Přírodovědecká fakulta
Faculty of Science

Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

Production and Biophysical Characterization of Phi8P4 Helicase

Bachelor thesis

Siroun Moura

Supervisor: RNDr. Zdeněk Franta, Ph.D.

Co-supervisor: Mgr. Tomáš Fessl, Ph.D.

České Budějovice 2020

Moura S., 2020: Production and Biophysical Characterization of Phi8P4 Helicase, Bc. thesis, in English; 41 pages, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

The goal of this thesis was cloning of Phi8P4 protein into expression vector with N-terminal his-tag, optimization of protein expression and large-scale production for further functional and crystal structure studies with RNA.

Declaration:

I hereby declare that I have worked on my bachelor's thesis independently and used only the sources listed in the bibliography. I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full form resulting from deletion of indicated parts to be kept in the Faculty of Science archive, in electronic form in a publicly accessible part of the IS STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages. Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defence in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

České Budějovice, 21 May 2020

……

Siroun Moura

Acknowledgement

I would like to thank the Institute of Chemistry members for their ongoing support and guidance. I would like to express my gratitude specially to:

- RNDr. Zdeněk Franta, Ph.D., Mgr. Tomáš Fessl, Ph.D., providing me the opportunity to work on this project.
- My family and friends for supporting me through the whole study process.

Abstract

Genome packaging of double-stranded RNA viruses into a procapsid is important for the assembly of many viruses. In case of double-stranded RNA (dsRNA) bacteriophages belonging to *Cystoviridae* family, the packaging is done by hexameric P4 helicase protein which uses the energy from ATP hydrolysis. P4 consists of 3 main domains: N-terminal, central core NTPase and C-terminal. The C-terminal domain is essential for RNA binding and loading. This thesis was focused on cloning Phi8P4 into expression vector pASK-IBA37plus (N-terminal Histidine-tag) and producing the protein under optimal conditions. Furthermore, testing its helical activity on RNA hairpin.

Table of contents

List of abbreviations

A – adenosine

AAA^+ - ATPases associated with diverse cellular activities

AMPcPP - α, β-methyleneadenosine-5'-triphosphate

APS – ammonium persulfate

ATP – adenosine triphosphate

ATPase - adenosine triphosphatase

ACF – autocorrelation function

bp – base pair

C – cytosine

dH_2O – double-distilled water

ddH_2O – double-distilled water

DNA – deoxynucleic acid

dsDNA – double-stranded DNA

dsRNA – double-stranded RNA

EDTA – ethylenediaminetetraacetic acid

E – FRET efficiency

F – forward

FT – flow-through

FRET – förster resonance energy transfer

FCS – fluorescence correlation spectroscopy

G – guanine

His – histidine

HPLC – high performance liquid chromatography

ICTV – international committee of the taxonomy of viruses

IMAC – immobilized metal affinity chromatography

IPTG – isopropyl β-D-1-thiogalactopyranoside

kb - kilobase

kDa – kilodalton

LB – lysogeny broth

L - large

MM – master mix

M - medium

Mg – magnesium

mRNA – messenger RNA

NC – negative control

NA – nucleic acid

NTP – nucleoside triphosphate

NTPase – nucleoside-triphosphatase

OD_{600} – optical density measured at 600 mm

PAX – periodic acceptor excitation

PBS – phosphate-buffered saline

PCR – polymerase chain reaction

PAGE – polyacrylamide gel electrophoresis

R – reverse

RecA – recombination protein A

RNA – ribonucleic acid

S.O.C. – super optimal broth with catabolite repression

SMFD – single-molecule fluorescence detection

smFRET – single-molecule förster resonance energy transfer

S – small

SF – super family

1 Introduction

Viruses are small infectious organisms which replicate using the apparatus and metabolism of their hosts [1]. The viruses infect broad ranges of hosts including animals, plants, archaea, bacteria, etc. A Virus particle consists of protein envelope and viral genome, which is made of nucleic acid and may contain more than one segments, single or double stranded DNA (ssDNA or dsDNA) or RNA (ssRNA or dsRNA) [1]. An example of different types of viruses are mentioned below (see Figure 1).

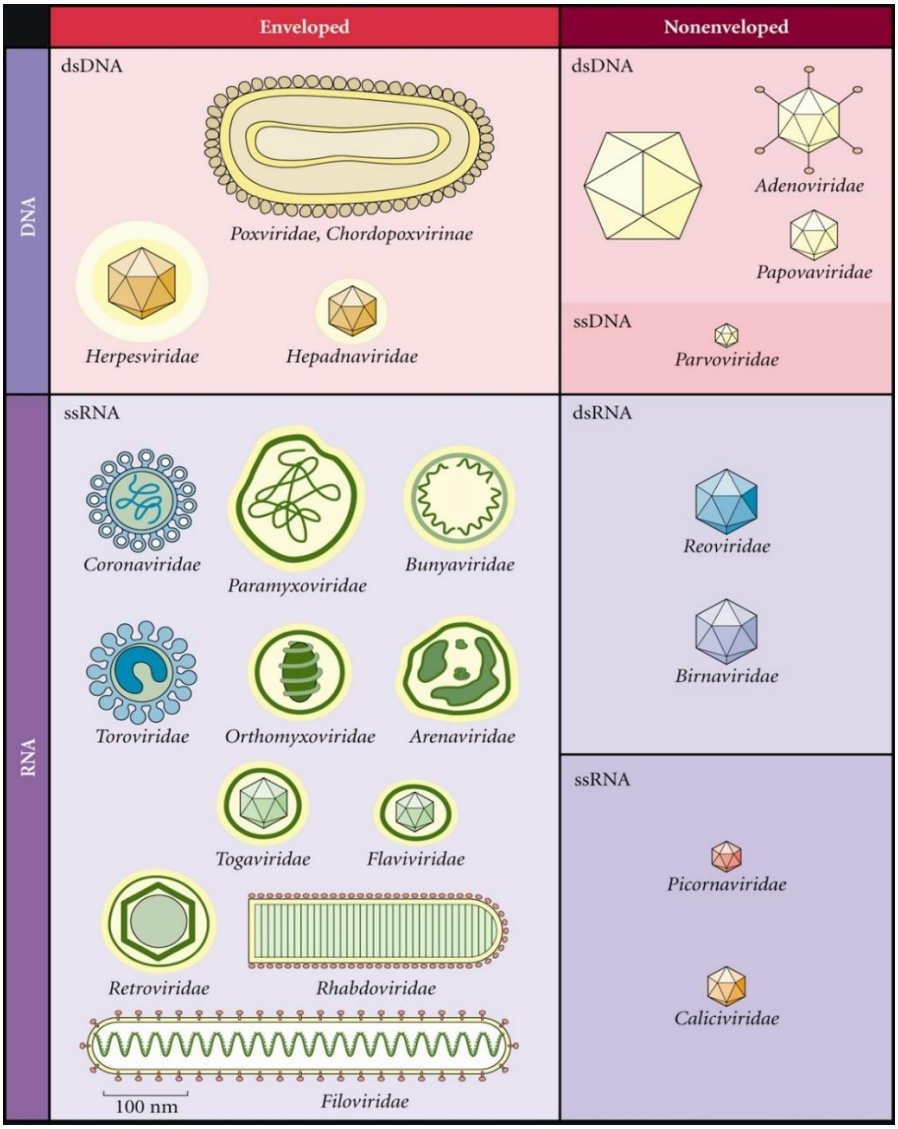


Figure 1: Virus classification. Available from <https://ppdictionary.com/viruses.htm>

RNA viruses hold RNA as their genetic material. The (-)-RNA strand is the strand responsible for replication and the (+)-RNA strand is responsible for synthesis [1]. Double-stranded RNA (dsRNA) is not affective as messenger RNA (mRNA) templates for translation and does not have the ability to function as template for host cell transcriptases, regardless of their stability.

Therefore, dsRNA viruses must have their own transcription and sometimes capping enzymes and deliver them to their host cells, allowing the synthesis of functional mRNAs to start the viral protein synthesis, ensuring the successful replication of these viruses [2].

1.1 Double stranded RNA (dsRNA) Viruses

Double stranded RNA viruses are larger groups of different pathogens, affecting a variety of plants, fungi and prokaryotes. Most of these viruses exhibit an icosahedral capsid structure and have similar structural and biochemical properties as an evidence of a common ancestry. Similarities are mainly in the innermost capsid layers and internal virion-attached enzymes (conserved viral proteins). Nowadays, eight different families of dsRNA viruses (see Table I) are described by the International Committee of the Taxonomy of Viruses (ICTV), having medical, veterinary or agricultural importance [2].

Table I: The family of dsRNA viruses [2].

Family	Number of genome segments	Type of virus particle	Host
Hypoviridae	1 (unpackaged)	~50-80 nm diameter, pleomorphic vesicles (no capsid)	Fungi
Totiviridae	1 (packaged singly)	~30-40 nm diameter, icosahedral	Fungi
Birnavirida	2 (co-packaged)	~60 nm diameter icosahedral, single shell	Fish, insects, birds, molluscs
Varicosavirus	2 (separately packaged)	~18 x 320-360 nm, rod shaped	Plants
Partitiviridae	2 (separately packaged)	30-40 nm diameter, icosahedral protein capsid	Fungi plants
Cystoviridae	3 (co-packaged, equimolar)	~85 nm diameter, three-layer structure, with an envelope surrounding a two layered icosahedral nucleocapsid	Bacteria (Pseudomonas)
Chrysoviridae	4 (packaged separately)	~30-40 nm diameter icosahedral protein	Fungi
Reoviridae	10,11 or 12 (co-packaged equimolar)	~70-90 nm diameter icosahedral (one, two or three layered protein capsid)	Insects, plants, fish, reptiles, birds, mammals, arachnids, fungi, arthropods crustaceans

1.2 The Family of *Cystoviridae*

Bacteriophages (also known as phages) are viruses which infect bacteria and their genomes may consist of RNA or DNA. Bacteriophages belonging to *Cystoviridae* family are characterized by having tri-segmented dsRNA with different sizes of the genome segment; L (large, 6.4 kbp), M (medium, 4.1 kbp) and S (small, 2.9 kbp). Each segment size refers to different protein types. One copy of individual genome segment is found in a virion and coding regions are surrounded by terminal non-coding regions which hold signals for genome transcription, packaging and replication [3]. Bacteriophage Phi6 (see Figure 2) is the main representative of *Cystoviridae* family whereas, bacteriophage Phi8 is the least studied case.

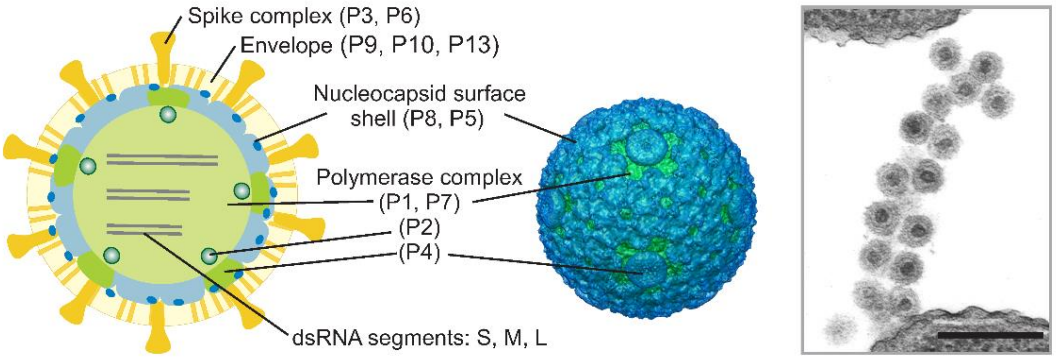


Figure 2: Cystoviridae (*Pseudomonas* phage phi6) [3]. Structure of Phi6 with specified location of the virion protein (left). 3D structure of the nucleocapsid (middle). Thin-section electron micrograph of Phi6 particles attached with pilus receptor of the host (right), the bar refers to 200 nm [3].

