xwdRLoTjhuu8 Durinskia baltica, Strain CSIRO CS-38
AAARKPKRASKIARGRMAKAMVLRGSKEKTSGGLKSDALMRNKRNKVVSKRASAAGKRAYKNIETWVDCMVE
ARRALHITGFAAVNGKSLQGKALYVKCKAL--
>Kryptoperidinium-foliaceum-CCMP1326-20130916|253544_1/14-115 [subseq from] 5KZF1b//Q5ULJzWo612X/Ww6V8w Kryptoperidinium foliaceum CCMP1326
VDGRRKQRSSKVARGRMAKAMVLRGSKERTSGGLRREALMKNKRGKVVSKRASAVGKLRYKNIEGWVDSLME
ARRALHITGFVSVNGKTLQGKALYVKAKAL--
>Kryptoperidinium-foliaceum-CCMP1326-20130916|38427_1/13-114 [subseq from] ivCPQwxELEDb18OGPRZ88bgDgFE Kryptoperidinium foliaceum CCMP1326
VAGTKSKRGSKVARGRMAKAMVLRGSKEKTSGGLKREALMRNKRGRVVSKRASAVGKLRFRNIETWVDSLME
ARRALHISGFVAVNGKTLQGKALYVKAKAL--

>Glenodinium-foliaceum-CCAP1116_3-20130913|48184_1/43-144 [subseq from] BpMwnWdxzi64T2FO6e19CGiMFFw Kryptoperidinium foliaceum
TAMAKKKTSTKIAKGRMAKALVLRGSKEKTSGGLRKEGLMKNKRGKVVSKRASAVGKQRFRNIETWVDSLME
ARRALHISGFVAVNGKTLQGKALYVKAKAL--
>Kryptoperidinium-foliaceum-CCMP1326-20130916|413494_1/16-117 [subseq from] 7p5cBLZES6Giow4guXUcNgo0tuQ Kryptoperidinium foliaceum CCMP1326
TAMKAKKRGSKIAKGRLAKALVLRGSKEKTSGGLRKEALMKNKRGKVVSKRASAVGKQRYRNIETWVDSLME
ARRALHISGFVAVNGKTLQGKALYVKAKAL--
>Kryptoperidinium-foliaceum-CCMP1326-20130916|413494_1/221-319 [subseq from] 7p5cBLZES6Giow4guXUcNgo0tuQ Kryptoperidinium foliaceum CCMP1326
---QAAPNGSKIAKGRLAKALVLRGSKEKTSGGLRKEALMKNKRGRVVSKRASAVGKQRYRNIETWVDSLVE
ARRALRISGFVAVNGKTLQGKALYVKAKAL--
>Kryptoperidinium-foliaceum-CCMP1326-20130916|248760_1/37-138 [subseq from] vealDqO+p+BoiGwKxKzrKyS9xUM Kryptoperidinium foliaceum CCMP1326
GGVGRPKKPSKIAKGRMAKALVLRGSREKTSGGLKKEALMKNKRGKVVSKRQSALGKLRYKNIEGWTDCFME
ARRALHITGFVAVNGKTLQGKALYVKAKAL--
>Crypthecodinium-cohnii-Seligo-20130904|17115_1/15-118 [subseq from] kIh5lGWizB9G7JXR3g68up680Is Crypthecodinium cohnii Seligo
AKVMKAKRVTKVARGRLAKALVFRGSKEKTVGGVRKDGLMKNARGRIVSKKASAAGKRRFQNIESWVDSLSE
ARKALHISGFVAVNGKSLQGKALYSKTKSIYE
>Crypthecodinium-cohnii-Seligo-20130904|83155_1/2-95 [subseq from] zg2L+Qc0mz6m0GZjY98Zu1R+gy8 Crypthecodinium cohnii Seligo
----------KVARGRLAKALVFRGSKEKTVGGVRKDGLMKNARGRIVSKKASAAGKRRFQNIESWVDCLSE
ARKALHISGFVAVNGKSLQGKALYSKTKSIYE
>MMETSP1374-20130617|572_1/54-155 [subseq from] k2ujEjTf4AisBZQ0AY5kPTxwQ6o Symbiodinium Mf12.5f/clade A3
KTVMKAKRVSKIASGRLAKAMVLKGTREKTSGGLKKESLMKNKRGKIVSKRAAAAGKRRFKNIEEWCKSVAE
ARKALHVEGFVAINGKSLQGKALYVKSKAL--
>Amphidinium-carterae-CCMP1314-20130924|94707_1/27-128 [subseq from] empZyiyVKmJllOhCGsnJl4wkh70 Amphidinium carterae, Strain CCMP1314
KTVMKAKRVSKIATGRLAKALVLKGTREKTSGGLKKDSLMKNKRGKIVSKRAAAAGKIRFQNIEEWCKSVAE
ARKALHVEGFVAINGKSLQGKALYVKSKAL--
>Amphidinium-carterae-CCMP1314-20130924|18310_1/52-153 [subseq from] hbW1rJuD/w3dX1M1jb9TjB/rAZE Amphidinium carterae, Strain CCMP1314
KTVMKAKRVSKIASGRLAKAMVLKGAREKTSGGLKKDALMKNKRGKIVSKRAAAAGKDRFKQIESWTKCLAE
ARKALHLEGFVAINGKSLQGKALYVKSKAL--
>Azadinium-spinosum-3D9-20130829|105038_1/33-135 [subseq from] idWtAtiLLFrqG6jVqSrJYzfCuoQ Azadinium spinosum 3D9
RQAMKAKRVSKIARGRFAKALVLRGSKEKTVGGLTRDALTKNKRGKIVSKRAAANGKRRYQNIAAWVDSVVE
ARKALQMQGFVAINGRSVQGKALYVKSKSLF-
>Pseudo_nitzschia-fradulenta-WWA7-20140214|103127_1/50-139 [subseq from] ke2TR3VYX5fw8HUPIXjIGAA6PY0 Pseudo-nitzschia fraudulenta, Strain WWA7
----------EGQVAKALVLRGSKEKTVGGLTRDALTKNKRGKIVSKRAAANGKRRYQNIAAWVDSVVE
ARKALQMQGFVAINGRSVQGKALYVKSKSLF-
>Pseudo_nitzschia-fradulenta-WWA7-20140214|10935_1/59-160 [subseq from] dnxFJGeZ920Jg3Zcey0FqBUxew4 Pseudo-nitzschia fraudulenta, Strain WWA7
MKTMKAKRVSKVARGRLAKSMVFRGKKEKTVGGLTKDGLVRNKRGKIVSKRAAANGSKRFKNIEAWVQSVAE
ARKALQMQGFVAINGKSVQGKALYVKSKAL--
>Pseudo_nitzschia-fradulenta-WWA7-20140214|14139_1/33-134 [subseq from] uTj45F7Xmq13qlfDOD7y9AYrauY Pseudo-nitzschia fraudulenta, Strain WWA7
SGSMKAKRVSKIAKGRLSKAMVMRGSKERTVGGLTKDTLFKNKRGKIVSKRASAAGKRRFKNIKAWADSVVA
ARKALQVQGFVAINGKTAQGKALYAKAKAL--
>Karenia-brevis-SP3-20130916|122837_1/17-118 [subseq from] RPKvty4nvqsVOlHKSI4kZaxyNGg Karenia brevis SP3
KEAKKNTKKSKVARGRCAKLLVMRGKKEKTAGGLTRDALMRNKVGKIVSKKRSALGKRQFKHIEEWVGSVST
ARKSLKLEGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-CCMP2229-20130916|267803_1/40-141 [subseq from] V4LJdl85N8GARTv8DRLIdHsAmTE Karenia brevis CCMP2229
KTVMKAKKVSKIAKGRLAKALVLNGKKEKTSGGLTRDALMRNKRGKVVSKRRSALGKRQFKNIEEWVGNVKS
ARKQLALEGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-SP3-20130916|144073_1/9-110 [subseq from] z2FMnZvxbBz5okoixKQnlwLo0b4 Karenia brevis SP3
KTVMKAKKVSKIAKGRLAKALVLNGKKEKTSGGLTRDALMRNKRGKIVSKRASALGKRQFKNIEEWVGNVKS
ARKQLALEGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-Wilson-20130916|378265_1/169-270 [subseq from] 8ojx+BlbcHSwBjZyQRwmwIEddaQ Karenia brevis Wilson
KAVMKAKKVSKIAKGRLAKALVLNGKKEKTSGGLKRDALMRNKRGKVVSKRRSALGKKQFKNIEEWVGNVKS
ARKQLALEGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-SP1-20130916|13274_1/14-115 [subseq from] bMTqczPFmvKhujlTQ4Mwi+olQR8 Karenia brevis SP1
KAVMKAKKVSKIAKGRLAKALVLNGKKEKTSGGLKRDALMRNKRGKVVSKRRSALGKRQFKNIEEWVGNVKS
ARKQLALEGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-SP3-20130916|45671_1/18-119 [subseq from] 65niHGdYrrtE5irC74Kkg7H+pjo Karenia brevis SP3
KTVMKAKKVSKIAKGRLAKALVFNGKREKTSGGLKRESLMRNKRGKVVSKRRSAVGKKQFKNIEEWVGNVTT
ARKQLALEGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-Wilson-20130916|253045_1/3-103 [subseq from] dBZ58JNSMuA4u+g5/MYhFF7MbzI Karenia brevis Wilson
-TVMKAKKVSKIAKGRLAKALVFNGKREKTSGGLKRESLMRNKRGKVVSKRRSAVGKKQFKNIEEWVGNVKS
ARKQLALEGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-SP1-20130916|93830_1/1-96 [subseq from] tisDUKinZR7KLxQVgBomqtrCmp4 Karenia brevis SP1
------KKVSKIAKGRLAKALVFNGKREKTSGGLKRESLMRNKRGKVVSKRRSALGKKQFKNIEEWVGNVKS
ARKQLALEGFVAINGKSLQGKALYVKAKAL--
>Karenia-brevis-SP3-20130916|225025_1/12-113 [subseq from] 5/HOhwjeWMdty9zIEnNaHY/U5jo Karenia brevis SP3

KAAMKAKKVSKIAKGRLAKAQVLRGTKEKTSGGLTRDSLMRNKRGKIVSKRQNAKGKQAFKHIEHWVSSVTS
ARKSLAIEGFVAINGKSLQGKALYVKAKAL--
>MMETSP0467-20121206|40143_1/17-118 [subseq from] 9nawdEgRspAO8Evz9syKjZDlBuY Mesodinium pulex, Strain SPMC105
TTAMKAKRVSKTAHGRLAKALVFRGSKEKTVGGLTKDMLIKNKRGKVVSKRASARGKRIFRNIESWVEAHME
ARRALKVSGFVPLNGKSLQGKALYVKAKSL--
>MMETSP0503-20121227|25466_1/42-143 [subseq from] +qEPeXa7cnbOPYHT7JX6DYsqXXY Heterocapsa rotundata, Strain SCCAP K-0483
TTAMKAKRVSKTAXGRLAKALVFRGSKEKTVGGLTKDMLIKNKRGKVVSKRASARGKRIFRNIESWVEAHME
ARRALKVSGFVPLNGKSLQGKALYVKAKSL--
>MMETSP0126-20121128|2750_1/46-147 [subseq from] ujE3stTlRNZ4qHMhHkPxHrqHBZo Strombidinopsis acuminata
MKAMKAKRVTKVARGRLAKALVLRGSXEKTVGGLTKDTLMKNKRGKXVSKKASAWGKRAFARIESWVEAHME
ARKALXVSGFVPVNGKTLQGKAIYVKAKAI--
>ht00438-ht/29-130 [subseq from] FW8+SxoeHiBtbX8k5ICKKqmMM4c Heterocapsa triquetra
MKAMKAKRVTKVARGMFAKALVLRGSKEKTVGGLTKDMLMKNKRGKVVSKKASAWGKRSFTRIESWVEAHME
ARKALRVSGFVPVNGKTLQGKAIYVKAKAI--
>MMETSP0126-20121128|20405_1/45-146 [subseq from] ocj2LtH94HvBQtaaJjPCh4fgOFU Strombidinopsis acuminata
MKAMKAKRVSKVARGRLAKALVLRGSKEKTTGGLTKDMLMKNKRGKVVSKKSSANGKRAFTRIESWVEAHME
ARKALRVNGFVPVNGKTLQGKAIYVKAKAI--
>Azadinium-spinosum-3D9-20130829|190571_1/32-133 [subseq from] 4CduYD/Vt6fvxxbdLwn2wfJQkI8 Azadinium spinosum 3D9
RKAMKSKRVSKVAKGRGAKAQVLKGKKEKTVGGLKRDDLIRNKRGKVVSKRACANGKRRYRQIEDWVEALMA
ARSALHVQGFVAVNGKTLQGKALYAKTKAL--
>Pseudo_nitzschia-fradulenta-WWA7-20140214|146692_1/34-135 [subseq from] tseJ0964zK41zadg2Uqf9TLsFrY Pseudo-nitzschia fraudulenta, Strain WWA7
GKAMKGKRVSKVAKGRMAKALVLKGKREKTTGGLKMDMLMRNKRGKVVSKKAAAAGKRKYKNIEAWVEAHVA
AREALHVRGFVAINGRTLQGKAIYAKAKAL--
>MMETSP0796-20121207|16892_1/13-115 [subseq from] /uBY8nKEz5sd8Na4bSArKmCla8U Pyrodinium bahamense
MRAMKAKTASKVARGKYAKVLVFRGRKEKTVGGVTADGLMKNKKGRVVNKRRSAHSKRLYRRLEEWVQSLME
ARKALHAKGFVAINGRTLQGKALYIKAKELL-
>MMETSP0796-20121207|17200_1/36-138 [subseq from] oJLhqT8PMUF6sQwVMy6qdxWdLIg Pyrodinium bahamense
MKAMKAKTVSKVARGKFAKVLVLRGRREKTAGGVTAEGLMKNKRGRIVSKRHSAHGKRAYRWVEGWIEALME
ARRALHAKGFVAVNGRTLQGKALYVKAKELL-
>MMETSP0796-20121207|60263_1/12-92 [subseq from] nLECmJlSOUC2h9TFqwMnoaQx+4s Pyrodinium bahamense
MKAMKAKRVSKVARGKFAKVLVFRGRREKTAGGVTADGLMKNRRGRIVSKRRSANGKRVYRLVEDWTEALME
ARKALHAKG----------
>MMETSP0796-20121207|16723_1/25-126 [subseq from] HIAIYkIY7qa1y4xaNPc1ZlZxLJY Pyrodinium bahamense
-KAKKAKKASKVARGRFAKALVFRGLRERTAGGVTAMGLMKNKRGRIVSKRRSANGKRDYRRLEDWTEACME
ARKALHAQGFVPVNGRTLQGKALYFKAKALL-
>MMETSP0796-20121207|17070_1/13-115 [subseq from] 8HqkMypCxSHKsV1C4ozqozGKBeU Pyrodinium bahamense
KNAMKAKKVSTVARGRAAKALVFGGRREKTAGGVTAXGLMKNKRGRIVSKRRSANGKRDYKRLKDWTEACTA
ARKALGTEGFVAVGGRTSQGKALYGKAKALL-
>MMETSP0796-20121207|22702_1/46-148 [subseq from] SGTrNLy8w57Sygkvaf+ICJy3HTw Pyrodinium bahamense
MKAMKAKRVSKVARGKFARVLVFRGRKERTTGGVKADGLMRNKRGRIVSKRRSANGRRRYRHIEDWTEAVME
ARVALHTKGFVAINGRTLQGKALYVKAKTLL-
>MMETSP0796-20121207|14424_1/42-143 [subseq from] f7Fbuyl3Zfu+abGvsiuHK1L/sz8 Pyrodinium bahamense
MKAIKAKRVSKVAHGRFAKVLVFRGRKEKTPGGVTADGLMRNKHGRIVSKRQSANGRRRYRQIEDWVESLME
ARAALHTKGFVAVNGRTLQGKALYVKAKAI--
>Lingulodinium-polyedra-CCMP1738-20130920|211923_1/34-136 [subseq from] /q7A0AiTkkKxibmveF/Z+YjjQrI Lingulodinium polyedra CCMP1738
MKAMKAKRVSKVARGRFAKALVLRGKKEKTAGGVKAEGLMRNKRGKIVSKRASAAGRQRYKSVQGWIEALME
ARQALRAKGFVAINGKTLQGKALYAKAKAIW-
>Lingulodinium-polyedra-CCMP1738-20130920|363565_1/39-140 [subseq from] wxYtnEC5PZLjyTKycLqmgPtDJi4 Lingulodinium polyedra CCMP1738
TKAMKAKRVSKIARGRFAKALVLRGKKEKTAGGVKADGLMRNKRGKIVSKKASAAGKRRYKSVQGWVEALMQ
ARQALSAKGFVAVNGKTLQGKALYLKAKAI--
>Lingulodinium-polyedra-CCMP1738-20130920|284085_1/38-139 [subseq from] Z4Ji1wFDmQDLBPfSZbIVVQmVNjM Lingulodinium polyedra CCMP1738
MKAMKAKRVSKVARGRFAKALVFKGRREKTAGGVKADGLMRNKRGRIVSKRASAAGRRRYKSIEGWVEALME
ARQALHTKGFVAVNGKSLQGKALYLKAKAI--
>Lingulodinium-polyedra-CCMP1738-20130920|363565_1/189-273 [subseq from] wxYtnEC5PZLjyTKycLqmgPtDJi4 Lingulodinium polyedra CCMP1738
----------PKALVFKGRREKTAGGVKADGLMRNKRGRIVSKRASAAGRRRYKSIEGWVEALME
ARQALHTKGFVAVNGKSLQGKALYLKAKAI--
>MMETSP1439-20131203|41489_1/42-143 [subseq from] fn1AnbirmcMJXu/Hau8gheagBFI Gonyaulax spinifera, Strain CCMP409
MKAMKAKKVSKIARGRFAKALVFQGRREKTVGGVRADGLMKNKRGRIVSKRASAAAKRRYKNVEAWIEALME
ARQALHTRGFVAVNGKTLQGKALYVKAKSI--
>MMETSP0328-20130328|2684_1/14-115 [subseq from] DbS7i+6N0rhnz/SfcvrhB66odf4 Alexandrium minutum
MKAMKAKRVSKVAKGRLAKALVFKGSREKTVGGVKAEGLMKNKRGRIVSKRASANGRRRFQQVEDWIEAIME
ARQVLHTKGFVAINGKTLQGKALYVKAKAI--
>MMETSP0328-20130328|6731_1/18-119 [subseq from] ojh94lotJEYn7RLVA3joPmEGAE0 Alexandrium minutum
MKAMKAKRVSKVAKGRLAKALVFKGRREKTVGGVKAEGLMKNKRGRIVSKRASANGRRRFQQVEDWIEAIME
ARQVLHTKGFVAINGKTLQGKALYVKAKAI--
>MMETSP1439-20131203|124555_1/22-123 [subseq from] CcBj5lLHDPBw9HTYSbxbmkcAv14 Gonyaulax spinifera, Strain CCMP409
MKVMKAKRVSKVAKGRFAKALVFRGRKERTVGGVKAEGLMKNKRGKIVSKRASAAGKRRFRFVESWVSALME

ARQALHTRGFVAVNGRTLQGKALYVKAKSI--
>MMETSP1439-20131203|48805_1/2-100 [subseq from] C3cxLX4ofVAaZpsovA2d2NFH86k Gonyaulax spinifera, Strain CCMP409
---MKAKRVSKVASGRLAKALVFKGRRERTVGGVKAEGIMINKRGKFVSKRASAAGKRRFRFVESWVSALME
ARQALHTRGFVAINGKTLQGKALYVKAKSI--
>MMETSP0790-20130122|8206_1/16-117 [subseq from] Zl+lXksctyB9jeFinE+2ezGZwtE Alexandrium catenella, Strain OF101
MKAMKAKKTPKVARGKFAKVMVLRGRREKTSGGVKAEGLMKNKRGRVVSKRASANGRRRYEQVEDWIEAVME
ARQALHTXGFVAINGRTLQGKALYVKAKAL--
>Alexandrium-monilatum-CCMP3105-20140214|108175_1/46-148 [subseq from] nszaYXiQEdEkjaJeq5Y+nzcbAB8 Alexandrium monilatum, Strain CCMP3105
MKAMKAKRVSKVARGRFARALVFRGRKEKTIGGLTVDSLMRNKRGKIVSKRASANGKRRYRNVEGWIDALME
ARQALHTKGFVAVNGKTLHGKALYLKARAIH-
>Lingulodinium-polyedra-CCMP1738-20130920|362017_1/155-256 [subseq from] H4mesAid42FpRw3j1svVz0K/w7Q Lingulodinium polyedra CCMP1738
MKAMKAKRVSKVARGRFARALVLRGTKEKTVGGLKQDALMRNKRGKIVSKRASAAGKRRYRQIEDWTEAVME
ARQALHAKGFVAINGKTLHGKALYVKAKAI--
>Lingulodinium-polyedra-CCMP1738-20130920|354700_1/3-104 [subseq from] URxeEY+8wWqIGYOFp2khn5KpbCk Lingulodinium polyedra CCMP1738
MKAMKAKRVSKVARGRFAKVLVLRGRKEKTVGGLTKDGLMRNKRGKVVSKRASAAGRRRYKQVEDWIEAIME
ARQLLHTKGFVAINGRTLHGKALYVKAKAI--
>MMETSP0790-20130122|9734_1/42-143 [subseq from] y1OPoccNnlMCxj0I3lX8GOmPbHk Alexandrium catenella, Strain OF101
AKGMKAKKKSKVAKGRLAKYLVLSGRKEKTVGGVKADGLMXNKRGKIVSKRSSAAGRRKYKQVEDWVEAVME
ARQALHTKGFVAINGKTLHGKALYLKAKAL--
>Alexandrium-temarense-CCMP1771-20130823|412281_1/47-148 [subseq from] sG8aDNYp/a+E40Oba+OlKu9dT6Q Alexandrium temarense CCMP1771
KKTMKAKRETKIARGRLARFLVLKGRKEKTVGGITADGLMRNKRGKIVSKRAAANGRRRYKQVEDWIEAVME
ARVALHTKGFVAINGKTLHGKALYLKAKAL--
>Alexandrium-temarense-CCMP1771-20130823|200014_1/43-144 [subseq from] ySIIjkmf789y4FVIh9uU/eeLxm8 Alexandrium temarense CCMP1771
MKSMKAKRVSKVAHGRLARYLVLKGRKEKTVGGITADGLMRNKRGKIVSKRASAAGRRRYKQVEDWIEAVME
AREALHTKGFVAINGKTLHGKALYLKAKAL--
>MMETSP0661-20131031|32804_1/20-121 [subseq from] nGZrUzgnXVs4XLJFzsWh0Om2O/8 Alexandrium margalefi, Strain AMGDE01CS-322
LKASKAKRVSKVAHGRFARALVFRGRKEKTVGGLTKEGLMRNKRGKIVSKRASALGKRKYANIEDWTEAVME
ARQSLHTKGFVAINGKTLHGKALYIKAKSL--
>MMETSP0661-20131031|57215_1/17-118 [subseq from] IE1IW1wDvyS0CMLdeBOBrpluKo4 Alexandrium margalefi, Strain AMGDE01CS-322
MKAMKAKRVSKVARGRFAKALVFRGRKEKTIGGVTKDGLMRNRRGRIVSKRASALGKRRYANIQDWTEAVME
ARRSLHTRGFVAINGKSLHGKALYLKAKSL--
>MMETSP0661-20131031|12281_1/17-118 [subseq from] iCKr3ODF3itaWgH7VlMLwlAUmXk Alexandrium margalefi, Strain AMGDE01CS-322
MKAMKAKRVSKVARGRFAKALVFRGRKEKTVGGVTKDGLMRNRRGRIVSKRASALGKRRYANIEDWTEAVMQ
ARRSLHTQGFVAINGKTLHGKALYLKAKSL--
>MMETSP0661-20131031|4208_1/17-118 [subseq from] ARNqoY7KiBGwA2h1mvylY3vZGzY Alexandrium margalefi, Strain AMGDE01CS-322
TRAMKAKRVSKVARGRFAKALVFRGRKEKTVGGVTKDGLMRNKRGRIVSKRASALGKRRYANIEDWTDAVME
ARRSLHTQGFVAINGKTLHGKALYLKAKSL--
>MMETSP0661-20131031|55850_1/43-144 [subseq from] KavIeZqGUsmevqWZVnmG3Vg0UiQ Alexandrium margalefi, Strain AMGDE01CS-322
MKAMKAKRVTKVARGRLAKALVFRGRKEKTVGGVTKEGLMRNKRGKIVSKRLSALGKRKYANVEGWIEAVMQ
ARQSLHTQGFVAINGKTLHGKALYVKAKTL--
>MMETSP0661-20131031|11987_1/43-144 [subseq from] yb/Ws6jXKqP3dNnyvW0VRNFOUN4 Alexandrium margalefi, Strain AMGDE01CS-322
MKTMKAKRVSKVASGRFAKALVFRGRKEKTVGGVTKDGLMQNKRGRIVSKRMSAIGKRRYANVEGWIEAVME
ARQSLHTQGFVAINGKTLHGKALYLKAKSL--
>MMETSP0661-20131031|59882_1/2-85 [subseq from] zEcMCvtrvgW4GCtsigLTBDQ8FZ8 Alexandrium margalefi, Strain AMGDE01CS-322
----------KALVFRGRKEKTVGGVTKDGLMRNKRGRIVSKRVSARGKRSYANVEGWVEAVME
ARQSLHTQGFVAINGKTLHGKALYLKAKSL--
>Alexandrium-monilatum-CCMP3105-20140214|377589_1/52-153 [subseq from] UUHsq8hhVaq5Mi9I+uyugv+xhbk Alexandrium monilatum, Strain CCMP3105
MKTMKAKRVTKIARGRFARALVYRGRKEKTVGGLTQDGLMRNKRGRIVSKRASANGKRRYRNIEDWTEAVVE
ARQALHTRGFVAINGKSLHGKALYLKAKAI--
>MMETSP1439-20131203|125007_1/21-122 [subseq from] /77xX+6Fyto3KcBJJvnEBtUPZAw Gonyaulax spinifera, Strain CCMP409
MKVMKAKRVSKVARGRFAKALVFKGAREKTVGGVKADGLMKNRRGKIVSKRQSAAGKRRYRQIADWTEAVME
ARHALHSRGFVAINGRTLQGKALYVKARSI--
>MMETSP1439-20131203|7825_1/52-153 [subseq from] IDKdydlKneyKBW7o4JFgQq6VUFg Gonyaulax spinifera, Strain CCMP409
MKVMKAKRVSKVARGRFAKALVFRGAREKTVGGVKAEGLMKNKRGKIVSKRQSAAGKRRYRQISDWTEAVME
ARHALHSKGFVAINGKTLHGKALYVKARSI--
>Alexandrium-monilatum-CCMP3105-20140214|198942_1/11-112 [subseq from] IJgbr0ewMyb08tzHav7n8/c2hE4 Alexandrium monilatum, Strain CCMP3105
RKAMKAKRVSKVARGRFARVLVFRGRKEKTVGGLTKDGLMRNKRGKIVSKRRSALGRRAYLNVESWIEAVVE
ARQALHSKGFVAINGKTLQGKALYIKAKAI--

>Alexandrium-monilatum-CCMP3105-20140214|19378_1/36-137 [subseq from] rr34QTUu/FBOXaSfXXyFG3ZZz0g Alexandrium monilatum, Strain CCMP3105
MKAMKAKRVSKVARGRFARALVFRGRKEKTVGGLTKEGLMRNSRGKIVSKRANAHGKRGYRNIEGWTEAVMA
ARQALHSTGFVAVNGKSLHGKALYLKAKAI--
>MMETSP0790-20130122|27041_1/47-148 [subseq from] vV/YwIzk5zy/DnqA5GJaYlsIeKo Alexandrium catenella, Strain OF101
MKVMKAKRVSKIAHGRFARAMVFKGRKEKTVGGVKADGLMRNKRGKIVSKRASAVGRRLYKNIEGWTEAVVE
ARKALHSRGFVAINGKTLHGKALYLKAKAI--
>MMETSP1436-20131217|5119_1/15-116 [subseq from] wrR8OmxyRK+MuZkiGUVRC6UPunA Alexandrium andersonii, Strain CCMP2222
MKAMKAKRVSKVAHGRLARALVFRGARERTVGGVTKDGLMRNKRGKIVSKRASAHAKRNYHRIEDWVEAVVE
ARQALHSKGFVAINGKTLQGKALYIKAKAI--
>MMETSP1436-20131217|27218_1/43-144 [subseq from] bennEdfF2EOTfWX1PGMCj8vuXhQ Alexandrium andersonii, Strain CCMP2222
GAAMKAKRVSKVAHGRFAKVLVFRGKKEKTVGGVTKEGLMINKRGKVVSKRASARGKRHYRFIEGWTEAVME
ARQALHTKGFVAINGKTLHGKALYVKAKAI--
>Lingulodinium-polyedra-CCMP1738-20130920|130882_1/52-153 [subseq from] 5ZLKqcwOUvOQLLV7EsiQn4ZhOLM Lingulodinium polyedra CCMP1738
MKAMKAKRVSKVARGRLAKALVFRGSREKTSGGLTIDCLMRNKRGKVVSKRASAAGRRKYHYVEDWTRAVME
ARQLLHTQGFVAINGRTLHGKALYVKAKAI--
>Lingulodinium-polyedra-CCMP1738-20130920|209754_1/42-143 [subseq from] KD+mMBmdbknRibprkcGbbrT5Md0 Lingulodinium polyedra CCMP1738
SKGMKAKRVPKIARGRFARVLVFRGKREKTAGGLRAEALMRNKRGRIVSKRASARGRRCYKNVEDWIEAVMA
ARQALHTKGFVAINGKTLHGKALYVKAKAI--
>Lingulodinium-polyedra-CCMP1738-20130920|60086_1/14-115 [subseq from] Aj4phQSI1DSV4H2TH9+dyEA2Cmw Lingulodinium polyedra CCMP1738
MKTKRAKRVSKIARGRFAKVLVLRGKKEKTAGGVTAEGLMRNKRGRVVSKRASANGRRRFKQIEDWVEAVMA
ARTALHTKGFVAINGQTLHGKALYLKAKAL--
>Alexandrium-monilatum-CCMP3105-20140214|8621_1/47-148 [subseq from] 8mMS/Ury82R+b6Hyytk1+9fuXUo Alexandrium monilatum, Strain CCMP3105
MKAMKAKRVSKVACGRFAKALVFRGSKEKTIGGLTREGLLRNKRGKIVSKRASARGKRCYGNIEGWLEACME
ARKALHSTGFVAINGRSLQGRALYLKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|20412_1/20-121 [subseq from] 0SeC8PBGnxvR4G+S6bTXOFv6K88 Alexandrium monilatum, Strain CCMP3105
MKAMKAKRVAKVARGRFARALVFRGSREKTIGGLTREGLMRNKRGKVVSKRASARGKRCYGKIEAWTEACME
ARKALHSSGFVAINGKTLQGKALYLKAKAI--
>Lingulodinium-polyedra-CCMP1738-20130920|215292_1/20-113 [subseq from] XZekHAeDEnb8VqiBOVbEcRMCskk Lingulodinium polyedra CCMP1738
QKVMKAKRVSKVARGRMAKALVFRGSRERTSGGLRADGLMRNKRGRIVSKRASAQGKRRYQNIEDWTEAVME
ARQALHTKGFVAINGKTLQGKA----------
>Lingulodinium-polyedra-CCMP1738-20130920|217847_1/13-101 [subseq from] oztibWJFZK0WGncm8JmaKPf3KXE Lingulodinium polyedra CCMP1738
QKVMKAKRVSKVARGRMAKALVFRGSRERTSGGLRADGLMRNKRGKIVSKRASAQGKRRYQMIEDWTEACME
ARLALHTKGFVAINGKT---------------
>MMETSP0796-20121207|55852_1/37-139 [subseq from] OzPLxoRPWofAH+kWj8iN9BZofOM Pyrodinium bahamense
MKAKKAKRVSKVARGRLAKVLVFKGRKERTVGGLKAEGLMRNKRGRIVSKRASACGKRRYRFVEKWVEAVMQ
ARKDLHTKGFVAINGKTLHGKALYLKAKAIL-
>MMETSP1439-20131203|57726_1/41-142 [subseq from] VC+e0wxoO2xed0U1i4l5KHlnq30 Gonyaulax spinifera, Strain CCMP409
KRAMKAKRVSKIARGRFAKVLVLRGKKEKTVGGLRAEMLIRNKRGRVVSKRRSALGRRRYQYIEGWVEAVME
ARQALHTRGFVAINGKTLHGKALYLKAKSI--
>Alexandrium-monilatum-CCMP3105-20140214|54643_1/31-132 [subseq from] j9E9xflvf3FLIpZefGKJEAaFINw Alexandrium monilatum, Strain CCMP3105
EKAMKAKRVSKVARGRCARAHVFRGRKEKTVGGLTRECLMRNKRGRIVSKRASARGKRQYRYVEDWTEAVME
ARRALRSRGFVAINGKTLHGKALYLKAKAI--
>MMETSP0228-20121227|11844_1/55-156 [subseq from] msWOebTP8WzGmADsN5BHoehIFAA Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKSMKAKKVSKIARGRLAKVLVFRGKHEKTVGGIRADGLMRNKRGKIVSKRQAANGRRRFSQIKEWVEAVME
ARGALHSKGFVAINGRTLHGKALYVKAKAI--
>MMETSP0661-20131031|55784_1/33-134 [subseq from] 8nm0fBJqE0dp07z8mfC5xRlzCPE Alexandrium margalefi, Strain AMGDE01CS-322
KAATKAKRTPKVARGKFAKVLVFRGRREKTVGGVTIDGLMKNKRGRIVSKRKSALGRRSFTQIADWVEAVME
ARQALHTRGFVAINGRTLQGRALYVKAKAI--
>Prorocentrum-minimum-CCMP1329-20131001|125446_1/31-132 [subseq from] kSO4qdvvwJJNdM7O6WulYx/d2qg Prorocentrum minimum CCMP1329
KRVMKSKRVSKIATGRFAKVMVFKGKKEKTSGGIKADGLMLNKRGKVVSKRASAAGKRKYGRIEGWVEALMA
AREALHVKGFVAINGKTLQGKALYVKTKAL--
>Prorocentrum-minimum-CCMP1329-20131001|33903_1/54-155 [subseq from] 8lFcyONdV8bJ/q5Snj1mISmNyUw Prorocentrum minimum CCMP1329
SRVMKAKGVSKIATGRFAKVMVFRGKKERTVGGLRADALMVNKRGKVVSKRKSAAGKRRYSKIEGWVEALMA
AREALHVKGFVAVNGKTLQGKALYVKTKAL--
>Prorocentrum-minimum-CCMP1329-20131001|230902_1/24-125 [subseq from] xBtivfeN/GNf+CxbySykQGPiJFg Prorocentrum minimum CCMP1329
KRVMKAKGLSKIATGRLAKVMVFKGRKEKTSGGIKAEGLMVNKRGKVVSKRASAAGKRRYRNVEGWVEAFMA
ARAALHVRGFVAINGKTLQGKALYVKAKTI--
>MMETSP0766_2-20121228|69473_1/16-118 [subseq from] nJdwNftTP8FEu9UkqcpzMEIuu+4 Gambierdiscus australes
MKAMKAKRVSKIARGRLAKVLVFKGRREQTVGGLKAEALTVNKRGRVVSKRKSALCRRNYKGAVEWTECVMA

ARVALHATGFLAVNGKSLKGKALYVKAKAMH-
>MMETSP0766_2-20121228|48244_1/46-147 [subseq from] kL3KXxmuvVgRFGGcHXAakwixptQ Gambierdiscus australes
GKAMKAKRVSKIARGRFAKVLVFKGKRAKTVGGLKAESLMVNKRGRVVSKRQSAVGRRNYRGADEWIESLMA
AREALHTKGFVAVNGKTMQGKALYLKAKAI--
>MMETSP0784-20121206|50820_1/11-113 [subseq from] VgIU+Hja7GFMClq2t/8vlgY5GgM Gymnodinium catenatum
SNKKKGKRAPTVAKGRYAKVLVLRGTRKKTTGGLTAESLMRNKRGKVVSKRASAAGKRAFKNVESWLEAVMS
AREALHVRGFVAINGRSLQGKALYVKTKALY-
>MMETSP0784-20121206|65803_1/34-136 [subseq from] 6RNhGsYrhrR8ABahmLtymIjXLXk Gymnodinium catenatum
AKTMKAKRVATVARGRYAKALVLRGSKKKTTGGLTQEALMRNKRGKVVSKRASAAGKRAYRRIENWVEALMH
AREALHVRGFVAINGRSLQGKALYVKTKALY-
>Karlodinium-micrum-CCMP2283-20140214|14530_1/8-109 [subseq from] GDMUCaHlLiqlmPrSb1Y9tLVTGgY Karlodinium micrum, Strain CCMP2283
MKVMKSKRVSKIAKGRYAKSVVLRGLKEKTIGGLTKDRLMKNKRGKVVSKKSSAFGKRAYKNIQDWVSSVVA
ARKALQVTGFVAINGKSLQGKALYVKSKAL--
>km09892-km/30-131 [subseq from] HM377VSIT+Sx7Ay4PqefEvHF/9Q Karlodinium micrum, Strain CCMP2283
VKVMKSKRVSKIAKGRYAKSVVLRGLKEKTIGGLTKDRLMKNKRGKVVSKKSSAFGKRAYSNIQDWVSSVVA
ARKALQVTGFVAINGKSLQGKALYVKSKAL--
>Karlodinium-micrum-CCMP2283-20140214|59596_1/8-109 [subseq from] F0yqS5+6QnEM7jQ7KqUb+F6MEHs Karlodinium micrum, Strain CCMP2283
MKVMKSKRVSKIAKGRYAKSVVLRGFKEKTIGGLTKDRLMKNKRGKVVSKKSSASGTRAYKNIQDWVSSVVA
ARKALQVTGFVAINGKSLQGKALYVKSKAL--
>Karlodinium-micrum-CCMP2283-20140214|39206_1/41-142 [subseq from] Kc40psnhHjOcS7HHoapsrZIzly0 Karlodinium micrum, Strain CCMP2283
MKVMKAKRVSKIAKGRLAKSVVLQGRKEKTIGGLKREHLMKNKRGKVVSKKSSAFGKRAYKNIQDWVSSVVA
ARKALQVTGFVAINGKSLQGKALYVKSKAL--
>Karlodinium-micrum-CCMP2283-20140214|31725_1/9-109 [subseq from] ZX5kqFGJDPfVdMjGj7Hibt+bGQ8 Karlodinium micrum, Strain CCMP2283
MKA-KSKRVSKIAKGRLAKSVVLRGKKDKTVGGLTRENLMRNKRGKVVSKKSSAFGKRAYKNIQDWVSSVVA
ARKALQVTGFVAINGKSLQGKALYVKSKAL--
>Karlodinium-micrum-CCMP2283-20140214|204385_1/9-109 [subseq from] UD+S5yfA5L6VWX7s6vwfv6TpQmI Karlodinium micrum, Strain CCMP2283
MKA-KGKRVSKIAKGRLAKSVVLRGMKDKTVGGLTRENLMKNKRGKVVSKKSSAFGKRAYKNIQDWVSSVVA
ARKALQVTGFVAINGKSLQGKALYVKSKAL--
>km09428-km/27-127 [subseq from] u/sVwXm8sFTfyl+Lf7nUr58P+GY Karlodinium micrum, Strain CCMP2283
TKT-KSKRMSKVAKGRCAKSVVLRGTKEKTAGGLTRDKLMRNKRGKIVSKKSSASGKRAYTNIQDWVSSVVA
ARKALQVTGFVAINGKSLQGKALYVKSKAL--
>Karlodinium-micrum-CCMP2283-20140214|19771_1/26-126 [subseq from] MiXoZlAjD3wcNqAB86SRt+rzi9M Karlodinium micrum, Strain CCMP2283
SKT-KSKRMSKVAKGRCAKSVVLRGTKEKTAGGLTRDKLMRNKRGKIVSKKSSASGKRAYKNIQDWVSSVVA
ARKALQVTGFVAINGKSLQGKALYVKSKAL--
>Pseudo_nitzschia-fradulenta-WWA7-20140214|161032_1/43-144 [subseq from] FbScB8Hhdpdq1hiY9zQKklSxpCo Pseudo-nitzschia fraudulenta, Strain WWA7
MKGMKAKRASKIARGRMAKAMVFKGKKERTVGGLKQDGLTKNKRGKIVSKRASAAAKLRYRHIEGWLEAVME
AREALHTKGFVAINGKTLHGKALYVKAKNI--
>MMETSP0227-20121206|26432_1/52-153 [subseq from] NOD1SKZ+5t+r1SiMw9fkKah9lG4 Polarella glacialis CCMP1383
KKAMKAKRVAKVAKGRMAKAMVLAGKREKTTGGLKQDNLTKNKNGKVVSKKASAHGKRAFRHIEGWVEAVMA
AREALHTKGFVAINGKSLQGQALYVKARQV--
>MMETSP1338-20131121|23823_1/35-135 [subseq from] Y7/bD7V8oaTAxAsrsNITbOSlAHk Pelagodinium beii
KKAMKVKQ-TKIARGRMAKALVFAGKREKTSGGMKQESLMKNKRGKIVSKKMSAHGKRRFSNIEAWVDAVVQ
AREALHTTGFVAINGKSLQGKALYVKAKQV--
>Alexandrium-temarense-CCMP1771-20130823|409899_1/395-496 [subseq from] sYZZa4tpWeu968QPZ3zti86eCBY Alexandrium temarense CCMP1771
KKCMKAKRVSKIARGRLARHLVLSGRREKTTGGITAGELMKNKRGKIVSKRASAAGKRSYGQVEDWIEAVTE
ARKVLQTKGFVAINGKTLHGKALYLKAKSL--
>MMETSP0661-20131031|28109_1/28-129 [subseq from] ZL2e+k3tjTkq0Ry0yEWsaRo9lU8 Alexandrium margalefi, Strain AMGDE01CS-322
MKQKQGKRVSKIARGRLARYLVLKGSREKTTGGLKADDLMRNKRGKIVSKRANAAGRRRYKLVEDWVEAVME
ARGALHTRGFVPINGKTLHGKALYLKAKAI--
>MMETSP1377-20130617|328_1/12-114 [subseq from] zoooJTR3LA+x97nHvE+WfsgcFqs Symbiodinium type D1a
GHAMKAKRVTQVARGRLAKSMVFSGKKAKTSGGLTKDLLMSNAKGKIVSKRQSAHGKKSFKHIEGWVEAVME
ARAAFNAKGFVAINGKTLQGKALYAKAKTIL-
>MMETSP1110-20121109|33761_1/11-112 [subseq from] OnVkt+sPWH1UkjlTnjcgt6pFt24 Symbiodinium, Strain CCMP421
GRAMKAKRVTQVARGRLAKSMVFSGKKAKTSGGLTKDLLMSNAKGKIVSKRQSAHGKKSFKNIEGWVEAVME
ARAAFNAKGFVAINGKTLQGKALYAKAKTL--
>Symbiodinium-sp-Mp-20130822|25502_1/54-155 [subseq from] sufUa+SvIbQ/alMTw0Nsqlp0908 Symbiodinium sp Mp
SRVMKAKRVSQIAKGRMAKSMVFNGKKVKTTGGLTKDFLMQNSRGKIVSKKASAHGRKSYKHIEGWVEAVME
ARAAFNAKGFIAINGKTLQGKALYAKAKTI--
>MMETSP1110-20121109|47073_1/29-130 [subseq from] VO6QZTZv140vARjXK8kmG/ExBco Symbiodinium, Strain CCMP421
KAAMKAKKVTKIARGRMAKSMVFKGRKAKTSGGLTKDSLMVNKRGKVVSKLRSAKGKKSFKHIESWVEAVME
ARAAFNAKGFVAINGKTLQGKALYAKAKAI--
>MMETSP0227-20121206|35215_1/53-154 [subseq from] IJOZr09VJj9kyOcQbBuE6ymY9gI Polarella glacialis CCMP1383
MKAMKAKRISIVAKGRMAKSMVFKGKKEKTIGGIKKEGLVVNKRGKVVSKKASAKGKRSYKLIESWVEAVME
ARQSFNAKGFIAINGKSLQGKALYAKAKAL--

