Interaction of DNA and proteins in virus PBCV-1 particles

Structural organization of DNA in *Chlorella* viruses

TECHNISCHE UNIVERSITÄT DARMSTADT

Vom Fachbereich Biologie der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades
eines Doctor rerum naturalium
genehmigte Dissertation von

Dipl.-Biol. Timo Wulfmeyer
aus Groß-Gerau

Berichterstatter: Prof. Dr. Gerhard Thiel
Mitberichterstatter: Prof. Dr. Felicitas Pfeifer

Eingereicht am 05.10.2012
Mündliche Prüfung am 26.11.2012
Darmstadt 2012
D17

1. Table of contents

2. Chapter 1 - General Introduction

2.1. DNA compensation - A challenging process

DNA aggregation reveals in all known living organisms a difficult challenge. This is due to two basic physical principles of the DNA polymer: charge and size.

The deoxyribonucleic acid (DNA) is made up of two sugar-phosphate chains wrapped around each other in a helix. These strands are held in place by four bases which are directed towards the inner helical structure: adenine (A), guanine (G), cytosine (C) and thymine (T) [1]. The sugar-phosphate backbone includes a negative charge, which has to be compensated to allow DNA packaging. The 46 human chromosomes have a contour length between 16 and 82 mm. In comparison the diameter of human cells ranges in length from only 1 to 30 μm. Each chromosome contains between 48 and 240 million base pairs (bp) (34 nm/bp) [2]. Hence the DNA has to be packaged in order to store it inside a nucleus. The problem of packaging of a long DNA molecule into a small volume is not only relevant for eukaryotic cells but also for prokaryotes and viruses. For example, the volume of a typical *Escherichia coli* nucleoid is roughly 10^4 times smaller than the volume of a freely coiling linear DNA molecule with the same length as the *E. coli* genome [3]. Even more evident is the problem of DNA packaging in viruses. The phage λ particles for example have a 48 kbp genome, which is approximately 16.5 μm long [4]. This long DNA polymer has to be stored inside a viral particle with an inner diameter of merely 55 nm. The dense packaging causes a density of genome per volume of 0.5 bp/nm^3. These examples show that all organisms have to overcome the same physical barriers of DNA condensation and DNA charge compensation.

In evolution different strategies were developed for solving this problem and for guaranteeing a successful DNA packaging; altogether these strategies aim to store as much relevant genetic information in little volume as possible.

As a representative of bacteria *Escherichia coli* shall be mentioned. These bacteria use a mixture of typical chemical aggregation processes, which contribute to a high genomic compaction factor The formation of the bacterial nucleoid is facilitated by macromolecular proteins which allow the genomic condensation. It is also supported by small multivalent cations, which neutralize the charge of the DNA [4].

One important protein, which is involved in DNA condensation, is the *Escherichia coli* curved DNA-binding protein A (CbpA). This is a nucleoid-associated DNA-binding factor creating complex. When it is bound to DNA, CbpA forms large, highly dense aggregates, which further protect the DNA from degradation by nucleases [30].

Eukaryotic cells for example contain a crucial protein complex for DNA packaging and charge neutralization, the so-called histones [5]. Eukaryotic chromosomes consist of a regularly repeating protein–DNA complex called the nucleosome. Each nucleosome consists of a protein octamer made up of two copies of the histone proteins H2A, H2B, H3 and H4. Together with the fifth histone, H1, it organizes about 146 base pairs of DNA. Further organization involves the assembly of nucleosomes into higher-order chromatin structures [6]. Histones are extremely conserved proteins and a phylogenetic attribute of all kind of animal cells [7]. Due to the high content of positively charged amino acids (Arg and Lys) and the resulting basic isoelectric point, histones act as spools around which the negatively charged DNA is coiled. The 14 contacts per histone-octamer cause a binding energy of approximately 59 kJ/mol (each histone–DNA contact stores ~ 4.2 kJ/mol of energy) [8]. This electrostatic interaction between protein and DNA enables a compact packaging, which is necessary to fit the large genomes of eukaryotic cells inside the nuclei [9]. The protein-DNA interaction is a mutual interplay of both molecules. The DNA itself exhibits different structural properties, which increase the binding between protein and DNA. For example, the aforementioned histones and histone-like proteins are not generally considered to bind to DNA with absolute sequence specificity. Preferentially, the binding affinity rises at DNA fragments, which are composed of a higher content of GC-rich sequences. This is due to the fact, that a high genomic GC content may differ in the width of the major and minor grooves and makes the protein accessible for the DNA [13].

On the side of the protein, interaction with the DNA is mostly favored by electrostatic parameters. In histones, the ratio of negatively charged residues to positively charged residues (Asp+Glu)/(Arg+Lys) is very low (approximately: 0.28). Histone H4 for example is composed of 13.6 % Arg and 10.7 % Lys residues. The composition and number of the positively charged amino acids causes a high ionic attraction to the negatively charged phosphate groups.

As a result, the chromosomal DNA, which is approximately 2700 times larger than the diameter of a cell, can be packaged into an even smaller cell nucleus and the charge is compensated [10].

In addition to histones there are also other conserved proteins with a high content of positively charged amino acids for interacting with the DNA. These interactions are involved in a wide range of cellular processes ranging from cell cycle to cellular differentiation and development [11]. Besides histones, there are RNA polymerases, chromatin remodelling proteins, general transcription factors, several co-factors and a range of sequence-specific factors, which interact in different fashion with the DNA [12].
The DNA/protein interaction is, in these cases, supported by a specific conformation of the protein and by the amino acid composition. These structural microscopic and macroscopic elements of the protein architecture increase the binding affinity between protein and DNA. Further DNA-binding motifs developed during evolution. The following paragraph describes some of the most relevant amino acid motifs, which are involved in protein/DNA interaction:

Helix-turn-helix

The Helix-turn-helix is characterized by two protein helices which interact with the DNA. The C-terminal helix lies within the major groove of the DNA double helix in comparison to the second protein helix which binds to the sugar-phosphate backbone [14].

Helix-loop-helix

The protein binding-motif consists of a short segment of an alpha helix connected to a longer segment of an alpha helix by a loop. Both protein helices connect to the DNA's major groove [15].

Leucine zipper

Proteins of this type form homo- or heterodimers. Two alpha helices, one from each monomer, form a coiled-coil structure at one end due to interactions between leucines that extend from one side of each helix. Beyond the dimerization interface the alpha helices diverge, allowing them to fit into the major groove of the DNA double helix [16].

Zinc finger

It consists of a segment of alpha helix bound to a loop by a zinc ion. The zinc ion is held in place by two cysteine and two histidine groups. The alpha helix lies within the major groove of the DNA double helix. Zinc finger motifs are often repeated in clusters [17].

Other than the extremely specific proteins with their conserved binding motifs, which are common in eukaryotic and prokaryotic organisms, also viruses developed processes to achieve a high yield of DNA condensation. Theoretical considerations have shown, that the DNA in phages is highly organized [18]. The mechanism depends on one important parameter, namely that of positively charged molecules in the capsid, which compensate the negatively charged phosphate backbone of the DNA [19]. The main cation phage λ uses for this purpose is Mg^{2+}; but also polyamines, for instance putrescine and spermidine, are employed for charge compensation [20]. The polyamines appear to be surrounded by the protein of the phage; i.e. they are inside the phage head, which cotains the DNA. Hence the polyamines appear to be associated with the DNA rather than with the protein of the phage [3, 20, 21, 22].

Besides cations and polyamines, there are also other positively charged molecules used for this task. Especially cationic and fusogenic lipids increase the genomic aggregation. By becoming a liposome, the condensed DNA takes a toroid-like structure. Apart of the efficacy of liposome complexes is presumably due to the compact state of the DNA, which protects it from nucleases and allows it to pass more easily through small openings. The lipid coating on the DNA may also increase its permeability through cell membranes using it as a very efficient agent for transfection of eukaryotic cells [23]. For example a liposome which includes a 48 kb DNA fragment yields a toroid with an inner diameter of about 60 nm [32].

One further process, which is very well established in DNA condensation, is the principle of overlapping genes. These overlaps are typically assumed to be a form of genome compression, allowing the virus to increase its repertoire of proteins without increasing its genome length. Concluding, two or more proteins coded for the same nucleotide sequence.

However, it is supposed that a negative relationship between overlap proportion and genome length among viruses with icosahedral capsids consists, but not among those with other types of capsids [29].

As a resume, it can be concluded, that charge compensation of a DNA polymer bears some similarities to protein folding.
Due to the formation of compact, regular structures, the many types of non-covalent interactions and the accessibility for a wide range of different interactions, both processes are difficult to achieve [23].