Bacteriophages (Phi6, Phi8, Phi12 and Phi13) infect plant pathogenic bacteria [4]. The virion is composed of three different layers. The outer lipid layer allowing the integration of the membrane proteins, is exposed to the surface binding receptor and causes fusion with outer membrane of the host. The nucleocapsid is found under the membrane and its structure consists of procapsid and a shell of protein P8. Phi6 consists of proteins, P1 (major structural protein,120 copies), P2 (RNA polymerase,~10 copies), P4 (hexamer packaging NTPase,~11-12 copies) and P7 (assembly cofactor, 12 copies) essential for forming an icosahedral procapsid (~50 nm in diameter) (see Figure 2) [3]. P2 functions as a replicase and transcriptase by using ssRNA or double-stranded dsRNA. During RNA packaging, the capsid is stabilized by P7, also this protein is important for the virion activity. Finally, RNA is pumped through the hexamer P4. Furthermore, the produced capsid can package (+)-strand RNA genomic precursors, synthesize the complementary (-)-strand RNA following the transcription of other (+)-strands. Once packaging and replication are complete, the procapsid filled with RNA is covered by ~600 P8 dimers (missing in Phi8), introducing a nucleocapsid (~58 nm in diameter). On the other hand, the lipid layer of the virus is obtained during maturation. Each

(+)-strand in the capsid is filled in 5’ to 3’ direction. The order of packaging is based on the *pac* sequence found on 5’ ends and protein P1 controls the selection of RNA sequence. Also, increase of internal pressure causes conformational changes of the viral capsid, affecting the RNA binding sites (activating or inactivating) and controlling the selection of the *pac* sequence [1]. P4 protein also acts as a passive channel for the exit of ssRNA transcripts [5]. For a summary of the *Cystoviridae* life cycle (see Figure 3) [1].

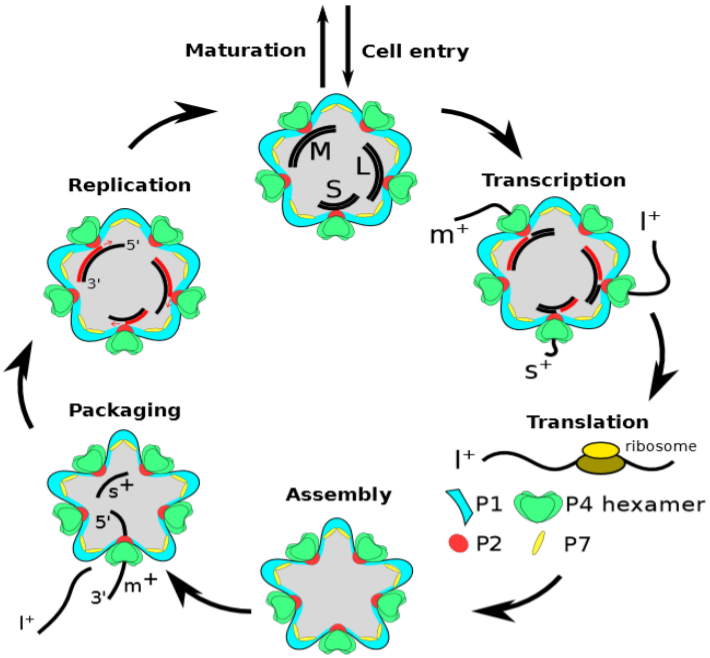


Figure 3: Scheme of the *Cystoviridae* core replication [1]. After cell entry, protein P2 transcribes the (+)-RNA. During transcription, newly formed (+)-strands (L^+,M^+ and S^+) are transported outside the procapsid by the passive portal P4. Translation of l^+ RNA by ribosome and producing protein P1, P2, P4 and P7 which co-assemble in forming empty procapsids. These proteins are presented in cyan, red, green and yellow. The internal pressure causes procapsid expansion. Afterwards, P2 replicates the (+)-RNA in the procapsid to yield double-stranded RNA (dsRNA). Modified from [6]. The red strands indicate RNA being synthesized by P2 [1].

The pure P4 hexamers of bacteriophage Phi8 show a helical activity in 5’ to 3’ direction along the ssRNA and structure like those of hexameric helicases [5].

1.2.1 *Cystoviridae:* bacteriophage Phi8 and its properties

The main work of this thesis focus on P4 protein from bacteriophage Phi8, which is a member of *Cystoviridae* family, with tri-segmented dsRNA genomes inside a polyhedral inner core surrounded by a lipid membrane [7] and composed of different types of proteins.

1.2.1.1 Structure of bacteriophage Phi8

The major structural difference between Phi8 and other bacteriophages is that protein P8 found in Phi8 does not form a shell around the inner core, making it a minor component of the membrane [8]. For instance, Phi6 enters the periplasmic space as the nucleocapsid is covered by P8 whereas, Phi8 enters as an inner core and the P8 is fused to the outer membrane of the host cell and is left behind. Also, in Phi8 the P1 protein (major structural protein of inner core) determines the packaging specificity. Additionally, P1 plays the same role as P8 in other bacteriophages, focused on the membrane interactions through which the viral membrane is obtained that is essential for forming the mature virion and also allowing the passage of core particle into the host cell through the cytoplasmic membrane [9]. Virions of Phi8 are in direct contact with the outer membrane of their host cells with no affinity to pilus [10]. When it comes to the infection the spheroplasts (a bacterium cell, lacking cell wall resulting in a spherical shape) of the host cell, the inner core of Phi8 is capable of accepting the lipid envelope without the need of forming a P8 shell first [11]. Protein P4 (NTPase) is the motor used for genome packaging. Hexamers are formed by P4 in absence of ATP and do not function as an actively until associated with empty inner procapsids. In Phi8 the gene for the ortholog of P7 protein is found at the 3' end of genome segment L. Additionally, in Phi8 the gene H is found at the usual place of gene 7 (see Figure 4), important for phage development, yet not necessary for in vitro genomic packaging and replication [9].

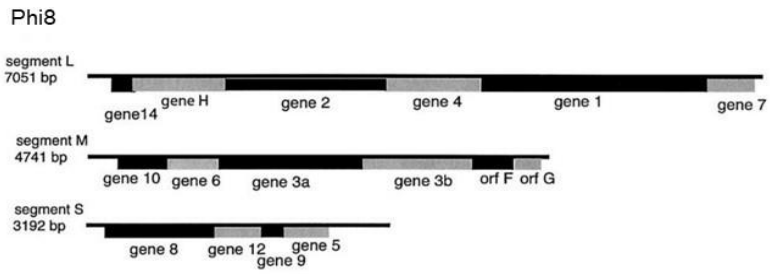


Figure 4: Genetic maps of genomic segments of bacteriophage Phi8. Modified from [9]. (+)-strand transcripts have pac sequences close to 5' ends upstream of the proximal genes. Note, the position of gene 7 in phi8 is different from those of other members of *Cystoviridae* [9].

1.3 Nucleic acid-based motors

1.3.1 Helicases

Helicases are molecular motors, which use the ATP hydrolysis to move along the nucleic acid (NA), separate the complementary strand of NA and unwind it. The NA base-pairs open and close under different physiological conditions (i.e. thermal fluctuations). Once helicase is added, it moves along the single-stranded (ssRNA) and binds to it excluding the complementary strand. Furthermore, helicases also unwind NA secondary structure (i.e. RNA hairpin). Both NA and NTP binding sites are specific for each helicase subunit and NA binding is affected by NTPase. Based on the oligomeric state, helicases are divided into two different groups: ring (hexamer) and non-ring (monomer and dimer). In case of non-ring helicases, two NA binding sites attach to each and release the NA as the distance between them changes [12]. Also, helicases can be divided depending on their unwinding direction (5' to 3' or 3' to 5') with respect to the strand they bind with and move along. According to their sequence, helicases are classified into six superfamilies (SF1-SF6) [13]. Members of one superfamily share poor sequence similarity and are narrowed to short sequence motifs of the AAA^+ or RecA-like structural domains [14]. The RecA-like core converts the energy produced by ATP binding and hydrolysis into a mechanical force important for DNA and RNA translocation. Hexameric helicases are case of interest, as we tested the activity of the packaging motor P4 which exhibits similar sequence and structure.

1.4 P4 protein

P4 is a protein of size ~35 kDa with a hexameric ring structure. The NTP binds through the external perimeter of the ring that is found at the interfaces between adjacent subunits, whereas the nucleic acid attaches through the central channel (see Figure 5) [15].

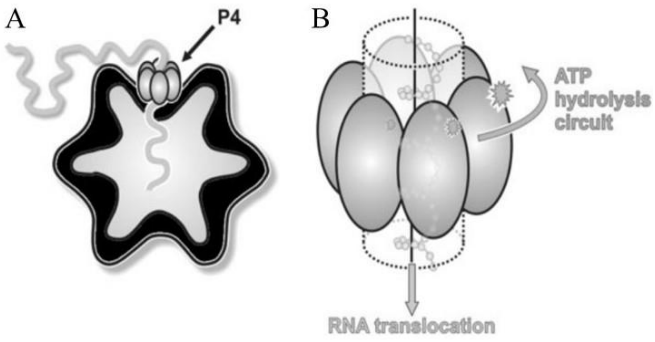


Figure 5: Cystovirus P4 protein, molecular packaging motor [15]. (A) a cartoon model showing the position of protein P4 hexamer (in grey color) on an empty cystovirus procapsid (in black color) during ssRNA packaging. (B) A cartoon model showing the mechanism of RNA translocation using P4. The energy from ATP hydrolysis is converted in a mechanical energy used for translocating the ssRNA [15].

1.4.1 P4 protein: Structure

P4 protein belongs to the superfamily SF4, characterized by five conserved motifs (H1, H1a, H2, H3 and H4) [16] and an arginine finger (see Figure 6) [15]. The motifs H1a and H2 together form a binding pocket where hydrolysis occurs, H1a binds to γ-phosphate which assists with the hydrolysis. H2 works with Mg^{2+} for catalysis. H1 is also known as P-loop, interacts with the α- and β-phosphates of the nucleotide attached to the catalytic site. The γ-phosphate sensor is found in H3. H4 includes the α6 helix which is directly connected to L2 loop that binds to RNA. According to studies based on the x-ray crystal structures of Phi12P4 in different states during ATP hydrolysis [15], it was revealed that α6 helix and L2 loop create a lever with "up" or "down" positions (see Figure 6C) [1]. The lever is found in both positions in the absence of nucleotide. During ATP hydrolysis, the P-loop faces conformational changes which effects the L2 loop, causing change in the levers' position from "up" to "down" [17]. This motion drags down to the tip of the lever ~6 Å, where lysine is located and binds RNA. The L1 loop is flexible acting as a grommet and fixing the RNA at the center of the channel [15]. The arginine finger contributes to the catalytic site from the neighboring subunit and by pointing into the binding pocket it neutralizes the negatively charged phosphates of the nucleotide and the transition state is stabilized resulting in a faster ATP hydrolysis [18]. The arginine finger can be found at different structural positions since it is not well conserved among helicase sequences.