>MMETSP1338-20131121|73159_1/13-101 [subseq from] cZ7dpneNOG6szmCWIkIKFIty7pY Pelagodinium beii
MKAKKAKRVSQVAKGRMAKAMVFKGKKVKTVGGLTKESLVVNKRGRVVSKRQSANGKRSYKNIESWVEAVLE
ARSLFNAKGFIAINGKS----------
>Prorocentrum-minimum-CCMP1329-20131001|261193_1/215-316 [subseq from] sf3d4RHWuSvNSbwIxlCOZfFNPSM Prorocentrum minimum CCMP1329
PRVMKAKRFSKIARGRFAKAMVFRGSKEKTVGGLTKDMLMKNKRGKVVSKKASAASKRKYKSIEAWTEATAS
ARRALQLKGFTAINGKSVQGKAMYVKAKTL--
>Prorocentrum-minimum-CCMP1329-20131001|93618_1/14-115 [subseq from] VL/Lyc3QQ21T3lpPxUWYl10YYB8 Prorocentrum minimum CCMP1329
PRVMKAKRFSKIARGRFAKAMVFRGSKEKTVGGLTKDMLMKNKRGKVVSKKASAASKKKYKSIEAWTEATVS
ARRALQLKGFTAINGKSVQGKAMYVKAKTL--
>Prorocentrum-minimum-CCMP2233-20131001|257896_1/54-152 [subseq from] PTCUq3pj/go3eEytdjLtG9qmjL8 Prorocentrum minimum CCMP2233
PRVMKAKRFSKIARGRFAKAMVFRGSKQKTVGGLTKEMLMKNKRGKVVSKKASAASKKKYKSIEAWTEATVS
ARRALQLKGFTAINGKSVQGKAMYVKX-----
>Prorocentrum-minimum-CCMP2233-20131001|255218_1/4-102 [subseq from] FKkvQKvXgTupC1oFIUPuB4HmRL8 Prorocentrum minimum CCMP2233
PRVMKAKRFSKIARGRFAKAMVFRGSKQKTVGGLTKEMLMKNKRGKVVSKKASAASKKKYKSIEAWTEATAS
ARRALQLKGFTAINGKSVQGKAMYVKX-----
>Prorocentrum-minimum-CCMP2233-20131001|257896_1/257-355 [subseq from] PTCUq3pj/go3eEytdjLtG9qmjL8 Prorocentrum minimum CCMP2233
PRVMKAKRFSKIARGRFAKAMVFRGSKEKTVGGLTKDMLMKNKRGKVVSKKASAASKKKYKSIEAWTEATVS
ARRALQLKGFTAINGKSVQGKAMYVKX-----
>Prorocentrum-minimum-CCMP1329-20131001|261193_1/54-152 [subseq from] sf3d4RHWuSvNSbwIxlCOZfFNPSM Prorocentrum minimum CCMP1329
PRVMKAKRFSKIARGRFAKAMVFRGSKEKTVGGLTKDMLMKNKRGKVVSKKASAASKKKYKSIEAWTEATAS
ARRALQLKGFTAINGKSVQGKAMYVKX-----
>Prorocentrum-minimum-CCMP1329-20131001|82158_1/14-107 [subseq from] TpoPGy1eQ87YJfeiO8aSP26Nnac Prorocentrum minimum CCMP1329
PRVMKAKRFSKIARGRFAKAMVFRGSKQKTVGGLTKEMLMKNKRGKVVSKKASAASKRKYKSIEAWTEATVS
ARRALQLKGFTAINGKSVQGKA----------
>Prorocentrum-minimum-CCMP2233-20131001|42914_1/14-115 [subseq from] DfmuStdahr6Sdi4C3h2tYwoFQ4s Prorocentrum minimum CCMP2233
PRVMKAKRFSKIARGRFAKAMVFRGSKQKTVGGLTKEMLMKNKRGKVVSKKASAASKKKYKSIEAWTEATVS
ARRALQLKGFTAINGKSVQGKALYVKAKAL--
>MMETSP0790-20130122|40122_1/48-150 [subseq from] r8kyqc4yYO47XbYPk1fvoCQ7Uyg Alexandrium catenella, Strain OF101
MKAMKAKRVPKIARGRFAKSQVLSGKREKTVGGVKADGLMKNKRGKVVSKRMSARARRFYHNVEPWIESIVE
ARQALHTKGFVLINGKTLVGKALYVKAKAIY-
>MMETSP0448-20130528|102950_1/36-137 [subseq from] Tt8YDwWShyo2K+Z93oJA0WKM4jg Heterocapsa triquetra
TEAMKAKRQVKVAKGRLAKALVFHGLRERTASGLRQDALTKNKRGKVVSKKKSASGKRIYKQVEDWVEAVME
ARTALHTQGFVAINGKSLQGKALYAKARAI--
>Alexandrium-monilatum-CCMP3105-20140214|28743_1/9-110 [subseq from] 1ykz+Ns3n/BM0qLv+3kxTb+X0h4 Alexandrium monilatum, Strain CCMP3105
KKARKPKRVPKIARGRSARALVFSGRREKTVGGVRREGLMRNTRGRIVSKRRNALGKRAFRNVEGWVQAIME
ARVALHSKGFVAINGKTLHGKALYLKAKSI--
>Alexandrium-monilatum-CCMP3105-20140214|28781_1/10-112 [subseq from] 93ckRwI6KRs4vwGm7U8ZFUxTF5k Alexandrium monilatum, Strain CCMP3105
SKAIKAKRMSKVAHGRLAKALVLRGSKEKTKSGLTRDGLMRNKRGKVVSKRASAHGHQIYKNIAGWIEAVRE
ARQVLHTQGFVAINGKTLHGKALYLKAKSIL-
>Alexandrium-monilatum-CCMP3105-20140214|407830_1/62-153 [subseq from] yNihVdMrB8iK363iHoPvb3pNWRQ Alexandrium monilatum, Strain CCMP3105
------------GEGPLREVLVLRGRKEKTVSGLTKDGLMRNKRGRVVSKRASAAGMRRYRRVEGWIESLME
ARQMLHTKGFVAVNGKSLHGKALYLKAKSIYE
>Durinskia-baltica-CSIRO_CS-38-20140214|16833_1/28-130 [subseq from] w1s4/QwJh0lT0khTOGgSbof9j+0 Durinskia baltica, Strain CSIRO CS-38
KGAMKVKRVSKVAKGRLAKAMVFKGKKLKTTGGVTKEGIMRNKSGKYVSKRRSATGRRMYKNIEGWVDTIME
ARKALHITGFVAVNGKSLQGRALYVKAKALL-
>MMETSP1338-20131121|44477_1/16-119 [subseq from] fQhvOBH+23MNmA4eLZLAmZ8Z1Xg Pelagodinium beii
SKARKAKRVSKVATGRMAKVLVMRGAKEKTSGGLKKDALMKNRNGKIVSKRKSAQGKKNFAKIEQWVQCIAE
AREKLHLTGFSAINGKTTAGKALYAKTKALYT
>MMETSP0227-20121206|65208_1/44-146 [subseq from] a0IQoPIS8CE+iHeKnEB5oKI1aV0 Polarella glacialis CCMP1383
MKVMKAKRVTKVARGRFAKAMVMRGSKEKTTGGLTKDSLTKNKSGKIVSKRASAAGKRNFRTIESWVQSVVE
ARQQLHLQGFVAMSGKSMAGKALYAKTRALY-
>MMETSP0766_2-20121228|37199_1/22-124 [subseq from] 4skb46MwXx3rlS+byxRKpGsN4cM Gambierdiscus australes
TKVRKSSGSSKVARGRFAKVLVLRGKKEKTVGGVTANGLMRNKRGKVVSKRQSALGRRNYRHVENWTLAFLA
AREALHTKGFVAINGKTLQGKALYVKTRAIL-
>MMETSP0766_2-20121228|38417_1/44-145 [subseq from] csvJNVtIiuIch8+A9XLq2qF0hKs Gambierdiscus australes
KKAGKATRGSKIARGKFSKVQVFRGKKEKTAGGLRQESLMRNKRGRIVSKRKNAIGKRLYHYAEDWIDAFMA
AREALHTKGFVAINGKTLQGKALYLKTKAI--
>Scrippsiella-hangoei_like-SHHI_4-20140214|126861_1/49-150 [subseq from] dhhHoToFb5o0cLc7SqnzOkw92uc Scrippsiella hangoei-like, Strain SHHI-4
-PARQHRPATKIARGRFARVLVLRGKKEKTVGGVTRDGLKRNKRGRVVSKKMSALAKAKYRHIEKWVDSLVA
ARKALHLQGFVAINGNSMQGKALYVKTKDLF-

>MMETSP0227-20121206|47407_1/53-154 [subseq from] aGzSxx32pZeOCzC6CAG4rruMfh4 Polarella glacialis CCMP1383
KKSMKAKRVSKVARGRLAKVMVFRGRRERTVGGLRKDSLVVNKRGRIVSKRASAHGKLSFKRIETWVQAVMA
ARRAVNAKGFIAINGKKPGGRELYAKAKEL--
>Ceratium-fusus-PA161109-20140214|40549_1/12-113 [subseq from] 2mWw1DDqXyovRPdGlYy3EnLFdjg Ceratium fusus, Strain PA161109
KKPMKAARLRKIARGRLAKAQVLAGKYEKTVGGIRAEGLMRNKRGKVVSKRQSARGKLAYKNVEDWIEAFMA
ARAALHCVGFVAINGQSLQGKAFYIKSKAL--
>Ceratium-fusus-PA161109-20140214|216767_1/28-129 [subseq from] rfflX9x7ESA/PBaDXuArkfkDurg Ceratium fusus, Strain PA161109
KKSIKVGRAKTIARGRFAKALVLRGKYQKTAGGIKAERLMRNKRGKVVSKRQSARGKIMYKNIEGWVEAFMA
ARAALHCTGFVAINGKTLQGKALYIKTKAM--
>Ceratium-fusus-PA161109-20140214|216767_1/196-297 [subseq from] rfflX9x7ESA/PBaDXuArkfkDurg Ceratium fusus, Strain PA161109
KTTMKRGQKRTIARGRFAKALVLAGKYQKTVGGVRAEGLMRNKFGRVVSKRRSARGKILYKNIEDWTEALMA
ARAALHCTGFVAINGKTLQGKALYIKNKAM--
>MMETSP0228-20121227|29246_1/26-125 [subseq from] Ko/xDdwARdygcNAjvAwXqIf43nE Protoceratium reticulatum, Strain CCCM 535
(=CCMP 1889)
--KKKSTKPMKVARGRLAKSLVLKGKCEKTVGGIRADGLMRNKRGKIVSKRGSANGKRRYKQVEDWIQAVVQ
ARDALHTKGFVAINGKSLIGKALYLKAKAI--
>Durinskia-baltica-CSIRO_CS-38-20140214|31785_1/17-118 [subseq from] K6N5Z3y+iUd1oK3zyvp/IZHNBwc Durinskia baltica, Strain
CSIRO CS-38
AKAAKTKRVGKIARGRMAKAMVLRGSKEKTAGGLSGGDAQEGERNRVVSKRASAAGMRAFKHIEAWVDCVVE
ARRALHITGFVAINGKSLQGKALYVKCKAL--
>Durinskia-baltica-CSIRO_CS-38-20140214|31277_1/17-118 [subseq from] jOAYW/eKTcY83ggYogwY0lQVuzo Durinskia baltica, Strain
CSIRO CS-38
AKAATTKRISKIARGRMAKAMVLRGSQEKTSGGLKRKAXDEEQAKQCVSKRASAAGMRAFKHIEAWVDCVVE
ARRALHMTGFVAINGKSLQGKALYVKCKAL--
>Durinskia-baltica-CSIRO_CS-38-20140214|11343_1/36-132 [subseq from] x7odA5R5HIL lZbGMKDhFCKU5gwU Durinskia baltica, Strain
CSIRO CS-38
GRQAEQDR-----GGRMAKAMVLRGSKEKTAGGLKREALMRNKRNRVVSKRASAAGMRAFKHIEAWVDCVVE
AHRALHMTGFVAINGKSLHGKALYVRCKAV--
>MMETSP0790-20130122|12236_1/20-122 [subseq from] /jVKDQ5MUC7b+pkd1IiX157WYrA Alexandrium catenella, Strain OF101
MKAKQRKRRSKVARGRVAKAMVFSGAREKTAGGLTKDGLVRNKEGRLVSKRRNALGRQKYHQVEDWVKAVVQ
ARQAVHATGFVAVNGKTVLGKAIYLKAKSIL-
>Crypthecodinium-cohnii-Seligo-20130904|6203_1/38-140 [subseq from] DptsgE2sIwAElpvu6VqRWrZe83g Crypthecodinium cohnii Seligo
MKAMKKKKLSIIARGRLAKSQVLKGRKVKTVGGLKKEALMKNKRGKVVSKRQNVHGNHAYKNVRGWVDSVMS
ARRSLGLEGFVAINGKNAQGKALYVKAKALW-
>MMETSP0253-20130528|9206_1/18-117 [subseq from] JaWnTbgh5g5SsdPT5rtMU35O2O8 Noctiluca scintillans
---SKAKPVSKVARGRYAKVLVLRGRKERTAGGLKADNLFKNKRGKVVSKKASAAGQRQYRNIEAWVEAHMS
ARAALRVTGFAPLNGRTLTGKALYVKTRALY-
>MMETSP0253-20130528|12949_1/11-109 [subseq from] 2TKFfDVS+mxSnQuKTR5I9UScKh8 Noctiluca scintillans
----KGKCAAKVARGRGAKALVLNGRRERTAGGLKADMLMRNRSGKVVSKRASAAGRRNFAKIEPWLQAVMT
ARECLRVTGFVAINGRTLQGKALYVKSRAIY-
>MMETSP1439-20131203|35891_1/52-153 [subseq from] BNOIglmncs7/P9/uqL9XfrGNpgg Gonyaulax spinifera, Strain CCMP409
KQVMKAKRVSRVAKGKFAKVLVFSGRRLKTQGGLTKDMLMKNKRGKIVSKRQSAHGKRTFHNVNVWCEALSV
ARRAMHISGFVAMNGKSLAGKALYVKVRAL--
>Ceratium-fusus-PA161109-20140214|123722_1/38-141 [subseq from] ImSqx+iCHu/zXIp09gsWmkdhmbY Ceratium fusus, Strain PA161109
QRVMKAKRVSKIARGRKAKLLVFRGAHEKTVSGLRADGLMMNKRGKIVSKRKSAIVSRRFETIKDWVDSIML
ARQQLNLNGMVLVNGKSSQGKALYLKAKAAYE
>Ceratium-fusus-PA161109-20140214|16001_1/13-113 [subseq from] za8Cs534W5UoFkfILEFFcBeJ+kw Ceratium fusus, Strain PA161109
QRVMKAKRVSKIARGRMAKALVLRGKHEKTAGGLRADGLMKNKRGKVVSKRQSAQNAR---KSKDWVDSIMT
ARQLLNMNGFVLVNGKSAQGKALYIKAKATYE
>Ceratium-fusus-PA161109-20140214|7638_1/816-916 [subseq from] VfS0TbWurGdYJ31z8uBoiofaxjg Ceratium fusus, Strain PA161109
QRVMKAKRVSKIARGRMAKALVLRGREKTSGGLRADGLMKNKRGKVVSKRKSALYAK---TTKDWIDSCMS
ARQLLNLNGMVCVNGKSSQGKALYLKAKAAYE
>Ceratium-fusus-PA161109-20140214|34624_1/35-135 [subseq from] YoBD3ZEflh4oMxJixSR5+QRkZa0 Ceratium fusus, Strain PA161109
KRVMKAKRVSKIARGRSAKLMVFRGKYEKTSGGVRADGLMKNKLGKVVSKSKSAWTRR---VLKNWTDSVMS
ARQLLNLEGFVAINGNSLEGKALYLKAKAAYE
>Karenia-brevis-CCMP2229-20130916|50553_1/14-116 [subseq from] pZBO7xUZRmrQ1x4uZoBKuc3umLE Karenia brevis CCMP2229
MKLMKRKRRSKIARGRMAKAMVMRGTKEKTVGGLTRMDLKRNKRGKIVSKKASAHGANMYKFVEEWVKAFME
ARADLGTAGFLAINGKSQQGKALYTKALAVY-
>Oxyrrhis-marina-LB1974-20131105|535_1/31-133 [subseq from] JuByjFCAIT4IfrRWR7qoi6Q5kEk Oxyrrhis marina LB1974
-KVKKVKRVSKVAKGKLAKAAVFKGSKAKTSTGLTASDLVKSKSGKIVSKKKSILAKKNFAKIGGWNKAVMA
ARKALGVKGFCAIGGKSAQGKALLAKARALYK
>Oxyrrhis-marina-20131105|151391_1/832-934 [subseq from] G6EJ2OlHfG11eCb8HJCYqTpOFao Oxyrrhis marina
-KVKKVKRVSKVAKGKLAKSAVFKGAKAKTSTGLTASDLVKSKSGKIVSKKKSALAKKNFAKIGGWNKAVMA
ARKALGVKGFCAIGGKSAQGKALLAKARALYK
>Oxyrrhis-marina-LB1974-20131105|7037_1/34-121 [subseq from] uDPq78zm17ktjXcpzWHzkhn2riE Oxyrrhis marina LB1974
-KVKKVKRVSKIAKGKLAKAAVFKGSKAKTSTGLTASDLVKSKSGKIVSKKKSALAKKNFAKIGGWNKAVMA
ARKALGVKGFCAIGGKS----------
>Favella_taraikaensis_FeNarragansettBay-20140227|13121_1/1-86 [subseq from] upsODks+ylYhHwJ62jYV9liukEM Favella taraikaensis,
Strain Fe Narragansett Bay
--------VSKIAKGKLAKSAVFKGAKAKTSTGLTASDLVKSKSGKIVSKKKSALAKKNFAKIGGWNKAVMA
ARKALGVKGFCAIGGKSAQGKA----------
>Oxyrrhis-marina-LB1974-20131105|4254_1/50-152 [subseq from] Vcr2oUeHiWldVdmLshC6YGN91GA Oxyrrhis marina LB1974
-KVKKVKRVSKVAKGKRAKAAVFKGSKAKTATGLTASDLVKSKSGKIVSKKKSALGKKNFAKIGGWTKACMA

ARKALGIKGFCAIGGKSAQGKALLAKARALYK
>Oxyrrhis-marina-20131105|149221_1/72-174 [subseq from] o3+qbtZALylJKlM+fL0jQauMcF4 Oxyrrhis marina
-KVKKVKRVSKIAKGKRARAAVFKGTKEKTATGMKASDLVKSKSGKIVSKKASATAKKNFAKLGGWLKAVTA
ARKALGVKGFCIIGGKSAQGKALLAKARALYK
>Oxyrrhis-marina-20131105|11209_1/17-103 [subseq from] bt1lhoTiAX6dIu62AD/UJLX0DBo Oxyrrhis marina
----------TRSEAFKGTKEKTATGMKASDLVKSKSGKIVSKKASAMAKKNFAKLGGWNKAVTA
ARKALGVKGFCIIGGKSAQGKALLAKARALYK
>Oxyrrhis-marina-LB1974-20131105|10285_1/34-136 [subseq from] VwDmc2RVeKQUKvg5Vr/lEulXi/M Oxyrrhis marina LB1974
-KVKKTKRVSKIASGKRAKSAVFKGRKEKTVGGLKASDLMKSKTGKIVSKKQSMSAKKNFAKLGGWNKAVMA
ARKALGVKGFCAVGGKSTQGKALLAKARALYR
>Oxyrrhis-marina-LB1974-20131105|9971_1/38-140 [subseq from] oWeehCu1ak9k3h8SvokrvyVd/KI Oxyrrhis marina LB1974
-KVKKTKRVSKIASGKLAKSAVFKGRKEKTVGGLKASDLMKSKTGKIVSKKQSMSAKKNFAKLGGWNKAVMA
ARKALGVKGFCAVGGKSAQGKALLAKARALYR
>Oxyrrhis-marina-LB1974-20131105|65374_1/34-133 [subseq from] +WbEU5FvrINRHdhe89RDwaCls6w Oxyrrhis marina LB1974
-KVKKAKRVSKTARGKLAKSAVFKGRKEKTVGGLKASDLIKSKTGKIVSKKRSMSAKKNFAKLGGWNKAVMA
ARKALGVKGFCAVGGKSAQGKALLAXRKS---
>Oxyrrhis-marina-20131105|19870_1/38-140 [subseq from] w6xX6bZCHCgeYzmVBrd7Wo1X9Uo Oxyrrhis marina
-KAKKVKRVSKIATGKRAKAAVFKGKKEKTSSGLKSSDLVKSKTGKIVSKKQSLRGKKQFAKLGGWNKAVMG
ARKALGVKGFCAIGGKSAQGKALLAKARSLYK
>Oxyrrhis-marina-20131105|11173_1/52-154 [subseq from] t+0OF5C4bk2/n9KB/AQDEsqemWg Oxyrrhis marina
-KVKKVKRVSKIAKGKRAKAAVFKGTKEKTATGIKKSDLMKSKSGKLVTKKQNAAGKKNYKRISGWTKACMA
ARKALGVKGFCPVGGKSAQGKALLAKTRALYK
>Oxyrrhis-marina-20131105|7472_1/34-136 [subseq from] mh8yfVeq+9wpvrSZLXIrejdOpZM Oxyrrhis marina
-KVKKVKRVSKIAKGKRAKAAVFKGTKEKTATGIKKSDLMKNKSGKLVTKKQNAAGKKSYKRISEWTKACMA
ARKALGVKGFCPVGGKSAKGKALLAKTRALYK
>Oxyrrhis-marina-LB1974-20131105|65491_1/46-143 [subseq from] DVyMgKOLxR/2BJvoviA60yNBmrc Oxyrrhis marina LB1974
-KVKKVKRVSKIAKGKRARAAVFKGTKEKTATGIKKSDLIKSKSGKLVTKKASAAGKKNYKRISGWTKACMA
ARKALGVKGFCPVGGKSAQGKALLAKT-----
>Oxyrrhis-marina-20131105|149221_1/288-390 [subseq from] o3+qbtZALylJKlM+fL0jQauMcF4 Oxyrrhis marina
-KVKKVKRVSKIAKGKRARAAVFKGTKEKTATGIKKSDLMKSKSGKLVTKKANAAGKKSFKNISGWLKAVTA
ARKALGVKGFCAVGGKSAQGKALLAKARALYK
>Oxyrrhis-marina-20131105|9804_1/118-220 [subseq from] 608bM0maN6ZOu9u3Z1Ey6m3UnPQ Oxyrrhis marina
-QTKKVKRVSKIAKGKRARAAVFKGTKEKTATGIKKSDLMKSKSGKLVTKKASANGKKSFKNISGWLKAVTA
ARKALGVKGFCAVGGKSAQGKALLAKARALYK
>Oxyrrhis-marina-20131105|18382_1/43-145 [subseq from] kwIVrpx0fBH/D1xgxb24WDaCTNU Oxyrrhis marina
-KVKKVKRVSKIAKGKRARSAVFKGTKEKTATGLKASDLMKSKSGKLVTKKASAAGKKKFKNISGWIKAVGA
ARKALGVKGFCAVGGKSAQGKALLARARALYK
>MMETSP0693-20131125|7023_1/10-109 [subseq from] lcd4sCv1NBVP/VJzJ5ZPbwxHQ/k Thalassionema nitzschioides, Strain Unknown
-KVKKVKRVSKIAKGKRARSAVFRGTKEKTATGLKASDLMKSKSGKLVTKKASAAGKKKFKNISGWIKAVGA
ARKALGVKGFCAVGGKSAQGKALLARARA---
>Favella_taraikaensis_FeNarragansettBay-20140227|2239_1/6-106 [subseq from] 8ciuKWj+ItID5CuQNdjCPVs+dlo Favella taraikaensis,
Strain Fe Narragansett Bay
-KVKKVKRVSKIAKGKRAKAAVFKGTKEKTATGIKKSDLMKSKSGKLVTKKASAAGKKKFKNISGWIKAVGA
ARKALGVKGFCAVGGKSAQGKALLARARAL--
>Oxyrrhis-marina-LB1974-20131105|9126_1/54-156 [subseq from] 3IfgWo1DT/RDn+CPtXMqcxPibvI Oxyrrhis marina LB1974
-KVKKVKRVSKIAKGKLAKAAVFKGKKEKTVAGVRKEDLIKSKTGKLVTKRASEAGKKKYKNVVGWIKAVTM
ARRSLGVTGFCAVGGKSAQGKALLAKARALYK
>Oxyrrhis-marina-LB1974-20131105|8670_1/65-167 [subseq from] 0XV9vTCvnQke5qfMi/eFQPnBuHc Oxyrrhis marina LB1974
-ESEEDKESEQNREWKRAKSAVFKGRKEKTVGGLKASDLIKSKTGKIVSKKQSMSAKKNFAKLGGWNKAVMG
ARKALGVKGFCAVGGKSAQGKALLAKARALYR
>Oxyrrhis-marina-LB1974-20131105|65491_1/309-411 [subseq from] DVyMgKOLxR/2BJvoviA60yNBmrc Oxyrrhis marina LB1974
-RVKKVRKVSVIAKGKLAKASVFKGKKVKTASGLKASDLTKSKSGKIVSKKASAAGKKAYAKIGAWTKAVMA
ARKALGVKGFCAVGGKSAQGRALLAKTRSLYK
>Oxyrrhis-marina-20131105|5373_1/58-160 [subseq from] BRLiAoNJJeuCv0xNX82n+XsKGCg Oxyrrhis marina
-KVKKVKRVSKIAKGKRARSAVFRGTKEKTSTGLAKSDLMLNKNGKLVSKKASAAGKKKFQSISGWSKAVSA
ARKALGLKTFCPIGGKSAQGTALLAKTRSFYR
>Oxyrrhis-marina-LB1974-20131105|65186_1/71-171 [subseq from] BpnVsUM1Ug6wz46aYkHGrmkbS0c Oxyrrhis marina LB1974
---SSKKRVSKVAKGKRAKAAVFAGTKEKTATGLKKSDLMKTKSGKLVTKKRHAAGKKAYKNVSAWTKATQQ
ARKTLGIKGFCPIGGKSAQGKALLNKIRSLYX
>Oxyrrhis-marina-CCMP1795-20131203|1672_1/14-102 [subseq from] wyRaLLZOvVC1RKdQZTBMAnNX2pw Oxyrrhis marina, Strain
CCMP1795
---SSKKRVSKVAKGKRAKAAVFAGTKEKTATGLKKSDXMKTKSGKLVTKKRHAAGKKAYKNVSAWTKATQQ
ARXTLGIKGFCPIGGKSAQG----------
>Oxyrrhis-marina-LB1974-20131105|10516_1/32-134 [subseq from] 6av1eTZ49ZzRefPHcQq83V6uU/o Oxyrrhis marina LB1974
-ASKKARKASKIAKGKRAKVAVFHGTKEKTPGGLKASDLVKSKRGKIVSQKKSALGKKNFAKLGAWNKAVAA
ARKAMGLKGFCAIGGKSAQGKALLEKARSLHR
>Oxyrrhis-marina-20131105|13230_1/36-138 [subseq from] 0k23tq8zdO/+v8RQCncmsI4DupU Oxyrrhis marina
-KAKKAKSVSKIARGKRAKAAVFNGTKEKMPSGLTASDLMRSKRGKIVSQKKNALGKKNFAKLSAWNRAVAA
ARKALGVKGFMAIGGKSAQGKALLAKARALHT
>Oxyrrhis-marina-LB1974-20131105|10205_1/34-133 [subseq from] nFpLhT4PtkF+zlPK8hu8f1h3fRo Oxyrrhis marina LB1974
-KVKKSKRVSKVAKGKRAKSAVLRGKHEKTPGGLKASDLMKAKSGKIVSKKQSAQGKKSFGE---WHKAVVA
ARKALGITGFCVIGGKSAQGKALLAKIRALYR
>MMETSP0467-20121206|14010_1/15-118 [subseq from] UqB1JCf8m/4ZP4tkG04So6p7RN8 Mesodinium pulex, Strain SPMC105

MKAMKAMKKSVIAKGKLQKVMVFNGFYEKTQGGLAKGDMHKSKRGKIVSKKSTAAGKKAYKNISGWTKACMA
AKKALGLKGFVAVGGKTAQGKAVYAKAKAIYN
>MMETSP0503-20121227|27800_1/13-98 [subseq from] 5zPcjwVRlus6WfavUs5/FZdq3zw Heterocapsa rotundata, Strain SCCAP K-0483
MKAMKAMKKSVIAKGSMKKAMVFNGFYSKTSSGHAKGDLHKSKTGKIVSKKATAAGKKAYKNISGWTKACMA
AKKALGLKGFVAIG----------
>MMETSP0467-20121206|11853_1/2-104 [subseq from] TaIH32fXWyY+lhsc5X+knjYDkrc Mesodinium pulex, Strain SPMC105
-KAMKAMKKSVIAKGARARAMVFAGAYAKTKTGLKKGDLHKSKRGKIVSKKATASGKKAYKNISGWVKACAA
ARKALGLKGFVAIGGKTAQGKAAYAKAKAIYN
>ht01324-ht/32-135 [subseq from] VNhFMkcGWb89ZxpTvS7/SxvlYpY Heterocapsa triquetra
MKVMKAMKKSKIAKGKMMRAMVMNGSYEKTSSGLKKSNLIKSKSGKIVSKASSAAGKKAYKRISGWTNACKA
AKKALGLSGFVPCGGKSAQGKAFYAKAKAIYN
>MMETSP0126-20121128|12372_1/1-92 [subseq from] 8Thkk2dEExmwfVCg4xtRFsFzaFE Strombidinopsis acuminata
----------AKGKMMRAMVMNGSYEKTSSGLKKSNLIKSKXGKIVSKAASAAGKKAYKRISGWTNACKA
AKKALGLSGFVPCGGKSAQGKAFYAKAKAIYN
>MMETSP1441-20131203|6962_1/2-98 [subseq from] SbsqqraE9x33YlQ4B85VCMNdtXU Heterocapsa arctica, Strain CCMP445
-------KKSVVAKGKMMRAMVFAGAYAKTKTGLKKADMHKSKTGKIVTKAQTAAGKKAYKHISGWTKACQA
AKKALGLKGFVPCGGKSAQGKALYAKAKAIYT
>MMETSP0467-20121206|7893_1/74-176 [subseq from] J4tA1H6CRG0abnZLQf/adu5Zf0k Mesodinium pulex, Strain SPMC105
MRAMKAKKVSVIAKGKLMRAMVFQGAYAKTKTGLRKADLHKNKDGKIVTKASTAAGKKAFKRISAWNRACKT
AKRTLGLKGFVPVGGKSAQGKALYKKARALY-
>MMETSP0503-20121227|27826_1/7-106 [subseq from] KbgJGnWe0pDid6eqgrShs/IqBfk Heterocapsa rotundata, Strain SCCAP K-0483
MKAMKAMKKSVIAKGQNKRAMVFRGFYAKTSTGLAKSDLIKNKNGKIVSKKRSANGKKAYKRVSGWIAACKA
AKKALGLKGFVAIGGKTAQGKAVYAKAK----
>Symbiodinium-sp-C15-20130923|75686_1/35-138 [subseq from] m9+8np/6YzE1wFUBT3QDmP4E2ak Symbiodinium sp C15
MKAMKAMKKSVIAKGKHCRASVLKGYKEKTYTGLKKADLMVNKRGKIVTKKAHASAKKQYAKISGWTKACQQ
ARKELGIKGFCPVGGKTAQGRALYAKTKSLYQ
>Symbiodinium-sp-C1-20140214|100176_1/56-159 [subseq from] DATeHyBIRzpN+FLqWJ/hHDPZ1qg Symbiodinium, Strain C1
MKVMKAMKKSVIAKGKHCRASVLKGYKEKTYTGLKKTDLMVNKRGKIVTKKAHASAKKMYGKISGWTKAFQQ
ARKELGIKGFCPVGGKTAQGRALYAKTKSLYQ
>MMETSP1377-20130617|18513_1/60-161 [subseq from] q2giLjQVA8LFwMshOCIBC+DnqYo Symbiodinium type D1a
MKAMKAMKVSKIAKGKLCRSSVLKGYKEKTYTGLKKTDLMKNKRGKIVSKKAHAASKKAFSNISGWMKAVQQ
ARKELGIKGFCPVGGKTAQGKALYAKTKSL--
>MMETSP0191-20121206|48533_1/28-129 [subseq from] m6Wd/y1VtZPHwqX1aMGtmyieUzM Sorites sp.
MKVMKAMKKSVIAKGKRAKVSVIKGFKEKTATGLKKADLMVNKRGKVVSKKAHAAAKKKFQAISGWLKAVTQ
ARKELGIKGFCPVGGKTAQGKALYAKAKSL--
>Symbiodinium-sp-CCMP2430-20130923|97717_1/30-132 [subseq from] cCslwsKjGyfiCgjZt8XMLSPgbjU Symbiodinium sp CCMP2430
MKAMKAMKVSVIAKGKRARASVFSGSKEKTYTGLKKSDLMKNKRGKIVSKKSHASGKSSYKMIKGWTEACQQ
ARKELGIKGFCAVGGKTAQGKALYAKAKSIY-
>MMETSP1110-20121109|18832_1/36-138 [subseq from] Yqy4zFWp9p8uTRf97UwniBz4T1E Symbiodinium, Strain CCMP421
MKAMKAMKVSVIAKGKRARASVFSGTKEKTYTGLKKGDLMKNKRGKIVSKKSHASGKSSYGQIKGWTEAVQQ
ARKELGIKGFCAVGGKSAQGKALYAKAKSIY-
>MMETSP1110-20121109|21198_1/65-166 [subseq from] fBzud4S6QFNV6D5oHwN6LoGVCLc Symbiodinium, Strain CCMP421
MKAMKAMKVSKIAKGKRARASVFHGTKEKTYTGLSKGDLKLNKRGKIVSKKSSAAGKKSYENISAWTKACQQ
ARKELGIKGFCPIGGKTAQGKALYAKVKSI--
>Symbiodinium-sp-Mp-20130822|20930_1/38-139 [subseq from] 8WasnQFTBFoox8Q23iBL4QOVwn4 Symbiodinium sp Mp
MKAMKAMKVSAIAKGPRARASVFTGGKEKTYTGIKKSDLMKNSRGRLVTKKSHKAGVKSYKHISKWGEATQK
ARKELGIKGFCPMGGKTAQGKALYAKTKSX--
>MMETSP0784-20121206|57436_1/9-111 [subseq from] ld7uu6EvpHbeIfoTuf4/YeeDLpc Gymnodinium catenatum
-KVMKAMKVSKIAKGPRAKAAVFLGSREKTQGGLTKADLRKNKSGKIVSKKASAAGKKAYSHISKWTQAVQK
AKKALGITGFAIVGGKTGQGKALYAKAKSIYE
>MMETSP0784-20121206|58208_1/33-135 [subseq from] qaY/zPAjhO2kGSqL75uWA4N+oMI Gymnodinium catenatum
-KVMKAMKVSKIAKGPRAKAAVFFGSKEKTLGGLTKADLRKNRSGKIVSKKSSAAGKKAYSHISKWTQACQK
AKKALGITGFAVIGGKTGQGKALYAKAKSIYE
>MMETSP0224-20130122|2371_1/19-120 [subseq from] 2Iqt6OcbNfJgO+yLEp/tKThE6ns Togula jolla, Strain CCCM 725
MKAMKAKRVSVIAKGKRAKTAVFAGTKEKTYTGLKKSDLMRSRTGKIVSKKSSAAAKKRFAKISAWSKAVQK
ARKAMHVKGFVPIGGKTAQGKALYAKAKSL--
>MMETSP0224-20130122|12723_1/59-160 [subseq from] 9le0F2zI4GeHQ3AzCbImwo2t4ms Togula jolla, Strain CCCM 725
MKGMKAMKVSVIAKGKRAKTAVFAGTKEKTYTGLKKSDLMRSRTGKIVSKKSSAAAKKRFAKLAAWSKAVQK
ARKAMGVKGFVPIGGKTAQGKALYAKAKSY--
>MMETSP0224-20130122|16753_1/63-164 [subseq from] hTfnHAxqOMYuHMZrSP25MhyzN/Y Togula jolla, Strain CCCM 725
MRAMKAMKVSVIAKGKHARSAVFHGTKAKTHTGLKKSDLIRSKSGKIVTKKSSAAAKKNFAKLAAWAKACQK
ARKALGITGFVPMGGKTPQGKALLAKAKTF--
>Symbiodinium-sp-Mp-20130822|189753_1/22-124 [subseq from] yIWKzeeRRrQqeWoXedHHRJaabOk Symbiodinium sp Mp
MRVMKKKAVSKIAHGKRAKVVVFTGKKEKTQSGIKKSDLMKNKRGRIVTKAQNAAGKKAYKNLSAWTEACMK
ARKELGIKGFCAVGGKTAQGKALYAKTKALL-
>Symbiodinium-sp-Mp-20130822|162596_1/45-146 [subseq from] qiZiHn4KLA10N3UqE6DZSmI8tec Symbiodinium sp Mp
MRSMKKKAVSKIATGKLSKAVVFRGNKEKTRSGIKKADLMKNKRGKIVTKAQNSAGKKAYKNLSAWTEACMK
ARKELGIKGFCPVGGKTSQGRALYAKAKAL--
>Symbiodinium-sp-Mp-20130822|185135_1/56-157 [subseq from] WZV3CM58nsZ4oIQHRC9uzBDmhWI Symbiodinium sp Mp
MRAMKKKAVSKIATGKLSKAVVFRGKKEKTRSGIKKGDLMKNKRGKIVTKAQNSAGKKAYKNVSAWTEACMK
ARKELGIKGFCPIGGKSAQGKALYAKAQAL--
>Symbiodinium-sp-Mp-20130822|2526_1/39-140 [subseq from] rMH5Q/Va3+8qKX9RzdTc6xM5h44 Symbiodinium sp Mp
MRAMKKKAVSKIATGKLSKSVVFKGNKEKTRSGIKKSDLMKNKRGKIVTKAQSSAGKKAYKNISAWTAAVVK