2.2. PBCV-1 - The prototype *Chlorella* virus

The present work attempts to uncover the mechanisms by which the *Paramecium bursaria chlorella* virus 1 (PBCV-1) stores its large dsDNA genome in its capsid. PBCV-1 belongs to the family of the *Phycodnaviridae*, i.e. a group of DNA viruses infecting algae. The members of this family are important ecological elements in aqueous environments within the virioplankton. Along with other viruses, phycodnaviruses play an important role in the dynamics of algal blooms, nutrient cycling, algal community structure, and possibly gene transfer between organisms [24, 25]. *Chlorella* viruses are a subgroup of the phycodnaviruses; they are large, icosahedral, double-stranded-DNA containing viruses that replicate in certain strains of the unicellular green alga *Chlorella*. DNA sequence analysis of the 330 kbp genome of PBCV-1 predicts 365 protein-encoding genes. PBCV-1 has a multilayer structure with an outer glycoprotein capsid that surrounds a lipid bilayer membrane and the dsDNA genome. Vp54 is a glycoprotein and the major component of the capsid and represents about 40 % of the weight of the total protein content in the virus. It forms an icosahedral capsomer, which has a pseudo-6-fold symmetry.

Using cryo-electron microscopy, it has been shown, that the virion has a diameter varying from 165 nm (measured along the 2-fold and 3-fold axes) and 190 nm (measured along the 5-fold axes). The capsid shell consists of 1680 donut-shaped trimeric capsomers plus 11 pentameric capsomers, one at each icosahedral edge (see Figure 2.2.1) [26]. The PBCV-1 glycoprotein shell is composed of 20 triangular units (or trisymmetrons) and 12 pentagonal units (or pentasymmetrons) at the 5-fold vertices [27]. Additional cryo-electron microscopic analysis of PBCV-1 revealed a 54 nm long narrow cylindrical structure at one edge of the icosahedral capsid. This so-called spike seems to attach to the host cell wall and keeps the particle in contact with it.

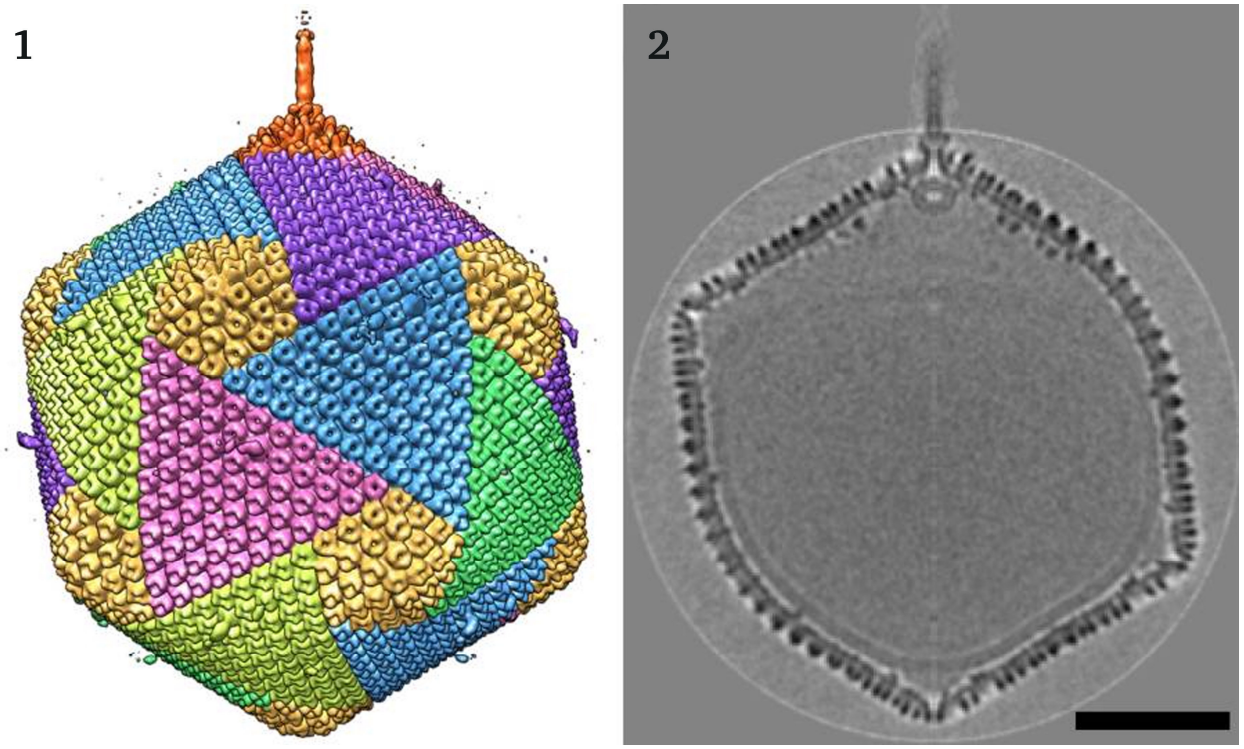

Figure 2.2.1: The icosahedral cryo-electron microscopic structure of PBCV-1.

1: Hexagonal arrays of major capsomers form trisymmetrons (blue, magenta, green) and pentasymmetrons (yellow). The unique vertex with its spike structure is at the top. 2: Central cross-section of the cryo-electron microscopy density (Scale bar: 50 nm) [26].

The attachment of PBCV-1 to the host receptor is a specific and highly affine process. Unlike in other viruses, the spike is probably not used for DNA delivery into the host. The spike diameter is at the outside of the capsid only 3.5 nm and hence to narrow for DNA to pass. The spike rather seems to serve for puncturing the cell wall of the host and for releasing cell wall digesting agents. It has been assumed, that the spike may furthermore play a role in the target-oriented navigation of the particle [26].

Circumstantial data suggest that, after a successful release of enzymes from the virus and after successful local digestion of the algal cell wall, the membrane of the virus and the plasma membrane of the host fuse due to the high internal pressure within the host cell [31, 34]. This enables the delivery of the viral DNA and virion-associated proteins into the host cell. There, DNA and proteins probably move to the nucleus where the early gene transcription begins after 5 - 10 min post infection (p.i.) The early mRNAs are transported to the cytoplasm for translation, and at least some early proteins presumably return to the nucleus to initiate viral DNA replication.

This begins 60 - 90 min p.i., followed by late gene transcription. Late mRNAs are transported to the cytoplasm for translation and many of these late proteins are targeted to the virus assembly centers where virus capsids are formed and DNA is packaged. The *Chlorella* cell wall is lysed, and the infectious PBCV-1 progeny is released at 6 - 8 h p.i. The processes of infection and replication are summarized in detail by [27].

The phycodnaviruses have some of the largest virus genomes ranging in size from < 170 to 560 kbp. In the case of PBCV-1 the genome consists of 330 kbp with an average length of 100 μm. This genome has to be packaged into a capsid with an inner diameter of about 80 nm resulting in a density of 0.15 bp/nm^3. In addition to the DNA, the virus also contains proteins. Measurements suggest that it contains 64 % protein, 23 % DNA and 7.5 % lipid [28]. In the context of these geometric parameters the process of DNA packaging in chloroviruses is similar to that in phages. In comparison, phage λ particles have a 48 kbp genome, which is approximately 16.5 μm long [3]. This causes a density of genome per volume of 0.5 bp/nm^3.
As already mentioned, the presence of lipids is an important mechanism involved in charge aggregation. With the help of the inner lipid bilayer, the viral structure creates a liposome in which the DNA is located. There is no doubt that the lipids in PBCV-1 play a role in charge compensation even though the molecular lipid composition is not known. It is assumed that especially the genomic charge at the surface, close to the lipid is compensated. As mentioned before, a 48 kb DNA fragment yields a toroid with an inner diameter of about 60 nm. Due to this fact, it is not possible to fully compensate the charge of the DNA, because for this purpose the inner viral diameter should have a size of about 640 nm.

The aspect that the packaging density of the chloroviruses is similar to that in phages implies that both types of viruses have a similar challenge for packaging their DNA into a small volume.

2.3. Aim of the work

The aim of the present work is to give an answer to the following question: “Do PBCV-1 virions contain proteins, which are responsible for DNA aggregation?”.

Recent experiments revealed that PBCV-1 does not include such a high concentration of polyamines like phage λ does. Due to this fact, PBCV-1 should have established another process for charge compensation which is not usual for viruses.