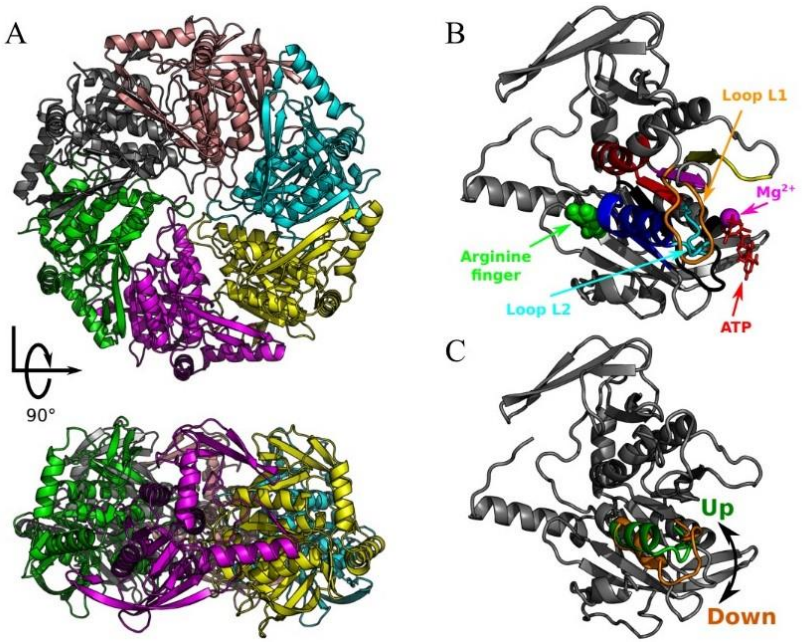


Figure 6: The fold of P4 proteins. As an example, structure of Phi12P4 is considered [15]. (A) Top and lateral view of the helicase. (B) Structural conservation of helicase motifs of P4 proteins. H1, H1a, H2, H3, H4, L1 and L2 are marked in black, yellow, magneta, red, blue, orange and cyan colors (in order). The arginine finger, Mg^{2+} ion and ATP analogue (AMPcPP) are marked in green, magneta and red colors. The L1 loops was modelled. (C) two different positions of the lever and L2 loop obtained from the x-ray crystal structure of phi12P4. The "up" position is marked in green, the "down" position is marked in orange [15].

Further on, P4 protein of bacteriophages is composed of three main domains: N-terminal region (110-150 residues), a central core NTPase domain (~160 residues) and a smaller C-terminal domain (~40-50 residues) (see Figure 7) [15]. Attachment of ssRNA to the primary binding site found on outer surface of P4 initiates the formation of the P4-RNA complex. As a result, the P4 hexamer ring opens and the RNA enters the central channel and binds to flexible charged loops. RNA binds directionally by P4, causing extension of the 5' terminus of the strand from wider C-terminal base of the P4 protein [19].

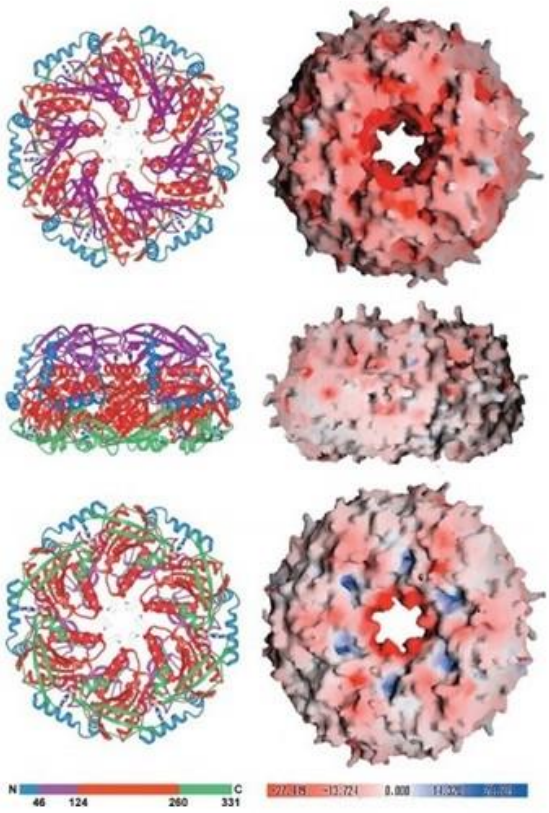


Figure 7: Structure of P4 hexamer [15]. Shown is the secondary structural elements and solvent accessible surface in top, side and bottom views. Different segments of P4 monomer marked by different colors: N-terminal safety pin motif (in blue), all β domain (in dark purple), conserved RecA-like ATP binding domain (in red) and antiparallel β strands and C-terminal helix (in green). Six molecules of AMPcPP, represented as ball-and-stick located in clefts between monomers. The solvent accessible surface of P4 (without nucleotides present) is marked in colors according to the electrostatic potential (defined in the key). The solvent-exposed face of the P4 is seen from the top view and the C-terminal face against located against the procapsid is seen from the bottom view. All representations are done by a previous study [20].

1.4.2 P4 protein: Function

P4 protein (NTPase) is essential for genome packaging into viral capsids, which protects the NA from the nuclease degradation occurring inside and outside the host cell. P4 is also important during a semi-conservative transcription to release the mRNA from the core particles (i.e. transport of mRNA in the opposite direction of packaging) [20]. There are three ways by which bidirectionality works: 1) Strand separation at replication fork driven by P4; 2) P4 acting as a passive but necessary channel for ssRNA that are newly transcribed; 3) Active export of RNA outside the core by P4 [7].

1.4.3 Phi8P4

As mentioned previously, Phi8P4 shares helical function and structure similar to those of hexameric helicases. The following properties indicate that mechanism of RNA packaging might be similar to translocation mechanism of the nucleic acid by helicases. Phi8P4 hydrolyzes both purines (A and G) and pyrimidines (T and C) based on previous studies [21]. The two advantages of Phi8P4 system regarding the RNA translocation reaction are: 1) Protein P4 shows a close coupling between ATP hydrolysis and RNA translocation; 2) the stability of P4 protein depends on presence of magnesium (Mg) without the need of di- or triphosphates [22]. The Phi8P4 and P4 protein of other bacteriophages are closely related to each other and Phi8P4 is a member of distinct subfamily of RecA-type ATPases according to the ATPase core domain, which is a Rossman-type nucleotide-binding domain made of a twisted seven-stranded β-sheet (parallel and antiparallel) surrounded by five helices. The residues (see Table II) of ATPase domain are essential for coupling ATP hydrolysis with RNA translocation and all are conserved except for one residue H4 missing in Phi8P4 [23].

Table II: Phi8P4 conserved residues and their functions. Modified from [23].

Amino acid	phi8P4	Function	Walker motif
Lysine	K116	Phosphate binding	H1
Serine/Threonine	T117	Phosphate binding	H1
Glutamate	E141	Catalytic base	H1a
Aspartate	D171	Coordinate Mg	H2
Aspargine	N216	Sensor motif	H3
Lysine	K185	RNA binding	H4
Serine	S237	Sensor motif II	H4
Arginine	R263	Arginine finger	
Glutamine		Base stacking	
Arginine	R266	Arginine finger	
Tyrosine/Phenylalanine	F247	Base stacking	

According to a previous study, the residues found in the N-terminal of Phi8P4 are visible in its crystal structure, covers the apical part of the hexamer P4 protein (see Figure 8) and has a higher secondary structure content. The first 12 and 31 residues that are invisible play an important role in stability of the protein. N-terminal also contains two helices separated by a four-stranded antiparallel β-sheet. P4 protein and P1 shell directly interact with viral lipid membrane because of lacking P8 nucleocapsid protein layer in Phi8P4. C-terminal of P4 protein consists of the least number of aa residues (~40-60), is usually located at the bottom

part of the hexamer protein with the importance for binding to the capsid protein P1. Nevertheless, the domain of each bacteriophage is different from each other. In Phi8P4 the L1 loop containing a motif (LKK) is crucial for RNA binding, as RNA enters the central channel through ring opening, therefor C-terminal domain is important for ATP hydrolysis, as it limits the diameter of the central channel and blocks the interface through which RNA is loaded [23].

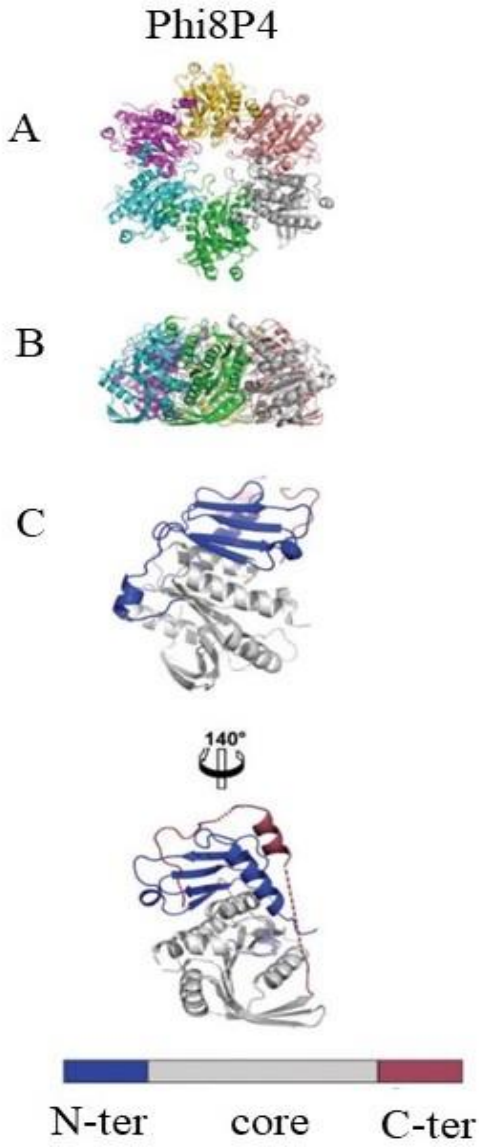


Figure 8: Overall structure of Phi8P4 protein. (A) Top view of the structure marked in different colored chains. (B) Side view of the structure. (C) The panel showing the P4 monomeric structures in two different orientations, The upper orientation of the monomer corresponds to the one in cyan in (B) whereas the lower orientation is after rotation of 140ºC with the purpose of showing the C-terminal domain. The core domain is in gray color, N-terminal domain in blue color and the C-terminal domain in red. Nucleotides, when present, are referred to as sticks of carbon, oxygen, nitrogen and phosphorus atoms in yellow, red, blue and orange colors, respectively. Finally, disordered regions of the protein are referred to as dotted lines. Modified from [23].

This thesis was focused on production, purification and assessment of helicase activity of Phi8P4. I did clone the protein into pASK-IBA37plus vector (with N-terminal Histidine-Tag) and produced His-taged Phi8P4 protein in *E. coli* expression. I did purify it using two different chromatography methods, namely Immobilized Metal Affinity chromatography - IMAC and Size Exclusion Chromatography - SEC. I did test its helicase function in unwinding the RNA hairpin in presence of ATP using single-molecule fluorescence techniques.

2 Goals

- Cloning of whole Phi8P4 gene into expression vector pASK-IBA37plus with N-terminal His-tag.
- Production and purification of Phi8P4 in *E.coli* expression system.
- Development of FRET assay for Phi8P4 helicase activity.

Table III: List of Buffers used and their composition.