ARKELGIKGFCPVGGKTPQGKALYAKAQAI--
>Symbiodinium-sp-C15-20130923|145161_1/25-127 [subseq from] AALMHdmCgFKwOqGl4V/pU/nUQho Symbiodinium sp C15
MKAMKKKAVSKIAKGKLAKAVVFKGNKEKTASGHKKSDLMMNKRNRIVTKKQNAAGKKAYKNISAWTTATLK
ARKELGIKGFCAVGGKTAQGKALYAKAKAIY-
>Symbiodinium-sp-C1-20140214|106326_1/212-314 [subseq from] 2k/qldqJSQhXU/tYuEGtIgowlP0 Symbiodinium, Strain C1
MKAMKKKAVSKIARGKLAKVVVFKGNKEKTASGHKKSDLMINKRNKIVTKKQNAAGKKAFKHISGWNTAVIK
ARKELGIKGFCAVNGKTAQGKALYAKAKAIY-
>MMETSP0191-20121206|27886_1/48-137 [subseq from] GAKb5vSAgsDD5R2qxKnuN9cHqF8 Sorites sp.
-RAMKKKAVSXIARGKLAKVVVFKGNKEKTASGHRKSDLMINKRNKIVTKKQNAAGKKAYQHISSWNKAVIK
ARKELGIKGFCAVNGKSAH----------
>MMETSP0191-20121206|24886_1/29-130 [subseq from] noEf74tQ1PnQ+LEXE8HRMRd92Js Sorites sp.
-RAMKKKAVSKIAKGKNAKSVVFKGNKEKTGSGFSKSDLMKNKRNKIVTKKQNAAGKKAYKNISAWTTATXK
ARKELGIKGFCPMGGKTAQGKALYAKAKAIY-
>Symbiodinium-sp-C1-20140214|104613_1/50-151 [subseq from] zL9dkE//T2wTedEyWnxnq3ejzOk Symbiodinium, Strain C1
-RAMKKKAVSKIARGKLAKVVVFKGNKEKTASGHKKADLMLNKRNKTVTKKQNAAGKKAYKNISAWTTAVTQ
ARKELGIKGFCLVNGKTAQGKALYARAKANY-
>MMETSP1377-20130617|12080_1/49-150 [subseq from] pPa5ie22Re3zHUXCVwsavy/SGIM Symbiodinium type D1a
-RAMKKKAVSKIARGKLAKVMVFKGNKEKTSGGNSRSDLMKNKRNKVVSKKQNAAGKKNFKYISGWNEAVIK
ARKELGIKGFCAINGGSAQGKALYAKAKAIY-
>MMETSP1377-20130617|750_1/29-130 [subseq from] ZldKK9G9TRTL+m905ye24WaIiMU Symbiodinium type D1a
-RVMKKKAVSKIARGKLAKVVVFKGNKEKTASGQSRSDLMKNKRNKIVSKKQNAHGKKAFKFISSWTSAVIS
ARKELGIKGFCAINGKTAQGKALYAKAKAIY-
>MMETSP1377-20130617|15842_1/23-124 [subseq from] Uem6zVNO3J0MUROVAnb/dR86FQw Symbiodinium type D1a
-RVMKKKAVSKIARGKLAKVVVFKGNKEKTAGGKTRSDLMKNKRNKVVSKKQNAAGKKAYANISTWTSSVIK
ARKELGIKGFCAVNGKTPSGKALYAKAKAIY-
>Symbiodinium-kawagutii-CCMP2468-20131203|4955_1/5-94 [subseq from] /jMICBHRstuzrIGJ3UQtOm54vtE Symbiodinium kawagutii
CCMP2468
-------HVXKIARGKLAKVVVFKGNKEKTASGFSASDLMKNKRNKIVTKKQNAAGKKAYKNISAWTTAVTK
ARKELGIKGFCAVNGKTAQGKALYA-------
>MMETSP1110-20121109|22567_1/23-124 [subseq from] AA9VTONLE/Nvn/CDvmMKyrH+DPc Symbiodinium, Strain CCMP421
-KVMKKKAVSKIAKGKNAKAVVFKGSXEKTATQLTKTDLMKNKRGRVVTKKQNAAGKKAFKNISAWTEAVTK
ARKELGIKGFCAVNGKTSQGKALYAKAKAIF-
>Symbiodinium-sp-C1-20140214|106326_1/25-126 [subseq from] 2k/qldqJSQhXU/tYuEGtIgowlP0 Symbiodinium, Strain C1
-RVMKKKAVSKIARGKSAKAVVFRGSKEKTATGFKQTDLLKNKRNKVVTKKQHAAGKKAYKNISAWTTAVTK
ARKELGIKGFCAVNGKTAQGKALYAKAKATY-
>MMETSP1110-20121109|31181_1/32-133 [subseq from] iVSpkstsaEeOmkxXE+v7WPyV4/E Symbiodinium, Strain CCMP421
-KAMKKKAVSKIAKGKRAKTAVFKGAKEKTATQLTKSDLMMNKRGKVVTKKKSAAGKKAYKYISAWNTAVQK
ARKELGIKGFCPVNGKTAQGKALYAKAKAIF-
>MMETSP0191-20121206|28726_1/29-130 [subseq from] LvYuTr/uGppSRjmfW5gbpf9JoAU Sorites sp.
-TVMKKKAVSKIARGKRAKVVVFKGNKEKTASGHSKADLMLNKRGRIVSKKQNAAGKKMYQNISGWTTALQK
ARKELGIKGFCAVNGNTAHGKALYAKAKAIY-
>Lotharella-globosa-CCCM811-20130918|475_1/15-116 [subseq from] kXyKyJoKOIAAb6NrO1eptd5d53s Lotharella globosa, Strain
CCCM811
MRVMKKKAVSKIARGKRAKATVFNGKKEKTQGGFRKADLVKNKRGKVVTKAQHAAGKKHYAKIASWTEAMMK
ARKELGIKGFCAIGGKTAQGKALYAKTKTL--
>Aurantiochytrium-limacinum-ATCCMYA1381-20130828|512_1/24-125 [subseq from] HoqHyM4YIqXMl2eI3gI6k84RFmM
Aurantiochytrium limacinum ATCCMYA1381
MRVMKKKAVSKIARGKRAKATVFNGKKEKTQGGFRKADLVKNKRGKVVTKAQHAAGKKHYAKIASWTEAVMK
ARKELGIKGFCAIGGKTAQGKALYAKTKTL--
>Aplanochytrium-sp-PBS07-20140214|7889_1/25-126 [subseq from] CkYE/zt1mw00qrsxGR2QN/d+OgI Aplanochytrium sp, Strain PBS07
MRVMKKKAVSKIARGKRAKATVFNGKKEKTQGGFRKADLVKNKRGKVVTKAQHAAGKKRYAKIASWTEAVMK
ARKELGIKGFCAIGGKTAQGKALYAKTKTL--
>Symbiodinium-sp-Mp-20130822|24742_1/152-253 [subseq from] c0Cl5mrIT/QIWURur0Oi0AC1RLo Symbiodinium sp Mp
MRVMKKKAVSKIARGKLAKAVVFQGKKEKTKGGFCKADLLKNKRGKVVTKAKHAAGKKSFVTIAPWNEALMK
ARKELGIKGFCPVGGKSAQGKALYAKTKAL--
>Symbiodinium-sp-C15-20130923|8043_1/29-130 [subseq from] QXI9Xa4ruMzifLqY4VeLClA1Rsg Symbiodinium sp C15
-TVMKKKAVSKIARGKRAKAAVFLGRKEKTASGHSRSDLMKTKHGRIVTKAKHAQGKKAYAYISAWTEALQK
ARKELGIKGFCAVNGKTAQGKALYAKTKAIF-
>Symbiodinium-sp-C15-20130923|143988_1/28-129 [subseq from] 0R+Z6T7FGGi4/0ZBPeEXE4ZGt10 Symbiodinium sp C15
-TVMKKKAVSKIARGKRAKIAVFKGKKEKTASGHSRSDLMKTKQGRIVSKAQHAHGKKAYAYISAWTEAVQK
ARKELGIKGFCAVNGKTAQGKALYVKAKAIF-
>Symbiodinium-sp-C1-20140214|106967_1/622-723 [subseq from] oenDk+3PJEyaPRkLwZyPoKGow9A Symbiodinium, Strain C1
-RVMKKKAVSKIARGKRAKIAVFKGKKEKTSSGHTRNDLMINKRGKVVSKKANAAAKNRFKNISGWNNAVIQ
ARKELGIKGFCAVNGKTAQGKALYVKAKALY-
>Symbiodinium-sp-C15-20130923|8174_1/27-128 [subseq from] nS3FWnRo7/YfJXmrupe82fff+Go Symbiodinium sp C15
-RVMKKKAVSKIARGKRAKIAVFKGKKEKTSSGNTRNDLMINKRGKVVSKKASAAAKKRFNYISGWNKAVIQ
ARKELGIKGFCAVNGKTAQGKALYVKAKALY-
>MMETSP1377-20130617|18719_1/29-130 [subseq from] ERDQajKVqpQJOngcs+Rp+JJVX5M Symbiodinium type D1a
-KTMKTKRVSKVAKGKRAKVVVFKGNKEKTASGIAKKDLMKNKTGKIVTKKSHAAGVKAYKQISAWTIACQG
ARKELGIKGFCPVGGKTAQGKALYAKAKAIY-
>Symbiodinium-sp-C1-20140214|71288_1/49-150 [subseq from] pZ/6K1TxO30h1jqxXkX7OTACtKI Symbiodinium, Strain C1
-KAMKKKRVSKVARGKRAKVAVFKGNKEKTASGHAKADLMVSKTGKVVSKKSHAAGKRAYKNIAAWTKACQA
ARTELGIKGFCAVNGKTAQGKAFYAKAKAIF-

>Symbiodinium-sp-C1-20140214|95707_1/61-162 [subseq from] RYBP71jok3OYq5JKePCz8Y0AKzY Symbiodinium, Strain C1
-KVMKKKAVSKIATGKRAKVAVFKGRKEKTVSGLKREHLMINKRGKVVSKKMNAAGKNRFKQISAWINACQS
ARNELGIKGMCCVGGKTVQGKALYVKAKAIF-
>Symbiodinium-sp-C15-20130923|125527_1/24-125 [subseq from] MxomX3QHj1ZHoDaQ5+cdxLt/6TY Symbiodinium sp C15
-KVMKKKAVSKIATGKRAKVAVFKGRKEKTVSGLKREHLMINKRGKVVSRKMNAAGKNRFKHISAWIDACQS
ARNELGIKGMCCVGGKTVQGKALYVKAKAIF-
>Symbiodinium-sp-CCMP2430-20130923|13530_1/20-122 [subseq from] gNkC+sWF6sASt2rSf/h1GQ51fIk Symbiodinium sp CCMP2430
MKAMKRKAISKIAQGKRAKLAVFKGSKEKTYTGLRKTDLMKSKSGRIVSKKQHAKGKALFQQAKKWVDAVIT
ARKELGLKGFCAVGGKSAQGKALYAKAKTIY-
>MMETSP1377-20130617|17651_1/20-122 [subseq from] x6wAVSfTIcfgCBKOjl28HtU9pfg Symbiodinium type D1a
MKAMKRKAISKIAQGKRAKLAVFKGSKEKTYTGLRKTDLMKSKSGRIVSKKQHAKGKALFQQAKKWVDAVIT
ARKELGLKGFCAVGGKSAQGKALYAKTKTIY-
>MMETSP1377-20130617|13384_1/46-148 [subseq from] 63E0MW6rO9MhTtYjUhSTz15z7zA Symbiodinium type D1a
MKAMKKKAISKIARGKLAKSVVFNGRKEKTYTGLHKTDLMKNKDGRIVTKKQHARGKALFQQAKKWVEAVST
ARKELGLKGFCAVGGKSAQGKALYAKAKAIY-
>MMETSP1110-20121109|13484_1/27-129 [subseq from] zL/bQYJlaLTS6zIv2Hg2y/xqoRY Symbiodinium, Strain CCMP421
MKAMKKKAVSKIARGKLAKAVVFNGRKEKTYTGLHKTDLMKNKSGRIVTKKQHAKGKXLFQQAKKWLDAVMT
ARKELGLKGFCAVGGKSAQGKALYAKAKAIY-
>MMETSP1110-20121109|41808_1/44-146 [subseq from] XTSbSaoOxXZ8CBkOdxdM4Utz3t0 Symbiodinium, Strain CCMP421
MKAMKKKAISKIARGKLAKSVVFNGRKEKTSTGLHKTDLMKNKYGRIVTKKQHAKGKALFQQAKKWLDAVMT
ARKELGLKGFCAVGGKSAQGKALYSKAKAIY-
>MMETSP1110-20121109|28355_1/26-128 [subseq from] xZfVnBC50XMuaFFCDSTzWCPEP7E Symbiodinium, Strain CCMP421
MKAMKKKAVSKIARGKLAKMVVFKGRKEKTSTGLQKAHLMKNKRGKVVTKKQHAIGAALFKKGQRWLEACMT
ARKELGVKGFCAIGGKSAQGKALYAKAKSLY-
>MMETSP1110-20121109|31536_1/28-130 [subseq from] u93RFEyO6dvJCj1XBb8uXzGGHQU Symbiodinium, Strain CCMP421
MKAMKKKAITNVARGKLAKMVVFKGRKEKTATGLQKAQLMKNKRGKVVTKKQHARGKALFQKAQRWLDAVMT
ARKELGVKGFCAIGGKSAQGKALYAKAKALY-
>MMETSP1110-20121109|24708_1/46-147 [subseq from] 79a6cKodmu7SnEgDvTyFgjwH98Y Symbiodinium, Strain CCMP421
TKVMKKKAVSKIARGKLAKAAVFQGRKEKTSTGLQKMHLMKNKYGKVVSKKKHASGK-QFQKGQKWISAVMT
ARKELGMKGFCAVGGKSAQGKALYAKAKALY-
>MMETSP1377-20130617|25676_1/44-145 [subseq from] rMKBqaziKS8KEDJFv+BAeVxlrmo Symbiodinium type D1a
MKVMKKKAVSKIARGKLAKAAVFQGRKEKTSTGLQKMHLMKNKYGKVVSKKKHASGK-LFQKGQKWISAVMT
ARKELGMKGFCAVGGKSAQGKALYAKAKALY-
>MMETSP1110-20121109|16400_1/26-127 [subseq from] ZC0ynD3WwOzTUEneJ0jWcK6bhV8 Symbiodinium, Strain CCMP421
MKVMKKKAISRIARGKLAKVAVFQGRKEKTATGLQKTHLMKNKHGKVVTKKKHASGK-LFQKGQKWISAVMT
ARKELGMKGFCAVGGKSAQGKALYAKAKALY-
>MMETSP1110-20121109|67004_1/46-147 [subseq from] LHys4kXx8HkpwnphdAw3gZyhcw4 Symbiodinium, Strain CCMP421
MKVMKKKAISRIARGKLAKVAVFQGRKEKTATGLQKTHLMKNKHGKVVTKKKHASGK-QFQKGQKWISAVMT
ARKELGMKGFCAVGGKSAQGKALYAKAKALY-
>Symbiodinium-sp-CCMP2430-20130923|7199_1/55-156 [subseq from] gUKWxvvzZf5HYGjsUxwfSvx5e3c Symbiodinium sp CCMP2430
MKAMKKKAVSKIATGKRAKSVVFRGSKEKTSGGLAKSQLMKNKRGKVVSKKMHAKGK-TIQKVQQWLDAVMT
ARKELGIKGFCAVGGKSAQGKALYAKAKALY-
>MMETSP1110-20121109|16207_1/48-149 [subseq from] Q3RsIh4iodVjZxb8DvQSNk6oe7s Symbiodinium, Strain CCMP421
MKVMKKKAVSKIAMGKRAKSVVFRGSKEKTSGGLAKSQLMKNKRGKVVSKKMHAKGK-TIQKVQRWLDSVMT
ARKELGIKGFCAVGGKSAQGKALYAKAKALY-
>MMETSP1377-20130617|18322_1/29-132 [subseq from] 6XMrCDvTZ4N1wN3dUir/zoQMR+A Symbiodinium type D1a
MKAMKAMKVSIIAKGTRAKWSVLKGTKEKTTGGLTKADLKKNKRGKIVSKKASANALKRFNTLGSWLKAVST
ARKELGIKGFCPVGGKTAQGKALYAKAKSIYT
>Symbiodinium-sp-Mp-20130822|138925_1/54-154 [subseq from] oywieE+5BikHpuWyLPR9Hqq2toI Symbiodinium sp Mp
-KAMKAMKVTKVASGKMRKSVVLRGSKEKTSGGLTKADLTRNKNGRVVSKKMSAAAKSRFASVGPWIAAVSK
ARKELGIKGFCAVGGKSAQGKALYAKAKSL--
>MMETSP0448-20130528|4808_1/37-138 [subseq from] XNmY/uWsN8U4HRJEdZzoP7hBz10 Heterocapsa triquetra
MKAMKAMKASPIAKGKRARYSVFAGYKEKTKTGLTKDKLMKNKRGKIVSKKAHAAGKKRYQKLAPWISAVSQ
AKKALGISGFATVGGKTAQGKAIYVKAKAL--
>MMETSP0228-20121227|16550_1/33-136 [subseq from] frtnfY1/CVXmZOuzgKzY60+JR/0 Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKAMKAMKKSKVATGKRARSTVFRGNKVETVGGLTKEKLVKNTRGRIVSKAQSAHAKRQFSKVGAWHQAVMA
AREALGVRGFCLIGGKSMQGRALYAKAKALYK
>MMETSP0228-20121227|11226_1/22-124 [subseq from] hfFa4SbkznCbiOfbT4L1R/RWukY Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKAMKAM-KSKVATGKRAKSSVFRGTKVKTRGGLTKDKLVKNKRGKIVSKAASAFSKSQFSKVGAWHKAVMA
ARTALGVQGFCAIGGKSAQGQALYAKAKSLYK
>MMETSP0228-20121227|26524_1/5-90 [subseq from] j3aKDGsmWZs4FJ9K6jqg7eSWly4 Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
-KAMKR--KSVIATGKRAKASVFRGTKVKTSGGLTKDKLVRNKQGKVVSKARSAHAKKXYTSVGAWNKAVMA
ARKALGLQGFCAIGGKS----------
>MMETSP1338-20131121|1509_1/4-107 [subseq from] P9ugmxegwM1bEq2fRnMQq62+zig Pelagodinium beii
MKAMKAMKKSMIATGKRAKASVFRGTKTKTSGGLTKEKLTKNKAGRIVSKARSSASKKLYAKIGLWNKAVMT
ARKELGIKGFQAVGGKSAQGKALYAKAKSLYK
>MMETSP1338-20131121|10_1/1-84 [subseq from] wGzUbOZAw/7rNxJ1WWxF3SDzydA Pelagodinium beii
----------KRAKASVFRGTKSKTSGGLTKDKLTKNKAGRIVSKAASAASKKKYAKIGLWNKAVMT
ARKELGIKGFQAVGGKSAQGKALYAKA-----
>MMETSP1338-20131121|16910_1/51-153 [subseq from] ByZgJVxf/tnUoPF8AgNhxFXpwGs Pelagodinium beii

MKAMKAMKKSNVAKGKRAKSSVFRGSKAKTSGGLTKDKLTKNKNGKVVSKAMSAASKKRYASIGKWNKAVQA
ARKALGIKGFVAIGGKTAEGKALYAKAKSIY-
>MMETSP1440-20131203|31_1/11-99 [subseq from] obndTQRJ1CvQNZQlHh4zXm1hMyc Polarella glacialis, Strain CCMP2088
-KVMKAMKKSTIATNKRAKASVFRGTKVKTVGGLTKEKLTKNRTGKVVSKAASAAAKKKYANLGGWTKAVQA
ARKALGVKGFAAIGGKSA----------
>MMETSP1440-20131203|2163_1/2-84 [subseq from] RkM4L2hSHK7MY86kKxZ47uc/YhM Polarella glacialis, Strain CCMP2088
----------VFRGSKAKTSGGLTKEKLTKNKGGKVVSKASSARAKRAYSKIGGWNTAVMA
ARKALAIKGFCAIGGKSAQGKALYAKAKSLYR
>MMETSP0227-20121206|22336_1/1-83 [subseq from] 9XxJIYLjJL7M8DMEWdWU+TgavFU Polarella glacialis CCMP1383
----------VFRGTKVKTGGGLTKDKLTKNRVGKVVSKASSAAAKKKYAKLGGWTKAVQA
ARKALXIKGFAAIGGKSAQGKALYXKAKSLYT
>MMETSP1440-20131203|182_1/3-86 [subseq from] HaWVBP/zabyraeCfW2vCw2qb1VU Polarella glacialis, Strain CCMP2088
---KKAMKKSTIAKGKRAKSSVFRGSKAKTSGGLTKDKLTKNKAGKVVSKASSAVAKKNYSKIGAWNKACSG
ARKALAIKGFCAIGG----------
>MMETSP1338-20131121|128090_1/135-229 [subseq from] Ijre49x83LZ1ZWj8SZo4Bq2gAns Pelagodinium beii
MKAMKAMKVSKVAKGKNAKSAVFRGTKEKTSTGMTKDKLMKNKRGKIVTKGKSAVAKKRFAAIGGWAKAVAE
ARKALGIKGFVAIGGKTAEGKAF---------
>MMETSP1338-20131121|128090_1/25-112 [subseq from] Ijre49x83LZ1ZWj8SZo4Bq2gAns Pelagodinium beii
MKAMKAMKVSKVAKGKNARRVVFAGKKEKTSTGMTKDKLIMNKRGKIVTKAASAVAKKRYASIGNWSKAVAE
ARKALGIKGFVAIVKD----------
>MMETSP1110-20121109|43514_1/45-145 [subseq from] 5/aVv1u02Uadl1lDQVokMi8Dqrs Symbiodinium, Strain CCMP421
-KPMKAMKISKIAKGKRAKVSVFKGYKEKTSSGMTKDKLTLNKNGRVVSKAQSALAKKRNS-LGKWNKAVSE
ARKALGIKGFCAVGGKTNQGKALYVKAKSLY-
>MMETSP1439-20131203|125299_1/24-126 [subseq from] 6dSDuh8wyvQxvHnXvXW0549u9zM Gonyaulax spinifera, Strain CCMP409
MKAMKAKKVSVIARGKMARAAVFRGSKQKTQGGMTKDKLVKNRHGKIVSKAASVRAKKAFANLKAWADAVKK
ARKALNLTGFVAIGGKSATGKALYAKAKALL-
>MMETSP1462-20131121|9570_1/45-146 [subseq from] okpTAsaJLiXTYCgW2hzW9PrHAC4 Brandtodinium nutriculum, Strain RCC3387
MKAMKKKRVSVVAAGKRAKSSVFSGRKQKTVGGLTKANLVKNKTGKIVSKKMSARAKQAYAGIKLWADAVKA
ARKALGLTGFVAIGGKSAAGRALYAKAKTL--
>Scrippsiella-trochoidea-CCMP3099-20130930|292959_1/34-134 [subseq from] w/3KTLM3tkTEkFG1hFIU7ojvGxA Scrippsiella trochoidea
CCMP3099
-KAMKAKKKSVIAKGKRARASVFLGRKEKTQSGLTQGQLMKNKAGKIVSKKASARAKKAYASAKKWVDAVKA
ARKALGLTGFVAVGGKSAVGKALYAKAKSL--
>Scrippsiella-trochoidea-CCMP3099-20130930|55203_1/60-160 [subseq from] cu7ARq8qwRjfEhR6PqP+8/Euakw Scrippsiella trochoidea
CCMP3099
-KAMKVKKVSVIAKGKRARAVVYSGRKQKTKSGLTQAQLTKNKVGKIVSKKASAAAKKNYAKAKAWVDAVKA
ARKALGLTGFVAVGGKSAAGKALYAKAKSL--
>Scrippsiella-trochoidea-CCMP3099-20130930|34776_1/31-131 [subseq from] cvQpljb7XP7w1CSY6odZhlEfpgw Scrippsiella trochoidea
CCMP3099
-KVMKAKRVSTIAKGKMARAAVFSGRKAKTVGGMTQATLTKNKHGKIVSKKASARAKRAYASIKAWADAVKA
ARKALGITGFVAVGGKSAAGKALYAKAKSL--
>Scrippsiella-trochoidea-CCMP3099-20130930|248899_1/21-115 [subseq from] B9338sG2JiVNhg5tPrk9h87tl2w Scrippsiella trochoidea
CCMP3099
-KAMKTKRISNIAKGKNARAVVFSGKKAKTMSGMTKAGLTKNKNGRIVSKKASAHAKRAYATIKAWADAVKA
ARKALGLTGFVAIGGKSAAGKALY--------
>MMETSP0766_2-20121228|147592_1/6-107 [subseq from] TJIbfyI6c9iodOcuOglTAP1RMaY Gambierdiscus australes
-KGMKLKKASVVAKGVRAKSTVFAGRKQKTSGGLTKDKLVKNKLGKVVSKARSAVSKVRYASIKKWAETVKA
ARKALGITGFVAIGGKSAEGKALYAKAKAMM-
>MMETSP0766_2-20121228|147592_1/185-286 [subseq from] TJIbfyI6c9iodOcuOglTAP1RMaY Gambierdiscus australes
-KGMKLKKVSVVAKGVRAKSSVFAGRKEKTVGGLTKDKLVKNKLGKVVSKARSAVSKARYASLKKWADMVKA
ARKALGITGFVAIGGKSAEGKALYAKAKAMM-
>Kryptoperidinium-foliaceum-CCMP1326-20130916|413584_1/147-248 [subseq from] ODy461jEzGqPaZ0ooMHCwZHZxxQ
Kryptoperidinium foliaceum CCMP1326
-KAMKAKRVSVVAKGVRAKSSVFSGRKVRTSGGLKKENLTQNRLGKVVSKKASAAAKTRYAKAKAWAEAVKA
ARKALGITGFVAIGGKSASGKALYAKAKSLF-
>MMETSP0766_2-20121228|64176_1/53-154 [subseq from] wA6F1jQDWYN+i97A4/fOyuAK7Y8 Gambierdiscus australes
MNAMKVKKTSAVATGVRAKSTVFLGKKTKTVGGLTKDKLVKNKLGKIVSKAASAVAKKRYASAKTWAEAIKA
ARESLGCKGFVAIGGPSAEGKALYAKAKSL--
>Scrippsiella-hangoei_like-SHHI_4-20140214|276844_1/206-307 [subseq from] EzyuW/lGPraIkAazuTmuayfImkQ Scrippsiella hangoei-like,
Strain SHHI-4
-KSMKAKKVSKVARGKGAKTKVFKGTKEKTVSGLKKESLTQNKFGKVVSKKSSARAKTNYANFKTWIDAVRA
ARKALGLTGFIPVGGTSAAGKALYAKAKSLL-
>Scrippsiella-hangoei-SHTV5-20131105|128628_1/341-442 [subseq from] 2di19EM7oyOhn/0xUpnnBXLMj+s Scrippsiella hangoei SHTV5
-KAMKVKKVSKVARGKGAKTKVFKGTKEKTVSGLKKESLTQNKFGKVVSKKSSARAKTNYANFKTWIDAVRA
ARKALGLTGFIPVGGTSAAGKALYAKAKSLL-
>Scrippsiella-hangoei_like-SHHI_4-20140214|276844_1/49-146 [subseq from] EzyuW/lGPraIkAazuTmuayfImkQ Scrippsiella hangoei-like,
Strain SHHI-4
-KSMKAKKVSKVARGKGAKTKVFKGTKEKTVSGLKKESLTQNKFGKVVSKKSSARAKTNYANFKTWIDAVRA
ARKALGLTGFIPVGGTSAAGKALYAKX-----
>MMETSP1098-20130426|2018_1/29-129 [subseq from] NmfCnAnXCppw+NN7kWllvY5u6D8 Spumella elongata, Strain CCAP 955
-KSMKVKKVSKVARGKGAKTKVFKGTKEKTVSGLKKESLTQNKFGKVVSKKSSARAKTNYANFKTWIDAVRA
ARKALGLTGFIPVGGTSAAGKALYAKAKSL--

>Peridinium-aciculiferum-PAER_2-20130926|15088_1/58-158 [subseq from] n+zrnBsAWJ3Xnknl4OhaCNdR1OY Peridinium aciculiferum PAER_2
-KAMTVKKVSKVARGKGAKTKVFKGTKEKTVSGLKKESLTQNKFGKVVSKKSSARAKTNYANFKTWIDAVRA
ARKALGLTGFIPVGGTSAAGKALYAKAKLL--
>Scrippsiella-hangoei_like-SHHI_4-20140214|276844_1/409-489 [subseq from] EzyuW/lGPraIkAazuTmuayfImkQ Scrippsiella hangoei-like, Strain SHHI-4
-KAMKVKQVSKVARGKGAKNKVFKGTKEKTVSGLKKESLTQNKFGKVVSKKSSARAKTNYANFKTWIDAVRA
ARKALGLTGF----------
>Scrippsiella-hangoei_like-SHHI_4-20140214|271989_1/49-135 [subseq from] YespCMqkc5ltfa5L0oiDc/cVARs Scrippsiella hangoei-like, Strain SHHI-4
-KAMKVKQVSKVARGKGAKTKVFKGTKEKTVSGLKKESLTQNKFGKVVSKKSSARAKTNYANFKTWIDAVRA
ARKALGLTGFIPVGGT----------
>Emiliania-huxleyi-379-20130905|30213_1/9-109 [subseq from] 7351iiwDqcCQDxL3rsCPs1tjlpo Emiliania huxleyi 379
-KAMKAKKVSVVARGRGAKARVFSGIKEKTSSGLKKDSLMKNKFGKVVSKKRSAVAKKNFAKLKTWSDAVKA
ARKALGLTGFVAIGGKSAAGKALYAKAKSL--
>MMETSP0693-20131125|2761_1/32-132 [subseq from] zX+SNFTTyLCJx7M5fOBGu5BJI60 Thalassionema nitzschioides, Strain Unknown
-KAMKAKKVNVVARGRGAKARVFSGIKEKTSGGLKKDGLTKNKFGKVVSKKRSAVAKKNFAKLKTWSDAVKA
ARKALGLTGFVAIGGKSAAGKALYAKAKSL--
>Scrippsiella-hangoei-SHTV5-20131105|127326_1/165-265 [subseq from] AhCcqAiIvYxWbRe5/577H6eJ0ak Scrippsiella hangoei SHTV5
-KAMKAMKVSVVARGRGAKAKVFSGIKEKTSGGLKKDGLMKNKFGKVVSKKRSSVAKKNFAKLKPWSDAVKA
ARKALSLKGFVAIGGKSAAGKALYAKAKSL--
>Peridinium-aciculiferum-PAER_2-20130926|177032_1/47-147 [subseq from] CEyC1gZieKTdGR7F2Bh4eYO/s0I Peridinium aciculiferum PAER_2
-KGMKAKKVSIVARGKGAKARVFSGKREKTSSGLKKDALMKNKFGKVVSKKKSAVSKQNYAKLKAWSEAVKA
ARTELGLTGFVAIGGKCAAGKALYAKAKSM--
>Scrippsiella-hangoei-SHTV5-20131105|38936_1/25-125 [subseq from] 8oQ6QaLIp/ARordTvrTHFYJ2Sak Scrippsiella hangoei SHTV5
-KGMKAKKVSIVARGKGAKSRVFSGKREKTSSGLKKDALMKNKFGKVVSKKKSAMSKQNYAKLKAWGDAVKA
ARIELGLTGFVAIGGKCAAGKALYAKAKSM--
>Scrippsiella-hangoei-SHTV5-20131105|15542_1/52-152 [subseq from] MPd929QZeENkVrN+IJjt+fwnzyc Scrippsiella hangoei SHTV5
-KGMKAKKVSIVARGKGAKARVFSGKREKTSSGLKKDALMKNKFGKVVSKKKSAVSKQNYAKLKAWGDAVKA
ARKELGLTGFVAIGGKSAAGKALYAKAKAM--
>Peridinium-aciculiferum-PAER_2-20130926|177447_1/19-119 [subseq from] 9/jWm8SYMsMlBqLWvtRmWkW3k08 Peridinium aciculiferum PAER_2
-GAMKAKKVSVVARGKGAKARVFSGKKEKTSSGLKKEALTKNKFGKVVSKKKSAVSKKNYAKLKAWGDAVKA
ARKELGLTGFVAIGGKSAAGKALYAKAKAM--
>Scrippsiella-hangoei-SHTV5-20131105|80557_1/19-119 [subseq from] gndN/vQx9/4QYe8Xmkh9opF7O1s Scrippsiella hangoei SHTV5
-GAMKTKKVSVVARGKGAKARVFSGKKETTSSGLKKEALTKNKFGKVVSKKKSAVSKKNYAKLKAWGDAVKA
ARKELGLTGFVAIGGKSAAGKALYAKAKAM--
>Scrippsiella-hangoei_like-SHHI_4-20140214|280545_1/202-286 [subseq from] hIPDj/VHvBsM8zSFAlwnH9U5Ln8 Scrippsiella hangoei-like, Strain SHHI-4
----------AALQVMSGKKEKTSSGLKKDALMKNKFGKVVSKKKSAMSKQNYAKLKAWGDAVKA
ARNELGLTGFVAIGGKCAAGKALYAKAKSM--
>Scrippsiella-hangoei_like-SHHI_4-20140214|10777_1/29-129 [subseq from] agh10ARZBtKSXEePKZdnx43sDNU Scrippsiella hangoei-like, Strain SHHI-4
-KAMKTKKVSVVARGRGAKARVFRGAKEKTVGGMKKDDLTKNKFGKVVSKKASARAKKNYANLKAWGDAVKA
ARKALGLTGFVAIGGKSAAGKALYAKAKSL--
>Scrippsiella-hangoei_like-SHHI_4-20140214|52578_1/55-155 [subseq from] ld1kjd8Pwt1vfVUqPwO1Tk6lfrI Scrippsiella hangoei-like, Strain SHHI-4
-KAMKTKKVSVVARGRGAKARVFKGAKEKTVGGMKKDDLTKNKFGKVVSKKASARAKKNYANLKAWGDAVKA
ARKALGLTGFVAIGGKSAAGKALYAKAKSL--
>Scrippsiella-hangoei-SHTV5-20131105|128628_1/29-126 [subseq from] 2di19EM7oyOhn/0xUpnnBXLMj+s Scrippsiella hangoei SHTV5
-KAMKTKKVSVVARGRGAKARVFRGAKEKTVGGMKKDDLTKNKFGKVVSKKASARAKKNYANLKAWGDAVKA
ARKALGLTGFVAIGGKSAAGKALYAKX-----
>Peridinium-aciculiferum-PAER_2-20130926|12766_1/55-155 [subseq from] vWbv7/oZRGJdtQdHMSaV2uOiIFE Peridinium aciculiferum PAER_2
-KGMKTKKVSVVARGRGAKSRVFRGAKEKTVGGMKKDDLTKNKFGKVVSKKASARAKKNYANLKAWGDAVKA
ARKALGLIGFVAIGGKSAAGKALYAKAKSL--
>Scrippsiella-hangoei-SHTV5-20131105|123233_1/51-148 [subseq from] DXDq6wKlKUoKyR1gPj0XlZSHzTI Scrippsiella hangoei SHTV5
-KAMKAKKLSVVARGRGAKAKVFKGSKEKTAGGLNKDALTKNKFGKVVSKQASARSKKNYATLKAWGDAVKA
ARKELNLTGFVPIGGKSAAGKALYAKX-----
>Peridinium-aciculiferum-PAER_2-20130926|179711_1/52-149 [subseq from] qlYRyyW3z9wDXOFecROycLmaCks Peridinium aciculiferum PAER_2
-KAMKAKKLSVVARGRGAKARVFKGSKEKTAGGLNKDALTKNKFGKVVSKKASARSKKNYATLKAWGDAVKA
ARKELNLTGFVPIGGKSGAGKALYAKX-----
>Scrippsiella-hangoei_like-SHHI_4-20140214|46468_1/53-150 [subseq from] x1IDjz7tLnjMb6i/rvoZyJXMupQ Scrippsiella hangoei-like, Strain SHHI-4
-KAMKAKKLSVVARGRGAKARVFKGSKEKTAGGLNKDALTKNKFGKVVSKKASARSKKNYASLKAWGDAVKA
ARKELNLTGFVPIGGKSAAGKALYAKX-----
>Crypthecodinium-cohnii-Seligo-20130904|6346_1/21-121 [subseq from] M/nLVySvEM4tZr4gk90Ba/Gkq08 Crypthecodinium cohnii Seligo
-KAMKAMKVSVIAKGPLAKSRVFKGKKEKTVGGLTKDKLTKNKFGKVVSKSASARAKKNFANLKSWGLAVKA
ARKALGITGFVPIGGKTASGKALYAKAKSL--
>Crypthecodinium-cohnii-Seligo-20130904|197744_1/27-127 [subseq from] ogxk0ZQs2gBQUokTEhmR+tzPsms Crypthecodinium cohnii Seligo

-KTMKAKRVSTVARGHGAKSRVFKGKKEKTSGGLTKDKLTLNKHGKVVSKARSATSKKNYANLKSWGLAVQA
ARKALGITGFVPIGGKSASGKALYAKAKSL--
>Crypthecodinium-cohnii-Seligo-20130904|20441_1/59-159 [subseq from] Ul6aGCsJF/AJsjetWoBQnxEILzs Crypthecodinium cohnii Seligo
-RVMKAKKVTVVAKGRGAKSRVFKGKKEKTVGGLTKDKLTLNKYGKVVSKGKSAAAKNRFANLKSWGLAVQA
ARKALGITGFVPIGGKSASGKALYAKAKSL--
>Crypthecodinium-cohnii-Seligo-20130904|44450_1/50-150 [subseq from] d93BRWtmPd0Zi6C9Z9vWu0ZYw/w Crypthecodinium cohnii
Seligo
-RVMKAKKVTVVARGRGAKSRVFKGKKEKTVGGLTKDKLTLNKYGKVVSKSASARAKKNYANLKSWGLAVQA
ARKALGITGFVPIGGKSASGKALYAKAKSL--
>Durinskia-baltica-CSIRO_CS-38-20140214|15479_1/41-142 [subseq from] 6Fwapd8D26PYKDIu0QBiRXUpFBU Durinskia baltica, Strain
CSIRO CS-38
MKAMKAMKKSMIAKGKRARSAVFSGRKVKTTAGLTKEKLTKNKSGRIVSKAASARAKRAYASVKKWAEAVRS
ARKALGLKGFVPIGGKSAAGKALYAKVKSM--
>Durinskia-baltica-CSIRO_CS-38-20140214|30639_1/2-102 [subseq from] uqijzvct8GucQFz5DA+RTTCtuPg Durinskia baltica, Strain CSIRO
CS-38
-KAMKAMKKSMIAKGKRARSAVFSGRKVKTTAGLTKEQLTKNKSGRIVSKAASARAKRAYASVQKWAEAVRS
ARKALGLKGFVPIGGKSAAGKALYAKVKSM--
>Durinskia-baltica-CSIRO_CS-38-20140214|232871_1/202-298 [subseq from] QRwUZkbwV4Lp4D0xOcKXbnA5Jc0 Durinskia baltica, Strain
CSIRO CS-38
MKAMKAMKKSMIAKGKMARSAVFSGRKAKTYTGLTKEKLTKNKSGKIVSKAASARSKRAYASVKKWAEAVRS
ARKALGLKGFVPIGGKSAAGKALYA-------
>Durinskia-baltica-CSIRO_CS-38-20140214|141957_1/65-166 [subseq from] 9mGUE+19s4nMytXE9BEGsiapV+E Durinskia baltica, Strain
CSIRO CS-38
MKAMKAMKHSMIAKGKRAKSSVFAGRKVKTSGGLTKEKLTRNKAGRVVSKAASARAKKAYAHVKKWADAVKA
ARKSLGITGFVPIGGKSGAGKALYAKAMSI--
>MMETSP1098-20130426|1944_1/13-113 [subseq from] qaHaV7TqqHb2vm/whexMGHtkekc Spumella elongata, Strain CCAP 955
-KAMKAMKVSKIAKGPRAKVSVFLGGKEKTSSGLTKAALTKSKTGRIVSKKKSAVGKKNYAGAKAWVDACKA
ARKALGLTGFVPVGGKSAPGKALYAKAKAL--
>Scrippsiella-hangoei-SHTV5-20131105|100999_1/2-91 [subseq from] oQtspQR8+4ZpRNzV/8XLDH+kgSI Scrippsiella hangoei SHTV5
---------SKIAKGPRAKVSVFLGGKEKTSSGLTKAALTKSKTGRIVSKKKSAVGKKNYAGAKAWVDACKA
ARKALGLTGFVPIGGKTATGKALYAKX-----
>Scrippsiella-hangoei-SHTV5-20131105|127122_1/343-443 [subseq from] 9YhyOkq4ylfXTH1sZkyiqpgmTIg Scrippsiella hangoei SHTV5
-KGMKAKKISKIAKGPRAKVSVFHGGKEKTSSGLTKANLTKSKQGRIVSKKKSAVAKKNYAGAKAWADACKA
ARKALGLTGFVPIGGKTAAGKALYAKAKSL--
>Scrippsiella-hangoei_like-SHHI_4-20140214|10117_1/70-167 [subseq from] 9in8tsBqZ61WEhSltxRByB4CO7E Scrippsiella hangoei-like,
Strain SHHI-4
-KGMKAKKISKIAKGPRAKVSVFHGGKEKTSSGLTKANLTKSKQGHIVSKKKSAVAKKNYAGAKAWADACKA
ARKALGLTGFVPIGGKTATGKALYAKA-----
>Scrippsiella-hangoei-SHTV5-20131105|127122_1/79-176 [subseq from] 9YhyOkq4ylfXTH1sZkyiqpgmTIg Scrippsiella hangoei SHTV5
-KGMKAKKISKIAKGPRAKVSVFQGGKEKTSSGLTKANLTKSKQGRIVSKKKSAVAKKNYAGAKAWADACKA
ARKALGLTGFVPIGGKTATGKALYAKX-----
>Scrippsiella-hangoei-SHTV5-20131105|17573_1/76-173 [subseq from] oFusNZmio6ec8K5u8IUt3GmsuhM Scrippsiella hangoei SHTV5
-KGMKAKKISKIAKGPRAKVSVFHGGKEKTSSGLTKANLTKSKQGRIVSKKKSAVAKKNYAGAKAWADACKA
ARKALGLTGFVPIGGKTATGKALYAKA-----
>Peridinium-aciculiferum-PAER_2-20130926|185667_1/190-287 [subseq from] fasYfmWJjZdd8BQEpmdoD2e6WeI Peridinium aciculiferum
PAER_2
-KGMKAKKISKIAKGPRARVSVFQGGKEKTSSGLTKAGLTKSKDGRIVSKKSSAVAKKNYVGAKAWADACKA
ARKALGLTGFVPIGGKTAAGKALYAKX-----
>Peridinium-aciculiferum-PAER_2-20130926|98847_1/15-115 [subseq from] 8LES/kt0zsUtCx6cbRfMjf+XFyQ Peridinium aciculiferum
PAER_2
-KGMKKKTVGKIATGPRMRAAVFRGGKAKTVGGLTKELLTKTTRGRLVSRKASAASKKKYVNAKKWCDACSA
ARKALGITGFVAIGGQTAAGKALYAKAKSL--
>MMETSP1098-20130426|77787_1/19-119 [subseq from] 3r/Vbqm0JWIYMOUzzCka+HncdiU Spumella elongata, Strain CCAP 955
-KGMKTKTVSKIAWGPRMRAVVFRGGKAQTPGGLTKESLTKTTRGRLVSKKASAASKKKYVNAKRWCDACSA
ARKALGITGFVPIGGQTAAGKAFHAKAKSL--
>Scrippsiella-hangoei_like-SHHI_4-20140214|188751_1/19-119 [subseq from] QgUscd+R+dslUIEmaN+uJK6ivkI Scrippsiella hangoei-like,
Strain SHHI-4
-KGMKTKTVSKIAWGPRMRAAVFRGGKGKTPGGLTKESLTKTTRGRLVSKKASAASKKKYVNAKRWCDACSA
ARKALGITGFVPIGGQTAAGKAFHAKAKSL--
>Azadinium-spinosum-3D9-20130829|217600_1/24-121 [subseq from] MgrnJpooZfLKRsoYaV5aOkfK2yU Azadinium spinosum 3D9
----GTAKKSKIARGKRARSAVFNGKKEKTGTGLTKANLIKNKNGKIVSKAASARAKRNWASIKKWAEAVKL
ARKALSITGFVPVGGKTPAGKALYAKTKSL--
>Pseudo_nitzschia-fradulenta-WWA7-20140214|208619_1/35-129 [subseq from] z4rdFoQWDO7tpwuENppC3/Ttn0Q Pseudo-nitzschia
fraudulenta, Strain WWA7
----GTAKKSKIARGKRARSAVFNGKKEKTGTGLTKANLIKNKNGKIVSKAASARAKRNWASIKKWADATKL
ARKSLSITGFVPMGGKTPAGKALYAKC-----
>Azadinium-spinosum-3D9-20130829|40682_1/24-121 [subseq from] DfyNf3HSrSYt6oSwVtGIZytskPk Azadinium spinosum 3D9
----GTAKKSKIARGKRARSAVFNGKKEKTGTGMTKADLIKNKRGKVVSKASSARAKRAYASIKKWAEAVKL
ARKALSITGFVPVGGKTPAGKALYAKTKSL--
>Azadinium-spinosum-3D9-20130829|53259_1/19-109 [subseq from] cIgkGQB4pJp0h17oEk8XuZCUkI8 Azadinium spinosum 3D9
----------IAKGKRARASVFNGKKEKTGTGMTKADLIKNKRGKVVSKASSARAKRAYASIKKWAEAVKL
ARKALSITGFVPVGGKTPAGKALYAKTKSL--
>Azadinium-spinosum-3D9-20130829|255773_1/96-191 [subseq from] Iz3/5QRq4bN0p0lHP3Lo8/wW6hs Azadinium spinosum 3D9