2.4. References

[1] **Richmond T. J., Davey C. A. (2003)**: The structure of DNA in the nucleosome core. *Nature* **423**, 145-150

[2] **Hayes J. J., Tulliust T. D., Wolffe A. P. (1990)**: The structure of DNA in a nucleosome. *Proc. Nati. Acad. Sci. USA* **87**, 7405-7409

[3] **de Vries R. (2010)**: DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins. *Biochemie* **92**, 1715–1721

[4] **Grayson P., Han L., Winther T., Philips R. (2007)**: Real-time observations of single bacteriophage λ DNA ejections in vitro. *Proc. Natl. Acad. Sci. USA* **104**, 14652-14657

[5] **Black B., Foltz D., Chakravarthy S., Luger K. (2004)**: Structural determinants for generating centromeric chromatin. *Nature* **430**, 578-82

[6] **Luger K., Mäder A.W., Richmond D.F. (1997)**: Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260

[7] **Khorasanizadeh S. (2004)**: The Nucleosome: From Genomic Review Organization to Genomic Regulation. *Cell* **116**, 259–272

[8] **Saha A., Wittmeyer J., Cairns B. R. (2006)**: Chromatin remodelling: the industrial revolution of DNA around histones. *Nature Reviews Molecular Cell Biology* **7**, 437-447

[9] **Kimura H. Cook P. (2001)**: Kinetics of Core Histones in Living Human Cells: Little Exchange of H3 and H4 and Some Rapid Exchange of H2B. *JCB* **153**, 1341-53

[10] **Shampo M. A., Kyle R. A. (1976)**: Albrecht Kossel. *JAMA* **235,** 1234

[11] **Bulyk M. L., Gentalen E., Lockhart D. J., Church G. M. (1999)**: Quantifying DNA–protein interactions by double-stranded DNA arrays. *Nat Biotechnol*. **6**, 573-7

[12] **Helwa R., Hoheisel J. D., (2010)**: Analysis of DNA–protein interactions: from nitrocellulose filter binding assays to microarray studies. *Anal Bioanal Chem* **398**, 2551–2561

[13] **Wellman S. E., Sittman D. B., Chaires J. B. (1994)**: Preferential binding of H1e histone to GC-rich DNA. *Biochemistry* **33***,* 384–388

[14] **Brennan R. G., Matthews B. W. (1989)**: The Helix-Turn-Helix DNA-binding Motif. *The Journal of Biological Chemistry* **4**, 1903-1906

[15] **Benezra R., Davis R. L., Lockshon D., Turner D. L., Weintraub H. (1990)**: The protein Id: a negative regulator of helix-loop-helix DNA-binding proteins. *Cell* **61**, 49-59

[16] **Landschulz W. H., Johnson P. F., McKnight S. L. (1988)**: The leucine zipper: a hypothetical structure common to a new class of DNA-binding proteins. *Science* **240**, 1759-1764

[17] **Pavletich N. P., Pabo C. O. (1991)**: Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809-817

[18] **Li Z., Wu J., Wang Z. G. (2008)**: Osmotic Pressure and Packaging Structure of Caged DNA. *Biophys. J.* **94**, 37-746

[19] **Bloomfield V. A. (1998):** DNA Condensation by Multivalent Cations. *Biopolymers/Nucleic Acid Sci.* **44**, 269-282

[20] **Ames B. N., Dubin D. T. (1960)**: The Role of Polyamines in the Neutralization of Bacteriophage Deoxyribonucleic. Acid. *J. Biol. Chem.* **235**, 769-775

[21] **Cordova A., Deserno M., Gelbart W. M., Ben-Shaul A. (2003)**: Osmotic Shock and the Strength of Viral Capsids. *Biophys. J.* **85**, 70-74

[22] **Yamada T., Furukawa S., Hamazaki T., Songsri P. (1996)**: Characterization of DNA-Binding Proteins and Protein Kinase Activities in Chlorella Virus CVK2. *Virology* **219**, 395 406

[23] **Bloomfield V.A. (1996)**: DNA condensation. *Current Opinion in Structural Biology* **6**, 334–341

[24] **Van Etten J. L., Lane L. C., Meints R. H. (1991)**: Viruses and viruslike particles of eukaryotic algae. *Microbiol Rev* **55**, 586–620

[25] **Yamada T., Onimatsu H., Van Etten J. L. (2006)**: Chlorella Viruses. *Adv.Virus Res.* **66**, 293-336

[26] **Cherriera M. V., Kostyuchenkoa V. A., Xiaoa C., Bowmana V., Van Etten J. L., Rossmann M. G. (2009)**: An icosahedral algal virus has a complex unique vertex decorated by a spike. *PNAS* **27**, 11085-11089

[27] **Van Etten J. L., Dunigan D. D. (2012)**: Chloroviruses: not your everyday plant virus. *Cell Press* **17**, 1–8

[28] **M. P., Burbank D. E., Xia Y., Meints R. H., Van Etten J. L., (1984)**: Structural proteins and lipids in a virus, PBCV-1, which replicates in a chlorella-like alga. *Virology* **135**, 308-15

[29] **Chirico N., Vianelli A., Belshaw R.**: Why genes overlap in viruses. *Proc. R. Soc. B* **277**, 3809

[30] **Cosgriff S., Chintakayala K. (2010)**: Dimerization and DNA-dependent aggregation of the Escherichia coli nucleoid protein and chaperone CbpA. *Mol Microbiol* **77**, 1289–1300

[31] **Mehmel M., Rothermel M., Meckel T., Van Etten J. L., Moroni A., Thiel G. (2003)**: Possible function for virus encoded K+ channel Kcv in the replication of *Chlorella* virus PBCV-1. *FEBS Lett.* **552**, 7-11

[32] **Reimer D. L., Zhang Y. P., Kong S. (1995)**: Formation of Novel Hydrophobic Complexes between Cationic Lipids and Plasmid DNA. *Biochemistry* **34***,* 12877- 12883

[33] **Zhang X., Xiang Y., Dunigan D. D., Klose T., Chipman P. R., Van Etten J. L., Rossmann M. G. (2011)**: Three-dimensional structure and function of the Paramecium bursaria chlorella virus capsid. *PNAS* **108,** 14837-14842

[34] **Thiel G, Baumeister D., Schroeder I., Kast S., Van Etten J. L., Moroni A (2011)**: Minimal art: or why small viral K(+) channels are good tools for understanding basic structure and function relations. *Biochimica et biophysica acta* **1808,** 580-8

3. Chapter 2 - PBCV-1 contains no appreciable amounts of cations or polyamines

3.1. Abstract

Condensation is defined as a decrease in volume of a molecule changing from a large to a compact state. In addition to condensation, the charge of a molecule poses also a difficult obstacle in dense packaging of molecules. A biological example for successful condensation of a charged molecule is the packaging of the long dsDNA genome of *Chlorella* viruses; these viruses package in case of PBCV-1 a ca. 100 μm long DNA molecule into a volume of only 1.7 * 10^{-7} m^3. The small amounts of polyamines, which are present in the *Chlorella* virus PBCV-1, are insufficient for charge compensation. To test whether the virus PBCV-1 uses like phages cations for charge compensation we measure here the content of cations in the particles of virus PBCV-1. Two independent methods, namely Energy dispersive X-ray spectroscopy (EDX) and Inductively coupled plasma-mass spectrometry (ICP-MS), are able to detect the DNA associated phosphate in the particles. The data indicate that PBCV-1 includes no sufficient amount of cations, which is required for the charge compensation of the DNA.

3.2. Introduction

Members of the phycodnaviruses contain some of the largest viral genomes, ranging in size from < 170 to 560 kbp. In the case of virus PBCV-1 the genome consists of a 330 kbp dsDNA molecule with an average length of 100 μm. Because the polymer molecule has to be packaged in a viral capsid with an inner diameter of only 80 nm the DNA has to be heavily condensed. The available space is in reality even more constrained since the capsid contains in addition to the DNA also many proteins. Measurements suggest that it is made of 64 % protein, 23 % DNA and 7.5 % lipid [1]. The challenge of DNA packaging in chloroviruses is in principle similar to that in many other virsues. On a more abstract level, all living organisms e.g. prokaryotes and eukaryotes have the same challenge for packaging their large genome in small volumes. The phage λ for example, has a 48 kbp genome, which is approximately 16.5 μm long [2]. This has to be packaged into a 300 times smaller viral coat.

The fact that the genomic density of the chlorovirus is similar to that in phages (Table 3.2.1) implies that both types of viruses have a similar challenge for packaging their DNA into a small volume.

Table 3.2.1: Physical characteristics of λ Phage and PBCV-1 concerning the DNA packing.