Name	Composition
Staining buffer	0.625g Coomassie 250 Brilliant Blue, 112.5 mL Methanol, 112.5 mL H_2O, 25 mL acetic acid
Destaining buffer	60% H_2O, 30% MeOH, 10% acetic acid
Elution buffer	20 mM Tris, 50 mM NaCl, 5mM $MgCl_2$ 1 M imidazole, 0.02% NaN_3, pH= 8.5
Equilibriation buffer	20 mM Tris, 50 mM NaCl, 5 mM $MgCl_2$, 0.02% NaN_3, pH= 8.5
Leammli sample buffer (4X)	0.5 M Tris, 3.5 mM SDS, 20 mL Glycerol, 0.04 mM Bromophenol blue
Lysis buffer	50 Mm KH_2PO_4, 400 mM NaCl, 100mM KCl, 10% glycerol, 0.5% TrironX-100, 10mM imidazole, pH =7.8
Resuspension buffer	20 mM Tris, 400 mM NaCl, 7.5 mM $MgCl_2$, 1x protease inhibitor (SIGMA Protease Inhibitor Cocktail tablets, EDTA-free), pH = 8.5
SDS-PAGE sample buffer (1X)	100 µL β-mercaptoethanol, 900 µL 4X Leammli sample buffer
SDS-PAGE transfer buffer (10X)	58.15g Tris, 29.3g glycine, 3.75g SDS, 800 mL dH_2O
TAE buffer (50X)	242g Tris, 57.1 mL acetic acid, 100 mL 0.5 M EDTA in H_2O
TBS buffer (10X)	60.5g Tris, 87.66g NaCl, 800 mL ultra-power water, pH= 7.6

3 Material and methods

3.1 Material

The plasmid carrying Phi8P4 gene with C-terminal His-tag was provided by Roman Tůma, University of Leeds.

3.2 The design of PCR primers

Two gene specific primers F and R as shown were designed and synthesized by Generi Biotech, s.r.o. BsaI restriction site is in bold and stop codon in red. Primers were used for the amplification of whole Phi8P4 gene, which was cloned into the pASK-IBA37plus expression vector.

Forward (F) – ATGGTA**GGTCTC**AGCGCGCTAGAAAAACGAAAGTTACA

Reverse (R) - ATGGTA**GGTCTC**ATATCTTATTTGTCAACTCCTTCAATGTA

3.3 Polymerase chain reaction (PCR)

3.3.1 Gradient PCR

To test the optimal annealing temperature of the designed primers, gradient PCR was done using TaKara Ex Taq DNA polymerase (TaKaRa Bio) in Biometra Tone Cycler (Analytik Jena). The reaction started with initial denaturation at 98°C for 30 seconds and composed of 30 cycles of the following 3 steps (see Table IV), and then final elongation at 72°C for 7 min.

Table IV: Gradient PCR reaction cycle at specified temperature and duration.

Step	Temperature	Duration / sec
Denaturation	95°C	15
Annealing	45-65°C	40
Elongation	68°C	60

The PCR reactions (20 µL) were prepared by using 2 µL of 10X Ex Taq buffer, 0.4 µL of dNTPs (10 Mm), 0.1 µL DNA polymerase TaKaRa Ex Taq (5U/µL), 2 µL of both F and R primers (10 µM each), 0.5 µL of template DNA and filled up with ddH_2O. For the visualization of results, 1% agarose gel was prepared by dissolving 0,5 g of agarose (SERVA Electrophoresis) in 50 mL of 1X TAE buffer (see Table III). PCR product was stained using Serva DNA Stain Clear G (SERVA Electrophoresis).

3.3.2 PCR using Q5 High-Fidelity DNA Polymerase

For the amplification of the cloning fragment proofreading Q5 High-Fidelity DNA polymerase (New England Biolabs) was used. PCR reaction (50 µL) were prepared by using 10 µL of 5X Q5 reaction buffer, 1 µL of dNTPs (10 mM), 0.5 µL Q5 High-Fidelity DNA Polymerase (2U/ µL), 2.5 µL of F and R primers (10 µM each), 10 µL of 5X Q5 High GC enhancer, 1 µL of template DNA and was filled up with ddH_2O.

The PCR was started with initial denaturation temperature at 98°C for 30 sec and was repeated in 30 cycles of the following 3 steps (see Table V):

Table V: PCR reaction cycle at specific temperature and duration.

Step	**Temperature / °C**	**Duration / sec**
Denaturation	98	10
Primer annealing	52	15
Elongation	72	30

The last PCR cycle is followed by a final elongation step 2 min at 72°C.

3.4 Purification of the PCR product

PCR amplicon was purified using NucleoSpin Gel and PCR Clean-Up (Machery-Nagel) according the the manufacturer´s instruction with one exception where water was used instead of the elution buffer). The concentration of the purified amplicon was measured by NanoPhotometer Pearl (Implen).

3.5 Restriction enzyme digestion

For the digestion of both DNA amplicon as well as the cloning vector, the restriction enzyme BsaI-HF was used. The reaction contained 1 µg of either DNA amplicon or cloning vector, 5 µL of 10X CutSmart Buffer which is specific for BsaI-HF, 1 µL of BsaI-HF restriction enzyme and was filled up with ddH_2O to total volume of 50 µL. The mixture was incubated at 37°C for 15 min followed by enzyme deactivation at 65°C for 20 min in Biometra TOne Thermal Cycler (Analytik Jena). Then, the digested plasmid was dephosphorylated by adding 2.5 µL of Shrimp Alkaline Phosphatase (New England Biolabs). The reaction was incubated at 37°C for 30 min followed by inactivation at 65°C for 5 min. To verify successful digestion, 1% agarose gel was used and Serva DNA Stain Clear G (SERVA Electrophoresis) as staining solution, followed by purification of digested plasmids and measuring the concentration.

3.6 Ligation

3.6.1 Instant Sticky-End Ligase Master Mix

Instant Sticky-End Ligase Master Mix (New England Biolabs) was used to ligate Phi8P4 into the pASK-IBA37plus cloning vector according to manufacturer's instructions. Briefly, first mixture with molar ratio of 3:1 (DNA amplicon: vector) was prepared through addition of 17.67 ng/3.75 µL of DNA amplicon and 20 ng/1.25µL of vector and combined with nuclease-free water to have a final volume of 5 µL and were finally mixed with 5 µL of instant sticky-end ligase master mix and placed on ice prior next step.

3.7 Transformation of *E.coli* Competent (One Shot TOP10) cells

The transformation was done according *One Shot TOP10* Chemically Component *E.coli* Product Information Sheet (Invitrogen by Thermo Fischer Scientific). Briefly, competent cells were thawed on ice for 20 min and 2 µL of ligation reaction was added and mixture was gently mixed by flicking. The vials were incubated on ice for 30 min followed by heat shock for 30 sec at 42°C and then kept on ice for 2 min. 250 µL of pre-warmed S.O.C. LB medium was added to the vial and the cell were shaken at 37°C for 1 hour at speed of 220 rpm. Next 50 µL and 200 µL of transformation reaction were spread on two pre-warmed LB agar dishes containing ampicillin (50 µg/mL) and incubated overnight at 37°C.

3.8 Colony PCR

The quickest way to confirm the presence of the DNA amplicon in TOP10 competent *E.coli* cells was by performing colony PCR using Top-Bio PPP Master mix. A sterile pipette tip was used to pick individual colony and the colony was transferred into an Eppendorf tube filled with 20 µL Milli-Q water and was well mixed. Each tube was labelled with the same number as the colony. PCR reaction (20 µL) mixture was prepared as follows: 1 µL of template (colony), 2 µL 10X Taq reaction buffer, 1 µL of F and R primers (10 µM each) 0.5 µL dNTPs (10 mM), 0.5 µL Taq Polymerase and was filled up with dH_20 to a final volume of. In total 11 reactions were prepared, 10 with colony and 1 just filled with water without any colony marked as negative control. The PCR was initiated with initial denaturation at 98°C for 6 min and was continued with 25 cycles of the following 3 steps (see Table VI):

Table VI: Colony PCR running conditions

Step	**Temperature / °C**	**Duration / sec**
Denaturation	94	15
Annealing of primers	52	15
Elongation	72	60

The last PCR cycle is followed by a final elongation step 7 min at 72°C.

3.9 DNA Plasmid isolation

Mini prep was made using the positive colony from colony PCR by adding it into 4 mL LB media containing ampicillin (50 µg/mL). Cells were left to shake overnight at 37°C and 220rpm. Next day, plasmid DNA was isolated using NucleoSpin Gel and PCR Clean-Up (Machery-Nagel) according to the instructions. Afterwards, the concentration of the pure plasmid was measured with NanoPhotometer Pearl (Implen) and sent for sequence verification to SEQme company.

3.10 Transformation to *E.coli* Competent BL21(DE3) cells

BL21(DE3) (Invitrogen) competent *E.coli* cells were transformed using sequence verified expression plasmid (see chapter 3.7) according manufacturers instructions.

3.11 Pilot expression

To determine the optimal conditions for the protein production the Pilot expression was done. Overnight culture (10 mL) containing LB media and ampicillin (50 μg/mL, final concentration) were prepared and inoculated with one colony of BL21(DE3) cell and was left to shake overnight at 37°C using the horizontal shaker with speed of 220 rpm. Following day, 500 μL of the overnight culture was added to 4 new falcon tubes each containing 10 mL LB medium with ampicillin (50 μg/mL). The tubes were shaken at 37°C with speed of 220 rpm until the OD_{600} was 0.5-0.8. Once the desired value was reached the protein production was induced by anhydrotetracycline (200 μg/L) and cells were shaken at 37°C and 18°C with speed of 220 rpm. 1 mL of uninduced samples was collected. Every two hours 1 mL of both induced and uninduced samples was collected in case of 37°C (time period of 6 hours) and 18°C (time period of 24 hours), centrifuged using the microcentrifuge for 30 sec at maximum speed and the pellets were frozen at -20°C.

3.11.1 SDS-PAGE

The cells from the pilot expression were thawn and resuspended in 500 μL of Lysis buffer (see Table III). Cells were lysed by 3 cycles of freezing in liquid nitrogen and thawing at 42°C. Samples were spun down using microcentrifuge at maximum speed for 5 min at 4°C in order to isolate the soluble and insoluble fractions. The supernatant was removed and mixed with 125 μL of 4X SDS-PAGE sample buffer and was boiled for 5 min. then the pellets were mixed with 500 μL of 1X SDS-PAGE sample buffer in 1X PBS (see Table III) and were boiled for 5 min. 12.5% SDS-PAGE gels were prepared (see Table VII). The first column of the gel was loaded with 10 μL of PageRuler Protein Prestained Ladder (Thermo Fischer Scientific) followed by uninduced samples and induced samples with different collection times. After the run, stacking gel was disposed and the resolving gel was stained overnight using staining buffer (see Table III). The gel was distained using destaining buffer (see Table III) to remove the extra staining buffer (Coomassie blue) from the gel.

Table VII: SDS-PAGE gel composition.

Resolving gel

	12.5%
30% Acrylamide	2.08 mL
H_2O	1.57 mL
1.5M Tris (pH = 8.8)	1.25 mL
10% SDS	50 µL
10% APS	50 µL
TEMED	5 µL

Stacking gel

	4%
30% Acrylamide	340 µL
H_2O	1.36 mL
0.5M Tris (pH = 6.8)	250 µL
10% SDS	20 µL
10% APS	20 µL
TEMED	2 µL

3.12 Phi8P4 protein production

The optimal conditions for the protein production were monitored by SDS-PAGE of the pilot expression experiment. The overnight culture (20 mL of LB medium with 50µg/mL Ampicillin) from freshly transformed *E. coli* competent BL21(DE3) cells was prepared and left to shake overnight with speed of 220 rpm at 30°C using the horizontal shakers. Next day, 15 mL of ON culture were transferred into 800 mL of LB medium mixed with 400 µL ampicillin (20 µg/mL) and were left to shake at 30°C. Once the OD_{600} of value 0.65 was reached, the culture was cooled down in the fridge for about 30 minutes and then the protein production was induced with anhydrotetracycline (200 µg/L) and the culture was shaken at 18°C, 200 rpm for 20 hours. The cell culture was collected by centrifugation using the centrifuge (SIGMA 3-30K) for 30 min at 4°C and 4000 rpm. The cell pellet was stored at -80°C prior purification.