-----KKEKSKIARGKRARSAVFNGKKEKTGTGLTKANLIKNKNGKIVSKAASARAKRNWASIKKWADATKL
ARKSLSITGFVPMGGKTPAGKALYAKTKS---
>MMETSP0766_2-20121228|31847_1/30-131 [subseq from] Qu47quQzrdkDJHGFyBcsEjX1Ji0 Gambierdiscus australes
NVTMKAKRVSTIAKGARAKFVVFSGRKTKTSGGLTKDKLVKNKQGRIVSRARSASSKRHYANLRKWANAVKA
ARKALAFKGFVPIGGQSAQGKALYVKAKSL--
>Durinskia-baltica-CSIRO_CS-38-20140214|29336_1/86-188 [subseq from] 3myVp9eNn8Fiz4PgpjGQ2C1Rtmo Durinskia baltica, Strain
CSIRO CS-38
MKAMKATKKSMVAQGVRAKVVVFAGQKKETSGGLTKNHLTKNKKGKIVSKAMSAASRKNFASLKKWAVATKK
ARKELGIKGFCPVGGKTAQGKALLAKIRAIL-
>Durinskia-baltica-CSIRO_CS-38-20140214|25566_1/89-191 [subseq from] u8aRYgdDMoEDQghiZ6fE0FpN1XY Durinskia baltica, Strain
CSIRO CS-38
MKAMKVTKKSKVAQGVRAKVVVFAGHKDETSGGLKKSNLTKNKKGKIVSKAMSAASRKNFASLKKWADATKK
ARKELRIKGFCPVGGKTAQGKALLAKVRAIL-
>Durinskia-baltica-CSIRO_CS-38-20140214|26499_1/22-123 [subseq from] B7HrNootq+hGSvbV7sbqYpvvr3I Durinskia baltica, Strain
CSIRO CS-38
MKAMKAMKKSIIAKGKLAKVVVFAGHKQKTVGGLTKDKLTKSKSGKIVSKAASAAGKKNFAKLKKWAMATKQ
ARKALGIKGFCPVGGKSAQGKALLAKVRSI--
>Durinskia-baltica-CSIRO_CS-38-20140214|232869_1/45-146 [subseq from] XV3rt6yMrPeu2UgrQQXqYaNdnB4 Durinskia baltica, Strain
CSIRO CS-38
KKAMKAMKKSIIAKGKLAKVVVFAGHKKKTVGGLTKDKLTKSKSGKIVSKAASAAGKKNFAKLKKWAMATKQ
ARKALGITGFCPVGGKSAQGKALLAKVRSI--
>Durinskia-baltica-CSIRO_CS-38-20140214|232871_1/8-106 [subseq from] QRwUZkbwV4Lp4D0xOcKXbnA5Jc0 Durinskia baltica, Strain
CSIRO CS-38
MKAMKAMKKSVIAKGKLARAVVFAGHKKKTVGGLTKEKLTKSKRGKIVSRAASAAGKKNFEKLKKWAMATKQ
ARKALGITGFCPVGGKSAQGNALLAKV-----
>Kryptoperidinium-foliaceum-CCMP1326-20130916|405412_1/384-485 [subseq from] 7/oCR3k9w52Ue3MPvYrA8OfhSW4
Kryptoperidinium foliaceum CCMP1326
MKGMKAMKTSKVATGKRARAVVLAGSKAQTSGGLTKSDLMRNKKGKIVAKSSSAASKRGFAKLKKWADATKK
ARKELKLTGFVPVGGKTPKGKALLAKVRAI--
>Kryptoperidinium-foliaceum-CCMP1326-20130916|405412_1/210-309 [subseq from] 7/oCR3k9w52Ue3MPvYrA8OfhSW4
Kryptoperidinium foliaceum CCMP1326
MKGMKAMKTSKVATGKRARAVVLAGSKAQTSGGLTKSDLMRNKKGKIVAKSSSAASKRGFAKLKKWADATKK
ARKELKLTGFVPVGGKSIVGKRL--ALRAA--
>Kryptoperidinium-foliaceum-CCMP1326-20130916|408955_1/460-561 [subseq from] wjy1MsFwh+rGncn6C06rhAOIPtY Kryptoperidinium
foliaceum CCMP1326
ASAMKAMKTSKVATGKRARAVVLLGYKAQTSGGLTKSDFKRNKKGKIVAKSSSAASKKNFAKLKKWADATKK
ARKELKITGFVPLGGKTPAGKALLAKVRAI--
>Kryptoperidinium-foliaceum-CCMP1326-20130916|411954_1/155-256 [subseq from] X8S9lqIyc5OlsE9+ahpkyvZPJfI Kryptoperidinium
foliaceum CCMP1326
MKSMKAMKESKVAKGMFRKVVVFKGGKAETSGGLTKSELKKNKAGRIVSKKASMAAKRNFSQLKKWADATKA
ARKVLGITGFVPCGGTSGKGKALLAKVRAI--
>Glenodinium-foliaceum-CCAP1116_3-20130913|99223_1/2-92 [subseq from] 4ypUyanY0pp+u8bWnMF1pcqs4F8 Kryptoperidinium
foliaceum
---------SKVAKGMFRKVVVFKGGKSETSGGLTKSALTKNKVGRIVSKKASMAAKRNFSQLKKWADATKA
ARKVLGITGFVPCGGTSGKGKALLAKVR----
>Glenodinium-foliaceum-CCAP1116_3-20130913|49703_1/53-154 [subseq from] 0Dk96y7kQyXDEaWHX2+hjknC3yQ Kryptoperidinium
foliaceum
-SMMKKKAVSKVAKGKLAKVVVFAGRKAKTSGGLAKSDLTKNKSGKIVSKKASAKGKNNFARLKKWADATKA
ARKALGITGFVPVGGRSGAGKALLAKVRAIL-
>Glenodinium-foliaceum-CCAP1116_3-20130913|14706_1/53-153 [subseq from] l3p3snKNq7aqTzcxXu4dQIiAdq8 Kryptoperidinium
foliaceum
-MVMKAKAVSNVATGKLAKLVVFNGRKAKTQGGLTKADLSKNKDGKVVSKKRSANGKKAFAKLKKWADATKA
ARKALGIRGFVPVGGSTAAGKALLAKVRSI--
>Glenodinium-foliaceum-CCAP1116_3-20130913|16498_1/80-180 [subseq from] CjKabA1JgbgZl+9Wgap8Pd3Why4 Kryptoperidinium
foliaceum
-MVMKAKAVSKVAKGKLAKAVVFSGRKAKTQGGLTKGDLAKNKDGKVVSKKRSANGKKAFAKLKKWADATKA
ARKALGIRGFVPVGGSTAAGKALLAKVRSI--
>MMETSP0251_2-20131101|1247_1/59-161 [subseq from] tXsPK8qeyI0EmoixsGBWEPE4uTA Prorocentrum micans, Strain CCCM 845
MKVMKAKRVSKVAKGKRQRAQVLRGSKEKTASGLTKDQLMRNKRGKIVSKKAHATRRKLYEKIKVWAECVNA
ARKALNLKGFVAINGKLAEGKALYAKAKALY-
>Prorocentrum-minimum-CCMP1329-20131001|248570_1/13-115 [subseq from] oMx05d7iq33wepxNDrv/lhJsjNw Prorocentrum minimum
CCMP1329
MKAMKAKRVSKVAKGKLGRAQVLRGTKEKTSTGLTKDKLMKNAKGKIVSKKQSAAGKKMYSNIKAWNECVAQ
ARKALNLKGFVAINGKKVEGKALYTKAKALY-
>Azadinium-spinosum-3D9-20130829|257580_1/200-301 [subseq from] kI0YbGC+HQhOpPedQMcH3f0zdX4 Azadinium spinosum 3D9
-GSMKSKRVSKIAKGKLARAMVLRGSKEKTATGLRKDKLFRNKFGKIVSKKASAASKKRYQTAKVWADCVKA
ARKSLNVTGFVAINGKRAEGKALYAKSKSLY-
>Azadinium-spinosum-3D9-20130829|91391_1/12-113 [subseq from] DKodFUgESUCwvr4651rPOzldks4 Azadinium spinosum 3D9
-GSMKSKRVSKIAKGRLARAMVLRGSKEKTTAGLTKDKLFRNKFGKIVSKKASAASKKIYQTAKVWADCVKA
ARKSLNVTGFVAINGKRAEGKALYAKSKSLY-
>Azadinium-spinosum-3D9-20130829|257580_1/12-113 [subseq from] kI0YbGC+HQhOpPedQMcH3f0zdX4 Azadinium spinosum 3D9
-GSMKSKRVSKIAKGKLAKAMVLRGSKEKTASGLTKDNLFRNKFGRIVSKKASAASKKRYQTAKVWADCVKV
ARKSLNVTGFVAMNGKRAEGKALYAKTKSLY-

>Pseudo_nitzschia-fradulenta-WWA7-20140214|13991_1/5-106 [subseq from] 6uOaY0csnAxcBtv6qElWrrCFf+8 Pseudo-nitzschia fraudulenta, Strain WWA7
-GSMKSKRVSKIAKGKLAKAMVLRGSKEKTASGLTKDNLFRNKFGRIVSKKASAASKKRYQTLKVWADCVKA
ARKSLNVTGFVAINGKRAEGKALYAKSKSLY-
>Azadinium-spinosum-3D9-20130829|259969_1/407-508 [subseq from] r0ufCGf2FvY4SLxcZZ/ediw0SxQ Azadinium spinosum 3D9
SRIMKSKRVSIIARGPRARAAVLAGRKAKTGTGLTKEKLMKNKYGKIVSKASSARAKKAYAAIKTWIEAVKT
ARKSLAVKGFVAINGKSAEGKALYAKAKAI--
>Azadinium-spinosum-3D9-20130829|259969_1/34-135 [subseq from] r0ufCGf2FvY4SLxcZZ/ediw0SxQ Azadinium spinosum 3D9
SRIMKSKRVSIIARGPRARAAVLAGRKTKTGAGLTKEKLMKNTYGKIVSKAASARAKKAYAAIKTWIEAVKT
ARKSLAVKGFVAINGKSAEGKALYAKAKAI--
>Pseudo_nitzschia-fradulenta-WWA7-20140214|40203_1/33-133 [subseq from] 5vabxEkyAEDx2vnPIRT57B39bvM Pseudo-nitzschia fraudulenta, Strain WWA7
-GVMKSKRVTIIARGKFARSAVLSGKKQKTVSGLTKDKLMRNKFGKIVSKAASAARKKAYANFKKWGDAVKA
ARKSLALSGFVAVNGKSAEGKALYAKAKSL--
>Prorocentrum-minimum-CCMP2233-20131001|255074_1/21-122 [subseq from] tIWGI03lZz3hUWFKCdxZ4HiGhbQ Prorocentrum minimum CCMP2233
MKAMKKKK-SIIAKGSLAKSQVLKGAKMKTSGGLTKDKLMKNKRGKVVSKKQHAAGKKNYAGLKKWTECVTK
ARSALGLKGFVAVNGKKPEGKALYAKAKALY-
>Prorocentrum-minimum-CCMP2233-20131001|255074_1/186-285 [subseq from] tIWGI03lZz3hUWFKCdxZ4HiGhbQ Prorocentrum minimum CCMP2233
---KGXTE-CVIAKGSLAKSQVLKGAKVKTSGGLTKDMLMKNKRGKVVSKKRHAAGKKIYAALKKWTECVTM
ARSVLGLKGFVAVNGKKPEGKALYAKAKALYS
>Prorocentrum-minimum-CCMP1329-20131001|55121_1/2-99 [subseq from] Bvij8RdvUvUVtMp8seIQ7FzpiGE Prorocentrum minimum CCMP1329
-----KKK-SIIAKGSLAKSQVLKGAKVKTSGGLTKDMLMKNKRGKVVSKKRHAAGKKIYAALKKWTECVTM
ARSVLGLKGFVAVNGKKPEGKALYAKAKALYS
>Prorocentrum-minimum-CCMP1329-20131001|9142_1/21-108 [subseq from] 4Clj1LMJnLhTXNZm9o+E3n0DMrI Prorocentrum minimum CCMP1329
MKAMKKKK-SIIAKGVQAKSQVLKGAKVKTSGGLTKDSLMKNKRGKVVSKKKHAAGKKLFAAVKKWGECVAK
ARVALGLKGFVAINGKK----------
>Prorocentrum-minimum-CCMP1329-20131001|1425_1/52-134 [subseq from] nQe2zjC92jFl0mxEBMbPlWwNKl0 Prorocentrum minimum CCMP1329
MKAMKKKSTSTIAKGVQAKSQVLKGAKVKTSGGLTKDGLMKNKRGKVVSKKKHAAGKKLFAAVKKWAECVAK
ARAALGLKGFV----------
>Prorocentrum-minimum-CCMP2233-20131001|27970_1/52-132 [subseq from] 313EbT5sZSNWyl7feALDiv9ec6k Prorocentrum minimum CCMP2233
MKAMKKKSTSTIAKGVQAKSQVLKGAKVKTSGGLTKDSLMKNKRGKVVSKKKHAAGKKLFASVKKWAECVAK
ARAALGLKG----------
>MMETSP0797-20121207|34022_1/26-129 [subseq from] 8rxjy8u3BuQ9kv5BQmfsxhmKajw Dinophysis acuminata
MKAMKAKRVSIIAKNKRAKVTVFHGRKTKTVGGLTKADLFKNKRGKIVSRKQSLSAKKRFSKLGAWNTALAT
ARKALDVKGFVAVNGKTAQGRALYAKAKSFYS
>MMETSP0797-20121207|62635_1/29-132 [subseq from] r/CLawspDUDtzHOq/foHMri9yJU Dinophysis acuminata
MKAMKAKRVSIIAKNKRAKVTVFHGRKTKTMGGLTKADLLKNKRGKIVSRKQSLSAKKRFSKLGAWNTALAT
ARKALDVKGFVAVNGKTAQGRALYAKAKSFYS
>MMETSP0797-20121207|50181_1/1-102 [subseq from] q2oSdWXwMN+YDJZfwVXFyqcvDBA Dinophysis acuminata
MXKM-XMKKSVIAKGKRAKVSVFKGTKLKTSGGLKKSDLKKNKAGKIVSAKRSAASKKSKKKIAAWGAAMSK
ARKALGIKGFCPCGGKSAQGKALLAKVRSFY-
>MMETSP0786-20121207|8618_1/5-106 [subseq from] TDnQAWhpqdVOstAE1uPB2RsRwr8 Thalassionema frauenfeldii
MKXM-AMKXSXIAKGKRGKVSVFKGSKVKTSGGLKKSDLKKSKSGKIVSAKASAAAKKSKKKIAAWAAATSK
ARKALGVKGFCPVGGKTPKGKALLAKVRSFY-
>MMETSP0798-20121207|5160_1/24-125 [subseq from] 132l1UfidEYSK0lJeEdyq2AHfGA Myrionecta rubra, Strain CCMP2563
MKKM-AMKKSIIAKGKRGKVSVFKGSKVKTSGGLKKSDLKKSKSGKIVSAKASAAAKKSKKKIAAWAAATSK
ARKALGVKGFCPVGGKTPKGKALLAKVRSFY-
>MMETSP0797-20121207|12805_1/24-125 [subseq from] G/GO9mL5m+VoC3AVFscbbzRmcw0 Dinophysis acuminata
MKKM-AMKKSIIAKGKRGKVSVFKGSKVKTSGGLKKSDLKKSKSGKIVSAKASAAAKKSKKKIAAWAAATSK
ARKALGIKGFCPCGGKSAQGKALLAKVRSFY-
>MMETSP0795-20121207|5064_1/26-127 [subseq from] Sj3tpAsTT8pUPSzFbYYO4icd71M Amoebophrya sp.
MKKM-AMKKSVVAKGKRAKSSVYRGTKVKTVGGLKKSDLKKNKAGKVVSAKASAAAKKRKKKIVAWGAAVVK
ARKALGVKGFCPIGGKTVQGKALLAKVRSFY-
>Crypthecodinium-cohnii-Seligo-20130904|38288_1/27-129 [subseq from] DKDbtTc8WSKR+8Yz9EomudfVbLo Crypthecodinium cohnii Seligo
MKAMKAMKKSVVARGHGAKARVFKGKKEKTTGGLTKEKLTINKAGKVVSKSKSAQAKSRYEGFKQWGLAISK
ARNALGITGFQTIGGSTAVGKALYAKARSLY-
>MMETSP0784-20121206|24030_1/192-294 [subseq from] QooQpuH4e6F03YQEBJq+XtnOzks Gymnodinium catenatum
MKGMKKKRTSIVARGPRAKASVFNGRKEKTTSGLTKAMLFKNKHGKVVSKAASAAGKKAFKNISKWTVALAS
ARKALNIKGFVPVNGNTAEGKALYAKAKSLY-
>MMETSP0784-20121206|24030_1/52-143 [subseq from] QooQpuH4e6F03YQEBJq+XtnOzks Gymnodinium catenatum
-RG----------KGLTGEAAVFNGAKEKTASGLTKAMLHKNKRGKVVSKASTAAGKKAFRHIAAWLKAVAS
ARKALNITGFVGINGKTAEGKALYAKAKSLY-
>MMETSP0661-20131031|60730_1/50-153 [subseq from] /ORg7k0oZShh+rs2xw4+/3+BwlI Alexandrium margalefi, Strain AMGDE01CS-322
MKAMKRPASSKIARGKRARHAVFRGAKEKTVSGLTQDKLMKNKRGKVVSKAASAAAKRRSPHFEVWGRAVAA
ARKALNITGFCAVGGKTLEGKALYIKAKALYR

>MMETSP0661-20131031|11644_1/23-125 [subseq from] 8U8BmwneoBWar09Zk3TVih0O0Js Alexandrium margalefi, Strain AMGDE01CS-322
TKAMKRRGTSKIARGKRAKSSVFRGTKEKTIGGLTKDKLMKNKRGKVVSKAASAAARKRGSHFEAWGRAVSA
ARKALGITGFCVVGGESVEGRALYLKAKSLY-
>MMETSP0661-20131031|10938_1/23-125 [subseq from] ireDC5PbE5OXlwCp5MQqmrsT0IA Alexandrium margalefi, Strain AMGDE01CS-322
MKAVRRRGTSKIARGKRARSSVFRGTKEKTIGGLTRDKLMKNKRGKVVSKAASAAARKRGSHFECWGRAVSA
ARKAMGITGFCVVGGESVEGRALYLKAKSLY-
>MMETSP0766_2-20121228|59706_1/23-125 [subseq from] 9nwuHBXEvE+Rt1XDBnPwJpHpg2k Gambierdiscus australes
-AAMKAKKVSIIAKGKRARSIVFKGRKEKTYTGLTKAQLMKNRTGKLVTKKMHALGKRCYARISNWTKACQQ
AKKEMGIEGFCLVGGSSQAGKTLYAKAKAIYT
>MMETSP0766_2-20121228|8046_1/76-178 [subseq from] zAxtQvSJ6b5d/VeOzXSUHbWegN4 Gambierdiscus australes
-EAMKKKTASVIAKGRLARAAVFGGSKKKTSSGVTKRQLMKNRTGRLVTKKKHKLGKKSYSKISSWGKAVQK
ARKELGVTGFCLVGGSTTSGKALYAKAKEIYA
>MMETSP0766_2-20121228|45654_1/22-123 [subseq from] Jaa2YyObQOw231gYSJuV3TSl+hU Gambierdiscus australes
-KAMKSKAKDKVASGRRARMVVFRGGADKTQTGITKAKLMQNKDGKVVTKKKHAAGKKAYQKISGWTKACQK
AKKELGITGFCVIGGKSQEGKALYAKAKAIY-
>MMETSP0766_2-20121228|67063_1/86-187 [subseq from] d0Io4GXEUpxkA9yar0//N6FCPC4 Gambierdiscus australes
-KPMKARRQSVIARGKFARLVVYTGRKERTKTGLTKASLMKNKKGKVVTVKMHAQGKQNYGRISGWTEACQQ
AKKALGIVGFSPVGGKSLQGKALYVKAKSIY-
>Glenodinium-foliaceum-CCAP1116_3-20130913|267469_1/239-339 [subseq from] We8EI3xU6S0GDRNRxjiEHMNa1mY Kryptoperidinium foliaceum
-EADAAMKVSKVATGKRAKVQIFKGRKERTAGGLKKSDLMTSKSGKIVSKARSAHMKKRYATLKKWADAVKS
ARKALGITGFVPVGGQTAVGKALLAKVRAT--
>Glenodinium-foliaceum-CCAP1116_3-20130913|46884_1/57-157 [subseq from] /de8GN1RGegW5FWXnRWTqnHDLos Kryptoperidinium foliaceum
-KPMRAMKVSKVATGKRAKAGVFKGRKERTAGGLKKSDLITSKSGKIVSKARSAHMKKRYATLKKWADAVKS
ARKALGITGFVPVGGQTAVGKTLLAKVRAT--
>Glenodinium-foliaceum-CCAP1116_3-20130913|267469_1/60-159 [subseq from] We8EI3xU6S0GDRNRxjiEHMNa1mY Kryptoperidinium foliaceum
-KPMRAMKVSKVATGKRAKVQIFKGRKERTAGGLKKSDLITSKSGKIVSKARSAHMKKRHASLLARADAVKS
ARKALGITGFVPVGGQTAVGKALLAKVRA---
>MMETSP1462-20131121|115_1/1-88 [subseq from] fVdKvmGz7eR0lnWbW+ryRTKuCB8 Brandtodinium nutriculum, Strain RCC3387
----------IAVGRRAKSSVFSGRKERTTGGLTKASLVKNRHGKVVSKKRSAFMTQRYAGLKAWAEAVKT
ARHALGLTGFVGIGGKTAAGKALYAKA-----
>Peridinium-aciculiferum-PAER_2-20130926|107080_1/33-134 [subseq from] Hc8QWy6IHH2YZK24RST1eMB7VqY Peridinium aciculiferum PAER_2
MKGKKAKKVSKVAKGKYAKFAVFSGSKERTGGGIRQDGLVKNSRGRVVSKKMSAAGKRNYEKLKAWCLAVKA
ARLALGLTGFVAIGGKTAAGKALYAKAKSL--
>Scrippsiella-hangoei-SHTV5-20131105|71966_1/30-131 [subseq from] pH9TizQcw2s7Wr2pDZryrKv6aBw Scrippsiella hangoei SHTV5
MKGTKAKKVSKVAKGKYAKFAVFSGSKERTGGGIKQDGLVKNSRGRVVSKKMSAAGKRNYEKLKAWCLAVKA
ARLALGLTGFVAIGGKTAAGKALYAKAKSL--
>Amphidinium-carterae-CCMP1314-20130924|158230_1/12-109 [subseq from] exhhYDp+fRgsK04zPZ5rm6ykAaU Amphidinium carterae, Strain CCMP1314
MKPMKPMKTSKIG----SKISVFKGTKEKTSSGLKKSDYIKNKNGKIVSKKKSALAKQNYGSLKAWANAVKA
ARKSLGVTGFVAIGGKSPEGKALYAKAKSL--
>Amphidinium-carterae-CCMP1314-20130924|158230_1/193-278 [subseq from] exhhYDp+fRgsK04zPZ5rm6ykAaU Amphidinium carterae, Strain CCMP1314
----------G----SKISVFKGTKEKTSSGLKKSDYIKNKNGKIVSLKMSALAKKRYGSLKAWADAVKA
ARKSLGVTGFVAIGGKSPEGKALYAKAKSL--
>MMETSP0689_2-20121128|46804_1/71-165 [subseq from] r72MYYWwN9dCR+mHbvN2+77gLps Amphidinium massartii
MKAMKPMKKANVG----SKVSVFNGKKEKTKTGKKKSDFVMNKRGKVVHKKASAAAKKKYASLKAWGDAVKA
ARKALGLTGFVPIGGKTPEGKALYAKX-----
>MMETSP1440-20131203|268343_1/143-240 [subseq from] 2xGxzo225zDmLlVFyqEHZX0W09s Polarella glacialis, Strain CCMP2088
-KAVKAMKKSTIANGKRRKVVVFRGTKLKTSGGLKRADLIKSKTGKVVSRKASAAGKKAYASIKGWTDAVQK
ARKELGVKGFVAIK----KGTALYKAAKAIY-
>MMETSP1440-20131203|268343_1/47-135 [subseq from] 2xGxzo225zDmLlVFyqEHZX0W09s Polarella glacialis, Strain CCMP2088
-KAVKAMKKSIIAKGKRGKSSVFRGTKVKTSGGLKKSDLIKSKSGKVVSRKSSAAGKKAYGNIKGWTVAVQK
ARKELGVKGFVAIK----KGTA----------
>MMETSP1440-20131203|372_1/8-89 [subseq from] PUFj9VyFkWWHENgYaxAajH5mDps Polarella glacialis, Strain CCMP2088
-KAMKAMKKSTIASGKRCTVSVFKGTKVKTSGGLKKADLIKSKTGRVVSRKGSAAGKKAYANIKGWTDAVQQ
ARKELGVKGFV----------
>MMETSP1338-20131121|16872_1/59-143 [subseq from] 2SNX0lA9Y+sQPjLl8gFZ18jic9w Pelagodinium beii
MKAMKAMKVSKIAKGKRGKSSVFRGTKEKTTSGIKKSDLMKSKSGKIVTRKAHAAGKKAYANIKGWTAAVQK
ARKELGVKGFVAV----------
>MMETSP0227-20121206|16099_1/43-140 [subseq from] euTOKBJ3kUIXZ0SgzaifjqgF9gM Polarella glacialis CCMP1383
-KTMKAMKVSKIARNKNAKATVFKGVKEKTASGLKKSDLMKSKSGKIVSKSQHAAGKKSYANIKGWTVAVQQ
ARKNLGVKGFVAVK----KGTALYKAAKAIY-
>MMETSP1440-20131203|14896_1/11-108 [subseq from] vOG6xV76d+C/tNNLpAxuWcJf/0w Polarella glacialis, Strain CCMP2088
-KVMKAMKVSNVAKGKLAKNVVFKGNKEKTKGGLKKSDLMKNKTGRIVTRKQHAKGKKAYKNISGWTAAVQK
ARKDLGVKGFVAIK----KGTALYKAAKAIY-
>MMETSP0227-20121206|25894_1/2-94 [subseq from] hcs7KAzRuHoz6Fbk8+v9q+VdV48 Polarella glacialis CCMP1383
----KAMKVSNVAKGKLAKNVVFKGNKEKTKTGLKKSDLMKNKTGRIVTRKQHAKGKKAYKNISGWTAAVQK

ARKDLGVKGFVAIK----KGTALYKAAKA---
>MMETSP0227-20121206|28724_1/11-95 [subseq from] X4aGJkhjnUB2+q6SSJiuA5svTu8 Polarella glacialis CCMP1383
-KVMKAMKVSXVAKGKLAKVVVFKGGKDKTKTGLKKSDLVKSKTGKIVSRKQHAQGKKAYSRIKAWTDACQK
ARKDLGIKGFVAIK----------
>MMETSP1440-20131203|694_1/36-116 [subseq from] tA/2gHzXceo1bu/2TYwAnBg8kf0 Polarella glacialis, Strain CCMP2088
-KVMKAMKVSKIAKGKLAKLVVFKGGKEKTATGLKKSDLMKTKTGKVVTRKQHAAGKKAYAHIKAWTTACQK
ARKDLGIKGF----------
>Azadinium-spinosum-3D9-20130829|259994_1/48-145 [subseq from] eJaontZCHTh73lCMx9XSebrAmfU Azadinium spinosum 3D9
-KVMKAKKVSQIAKGVRARATVFTGGKEKTYTGLKKTDLMKSKTGKIVTRKSHAAGVKAYKHIKGWTAAVQK
ARKDLGIKGFVAVK----KGTALYTAAKAIY-
>Pseudo_nitzschia-fradulenta-WWA7-20140214|208229_1/43-140 [subseq from] kd8AfLjm4OT7PiDcvVUobtjPC5g Pseudo-nitzschia
fraudulenta, Strain WWA7
-KVMKAKKVSQIAKGVRARAAVFTGGKAKTYTGLKKTDLMKSKTGKIVTRKSHAAGVKAYKHIKGWTAAVQK
ARKDLGIKGFVAVK----KGTALYTAAKAIY-
>MMETSP1338-20131121|125355_1/145-239 [subseq from] smcK9fWG3yLMzsNuZHVmJBrRvcg Pelagodinium beii
MKAMKAMKAMKVGK----KYSVFSGNKEKTSGGLKKSDLTKNKRGKVVSKNSSAHGKKAYKHIKAWTDACQK
ARKELGIKGFVAIK----KGTPFYKAAKAYY-
>MMETSP1338-20131121|125355_1/44-125 [subseq from] smcK9fWG3yLMzsNuZHVmJBrRvcg Pelagodinium beii
MKSMKAMKAMKVGK----KFSVFSGNREKTSGGLKKSDLTKNKRGKVVARSLSAHGKKAYKNIKAWTAACQK
ARKELGIKGFVAMK----------
>Ceratium-fusus-PA161109-20140214|213847_1/207-309 [subseq from] pDt41hkHm08OnVfmY1ofbcrWUI4 Ceratium fusus, Strain
PA161109
-KVVKVKTVSIIATGPRARMSVWRGTKTKTVGGLTKESLIKNKFGAIVSKKASAAKKKAYATIKIWAEAVKK
ARKQLKITGFQAIGGKTAQGKKLYEKAKALYK
>MMETSP1462-20131121|11489_1/16-116 [subseq from] 394VYkT/fN5bBk21onSgjPhpUrI Brandtodinium nutriculum, Strain RCC3387
-TKAAAKKPTTIARGAHAKVMVLRGTKTKTVGGLTKKDLVKNKYGKVVSKAASQASKAAYRKIKAWATAVQK
ARKSLKLKGFVPIGGKSAAGKKLYKAAKAL--
>Karenia-brevis-CCMP2229-20130916|260105_1/38-139 [subseq from] 2+2ueNBEbpUImQYRQjW5vtJASRU Karenia brevis CCMP2229
-AAMKAKGMSKIARGKLAKKQVFLGHKQKTGGGLKRDGILKNKYGKLVAKKKSAVAKKNYSKFEAWIKAVQK
ARKSLGLTGFFGINGKSAQGKAIYAKAKAIY-
>Karenia-brevis-Wilson-20130916|39900_1/7-109 [subseq from] BqDAIN81ROHzZ/YE5DV4yW20HPI Karenia brevis Wilson
TTVMKKKSTSTIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGRIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYAKAKAIYT
>Karenia-brevis-SP1-20130916|225068_1/7-109 [subseq from] vhD5p53uczPAkVTHhbu/G3GoQAA Karenia brevis SP1
TTVMKKKSTSIIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGRIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYAKAKSIYT
>Karenia-brevis-CCMP2229-20130916|58785_1/7-109 [subseq from] 8+I5N7Th6af2XAc2zchkieg312Q Karenia brevis CCMP2229
TTVMKKKSTSIIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGRIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYAKAKAIYT
>Karenia-brevis-SP1-20130916|217403_1/7-109 [subseq from] Je6/pg959bfCoKykOy93hy2jc+U Karenia brevis SP1
TTVMKKKSTSIIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGRIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGRAVYAKAKAIYT
>Karenia-brevis-Wilson-20130916|345171_1/16-118 [subseq from] 31Q2ux4ujYScc+++B7qeD2MygR4 Karenia brevis Wilson
TTVMKKKSMSTIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGKIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGRAVYAKAKAIYT
>Karenia-brevis-SP3-20130916|254186_1/218-320 [subseq from] P9YCZYgXBWpcdqGPYv1tZEkYwdQ Karenia brevis SP3
TTVMKKKSMSTIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGKIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYAKAKAIYT
>Karenia-brevis-Wilson-20130916|40379_1/15-117 [subseq from] sVtfVYtr5//c0Nf3jklMgvfMToY Karenia brevis Wilson
TTVMKKKSMSTIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGKIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYAKAKSIYT
>Karenia-brevis-SP1-20130916|5218_1/7-108 [subseq from] /DUODhzGiQi1KkbF3+nSlT3IVso Karenia brevis SP1
TTVMKKKSMSIIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGRIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYAKAKAIY-
>Karenia-brevis-SP3-20130916|254186_1/7-106 [subseq from] P9YCZYgXBWpcdqGPYv1tZEkYwdQ Karenia brevis SP3
TTVMKKKSTSIIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGKIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYAKAKX---
>Karenia-brevis-SP1-20130916|88500_1/14-113 [subseq from] yoj3AeXsBqT4Q5hSPXpjaKCpkd0 Karenia brevis SP1
TTVMKKKSMSTIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGRIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYAKAKR---
>Karenia-brevis-SP1-20130916|288451_1/7-105 [subseq from] rarlonJHfZjJhJSmhvQXyuG3zb8 Karenia brevis SP1
TTVMKKKSTSIIAKGRFAKAQVLNGSKTKTSGGITRDGITKNKFGRIVSKKKSLVSKRRYSKLEAWNKALST
ARKSLGVSGFLAVNG-SATGKAVYXEGK----
>Lingulodinium-polyedra-CCMP1738-20130920|127649_1/79-176 [subseq from] pZ9rp+DHkFvmN9pZ24wfutsjcA0 Lingulodinium polyedra
CCMP1738
TKKARNRRTSKVAKGRGAKSKVLAGLKERTVGGLQLKDIIKNRYGKVVGKKRSAHSQRN-----GWAQTVSA
ARKALGLTGFVLIN-SGLEGRALYAKAKAMYQ
>MMETSP0796-20121207|34664_1/7-103 [subseq from] SDFpAN6Xoz8o0M1qJ4yTAC1cKh0 Pyrodinium bahamense
MKVMKTM-KSKIARGKYRKALVLRGSREKTVGGLKAQDLTKNKYGKVVAKRASARAKNN-----PWAKAIAA
ARKALNLKGFVAIN-SGPEGQALYAKAKTLYN
>MMETSP0796-20121207|13200_1/59-155 [subseq from] v4WJ8A0Jp2nLodaJCKQqmN3FkCo Pyrodinium bahamense
MKAMKAR-KSKIARGRLAKAMVLRGAKEKTAGGLKAQDLTKNKYGKIVAKKRSAMGKKN-----PWMQACIM
ARKALGITGFVALN-SGPEGKALYAKAKSLYN

>MMETSP0796-20121207|59053_1/55-151 [subseq from] V5stmcxw5oR89SNlruQ5nD6m3yc Pyrodinium bahamense
MKAMKAM-KSKIARGRLAKAMVLRGSKEKTAGGLKAQDLTKNKYGKIVAKKRSAMAKKN-----PWMQACMV
ARKALGITGFVALN-SGPEGKALYAKAKSIYN
>Alexandrium-fundyense-CCMP1719-20130923|96149_1/21-117 [subseq from] oa7QiLSAXn+1ZARsXS1mE68z/5s Alexandrium fundyense
CCMP1719
MKAMKAM-KSKIARGRLSKSQVFKGSKEKTAGGLKASDIIKNKRGKYVSKKASARSKNN-----SWMQSIAA
ARKALGLVGFVAIN-KGPEGKALYAKAKALHE
>Alexandrium-temarense-CCMP1771-20130823|44480_1/36-132 [subseq from] B676qfdWN6orcCG/AjNbup1sjg0 Alexandrium temarense
CCMP1771
MKAMKPK-TSKIARGKLSKVMVYKGRREKTVGGLKASDIIRNKQGKFVSKKVSARSKNS-----SWMKATAA
ARKALGLTGFVAMN-SGPEGKALYAKAKELHR
>Alexandrium-fundyense-CCMP1719-20130923|7370_1/11-106 [subseq from] 9oDlxfdX/U6YYZ/7lIeJEOMwYWk Alexandrium fundyense
CCMP1719
MKAMKAK-TSKIARGKLRKAMVYKGRREKTSGGLKASDIIKNKRGKYVSKKSSARSKNS-----SWIKAVAA
ARKALGLTGFVAIN-SGPEGKALYAKAKALH-
>MMETSP0328-20130328|8943_1/17-113 [subseq from] ia3ZtdQyVtQDOhqHMyS6R+QQyjU Alexandrium minutum
MKAMKAMRKSMIGKGKLAKSMVYKGAKKKTVGGLKASDITKNKYGKYVSKKASARGKTN-----VWAKAIAA
ARKALGLKGFVLIN-KGPEGKALYAKAKSIM-
>MMETSP1436-20131217|462_1/55-149 [subseq from] ahK/hfkxdOWq+UFfOR6VSfBfreU Alexandrium andersonii, Strain CCMP2222
MKSMKAM-KSKIGKGRVAKSQVYKGKKEKTVGGLKASDITKNKYGKFVSKKRSAKGKTN-----VWAKAIGA
ARKALGLKGFVPIN-KGPEGKALYAKAKTL--
>MMETSP1436-20131217|1835_1/58-152 [subseq from] gL4eUc114JRqoLfe7qXM93FdRKs Alexandrium andersonii, Strain CCMP2222
MKAMKAM-KSKIGKGKLAKSMVYKGKKEKTVGGLKASDITKNKYGKFVSKKASARGKSN-----TWAKAIGA
ARKALGLKGFVPIN-KGPEGKALYAKAKTL--
>MMETSP0328-20130328|5149_1/53-147 [subseq from] g71Vw1sEcMC/PmrOom42Pmf9tdk Alexandrium minutum
MKAMKAM-KSXIGKGRLAKAQVYKGRKEKTVGGLKASDITKNKYGKLVSKKASARGKTN-----AWAQAIAA
ARKALGLKGFVLIN-KGPEGKALYAKAKSI--
>MMETSP0790-20130122|8762_1/23-117 [subseq from] E5aSVGcZhKdIXB0W9KFdwLuDDG0 Alexandrium catenella, Strain OF101
MKAMKAM-KSIVARGKLSKSQVYKGRKVKTSGGLKATDITKNKYGKFVSKKLSARGKTN-----VWAKAIAA
ARKALALKGFVPIN-KGPEGKALYAKAKTI--
>MMETSP0790-20130122|8723_1/86-180 [subseq from] ohNEO22DQbRtzhMBDNTGqJOsM9Q Alexandrium catenella, Strain OF101
MKAMKAM-KSIIAKGKLSKSQVYKGSKVKTSGGLKASDITKNKYGKYVSKKVSARSKTN-----VWAKAIAA
ARKALGLKGFVPIN-KGPEGKALYAKAKTI--
>MMETSP0790-20130122|64154_1/8-102 [subseq from] Wt1YBPYPtqhLgZrVnd1+9DxHdi4 Alexandrium catenella, Strain OF101
MKAMKAM-KSIVAKGKLSKSQVYKGRKVKTVGGLKASDITKNKYGKYVSKKASARAKTN-----VWAKAIAA
ARKALGLKGFVPIN-KGPEGKALYAKAKTI--
>MMETSP1436-20131217|118964_1/2-92 [subseq from] bxyEHI4OaTFUsoGG/ytDwnBIlA0 Alexandrium andersonii, Strain CCMP2222
---MKAM-KSKIGKGRMAKRQVYSGRREKTVGGLKASDITKNKYGKLVSKKRSAKGKTN-----AWAKAIGA
ARKALGLKGFVPIN-KGPEGKALYAKAKX---
>Alexandrium-fundyense-CCMP1719-20130923|9364_1/23-118 [subseq from] p6O69o45mXymG7Y62fBDzJCMnyo Alexandrium fundyense
CCMP1719
MKAMKAR-KSKIAKGKLSKAMVYKGAREKTVGGLKASDITKNKYGKFVSKKASARAKNN-----SWAKAIAA
ARKALGITGFVLIN-KGPEGKALYAKAKALQ-
>Alexandrium-temarense-CCMP1771-20130823|408128_1/69-164 [subseq from] mByb+53MTtskmaXZ0FPRAKS+Kbg Alexandrium
temarense CCMP1771
MKAMKAM-KAKIARGKLSKAMVYKGAREKTVGGLKASDITKNKYGKFVSKKNSARGKNN-----SWAKAIAA
ARKALGITGFVLIN-KGPEGKALYAKAKALQ-
>Alexandrium-temarense-CCMP1771-20130823|7675_1/58-153 [subseq from] Rw0G1QDpVGtgZx/47+fpfm/snEk Alexandrium temarense
CCMP1771
MKAMKAM-KAKIARGKLSKAMVYKGAREKTVGGLKASDITKNKYGKLVSKKRAARGKNN-----SWAKAIAA
ARKALGITGFVLIN-KGPEGKALYAKAKALQ-
>MMETSP0661-20131031|10705_1/49-143 [subseq from] 9ei2nzV4Cu3klxndkh+IbqrHnpQ Alexandrium margalefi, Strain AMGDE01CS-322
MKAMKAM-KSKIARGRLSKAQVYKGRKEKTVGGLKASDIIRNKYGKFVSKKASARAKNN-----RWAKAIAA
ARKALGLTGFVLFN-KGPEGKALYAKAKAL--
>MMETSP0661-20131031|9779_1/55-149 [subseq from] YurDFM/xVrVk62lARirao/6lre4 Alexandrium margalefi, Strain AMGDE01CS-322
MKAMKAM-KSKIAKGRLSKAQVYKGRKEKTVGGLKASDITRNKYGKFVSKKASARAKNN-----RWAKAIAA
ARKALGLTGFVLFN-RGPEGKALYAKAKAL--
>MMETSP0661-20131031|61714_1/29-122 [subseq from] 1H3Zcfe1Jskzn+ydHqYkD2+rCdg Alexandrium margalefi, Strain AMGDE01CS-
322
MKAMKAM-KSKIAKGRLSKAQVYKGQKEKTVGGLKASDITRNKYGKFVSKKASARAKNN-----RWAKAIAA
ARKALGLTGFVLFN-RGPEGKALYAKAKA---
>MMETSP0661-20131031|10061_1/64-158 [subseq from] V8c7jN7kKANqNHsA7pu7o0u9qAg Alexandrium margalefi, Strain AMGDE01CS-
322
MKAMKAM-KSKVAKGRLSKSQVYKGAREKTSGGLKASDITRNKYGKYVSKKVSDRARNS-----PWPKAIAA
ARKALGLTGFVPIN-KGPEGKALYAKAKAL--
>MMETSP0661-20131031|168_1/29-123 [subseq from] mgyX/X0OoyvceBgw+a0LllbqJI0 Alexandrium margalefi, Strain AMGDE01CS-322
MKAMKAM-KSKIAKGRLSKSQVYKGRKEKTSGGLKASDITRNRYGKFVSKKASARSKSS-----PWAKAIQA
ARKALGLTGFVPIN-SGPEGKALYAKAKSL--
>MMETSP0661-20131031|12557_1/32-126 [subseq from] m3oEzYl7E6GUE9pICRSvy1ZYlDk Alexandrium margalefi, Strain AMGDE01CS-
322
MKAMKAM-KSKIARGRLSMAMVYKGRKEKTVGGLKASDITRNKYGKFVSKKRSASAKSN-----VWARAIAA
ARKALGLKGFVPIN-RGPEGKALYAKAKAL--