	λ Phage	**PBCV-1**
Genome [kbp]	48	330
Diameter [nm]	55	190
DNA linear [μm]	16.5	100
Ratio [bp/nm^3]	0.5	0.15

As a consequence of the dense packaging of DNA, viral capsids are often highly pressurized. Recent experiments have established, that the forces inside these capsids reach high values of up to 50 bars when the genomes are fully packaged [3]. The high pressure in the stored DNA in turn is the driving force for the rapid ejection of DNA from the virus particle into the host. Recent experiments have shown that the phage λ ejects its DNA because of this pressure with an initial velocity of 60 bp/s [4]. As a result the entire DNA can be transferred within a few seconds from the phage into the host [5].

The mechanism of DNA packaging in phages depends on one important structural characteristic namely the availability of positively charged molecules in the capsid, which are able to balance the negatively charged phosphate backbone of the DNA [6]. Mainly, with the help of cations and polyamines, phages can in this way neutralize more than 90 % of the charge [3]. The compensation of these electrostatic forces is essential for the effective condensation of the DNA molecule. Without these cationic molecules the strong repulsive interactions between the negatively charged DNA, which become relevant when the segments in the DNA molecule approach closely, would otherwise prevent an efficient condensation.

During packaging into virus particles the high molecular weight DNA undergoes a dramatic aggregation to a compact, usually highly ordered structure. The main cation used for this purpose in T4 bacteriophages is Mg^{2+}, but in addition, also polyamines like putrescine and spermidine are employed for charge compensation [6, 7]. The enthalpy of binding of spermine to DNA is 0 J/mol, compared with -1260 J/mol for poly(L-lysine) and +1470 J/mol for Mg^{2+}. Thus polyamine binding is an exothermic, entropy-driven reaction. It is defined as an energetically favourable reaction that occurs spontaneously when polyamines bind the DNA phosphate [8].

In case of virus PBCV-1, sequence analysis revealed an open reading frame, A237R, which encodes a protein with 34 % amino acid identity to the homospermidine synthase from *Rhodopseudomonas viridis*. This protein catalyzes the NAD^+-dependent formation of homospermidin. But it is unlikely that these polyamines are important for the neutralization of PBCV-1-DNA, because the total number of polyamine molecules per virion (~ 539 molecules) is too small for significantly neutralizing DNA (~ 660000 nucleotides) of virus PBCV-1. Hence, despite of the presence of a polyamine-synthesizing-enzyme, merely 0.2 % of the viral phosphate residues could be compensated by polyamines (Table 3.2.2). From these considerations we can conclude that polyamines can be neglected in the balance of total charges in the virus particle [9]. The function of a virus coded putative homospermidine synthase remains unknown and its role in the virus life cycle is still obscure [10].

Table 3.2.2: Polyamine content of PBCV-1.

Merely 0.2 % of the total phosphate residues could be neutralized with the polyamines [9].

Polyamine	Molecules/virion	Total polyamine charge	Phosphate residues neutralized (%)
Putrescine	277	554	0.08
Cadaverine	32	64	0.01
Spermidine	196	588	0.09
Homospermidine	34	102	0.02
Total	539	1308	0.2

It has already been mentioned that the bacteriophage T4 uses also cations to compensate the negative charges of its DNA [11]. If PBCV-1 would do the same, the total amount of cations in the particle should be equal to the phosphate concentration. To address the question whether the PBCV-1 particle indeed contains sufficient amounts of cations for compensation of the charged DNA we use here energy dispersive X-ray spectroscopy (EDX) and the inductively coupled plasma mass spectrometry (ICP-MS) to estimate the cation content in the particles. The results from both methods show that the PBCV-1 particles contain no appreciable concentrations of cations; charge compensation of the DNA must rely on other mechanisms.

3.3. Material and Methods

3.3.1. Strain and culture conditions

The production and purification of the *Chlorella* viruses and the isolation of the DNA was done as previously described [10, 12].

3.3.2. Virus quantification

With the help of plaque assays it was possible to determine the amount of viral particles (plaque forming unit, PFU) [38]. With this quantification it was possible to use same concentrations of freshly isolated viral particles (10^{11} virions/ml) for each analytical method. The host cells, *Chlorella variabilis*, (formerly named *Chlorella* NC64A), were concentrated to a final density of 4.0 x 10^8 cells/ml. The MBBM soft agar (0.75 %) tubes were liquefied and kept in a 48 °C water bath. Agar (1.5 %) plates were produced under sterile conditions and 1 μl tetracycline (10 mg/ml) was added. To each tube with soft agar 300 μl of the concentrated *Chlorella variabilis* and 100 μl of the diluted PBCV-1 virus were added. The tubes were briefly mixed and poured onto the plate. The dish was then gently rotated until the entire surface of the plate was covered with soft agar. The plates were subsequently incubated at 25 °C in continuous light. After 3 - 4 days the plaques were ready to count.

3.3.3. Energy dispersive X-ray spectroscopy (EDX)

The Energy dispersive X-ray spectroscopy (EDX) is an analytical technique to characterize the chemical composition of a sample. In this technique, a probe is exposed to an electron beam. This results in the emission of x-radiation with a distinct energy. This energy of the emitted x-rays is characteristic for each atom and the height of the peaks is proportional to the concentration of the elements. Here, viral particles were excited by an external electron beam to analyze the element specific x-ray emission. 50 μl of PBCV-1 or 50 μl 100 mM KCl, 50 mM NaCl solution were dried onto a carbon covered, copper grid (3.5 mm).
The transmission electron microscope (TEM) Philips CM12 with an accelerating voltage of 120 kV and an EDX-Detector (EDAX^{TM}) was used to characterize the resulting emission spectra.

For the measurement of embedded PBCV-1 virions highly concentrated particles were fixed with a cacodylate-buffered (pH 6.8) 2 % glutaraldehyde, 2 % formaldehyde (freshly prepared from paraformaldehyde) solution. After washing in buffer, samples were post-fixed in OsO4 (2 % in the same buffer), dehydrated in a graded acetone series, and embedded in Spurr's resin. Ultrathin sections were obtained with diamond knives, post-stained with uranyl acetate and lead citrate and examined with a Zeiss EM 109 transmission electron microscope. [14]

3.3.4. Inductively coupled plasma-mass spectrometry (ICP-MS)

ICP-MS (inductively coupled plasma-mass spectrometry) is a conventional spectroscopic technique for measuring trace amounts of molecules; it uses inductively coupled argon plasma as the source of atomic emission. It can in principle be used for the determination of all elements and for analyzing the molecular composition of complex samples. A solution with PBCV-1 particles in a buffer (50 mM Tris-HCl, pH 7.8) was diluted with highly concentrated sulphuric acid, 95 % ultrapure (1:2). Samples were then boiled in a sterile melting pot until the solution was fully evaporated. The resulting condensation was re-suspended in 65 % nitric acid, again vaporized and finally stored in ultrapure LS-MS-H_2O. The measurements were kindly performed by Merck KGaA (Darmstadt, Germany).

3.4. Results and Discussion

3.4.1. PBCV-1 contains no significant amounts of cations

To examine the mechanism by which virus PBCV-1 neutralizes the charges of its DNA and to address the question whether PBCV-1 contains significant concentrations of ions, energy dispersive X-ray spectroscopy (EDX) on virus particles was performed. A spectrum of the virus particles and a control solution (reference salt solution: 100 mM KCl, 50 mM $CaCl_2$) are shown in Figure 3.4.1.1 (black and grey graphs, respectively). Because the measurements of the salts were done on crystals, it is not possible to compare the intensity values of the spectra quantitatively; crystallisation leads to an unknown stoichiometry of the salts. Intensity values were scaled to the carbon peak (maximum of the peak is not displayed) for each recording and are presented as a function of the accelerating voltage [keV] in Figure 3.4.1.1.

The reference spectrum reveals the profile for the dominant cations and anions. Notably the spectrum of the virus particle reveals no significant signal at the energies of the respective cations. The copper (Cu) and silicium (Si) peaks in the spectrum are generated by the surface of the sample carrier on which the samples were transferred. The high values of carbon (C), nitrogen (N) and oxygen (O) are due to the organic composition of PBCV-1 and of the carbon cover. In contrast, the concentration of possible DNA neutralizing cations (i.e. sodium, potassium and calcium) is very low.