3.13 Phi8P4 protein purification

The cell pellets were defrosted on ice and resuspended in 10 mL of resuspension buffer (see Table III), which was supplemented with 1x protease inhibitor cocktail (SIGMAFAST Protease Inhibitor Cocktail tablets, EDTA-free by Sigma Aldrich). In order to lyse the cells, the cell suspension was passed twice through the cell disruptor press (Stansted Fluid Power) and the produced lysate was then centrifuged using the HIMAC CP 90WX preparative ultracentrifuge with speed of 25000 rpm at 4°C for 1 hour. The supernatant was then used for IMAC purification.

3.13.1 Purification by immobilized metal affinity chromatography (IMAC)

The Phi8P4 protein purification was done using ÄKTA Pure system (GE Healthcare) fitted with HisTrap HP 5 mL column (GE Healthcare) which was washed with de-gassed dH_2O 25 min (5X volume of the column) with a flow rate of 5 mL/min. Then, it was equilibrated with equilibration buffer (see Table III) 25 mL, elution buffer (see Table III) 25 mL and again with equilibration buffer 25 mL. Afterwards, the sample was loaded on the column by the sample pump with flow rate of 1 mL/min. Phi8P4 protein was eluted using elution buffer gradient from 0% to 100% elution buffer over 20 minutes. Once the UV signal started to rise 2 mL fractions were collected. Finally, the system was again equilibrated using the equilibration buffer, washed with H_2O and kept at 20% ethanol. For the analysis of the collected fractions with the purified protein SDS-PAGE was used.

3.13.2 Purification by size exclusion chromatography (SEC)

Phi8P4 protein was further purified using ÄKTA Pure system (GE Healthcare) fitted with Superdex 200 Increase 10/350 GL column (GE Healthcare). The column was first washed with H_2O and equilibration buffer (see Table 1). The protein sample from IMAC purification was loaded on the column and the oligomeric state of protein was analyzed using multiangle laser light scatter detector *Dawn8+* (Wyatt Technology). 1 mL fractions of the purified proteins were collected. Finally, the SDS-PAGE was used for the validation of the protein purity.

3.14 Single-molecule fluorescence detection (SMFD)

3.14.1 Advantages of using SMFD

One of the greatest advantages of SMFD is the capability of revealing real statistical distributions of statically heterogenic and dynamic populations of macromolecules. In other words, we do not detect only average values, which are usually much less sensitive conformational changes [24]. In case of our RNA hairpin, using SMFD, we can precisely determine the ratio of folded and unfolded hairpins, as well protein-bound RNA conformation.

3.14.2 Single-molecule förster resonance energy transfer (smFRET)

SmFRET is a distance-dependent physical process by which energy is transferred nonradiatively from an excited molecular fluorophore (the donor) to another fluorophore (the acceptor) by means of intermolecular long-range dipole–dipole coupling [25]. SmFRET can be used as a "spectroscopic ruler" at angstrom distances (10–100 Å)". Used as a technique, it is highly efficient if the donor and acceptor probes are positioned within the Förster radius (R_0: the distance at which half the excitation energy of the donor is transferred to the acceptor, typically 3–6 nm) (see Figure 9). The FRET efficiency (E) is dependent on the inverse sixth power of intermolecular separation (r) (see equation 1) [25]. This makes it a common technique of choice for investigating a broad variety of biological phenomena that produce changes in molecular proximity [26].

$$E = \frac{1}{1+\left(r/R_0\right)^6} \qquad \text{eq. 1}$$

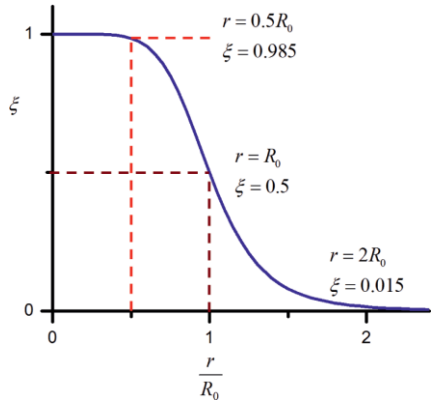


Figure 9: FRET efficiency (E) as a function of the donor-acceptor (D-A) distance (r) normalized with respect to the Förster radius (R_0) [27]. ξ refers to FRET efficiency (E).

Our technological implementation of smFRET is called Periodic Acceptor eXcitation (PAX) with two lasers green (cw laser diode, λ = 532, CrystaLaser) and red (pulsed laser diode, λ = 638 nm, PicoQuant) [28]. PAX yields information about labelling stoichiometry (S), so for further analysis we consider only fully labelled molecules with exactly one donor and one acceptor molecule (S~0.2-0.7) (see Figure 18). This helps to remove population of bleached donor molecules, which would otherwise be wrongly interpreted as a low FRET population.

3.14.3 Fluorescence correlation spectroscopy (FCS)

From early 1970s FCS has gradually developed as a method for the study of conformation dynamics, and molecular diffusion in solution and on membranes and chemical kinetics [29]. In FCS we measure the fluctuations of the fluorescence intensity which are caused by random effects (i.e. noise) as well as other effects such as chemical, biological, and physical on the target fluorophore. These slight fluctuations serve as an input information and are, decoded using the correlation algorithm to yield meaningful data. Here we use FCS to detect macromolecular structural dynamics and protein binding. Huge FCS advantage, is that in a single measurement it can also possess a large temporal linear dynamic range, from nanoseconds to minutes and can be specifically calculated for distinct populations of molecules of interest, defined for example by different FRET characteristics. For details (see Figure 10).

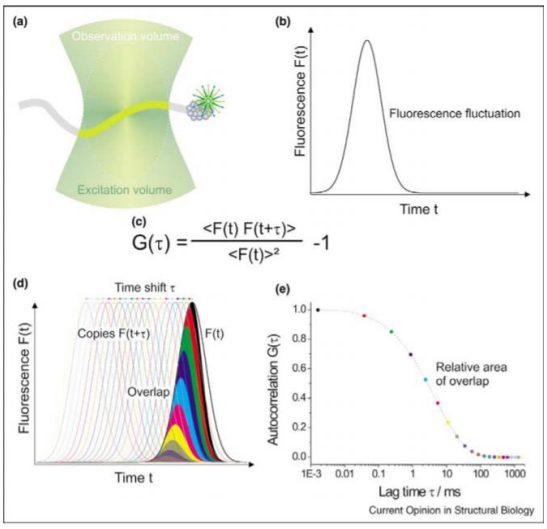


Figure 10: Fluorescence Correlation Spectroscopy captures differences in diffusivity of biomacromolecules. Molecules diffusing through a tight confocal volume (a) yield small fluctuations of the fluorescence intensity F(t) **(b).** These fluctuations are analyzed by autocorrelation function (G (τ)) **(c).** G(τ) is calculated shifting the lag time (t), which produces copies of shifted fluctuations f(t+τ) of the original signal F(t) (d). Both are multiplied by the area under the formed curve (overlap integral) which equals the value of the autocorrelation for this specific lag time as shown in **(e)** [30].

The autocorrelation signal G(τ), is calculated with respect to the self-similarity of the fluorophore diffusing in and out of the probe volume after lag time (τ). A mathematical model can be used to describe the autocorrelation curve, which provides quantitative information (i.e. diffusion time, number of molecules in the probe volume, diffusion coefficient, or molecular brightness). For measurements in solution, the probe volume is defined by a 3D Gaussian profile, in which a 3D Brownian diffusion model can be used (see equation 2) [29]:

$$G(\tau) = \frac{1}{<N>}\left(1+\frac{\tau}{\tau_D}\right)^{-1}\left(1+\frac{\tau}{S^2\tau_D}\right)^{-1/2} \quad \text{eq. 2}$$

where $<N>$ is the average number of particles in the detection area, and τ_D is the diffusion dwell time. This 3D diffusion model includes an S term, which defines the shape of the probe volume (the ratio of the z and xy radii, wz/wxy) [29].

3.15 Using SMFD to develop activity essay of Phi8P4 protein with N-terminal histidine-tag

After the Phi8P4 protein was produced and purified, the next step was to determine whether it kept its helical activity. To answer this question, we designed a short RNA hairpin (see Figure 11) with FRET labels (see table VIII), which should report on unfolding caused by Phi8P4.

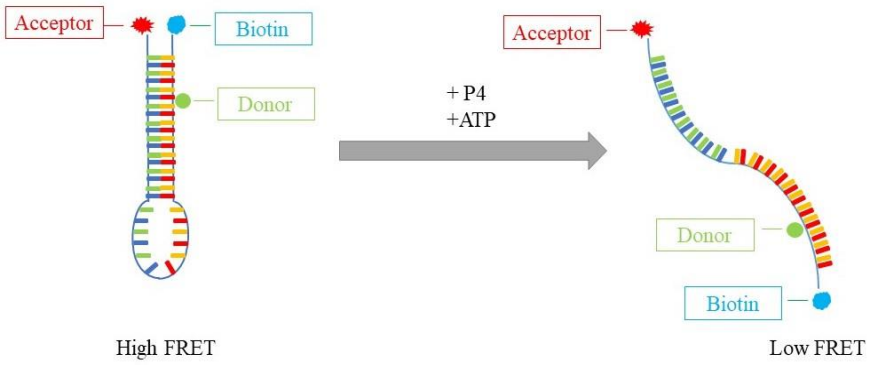


Figure 11: Opening of RNA hairpin. Closed form of the RNA hairpin (left) yields high-FRET, after addition of the helicase and ATP, the RNA hairpin opens (right) resulting in low FRET. Donor (iCy3) dye is shown in green, acceptor dye (ATTO 647N) is depicted in red and biotin tag is blue.

Table VIII: Sequence of the RNA hairpin. Biotin tag and organic fluorophores Cyanine 3 (Cy3 in green) and ATTO 647N (red) were attached during the commercial synthesis (Integrated DNA technologies, Iowa USA).

/5BiotinTEG/rCrU rCrCrG rCrGrU rGrUrC rUrCrC /iCy3/rUrU rCrCrC rArUrC rCrGrC rGrUrC rCrGrC rGrArA rCrGrC rGrGrA rCrGrC rGrGrA rUrUrU /3ATTO647NN/

Dependence of the helical activity on ATP concentration was determined from quantification of populations of folded and unfolded hairpins derived from smFRET experiments

To assess whereas ATP hydrolysis is needed for Phi8P4 binding, we used FCS to measure changes in diffusion coefficients of the hairpin upon addition of Phi8P4. Later we used smFRET-filtered FCS to determine diffusion coefficients of individual smFRET populations. All data analysis was performed in locally modified version of FRETBursts package [31]. All FCS and smFRET images were created using Python Matplotlib libraries [32].

All smFRET and FCS experiments were performed in the equilibration buffer (see Table III). Final concentration of labeled RNA hairpin was 80 pM. In experiments with phi8P4 and ATP, the final concentrations were 20 µg/mL of phi8P4 and 2.5 mM ATP.

4 Results

4.1 Cloning of Phi8P4

Plasmid carrying the whole sequence of Phi8P4 was a generous gift of Dr. Roman Tůma, University of Leeds and University of South Bohemia. The gene specific primers encoding whole phi8P4 (length: 963 bp, 321 aa) were designed with specific restriction site, suitable for cloning into pASK37plus expression vector (see Appendix 8.1).

The optimal annealing temperature of both primers was tested using the gradient PCR with temperature range between 55°C and 65°C (data not shown). The optimal annealing temperature was set to 52°C and was used for amplification of Phi8P4 gene with Q5 polymerase. Resulting PCR product was resolved using 1% agarose gel (see Figure 12) and purified from the gel yielding the DNA concentration of 42.3 ng/μL. Both the DNA amplicon and pASK-IBA37plus vector (for more details see Appendix 8.1) were digested using BsaI-HF restriction enzyme. The success of restriction reaction was verified using 1% agarose gel electrophoresis. Number 1 (see Figure 13), a clear band is seen at approximately 3000 bp is observed referring to cleaved linearized plasmid which migrated slower than the

original supercoiled un-cleaved recombinant DNA.