>Alexandrium-temarense-CCMP1771-20130823|266904_1/4-93 [subseq from] DJckNEsLizTnr+4sMlQUdpfkBqY Alexandrium temarense CCMP1771
---MKAL---KIARGKMSKSMVFKGAREKTVGGLKASDIFKNKYGKLVSKKVSAIGKKN-----SWAKAIAA
ARKALGITGFVLIN-KGPEGKALYAKAKEL--
>Alexandrium-monilatum-CCMP3105-20140214|412741_1/142-236 [subseq from] c2TsMUiWB4kF+34H8xWFHovD4XI Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKVATGRLAKSMVYKGSKAKTSGGLKASDIMRNKRGKFVSKRKSAAGLNF-----PWPKAVAA
ARAALGIKGFVCIN-SGPEGKAIYDKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|412741_1/4-98 [subseq from] c2TsMUiWB4kF+34H8xWFHovD4XI Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKVAKGRLAKSQVYKGTKAKTSGGLQASDIMRNKRGKFVSKKKSAVGRNL-----PWPKAIAA
ARKALGITGFVVIN-RGPEGKALYAKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|412741_1/328-422 [subseq from] c2TsMUiWB4kF+34H8xWFHovD4XI Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKVAKGRLAKSQVYKGTKVKTSGGLQASDIMRNKRGKFVSKKKSAAGMNF-----PWPKAVAA
ARKALGITGFVCIN-SGPEGKALYAKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|56750_1/38-132 [subseq from] UFxfaY8gmJqCNDhq5Aq/x2VLfcQ Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKVAKGRMAKSQVFKGTREKTSGGLKASDIVRNKTGKFVSKKRSAASKAL-----PWTKAIAA
ARKALGLTGFVPIN-RGPEGKALYVKAKAM--
>Alexandrium-monilatum-CCMP3105-20140214|280847_1/13-107 [subseq from] Lq43eCldtFfPrKjo/3XD7Y23FFk Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKVAKGRMAKSQVFKGTREKTSGGLKASDIVRNKFGKFVSKKRSAVSKAL-----PWPKAIAA
ARKALGLTGFVPIN-RGPEGKALYVKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|1447_1/49-143 [subseq from] mEGuikgxLmMa98Wm6H7GAygjpkE Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKIAKGRMAKSQVYKGTREKTSGGLKSSDIIRNKNGKFVSKKRSAVSKSL-----AWPKAISA
ARKALGLTGFVPIN-RGPEGKALYVKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|8715_1/29-123 [subseq from] cKWIGzoljPgG8ZWCGIl2opiWWhM Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKVAKGKMAKSQVYKGTREKTSGGLKASDIIRNKSGKFVSKKRSALSKTL-----AWPKAIAA
ARKALGLTGFIPIN-RGPEGKALYAKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|410752_1/33-127 [subseq from] 4oVo/08YRqaMUm3dkvh/LVuVapM Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKIAKGRMAKSQVYKGTREKTSGGLKASDIVRNRFGRFVSKKQSAAAKNR-----PWPKAIAA
ARKALGLTGFVPIN-RGPEGKALYAKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|45994_1/58-152 [subseq from] DFm6+8BKYSpfdhJ+91MmpJj+0a4 Alexandrium monilatum, Strain CCMP3105
MRAMKAR-KSKVARGKLAKSQVYKGTREKTSGGLKASDIIRNKSGKFVSKKRSAAGKNR-----PWPKAIAA
ARKALGLTGFVAIN-RGPEGKALYAKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|57659_1/23-116 [subseq from] /L70FDqZibQLAdiVFYWpsKoK3O0 Alexandrium monilatum, Strain CCMP3105
MKAMKVK-KSKVAKGKMAKSQVYKGRRERTSGGLKASDIIRNKSGKFVSKRASAAGKNK-----PWPKAVAA
ARKALGLTGFVAIN-RGPEGKALYAKAKX---
>Alexandrium-temarense-CCMP1771-20130823|53598_1/34-128 [subseq from] Y8GXd//+JgACCZpq/gp4+B6yaoU Alexandrium temarense CCMP1771
MKAMKVM-KSKIAAGRLSKAMVYKGAREKTTGGLKASDIIRNKKGKFVSKKLSARGKSN-----PWAKATAA
ARKALGIKGFVAFN-KGPEGKALYAKIKEL--
>Alexandrium-temarense-CCMP1771-20130823|21189_1/34-128 [subseq from] wkhdecZLdGQDbeX9WHtRTZ89NYk Alexandrium temarense CCMP1771
MKAMKVM-KSKIATGRLSKAMVYKGAKEKTVGGLKASDITKNKYGKLVSKKLSARGKSN-----PWAKATAA
ARKALGITGFVPFN-KGPEGKALYAKVKGI--
>MMETSP1436-20131217|118890_1/39-133 [subseq from] afqpkYs65bZhvqcGb0aZQbvCv/A Alexandrium andersonii, Strain CCMP2222
MKAMKVM-KSKIGKGRLAKSQVFKGTKEKTVGGLKRSDIVKNKYGKFVSKKASQRGKTN-----AWATAIAA
ARKALGLRGFVLIN-KGADGKALYTKAKAI--
>MMETSP1436-20131217|21720_1/26-106 [subseq from] /fIiAWH4zHKhohrd/5UpiDNhydU Alexandrium andersonii, Strain CCMP2222
MKAMKAM-KSKVGKGRLAKSQVYKGTKEKTVGGLKRSDIVKNKYGKFVSKKASQRGKTN-----AWAKAIAA
ARKALGLRGFVLIN-K----------
>MMETSP1436-20131217|118788_1/20-114 [subseq from] G16gX5lYzxTCcu3LWRxno3uKsp0 Alexandrium andersonii, Strain CCMP2222
MKAMKAM-KSKIGKGKLAKSQVFKGRKEKTAGGLKASDITKNKYGNYVSKKRSAQGKNR-----PWPKAIAA
ARKALGLTGFVPIN-KGPEGKALYIKAKEL--
>MMETSP1436-20131217|1589_1/16-92 [subseq from] 2V2mNB8G6BUmWxbVXCnpP6wNOBg Alexandrium andersonii, Strain CCMP2222
MKAMKAM-KSKIGKGKLAKSQVYKGRKEKTSGGLKASDITKNKYGKFVSKKMSALGKGR-----PWPKAIAA
ARKALGLKGFV----------
>MMETSP1436-20131217|840_1/20-96 [subseq from] mo6StFUBZEUK/kzOVHjihA4idAk Alexandrium andersonii, Strain CCMP2222
MKAMKAM-KSKIARGKMAKVQVYKGRNEKTTGGLKASDITKNKQGRYVSKKMSALAKNR-----PWPKAIAA
ARKALGLTGFV----------
>MMETSP0661-20131031|58413_1/53-147 [subseq from] PbHqYaTAl7T9x8KNr7p9nDQeGFI Alexandrium margalefi, Strain AMGDE01CS-322
MKAMKAM-KSKIAKGRLSKAMVYTGRKVKTSGGLKASDIMRNKYGKYVSKKRSAHSKSN-----AWPKAIQA
ARKALGLTGFVAIN-KGPEGKALYAKAKAL--
>MMETSP0661-20131031|12010_1/35-129 [subseq from] 2PQAElrI52BeDmX+PNwsl5ggaFo Alexandrium margalefi, Strain AMGDE01CS-322

MKAMKAM-KSKIAKGRLSKAMVYMGRREKTSGGLQASDIMKNKYGKYVSKKRSAHSRSN-----AWPKAIQA
VRKALGLTGFVAIN-KGPEGKALYARAKAL--
>MMETSP0328-20130328|6273_1/10-98 [subseq from] B3kRSn56Ubwbj3/aPc/fy/yt0Aw Alexandrium minutum
MKAMKAKRVATXGKGKLAKSMVFKGSKAKTSGGLKASDITKNKYGKYVSKKASARGKTN-----VWAKAIAA
ARKAXALKGFVLIN-KGPEGKAL---------
>Alexandrium-monilatum-CCMP3105-20140214|115817_1/70-164 [subseq from] WRnxVrHwlCM/m0xVPoLuAtyZ1/Q Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKIATGPRAMSLVYKGAREKTSGGLRAGDIVKNKYGRFVSKRASAQGRTN-----AWAKATAA
ARKALGLTGFVPLN-KGPEGKALYAKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|412490_1/37-131 [subseq from] ZJ2m28Fu886yK1ey//dwBTPgTdc Alexandrium monilatum, Strain CCMP3105
MKAMKAM-KSKIARGSRAMSMVYKGAREKTVGGLRAADIVRNKYGMFVSKKASARGKNN-----AWAKATAE
ARKALNLKGFVPLN-KGPDGKALYEKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|31108_1/31-125 [subseq from] m6AmMTKPmUyo6/BpKEGIs/igHxc Alexandrium monilatum, Strain CCMP3105
MKAMKAM-MSKIARGPRAMSMVYKGAREKTVGGLRPADIVKNKYGMFVSKKASARGKNN-----AWAKATAE
ARKALGLTGFVALN-KGPEGKALYEKAKAI--
>Alexandrium-monilatum-CCMP3105-20140214|207054_1/64-159 [subseq from] ME77+eG7KAO6A8q1k70/F1zs+DE Alexandrium monilatum, Strain CCMP3105
MKAMKAR-TSKIARGRLAFSQVYKGRKEKTAGGLRAKDIVMNKYGKLVSKKFMAKGKAA-----PWPKAIAA
ARKALGLRGFVAIN-SGPEGKALYAKAKALM-
>Alexandrium-monilatum-CCMP3105-20140214|47997_1/45-139 [subseq from] BFx39fu7di/XolX8xUYBnijJWgg Alexandrium monilatum, Strain CCMP3105
MKAMKAG-TSKIARGRMAFSQVFKGRKEKTAGGLKAKDIVKNKYGKLVSKKFMAKGKAA-----PWPKAITA
ARKALGLTGFVAIN-RGPEGKALYAKAKAL--
>Alexandrium-monilatum-CCMP3105-20140214|9053_1/70-164 [subseq from] Xus/V4wm5PccqyZ7z4ND3iioEMY Alexandrium monilatum, Strain CCMP3105
MKAMKAR-TSKIARGRMALSQVYNGSKEKTAGGLKAKDIVKNKYGKFVSKKFRAKGKAA-----PWPKAVTA
ARKALGLTGFVAVN-RGPEGKALYAKAKAL--
>Alexandrium-monilatum-CCMP3105-20140214|57222_1/64-159 [subseq from] +EgNAAmtENcGBUmUE8T+CqwmnaI Alexandrium monilatum, Strain CCMP3105
MKAMKAR-TSKIARGRMAFLLVYKGRKEKTVGGLQAKDIVKNKYGKLVSKKFMAKGKTA-----PWPKAISA
ARKALGLTGFVTIN-KGPEGKALYVKAKALM-
>Alexandrium-monilatum-CCMP3105-20140214|194990_1/66-160 [subseq from] E5SLkUAHKqeRYaYaiMZFE8sgF4E Alexandrium monilatum, Strain CCMP3105
MKAMKVR-KSKVGKGKLGKSQVYKGAKERTVGGLRASDITRNKYGRYVTKRASAVGKTN-----AWAIAIAA
ARKELGITGFVLVN-HGPEGQALYAKAKSL--
>MMETSP0790-20130122|27734_1/67-162 [subseq from] XEcuLJB+CV2s+XMd/vchPP+6Vjo Alexandrium catenella, Strain OF101
MKVMKAT-TSVIARGRLSKSEVYKGNKAKTAGGLKAGDIIRNKYGKFVSKKVSMRFKNN-----SWAKAVAA
ARKALGLTGFILIN-RGEVGEALYAKAKTIM-
>Alexandrium-fundyense-CCMP1719-20130923|96151_1/81-175 [subseq from] aUe87Oetb98Vlhbx/7ZT3QAV3Ik Alexandrium fundyense CCMP1719
MKAMKAM-KSAIARGKLSKSMVVKGTRAKTAGGLKAADITKNKWGKLVSKKRAANGMTN-----PWTQAVAA
ARKALGLTGFVLMN-KGAKGKSLYVTAKAL--
>MMETSP0766_2-20121228|48364_1/71-167 [subseq from] rbchGaqe1cnEhEYmZNjadmfR/gI Gambierdiscus australes
MKPMKAMKVLKIARGRGAKSMVLTGVREKTVGGLKAKDLVKNKYGKVVSKLNAARGRSN-----SWSKAIAA
ARKELSLTGFVAIN-KGSDGKALYAKAKTLF-
>Lingulodinium-polyedra-CCMP1738-20130920|32164_1/50-146 [subseq from] S2x7troh9ft2RQlsrKM5LSTpACE Lingulodinium polyedra CCMP1738
MKAMKAKRVSKIAKGRGARAKVFSGKKEKTIGGLQAKDLTKNRYGKVVAKKRSAIGRAN-----PWAKAIAA
ARKALGLKGFVAIN-GGPEGKALYAKAKSIY-
>Lingulodinium-polyedra-CCMP1738-20130920|18953_1/59-155 [subseq from] PA6FTEMtR1c6lppj29Na/3WOT64 Lingulodinium polyedra CCMP1738
MKAMKAKRVSKIAKGRGARAKVFSGKKEKTIGGLQAKDLTKNKYGKVVAKKRSAIGRAN-----PWAKAIAA
ARKALGLKGFVAIN-GGPEGKALYAKAKTIY-
>Alexandrium-monilatum-CCMP3105-20140214|101830_1/17-113 [subseq from] nUXl6wzqG/oqD7WEStI8c3yMzT8 Alexandrium monilatum, Strain CCMP3105
MKAAKKTRVSKIARGKRSKAMVCKGSKEKTAGGLQASDITRNRYGKLVAKKASERSKNN-----AWPRAIAS
ARRALGLTGFILIN-KGPEGKALYAKAKEIH-
>Ceratium-fusus-PA161109-20140214|196309_1/53-148 [subseq from] la/sahS3jrCo0EP1gkvS6jzVArk Ceratium fusus, Strain PA161109
MKMMKKT-MSKIAKGRGAKSKVLLGQKEKTVGGLRAGDLIRNKWGKVVSKKRSAKGKMS-----GWMKACSA
ARKALGLTGFVVIN-RGPEGKALYAKAKVIY-
>Alexandrium-monilatum-CCMP3105-20140214|56203_1/110-204 [subseq from] 2Wqe/vipi7E+Yc7Vrwlmxx9YOI8 Alexandrium monilatum, Strain CCMP3105
LKATRAR-RLRVARGRMAKSQVVKGTKDKTAGGLKASDIVLNRNGKLVSKKRSAVSSSL-----PWPRAVAA
ARKALGLTGFVLIN-RGPEGKALYAKVKAL--
>Alexandrium-temarense-CCMP1771-20130823|49793_1/45-141 [subseq from] rajoD7y+7+dSvZHBNrujpNBQ4tc Alexandrium temarense CCMP1771
MKAMKAR-TSKIARGRCAKSLVYKGAREKTVGGLKASDIMRNKYGKYVSKKRSAQGKNS-----SWSKAISD
ARKALGMTGFVLIN-FWPEGEVLYAKAKALHD
>MMETSP0661-20131031|49894_1/79-174 [subseq from] TazIa0IWu1M4/JiK8dv1uI9R7II Alexandrium margalefi, Strain AMGDE01CS-322
MKSRKAM-KSKVARGRMSKVLVAKGAREVTFGGLHASDIIRNKWGRYVSKKRSSQGKTN-----AWAQATAA
ARKSLGLTGFVVFN-RGPEGKAFYEKVKALM-

>MMETSP0790-20130122|27832_1/73-167 [subseq from] U/OdKLmZoFklmF0Zik9u0ezSQLA Alexandrium catenella, Strain OF101
MKAMKAM-KSKVAKGRLCKSQVVKGRQERTSGGLKVADITTNKYGKFVYKRKSAHGKTN-----AWARAIEA
ARKALGITGFVLIN-KGPQGQALYVKAKTL--
>MMETSP0790-20130122|10760_1/56-150 [subseq from] g3recKJB/ZET2D6GT4XEJOl6f4s Alexandrium catenella, Strain OF101
MKAMKAR-KSKIARGAKRKSQVIKGRMEKTSGGRKASDICTNKYGKLVFKKASAHSKNL-----RWPKAVKA
AREALHITGFVLLN-KGPEGKALYAKAKLI--
>Alexandrium-temarense-CCMP1771-20130823|43928_1/12-106 [subseq from] xnhEOTJnpYAw9F678U7wDtrk7iU Alexandrium temarense CCMP1771
-KAMKAKKPPKIARGKMRKAMVYKGTREKTPSGLKASDITKNKYGVLVSKKKAALGKKN-----PWTLAVAA
ARKDLGITGFVLMN-RGLKGKALYAAAKAL--
>MMETSP0228-20121227|47215_1/12-106 [subseq from] J7qWyTcf7ZMezrEp8U3Mn/BkNUo Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKGMKAAKKPIRAKGRMARSQVLRGLKGKTASGLTASDLTKNKAGRIVSKKRSAVAKRN-----PWNAAVAA
ARKALGLTGFVAIN-SGPVGKALYAKARA---
>Lingulodinium-polyedra-CCMP1738-20130920|203580_1/62-159 [subseq from] UiufUkXAcqO9HxDqanuTwHvPMsQ Lingulodinium polyedra CCMP1738
MKAMKVKRTSIIARGRGAKAKVLTGAKVKTGGGLLAKDLVRNKYGKVVVKKASVKGRVN-----PWMKATAA
ARKALGIKGFVPMN-SGPEGKALYAKAKSLYK
>Ceratium-fusus-PA161109-20140214|213844_1/661-758 [subseq from] isn1KND2iT6PHQbCWpmSnOwhWZQ Ceratium fusus, Strain PA161109
TKKMKVKRVSKIARGHGAKARVLKGKKEKTVGGLTASDLFRNKRGKVVSKKQSAKMRNS-----PWIKALAE
ARKSLGLTGFVAVN-SGPEGKAPYAKAKVIYT
>Ceratium-fusus-PA161109-20140214|108585_1/43-139 [subseq from] Fr+/JODP6YmuiFnOacaEIAxNe5o Ceratium fusus, Strain PA161109
MKKIKIKRVSKIARGHGAKSRVLSGKKAKTVGGLTASDLFRNKRGKVVGKKQSAKMRNS-----PWIKALAE
ARKSLGLTGFVLIN-SGPDGRALYAKAKVVY-
>Ceratium-fusus-PA161109-20140214|104785_1/44-140 [subseq from] fYsIwbU0PsTO0u9zeCurk2jfbZ0 Ceratium fusus, Strain PA161109
MKGMKKKRVSKIARGFGAKARVLAGKKEKTAGGLTANDLYRNKRGKVVNKKRSFKMKNH-----PWMKALSE
ARKALGLVGFVAIN-KGLDGKALYAKAKVIY-
>Azadinium-spinosum-3D9-20130829|156937_1/16-111 [subseq from] P7VP2lcQPCUq/gj+JTpjRS0UpjU Azadinium spinosum 3D9
-KTMKGKRVSKIARGRSAKSQVLKGRKEKTGSGLTAKDLMKNKYGKVVSRKMSALGMRR-----AWPKAVQQ
ARKALSLKGFCAINGKSAEGKALYAKAKSI--
>Azadinium-spinosum-3D9-20130829|230762_1/40-135 [subseq from] 2toj8FmuvmE/sxTacWROAr6iPow Azadinium spinosum 3D9
-KTMKGKRVSKIARGRSAKSQVLKGRKEKTGSGLTAKDLIKNKYGKVVSRKMSALGMRR-----PWPKAVQQ
ARKALSLKGFCAINGKSAEGKALYAKAKSI--
>Azadinium-spinosum-3D9-20130829|91647_1/2-86 [subseq from] d+3vOKLoHMOSkx7UEbSxjsD4WV8 Azadinium spinosum 3D9
----------ARGRSAKSQVLKGRKEKTGSGLTAKDLMKNKYGKVVSRKMSALGMRR-----PWPKAVQQ
ARKALSLKGFCAINGKSAEGKALYAKAKSI--
>Azadinium-spinosum-3D9-20130829|142329_1/1-84 [subseq from] UwssyJmwkd4UEP/0eshXOtj7mzk Azadinium spinosum 3D9
----------RGRSAKSQVLKGRKEKTGSGLTAXXXXXNKYGKVVSRKMSALGMRR-----PWPKAVQQ
ARKALSLKGFCAINGKSAEGKALYAKAKSI--
>Karlodinium-micrum-CCMP2283-20140214|4851_1/23-124 [subseq from] SXcybBm0P/cEmstG53R41wkySaw Karlodinium micrum, Strain CCMP2283
-VAMKSSKKSKIAKGSRAKYVVYSGSKEKTSSGLTKSMLMKNKRGKIVSKKAHSAGKKAFKLLAKWSQAAAT
ARKELKITGFVPMNGKSAAGKAWYAKTKAIY-
>km04844-km/20-121 [subseq from] Ep0l6AH1t9Sv9qgFfCFNmwMAzyg Karlodinium micrum, Strain CCMP2283
-VAMKSSKKSKIAKGSRAKYVVYSGSKEKTSSGLTKSCLMKNKRGKIVSKKAHSAGKKAFKLLASWSQAAAT
ARKELKITGFVPMNGKSAAGKAWYAKTKAIY-
>Symbiodinium-sp-C1-20140214|96990_1/39-140 [subseq from] lanlIY5f+sBqQ7U3VGhI07z0pkM Symbiodinium, Strain C1
-KVMKAKRVGQATHGRLSKLQVFRGTKAKTTGGLMKDSLVRNKHGRVVSKRLSENGKERFKNLRPWKEACME
ARAAFNSKGFIAFNGKSLQGKALYAKAKAIY-
>MMETSP0796-20121207|21111_1/31-133 [subseq from] u4u1d3CniWO4JctRzShY/djha64 Pyrodinium bahamense
AAAMKTKSKSTIAKGSRAKFMVMRGWKVKTSGGLRKDDIMKNPRGRLVSKRNSALSKKNFAKLGFWNKAVKE
ARDSLKLSGFVAVNGPTKEGKALYAKAKALF-
>MMETSP0790-20130122|41047_1/98-199 [subseq from] 9EiUOALHsg5pZjl3MeYXMxUWKOo Alexandrium catenella, Strain OF101
-KAKKGVQTSVIAKGRLAKAMVSMGRKVKTAGGLTQESIMKNPRGRLVSRKASAAAKVRFEKLHVWNKAVSD
ARQTLRLQGFVAINGGTPMGKALYAKARALF-
>MMETSP0328-20130328|3293_1/19-115 [subseq from] Jt0U+JVCb4ujTk2Y7yDm7Dgz/QM Alexandrium minutum
--QRKVISNSQVARGRLAKALVLRGTKAKTAGGLTQEGIMKNPRGRLVSKKASAAAKVRFEKLTVWNKAVSD
ARKALRLEGFVAINGGTPMGKALYAKA-----
>Alexandrium-temarense-CCMP1771-20130823|158272_1/56-149 [subseq from] PZ7fsuYicBADTB3ffHBYRA/jumw Alexandrium temarense CCMP1771
------VQDR---GGRRAKIAVARGTKVKTAGGVTKEGIMKNPRGRLVSRKASAAAKVRFKKLQHWNQAVSE
ARKQLRLAGFVAINGGSCMGKALYCKARSLF-
>Alexandrium-monilatum-CCMP3105-20140214|28015_1/82-182 [subseq from] 3DbeDI3aa5hHtRtc/GZTYnFCuAA Alexandrium monilatum, Strain CCMP3105
RHSRK--GKTIIARGRLAKVHVFRGYKEKTSGGLTADRIGKNARGRLVSKAASAASKVRFAKLCVWNQAVSD
ARKLLRLEGFVAINGGTPMGKALYAKARSLF-
>MMETSP0766_2-20121228|43157_1/64-164 [subseq from] S0lDoGaEvRo8Y5k5c/MAdDqdMl4 Gambierdiscus australes
--AMKAKAMKKVARGKFAKALVLKGSKDRTSGGLKKDDLMRNPRGKTVSKARSAVAKARFANLYLWNEALAA
ARKAMSLTGFVAINGKTPMGKALYAKARALY-
>Crypthecodinium-cohnii-Seligo-20130904|5753_1/32-128 [subseq from] t/ca2+MWWsp+YMRL9EBlUMSyN+E Crypthecodinium cohnii Seligo
-MKMKAMKKSIVG----SKSRVYNGTKVKTIGGLTKDSLIKNKNGKVVSKKASAASKARYANVKDWAMACKA

ARKALGLTGFVPIGGKTAAGKALYAKAKSL--
>Crypthecodinium-cohnii-Seligo-20130904|17527_1/21-117 [subseq from] cMnZTMjR2OYafQabRC38x4eulhI Crypthecodinium cohnii Seligo
-MKMKAMKKSIVG----SRTRVFHGTKVKTVGGLTKDGLIKNKNGKVVSKKASAASKARYANVKDWAMACKA
ARKALGLTGFVPIGGKTAAGKALYAKAKSL--
>MMETSP0796-20121207|39858_1/44-143 [subseq from] 5elTLGxRXgFMN7xL4yni9xbJqUU Pyrodinium bahamense
--GKKAPKGGIIAVGARARLAVFTGRKQKTKTGLTKASLMKNRVGKIVSKKASAIAKKKYHAFKAWAEAVKA
ARRSFGFEGFVPVGGVSVQGRALYAKAKAL--
>MMETSP0228-20121227|49108_1/54-156 [subseq from] 6LljJqdDtTnYY1pp8hu0SX+CKYQ Protoceratium reticulatum, Strain CCCM 535
(=CCMP 1889)
MKAAKQKSKSVIATGKRARFVVFSGSKKKTRGGLTKDKLTKSKRGKVVSKAVSSKSKKTYQTLSDWTKACQR
ARVALGVKGFCPCGGKSPEGKRLYAQARAFY-
>MMETSP0766_2-20121228|54526_1/60-162 [subseq from] p82CSbM/YcMPSSfRBV0FAAS43WE Gambierdiscus australes
-KVMKAMKVSAIAKGPLGRLLVYDGRKEKTHSGITKNQLMKNRVGKLVTKKAHAKGKKAFARISAWAKAHKK
ARKELGVQGFCPMGGKTKKGQALYKMTKLIYD
>Karenia-brevis-SP1-20130916|45175_1/4-103 [subseq from] y95p8YHdEPzmrJC7glbsuEaEb4A Karenia brevis SP1
MKVMKSKKVSKIANGKRAKAAVLTGRKEKTTGGLQATDLMKNFRGKTVSKKKVALGKRNFKKLFKWNAAVKK
ARESLGLEGFVGIK----KGSPLYAKAQAFYR
>Karenia-brevis-SP3-20130916|46283_1/4-103 [subseq from] aLwTwIeM+mYAOqS/0FF88vVVLgg Karenia brevis SP3
MKVMKSKKVSKIANGKRAKAAVLTGRKEKTTGGLQATDLMKNFRGKTVSKKKVALGKRNFKKLFKWNAAVKK
ARESLGLEGFVGIK----KGSPLYAKARAFYR
>Karenia-brevis-SP1-20130916|46170_1/4-99 [subseq from] TL5/YQ2W9j2rEufW8CXX6FrcUjQ Karenia brevis SP1
MKVMKSKKVSKIANGKRAKAAVLTGRKEKTTGGLQATDLMKNFRGKTVSKKKVALGKRNFKKLFKWNAAVKK
ARESLGLEGFVGIK----KGSPLYAKAX----
>Karenia-brevis-SP3-20130916|138309_1/7-101 [subseq from] gt6CkCwbk1Xm6lDwp5jh+SazAlU Karenia brevis SP3
MKVMKSKKVSKIANGKRAKAAVLTGRKEKTTGGLQATDLMKNFRGKTVSKKKVALGKRNFKKLFKWNAAVKK
ARESLGLXXXXGIK----KGSPLYAKA-----
>Karenia-brevis-Wilson-20130916|239524_1/4-98 [subseq from] BQOiCqUsKxGTes2E5zpBYAW693s Karenia brevis Wilson
MKVMKSKKVSKIANGKRAKAAVLAGRKEKTTGGLQATDLMKNFRGKTVSKKKVALGKRNFKKLFKWNAAVKK
ARESLGLEGXXGIK----KGSPLYAKA-----
>Karenia-brevis-SP3-20130916|43362_1/1-97 [subseq from] aWFRuA8zZrj+ZT/4+OZzXVpc9lY Karenia brevis SP3
---MKSKKVSKIANGKRAKAAVLTGRKEKTTGGLQATDLMKNFRGKTVSKKKVALGKRNFNKLFKWNAAVKK
ARESLGLEGFVGIK----KGSPLYAKAQAFYR
>Karenia-brevis-Wilson-20130916|278949_1/85-173 [subseq from] 54Qu5GvbeivG38wvGXBM8iUBbUQ Karenia brevis Wilson
----------DANGKRAKAAVLTGRKEKTTGGLQATDLMKNFRGKTVSKKKVALGKRNFKKLFKWNAAVKK
ARESLGLEGFVGIK----KGSPLYAKAQAFYR
>Karenia-brevis-SP1-20130916|14312_1/4-103 [subseq from] vczCvNpbgYccigpxcNYfOC44GvU Karenia brevis SP1
MKVMKSKKVSKIANGKRAKAAVLTGRKEKTTGGLQATDLMKNFRGNMRRLEEGSLGKRNFKKLFKWNAAVKK
ARESLGLEGFVGIK----KGSPLYAKAQAFYR
>Ceratium-fusus-PA161109-20140214|37837_1/30-128 [subseq from] Be7KHW5X/Z5+KTojx7Zi7rkuk4A Ceratium fusus, Strain PA161109
MKVMKKKSVSKIAQGKLAKAMVLRGHKAKTTGGLMASDLIKNKQGKVVSKKASAASKKRHAKLGPWNKAVKM
ARKTLKITGFAAIK----KGSPLHTKAKELY-
>MMETSP0797-20121207|18343_1/14-106 [subseq from] 1fLsOpSLof8Ew/r5lnGbP9bGcaM Dinophysis acuminata
MKAM-KKTVSKIAKGRYAKAMVLRGSKAKTTGGLQAKDLIKNKYGKIVSKKASAATK-----NNPWIVAVKK
ARSALKIKGFCAIK----KGTPLYAKAKEFF-
>MMETSP0790-20130122|61422_1/53-154 [subseq from] VyCD0uZkNEZ1j/Xl6oBk82i2sNs Alexandrium catenella, Strain OF101
MKAMKVKRVSKVARGKYAYSLVLRGKKERTAGGITKDGLMKNSKGKLVSKRASAGAMRTKKFIKAWIEAVKE
ARQVLHLTGMVPVGGKTPMGKTLYLTAKRI--
>MMETSP0467-20121206|39689_1/88-190 [subseq from] XUQkXaT7ZkG9BTbDpY7dt1PxlFQ Mesodinium pulex, Strain SPMC105
MKAMKAMKAKRVAFGPLARYSVLLGRRERTKGGLTKDDLTKNRKGKIVSKRVSELGKRAYKNVEPWSEALQE
AKHAMGLSGMVLVGGKTALGQALYAKAKTIL-
>MMETSP1110-20121109|20795_1/38-140 [subseq from] 6PWmcG1uVRPJ+96Ou7ABgA6yZAA Symbiodinium, Strain CCMP421
MKVMKKKAVSQIAKGRVAKALVFKGRKEKTRGSFHQTDLMKNKRGKIVTIKKHEAGMKQYHQISAWTDAVRK
AREELGMTGFIPVNGNNWQGKALYVRTKAIF-
>MMETSP0253-20130528|34991_1/90-191 [subseq from] B1p7AtDUEZebBVjlhPgqJjeJnS0 Noctiluca scintillans
KAARKPRGSKI-ARGSMAKVLVFRGKKVKTSGGMTADMLMINKRGRVVSKKKHALGVHFFCRISTWVDAVLE
ARKQLNIVGFVPINGKTLAGRSLYVKSRAIY-
>MMETSP0784-20121206|23127_1/53-152 [subseq from] IJcRMCnSNmv56z5+WQYG4qMqh78 Gymnodinium catenatum
--NSKTRRSKIVARGRLARFMVFSGRRVKTTGGLTSEQLMKNKRGKIVSKKQHARGLRLLWRIRPWLESVKA
ARSQLKIVGFVPVNGRTLSGKALYVRAKKL--
>MMETSP1439-20131203|48425_1/26-125 [subseq from] kABVaQj1Om1tUvlV/dHMFY3oYeQ Gonyaulax spinifera, Strain CCMP409
TKAMKGRRVTETAKGRMAKVLVYRGTRAKTPGGLTAEKLTKNKDGRVVSFRKAAHGERVFAR---WREAVQD
ARKAMNAKGFVALNGKSILGKAIYLKARAIY-
>MMETSP0228-20121227|63676_1/16-95 [subseq from] LVTA9zd7kVvlq/KhAL66RsvjUdI Protoceratium reticulatum, Strain CCCM 535
(=CCMP 1889)
-KTGKAKKVCKVATGKFAKVLVFGGRREKTVGGLRAEGLVRNKRGKIVSKRRSADGRRRFGQVRGWIEALME
ARRALHITG----------
>Ceratium-fusus-PA161109-20140214|88209_1/277-371 [subseq from] 6eZamkb1WaWiiBcrUFGn+SrqIS0 Ceratium fusus, Strain PA161109
MKAMKKK-VSIIAKGRFARRQVFEGKKVRTMGGIKVADLVKNKSGKLVTKKASAKGKKS-----AWAIATKK
AREELAVTGFVPIGGKTPEGKALLAKIKX---
>MMETSP0689_2-20121128|16822_1/26-123 [subseq from] H5dAgscZpHS9eSMO3jrczkx4P7M Amphidinium massartii
AVPMKAMK---AMKTTGNRASVFSGSRVKTAGGLKKSSLTKNKRGKIVGKAASANGKKQYNRVKKYAEALKK
ARKALGVTGFVPIGGKTKAGKDLYKKMQE---
>MMETSP1374-20130617|2655_1/2-91 [subseq from] zzq+Im6Vj7i10eeP0oxIGmEMuK8 Symbiodinium Mf12.5f/clade A3

----------AKKTGSKFSVFKGSKDKTIGGLTKSKLTKNKSGKIVSKAQSANGKKAYKHIKKWADALKK
ARKDLGITGFVPVGGKTKAGKDLYKKVKEI--
>MMETSP1374-20130617|23250_1/2-91 [subseq from] /pHmMf8BDJtF9hIdht9dYXALX8o Symbiodinium Mf12.5f/clade A3
----------AKKTGSKASVFKGSKEKTVGGLKKSQLMKNKSGKVVSKATHAHGKKQYKNIKKYADALKK
ARKDLGITGFVPIGGKTKAGKDLLKKIREI--
>Amphidinium-carterae-CCMP1314-20130924|15081_1/4-92 [subseq from] CD0jH59HRya0YHQK0urFWCk2u3I Amphidinium carterae, Strain CCMP1314
----------KKVGKKASVFKGTKEKTSGGLTKSKLTKNKRGKVVSKAASAQGKKAYNHIKKWAEALKK
ARKELGIKGFVPIGGKTKAGKDLLAKVKEI--
>Amphidinium-carterae-CCMP1314-20130924|25323_1/39-127 [subseq from] 7tVUAtqEZ+GP/zjEfXZmAWRHlJI Amphidinium carterae, Strain CCMP1314
----------KKVGKKVAVFKGNREKTGGGLTKSKLKKNKDGKIVSKAASANGKKYYKHIKKWADALKK
ARKDLGITGFVPVGGKTKAGKDLLKKVKEI--
>MMETSP1374-20130617|15097_1/32-121 [subseq from] dhyTeP7NL37bhCo+KiaV3U9xyEQ Symbiodinium Mf12.5f/clade A3
----------KKATGSRASVLKGTRTKTVGGLKKSDLKKNKRGKIVSKAASDSGKKSYSRIKKYAEALKK
ARKDLGITGFVPIGGKTKAGKDLLKKMREI--
>MMETSP0689_2-20121128|15612_1/54-149 [subseq from] C0sjxHLhnF27BUa/yLSgjARvTbs Amphidinium massartii
--AMKPMK---AMKSVGAKFSVLKGSKTKTSGGLTKDKLAKNKRGKVVSKASSAAGKKAYKNIKSYADALKK
ARKELGITGFVPVGGKTKAGKDLYKKMQE---
>MMETSP0253-20130528|6017_1/35-137 [subseq from] OqJeIeioLiZrgjGIombLUbFRGqg Noctiluca scintillans
TKVMKESRVSKIARGVRMRAAVFRGLKDKTYTGLKRTDLKMNRKGKIVPKKRSALGKKNFKKLEAWGKACTA
ARKELNIVGFVAMDS-TPQGRALYAKTKAIYT
>MMETSP0253-20130528|95269_1/2-101 [subseq from] D5TxRPLDs7QgV+fCVwhCiZVwZp4 Noctiluca scintillans
--AMKEHKGSSVARGVRMRYAVYSGRKDKTYTGLKKSDLKYSKKGKIVSKKMSIASKKHFKKLEAWGKACTA
ARKELHLVGFVAMDS-TLQGRALYAKTKAIY-
>MMETSP0253-20130528|44304_1/24-124 [subseq from] Bi5lgOzKyXdxYOPOu8I99hwnWPM Noctiluca scintillans
--SMKVKRPSVVARGSRMRVAVFLGRKHKTYTGLTKNDLMKNADGRIVSKRSSALSRKRWVRCQEWNKAVVA
ARKQLSLVGFVAVNSPSPQGRALYVKARSIF-
>MMETSP0253-20130528|1359_1/15-97 [subseq from] xr4dZOzqTmAvmuM+vendJMB2iF4 Noctiluca scintillans
--SMKAKKASAIARGSRMRLVVFRGRKEKTYTGLSRSDLKLSKSGKIVTKKSSAAARKRWERPQQWIKAVVS
ARKQLSVSGFVAV----------
>MMETSP0228-20121227|39093_1/106-205 [subseq from] qM+PRu9t52skPSpr3MpixkVlPD8 Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKAMKVKKTSKIARGKRRYSQVLLGKKEKTAAGLTKADLMRNKAGRIVSRRRSAASRRSFSHIEGWVGCCKR
ARVELGLTGFVTLR----KDSEFYKRAVQLYK
>MMETSP1377-20130617|61067_1/283-380 [subseq from] A1UGUzFlry1AScbBzgag+DhR9D8 Symbiodinium type D1a
-TPMKKAKVSVIAKGKRAKSSVWQGKKEKTSTGLKKEHLKKNKDGKLVTKAKSEAGRKNFKTVARWISACKQ
ARAELGITGFVAVK----KGTPLYLKAKELY-
>MMETSP1338-20131121|28166_1/259-356 [subseq from] lfjwUXQrmZGeDy10TqbVdzdqjo0 Pelagodinium beii
KKPKKKA-ASIIAKGKRARASVFAGKKQRTKTGLKKADLKKNKDGKIVTKKKSALGKKQFKNVAKWIGACKQ
ARDQLGITGFAAVK----KGTPLYNLAKELY-
>MMETSP1110-20121109|51797_1/290-385 [subseq from] Z0KFRunfz2ItNVoTI+0RvchD99g Symbiodinium, Strain CCMP421
MKILKS---TTVAKGPRAKSQVWRGKKTKTSTGLKVDDMKKNMNGKLVSKKKSAAGRKAFGRVACWIAACKQ
ARAELGITGFVAIK----KGTALYKKAKALY-
>MMETSP1110-20121109|30830_1/183-279 [subseq from] +U5mu3RAmHJAO2ZmRXuzLcqFs6M Symbiodinium, Strain CCMP421
MKLSKPK-VCSIASGPRAKSLVWCGKKTKTSTGLKKEDMTKNKFGRLVSKRRSAAGRTAFGRVSRWIAACKQ
ARAELGITGFVAIT----KGSELYEKAKEF--
>Symbiodinium-sp-Mp-20130822|139325_1/258-354 [subseq from] aW9MgjnxyPVKLTwmwRJKg/5jkgE Symbiodinium sp Mp
MRSSKTRE--SLATGKRARSKVFLGKFDKTSTGLKKEDLKRNKHGKVVSKKKSETGKKNFAHVARWISACKT
ARAQLGITGFAAVR----KGSELYKRAKELY-
>MMETSP1377-20130617|26314_1/260-355 [subseq from] 2L4vprW4Pidm1w7HiINqdtpV5mg Symbiodinium type D1a
--AKPAKKVAKVATGKDAKAQVFQGKFTRTSTGLKKKDLKKNKYGKVVSKRLSEAGRKNFVRVACWIAACQK
ARDELGITGFVAIR----KDTELYKKAKEL--
>MMETSP1110-20121109|17998_1/258-352 [subseq from] P70o5j2d9TjLcEA9uaSUB2r9gLA Symbiodinium, Strain CCMP421
--AAKAKKV-KVATGKDAKAQVFQGKFTRTSTGLKKRDLKKNKYGKVVSKRLSEAGRKNFARVACWIAACQK
ARDELGITGFVAIR----KDTELYKKAKEL--
>Scrippsiella-trochoidea-CCMP3099-20130930|54339_1/228-326 [subseq from] ZLBbOP+0rukGoq3udvClc4JnotY Scrippsiella trochoidea CCMP3099
-KSMKHMKPMKVATGKNSKAKVFLGTFEKTPSGLKKEHLTKNKDGKIVSKKMSAIGKLRYSNVKSWIEACKK
ARAELGLTGFVATK----KGTPFYTKAREIYE
>MMETSP0224-20130122|15834_1/396-494 [subseq from] Q97nKBL/xdJRfY/AY57KsFMA0ws Togula jolla, Strain CCCM 725
MKAMKAMKASKIAKGKRGRSLVMRGLKEKTQYGLKASDLMKNRKGKVVTKKQSAAGKKCFKHIESWIQACSQ
ARKELGIVGFQPVK----KGTAFYEKAKEIL-
>Kryptoperidinium-foliaceum-CCMP1326-20130916|236076_1/285-382 [subseq from] QHnflz9VrY6dAP5UJeQ9l35F8tA Kryptoperidinium foliaceum CCMP1326
-ALKKAMKKAIKPKKVGSRALVLKGLREKTKTGLRADALMRSKSGKIVSKKMSALGKRLFKHIGTWIAAVKS
ARMELGITGFVAIK----KGTPFYESARKKY-
>Kryptoperidinium-foliaceum-CCMP1326-20130916|17277_1/349-446 [subseq from] /9pri/16W6Xw2nZw5qRWmUVyFko Kryptoperidinium foliaceum CCMP1326
-AMKKAMKKAIKPKKVGSRALVLKGLREKTKTGLRADALMRSKSGKIVSKKMSALGKRLFKHIGTWIAAVKS
ARMELGITGFVAIK----KGTPFYESARKKY-
>Kryptoperidinium-foliaceum-CCMP1326-20130916|226914_1/130-227 [subseq from] QsSN1AnylHT9wiZqJ0OBbX0fPvs Kryptoperidinium foliaceum CCMP1326