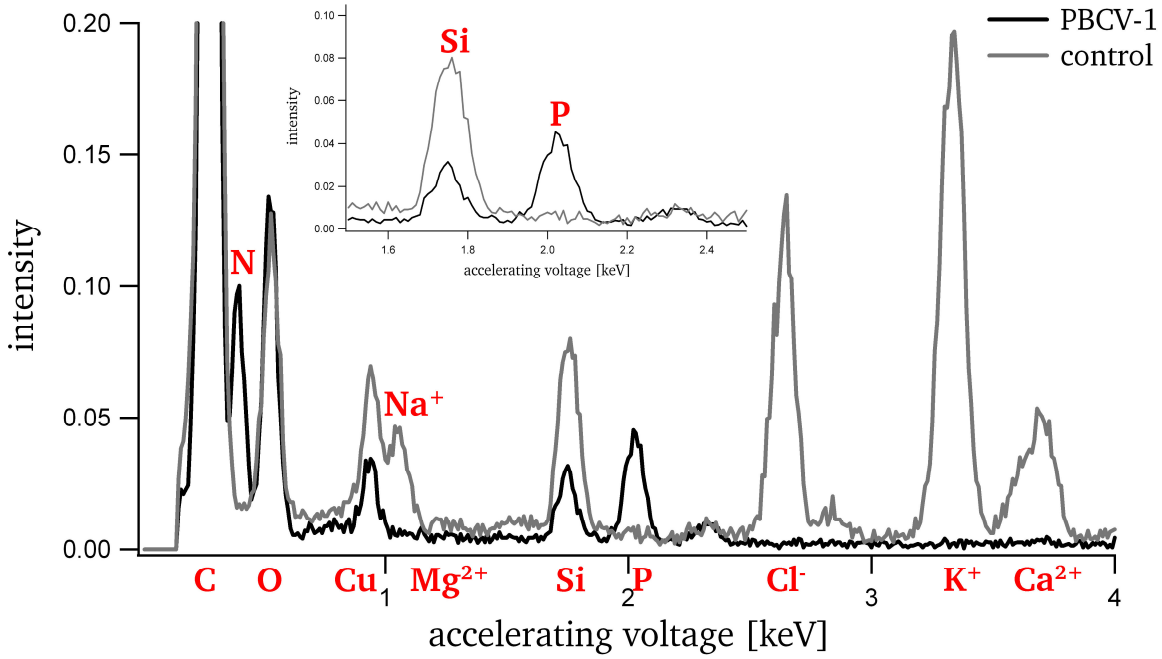

Figure 3.4.1.1: EDX analysis of the molecular composition of PBCV-1 virions and salt solution.

EDX spectra of Cu coated PBCV-1 (black) particles and a control solution (100 mM KCl. 50 mM $CaCl_2$, grey). The intensity is presented as a function of the accelerating voltage [keV]. Inset: Magnification of the phosphate peak. Due to the DNA a distinct phosphate peak in the viral sample was detectable.

C: carbon, N: nitrogen, O: oxygen, Cu: copper, Na^+: sodium, Mg^{2+}: magnesium, Si: silicium, P: phosphor, Cl^-: chloride, K^+: potassium, Ca^{2+}: calcium

In contrast to the reference, the spectrum of the virus particle also exhibits a large phosphate peak. The majority of phosphate in PBCV-1 comes from the DNA and the lipids. It is assumed, that each virus generally contains with its dsDNA approximately 660000 phosphate residues, which have to be compensated. The virus particle is roughly spherical with an inner diameter of 80 nm, which translates into a volume of 2.7 x 10^{-19} l. This corresponds to a concentration of about 3.3 M phosphate.

In the EDX spectra of PBCV-1, (Figure 3.4.1.1, inset) phosphate can clearly be detected in the viral capsids. If cations were compensating the charge, we would expect a signal of a similar magnitude from the respective cations. However, the spectra reveal that the major cations are absent in these recordings. The same type of experiment was repeated with thin sections of embedded virions. This procedure should eliminate possible artefacts from the external solution and increase the relative signal from the interior of the virus particles. Because the viral particles were fixed in a resin it was possible to collect the emitting energy from a single virus (Figure 3.4.1.2 A). This should increase the specificity of the recording.

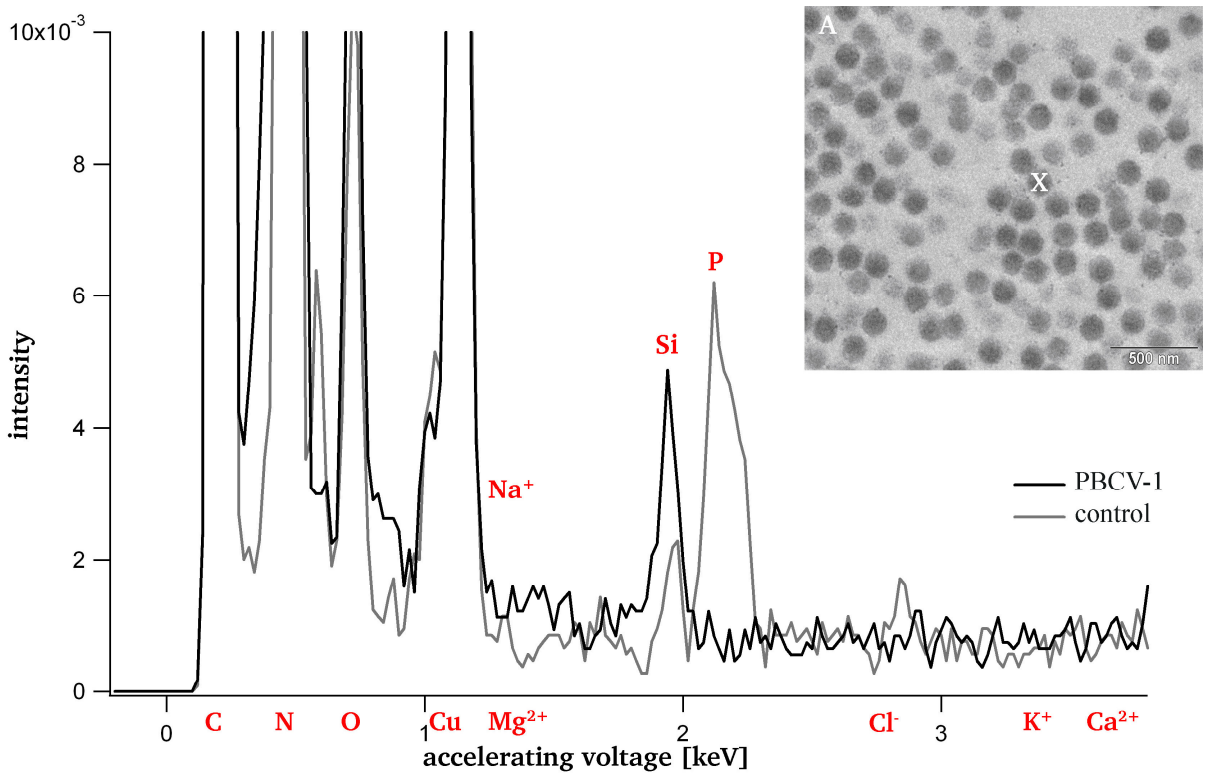

Figure 3.4.1.2: EDX analysis of the molecular composition of an embedded PBCV-1 virion.

EDX spectrum of embedded PBCV-1 virions (black) and a control (grey, resin). The intensity is presented as a function of the accelerating voltage [keV]. Inset: TEM-image of embedded PBCV-1 virions. For determination of the cationic content of a single viral particle (X), one virion was focused and the emitting energy spectra was detected.

C: carbon, N: nitrogen, O: oxygen, Cu: copper, Na^+: sodium, Mg^{2+}: magnesium, Si: silicium, P: phosphor, Cl^-: chloride, K^+: potassium, Ca^{2+}: calcium

The spectra show that it is not possible to resolve each signal; the intensity of highly concentrated elements such as carbon, nitrogen and oxygen mask smaller signals. This fact results in a shift along the x-axis.

The data confirm the aforementioned results of measurements on the whole particles: the phosphate is again detectable but there is no signal, which can be correlated with a relevant cation. In comparison with the control solution there is no major concentration of free cations in the viral particles; also the presence of complexed cations such as in cationic lipids can be excluded. The data do not priory exclude the presence of any cation in the virion; but their concentration must be below the sensitivity of the EDX method and much smaller than the phosphate concentration in the virion. Collectively the data stress that the concentration of cations is too small for compensating the viral DNA. Hence, other mechanisms must be employed for this task. Further experiments reported below substantiate the hypothesis that the storing of DNA in the particles of chloroviruses is different from that of phages.