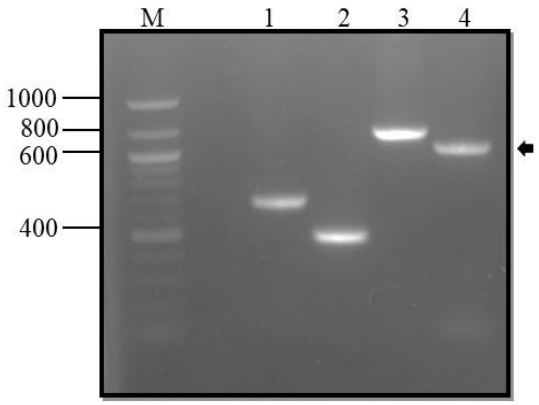


Figure 12: PCR product. Marker used (M): 100 bp DNA Ladder (New England Biolabs). The arrow refers to target amplicon labelled number 4, number 1-3 are not related to this experiment.

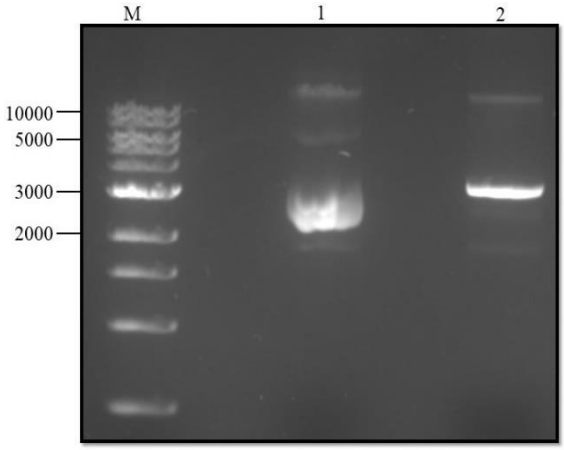


Figure 13: Restriction Analysis. Marker used (M): 1 kb DNA Ladder (New England Biolabs). Number 1 refers to uncleaved plasmid, number 2 refers to cleaved plasmid.

After successful amplification and digestion, Instant Sticky-end MM was used for ligation of the recombinant DNA as described in methods and materials (see chapter 3.6.1).

4.1.1 Transformation of *E.coli* competent (One Shot TOP10) cells and colony PCR

After thawing *E.coli* competent (One Shot TOP10) cells on ice, the ligation mixture was added and then incubated on ice followed by heat shock at 42°C. The cells were treated with warmed S.O.C. LB media and shaken using horizontal shakers. The cells were spread on the LB medium petri-dishes with Ampicillin (50 µg/mL) and were incubated overnight at 37 °C in static incubator. Next day, individual colonies were picked for colony PCR. In total 11 samples were prepared, 10 colony samples and 1 negative control (NC). TOP-Bio PPP master mix was used as described in chapter 3.8 and for result analysis, 1% agarose gel (see Figure 14) was used. Only 1 colony with a band of size ~1000 bp was positive. Mini preps were prepared using colony number 7 (positive colony), grown overnight. The plasmid DNA was isolated and measured (110 ng/mL) and then was sent to SEQme company for sequencing and the results confirmed the presence of Phi8P4 sequence in pASK-IBA37plus vector without any frameshift.

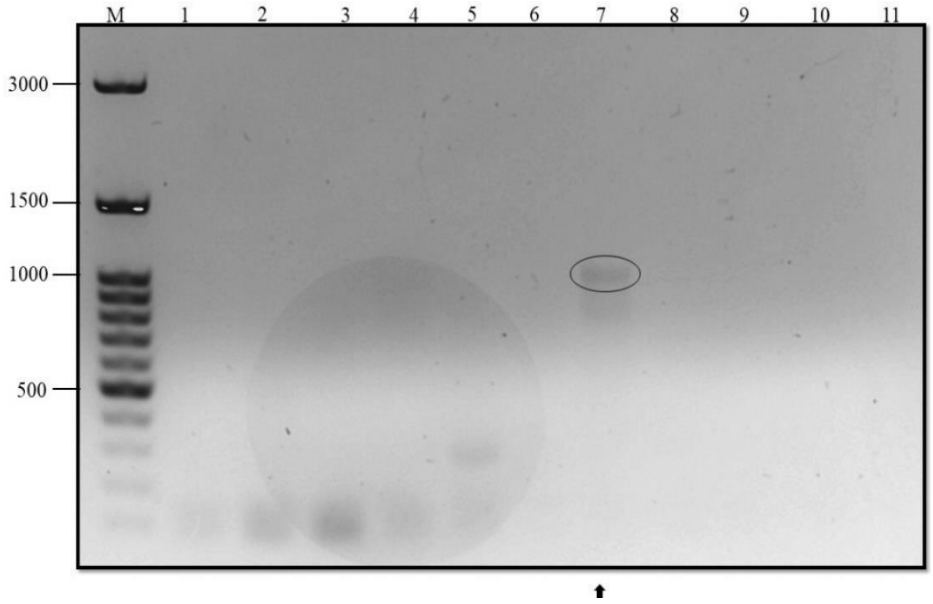


Figure 14: Gel electrophoresis colony PCR samples. Marker used (M): 100 bp DNA Ladder H3 RTU (Nippon Genetics). The positive colony (number 7) is marked by arrow and circle.

4.2 Protein production

4.2.1 Pilot expression

Colony number 7 was further transformed into *E.coli* competent BL21(DE3) cells as described previously (see chapter 4.2). The overnight cultures were prepared by picking up colonies from the petri-dishes, inoculated with LB media containing ampicillin (50 µg/mL) and were shaken overnight at 37°C. The next day, 500 µL of the overnight culture was mixed with new 10 mL prepared LB medium and ampicillin and left to shake at 37°C until the desired of OD_{600} of 0.7 was reached. After cooling down and splitting into 2x 10 mL cultures uninduced and induced by anhydrotetracycline (200 µg/L), the Phi8P4 protein production was initiated as the cultures were incubated at both 18°C (total 24 hours) and 37°C (total 6 hours). After collecting 1 mL samples (uninduced and induced) every two hours, they were centrifuged, and the pellets were separated and were kept at -20°C. Furthermore, the cells were defrosted, lysed, and thawn at 42°C, following centrifugation using microcentrifuge at 4°C. The pellets were mixed with 1X SDS-PAGE sample buffer in 1X PBS and the supernatant was mixed with 4X SDS-PAGE sample buffer, followed by boiling. SDS-PAGE was used for analyzing the samples (see Figure 15). The highest amount of the protein of expected size ~35 kDa, was present in soluble fraction of *E. coli* cells upon induction with

anhydrotetracycline and incubation at 18°C for 24 hours. These conditions were used for large scale production.

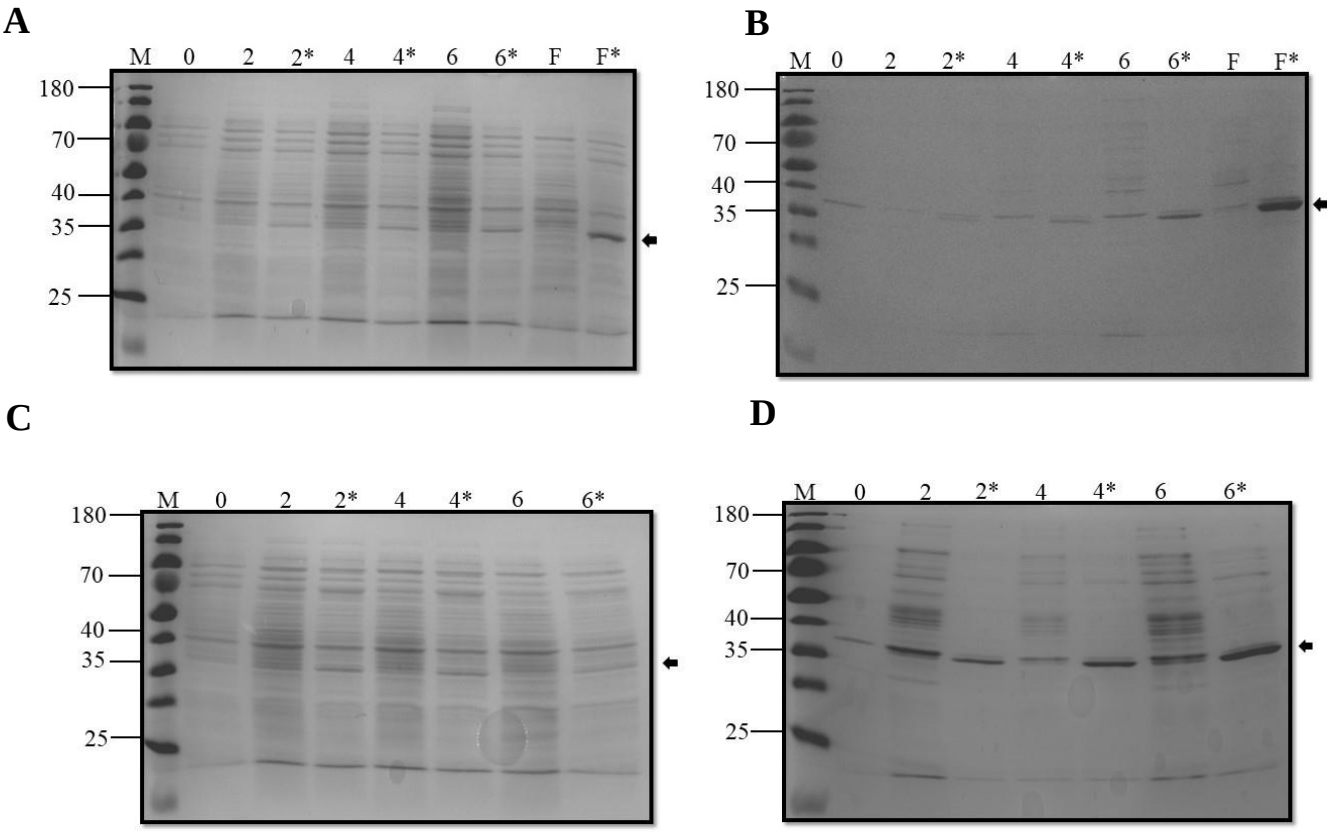


Figure 15: SDS-PAGE analysis of protein expression at 18°C (A – soluble fraction, B – insoluble faction) and 37°C (C – soluble fraction, D – insoluble fraction) collected at different time intervals. Marker used (M): PageRuler Protein Ladder (Thermo Fischer). Each number corresponds to the hour at which 1 mL of culture was collected, (*) indicates induced cultures and number only indicates uninduced cultures. F: corresponds to 24 hour time interval. Arrows refer to the Phi8P4 protein.

4.2.2 Large-scale protein production

After the optimal conditions for Phi8P4 protein was figured to be at 18°C (soluble fraction) with speed of 220 rpm, large-scale protein production was done. *E.coli* competent BL21(DE3) cells were transformed by Phi8P4 plasmid, plated and the single colony was used to inoculate 20 mL of LB media with 50 µg/mL ampicillin. The cultures were left to shake overnight at 30°C with speed of 220 rpm. Next day, the 15 mL of overnight culture was mixed with 800 mL of new liquid LB media containing ampicillin (50 µg/mL) and was left to shake until the OD_{600} of value 0.65 was reached and were left to cool down and protein production was

initiated by induction with anhydrotetracycline (200 µg/L) and shaking for 20 hours at 18°C. Then, the protein cultures were centrifuged for 30 min with speed of 4000 rpm at 4°C. The pellets were stored at -80°C prior further use.