-PMKKPMKKPMKAKKVGARALVLKGLREKTKTGLRADALMKSKSGKVVTKKSSALGKRQFKHIGTWLAAVKS
ARMELGITGFVAIK----KGTAFYESARKKY-
>Kryptoperidinium-foliaceum-CCMP1326-20130916|399042_1/25-122 [subseq from] p3tpPfrh+ZfhJ48zHeUCZcot6eQ Kryptoperidinium foliaceum CCMP1326
MKVMKVMKKA-KVKKAGSKALVLKGLREKTKGGLKAGDLMKSTSGKVVSKKMSAAGKRSFQHIGVWVAACKA
ARIELGLSGFVAVK----KGTPFYDAVKARY-
>Kryptoperidinium-foliaceum-CCMP1326-20130916|421782_1/296-393 [subseq from] GwRLLtBT1TC9pEmtbJkUWqUZGeo Kryptoperidinium foliaceum CCMP1326
-VMKKAMKKAVKAKKVGSKALVLKGMRERTKGGLKAADLMKNKSGKVVIKSRNAAGKRRFQHIGVWLAAFKA
ARTELGLSGFVKVK----KGTPLYRATKVRY-
>Karlodinium-micrum-CCMP2283-20140214|30935_1/193-286 [subseq from] +30SnFnLyVDr/N9o6UygviQjXno Karlodinium micrum, Strain CCMP2283
KKVMKTRKVSIIAKGKKARVQVWRGKKVKTTGGLKKDQLVKSRKKKIVSKKRSEKGRE-----SKWSRATQK
AYKVKGYSGFKPIK----KGTSFYEKAKEVM-
>Karlodinium-micrum-CCMP2283-20140214|149797_1/207-300 [subseq from] E/8eLN3JoT47YqbOxH54uugeBUY Karlodinium micrum, Strain CCMP2283
KKVMKTRKVSIIAKGKKARVQVWRGKKVKTTGGLKKDQLVKSRKKKIVSKKLSEKGRE-----SKWSRATQK
AYKVKGYSGFKPIK----KGTSFYEKAKEVM-
>Alexandrium-temarense-CCMP1771-20130823|410993_1/690-783 [subseq from] tG4qYIt8s3KUKhRK1uRQIhfv23A Alexandrium temarense CCMP1771
MKVMKAMKVSTVAKGKKAKVEVYKGKKLKTTSGLKKDDLVKSKGGKIVSAKKSAAGKK-----SKWAIASAK
ARAEKGYTGFKGIK----RGGSFYNKAKEIM-
>MMETSP1439-20131203|2338_1/234-326 [subseq from] B451PTCM4XFn2gyHncMllDt9U4w Gonyaulax spinifera, Strain CCMP409
KAAMKPMKVSKIATGKKAKVQVYKGAKVKTKKGLKKEDLVKSKGGKIVSAKKSQQGKD-----SKWAKATAK
ARAAKGYTGFKPIK----RGGSFYEKAKEI--
>MMETSP0797-20121207|9138_1/264-356 [subseq from] 05CFIK4O9LCYUj4wQ+wHO/U0Vt4 Dinophysis acuminata
KKVMKPMKVSIVATGKRAKVQVFRGQKTKTKKGLKKDDLVKSKGGKIVSVKRSEAGKA-----SKWNKATAK
ARAAKGYTGFKAIK----KGTSFYEKAKEF--
>MMETSP0797-20121207|77426_1/248-330 [subseq from] pWUl+/AZ8UdGt7gTL5a4d6RsAnA Dinophysis acuminata
KKVMKPMKVSIVATGKRAKVQVFRGQKTKTKKGLKKDDLVKSKSGKIVSVKRSEVGKA-----SKWNKATAK
ARAAKGYTGFKAIK----KG----------
>MMETSP0797-20121207|89641_1/208-290 [subseq from] iyKvs8YziBruQ5Z8AeYzEF71Bjk Dinophysis acuminata
KKVMKPMKVSIVATGKRAKVQVFRGQKTKTKKGLKKDDLVKSKSGKIVSVKRSEAGKA-----SKWNKATAK
ARAAKGYTGFKAIK----KG----------
>MMETSP0796-20121207|65671_1/208-300 [subseq from] SKdpLwqzq1ZwElzW8H+e4Os1zdQ Pyrodinium bahamense
MKVMKMKRTSMIAKGKMAKVEVWRGKKLKTAKGLKKEDLIKNKEGKIVSAKKSQTGKD-----SKWNKAVAK
ARAVKGYTGFKAIK----KGTSFYEKAKEV--
>MMETSP0766_2-20121228|58736_1/211-303 [subseq from] UelLShjHszxES6grow4WHiRd9nY Gambierdiscus australes
KAKAKIKAKSKIAKGKKAKSQVFKGVKARTTSGLKKEDLMKNKKGKVVSTKMYEMGKK-----NKWTQATAK
ARALKGYTGFKPLK----KGGSFYEKTKEV--
>MMETSP1462-20131121|48045_1/212-302 [subseq from] ckEkGxt1boWQXPwB9t1lu6XUYO4 Brandtodinium nutriculum, Strain RCC3387
--KAKTAKATAKATAKSDKLRVYKGDKLKKKSGLTKTDLVRNKDGKVVSAKRSEIGRN-----SKWAKATAK
ARAIKGYIGFKSIK----KGSSFYEKAREV--
>Crypthecodinium-cohnii-Seligo-20130904|83812_1/225-316 [subseq from] 11weYt39vUiIKSpAkTrY4X9EG+M Crypthecodinium cohnii Seligo
-KASKPKKESKIAKGKKAKELVWKGLKEKTSTGLKKEHLVQNKFGKVVSAKKSAAAKT-----NKWAIATSR
ARKELGFSGFRPLR----KGDEFYVKAKEI--
>Prorocentrum-minimum-CCMP1329-20131001|41293_1/343-437 [subseq from] D7hgtDeVa2V7pAc/q6PCepbtK0E Prorocentrum minimum CCMP1329
MKAAKAKRISKV----GTKAQVLAGKREKTKWGLRKEHLVKSKKGKTVTKRKSAVGKKAFQNLAKWILSFRR
ARAELGLTGFVLVK----KGTALYNKTKEYY-
>Prorocentrum-minimum-CCMP1329-20131001|1849_1/295-389 [subseq from] /+X/01yM3JRQrEF/Yr9bm+J+rBI Prorocentrum minimum CCMP1329
MKAAKAKRISKV----GTKAQVLAGKREKTKWGLRKEHLVKSKKGKTVTKRKSAAGKKTFQNLAKWILSFRR
ARAELGLTGFVLVK----KGTALYNKTKEYY-
>Prorocentrum-minimum-CCMP1329-20131001|55379_1/30-124 [subseq from] 0Yiu0DHCRoA2VBZcD5PLgx7l7yU Prorocentrum minimum CCMP1329
MKAAKAKRISKV----GTKAQVLAGKREKTKWGLRKEHLVKSKKGKTVTKRKSAAGKKTFQNLAKWILSFRR
ARAELGLTGFVLVK----KGAALYSKTREYY-
>Prorocentrum-minimum-CCMP1329-20131001|238994_1/287-378 [subseq from] HF/4ev7RFDdUVrxK5EAzDPHBgCo Prorocentrum minimum CCMP1329
---VKARRLSKV----GTKAQVLAGKREKTKWGLRKEHLVKSKKGKTVTKRKSAVGKKAFQNLAKWILSFRR
ARAELGLTGFVLVK----KGTALYNKTKEYY-
>Prorocentrum-minimum-CCMP1329-20131001|14638_1/249-340 [subseq from] RuvQMOqqxIYjgkHfMmkNhhq1lNQ Prorocentrum minimum CCMP1329
---VKARRLSKV----GTKAQVLAGKREKTKWGLRKEHLVKSKKGKTVTKRKSAAGKKTFQNLAKWILSFRR
ARAELGLTGFVLVK----KGAALYSKTREYY-
>Scrippsiella-hangoei_like-SHHI_4-20140214|53260_1/240-333 [subseq from] kzzQaZ/ZAYbAWxlPmkhKl6mJIWQ Scrippsiella hangoei-like, Strain SHHI-4
MKAMKVKKESMI----GSKKAVFHGKKEKTKTGLKSSDLMKNKDGKIVFKRRNAHGRKIFENLGRWVAAVSK
ARAELGLTGFALVE----KGSALYAKTKEL--
>Scrippsiella-hangoei-SHTV5-20131105|19734_1/255-348 [subseq from] 4OM448BisVRbzDo0xt8ESpx5qrU Scrippsiella hangoei SHTV5
MKAMKVKKESKI----GSKKAVFHGKKEKTKTGLKSSDLMKNKDGKVVFKRRNAHGRKIFENLGRWVAAVSK

ARAELGLTGFVLVK----KGSALYAKTKEL--
>MMETSP0796-20121207|36884_1/262-355 [subseq from] wGLCfzusFGjpS0eJY/nF+nIQn50 Pyrodinium bahamense
MKAMKVKRTTKI----GSRRQVLCGWKEKTTKGLKKADFIKNKDGKWVTKKRSAIGRSRYANIAPWINACKK
ARQELGLTGFVPVK----KGSALYEKAREL--
>Glenodinium-foliaceum-CCAP1116_3-20130913|15436_1/279-372 [subseq from] uLWwS4RgRpaBC5z+ri9DFBQHlOs Kryptoperidinium foliaceum
----RAVACSVCGRTVGTKEEVLSGACARTAGGLQAKGLVRNSSGKVVSKKLSALGKRAFQHIGTWLAAVKA
ARLELGLVGRGAIR----KGTPLYESARAL--
>Karlodinium-micrum-CCMP2283-20140214|23974_1/262-353 [subseq from] RDv96yytb3qE/KzCoA8I6UCJQW8 Karlodinium micrum, Strain CCMP2283
---MKSLKMSNV----GSRKSVLAGKKIKTKTGLTKDQLVKNKVGNICSRAKNAIGYKHFKNIEKWVNALKT
ARRELGLTGFVAVK----RGTEYYNKTMSLY-
>Karlodinium-micrum-CCMP2283-20140214|211202_1/621-712 [subseq from] 2UmRDRPHaUEcmOXZTbBKVbavc+Y Karlodinium micrum, Strain CCMP2283
---MKAMKLSKV----GSRRSVFAGKKVKTKTGLKKEQLVKSKSGKICSSKKVAHGHKHFKNIEKWVNAMKT
ARKELGLTGFVAVK----KGTAYYDKTKTLY-
>Karlodinium-micrum-CCMP2283-20140214|128717_1/155-246 [subseq from] Un1wkYQFRs8LQvEqneiGg0y0oSk Karlodinium micrum, Strain CCMP2283
---MKAMKLSKV----GSRRSVFAGKKVKTKTGLKKEQLVKSKSGKICSSKKVAHGHKHFKNIEKWVNAMKT
ARKELGLTGFVAVK----KGTAYYDKTKALY-
>Karlodinium-micrum-CCMP2283-20140214|157913_1/154-253 [subseq from] EvQlGtNwKh2JbQ3MDYC3YUqrQAk Karlodinium micrum, Strain CCMP2283
MKAMKTSKKKKIAKGKRGKSMVYKGKFEKTVGGLTKDALVKSRAGKIVSKRLHAHGKKSYSNIKSWVEAFLK
ARSELGLNGFVAIK----KGSALYSKTMELYK
>km09403-km/58-157 [subseq from] dtZgJij+NksFIVcVP0heoN+0E3A Karlodinium micrum, Strain CCMP2283
MKAMKTSKKKKIAKGKRGKSMVYKGKFEKTVGGLTKDALVKSRAGKIVSKRLHAHGKKSYSNIKSWVEAFLK
ARSELGLNGFVAIK----KGSALYAKTMELYK
>Karenia-brevis-CCMP2229-20130916|223851_1/137-233 [subseq from] 62VhiHmAt5iYA7g83+mam175MCo Karenia brevis CCMP2229
MKVMKTS-KRKIAKGKGAKRLVYKGKFEKTVGGLTKDQLTKSKAGNIVSKRMQARGKIAYVNIKSWVTAFMK
ARTELGLSGFVPVK----KGLPLYAKTKEL--
>Karenia-brevis-SP1-20130916|120678_1/137-233 [subseq from] G4oIzyAKqOAUMQ1quVbRZ4+Mg1M Karenia brevis SP1
MKVMKTS-KRKIAKGKGAKRLVYKGKFEKTVGGLTKDQLTKSKAGKIVSKRMQARGKIAYLNIKSWVTAFMK
ARTELGLSGFVPVK----KGLPLYAKTNEL--
>MMETSP0797-20121207|41234_1/147-242 [subseq from] EAtFnURKxTVUQQd0Nr9HKtD/ZZg Dinophysis acuminata
TKAAGVR-KAKVARGKRGKALVYRGRFERTPGKLRKEDLARSKSGKIVSKRMQACGQRAYGNIRAWVDAMMG
ARAQLGLSGFVPIR----RGSPLYLATKA---
>Karlodinium-micrum-CCMP2283-20140214|14155_1/92-173 [subseq from] MXYauG9prWR7Ic4tQ+YU8c5cSNQ Karlodinium micrum, Strain CCMP2283
----------SKLDVFLGRKEKTVGGLTKIDLVRSSSGKIVSKKLQAHAKRSFKNIKPWNDALAK
AKKELGVSGFILVK----KTSAVYKRARRIY-
>Alexandrium-monilatum-CCMP3105-20140214|42973_1/19-113 [subseq from] nNFHIBdcYtykoGI32dS94k1nUMo Alexandrium monilatum, Strain CCMP3105
-RAMKRRKKVIIARGRFSRGDVFNGKKLKTVGGLRVCDMMRNRYGKIVSKRASARAK-----LNIWSKAISA
ARRALGITGWVNIN-QGPKGVAIYQKAKML--
>MMETSP1374-20130617|26929_1/224-322 [subseq from] ju320ykhHlwUjFXypHnKGnYn4Mo Symbiodinium Mf12.5f/clade A3
MKAMKPMKVNKVAYGRGAKARVYKGLKTRTSGGLTKDKVIKNKAGKFVSKALSEAAKKRMATVKYW-QAIKK
ARAELGLTGFVAIK----KGEPLYKRAMEIYK
>MMETSP0689_2-20121128|31647_1/246-344 [subseq from] T82metQg+e40q1eqSdj3mv/EA6s Amphidinium massartii
MKAMKVMKPSIIARGKRAKVQVYSGKKVKTSGGTTKDMLMRNKAGDIVSKRMSLAAKKRYTTIKPWVEATKK
ARTALGITGFVAMK----KGSEFYKKTKEFY-
>MMETSP0253-20130528|11663_1/194-292 [subseq from] 6kzpbZMtDV+/pR4ADbBP7sTHpK8 Noctiluca scintillans
MKTMKATAMKTIARGRLAKVRVFQGKFKKTSGGLTKDALMKNSRGKVVSKAKSAKSKKLYRKLQPWHESWKN
AKAELGLTGFALCK----KGTPFYIKIMEKY-
>Lingulodinium-polyedra-CCMP1738-20130920|61392_1/289-387 [subseq from] CK+k0Di7C/szpcEbJWe36MSw+U4 Lingulodinium polyedra CCMP1738
MKAMKKARVSKIARGRYMYLQVLRGKKEKTPAGLKKDDLLKNAHGKVVSKKRSAWSKTQYKHIAKWVDSVMQ
ARKELSITGWLVPR----KGTPLYIRAKELF-
>Crypthecodinium-cohnii-Seligo-20130904|197819_1/294-389 [subseq from] 72V9tBljHPGEMXOIdQCdt1yIhuo Crypthecodinium cohnii Seligo
KKAMKK-KVDKRRIQRTL-VAVWRGKLKRTPGGLKKDDLMKNARGKIVSKKKSAKGKEQFKRVSVWIAALVK
ARKELGLTGKDVPT----KGSPLYKKVKEL--
>MMETSP1338-20131121|94253_1/39-137 [subseq from] afESDyUdRHFQFB5P9mqeb9MRnow Pelagodinium beii
MKAVKKR-KSKIASGQLRKAHVLQGVKVKTSGGLTADKLARNKEGRVVSKAQSEVGKKYFESLGKWVAACKA
AREALGVRGFCAVGGTSLQGKALYIKAK----
>MMETSP0228-20121227|31718_1/2-84 [subseq from] jUINxG8ZqLYG2LcV2TdtrxRrk5s Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
----------GKLAKASVLRGNEVKAXDGLTKDRLAKNRFGRIVSKAASVRARKFFSAFGAWNKACAA
VCEALGVQAFCPTGGSTAQGKARYA-------
>MMETSP1462-20131121|24672_1/49-145 [subseq from] 4t/dGIS6eDgFKg/Hu77HgnKJWHU Brandtodinium nutriculum, Strain RCC3387
AAKRKRR-PSKIARGKMAKSAVFRGDKAKTSGGLVKAQIAQNSEGRYVYKSLSAVARKRYENLKAWVDACKA
ARAAMGLTGFVAVGGKTATGRALHAQ------
>Alexandrium-temarense-CCMP1771-20130823|4222_1/38-132 [subseq from] mZ1qV0gpZqWBz108vevIHPIkJ/o Alexandrium temarense CCMP1771

---MKAVVKATIASGRLSKAMVYQGRRDKTAGGLKASDIIRNKRGKFVFKKASLRGVNA-----SWIKALAA
ARKALGLKGFVKI-NKGPDGQALYAKAKALHQ
>MMETSP1462-20131121|21210_1/120-212 [subseq from] xPMd3x0lrZYQgsBqqF9jCPTGzwQ Brandtodinium nutriculum, Strain RCC3387
----KAKKAAARAKSRYAKAMVARGTLAQTPGGLTKKDIVRSKSGKFVSKKQSEQGRAS-----PWIKAVAA
ARKALQIKGFAVVGGKSKQGKELYAKAKSL--
>MMETSP1462-20131121|24689_1/191-283 [subseq from] h5GhGXnEWuk7rH5BwWlUf2/e+oc Brandtodinium nutriculum, Strain RCC3387
----KAKQAAARLRSRSAKAMVFRGFLAQTAGGLTKNDIVRSKSGKFVSKKASEKGKAN-----RWIQAVAA
ARQALKIKGFAVVGGKTPQGKELYAKAKSL--
>MMETSP1462-20131121|10633_1/66-158 [subseq from] oX6lVs2mw71bu6XqbZJXfOXo61w Brandtodinium nutriculum, Strain RCC3387
----KAKKEAARARSPRARMMVARGTLAQTAGGLTKKDLVKNKGGMYVSKARSAASKKS-----PWIKAVAK
ARNALGIKGFAVVGGKTKEGMALYKKAKSF--
>MMETSP1462-20131121|11379_1/150-236 [subseq from] j7S8fCSvsTvH2xp614ZQUbZfswI Brandtodinium nutriculum, Strain RCC3387
------K-----MPPNKVKLLVVQGKMPETGGGLKKKDFVKNKYGKIVSKKAQKHAKGN-----PWMKAVVA
ARKALGVKGFAVVGGKTKQGKALYTKAKSLM-
>MMETSP1462-20131121|157924_1/86-171 [subseq from] 52rGdYFVuj4A+jczy/LCs7Nirw8 Brandtodinium nutriculum, Strain RCC3387
------K-----MPSSRIKALVAQGKLSETKGGLKKKDLVKNKAGKLVSKKASASRKNS-----PWIKAVAA
ARKALGVKGFAVVGGKTKEGKALYAKAKSL--
>MMETSP0228-20121227|26920_1/25-127 [subseq from] 5U1fhF5tgLVQfDOqTs0W7CU37rg Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
-QAAQVKAMPVIARAPRAKSAVFQGNRVTTHTGFTKERLTNSIRDSIVSKAASAVAQKSFPKIGSWNKAVAA
ARKALGIQGFAAIGGKSAQGKALYSKIETLYE
>Lingulodinium-polyedra-CCMP1738-20130920|354502_1/61-152 [subseq from] u90gDkkWoWqpv4rJHxhahKCDPeg Lingulodinium polyedra CCMP1738
MKSKKAKRVSTVAIGKMARAKVIQGWKVRTAGGLRANDLMRNQHGRIVPKKQHALGK-----IHPWIKALSA
ARKALGFKGFIAIN-RGPEGKALYAK------
>MMETSP0228-20121227|9328_1/47-144 [subseq from] hi8JwUmgutMYPReyQQCY7OoU9MU Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKMMKAKRVSKIALGRMAKYLVFKGSKAKTSGGLQAKDIKKNKNGKLVSKKLSAAGKS-----HPWMAACAA
ARKALGLKGXVPMN-SGPDGKAXXAKAKSLYK
>MMETSP0228-20121227|45795_1/69-166 [subseq from] Txh1LF5aBmmpgS5aQrJLwkGOsLk Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKMMKAKRVSKIATGRMAKFLVFKGSKAKTSGGLQASDIKKNKSGKLVSKKLSAASKS-----HPWMAACAA
ARKALGLKGFVVMN-RGPEGKAYYAKAKSLYK
>Karenia-brevis-CCMP2229-20130916|173497_1/12-106 [subseq from] TFeEDM/Z5YOfgwINSTB4ejru+gg Karenia brevis CCMP2229
-------TSKTIATGKYKKAVVFRGSKAKTSGGLTKADLKKNKNGKIVSKKASARAKSLYKDIGPWVKAVSV
AKKELGITGCVYVN-SPGQGRTLYKKAKEIY-
>Karenia-brevis-SP1-20130916|291446_1/609-701 [subseq from] mM7IMxrkercwSUZ7WNjIqqglFN4 Karenia brevis SP1
---------KSIATGKYKKAVVFRGSKAKTSGGLTKADLKKNKNGKIVSKKASARAKSLYKDIGPWVKAVSV
AKKELGITGCVYVN-SPGQGRTLYKKAKEIY-
>Peridinium-aciculiferum-PAER_2-20130926|11815_1/188-283 [subseq from] qYj/QRGKHIZ00xLS2C5MEOhn+Y0 Peridinium aciculiferum PAER_2
--AKKAKKPSIIARGVRRNYLVFKGTKEKTVGGNTKADIMKNKWGRYVSKKKHEIGMKN-----PWIAAVMA
ARKALKITGFKTIGGQSEAGKKLYAKAKSFF-
>MMETSP0790-20130122|8298_1/297-380 [subseq from] 9Mb3FVWe+Nq1ZKUJ/tCeKodWR6E Alexandrium catenella, Strain OF101
----------QGQAVQILVFTGAKERTIGGLKATDIVKNRYGKCVSKKASNRAKNN-----IWANAIKE
ARAELGITGFCTIGMGAE-GKALYHKAKTIL-
>Prorocentrum-minimum-CCMP1329-20131001|12550_1/43-143 [subseq from] +GtWF9g2NJBCKEi0lBOhQCEPBgM Prorocentrum minimum CCMP1329
--ASRGKKFSKVASGPHARRQVMSGIKEKTKGRLTRDLLVKNKDGKIVSIKASAAGRKRYLSIKLWFESVSA
ARKALRLTKFVAVNGKSPEGQALYASAKALY-
>MMETSP1110-20121109|63058_1/370-467 [subseq from] PhMB0NcbLVRsdqFohAgF0Z+W9N4 Symbiodinium, Strain CCMP421
LSAAKRREVMKAR------KEVMQGMRKKTKGGLTKEDLIVNKRGRVVSKKKHEQGKRSFQKLSSWVDAVKA
ARDKLAVIGFRPVGGKSQEGKELYKEAKEIYE
>MMETSP0191-20121206|3610_1/45-142 [subseq from] Hf0Va29et0rGIstOw1tUlSytXNQ Sorites sp.
LKAARRREVMKAR------REVMQGLREKTSGGLTKDDLMVNKRGKIVSKKRHERGLKLFSRLSSWIDAVKS
ARSKLSILGFRPVGGRSKEGKELYMEAKQIYN
>MMETSP0191-20121206|838_1/3-90 [subseq from] kAEMPXQAe8QaNdJuXOGZTwy2SoM Sorites sp.
----------KAR------REVMEGLREKTSGGLTKDDLMVNKRGKIVSKKRHERGLKLFSRLSSWIDAVKS
ARSKLSILGFRPVGGRSKEGKELYMEAKQIYN
>Symbiodinium-sp-C15-20130923|40249_1/273-370 [subseq from] u6aAXVFvKSLisMMrTWjZJI6JfgA Symbiodinium sp C15
LSASRRREVMKVR------RQVMKGLREKTSGGLTKDDLMVNKHGKIVSKKQHERGLKLFSRLSSWVDAVQS
ARSRLSILGFRPVGGRSKEGKELYTEAKKIYQ
>Symbiodinium-sp-C1-20140214|96592_1/278-373 [subseq from] SM9FUAX1OW1VZX9pdJusc7j7TfI Symbiodinium, Strain C1
--AARRREVMKVR------RQVMKGLREKTSGGLTKDDLMVNKHGKIVSKKQHERGLKMFSRLSSWVDAVQS
ARSKLSILGFRPIGGRSKEGKELYMEAKKIYQ
>MMETSP1338-20131121|32375_1/76-173 [subseq from] HyVmKw/YPk1uYHIqz/eVc/ssho0 Pelagodinium beii
MQAKQRRAMMKAR------RMVMSGKKKKTTGGLTKDDLMVNKRGKIVSKKQSARGQKTYALLKTWQEAVMT
ARSKMNVQGFIPINGRTKEGQELYQTAKKIYE
>MMETSP1440-20131203|4564_1/122-217 [subseq from] 1wlPEEp5m8E+pOSL7091A11UNII Polarella glacialis, Strain CCMP2088
--AVKRREMMKVR------RLVMEGKKEKTSGGLKKDDLIINKRGQVVSKRQSARGKELYAKLKGWVDAVAL
ARSKMALIGFIPAGGRTEQGQELYKTAKDIYD
>MMETSP0227-20121206|11678_1/123-218 [subseq from] A0p87DrxVP+2CXfVPijENnasBwo Polarella glacialis CCMP1383
--AVKRREMMKVR------RLVMEGKKEKTSGGLKKDDLIINKRGKVVSKRQSARGKELYAKLKGWVDAVAL

ARSKLALIGFIPAGGRTEQGQELYKTAKDIYD
>MMETSP1441-20131203|18526_1/9-105 [subseq from] ajMqbYSfRm7NaGqLE1+UXfCR018 Heterocapsa arctica, Strain CCMP445
MMAKHRREMMKAR------RLVLEGKLTKTSGGLTKDDLVLNKSGKIVSKKASESAKARYATLASWNRAVME
SRAELGLTGFVAIRGATDAGQKLYETASKRY-
>MMETSP0503-20121227|41245_1/383-479 [subseq from] m7PXJuFolyU4B1ytcWd4cRsPgGw Heterocapsa rotundata, Strain SCCAP K-0483
MTARRRREMMKAR------RLVLEGKLKKTKGGLTKDDLILNKQGKIVSKKASEKAKARYAELKSWNQAVMD
ARAQLGLKGFVAVNGPSEEGQKLYQVASKRY-
>MMETSP0448-20130528|23294_1/8-100 [subseq from] hBkPSTPxAmpulFU/FjVt0yTwJDE Heterocapsa triquetra
-----RREIMKAR------RQVLEGKLKRTSGGLTKDDLVLNKQGRIVSKRASQRAKARYAMLGAWNKAVME
ARAELGLTGFVAINGQTTEGQKLYRVASELYQ
>Scrippsiella-trochoidea-CCMP3099-20130930|321723_1/405-502 [subseq from] TmjIgd/efP2KGcS7tXzLd67y+YQ Scrippsiella trochoidea CCMP3099
MKGLKRRHMMQAR------RKVMKGLLEKTPTGLTKDDLVKNKYGRVVSKKQHERGKMNYAKLKGWADAVRA
ARSALDISGFVAIGGSSETGKTLYKKAKELYE
>Azadinium-spinosum-3D9-20130829|21605_1/842-939 [subseq from] SO+WKOwaCXx6OaeMpj4cX3VMICo Azadinium spinosum 3D9
IKAAHWRGIMKAR------REVVNGTRAKTKSGLTKDDLIINKRGKIVSKKKSLAGQAAYAKLKGWTDAVAA
ARAELQISGFLAVNGGTELGQKLYLKAKAIYE
>Prorocentrum-minimum-CCMP2233-20131001|258029_1/396-492 [subseq from] 0AnaCeTxftTp54iL53bsBnWfz28 Prorocentrum minimum CCMP2233
MQARSRRELMAMR------RRVYLGKLEKTSGGLTRDDLIMNKRGQVVSKKRNKAGLEAYKKLKAWVDAVQA
ARASLNVTGFQAVGGQTEAGQRLLEKARAVY-
>Lingulodinium-polyedra-CCMP1738-20130920|358111_1/581-674 [subseq from] 3Qsi3GORvalDJBv/2evexL7xUvA Lingulodinium polyedra CCMP1738
LKAEKRKFMFQAR------REVFVGERDRTPSGLSKDDLMLNRAGRVVSKAKSMAAKQR---LRTWWESVKA
AKEALNISGFQIIGGETEAGQELYRKAREIY-
>MMETSP0796-20121207|60999_1/728-823 [subseq from] x0e8WyUHEE3zmQMFKJh9Rk2NZis Pyrodinium bahamense
--AKTRRRLMQLR------REVLEGKRLSTPSGLKKDDLMVNKRGEVVSKLKYEQGQKHYAALAKWVEAVKQ
ARKVLDIKGFAPIGGKSESGQLLYRKAREIYD
>Alexandrium-monilatum-CCMP3105-20140214|43838_1/63-160 [subseq from] gLJlYGx54xiodaWfGCQpfwlOq9E Alexandrium monilatum, Strain CCMP3105
LKADKRRLLIRAR------RDVFKGKAGRTSGGLTRDHLTKNKRGRVVSKARSAHGKRIYATLKSWIDAVMA
ARAELGVSGFVAVKGQTHSGQELYRKARAIYD
>MMETSP0228-20121227|54781_1/135-226 [subseq from] iuejsonb6BGBC1D/fcczFpHH9rc Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
-KAAKRRLMMQAR------RDVMAGRRDRTPGGLTKDQLVRNKHGRIVSKAKSEAGKQAYARLRGWVEAVVA
ARAELKISGFSPVGGGSETGQRLYGRA-----
>MMETSP0790-20130122|20240_1/112-208 [subseq from] F2PeQS1mpJF9jIkPRHeEWDdIsYQ Alexandrium catenella, Strain OF101
LKAEGRRRLMKMR------REVLMGKRAKTPGGLRQDNLVKNKRGRVVSKAASRASKESYAKLATWTEACVT
AKAELGLSGFVPVGGRSAAGQELHRRARAIY-
>MMETSP0228-20121227|53345_1/236-334 [subseq from] CoCIFqKO7Yfb7HNhAciysNcwCQ8 Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKPMKMKPMKKVLKGSRGKRMVVQGKFAKTRGGLKADALVASKKGKIVSKKGQAHGVKAFANIKQWTEATKK
ARQELGLVGFHLV----KKGSALYKRIKEIM-
>MMETSP0228-20121227|8347_1/236-334 [subseq from] U8eQR+d6Pb0RHdypYKrlkYIvl9I Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
MKPMKMKPMKKVLKGSRGKRMVVQGKFAKTRGGLKADALVASKKGKIVSKKSQAHGVKAFANIKQWTEATKK
ARQELGLVGFHLV----KKGSALYKRIKEIM-
>Karenia-brevis-CCMP2229-20130916|115263_1/202-299 [subseq from] zFZT8doGtJXnmUxDeRBEtlQqmTg Karenia brevis CCMP2229
MKAMKKGKVSKIARGKFSKRKVLKGMFAKTSGGLVKQDLMMNKKGKVVSKKQHALGKKAYLNIKGWTDSYMK
ARAELGISGFALL----RKDTAVYAKTKEL--
>Karenia-brevis-SP3-20130916|231030_1/202-299 [subseq from] gT+/EMucIaB00xjOiNdnDcc801w Karenia brevis SP3
MKAMKKGKVSKIARGKFSKRKVLKGMFAKTSGGLVKQDLMMNKKGKAVSKKQHALGKKAYLNIKGWTDSYMK
ARAELGISGFALL----RKDTAVYAKTKEL--
>Karenia-brevis-SP1-20130916|21121_1/207-304 [subseq from] TyJY+ayB+Cq9Qoe1Q81RNnGx/cI Karenia brevis SP1
MKAMKKGKVSKVARGKFSKRKVLKGMFAKTSSGLVKQDLMVNKKGKAVSKKQHALGKKAYLNIKGWTDSYMK
ARAELGISGFALL----RKDTAVYAKTKEL--
>Karenia-brevis-Wilson-20130916|365171_1/181-278 [subseq from] Y3MEwEX2L+T23QT/B3211R2UeuA Karenia brevis Wilson
MKARKKGKVSKVARGKFSKRKVLKGMFAKTSGGLVKQDLMVNKKGKAVSKKQHALGKKAYLNIKGWTDSYMK
ARAELGISEFALL----RKDTAVYAKTKEL--
>MMETSP0766_2-20121228|57444_1/264-358 [subseq from] A5ve1M4TdISv5YXbqc5Y+/8hyQg Gambierdiscus australes
-KKLKK-KASKIAKGKFS--KVMKGEKVKTRGGLKKEDLVINKRGKVVSKKRSERGNSIYDRVKPWIEACNK
ARAELGITGWVNV----KKGSELYNKAKEIY-
>km06049-km/27-123 [subseq from] 9cU4v705UJ1dFeSwkReOeknFh3Q Karlodinium micrum, Strain CCMP2283
---MKAARASKIATGRFANAQVYLGRKERTTGGVKKVGLGKNKWGQIVSKKRSAASRKK---LAGWVEACQK
ARKDLNITGFAIIGGKSLVGKALYAKALSYY-
>Karlodinium-micrum-CCMP2283-20140214|22567_1/33-129 [subseq from] ZMK+JExhWvycje8B3XXJyT6tJoQ Karlodinium micrum, Strain CCMP2283
---MKAARASKIAAGRFANAQVYLGRKERTTGGVKKVGLAKNKWGQIVSKKRSAASKKK---LAGWVEACQK
ARKDLNITGFAIIGGKSLVGKALYAKALSYY-
>Symbiodinium-sp-C1-20140214|7134_1/115-216 [subseq from] JzAw1RbKaHVBSVwwR5WYgvIygGA Symbiodinium, Strain C1
-AAKKKKGNPKVAVGRWAKWRVYVGWKEKTVGRLKKHNLKTNRRGRVVYKKMSDQAKERMKNISTWYQSVKE
ARQVLKLVGFVPVGGKTAAGQALLKQTKMIY-
>MMETSP0191-20121206|72365_1/50-151 [subseq from] 3N19JQ56BvHPpICQD6ZlG2QbvBA Sorites sp.

-KSKKTTTNPKVAVGRWAKWRVYAGWKEKTVGRKKKPDLKLNRRGRVVYKTMSDQAKERMKNIGTWYQCVKE
ARQALNLVGFVPVGGKTAAGKALLIKTKVIY-
>MMETSP1110-20121109|22104_1/8-103 [subseq from] yxfqoalFhGjLfeFD4OtcTy2jL7U Symbiodinium, Strain CCMP421
-------KAGAVARGRWARWRVFIGVKEKTVGRLNRTNLTLNRRGKVVSKKSSEQAKSRFQHLKTWFEAVKA
AREALGLRGFVPVGGKTIAGKALYLRAKVIF-
>MMETSP0797-20121207|34554_1/21-113 [subseq from] BvBdRXXySXzoofnv9rcoLf8VWg0 Dinophysis acuminata
-KXMKAKR-----SKRMAKVVAFRGGSTGGATTLKQGDLIKNXHGRIVSKKQSQLAKKRYASIGKWTAAVQK
ARKALGTKGFVAXK----KGSPLYKKAREFF-
>MMETSP0790-20130122|48890_1/115-206 [subseq from] ie3zu66oDqiqG2VFsqp85ggOCpE Alexandrium catenella, Strain OF101
---RRKKP-----TKKMAKVVVFRGGDTDGVTKLRKEDLVKNSAGRIVSKKASEAAKKRYADIGKWTACVKR
AREQLGLEGFVKVK----KGSPVYEQAMAYYK
>MMETSP0661-20131031|50927_1/98-189 [subseq from] HER4ZdZTcLFJqgzFdcaC7bLNur8 Alexandrium margalefi, Strain AMGDE01CS-322
---RRKKP-----SKKMAKVVAFRGGDTGGVTKLRREDLVKNSAGRIVSKRASEAAKKRYADIGKWTACVKR
AREELGLEGFVKVK----KGSPVYEQAMAYYK
>Alexandrium-monilatum-CCMP3105-20140214|6733_1/426-516 [subseq from] gd/Jmgltwr5hXYlu11EJOuskeCk Alexandrium monilatum, Strain CCMP3105
----RKKP-----SKKMAKAVVFRGGDTDGVTKLRKEDLTKNSQGKIVSKRASEAAKKRYAEIGRWTAAVKR
ARDELGLEGFVKVK----KGSPVYEQAMVYYK
>MMETSP0796-20121207|78150_1/374-460 [subseq from] maNYg4tRPYf5h9Us/edscrQpvR8 Pyrodinium bahamense
----------SKKMAKAVVFRGGDTEGVTKLRKEDLIKNSSGKIVSKKASEAAKKRYADIGKWTEAVKR
AREELGLEGFVSVK----KGTPVYEKAREYYK
>MMETSP0228-20121227|22299_1/296-381 [subseq from] 9H0sTeBkMfGKC8tK2/V/02X4Z84 Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
----------SKKMAKRVAFQGGDTEGKTKLKKEDLVKNASGKIVSKKASEASKKRYAEIGKWTDAVKR
ARQELGLQGFVSVK----KGSPVYEKAREYY-
>Alexandrium-temarense-CCMP1771-20130823|21540_1/450-540 [subseq from] niCIIQxxXMuMCV94Br2OV9PrALM Alexandrium temarense CCMP1771
----RRKP-----SKKMAKVVVFRGGDADGATKLRKEDLVKNAAGKVVSKKASEAAKKRYLDIGKWTAAVKR
ARDELGVEGFVKVK----KGTPVYEKAQEFYK
>Kryptoperidinium-foliaceum-CCMP1326-20130916|15297_1/314-403 [subseq from] GT/Yv9XKhGb651bU1GiZqUz7FJQ Kryptoperidinium foliaceum CCMP1326
-----QQP-----SKKMAKAVAFRGGDTDGLTKLRKEDLVKNRTGKIVSKKASEAAKKRYAEISKWVEAVKR
ARAEMGLAGFVAVK----KGTPVYEKAQEYYK
>Durinskia-baltica-CSIRO_CS-38-20140214|1746_1/100-187 [subseq from] KH2DZOr8inJGjuCR2To6jdsKA8M Durinskia baltica, Strain CSIRO CS-38
-------Q-----TKKMVKAVAFRGGDTDGLTRLRQDDLVKNRSGKIVSKRASEAAKKRYADVSKWVEAVKR
AKAEMGLEGFVAVK----KGTPVYEKAREFYK
>Scrippsiella-trochoidea-CCMP3099-20130930|4496_1/76-167 [subseq from] YBnoszuPQDSWCWmxOevUoppPKXk Scrippsiella trochoidea CCMP3099
-VTRRKKQ-----SKKMAKAVAFRGGDTDGVTRLKKEDLVKNSAGKIVSKKASEAAKKRYADASKWNEAVKR
ARTEMGLAGFVAVK----KGTPVYEKAREI--
>Scrippsiella-hangoei_like-SHHI_4-20140214|279157_1/429-514 [subseq from] Gog0+4FR9bLc+XjT0Rx1szI5Kn0 Scrippsiella hangoei-like, Strain SHHI-4
----------SKKMAKAVAFRGGDTDGVTRLKKEDLVKNSHGKIVSKKASEAAKKRYADAAKWNEAVKR
ARTEMGVSGFCAVK----KGTPLYERAQAIL-
>Scrippsiella-hangoei-SHTV5-20131105|68234_1/467-551 [subseq from] HtJgffoXA4pVgej2PUJlLGb0MSc Scrippsiella hangoei SHTV5
----------SKKMAKAVAFRGGDTDGVTRLKKEDLVKNSHGKIVSKKASEAAKKRYADAAKWNEAVKR
ARIEMGVSGFCAVK----KGTPLYERAQAI--
>Peridinium-aciculiferum-PAER_2-20130926|12253_1/165-249 [subseq from] w1fwTjzswVAqokygcnczJ0+CPTY Peridinium aciculiferum PAER_2
----------SKKMAKAVAFRGGDTDGVTRLKKEDLVKNSNGKIVSKKASEAAKKRYADAAKWNQAVKR
ARTEMGVSGFCAVK----KGTPLYERAQAI--
>Crypthecodinium-cohnii-Seligo-20130904|21511_1/476-561 [subseq from] aIXkebtekY8Uk7vgr1U4K96H+Hg Crypthecodinium cohnii Seligo
----------SKKMAKAVAFHGGDAEGLTRLKKDDLVKNSQGKIVSKRASEAAKRRYADAAKWTEAVKR
ARTEMGIQGFYAVK----KGTPVYDKAKEIY-
>MMETSP1462-20131121|7885_1/94-178 [subseq from] CNBYexwgBPSOsUQes+fsZAoomkU Brandtodinium nutriculum, Strain RCC3387
-------K-----SKKMAKAVAFRGGDTDGVTRLKKSDLVKNSQGKIVSKRASEAAKKRFADIAKWMEAVKK
AKVAMGLEGFVKVK----KGSPVYEKAQE---
>MMETSP0784-20121206|65506_1/391-481 [subseq from] B63I9qdo8LepIy5qbLO+iwf9lVo Gymnodinium catenatum
----RKKP-----SKKMAKIVAFRGGDTDGVTRLRKEDLVKNSVGKIVSKKKSESSKKKYKDLSKWTVAVKR
AREELGLEGFVAVK----KGSPVYEKAKAYYE
>MMETSP0784-20121206|30352_1/439-527 [subseq from] 8sVR6+xZXTVyPAVx0vwgdBaY4kA Gymnodinium catenatum
------KP-----SKKMIKAIAFRGGDTDGATSLRKEDLMKNSAGRIVSKRKSEAAKKRFQDIGKWTLAVKR
AREELGLQGFVAVK----KGTPVYDRARELYQ
>Azadinium-spinosum-3D9-20130829|67964_1/78-169 [subseq from] uCzBIv4sszKeVR+hICFx3BgANRk Azadinium spinosum 3D9
---RRKKP-----SKKMVKVIAFRGGDTDGRTKLKSEDLIKNSQGKIVSKKASEAAKKRYADGCRWNDAVKR
ARKEMGVQGFVAVK----KGTPLYEKAQEYFK
>Ceratium-fusus-PA161109-20140214|990_1/546-631 [subseq from] Pj1fTB1oxrHnjtSRdwXao6KwaiU Ceratium fusus, Strain PA161109
----------TAKMAKAIVFKGGDAEGHTSLRKEDLVKSRSGKIVSKKQSEAAKKRFAEIGKWSTAVKR
AREELGLSGFVAVK----KGSAVYDKAREYY-
>MMETSP0448-20130528|93921_1/83-169 [subseq from] 8EDfXkBhW2VFaCRDl0YReGjnuWE Heterocapsa triquetra