3.4.2. The content of Mg^{2+} is one fifth of the total phosphate concentration

To verify the EDX data, we also determine the amount of cations in the virions of PBCV-1 with an independent and more sensitive method. The ICP-MS technique is suitable to detect even trace amounts of a molecule in a solution. The data from the ICP-MS measurements show that the viral particles contain calcium, magnesium, potassium and phosphor. Intensity values were scaled to the phosphor.

As in the EDX spectra, the ICP measurements also show that the content of phosphor is higher than that of the cations. Hence, the phosphor content of the particles is 5 times higher than that of each cation, which is associated with the particles (Figure 3.4.2.1). The data agree with the EDX measurements that the phosphor in the particle is not compensated by any single cation.

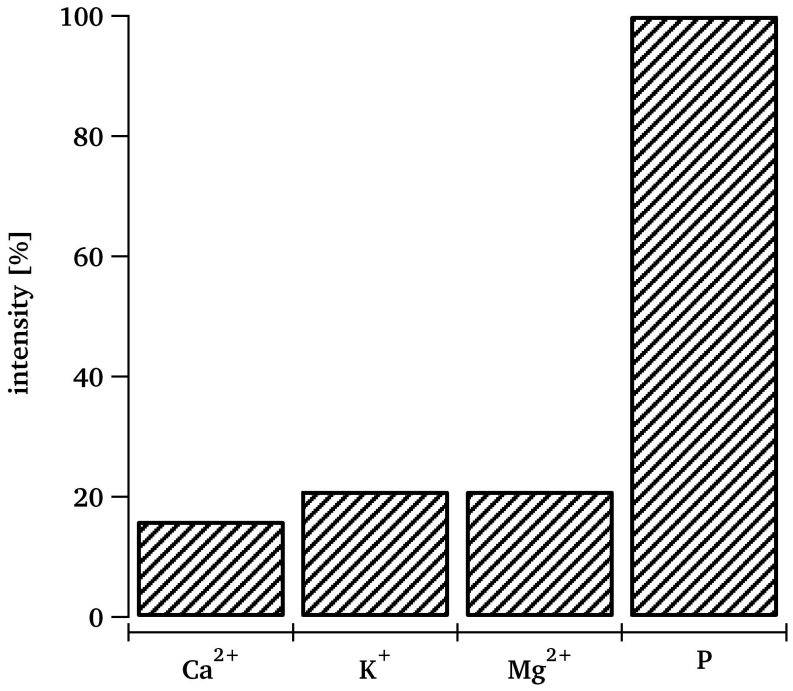

Figure 3.4.2.1: Molecular concentration of common cations and phosphor of PBCV-1.
The viral molecular composition referred to calcium, magnesium, potassium and phosphor is measured using the ICP-MS method. The data revealed that the amount of each cation in PBCV-1 is one fifth of the total phosphor concentration. Table: The total concentration of cations in PBCV-1is comparable to those in the reference (Tris-HCl). Ca^{2+}: calcium, Mg^{2+}: magnesium, K^{+}: potassium, P: phosphor

The most common DNA-neutralizing cation Mg^{2+} [15] is only present at trace amounts. Hence, Mg^{2+} can only compensate about 21 % of the phosphate. To sum up, it turns out that all cations can account for about 58 % of the charge compensation.

3.5. Conclusion

The data reveal that cations are even detectable in a quantitatively fashion but the concentration is not sufficient to accomplish the whole genomic condensation. Only one fifth of the total phosphor concentration can be neutralized by using Mg^{2+}, which is the most common DNA compensating cation. But taken all cations together, interestingly 58 % of the total phosphate can be neutralized.

Due to that fact it seems possible PBCV-1 should use another mechanism or established a mixture of DNA aggregation processes like *Escherichia coli* does.

3.6. References

[1] **Skrdla M. P., Burbank D. E., Xia Y., Meints R. H., Van Etten J. L., (1984)**: Structural proteins and lipids in a virus, PBCV-1, which replicates in a chlorella-like alga. *Virology* **135**, 308-15

[2] **Grayson P., Han L., Winther T., Philips R. (2007)**: Real-time observations of single bacteriophage λ DNA ejections in vitro. *Proc. Natl. Acad. Sci. USA* **104**, 14652-14657

[3] **Bloomfield V. A. (1998)**: DNA Condensation by Multivalent Cations. *Biopolymers/Nucleic Acid Sci.* **44**, 269-282

[4] **Kindt J., Tzlil S., Ben-Shaul A., Gelbart W. M. (2001)**: DNA packaging and ejection forces in bacteriophage. *Proc Natl Acad Sci USA* **98,** 13671–13674

[5] **Wellman S. E., Sittman D. B., Chaires J. B. (1994)**: Preferential binding of H1e histone to GC-rich DNA. *Biochemistry* **33***,* 384–388

[6] **Shampo M. A., Kyle, R. A. (1976)**: Albrecht Kossel. *JAMA* **235,** 1234

[7] **Ames B. N., Dubin D. T. (1960)**: The Role of Polyamines in the Neutralization of Bacteriophage Deoxyribonucleic. Acid. *J. Biol. Chem.* **235**, 769-775

[8] **Ross P. D., Shapiro J. T. (1974)**: Heat of interaction of DNA with polylysine, spermine, and Mg^{2+}. *Biopolymers.* **13**, 415-416

[9] **Van Etten J. L., Meints R. H., Burbank D. E., Kuczmarski D., Cuppels D. A., Lane L. C. (1981)**: Isolation and characterization of a virus from the intracellular green algae symbiotic with *Hydra viridis*. *Virology* **113**, 704–711

[10]] **Kaiser A., Vollmert M., Tholl D., Graves M. V., Gurnon J. R., Xing W., Lisec A. D., Nickerson K. W., Van Etten, J. L. (1999)**: Chlorella virus PBCV-1 encodes a functional homospermidine synthase. *Virology* **263**, 254–262

[11] **Li Z., Wu J., Wang Z. G. (2008)**: Osmotic Pressure and Packaging Structure of Caged DNA. *Biophys. J.* **94**, 37-746

[12] **Van Etten J. L., Burbank D. E., Xia Y., Meints R. H. (1983)**: Growth cycle of a virus, PBCV-1, that infects chlorella-like algae. *Virology* **126**, 117–125

[13] **Van Etten J. L., Burbank D. E., Kuczmarski D., Meints R. H. (1983)**: Virus Infection of Culturable *Chlorella*-like algae and Development of a Plaque assay. *Science* **219**, 994-996

[14] **Greiner T., Frohns F., Kang M., Van Etten J. L., Käsmann A., Moroni A., Hertel B., Thiel G. (2009)**: Chlorella viruses prevent multiple infections by depolarizing the host membrane. *Journal of General Virology* **90**, 2033–2039

[15] **Bloomfield V. A. (1996)**: DNA condensation. *Current Opinion in Structural Biology* **6**, 334-341

4. Chapter 3 – Structural Organization of DNA in *Chlorella* Viruses

4.1. Abstract

Chlorella viruses have icosahedral capsids with an internal membrane enclosing their large dsDNA genomes and associated proteins. Their genomes are packaged in the particles with a predicted DNA density of ca. 0.2 bp/nm^3. Occasionally, infection of an algal cell by an individual particle fails and the viral DNA is dynamically ejected from the capsid. This shows that the release of the DNA generates a force, which can aid in the transfer of the genome into the host in a successful infection. Imaging of ejected viral DNA indicates that it is intimately associated with proteins in a periodic fashion. The bulk of the protein particles detected by atomic force microscopy have a size of ~ 60 kDa and two proteins (A278L and A282L) of about this size are among 6 basic putative DNA-binding proteins found in a proteomic analysis of DNA-binding proteins packaged in the virion. A combination of fluorescence images of ejected DNA and a bioinformatics analysis of the DNA reveal periodic patterns in the viral DNA. The periodic distribution of GC rich regions in the genome provides potential binding sites for basic proteins. This DNA/protein aggregation could be responsible for the periodic concentration of fluorescently labeled DNA observed in ejected viral DNA. Collectively the data indicate that the large *Chlorella* viruses have a DNA packaging strategy that differs from bacteriophages; it involves proteins and share similarities to that of chromatin structure in eukaryotes.

4.2. Introduction

Chloroviruses in the family *Phycodnaviridae* have a long evolutionary history, possibly dating back to the time when eukaryotes arose from prokaryotes [1-3]. They are predicted to have a common ancestor with the poxviruses (e.g. vaccinia virus), asfarvirues, iridoviruses, ascoviruses and mimiviruses [1, 4, 5]. Collectively, these viruses are referred to as nucleocytoplasmic large DNA viruses (NCLDVs). PBCV-1 virions, the prototype chlorovirus, are large icosahedral particles (190 nm in external diameter) that have an internal lipid bilayered membrane [6]. However, the particles have more surface features than was originally thought.