4.3 Protein purification

4.3.1 Purification of Phi8P4 by immobilized metal affinity chromatography (IMAC)

The Phi8P4 cells were removed from -80°C freezer, thawn and resuspended using the resuspension buffer (contains 1x protease inhibitor) Cells were lysed by French press (twice) and the resulting cell lysate was centrifuged for 1 hour by the HIMAC CP 90WX preparative ultracentrifuge with speed of 25000 rpm at 4°C. The Phi8P4 protein was loaded on HisTrap column and eluted using gradient (0-100% imidazole) of elution buffer for 20 min. 2 mL fraction were collected and analyzed by SDS-PAGE (see Figure 16).

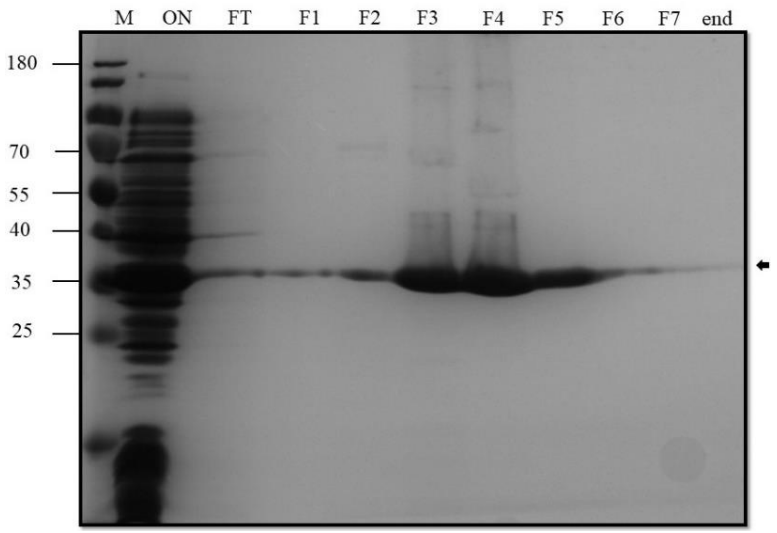


Figure 16: SDS-PAGE of Phi8P4 purification by immobilized metal ion affinity chromatography (IMAC). Marker used (M): PageRuler Prestained Protein Ladder (Thermo Fischer). ON: fraction containing supernatant of phi8P4, which was loaded on the column. FT: flow-through fraction. F1- F7: fractions collected during gradient. End: final fraction collected in case there was some protein present. The arrow refers to phi8P4 protein (size: 35 kDa).

4.3.2 Purification of Phi8P4 protein by size exclusion chromatography (SEC)

Since the fractions F3 & F4 from IMAC contained the majority of purified Phi8P4 protein they were poured into one fraction and were concentrated to a volume of 1 mL and were used for further purification by size exclusion chromatography (SEC). 1 mL fractions were collected. The analysis was done by SDS-PAGE (see Figure 17).

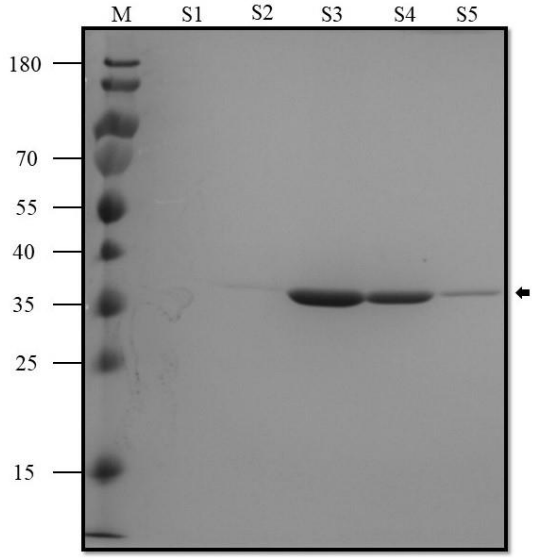


Figure 17: SDS-PAGE of Phi8P4 purification by size exclusion chromatography (SEC). Marker used (M): PageRuler Prestained Protein Ladder (Thermo Fischer). S1-S5: fractions obtained from purifying F3 & F4 from IMAC. The arrow refers to pure phi8P4 protein (size: 35 kDa).

4.4 Activity assay of Phi8P4 protein with N-terminal histidine-tag

After pure hexametric Phi8P4 protein (full length) was produced, the next step was to test its main activity as a helicase in unwinding the RNA hairpin loop.

4.4.1 Helical activity of Phi8P4 determined by RNA hairpin opening

The sample of RNA hairpin (80 pM) resulted in abundant population of high-FRET species in closed state (see Figure 18).

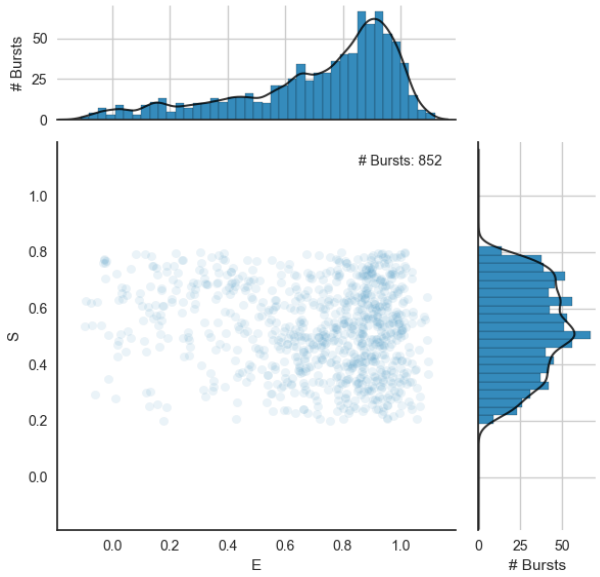


Figure 18: Determination of RNA hairpin conformational state shown as FRET Efficiency (E) vs. Stoichiometry (S). Here we show RNA hairpin in equilibration buffer. Conformational space is clearly dominated by closed state of RNA hairpin. We use this 2D representation to prove that for final FRET analysis, we take into account only completely labelled molecules. In other words, we exclude RNA loops with donor or acceptor only dyes.

RNA hairpin was found mostly in closed state (see Figure 19A). Addition of Phi8P4 protein (20 μg/mL) to the RNA hairpin mixture (80 pM) did not yield any significant change in FRET, in other words, the RNA hairpin did not open (see Figure 19B). Only after addition of 2.5 mM ATP to the mixture, RNA hairpin showed a significant shift and in FRET and yielded abundant population of low FRET species. This we interpret as hairpin opening caused by ATP driven hydrolysis of Phi8P4 (see Figure 19C).

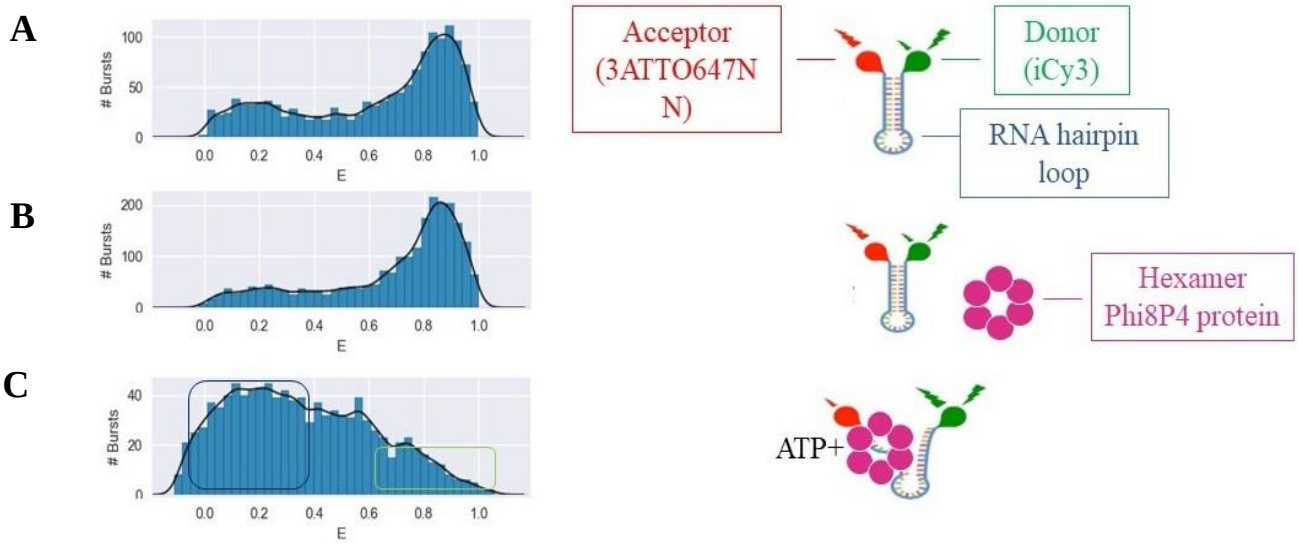


Figure 19: FRET efficiency (E) histograms. (A) RNA hairpin alone yields dominant high FRET population indicating that majority of hairpins are in closed state. (B) After addition of phi8P4 protein (20 μg/mL, with N-terminal his-tag) the histogram hasn't changes, indicating, that phi8P4 didn't trigger any change in RNA conformation. (C) Addition of 2.5 mM ATP yields significant changes in FRET histogram, showing dominant peak at low FRET values (i.e. open configuration), compare populations indicated by black and green rectangles.

4.4.2 ATP induced binding of Phi8P4 helicase to RNA hairpin

Due to smFRET, we have already detected conformational states of RNA hairpins (see figure 19). Using derived FRET regions as computational filters for FSC, we can now determine diffusion coefficients for each state. This allows us to further link conformational states of RNA to protein binding.

We have constructed a set of filters based on two FRET regions for (i) open and (ii) closed hairpin, these were derived from data (see Figure 20) (black line – open hairpin, green line – closed hairpin).

Linearized RNA showed comparatively longer diffusion times than closed hairpins (see figure 20), indicating presence of bound Phi8P4 hexamers.

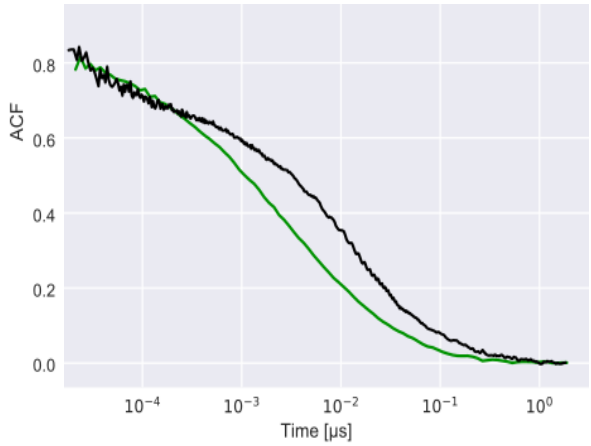


Figure 20: Diffusion time measured by fluorescence correlation spectroscopy. Black line: autocorrelation function of low FRET population. Mixture of RNA hairpin with phi8P4 protein and ATP shows an increase in diffusion time of the ATP dependent complex formation followed by increase in size. Green line: autocorrelation function of high FRET population.shows closed state of RNA hairpin with significantly shorter diffusion time.

5 Discussion

The findings of this thesis show that the produced and purified Phi8P4 protein with N-terminal Histidine-Tag exhibits a ring structure and helical activity driven by ATP hydrolysis like those of hexameric helicase family.

Our main task was to produce a functional Phi8P4 protein and test its helical activity with labelled RNA. We cloned the plasmid holding the whole sequence of Phi8P4 into pASK-IBA37plus expression vector with N-terminal His-tag by gene specific primers (forward and reverse) with specific restriction site, as we wanted to have the C-terminal native important for RNA binding, supported by a previous study which showed that RNA binds to P4 protein through its central channel and the C-terminal domain controls the diameter of the central channel, blocking the interface through which RNA loading and binding [23].