----------SKRMARVVAFHGGDAEGGTRMRKEDLIKNSSGKIVSKKMSEAAKKRFQDISQWSASVKR
ARTELGLAGFVAVK----KGSPVYEKARAYYE
>Prorocentrum-minimum-CCMP1329-20131001|261503_1/376-467 [subseq from] reYC47fhkiAD6LuKnnyhZk2SPzU Prorocentrum minimum CCMP1329
---KRRKP-----SAKAAKRVAFHGGDTEGRTKLRKEDLMMNKAGKIVSKKQSDAAKKRYSDIGKWAVAVKR
AREELGIQGFAAVK----KGTPIYEKAREFYE
>MMETSP0766_2-20121228|17402_1/21-111 [subseq from] EUh/b1rwiGBdo+TFWXxyKdepmxA Gambierdiscus australes
---RRRQP-----TKKMAKAVAFRGGDTDGLTKLKREDLMRNRSGRIVSKRASEAAKKRGNPLGKWTMAVKR
AREELQLEGFVSLK----KGSPVYDKAMEYY-
>Amphidinium-carterae-CCMP1314-20130924|1945_1/387-478 [subseq from] OPpo5IzgM8iP6yfVKYIcmzot2qA Amphidinium carterae, Strain CCMP1314
---KNKKP-----TKKQAKALAFKGYDTGGATKLTKDDLTKNKHGKVVSKRASEASKKRFAEFGKWTEAVKR
ARIDLGIVGFAAAK----KGSPLYDKAQEYYK
>MMETSP1374-20130617|48538_1/364-452 [subseq from] qKtDihITUeXwPwj3CzzTYVaGd0k Symbiodinium Mf12.5f/clade A3
-----GKP-----TKKEAKAIAFKGYDTGGATKLTKDDLCRNKNGKVVSKKASEVAKKRFADLAKWSDAVKR
ARADLGIVGFAAAK----RGSPLYEKAQEYF-
>MMETSP1338-20131121|15174_1/397-484 [subseq from] jaEQ+nuunX6m/60fzpjeGRWiPNw Pelagodinium beii
-------P-----SKKMARAVAFRGGDAGGKTHLKKDDLVKNSQGKIVSKKASEAAKKRVNELGKWTEAVKR
ARTDLGVAGFEAVK----KGSALYEKAQEHYK
>Symbiodinium-sp-C1-20140214|1343_1/421-506 [subseq from] 03IwJ/ghN/2pTAF3gWpyC6BugWI Symbiodinium, Strain C1
----------SKKMARAVAFRGGDTGGVIRLQKEDLTKNSRGRIVSKRASEASKKRVNEIGKWTEAVKR
ARTELNVSGFAVVK----KGTELYEKAQEHF-
>Symbiodinium-sp-CCMP2430-20130923|33300_1/1-84 [subseq from] 4GHvNfHXeSwymuO4nPKj5m2mJ0k Symbiodinium sp CCMP2430
----------MAKAVAFRGGDTGGVIRLKKEDLVKNSRGKIVSRRASEASKKKSNEFRAWTAAVKR
AREELGVSGFAVVK----KGTPLYDKAQEIYK
>MMETSP0797-20121207|31586_1/129-217 [subseq from] JrsiDZcxEG8EP1z1dfD2oQWjmnQ Dinophysis acuminata
-----KKP-----SKQRAKVIAFYGGDTGGVTKLKAQDLIKNKHGKIVSRKASEASRKRFADTSKWSEAVKK
ARAELGVEGFVALK----KGTPFYERAQHYY-
>Karlodinium-micrum-CCMP2283-20140214|86869_1/416-505 [subseq from] jV8xo1f0vKP5PW40FsfP3DbLsgM Karlodinium micrum, Strain CCMP2283
-----KKE-----SKKRARILAFRGGDTGGRTRLGKDDLMRNSQGKIVSKRASESAKRRMQAFGKWTDAVKR
AKEELGLKGFVKVK----KGTDVYEKAQAIFK
>Karenia-brevis-SP1-20130916|215369_1/227-316 [subseq from] OquGM8reAl5Ci4BPDnw34n8qm9M Karenia brevis SP1
-----KKP-----SKKMAFRVAFHGGDASEHTKLRASDLIKNQYGKIVSKKASEAARKKVQSIGKWTQCVKK
ARQEMNITGFLAIK----KGTPVYERAQEIYK
>Scrippsiella-hangoei_like-SHHI_4-20140214|32433_1/139-230 [subseq from] bWFQXEr207NgDFtEOIPD6k2B/Bs Scrippsiella hangoei-like, Strain SHHI-4
----------VTAVGPYSKSMVWPGARVKTSCGLTKKDLKRNARGRLVSKRASNASHKKKHALDAWNEACRR
ARLELHCHGWVPVGGKTVQGKQLLRRARAI--
>Scrippsiella-hangoei-SHTV5-20131105|13169_1/165-256 [subseq from] Zg4ije29SWVI35KT57+VaP1zzsU Scrippsiella hangoei SHTV5
----------VTAVGPYSKSMVWRGARVKTSGGLTKKDLKRNARGRLVSKRASNASHNKKHALDAWNEACRR
ARLELHCHGWVPVGGKTVQGKQLLRRVRAI--
>MMETSP1439-20131203|74199_1/118-207 [subseq from] EqD/1biuQBurUFiiX5L5cBokH5Y Gonyaulax spinifera, Strain CCMP409
---------VMTA-----RREVLRGKITRTRGGLKSEDLVKNNRGKIVSKKSSDSSKRRYAALGKWMDAVKE
ARLALQITGFVPVGGKTAKGQELHARAKKIYE
>MMETSP0253-20130528|38973_1/19-118 [subseq from] b+W83bGZ1sJBnYbolmxRk/vyjFw Noctiluca scintillans
---TKKVAQANYKRKGWAMRLVFNGTKEKTSGGLKKSDLMKNKAGRVVSKKMHANGKLRYARLDAWSRGLVE
ARRELNLKGFVAVRSDSQSGTSLYARVAELY-
>MMETSP0253-20130528|32960_1/17-113 [subseq from] m3fEtqbEoUfMt2SQhYcXNioL7A4 Noctiluca scintillans
------VAQKHYKKKNYAKRLVWAGSKKRTSGGLGKADLMKSKTGKIVSKKQHANGKVRYIKLDAWSRGVVA
ARKEMNIQGFVPIGSNASTGTALYARVAELY-
>MMETSP0467-20121206|8602_1/74-171 [subseq from] v+SUYP4kmSc+wk+SqT3V+RqDmLI Mesodinium pulex, Strain SPMC105
---RASIAPTDWKRGRAAYRLVWTGRKKKTSGGLKRENLVMNKRGRVVSKRRHNLAKVSFKRLDAWNHCMMR
ARDELKLSGFVKVRGGTKQEAELYAKSEE---
>Scrippsiella-trochoidea-CCMP3099-20130930|26211_1/82-178 [subseq from] 5q+yGkZI5KMS2oX8ErJMq+rO1kI Scrippsiella trochoidea CCMP3099
---AKGVAQSSFARRNWAKRLVFKGKKVKTSTGLFKRDLMRNARGSIVSKRKSAIGRQRYHELDLWTRSFMQ
TRSEQGIKGFVLCKAGNSTERSLFSTTR----
>Durinskia-baltica-CSIRO_CS-38-20140214|147837_1/70-163 [subseq from] ybfAyxFprVG1K6nEaufXm/A6SxI Durinskia baltica, Strain CSIRO CS-38
---AKGVAQSSFARRNWAKRLVFLGKKVKTSTGLKKRDLVRNKRGSIVSKRKNAIGRRRYHELDLWTQSLMQ
VRDDKNIKGFMLCKGGTRAERELYL-------
>Durinskia-baltica-CSIRO_CS-38-20140214|45334_1/37-130 [subseq from] +NLhrW7K7ULJucxTtR3zv1mdNwA Durinskia baltica, Strain CSIRO CS-38
---ASGVAQSNYTRRNWAKRLVFLGKKDKTNTGLVKRDLMRNKRGTIVSKRKSALARRRYHEIDLWTEALLQ
VRDDQNIKGFLKCKKGTAAERMLYQ-------
>Peridinium-aciculiferum-PAER_2-20130926|92832_1/80-167 [subseq from] DWN9fhD8J4cpY+/hCD6Qc8wroxc Peridinium aciculiferum PAER_2
---AKGIAQSAFSRKNWAKRLVWKGTKVKTSTGLAKRDLCVSKRGGIVSKRKSSIAKSRYVSLDLWTTSLMQ
AREQHGAVGFVLCKAGTKL----------
>Scrippsiella-hangoei_like-SHHI_4-20140214|23282_1/80-166 [subseq from] 1GFZuUaV3bqXsBadgdS6QUlBqeU Scrippsiella hangoei-like, Strain SHHI-4
---AKGVAQSSFSRKNWAKRLVWKGTKVKTSTGLAKRDLCVNKRGGIVSKRKSSIAKQRFVSLDLWCKSLCQ

ARDQHGAVGFVKCKAGTK----------
>Scrippsiella-hangoei-SHTV5-20131105|77782_1/78-164 [subseq from] dsQINF2EfY1mwGMKnnAXGk/SP0w Scrippsiella hangoei SHTV5
---AKGVAQSSFSRKNWAKRLVWKGTKVKTSTGLAKRDLCVNKRGAIVSKRKSSTGKQRFVSLDLWCKSLCQ
ARDQHGAVGFVKCKAGTK----------
>Kryptoperidinium-foliaceum-CCMP1326-20130916|259_1/34-122 [subseq from] OaNoYghGnUac9CYM1JRNk5i4pB4 Kryptoperidinium foliaceum CCMP1326
---SRTIAQAQYKRKNWAKRLVFFQKKVKTSGGLKKRDLVRNRAGKVVSKRKSAIGVRRYHEIDLWAQALTD
VRAEMNVTGWLLCKAGTKQE----------
>MMETSP0796-20121207|787_1/134-214 [subseq from] /77hMiYpaJPZfshHLDwS2bKeusI Pyrodinium bahamense
--------QNQYKRKHHAKRLVFEGKCLKTSGGLHKRDLKRNKRGSIVSKRLSQLASRRYSELDLWVESFME
VRNERGLTGFVPLKGGA----------
>Ceratium-fusus-PA161109-20140214|4555_1/42-133 [subseq from] trGmysJekVKDdnk7Uh4rBdcHwyk Ceratium fusus, Strain PA161109
-------QQVKIAKGKNMRAAVFLGSKVKTASGHKKSDLMKSKSGKIVTKKAHSAGKKAYANIKKWTESCKK
VRKEMRLKGFVPCKKGT----AFYKAVKALY-
>MMETSP0224-20130122|44910_1/300-397 [subseq from] 7UVOlZ8+hAE/s/llQa/lrutRPRA Togula jolla, Strain CCCM 725
MKKKKVKKVSMIARGKLMRSRVLKGEKLKTYTGLKASDLMKNIYGKVVTKRQSEAGKKAFKNIEPWMKACVQ
ARRDFCIVGFRAVK----KGTAFYARAKEI--
>MMETSP0467-20121206|45446_1/2-91 [subseq from] nb+xcsnBK2Patayz4uf3CzbT3ys Mesodinium pulex, Strain SPMC105
---MRAKKASDIAKGKRARASVWGGKASRTVSGLTKDNLMRNKRGTIVPKKKHANGKKAFSNLGPWLKACQE
ARKALGLSGFIPIKGRTPQGK----------
>Scrippsiella-trochoidea-CCMP3099-20130930|52010_1/87-189 [subseq from] k5Y4gblgmpvFWFLu8G2oOKTu1oM Scrippsiella trochoidea CCMP3099
MKAGKARNTRTVPRFKVQRRQVFFGREGKTRGKMEQKDIMRNAYGRLVSKKASASSKRTHMTAKPWSDSVKE
ARKELGLTGFVLLGGKTAAGKALLAKVRAIY-
>Scrippsiella-trochoidea-CCMP3099-20130930|169056_1/75-176 [subseq from] gK0mw2o3Q04FfxeFYWbZE0v6mF8 Scrippsiella trochoidea CCMP3099
-RAAKVKGLYTISKAKLERVQVFLGRKTSTRGRREKKDFIKNKYGRIVSKKVSANSKRNHANAKPWVDSIKQ
ARKELGLTGFVLLGGKTAAGKALLAKVRAIY-
>MMETSP0795-20121207|7934_1/18-117 [subseq from] jmmQfe8guIQBdlDUni5m4YNE9vs Amoebophrya sp.
-KSMK--RGCTVAKGPGARQKVWNGEAVRTKAGLYKHHLKKDARGKIVALSRSELSKKHYKKIERWNKACGK
ARTRLHLEKWTPVGGKTPQGKCVIAIARDVY-
>a|EV_3300002603.a:JGI24921J38492_100053_5/15-71 [subseq from] # 3777 # 4028 # -1 #
ID=55226_5;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.603
-----------SGLKRNRTGEIVSIKKSADAKKKSAR-----KAVDA
AKKKLGMGEFVPIK----KGTPLYREAKKI--
>MMETSP0796-20121207|8929_1/147-246 [subseq from] z22dCrs6h48vucyS13F9kMUBcVA Pyrodinium bahamense
PKRSKAK-KARVALRKRERASVFAGRLERTRSGLAQVDLERNRAGKVVNKAKSRIGKKLDR-VKPWRESVMR
AREEMSVQGFVAVKGDTAIGRALYEKAQRL--
>MMETSP0795-20121207|16073_1/175-276 [subseq from] 3BaIYgCLHgwX4w71w4Ifxl0NIss Amoebophrya sp.
MKKMKKK-KDKFKRRTTKKWLVYQGRRLRTTGGLTKEGLKRNKRGKIVSRKAAAATMRKPQTVLNWAGCLKA
ARQLLGLVGFVPCGGKTEAGQRLYRETRRLY-
>MMETSP0689_2-20121128|31461_1/314-410 [subseq from] yDeBtsG1S2RZzs8peVEznwJDLLY Amphidinium massartii
-KKARAKKKTKL-KRRSVNSLVLKGSLPRTKGGLKAEKLVKNAKGRVSSKKMFDRGMKTWKNLGPWVEAMLK
ARAELGLVGMVGCK----KGTPFYTKARELY-
>a|EV_3300001348.a:JGI20154J14316_10000048_77/6-75 [subseq from] # 61496 # 61744 # -1 #
ID=29546_77;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.369
----KAV----------TKLEVWKGKAKHTSGGLTKDKLMKNKRGKVISKKKHAAG----------IKM
ALKKLGYTTFGVFR----HGKKVTKKSKK---
>Karenia-brevis-SP1-20130916|24438_1/39-131 [subseq from] XdV6TmzK8FAiXFhS7LSXqJ5H8NU Karenia brevis SP1
---KRSKRQ-SVTRPGYAYRLIWEGKKTRTHSGLTRKDFKLNAKGRVVPIRRSAHSRNNWQNARSWSMSLKQ
VRKDMGLEGQVWPKKSGSDVQKIYY-------
>Karenia-brevis-Wilson-20130916|14999_1/62-154 [subseq from] KaTNxzFngGOT5G80fd4BZtNEdHI Karenia brevis Wilson
---KRSKRQ-SVTRPGYAYRLIWEGKKTRTHSGLTRKDFKLNAKGRVVPIRRSAHSRNNWQKARSWSMSLKQ
VRKDMGLEGQVWPKKSGSDVQKIYY-------
>Karenia-brevis-SP3-20130916|20271_1/26-118 [subseq from] 3t0P2lerSPQK/4y0V7DggXwem44 Karenia brevis SP3
---KRSKRQ-SVTRPGYAYRLIWEGKKTRTHSGLTRKDFKLNAKGRVVPIRRSAHSRNNWQRARSWSMSLKQ
VRKDMGLEGQVWPKKSGSDVQKIYY-------
>Azadinium-spinosum-3D9-20130829|9955_1/74-161 [subseq from] JQu+wdZN/MsJD35P2cugrwSbbwQ Azadinium spinosum 3D9
--RKRIARTCDFSSRARMLRAVWKGKFQKTPGMLTKADLRLNKRGRVVPIRKSADGQRKAEAMRDWFSSVSW
ARKELGIVGCLLKKKNGS----------
>MMETSP1439-20131203|8963_1/150-232 [subseq from] dlFI5UvicXs4NmZ3ALm/QQ5atbM Gonyaulax spinifera, Strain CCMP409
----------AMRRAVWEGRMKYTSGRLTKNDLELNSKGAIVPKRRSRHSRVLCKAMADWYDALRE
ARSRLQLEGFVLVKKSGNQQQALYRAA-----
>MMETSP0228-20121227|18568_1/112-194 [subseq from] kl4Keg+z3GjZFqq1BI1eYqZwNkc Protoceratium reticulatum, Strain CCCM 535 (=CCMP 1889)
----------EMRKAVWEGKYYRTSGRLTKEDLFVNPKGIIVPKRRSEHSQRVCKAMSDWYDAMRY
ARKKLGVTGFVLVKRNGTLQDALYRAA-----
>Prorocentrum-minimum-CCMP2233-20131001|47985_1/166-253 [subseq from] MQkZqsfOc9tsyBZg4EhMTIuCBII Prorocentrum minimum CCMP2233
-------R-----SRAYMYRAIWQGRFKKTKHGIEKKDLCVSASGRIVSKKRSVHCRKLGTLWRMWCSATSQ
AREALGATGFVKVKRAGTEEATLFKQVA----
>Prorocentrum-minimum-CCMP2233-20131001|195159_1/184-260 [subseq from] j6Qw8o57AsNiDwQ5svFafMY2d3E Prorocentrum minimum CCMP2233

-------R-----SHAYMYRALWQGRFKRTTGGLTKDDLTVSITGKIVPKRRSAHCRKIGTEWRHWLRATRD
ARAALGVDGFVKVKKNG----------
>MMETSP0227-20121206|45523_1/147-236 [subseq from] 95zD/sRXCxnmyXG7IX6I2wPzjoo Polarella glacialis CCMP1383
---MKSIRVRDYKKKGHAYKLVFEGKKKRTSGKLFKRDLMRNKRGRVVSKKRHARGCKLYKELDKWMAAVQQ
TKEDLGLTGFVLMKKPENGGK----------
>MMETSP1440-20131203|5516_1/116-205 [subseq from] P7j7faorUBg+CuUSLtEc7d0T1EY Polarella glacialis, Strain CCMP2088
---MKSIRVRDYKKKGHAYKLVFEGKKKRTSGKLFKRDLMRNKRGRVVSKKRHARGCKLYKELDKWMAAVQQ
TKEDLGLTGFVLLKKPENGGK----------
>MMETSP1440-20131203|17264_1/17-99 [subseq from] 5FNb9QuuIoW4bC/6g+TaDyyHLOw Polarella glacialis, Strain CCMP2088
---KKEKKVSKIAKGKLAKSVVSKDDKEMIATVLKKTDLMKTKKGKFVSRMPHDSGKKDYAHIKDWTTAYQN
ARKDFGIKGCVAIK----------

Appendix 3: Multiple alignment of viral DVNP sequences

>a|RefSeq_gi|197322349|ref|NC_011183.1|_Feldmannia_species_virus,_complete_genome_116/2-80 [subseq from] # 106737 # 107039 # -1 # ID=267574_116;partial=00;start_type=ATG;rbs_motif=GGA/GAG/AGG;rbs_spacer=5-10bp;gc_cont=0.512
----------KTIGSRANVFHGTARKTSGGLVKGDLKLNKNGSIVSKKNSLRAKRTSPL-----KAWRK
AVKNTY-EGFCII----KKGTPFYREVNAE--
>a|RefSeq_gi|13242472|ref|NC_002687.1|_Ectocarpus_siliculosus_virus_1,_complete_genome_81/5-83 [subseq from] # 80854 # 81222 # 1 # ID=265518_81;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.499
----------SNVGSRAQVMHGTAKKTSGGLEKEDLAYNKSGSIVSKKKSLEAKKSNGL-----KLWRK
AMQEVS-MDFVKP----KRGSVVHAKVFAE--
>a|EV_3300002603.a:JGI24921J38492_100884_35/24-95 [subseq from] # 32918 # 33229 # 1 # ID=55249_35;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.413
----------QRQGSRAQVMHGTALQTTGGLQKSDLVYNRSGKIVSKAKSKTAKKENS-------RLKK
K-----QRGFGVN----KL---GGKKTRKNRK
>a|EV_3300003941.a:Ga0007832_1000003_17/13-84 [subseq from] # 11447 # 11758 # -1 # ID=73845_17;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.571
----------RAIGSRAQVMHGTAHHTKGGLTKKHLKYNKYGKIVSARKSALAKSKG---------LRK
W------TGRWTI----GKPKVKHGKAKKGKK
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10000866_26/15-84 [subseq from] # 21650 # 21955 # 1 # ID=24018_26;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.618
----------RAVGSRAQVMHGTAHHTKGGLTKKGLKYNKYGKIVSTRKSASAKSKG---------LKK
W------TGRWTI----GKPKVAHGKAKKG--
>a|EV_3300002835.a:B570J40625_100000564_70/3-70 [subseq from] # 59978 # 60211 # -1 # ID=61205_70;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.556
----------AVGSRAEVMHGTADHTSGGLTSGDLKYNKWGRIVSRRKSASARKEK---------LHK
A------AGFGAV----K---IGDKKTRRHRK
>a|EV_3300003431.a:JGI25913J50563_1000049_27/9-71 [subseq from] # 17630 # 17905 # -1 # ID=67894_27;partial=00;start_type=ATG;rbs_motif=ATA;rbs_spacer=3bp;gc_cont=0.536
----------RIGSRAQVMHGTAHHTTGGLTKKALKYNKWGKIVSVKKSMKAKKEK---------LEK
M------TGFGVV------------KTKKGKK
>a|EV_3300000439.a:TBL_comb48_EPIDRAFT_1000293_37/5-64 [subseq from] # 23632 # 23850 # -1 # ID=23238_37;partial=00;start_type=ATG;rbs_motif=AAA;rbs_spacer=3bp;gc_cont=0.434
------M------KKVGSRAQVMHGTAVKTSGGLTRKDLKYNKSGRIVSVKKSHTAKKDK---------LEK
H------TGFGAI------------KH-----
>a|EV_3300000115.a:DelMOSum2011_c10000012_52/4-64 [subseq from] # 65236 # 66042 # -1 # ID=13735_52;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.325
------K------QTTGSRAQVMHGNATKTSGGLTKSQLKYNKQGKIVSKKASALAKRNN---------LVK
A------TGFGIS------------MKG----
>a|EV_3300000115.a:DelMOSum2011_c10000222_32/4-63 [subseq from] # 24235 # 24906 # -1 # ID=13176_32;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.293
------K------QTTGSRCQVMHDNATKTSGGLTKSQLKYNKQGKIVSRKASALAKNNN---------LVK
A------TGFGVS------------KM-----
>a|EV_3300000115.a:DelMOSum2011_c10000222_19/4-64 [subseq from] # 12689 # 12982 # -1 # ID=13176_19;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.323
------K------QITGTRAQVMHGTANKTSGGLTKSQLKYNKQGKIVSRKASALATKNN---------LVK
A------TGFGIV------------KTR----
>a|EV_3300000115.a:DelMOSum2011_c10000130_48/4-64 [subseq from] # 44379 # 44987 # -1 # ID=13737_48;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.325
------K------QTTGSRAQVMHGTARKTTGGLTKSQLKYNKQGKIVSRKASTLAKKNN---------LAK
A------TGFGSK------------KMK----
>a|EV_3300002599.a:JGI24920J38491_100212_2/4-67 [subseq from] # 1062 # 1331 # 1 # ID=55155_2;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.400
----------KTYGSRAEVFHGKAKKTSGGLIKKNLKKNKWGRIVSVRLSNLAKKQK---------LEK
A------AGFGYV------------KTGKAKK
>a|EV_3300002466.a:JGI10211J34971_10001083_5/7-67 [subseq from] # 2287 # 2514 # -1 # ID=48180_5;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.430
------M------MSVGSRAQVWHGTAKKTSGGLTKKDLKKNKAGRIVSVKMSNRAKKER---------LEK
A------PGFLLF------------K----KK
>a|EV_3300001349.a:JGI20160J14292_10006195_13/1-66 [subseq from] # 7846 # 8097 # 1 # ID=30186_13;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.409
----------MTVGSRAQVFHGTVNKTTGGLEKKDLMKNKHGRIVSVKKHKTAKKEN---------LKK
A------TGFGSF----KP------KTKKNKT
>a|EV_3300003591.a:JGI26250J51715_1000278_15/1-69 [subseq from] # 15099 # 15344 # 1 # ID=69048_15;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.402
----------MTVGSRAQVFHGTVDQTTGGLEKKDLMKNKHGRIVSLKKHKTAKKNN---------LKK
A------TGFGSF----KK---EDNKTRKNKS
>a|EV_3300001348.a:JGI20154J14316_10003085_10/4-70 [subseq from] # 7585 # 7833 # -1 # ID=29569_10;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.382
----------VGSRAQVFHGNADKTSGGLEKKDLMKNKHGRIVSVRKHKTAKKEK---------LEK
A------AGFGFI----KK---EKSKTRKNKS
>a|EV_3300002933.a:G310J44882_10000138_18/19-89 [subseq from] # 20450 # 20965 # -1 # ID=62702_18;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.335
------D------MLEGSRAQVNNGTAYKTTGGLKKTDLFQNKNGRIVSRKKHITATKEK---------LVK

A------TGFGYV----KLD--GSRKSKKSRK
>a|EV_3300001352.a:JGI20157J14317_10001110_24/20-89 [subseq from] # 15940 # 16251 # 1 #
ID=30986_24;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.311
------D------MLIGSRASVWHGTAYKTSGGLKKEHLMMNKNGRIVSKKKHNTAKKEK---------LVK
A------TGFGAV----YV---GTKKSRKSRK
>a|EV_3300002928.a:G310J44881_10000069_45/20-88 [subseq from] # 48837 # 49460 # 1 #
ID=62523_45;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.418
------E------MLEGSRAQVMHGTAFKTSGGLKKHELLQNKSGRIVSRKKHGTAKREN---------LVK
A------TGFGAV----KL---GS-HSRSRRS
>a|EV_3300000439.a:TBL_comb48_EPIDRAFT_1000102_2/21-89 [subseq from] # 2027 # 2419 # -1 #
ID=23023_2;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.356
-------------TLEGSRAQVMHGTAYKTSGGLKKSDLLQNKNGRIVSKKKHGTAKKEN---------LVK
A------TGFGAV----KL---GSKSSRSKKS
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10000207_51/20-85 [subseq from] # 52594 # 53046 # 1 #
ID=24832_51;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.406
------E------KLEGSRAQVVHGTAFKTSGELTKKDLLQNKNGRIVSRKKHVLAKKEK---------LVK
A------TGFGAV----RL---SGHHSS----
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10002566_8/20-89 [subseq from] # 9701 # 10138 # -1 #
ID=24879_8;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.395
------D------MLIGTRAQVWHGTAYKTSGGLTKANILQNKNGRIVSRSKHASAKREN---------LVK
A------TGFGAV----KLN--GKSSRRHSK-
>a|EV_3300002933.a:G310J44882_10000034_12/20-88 [subseq from] # 26220 # 26696 # 1 #
ID=62618_12;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.491
------E------MLIGTRAQVWHGTAFKTSGGLTKGHLLQNKAGRIVSRTKHATAKKEK---------LVK
A------TGFGFV----ML---NKSSSKSKS-
>a|EV_3300000736.a:JGI12547J11936_1000035_20/20-87 [subseq from] # 19160 # 19576 # 1 #
ID=26298_20;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.400
------E------MLMGSRAQVWHGTAYKTSGGLTKNNLMQNKAGRIVSKDKHMTAKKEK---------LLK
A------TGFGFV----KL---GKSSKHHG--
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10000539_28/21-87 [subseq from] # 31404 # 31709 # -1 #
ID=24296_28;partial=00;start_type=ATG;rbs_motif=AATAA;rbs_spacer=7bp;gc_cont=0.497
-------------ILIGTRAQVWHGTAYKTTGGLTKTHLFQNKAGRIVSAKKHATAKKEK---------LIN
A------TGFGFV----KI---GTKKRRGK--
>a|EV_3300003429.a:JGI25914J50564_10000155_14/20-89 [subseq from] # 14384 # 14704 # 1 #
ID=67532_14;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.402
------E------RLMGSRAEVWHGTAYKTSGELCKHHLMQNKHGRIVSKAKHETAKKEK---------LLK
A------TGFGFV----KV---GTKHNRSNGK
>a|RefSeq_gi|944325232|ref|NC_028104.1|_Yellowstone_lake_mimivirus_DNA,_complete_genome,_isolate:_1_95/22-88 [subseq from] #
67738 # 68175 # 1 # ID=265899_95;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=7bp;gc_cont=0.429
-------------LVGTRAQVWHGTAYKTSGGLCKPDLMQNHAGRIVSKAKHATAKKEM---------LLK
H------TGFGFV----KL---SGHHGRKSR-
>a|EV_3300001347.a:JGI20156J14371_10000545_3/2-67 [subseq from] # 1241 # 1477 # 1 #
ID=29246_3;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.329
---------------VGSRAQVWHGTAKKTSGGLTRKDLLKNKAGRIVSKRKHFTAKKDK---------LVK
A------TGFGFI----KK---DSKK-KKSKT
>a|EV_3300003429.a:JGI25914J50564_10000017_56/22-80 [subseq from] # 51409 # 51768 # 1 #
ID=67511_56;partial=01;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.381
-------------LIGTRAQVWHGTAYKTTGGLTHSDLMKNKSGRIVSKAKHNTAKKDK---------LIK
A------TGFGFV----KK---T---------
>a|EV_3300000119.a:KGI_S1_ANT01_95mDRAFT_c10000266_2/21-89 [subseq from] # 1457 # 1762 # 1 #
ID=18474_2;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.330
------D------MLFGSRAQVWHGTAYKTKGQLKKENLMMNKRGRVVSKRKHNTAKRDK---------LIK
A------TGFGFV----RK---DGSSSKSKK-
>a|EV_3300002466.a:JGI10211J34971_10000114_34/20-88 [subseq from] # 24605 # 24910 # -1 #
ID=47326_34;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.265
------D------KLIGSRAQVWHGTAYKTKGDLVKDDLMMTKRGKIVSKRKHFTAKKEK---------LEK
H------TGFGYV----KK---TPKKHRTKK-
>a|EV_3300002466.a:JGI10211J34971_10000183_14/20-86 [subseq from] # 10580 # 10933 # -1 #
ID=48585_14;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.266
------E------KLIGSRAQVWHGTAYKTKGDLRKEDLLMTKRGKIVSKRKHFTAKKEK---------LQK
H------TGFGYV----KK---TPLKNKH---
>a|EV_3300001348.a:JGI20154J14316_10000227_34/24-89 [subseq from] # 28156 # 28473 # 1 #
ID=29508_34;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.314
----------------GSRAQVWHGTAYETTGGLKKGDIKMNKHGRIVSTKKSSTAKKEK---------LEK
A------TGFGFV----KK---DSSSSRKSRK
>a|EV_3300001348.a:JGI20154J14316_10000269_26/24-87 [subseq from] # 24177 # 24506 # 1 #
ID=29754_26;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.385
----------------GSRAQVWHGTAYETTGGLRKGDLKMNKHGRVVSTKKSATAKKEK---------LEK
A------TGFGFV----KK---SSSS--KSRK
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10000225_37/19-86 [subseq from] # 36914 # 37456 # -1 #
ID=24423_37;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.389
------D------MLEGSRASVFHGTAYKTAGELTKSAFIMNKNGRIVSEKKHNTAKREM---------LVK
A------TGFGYV----RL---NGTKSQGR--

>a|EV_3300002601.a:JGI24917J38488_100824_10/22-89 [subseq from] # 13915 # 14235 # 1 #
ID=55211_10;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.502
----------ELEGSRAKVWHQTAYKTSGGLLKHHLIMNKHGRIVSKKKHGTAKKEK---------LEK
A------TGFGSV----HEG--D-SKGKKSR-
>a|EV_3300001352.a:JGI20157J14317_10000570_27/22-88 [subseq from] # 32152 # 32490 # -1 #
ID=31109_27;partial=00;start_type=ATG;rbs_motif=TATAA;rbs_spacer=4bp;gc_cont=0.507
----------HLEGSRAQVWHQTAYKTSGGLLKHHLIMNKHGRIVSKKKHGTAKKEK---------LEK
A------TGFGFV----KSE--T-KKVKKG--
>a|EV_3300002605.a:JGI24916J38487_100953_30/9-74 [subseq from] # 30269 # 30538 # -1 #
ID=55309_30;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=7bp;gc_cont=0.385
----------TGSRAQVWHGSAKKTSGGLEKRHLLMNKNGRIVSKKKHESAKREK---------LER
A------TGFGFV----KTH--NVKKSKKS--
>a|EV_3300002835.a:B570J40625_100003638_9/25-90 [subseq from] # 10965 # 11267 # -1 #
ID=61018_9;partial=00;start_type=ATG;rbs_motif=TAAA;rbs_spacer=5bp;gc_cont=0.574
----------GSRAQVFHGTAYKTDGGLTKEKLLMNKNGRIVSAKKHATAKREK---------LAK
H------AGFGAV----RIS--DLRKTKSRK-
>a|EV_3300002835.a:B570J40625_100010800_3/25-89 [subseq from] # 1972 # 2277 # 1 #
ID=61081_3;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=15bp;gc_cont=0.536
----------GSRAQVFHGTAYKTDGGLTKEKLLMNKYGRIVSAKKHATAKREK---------LAK
H------AGFGAV----RIS--DLK--KSRKN
>a|EV_3300003413.a:JGI25922J50271_10000002_5/25-88 [subseq from] # 5362 # 5673 # 1 #
ID=67134_5;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.577
----------GSRAQVFHGTAYKTDGGLTRDKLLMNKHGRIVSAKKHATAKKEK---------LEK
H------AGFGAV----RIS--ELKKTRS---
>a|EV_3300003413.a:JGI25922J50271_10000021_5/22-89 [subseq from] # 8145 # 8441 # -1 #
ID=67073_5;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.468
----------RLVGSRAQVWHQTAYKTSGGLTRAQLVMNKHGRVVSAKKHTTAKREK---------LVK
H------AGFGAV----KIG--AAKKTRKS--
>a|EV_3300003430.a:JGI25921J50272_10000003_73/39-106 [subseq from] # 86254 # 86622 # 1 #
ID=67740_73;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.442
------T------MLEGSRAQVWHDTAYKTPGGLTKSELVFNKHGRIVSAKKHATAKKEN---------LRK
Y------AGFGAI----KIN--SKTGRRA---
>a|EV_3300001354.a:JGI20155J14468_10000346_9/22-89 [subseq from] # 5640 # 5966 # -1 #
ID=31447_9;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.404
----------LEGSRAQVWHHTAYKTSGGLTRSHLMKNKHGRIVSKKKHTTAKRQK---------LEK
A------PGFVVM----RRS--MRKKKGKSK-
>a|EV_3300003591.a:JGI26250J51715_1000395_2/268-336 [subseq from] # 1039 # 2082 # 1 #
ID=69052_2;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=11bp;gc_cont=0.332
------K------ELFGSREQVHNGTAYKTKAGLTNDDILMNKWGRLVSAKKHETAKKEM---------LEK
H------AGFGYV----KK----GKNTRKNKK
>a|EV_3300001348.a:JGI20154J14316_10000788_30/20-86 [subseq from] # 27457 # 27738 # -1 #
ID=29760_30;partial=00;start_type=ATG;rbs_motif=AAAA;rbs_spacer=12bp;gc_cont=0.333
------K------ELFGSREQVYNRNAYKTKAGLTFDDILMNKRGRIVSAKKHKTAKKEQ---------LEK
Y------AGFGYV----KK----NK-TRKKK-
>a|EV_3300000188.a:SI60aug11_150mDRAFT_c1000054_20/22-89 [subseq from] # 24991 # 25296 # 1 #
ID=19614_20;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=9bp;gc_cont=0.359
----------ELFGSRTQVMNQTAYKTSGNLSKNDLMMNKWGRIVSAKKHKTAKKEK---------LEK
A------AGFGYV----KK----TRRTRKNTS
>a|EV_3300001348.a:JGI20154J14316_10000342_22/22-87 [subseq from] # 28871 # 29173 # 1 #
ID=29510_22;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.370
----------ELFGSRTQVMNKTAYKTSGNLTKNDLLMNKWGRIVSAKKHKTAKKEK---------LEK
A------AGFGYV----K------RKTRKNAT
>a|EV_3300002599.a:JGI24920J38491_100388_28/21-86 [subseq from] # 23951 # 24247 # -1 #
ID=55163_28;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=10bp;gc_cont=0.492
----------ELFGSRTQVMNGNAYKTSGNLAKKDLLMNKWGRIVSAKKHKTAKKEK---------LEK
A------AGFGYV----K------KQTRKNRK
>a|EV_3300002447.a:JGI24768J34885_10000087_16/21-87 [subseq from] # 17025 # 17315 # -1 #
ID=46518_16;partial=00;start_type=ATG;rbs_motif=TAAA;rbs_spacer=9bp;gc_cont=0.416
------K------ELFGSREQVVNGTAYKTSGELTVEDLVMNRWGRIVSAKKHKTAKKEK---------LEK
Y------SGFGYV----K------KGTRKAKR
>a|EV_3300003430.a:JGI25921J50272_10000258_5/20-86 [subseq from] # 4961 # 5245 # 1 #
ID=67696_5;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=13bp;gc_cont=0.460
------K------ELRGSRAQVHNRTAYKTTGGLTYSDLVMNKWGRIVSADKHRTAKKEK---------LEK
A------AGFGYV----KK-----NGTRKVK-
>a|EV_3300002447.a:JGI24768J34885_10000087_15/2-63 [subseq from] # 16565 # 16933 # -1 #
ID=46518_15;partial=00;start_type=ATG;rbs_motif=TAAA;rbs_spacer=8bp;gc_cont=0.472
----------QTVGSRAQVWHGTAQHTSGGLTKKNLIKNKWGRIVSAKKHKTAKKQK---------LEK
A------AGFGVV----K------RETM----
>a|EV_3300003591.a:JGI26250J51715_1000244_8/25-92 [subseq from] # 8611 # 8919 # 1 #
ID=69041_8;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.372
----------QLEGSRAQVWHGTAFRTPGGLKKHNLMQNKRGRIVSLRKHKTAKKQK---------LEK
A------TGFGFV----KK----HKKTHKKKH

>a|EV_3300000115.a:DelMOSum2011_c10001435_7/23-88 [subseq from] # 7351 # 7707 # 1 #
ID=13074_7;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.387
----------LIGSRAQVHHGTAYKTSGGLTKGDLVKNKNGRIVSKKVQATAKKQR---------LAK
A------AGFGAV----KMP-LGKAKSK----
>a|EV_3300000115.a:DelMOSum2011_c10000253_6/23-88 [subseq from] # 6780 # 7082 # 1 #
ID=12944_6;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=6bp;gc_cont=0.475
----------LIGSRAQVHHGTAYKTAGGLTKEKILMNKNGRIVSRKKHATAKKEK---------LEK
A------AGFGAV----RKD--GKNVTRR---
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10000893_17/20-88 [subseq from] # 15617 # 15985 # 1 #
ID=24847_17;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.480
------E------MLVGSRAQVHHGTAYKTSGGLKHHEILQNKNGRMVSRLKHTTAKKER---------LAK
H------AGFGPV----RMG-SHASTSRK---
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10000680_30/20-89 [subseq from] # 24188 # 24580 # 1 #
ID=24428_30;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.295
------E------KLEGSRAQVYHGTAYKTSGNLKKPEIMMNKNHRIVSKRKHFTAKKER---------LIR
A------TGFGFV----KI---GKMKIGKSKK
>a|EV_3300002933.a:G310J44882_10000043_30/31-98 [subseq from] # 26478 # 26876 # 1 #
ID=62653_30;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.361
------Q------NLTGSRKQVYHGTAYKTTGGLTKDLLFFNKTNRIVSKKKHFSAKKEK---------LLR
H------TGFGAV----RI---G--KIGKRKM
>a|EV_3300002835.a:B570J40625_100008891_18/23-89 [subseq from] # 17680 # 17991 # -1 #
ID=60269_18;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.375
----------LIGSRAQIWHGTAYKTKGGLTKADLLMNKRGHVVSKKLYNRAKREK---------LEH
A------TGFGWV----R----DGTKRKGRKS
>a|EV_3300002835.a:B570J40625_100000596_19/23-88 [subseq from] # 33298 # 33606 # -1 #
ID=61741_19;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.353
----------LNGSRAQVWHGTAYKTKGNLKKPDLLMNKRGHVVSRKVYNRAKREK---------LEK
A------TGFGWV----RH---DNSKTRRR--
>a|EV_3300002835.a:B570J40625_100002705_16/21-88 [subseq from] # 15127 # 15600 # -1 #
ID=59448_16;partial=00;start_type=ATG;rbs_motif=TATAA;rbs_spacer=4bp;gc_cont=0.367
----------ELFGSRKQVWFGSAYKTSGGLTKDDLFYNSSRRIVSLKKHTTAKKEK---------LLK
Y------TGFGYI----RI----KSKSRKNKK
>a|EV_3300001347.a:JGI20156J14371_10001367_3/36-105 [subseq from] # 7015 # 7380 # 1 #
ID=29434_3;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=8bp;gc_cont=0.484
------M------SKTGSRRQVFGTSAHKTSGGLTKKDLVMSH-GRIVSKKKHLTAKKEK---------LEK
H------AGFGFV----KRK--TRRATRKNKS
>a|EV_3300003581.a:JGI26257J51711_1000485_2/36-105 [subseq from] # 171 # 536 # 1 #
ID=68879_2;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=7bp;gc_cont=0.470
------M------SKTGSRRQVFHKSAYATSGGLTKKDLVLSH-GRIVSKKKHLTAKKEK---------LEK
H------AGFGFV----KRK--TRRATRKNKS
>a|EV_3300002242.a:KVWGV2_10199049_16/8-78 [subseq from] # 20905 # 21171 # 1 #
ID=46205_16;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=9bp;gc_cont=0.371
------K------TTIGTRRMVHSGTAEHTSGGLTKKDLIKNKWGRIVSRSKHESAKKEQ---------LKK
Y------AGFGAV----KMK--KSRKNRKTKS
>a|EV_3300000137.a:LP_F_10_SI03_10DRAFT_c1000045_34/20-88 [subseq from] # 29592 # 29891 # 1 #
ID=18798_34;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.233
------K------KTFGSRREVYNGTAEKTTGGLFKKDLFKNKNGRIVSKKKHFTAKKEK---------LEK
F------TGFGHI----S-K--EDKKKKKKK-
>a|EV_3300000224.a:SI34jun09_10mDRAFT_1000003_84/20-85 [subseq from] # 99602 # 99901 # -1 #
ID=20027_84;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.287
----------GARRQVWAGTAYMTEGGLTRDKLHYNKRGRIVSKKKFNTAKKEK---------LQK
H------TGFGYV----KKG--TKSKSKSKS-
>a|EV_3300003429.a:JGI25914J50564_10000113_13/21-85 [subseq from] # 12908 # 13270 # -1 #
ID=67598_13;partial=00;start_type=ATG;rbs_motif=TATAA;rbs_spacer=4bp;gc_cont=0.347
------K------KLIGTRAEVWHGTAYKTAGKLTKGQLMKSKIGRIVSKSKHDSSVKEK---------LQK
Y------AGFGPV----KIDDVAP--------
>a|EV_3300003395.a:JGI25917J50250_1000375_1/22-81 [subseq from] # 152 # 424 # 1 #
ID=65891_1;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.531
----------ELIGSRRKVWNGTAFKTHGQLHRSDLMMSKKGEIVSKKKHFTAKREK---------LEK
A------AGFKLF----------TRK---
>a|EV_3300003413.a:JGI25922J50271_10000122_12/21-87 [subseq from] # 16160 # 16465 # 1 #
ID=67117_12;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=4bp;gc_cont=0.405
----------HLIGSRKQVWMGSAYKTDGQLVKSNFIMNKHGRIVSAKKHASAKKEN---------LVK
A------TGFGYV----LL---GDKKSRRR--
>a|EV_3300001348.a:JGI20154J14316_10000180_14/14-78 [subseq from] # 14918 # 15589 # -1 #
ID=29641_14;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=6bp;gc_cont=0.531
----------MVIGSRRQVHNGTAHHTSGGLTKDKLFMTKNGRIVSKAKHFSAKKDN---------LVK
A------TGFGFV----LK---G--KSRKS--
>a|EV_3300000151.a:SI53jan11_200mDRAFT_c1000021_34/8-75 [subseq from] # 36682 # 37017 # 1 #
ID=19021_34;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.357
------S------KKVGSRRQVFNGSAEKTSGGLKKSDLLLNKNGRIVSKKASASAKKNG---------LVK
A------AGFGAF----K----DGEKVKNAN-