One of the PBCV-1 vertices has a 560 Å long spike structure; 340 Å protrudes from the surface of the virus. The part of the spike structure that is outside the capsid has an external diameter of 35 Å at the tip expanding to 70 Å at the base [7, 8]. The spike structure widens to 160 Å inside the capsid and forms a closed cavity inside a large pocket between the capsid and the internal membrane enclosing the virus DNA. Therefore, the internal virus membrane departs from icosahedral symmetry adjacent to the unique vertex. Consequently, the virus DNA located inside the envelope is packaged non-uniformly in the virion. In addition to the spike, external fibers extend from some virus capsomers. PBCV-1 infection resembles infection by tailed bacteriophages because its genome must cross the cell wall (and membrane) of its host *C. variabilis* to initiate infection. The PBCV-1 spike first contacts the host cell wall [8] and the fibers aid in holding the virus to the wall. The spike is too thin to deliver DNA and, therefore, it probably serves to puncture the wall and is then jettisoned. Following expansion of the hole in the host wall by a virus-packaged enzyme, the viral internal membrane presumably fuses with the host membrane, facilitating entry of the PBCV-1 DNA and virion-associated proteins into the cell, leaving an empty capsid attached to the surface [9]. This fusion process triggers rapid depolarization of the host membrane [10], possibly by a virus encoded K^+ channel (named Kcv) predicted to be located in the internal membrane of the virus, followed by rapid release of K^+ from the cell [11] and altered secondary active transport of solutes [12]. The rapid loss of K^+ and associated water fluxes from the host reduce its turgor pressure, which may aid ejection of viral DNA and virion-associated proteins into the host [13]. A property that all NCLDVs, including the chloroviruses share with dsDNA bacteriophages and other DNA viruses, is that they package a dsDNA genome into a geometrically confined capsid. An example of DNA packaging is the 48.5 kb genome of bacteriophage λ. Its extended linear form of 16.5 μm [14] is compressed into a capsid with an inside radius of 27.5 nm [15], creating a DNA density inside the particle of 0.6 bp nm^{-3}. This value approaches the maximal theoretical density for DNA packaging and the DNA is almost at crystalline density inside the phage head [16]. A lower DNA packaging density occurs in the two NCLDVs, vaccinia virus and mimivirus. Both viruses package their large DNA genomes with a density of ~ 0.05 bp nm^{-3} [17, 18]. Our estimates of DNA packaging density place the chlorovirus PBCV-1 between phage λ and vaccinia virus and mimivirus. The ~ 330 kb genome of virus PBCV-1 is compressed into a capsid with an inner radius of about 72 nm providing a DNA density of ~ 0.2 bp nm^{-3}. DNA packaging density has implications for virus infection.

Experiments and theoretical calculations indicate that the high DNA packaging density in phages generates enormous internal pressure in the particles, ranging up to 50 bars [19]. This pressure serves as a driving force for the rapid ejection of DNA from the virus particle. For example, phage λ expels its DNA with an initial velocity of 60 kbp/sec, which then decreases as the residual amount of DNA in the particle decreases [20]. As a result, the entire DNA can be propelled from the capsid in ~ 1.5 sec under optimal conditions [14]. In another example, phage T5 DNA is expelled in a stepwise fashion at a rate reaching 75 kpb/sec [21]. This pressure driven DNA ejection provides at least part of the energy required for the transfer of the DNAs into their hosts [22].

PBCV-1 is unique among the NCLDVs in that it uncoats its DNA at the cell surface and leaves an empty capsid on the outside of the cell wall, similar to many tailed bacteriophages. Consequently, PBCV-1 may use similar mechanical forces as phages to eject its genome into its host cell [13]. In contrast, most NCLDVs are not faced with a cell wall and they initiate infection by either an endocytotic or an envelope fusion mechanism with the host plasma membrane; they then uncoat inside the cell. Consequently, most NCLDVs do not require a high DNA density to initiate infection. In fact, when DNA is released from the vaccinia capsid it does not burst out but rather pours out like a thick fluid [17], suggesting that forced ejection of vaccinia DNA is not important for its infection.

Phage DNA packaging depends primarily on two parameters: the function of motor proteins and cations. Dense packaging of DNA requires that 90 % of its charge is neutralized [23]. Evidence for charge neutralization of densely stored DNA in phages existed more than 50 years ago. While phages typically use cations to neutralize their DNA, some phages use polyamines, such as putrescine and spermidine in addition to cations [24]. There is no evidence indicating basic proteins contribute much to neutralizing phage dsDNA genomes [25, 26]. Indeed the dense DNA packaging in many phages leaves little room for DNA-binding proteins [27]. As mentioned above, DNA packaging in NCLDVs faces similar challenges to those in bacteriophages. Indeed, many dsDNA viruses use proteins for DNA packaging. *Polyomaviridae* [28] and *Papillomaviridae* [29] for example, can functionally co-opt host histone proteins. Other dsDNA viruses (*Adenoviridae, Asfarviridae, Baculoviridae*) express small arginine rich protamine-like proteins with putative DNA condensation functions [30,31].

Currently, little information is available on the mode of DNA packaging in the large chloroviruses.

At least in the case of PBCV-1, and presumably other chloroviruses, DNA neutralization may also employ proteins because PBCV-1 virions contain 148 different viral-encoded proteins [46], some of which have been described as DNA-binding proteins [32]. This large number suggests that DNA-binding proteins play a role in the organization and packaging of chlorovirus DNA genomes. Here we report a procedure for releasing PBCV-1 DNA from the virus particle and analyze its structural properties.

4.3. Material and Methods

4.3.1. Materials

Chlorella NC64A (recently named *Chlorella variabilis* [41]) and virus PBCV-1 were grown and isolated as previously described. Viral DNA was isolated from particles by two procedures: i) hyper-infection and ii) by osmotic shock. In the first procedure, isolated PBCV-1 particles were incubated for 10 min in standard modified Basic Bolds medium (MBBM) [42] with 30 mM 4′,6-Diamidino-2-phenylindol (DAPI) to label its DNA. C. *variabilis* cells were inoculated in MBBM containing 30 μM DAPI with the fluorescently labelled PBCV-1 at a multiplicity of infection (m.o.i) of $\sim$ 100 [42]. This process leads to the dynamic release of DNA from a few virus capsids. For DNA release via osmotic shock, a PBCV-1 suspension (8 x 10^{10} PFU/ml) was incubated for 1 h in 0.5 M KCl solution and then rapidly transferred to 60 mM KCl. For microscopic imaging experiments, the DNA was then labelled by adding 30 μM DAPI to this solution. For other experiments, we transferred the particles to fresh mica for atomic force microscopy (AFM) imaging or kept the diluted solution for 1 h before separating the soluble proteins from the released DNA according to [38]. This procedure resulted in the viral DNA being concentrated in the pellet by a factor > 30. The pellet containing DNA and associated proteins was re-suspended in 10 μl of distilled water.

4.3.2. Fluorescence microscopy

PBCV-1 particles and ejected fluorescently-labelled DNA were imaged on a Zeiss Axioskop 40^{TM} epifluorescence microscope.

Samples were excited at 358 nm and fluorescence detected through a 461 nm filter. The images were recorded with a sensitive electron multiplying charged coupled device and a digital camera Andor LUCA^{TM}.

4.3.3. AFM imaging

Images of ejected DNA were obtained with an AFM (Asylum Research^{TM}, MFP-3D^{TM}) using a cantilever (AC160TS, Standard Si cantilever, Olympus) with a < 10 nm tip. Fifty μl of a solution containing PBCV-1 particles or viral DNA (10 ng/ml), which was released from the capsids by osmotic shock, were incubated for 5 min on a smooth mica surface. The preparation was washed twice with 1 ml distilled water and dried by high air pressure (1 min, 2 bar). For imaging, the tapping mode was used, which reduces sample exposure. The volume of particles was analyzed using the particle analysing tool Image Processing Software, Image Metrology A/S with standard settings.