The agarose gel results of the digested DNA plasmid showed slower migration speed than the original one, because of the net like structure of agarose gel and it is harder for a linearized structure to pass through.

The expression vector pASK-IBA37plus is known for its high protein yield, induced by tetracycline and nearly no leakage expression. Regardless, not commonly used compared to pSJ1B vector with IPTG induction used for producing Phi8P4 [33].

Then we produced Phi8P4 protein with size ~35 kDa using *E.coli* expression cells, due to its low cost, easy transformation and fast cell growth. In our study, we used similar conditions as a previous study conducted for production of the protein with C-terminal domain [33] and the protein was present in all conditions (37ºC and 18ºC) but in low amounts.

As the artificially added His-tag has a high affinity towards nickel, we purified Phi8P4 protein using a Histrap nickel column, eluted by imidazole containing buffer, since imidazole will compete with Histidine-Tag in order to bind to nickel and allow for the protein molecules to elute from the column. Followed by size-exclusion chromatography which separates the protein based on its size and shape. Both purification methods were already used in a study for Phi8P4 with C-terminal domain [19].

Furthermore, we observed that pure Phi8P4 protein bound to RNA hairpin did not show any influence on FRET efficiency (E) and therefor no conformational change of RNA hairpin. But once ATP was added to the protein mixture, it caused unwinding of RNA hairpin. All these data could be used to monitor the helical activity of Phi8P4 in presence of ATP. Our result is

also supported by a previous study, stating that the isolated P4 proteins of bacteriophage Phi8 exhibits a helical activity [23].

Based on FCS measurement of diffusion time of the complex between RNA loop and Phi8P4 protein, we observed that binding of the Phi8P4 protein to RNA hairpin happens in open state of the loop and that it is ATP dependent.

6 Conclusion

Genome packaging into a procapsid is the most important step in a life cycle of a bacteriophage, it is done by P4 hexamer protein with an energy from ATP.

Since the activity of previously produced Phi8P4 protein with C-terminal domain was measured, yet the crystal structure with RNA and the mode of action were not clear, the work of this thesis was focused on cloning and production of Phi8P4 using pASK-IBA37plus (with N-terminal Histidine-Tag). The main work included optimization of protein production conditions. Cloning of the protein using *E.coli* BL21(DE3) seemed a bit challenging, due to only a single positive colony produced, but most importantly the production of Phi8P4 protein under all conditions and sufficiently in soluble fraction was successful even if in low amounts. The second important goal, testing the helical activity of Phi8P4 on RNA hairpins showed that ATP is not needed for the protein to bind and it was also required for the protein to unwind the RNA hairpin.

Nevertheless, to fully understand the mechanical action of ATP-dependent helicase activity of this molecular machine, further research is needed. We propose to employ smFRET on individual RNA loops and/or Phi8P4 proteins to dissect individual steps and elucidate the mechanical action of this vitally important process in genome packaging of this bacteriophage.

7 References

[1] Radou, G. (//2015). Helicase Functional Dynamics from Low-Resolution Experimental Data and Simulation. University of Leeds, Leeds, England.

[2] Mertens, P. (2004). The DsRNA Viruses. Virus Research 101, 3-13. https://doi.org/10.1016/j.virusres.2003.12.002

[3] Poranen, M.M., Mäntynen, S., and ICTV Report Consortium (2017). ICTV Virus Taxonomy Profile: *Cystoviridae*. *Journal of General Virology* 98, 2423–2424.

[4] Semancik, J.S., Vidaver, A.K., Van Etten, J.L. (1973) Characterization of segmented double-helical RNA from bacteriophage phi6. J Mol Biol 78(4):617–625

[5] Kainov, D.E., Lísal, J., Bamford, D.H., Tuma, R. (2004). Packaging motor from double-stranded RNA bacteriophage phi12 acts as an obligatory passive conduit during transcription. Nucleic Acids Res 32(12):3515–3521.

[6] Kainov, D.E. *et al* (2003). RNA packaging device of double-stranded RNA bacteriophages, possibly as simple as hexamer of P4 protein. *J. Biol. Chem* 278, 48084-91.

[7] Mindich, L., 1999. Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage φ 6. Microbiol. Mol. Biol. Rev. 63, 149–160.

[8] Hoogstraten, D., Qioa, X., Sun, Y., Hu, A., Onodera, S., Mindich, L., (2000). Characterization of phi8, a bacteriophage containing three double stranded RNA genomic segments and distantly related to φ 6. Virology 272, 218–224.

[9] Sun, Yang & Qiao, Xueying & Qiao, Jian & Onodera, Shiroh & Mindich, Leonard. (2003). Unique properties of the inner core of bacteriophage φ8, a virus with a segmented dsRNA genome. Virology. 308. 354-61. 10.1016/S0042-6822(03)00013-8.

[10] Mindich, L., *et al* (1999). Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. J. Bacteriol. 181, 4505–4508.

[11] Olkkonen, V.M., Ojala, P., Bamford, D.H., (1991). Generation of infectious nucleocapsids by in vitro assembly of the shell protein onto the polymerase complex of the dsRNA bacteriophage φ 6. J. Mol. Biol. 218, 569–581.

[12] Lisal, J. (2015). Mechanism of RNA Translocation by a Viral Packaging Motor (Academic dissertation). University of Helsinki and Vikki Graduate School in Bioscinces, Helsinki, Finnland.

[13] Singleton, M.R., Dillingham, M.S., Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76:23–50.

[14] Ye, J., Osborne, A.R., Groll, M., Rapoport, T.A. (2004). RecA-like motor ATPases - lessons from structures. *Biochim Biophys Acta* 1659(1):1–18.

[15] Mancini, E.J., Kainov, D.E., Grimes, J.M., Tuma, R., Bamford, D.H. and Stuart, D.I. (2004). Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. Cell, 118, 743–755.

[16] Ilyina, T.V., Gorbalenya, A.E., Koonin, E.V. (1992). Organization and evolution of bacterial and bacteriophage primase-helicase systems. J Mol Evol 34(4):351–357.

[17] Lísal J, Tuma R (2005) Cooperative mechanism of RNA packaging motor. J Biol Chem 280(24):23157–23164.

[18] Kötting, C., Kallenbach, A., Suveyzdis, Y., Wittinghofer, A., and Gerwert, K. (2008). The GAP arginine finger movement into the catalytic site of Ras increases the activation entropy. Proc Natl Acad Sci 105(17):6260–6265.

[19] Astier, Y., Kainov, D.E., Bayley, H., Tuma Roman, and Howorka, Stefan (2007). Stochastic detection of motor peotein-RNA complexes by single-channel current recording. *ChemPhysChem* 2007, 8, 2189 – 2194. Doi:10.1002/cphc.200700179

[20] Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281-296.

[21] Pirttimaa,M.J., Paatero,A.O., Frilander,M.J. and Bamford,D.H. (2002) Nonspecific nucleoside triphosphatase P4 of double-stranded RNA bacteriophage ϕ6 is required for single-stranded RNA packaging and transcription. *J. Virol.*, 76, 10122–10127.

[22] Lisal, J., Kainov, D.E., Bamford, D.H., Thomas, G.J., Jr., and Tuma, R. (2004). Enzymatic mechanism of RNA translocation in double-stranded RNA bacteriophages. *J. Biol. Chem.* 279, 1343–1350

[23] Omari K El, et al. (2013). Tracking in atomic detail the functional specializations in viral RecA helicases that occur during evolution. Nucleic Acids Res 41(20):9396–9410.

[24] Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., & Ha, T. (2008). Advances in Single-Molecule Fluorescence Methods for Molecular Biology. *Annual Review of Biochemistry* 77, 51-76. https://doi.org/10.1146/annurev.biochem.77.070606.101543

[25] Sekar, R. B., & Periasamy, A. (2003). Fluorescence Resonance Energy Transfer (FRET) Microscopy Imaging of Live Cell Protein Localizations. *The Journal of cell biology*, *160*(5), 629–633. doi:10.1083/jcb.200210140

[26] Dos Remedios, C.G., &. Moens, P.D.J. (1995). Fluorescence Resonance Energy Transfer Spectroscopy Is a Reliable "Ruler" for Measuring Structural Changes in Proteins: Dispelling the Problem of the Unknown Orientation Factor. *Journal of Structural Biology*, 15(2), 175-185. https://doi.org/10.1006/jsbi.1995.1042

[27] Gaponenko, S., & Demir, H. (2018). Energy Transfer Processes. In *Applied Nanophotonics*, 210-226. Cambridge: Cambridge University Press. Doi: 10.1017/9781316535868.008

[28] Doose, S., Heilemann, M., Michalet, X. et al. Eur Biophys J (2007) 36: 669. https://doi.org/10.1007/s00249-007-0133-7

[29] Tian, Y., Martinez, M. M., & Pappas, D. (2011). Fluorescence correlation spectroscopy: a review of biochemical and microfluidic applications. *Applied spectroscopy*, 65(4), 115A–124A. doi:10.1366/10-06224

[30] Haustein, E., Schwille, P. (2004). Single Molecule Spectroscopic Methods. *Science Direct* 14, 531-540. https://doi.org/10.1016/j.sbi.2004.09.004

[31] Ingargiola, A., Lerner, E., Chung, S., Weiss, S., & Michalet, X. (2016). FRETBursts: An Open Source Toolkit for Analysis of Freely-Diffusing Single-Molecule FRET. *PLoS ONE* 11(8): e0160716. https://doi.org/10.1371/journal.pone.0160716

[32] Hunter, J.D. (2017). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 9(3), 90-95. doi: 10.1109/MCSE.2007.55

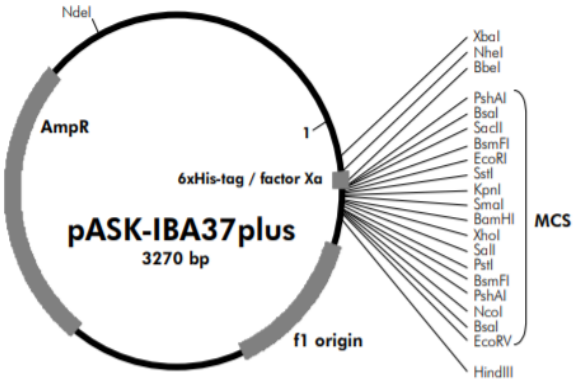
[33] Kainov, D.E., Butcher, S.J., Bamford, D.H., & Tuma, R. (2003). Conserved Intermediate on the Assembly Pathway of Double-Stranded RNA Bacteriophages. *J. Mol. Biol*. 328, 791-804. https://doi.org/10.1016/S0022-2836(03)00322-X

8 Appendix

8.1 pASK-IBA37plus (features)

Description	Plasmid for expression under transcriptional control of tetracycline as promotor/operator. Expression and localization of the recombinant protein in cytoplasm.
Affinity tag	Usage of 6xHistidine-tag for purification of recombinant protein. Affinity tag is fused to N-terminal of the recombinant protein and may be removed when cleaved with factor Xa.
Bacterial Expression	Induced expression by addition of anhydrotetracycline in *E.coli* shaking culture (A_{550} = 0.5) with final concentration of 200 µg/L.
Expression strain	Any *E.coli* strain, the tet-promoter works independently from genetic background of *E.coli*.
Resistance	Ampicillin.

	from bp	to bp
promoter	37	72
forward primer binding site	57	76
6xHistidine-tag	160	195
factor Xa cleavage site	196	207
multiple cloning site	208	284
reverse primer binding site	352	368
f1 origin	381	819
AmpR resistance gene	968	1828
tet-repressor	1838	2461



(IBA Lifesciences, available from IBA Solution for Life Sciences. Available from <https://www.iba-lifesciences.com/home.html>