>a|RefSeq_gi|939177347|ref|NC_028094.1|_Chrysochromulina_ericina_virus_isolate_CeV-01B,_complete_genome_77/20-85 [subseq from] # 72700 # 73029 # -1 # ID=265999_77;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.3
------D------KLEGSRAEVGHGTAYKTSGGLLAKDLVYVRN-RWKSKKKHETAKREQ---------LQK
H------AGFGYI----KK---TPKKSRS---
>a|EV_3300001352.a:JGI20157J14317_10000579_31/16-85 [subseq from] # 29571 # 29960 # -1 # ID=30983_31;partial=00;start_type=ATG;rbs_motif=GGA/GAG/AGG;rbs_spacer=5-10bp;gc_cont=0.410
------K------SLVGTRAEVMHGTAFKTGYGLTKINLKYNKNGRIVSKTKSAKGPALK--------QLRD
A------TGFGAV--------KTAKHGRKSKS
>a|EV_3300000115.a:DelMOSum2011_c10000612_19/35-102 [subseq from] # 16892 # 17281 # -1 # ID=12625_19;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.344
------K------MLVGSRSQVMNGTAYKTSYGLTKKHLKYNKQGRIVSVAKSSKKGKLA--------QLRR
A------TGFGAV--------KIGSKTRKK--
>a|EV_3300000167.a:SI39nov09_120mDRAFT_c1000389_6/21-91 [subseq from] # 4309 # 4650 # -1 # ID=19217_6;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.313
------Q------MLVGSRAQVMHETAYKTAGGLTKKGLKKNKHGKIVSRAKSSKGPQMK---------LHD
K------TGFGAI----KKG-KKGKKGRKKT-
>a|EV_3300001352.a:JGI20157J14317_10000162_42/23-90 [subseq from] # 37541 # 37867 # -1 # ID=31071_42;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=8bp;gc_cont=0.401
--------------LVGSRAQVMHKTAYKTTGGLTKKNLKKNKHGKIVSRAKSAKGPQMK---------LTD
K------TGFGAI----KKK-KRGKTAKKK--
>a|EV_3300001355.a:JGI20158J14315_10000795_11/18-83 [subseq from] # 7914 # 8345 # -1 # ID=31869_11;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.407
--------------LIGSRAQVMHGTAYKTSGNLKKKDLKYNKHKKIVSRAKSAKGPQMK---------LHN
K------TGFGFV------R-KDGKGSRKS--
>a|EV_3300000439.a:TBL_comb48_EPIDRAFT_1003900_3/30-91 [subseq from] # 525 # 821 # 1 # ID=22957_3;partial=00;start_type=ATG;rbs_motif=ATA;rbs_spacer=4bp;gc_cont=0.572
------K------MASGSKAQVWHGTARHTSGGLTKKDLMRHK-GKIVSRRKHAAGLKAI--------KLRK
L------AGFKLF--------------RKQK-
>a|EV_3300003429.a:JGI25914J50564_10000001_135/21-89 [subseq from] # 118022 # 118930 # 1 # ID=67510_135;partial=00;start_type=ATG;rbs_motif=GGA/GAG/AGG;rbs_spacer=5-10bp;gc_cont=0.352
------E------KLIGSRAQVWHGTAYKTSGGLTKTKLFKNKNGRIVSKNKHIQSTKEK---------LLK
Y--------YGYKS--FTLS---KKKTRKLK-
>a|EV_3300001349.a:JGI20160J14292_10002549_9/21-97 [subseq from] # 8599 # 8934 # 1 # ID=30023_9;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.411
------D------MLVGSRRQVWNGTAFKTQPGLTRKHLMQKKSGRIVSKKKSAQAKTKK---------LGA
YIT----KKKGKKF--MTKSKGMKKRTRRNKK
>a|EV_3300001349.a:JGI20160J14292_10000759_26/23-96 [subseq from] # 26497 # 26856 # 1 # ID=30213_26;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.389
-------------TLVGSRAQVWHGTAYKTNPGLTRKHLMQKKDGKIVSRKKSQTAKKQK---------LGK
YID----RRKGKKF--MAKGKTMKNNKRKG--
>a|EV_3300001460.a:JGI24003J15210_10000188_15/21-89 [subseq from] # 15460 # 15774 # 1 # ID=35977_15;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.400
------D------KLEGSRAQVWHGTAYKTSGDLKKSDL-KMHNGRIVSKKKSELARSQK---------LKG
HLQPKG-GGKGKK----------KKGTKKR--
>a|EV_3300000115.a:DelMOSum2011_c10000381_7/2-62 [subseq from] # 7536 # 7889 # -1 # ID=13179_7;partial=00;start_type=GTG;rbs_motif=TATAA;rbs_spacer=4bp;gc_cont=0.398
-------------KTIGTRAKVWHGTAEKTAGGLKKGDLMKNKRGEIVSKKKHALGLKAK---------LHN
A------AGFG---------------TKKQAT
>a|EV_3300001348.a:JGI20154J14316_10000781_4/21-85 [subseq from] # 3366 # 3698 # -1 # ID=29897_4;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.390
-------------ELKGSRAKVWHGVAAETNGGLTKSQLKKNKHGEIVSKAKSEKGPQLK---------LTN
K------TGFGFV-----------KKVKKGKK
>a|EV_3300003617.a:JGI26082J51739_10000819_25/2-69 [subseq from] # 18212 # 18475 # -1 # ID=69432_25;partial=00;start_type=ATG;rbs_motif=TATAA;rbs_spacer=4bp;gc_cont=0.341
-------------KTFGSRAEVWHGNAKKTSGGLTKKDLIQNKWGEIVSRKKHITAKKEKLE----------
--KY---AGFGAI----KKSPKTKKS-IK---
>a|tr|A0A0P0YM63|A0A0P0YM63_9PHYC/1-78 Uncharacterized protein OS=Yellowstone lake phycodnavirus 1 OX=1586713 PE=4 SV=1
----TAT------QAVGSRAQVMSGTAHHTSGGLTKKDLKYS-SGEIVSKDKSKSEKKNPIE------AVKK
AKKD---ITFVLV----SKGTDLYKR-AKEI-
>a|EV_3300003940.a:Ga0007839_1000047_4/3-79 [subseq from] # 1545 # 1811 # -1 # ID=73746_4;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.566
-----HT------QAVGSRAQVMNGTAHHTAGGLVKKDLKRNKTGEIVSKDKAKGAKANPIK------AVMK
AKKE---IGMAFP----KKGSELYKT-AKA--
>a|RefSeq_gi|944325656|ref|NC_028110.1|_Yellowstone_lake_phycodnavirus_2_DNA,_complete_genome,_isolate:_2_77/2-69 [subseq from] # 60442 # 60702 # -1 # ID=265901_77;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.
----THT------QAVGSRAQVMHGNAHHTAGGLVKKDLKKSKTGEIVSKDKAKSEKKNPIV------AVAK
AKKE---I----------KGFALVQ-------
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10004411_11/1-70 [subseq from] # 8108 # 8350 # 1 # ID=24481_11;partial=00;start_type=ATG;rbs_motif=AATAA;rbs_spacer=13bp;gc_cont=0.572
-------------MTVGSRAQVYHGNATHTAGGLTKKDLKMKD-GELVSKKKAKGSKSNPIK------AVAK
AKKE---I----------TGFALVQGRAREI-
>a|EV_3300002835.a:B570J40625_100000451_60/1-66 [subseq from] # 45366 # 45614 # -1 # ID=60447_60;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.542

----------MTVGSRAQVYHGNATETSGGLKKKDLKMKKTGEIVSKAKSKDEKKNPIQ------AVAK
AKKA---I----------KGFALVQG------
>a|RefSeq_gi|944325416|ref|NC_028108.1|_Yellowstone_lake_phycodnavirus_3_DNA,_complete_genome,_isolate:_3_115/1-65 [subseq from]
75315 # 75557 # 1 # ID=265900_115;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0
----------MTTGSRAQVFHGNADQTAGGLKKKDLKMIK-GEIVSKAKAKTEKKNPIK------AVAK
AKKD---I----------KGFALVQG------
>a|RefSeq_gi|944325416|ref|NC_028108.1|_Yellowstone_lake_phycodnavirus_3_DNA,_complete_genome,_isolate:_3_39/1-65 [subseq from]
23205 # 23447 # -1 # ID=265900_39;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.
----------MTIGSRAQVFHGNADQTAGGLKKKDLKMVK-GEIVSKAKAKTEKKNPIK------AVAQ
AKKE---I----------KGFALVQG------
>a|EV_3300000553.a:TBL_comb47_HYPODRAFT_10003945_23/1-66 [subseq from] # 14003 # 14251 # 1 #
ID=24893_23;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.506
----------MTIGSRAQVYHGNADRTAGGLTKKDLKKDSTGEIVSKSKSKGEKTNPIK------AVEK
AKKE---I----------KGFALVKG------
>a|EV_3300003388.a:JGI25910J50241_10000262_14/1-65 [subseq from] # 10918 # 11166 # 1 #
ID=65471_14;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.578
----------MTVGTRAQVYHGNXTETAGGLTKKDXKKDKSGELVSKAKSKDAKTNPIK------AVQK
AKKE---I----------KGFALVQ-------
>a|EV_3300002605.a:JGI24916J38487_100502_2/4-67 [subseq from] # 311 # 619 # 1 #
ID=55287_2;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.570
----------VGSRAQVFHGTATRTTGGLTKKDLRMVD-GRIVSKDKSKGAKTAPLR--LWLNAVAQ
AKKT---G----------KPKAM---------
>a|RefSeq_gi|985756931|ref|NC_010191.2|_Ostreococcus_tauri_virus_OtV5,_complete_genome_131/3-80 [subseq from] # 112140 # 112409 #
1 # ID=266206_131;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.507
------T------KPFGSRAEVFHGTAEKTTGGLRAKDLMLDNDGQIKSVAAHKAALARMKE---GKKHLTK
VFK----PGFALQ----KEGTKEYKKKIK---
>a|RefSeq_gi|939185678|ref|NC_028092.1|_Ostreococcus_mediterraneus_virus_1_isolate_OmV1,_complete_genome_137/7-84 [subseq from]
116368 # 116649 # 1 # ID=265820_137;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=
------T------KPVGSRAEVFHGTAAKTSGGLTQRDLMLDNDNQIKSVAAHKAALARMKE---GKKHLTK
VFK----PGFALQ----KEGTKEYKKKIK---
>a|EV_3300003617.a:JGI26082J51739_10005259_3/3-80 [subseq from] # 1237 # 1506 # -1 #
ID=69339_3;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=11bp;gc_cont=0.507
------T------KPIGSRAEVFHGTAAKTSGGLTQKDLMLDNDGQIKSVAAHKAALKRMRE---GKKHLTK
VFK----PGFALQ----KEGTAAYKKKMK---
>a|RefSeq_gi|314055095|ref|NC_014789.1|_Ostreococcus_tauri_virus_2,_complete_genome_129/3-80 [subseq from] # 104119 # 104388 # 1 #
ID=266514_129;partial=00;start_type=ATG;rbs_motif=GGA/GAG/AGG;rbs_spacer=5-10bp;gc_cont=0.504
------T------KPIGSRAEVFHGTAEKTSGGLRAKDLMLDNDGQIKSVQAHKSALDRMKE---GKKHLTK
VFK----AGFALQ----KEGTAAYKKKMK---
>a|RefSeq_gi|313843970|ref|NC_014766.1|_Ostreococcus_lucimarinus_virus_OlV1,_complete_genome_148/2-78 [subseq from] # 114533 #
114796 # -1 # ID=267714_148;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.515
----------KPIGSRAEVFHGTAEKTSGGLRAKDLMLDKDGQIKSVAAHQAALDRMKE---GKKHLTK
VFK----PGFALQ----KEGTKDYKKKMK---
>a|EV_3300002040.a:GOScombined01_104207045_3/3-79 [subseq from] # 372 # 635 # -1 #
ID=43892_3;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=11bp;gc_cont=0.481
----------KLIGSRAEVFHGTAERTSGGLRSKDLMLGKDGQIKSIAAHDAALERMRE---GKKSMVK
VFK----PGFTLQ----KEGTAAYKKKIK---
>a|EV_3300003345.a:JGI26080J50196_1000041_46/1-77 [subseq from] # 33191 # 33448 # -1 #
ID=64483_46;partial=00;start_type=ATG;rbs_motif=AAA;rbs_spacer=11bp;gc_cont=0.504
----------MTVGTRAEVFHGTAEKTAGGLAKKDLMMGKDGRIKSKAAHAAALKRMKE---GKKAMVK
VFK----PTFKLQ----KAGTAEYEKKIA---
>a|EV_3300001354.a:JGI20155J14468_10001282_14/1-76 [subseq from] # 10861 # 11115 # 1 #
ID=31415_14;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=7bp;gc_cont=0.518
----------MTVGSRAEVFHGNADKTPGGLSKKDLIM-KDGRVVSKAASKAALARMKE---GKKAMVK
VFK----PSFKLQ----KSGTVEYEKKIA---
>a|GlobVir_NODE_395_length_22993_cov_3.83634_ID_789_27/1-76 [subseq from] # 15856 # 16110 # -1 #
ID=235257_27;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=7bp;gc_cont=0.518
----------MTVGSRAEVFHGNADKTAGGLTKKDLKM-KDGRIVSKAASKAALDRMSE---GKKAMVK
VFK----PTFKLQ----KAGTKEYEKKIA---
>a|EV_3300001460.a:JGI24003J15210_10000041_48/2-76 [subseq from] # 33672 # 33923 # -1 #
ID=35779_48;partial=00;start_type=ATG;rbs_motif=GGA/GAG/AGG;rbs_spacer=5-10bp;gc_cont=0.504
----------VVGSRAQVFHGTADRTTGGLEKKNLVL-KDGRIVSKTASQAALKRMSE---GKKHLTK
VFK----PKFKLQ----KEGTAAYKKLVK---
>a|EV_3300002040.a:GOScombined01_103061127_19/2-75 [subseq from] # 11022 # 11270 # -1 #
ID=43986_19;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.526
----------IGSRAEVFHGTADMTAGGLEKKDLKM-KDGRIVSKAASEAALKRMDE---GKKAMVK
VFK----PKFKLQ----KEGTVAYKKLIK---
>a|RefSeq_gi|313768203|ref|NC_014767.1|_Micromonas_sp._RCC1109_virus_MpV1,_complete_genome_139/2-76 [subseq from] # 108301 #
108555 # 1 # ID=267715_139;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.506
----------IGSRAEVFHGTADKTAGGLMKKDLMQDKDGRIKSKAAHDAAMKRMKE---GKKAMVK
VFK----PGFKLQ----KEGTKAYKKLIK---
>a|EV_3300003216.a:JGI26079J46598_1000066_45/2-75 [subseq from] # 31066 # 31323 # -1 #
ID=63737_45;partial=00;start_type=ATG;rbs_motif=AAA;rbs_spacer=13bp;gc_cont=0.512
----------IGSRAEVFHGNADRTAGGLTKKDLVRDKDGRIKSKAARDAAKKRMKE---GKKAMVK

VFK----PDFKLQ----REGTKGYDKLI----
>a|EV_3300002603.a:JGI24921J38492_100802_13/2-78 [subseq from] # 7853 # 8110 # 1 #
ID=55248_13;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.450
----------KTIGTRAEVFHGNADHTPGGLGKKDLIKGEDGRIKSRAASCAATARMRE---GKKALVK
VFK----PNFKLQ----KEGTKDYAKKIK---
>a|EV_3300001960.a:GOS2230_1019982_17/2-78 [subseq from] # 13452 # 13709 # 1 #
ID=43613_17;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.457
----------KTIGSRAEVFHGNADHTSGGLRKKDLIQGEDGRIKSKAASCSALIRMRE---GKKALVK
VFK----PSFKLQ----REGTKEYAKKIK---
>a|EV_3300002605.a:JGI24916J38487_100703_14/2-78 [subseq from] # 8565 # 8822 # 1 #
ID=55295_14;partial=00;start_type=ATG;rbs_motif=TATAA;rbs_spacer=3bp;gc_cont=0.492
----------KNIGSRAEVFHGTAAKTPGGLVKKDLIRGGDGRIKSKAASCAAMDRMRE---GNKAMTK
VFK----PNFKLQ----KEGTKEYEKKIK---
>a|EV_3300003592.a:JGI26246J51724_1002006_8/2-78 [subseq from] # 5546 # 5803 # -1 #
ID=69067_8;partial=00;start_type=ATG;rbs_motif=TATAA;rbs_spacer=3bp;gc_cont=0.527
----------KNIGSRAEVFHGTATKTAGGLTKKELIRGGDGRIKSKAASCAALDRMRE---GKKAMTK
VFK----PTFKLQ----REGTKEYEKKIK---
>a|EV_3300001347.a:JGI20156J14371_10000993_13/3-78 [subseq from] # 13719 # 13982 # 1 #
ID=29405_13;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.481
------N------KMIGSRAQVFHGTADQTAGGLKKKDLILDN-GEIKSKAAQQAALARMKE---GKKHLVK
VFK----PKFKLQ----KEGTKAYKTKV----
>a|EV_3300003409.a:JGI26088J50261_1000143_19/2-76 [subseq from] # 19436 # 19693 # 1 #
ID=66364_19;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.465
----------IGSRAQVFHGTADQTAGGLKKKDLILDKDGQIKSKAAQEAALARMKE---GKKHLTK
VFK----PKFKLQ----KEGTKAYDKKVK---
>a|EV_3300003617.a:JGI26082J51739_10000004_82/2-76 [subseq from] # 52716 # 52970 # -1 #
ID=69423_82;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=11bp;gc_cont=0.475
----------IGSRAQVFHGTADQTAGGLKKKDLILGNDGQIKSKAAQQAALARMKE---GKKHLTK
VFK----PKFKLQ----KEGTKAYDKKIA---
>a|EV_3300000137.a:LP_F_10_SI03_10DRAFT_c1000078_26/2-78 [subseq from] # 18933 # 19193 # 1 #
ID=18800_26;partial=00;start_type=ATG;rbs_motif=GGA/GAG/AGG;rbs_spacer=5-10bp;gc_cont=0.387
----------QTFGSRAEVFHGTALKTTGGLEKKDLLQDKYGSIVSKAARESALKRMRE---GKKALVK
VFK----PKFGLQ----KEGTKKYKTLIK---
>a|EV_3300001353.a:JGI20159J14440_10004618_4/2-78 [subseq from] # 1369 # 1629 # -1 #
ID=31211_4;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.379
----------QTFGSRAEVFHGTALKTTGGLAKSDLMQDKYGRIVSKAARKSAIERMRE---GKKALVK
VFK----PKFGLQ----KEGTKKYKTLIK---
>a|RefSeq_gi|313767999|ref|NC_014765.1|_Bathycoccus_sp._RCC1105_virus_BpV1,_complete_genome_131/2-78 [subseq from] # 103167 #
103427 # 1 # ID=267713_131;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.398
----------QTFGSRAEVFHGTAMKTTGGLTKSDLTQDKYGAIISKAARKAALARMAE---GKQHLVK
VFK----PKFGLQ----KEGTKKYKTLVK---
>a|EV_3300003345.a:JGI26080J50196_1000429_2/2-78 [subseq from] # 1538 # 1795 # 1 #
ID=64445_2;partial=00;start_type=ATG;rbs_motif=GGA/GAG/AGG;rbs_spacer=5-10bp;gc_cont=0.508
----------KTFGSRAEVFHGTAEKTAGGLKKKDLFQDKYGAIKSKAASKAALTRMEE---GKKAMVK
VFK----PSFKLQ----KEGTAAYKKLIK---
>a|EV_3300001353.a:JGI20159J14440_10001602_5/2-78 [subseq from] # 1930 # 2193 # -1 #
ID=31277_5;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.360
----------KTFGTRAEVFHGAASKTAGGLTKKQLFQDKSGRIKSKGASCAALLRMSE---GDKHLVK
VFK----PGFKKQ----KSGTKTYKKLVK---
>a|EV_3300002605.a:JGI24916J38487_100289_43/3-80 [subseq from] # 29056 # 29322 # 1 #
ID=55279_43;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.509
------T------EATGSRAEVFHGAAMHTPGGLVKGDLVQDKYGNIKSKAAVAAAKKRMKE---GASSMVK
VFK----PGFKLA----KKGTKKYKTLIK---
>a|EV_3300002605.a:JGI24916J38487_100790_36/5-80 [subseq from] # 23826 # 24095 # -1 #
ID=55298_36;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.485
----------TEGSRAEVFHGTAAHTSGGLAKKDLVQDKYGNIKSKAAVAAAKKRMDE---GKKAMVK
VFK----PGFKLA----KKGTKKYKTLIK---
>a|EV_3300001354.a:JGI20155J14468_10000578_44/5-80 [subseq from] # 27776 # 28042 # 1 #
ID=31362_44;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.513
----------TEGSRAEVFHGTAKHTPGGLVKSDLVQDKYGSIKSKAAVAAAKKRMEE---GAKAMVK
VFK----AGFKLA----KKGTKKYKTLIK---
>a|EV_3300001346.a:JGI20151J14362_10002042_22/5-80 [subseq from] # 12260 # 12529 # -1 #
ID=29041_22;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=4bp;gc_cont=0.533
----------AEGSRAEVYHGTAKHTPGGLVKKDLTQDKYGNIKSKAAVAAAKKRMKE---GTSAMVK
VFK----PGFKLA----KKGTKKYKTLVK---
>a|EV_3300003621.a:JGI26083J51738_10000002_59/5-81 [subseq from] # 38763 # 39035 # -1 #
ID=69956_59;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.513
----------TEGTRAEVFHGTAAHTAGGLAKKDLVQDKYGNIKSKAAVAAAKKRVAE---GANAMVK
VFK----PGFKLA----KKGTAGYKKLIKK--
>a|EV_3300003592.a:JGI26246J51724_1000880_8/6-82 [subseq from] # 4479 # 4754 # 1 #
ID=69074_8;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.533
----------TEGTRAEVFHGTAAHTAGGLAKKDLVQDKYGNIKSKAAIAAAKKRMEE---GAKAMVK
VFK----PGFKLA----KKGTASYKKLVKK--

>a|EV_3300003620.a:JGI26273J51734_10000696_19/6-82 [subseq from] # 9293 # 9565 # -1 #
ID=69792_19;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=8bp;gc_cont=0.451
----------RIGSRAQVFHGTAKSTAGGLKAGDLIRDDDGQIKSKKAQMSALARMKE---GKSAMVK
VFK----PSFKLQ----KEGTAAYKKLIKK--
>a|EV_3300003427.a:JGI26084J50262_1000245_31/3-79 [subseq from] # 21244 # 21507 # -1 #
ID=67287_31;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.504
----------KEGTRAQVFHGTAERTAGGLTKSDLTQGKDGRIKSKAAVEAAIRRMEE---GNKSMVK
VFK----AGFEKQ----RAGTAEHKKLVKK--
>a|EV_3300003345.a:JGI26080J50196_1000058_49/1-76 [subseq from] # 34276 # 34542 # -1 #
ID=64484_49;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.442
----------MTVGSRAEVFHGNADKTSGGLTKKDLTM-KDGRIISKAASKAAKKSL-N---PFQAFID
IAKE---EGFCLV----SKNTKTYKKIIK---
>a|EV_3300000101.a:DelMOSum2010_c10005205_4/1-76 [subseq from] # 3183 # 3449 # -1 #
ID=11305_4;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.472
----------MTVGSRAEVFHGNANATSGGLTKKDLMM-KDGRIISKAASKAAKKSL-N---PFMAFID
LAKE---EDFCLV----KKGSKTYKKIIK---
>a|EV_3300003617.a:JGI26082J51739_10001635_15/2-75 [subseq from] # 13473 # 13736 # 1 #
ID=69248_15;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.504
----------IGSRAEVFHGTADKTSGGLEKKDLMM-KDGRIVSKAASKAAKKSL-N---PFKAFID
LAKE---AGFCLV----SKDTKTYKKIIK---
>a|EV_3300002605.a:JGI24916J38487_100588_24/2-75 [subseq from] # 22527 # 22790 # 1 #
ID=55291_24;partial=00;start_type=ATG;rbs_motif=AAA;rbs_spacer=13bp;gc_cont=0.511
----------IGSRAEVFHGNADKTPGGLAKKDLMM-KDGRIVSKAASKAAKKSL-N---PFKAFID
LAKE---AGFCLV----SKDTKTYKKIIK---
>a|GlobVir_NODE_2376_length_6113_cov_2.87224_ID_4751_4/2-76 [subseq from] # 1852 # 2118 # -1 #
ID=199095_4;partial=00;start_type=ATG;rbs_motif=AAA;rbs_spacer=13bp;gc_cont=0.479
----------ITGSRAEVFHGTADKTSGGLEKKDLVM-KDGRIISKAASKAAKKSL-N---PFKAFVD
LAKE---EAFCLV----KKGSKTYKKIIK---
>a|EV_3300001353.a:JGI20159J14440_10003197_3/2-74 [subseq from] # 1815 # 2078 # 1 #
ID=31355_3;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.511
----------IVGSRAEVFHGNADETSGGLEKKDLKM-KDGRIISKAASKAAKKSL-N---PFKAFIE
AAKK---AGFGLV----KKGTKAYKKI-----
>a|EV_3300001354.a:JGI20155J14468_10002051_17/2-76 [subseq from] # 9047 # 9328 # -1 #
ID=31400_17;partial=00;start_type=ATG;rbs_motif=TATAA;rbs_spacer=3bp;gc_cont=0.468
----------IGSRAEVFHGTADSTSGGLMKKDLTM-KDGRIISKAASKAAKKSL-N---PFQMFID
EAKK---DGFCAM----SNDTKAYEKLMKK--
>a|EV_3300001720.a:JGI24513J20088_1000048_20/1-74 [subseq from] # 15652 # 15906 # 1 #
ID=42531_20;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.482
----------MTVGSRAEVFHGNANQTSGGLTKKDLKM-KDGRIISKAASKASKKAYKT---AFKAFVD
AAK----AGFHRV----KEGTKAYDKL-----
>a|EV_3300002040.a:GOScombined01_100117075_16/2-73 [subseq from] # 14250 # 14504 # 1 #
ID=43900_16;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.537
----------IGSRAQVWHGNADKTSGGLKKKDLKM-KDGRIISKAVSKKSKAMKDH---PFQAFIK
KAK----AGFHRV----SKGTKEYKKM-----
>a|EV_3300001346.a:JGI20151J14362_10000631_15/4-65 [subseq from] # 12521 # 12949 # -1 #
ID=28783_15;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.312
----------TNAQRAKVFHGTLDKTSGGLTKNDLKKNEYGKIVSKKASSAARKNN----------
-------LGFKQT----KKGTKDYKES-----
>a|EV_3300003345.a:JGI26080J50196_1003604_6/4-79 [subseq from] # 2756 # 3043 # 1 #
ID=64517_6;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.524
----------KRVGSRAEVLHGTAHHTGGGLKKGDLFLDKDGRIKSKDATKAANSRLRE---GTAHFTK
VFK----PGFSLQ----KQGTKDYKKER----
>a|EV_3300002605.a:JGI24916J38487_100196_43/5-67 [subseq from] # 32127 # 32372 # 1 #
ID=55272_43;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.533
----------KTIGSRAEVFHGTAEKTTDGLKKKDLFRGKDGRIKSRKASKATK----------K
SMD----DHFRAY----RTEAKKGKK------
>a|EV_3300001348.a:JGI20154J14316_10000329_5/6-82 [subseq from] # 2669 # 2956 # 1 #
ID=29596_5;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.278
------T------KRYGTRAEVWHGKARMTQGRLTKEDFIMNEYGYIVSKRKSKKSK-KNPR---QKGLLQK
KKSK---NGFGP--------QKTKNKTNKKRI
>a|EV_3300001351.a:JGI20153J14318_10002797_9/8-76 [subseq from] # 5069 # 5320 # -1 #
ID=30679_9;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=8bp;gc_cont=0.290
----------RYGTRAEVMSGTARMTQGRLTKSDFMYNTKGYIVSKKKSIKMKGENLK---KQKYLQ-
-------TGFGPA-------KKTIKKKNK---
>a|EV_3300001354.a:JGI20155J14468_10002073_7/21-86 [subseq from] # 4559 # 4867 # -1 #
ID=31530_7;partial=00;start_type=ATG;rbs_motif=TAAA;rbs_spacer=8bp;gc_cont=0.430
------T------QYIASRASVYHGTAFKTSGGLQKKDLMMNKNGRIVSLKKHKTAKNEK---------LEK
A------TGFGYV----KKD------TKKRK-
>a|EV_3300002447.a:JGI24768J34885_10000109_15/28-89 [subseq from] # 11721 # 12032 # -1 #
ID=46528_15;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.590
----------RKVFKGTLLKTIGGVTRSGIMKRRDGKLISIKKHKAGKKSH---------EMR
ARDP---ADFKIA----KKP------TTRRS-

>a|EV_3300002601.a:JGI24917J38488_100992_9/24-85 [subseq from] # 10223 # 10606 # 1 #
ID=55219_9;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=5bp;gc_cont=0.365
----------CGTRLQVWRGMANETSGGLKKKDLVKNKWGRIVSRKKYHSASKDN---------LEK
H-----VNNF----------VKAPRKTRK---
>a|EV_3300001348.a:JGI20154J14316_10003916_3/9-73 [subseq from] # 1004 # 1285 # 1 #
ID=29524_3;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.337
----------KREVWEGRRSKTASGLKKKDLMKNKDGKVVSKKQHKRGQEL----------MMK
E-----LADFK------KRSSKTPKRSNKSRS
>a|EV_3300001348.a:JGI20154J14316_10001692_3/9-75 [subseq from] # 1171 # 1452 # 1 #
ID=29559_3;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.319
----------KREVWEGKRSKTASGLKKKDLMKNKDGKVVSKKQHKRGQELR--------MMKK
D-----LADFK------KRSSKTPKRSKKSRS
>a|EV_3300001348.a:JGI20154J14316_10000588_35/9-72 [subseq from] # 28966 # 29226 # -1 #
ID=29856_35;partial=00;start_type=ATG;rbs_motif=AGGA;rbs_spacer=5-10bp;gc_cont=0.272
----------KELVWNGEKKKTSSGLTKKDLMVNKNGKVVSKKQHKKGQELY--------MMKK
E-----LAEFQ------SKSSKS---KSKKRS
>a|EV_3300003427.a:JGI26084J50262_1001307_13/4-74 [subseq from] # 8533 # 8781 # 1 #
ID=67371_13;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=10bp;gc_cont=0.470
----------SRAEVFHDGTLKTAGGLIHXDLMMK-DGHIVSKARSKASKKALSNFKVFVQ---A
AKD-----GFELSKKRTKKYKEMTK-------
>a|EV_3300002835.a:B570J40625_100000501_34/4-81 [subseq from] # 30680 # 30973 # 1 #
ID=60110_34;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.517
------QT-----KTIGSRAAVLRGKADMTAQGHTADQLTKNSQGRIVGRSASSASKNN--SLAKWRECVTA
TAGDTGAKG-----------SEHYERAMKM--
>a|EV_3300002835.a:B570J40625_100000733_43/2-74 [subseq from] # 34808 # 35074 # -1 #
ID=61594_43;partial=00;start_type=ATG;rbs_motif=ATA;rbs_spacer=6bp;gc_cont=0.461
----------ASRAVVLHTKKEKTAGGLTSNKLTVSSSGEIVSKAKMKQGKESTAD-------TRL
CRRDLK---MVLFN--VKDGKALYECVKE---
>a|EV_3300002835.a:B570J40625_100002932_24/2-73 [subseq from] # 15340 # 15606 # -1 #
ID=60701_24;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.517
----------ASRALVLHKKESTTAGGLTSKKLTVSSSGEIVSKAKAKQGKESTAD-------TRL
CRKDLK---MVLFN--VKDGKALYACVK----
>a|EV_3300002605.a:JGI24916J38487_100974_27/40-102 [subseq from] # 30537 # 30863 # 1 #
ID=55312_27;partial=00;start_type=ATG;rbs_motif=TAAA;rbs_spacer=9bp;gc_cont=0.391
----------ELFGSRQKVWNGTAYKTTGNLVKTDLVQNSRRRIVSKKKYYQGKKE----------RT
KSRLFKTAGFGPVKR----------
>a|EV_3300001352.a:JGI20157J14317_10001101_15/51-113 [subseq from] # 16945 # 17307 # 1 #
ID=30912_15;partial=00;start_type=ATG;rbs_motif=TAA;rbs_spacer=10bp;gc_cont=0.364
----------QLFGSRQQVWNKTAYKTTGNLKRDDLTQNSRRRIVSKKKYIQGKKE----------VS
KKRLFKTAGFGAVKR----------
>a|EV_3300001352.a:JGI20157J14317_10000177_45/7-77 [subseq from] # 43472 # 43975 # -1 #
ID=30909_45;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.258
----------TKKIGNRSEVWNNLARQTPGGLQKKDLFISKTGKITSKKASLSSKKR----------IQ
GKCKMCKYG-KAILTEVKTN------VKE---
>a|EV_3300003427.a:JGI26084J50262_1000202_33/8-65 [subseq from] # 24135 # 24371 # -1 #
ID=67381_33;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.397
----------ERLRVLRGEIPKTSGGLTAKDLFPDG----TSKKASMAAKKRPP--------ILF
ATGEFQRVGYKKIMK----------
>a|EV_3300001352.a:JGI20157J14317_10006641_5/5-75 [subseq from] # 2917 # 3522 # 1 #
ID=30934_5;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.203
----------KIIGTREEVYKSQAQKTIGGLTKIDIIENNKTTYISKK--LSDKMRLQN------TIRE
KNPNFFRTHVSSIHKNSNSN----------
>a|EV_3300001349.a:JGI20160J14292_10002329_9/12-65 [subseq from] # 7150 # 7425 # -1 #
ID=30102_9;partial=00;start_type=ATG;rbs_motif=None;rbs_spacer=None;gc_cont=0.446
----------GLTQEHFVK-KEERIVSCKKCKGAANRILKNPDWTGLA--
------MERFARVR------KQQKDKEKS---

Appendix 4: Relevant DVNP sequences used in this study

>*Hematodinium* DVNP.6 full length protein sequence
MAKAMKKAMKAKKSSAKKGKKSSKRHSKVAKGKRAKSSVFRGTKEKTSGGLKKSDLTRNKSGK
IVSKKASEAAKKQFKKSGLSKFTEAVKKARKVLGIKGFQAVGGKSAKGQALLKKAREFYKK

>*Ostreococcus lucimarinus* virus DVNP full length protein sequence
MSTKPIGSRAEVFHGTAEKTSGGLRAKDLMLDPNDGQIKSVQAHKSALERMKKEGKKHLTKVFKA
KKGKFALQPKEGTAAYKKKMKKMA

>pTXB1-DVNP.6 translation
MAKAMKKAMKAKKSSAKKGKKSSKRHSKVAKGKRAKSSVFRGTKEKTSGGLKKSDLTRNKSGK
IVSKKASEAAKKQFKKSGLSKFTEAVKKARKVLGIKGFQAVGGKSAKGQALLKKAREFYKK^CITG
DALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVT
GTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPGL
VRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSHATGL
TGLNSGLTTNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ*
^: cleavage site

>pTXB1-vDVNP translation
MSTKPIGSRAEVFHGTAEKTSGGLRAKDLMLDPNDGQIKSVQAHKSALERMKKEGKKHLTKVFKA
KKGKFALQPKEGTAAYKKKMKKMA^CITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDR
HGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQ
RSAFSVDCAGFARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTD
AGVQPVYSLRVDTADHAFITNGFVSHATGLTGLNSGLTTNPGVSAWQVNTAYTAGQLVTYNGKT
YKCLQPHTSLAGWEPSNVPALWQLQ*
^: cleavage site

>pTXB1-NΔ-DVNP.6 translation
MSSVFRGTKEKTSGGLKKSDLTRNKSGKIVSKKASEAAKKQFKKSGLSKFTEAVKKARKVLGIKGF
QAVGGKSAKGQALLKKAREFYKK^CITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHG
NPVLADRLFHSGEHPVYTVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRS
AFSVDCAGFARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAG
VQPVYSLRVDTADHAFITNGFVSHATGLTGLNSGLTTNPGVSAWQVNTAYTAGQLVTYNGKTYK
CLQPHTSLAGWEPSNVPALWQLQ*
^: cleavage site

>Hematodinium DVNP.6 optimised ORF
ATGGCAAAGGCCATGAAGAAAGCCATGAAAGCAAAAAAAAGCAGCGCCAAAAAAGGCAAAA
AATCCAGCAAACGTCATAGCAAAGTTGCCAAAGGTAAACGTGCAAAAAGCAGTGTTTTTCGTG
GCACCAAAGAAAAAACCAGCGGTGGTCTGAAAAAAAGCGATCTGACCCGTAATAAAAGCGGC
AAAATTGTTAGCAAAAAAGCAAGCGAAGCCGCAAAGAAACAGTTCAAAAAAAGTGGCCTGAG
CAAATTTACCGAGGCAGTTAAAAAAGCACGTAAAGTGCTGGGTATTAAAGGTTTTCAGGCAGT
TGGTGGTAAAAGCGCCAAAGGTCAGGCACTGCTGAAAAAAGCCCGTGAATTTTACAAG

>Viral DVNP optimised ORF
ATGAGCACTAAGCCTATTGGTTCCCGCGCGGAAGTTTTTCATGGCACCGCGGAGAAGACTAGC
GGCGGTCTGCGCGCGAAAGACTTAATGCTCGATCCTAATGACGGACAAATCAAGTCTGTTCAG
GCCCACAAGAGTGCCCTCGAGCGCATGAAGAAAGAGGGGAAGAAACATCTTACCAAGGTGTT
CAAAGCGAAGAAGGGTAAATTCGCGTTGCAGCCGAAAGAAGGCACTGCGGCCTACAAGAAGA
AGATGAAGAAGATGGCC

>pET100/D_Topo_DVNP
CAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACA
AGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGC

GCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGA
GATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCC
CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCGGGGTTCTCATCATCA
TCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGACGA
TGACGATAAGGATCATCCCTTCACCATGGCAAAGGCAATGAAAAAGGCGATGAAAGCCAAGA
AGTCCAGCGCCAAAAAGGGCAAGAAGTCCTCCAAGCGTCACTCCAAAGTCGCGAAGGGTAAG
CGCGCGAAGTCTTCGGTTTTCCGAGGCACGAAGGAAAAGACCTCAGGCGGCTTGAAAAAAAGT
GACCTGACGCGCAACAAAAGTGGCAAGATCGTTTCTAAAAAAGCTTCGGAAGCCGCGAAGAA
GCAGTTCAAAAAGAGCGGTCTTTCAAAATTCACGGAGGCCGTCAAGAAAGCCCGGAAAGTTCT
TGGAATCAAGGGATTCCAGGCCGTCGGTGGAAAGTCGGCAAAAGGGCAAGCGTTGCTGAAAA
AAGCCCGCGAATTCTATAAGAAGTAGAAGGGCGAGCTCAACGATCCGGCTGCTAACAAAGCCC
GAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCT
CTAAACGGGTCTTGAGGAGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCCGCAAGAG
GCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGAT
GACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGAT
AAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAATTAATTCTTGAAGAC
GAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGAC
GTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACAT
TCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
AAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCC
TGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACG
AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGA
ACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGAC
GCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCA
CCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTA
ACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGA
ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGC
GCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGG
AGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGA
TAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAA
GCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAG
ACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCA
TATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTT
TGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTA
GAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAA
AAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAA
GGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTG
GCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGAT
AAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGAC
CTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC
CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG
ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTT
ACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGT
GGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCG
CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCT
GTGCGGTATTTCACACCGCAATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAA
GCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAAC
ACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACC
GTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTG
CGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCA
GCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGC

GGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATG
ATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTA
CTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCAC
TCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCA
TCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTAC
GAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAG
TCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGC
CTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGC
CCGAGATGCGCCGCGTGCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGT
TGGTTTGCGCATTCACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCC
GTTAGCGAGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCGCGACGC
AACGCGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCGTTCCATGTG
CTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGCGGTCCAATGATCGAAGTTAGGCTGGTA
AGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTGGACAGCATG
GCCTGCAACGCGGGCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAATGGGGAAGGCCAT
CCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGAT
AATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGC
GTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTC
CTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGAC
AGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAA
GGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGC
GTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGC
CAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGA
GACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCAC
GCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGA
GCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCG
GTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACG
ATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCC
GTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGAC
GCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCA
GATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTG
GTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATC
CTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCAC
CGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGT
TGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAG
GTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATG
TAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTG
GTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGT
TACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGA
AAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATT
AGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATG

Appendix 5: Double-stranded DNA oligomers used in isothermal calorimetry

20-mer
5’-ACGCCTGAAGAGTCTGGTGA-3’
3’-TGCGGACTTCTCAGACCACT-5’

36-mer
5'-ATCAAGCTACGCCTGAAGAGTCTGGTGAGCAAGGGT-3'
3'-TAGTTCGATGCGGACTTCTCAGACCACTCGTTCCCA-5'