4.3.4. Bioinformatic analysis of PBCV-1 genome

The experimental results indicated that one or more DNA-binding proteins were involved in organizing the viral DNA in nearly equidistant units; if so, the mechanism would most likely require periodic binding motifs in the DNA. To identify protein binding domains in PBCV-1 DNA, we analyzed its genome sequence for periodic signals using our own code [43, 44], BioPyhton [45] and NumPy for Fast-Fourier-Transformations (FFT). First, we scanned the genome for 6 and 8 bp motifs including two wildcard characters at most, which match any nucleotide. We computed the Hamming distance of all possible fragments (330742 fragments) to all possible, potential motifs (15361 and 311297 motifs, respectively). This produced a string of integers – representing the hamming distance - ranging from zero to N, where N is the motif length. Any periodic or semi-periodic structures in this data are revealed by Fourier analysis as peaks in the Fourier components. We restricted our analysis to small motifs because of computational constraints. Therefore, we could not detect signals for larger motifs directly, e.g., two binding motifs connected by a highly variable region. This prompted us to not look for one particular motif with the most pronounced peak in the Fourier spectrum. Instead, we counted how often a particular length scale appeared and averaged over all motifs.

This protocol has an additional advantage because it also reveals periodicities of much larger motifs; e.g., consider the simple case of two motifs separated by some variable genomic region. Separately, these two motifs would exhibit the same periodicity. However, if we analyze the data averaged over all motifs, we also uncover the "synchronization" of the motifs along the chain and therefore, the correlated periodicity. We also analyzed the motifs that contributed most, for their relative content of nucleotide types. The measurements were kindly performed by Kay Hamacher (TU-Darmstadt, Germany).

4.3.5. DNA bound protein

In a further analysis, we separated DNA bound proteins from the re-suspended pellet by SDS PAGE. Peptide map fingerprinting (PMF) of proteins from SDS-PAGE gel slices was performed using standard procedures, with treatment of the proteins with dithiothreitol and iodoacetamide followed by trypsin digestion. Peptides eluted from the gels were purified on ZipTip C18 columns (Millipore) and applied to a stainless steel target together with a-cyano-4-hydroxycinnamic acid as a matrix. The peptides were analyzed in a reflectron mode using a Shimadzu Biotech Axima Performance MALDI-TOF mass spectrometer. Calibration was done via nearest neighbor external standards, using 8 peptides (Sigma Aldrich) with m/z from 757.4 to 3657.9. Mass lists from the individual PMF spectra were submitted to an in-house Mascot Server PMF search engine using the NCBInr database. The taxonomy was limited to C. *variabilis* and PBCV-1 virus. Additional search parameters were set to monoisotopic mass, charge 1+, maximum of 1 missed cleavage, peptide tolerance of 0.3 m/z and $p < 0.05$. The root mean square (RMS) errors on the peptide mass matches ranged from 21 – 102 ppm. As a control, all searches were repeated using the decoy database generated by the Mascot Server software, using the same settings. In some cases, high energy CID MS/MS sequencing of the peptides was employed (using the same samples and instrumentation) to confirm protein identification. The measurements were kindly performed by Robert Shoeman (MPI für medizinische Forschung, Heidelberg, Germany).

4.3.6. Virus purification scheme for the proteomic study

Essentially, the virus was purified as previously described [42] with the following modifications.

Prior to sucrose density gradient separation, the virus-infected cell lysate was clarified by first incubating with 1 % (v/v) NP-40 detergent at room temperature for 1 – 2 hr with constant agitation, concentrated by centrifugation in a Beckman Type19 rotor at 17,000 rpm, 50 min, 4 °C. The pellet fraction was solubilized in virus storage buffer (VSB) (50 mM Tris-HCl pH 7.8), loaded onto a 10 – 40 % (w/v) linear sucrose density gradient made up in VSB and centrifuged in a Beckman SW28 rotor for 20 min at 20000 rpm at 4 °C. The virus band was identified by light scattering, removed from the gradient and pelleted. The resuspended virus was "washed" with 50 μg/ml proteinase K in VSB for 4 hr at 25 °C to disassociate and degrade contaminating proteins. The proteinase K treated virus was layered onto a 20 – 40 % linear gradient of iodixanol (OptiPrep^{TM}, Axis-Shield, Oslo, Norway) in VSB and centrifuged 20000 rpm in a Beckman SW28 rotor for 4 h at 25 °C for isopynic separation. The virus band was removed by side-puncture of the centrifugation tube, diluted approximately 10 fold with VSB, then concentrated by centrifugation in a Beckman Ti50.2 rotor at 27000 rpm for 3 hr at 4 °C. The pellet fraction was resuspended in VSB, then filter-sterilized.

4.3.7. Whole virion proteome

Essentially, virion proteins were evaluated as described in [12]. Gradient purified, protease-washed PBCV-1 particles were disrupted with 1 % SDS/5 mM dithiothreitol and the proteins were separated by SDS-PAGE. The gel was comprehensively evaluated for viral proteins by mass ion analyses of peptides eluted from trypsin-digested gel slices. MS/MS data were processed using Masslynx software to produce peak lists for database searches with MASCOT (Matrix Science). Database searches were done against the newly re-sequenced and annotated PBCV-1 genome (Gen-Bank accession number JF411744.1). Protein identifications were based on random probability scores with a minimum value of 25, the value for $p < 0.05$ confidence. Approximate, relative quantitation of the proteins was determined using the exponentially modified protein abundance index (emPAI) [40]. This method uses the number of observed peptides compared to the number of observable peptides giving a ratio that is directly proportional to relative abundance of the protein in the mixture when adjusted exponentially (emPAI= $10^{\text{Nobserved/Nobservable}}$-1). We assumed that the major capsid protein (A430L) is present in 1440 copies per virion for these calculations and other protein abundances were estimated from this value.

4.4. Results and Discussion

4.4.1. Ejection of PBCV-1 DNA from capsids

Phage DNA release is often triggered by an interaction between phage tails and host receptor. This interaction causes the DNA to rapidly expand and the DNA catapults out of the capsid [14]. The host receptor for virus PBCV-1 is unknown, although circumstantial evidence suggests it is carbohydrate [33]. Still DNA release from PBCV-1 can be achieved by infecting *C. variabilis* cells with a high m.o.i. (e.g. 100). Under these conditions it occurs as if the DNA is released from the virus particle but not able to enter the host; as a consequence the particle is dynamically catapulted away from the host cell leaving an unraveled, quasi-linear DNA polymer tethered to the cell. In some images it is possible to see the capsid at the end of the DNA thread projected away from the host (data not shown). The reason for the release of DNA into the medium is not known. However we know from other studies that usually only one virus infects the host cell, while the remaining viruses are excluded [34]. The fact that DNA release into the medium is only apparent at very high m.o.i., suggests that most viruses under these circumstances are not able to eject their DNA into the host because of a yet unknown exclusion mechanism. This release of DNA into the medium must be fast, because it is possible to detect isolated DNA molecules already within 5 min post infection. Figure 4.4.1.1 A shows a fluorescent image with a host cell and the unfolded DNA polymer from a virus particle. The DNA dye DAPI produces bright staining of the nucleus of the *Chlorell*a cell in the lower right part of the figure; in the upper left part of the figure the unraveled DNA from a virus capsid protrudes as a nearly linear structure from the cell. In the case of phages, it has been argued that the release of the genome can be explained on the basis of Brownian motion [35]. Such a process, however, cannot account for the observed DNA ejection from virus PBCV-1. Brownian movement would not generate the sort of straight lines and would also be much slower [22]. Hence, the present data stress the importance of an osmotic pressure in the dense environment of the capsid, which creates a driving force for DNA ejection. In this sense, PBCV-1 DNA ejection into the medium resembles that of many phage genomes [14, 21]. The forceful ejection of DNA from the PBCV-1 particle is consistent with the concept that PBCV-1 depends, at least to some extent, on these mechanical forces to eject its DNA against the turgor pressure of the *Chlorella* host cell.

Frequently, we observed two DNA strands under the same conditions, which projected away from the host at a common point of origin (Figure 4.4.1.1 C). This phenomenon was not only observed once, but in ~ 20 % of the ejected DNA molecules. Since the surface of a *Chlorella* cell is ca. 500 times larger than that of a virus, it is statistically rather unlikely that two virus particles independently infect at an m.o.i of 100 a host cell as frequently in the same spot. Hence, it is more likely that the two DNA polymers were not from separate viruses but from a single virus. This interpretation is supported by the electron microscopic images depicted in Figures 4.4.1.1 E, F. These images show a C. *variabilis* cell with the typical hole in the wall, which a virus digests in the course of infection. From this location, two linear structures project away from the host in an angular fashion. The projecting structures are most likely unfolded viral DNA because the half width of their cross section is < 10 nm. The combination of electron microscopic and fluorescent images suggests that PBCV-1 DNA might not in all cases enter its host initially by either of its termini. This scenario would suggest that packaging of the DNA in the virion differs from ejection because it is unlikely that DNA packaging begins in the middle of the genome.

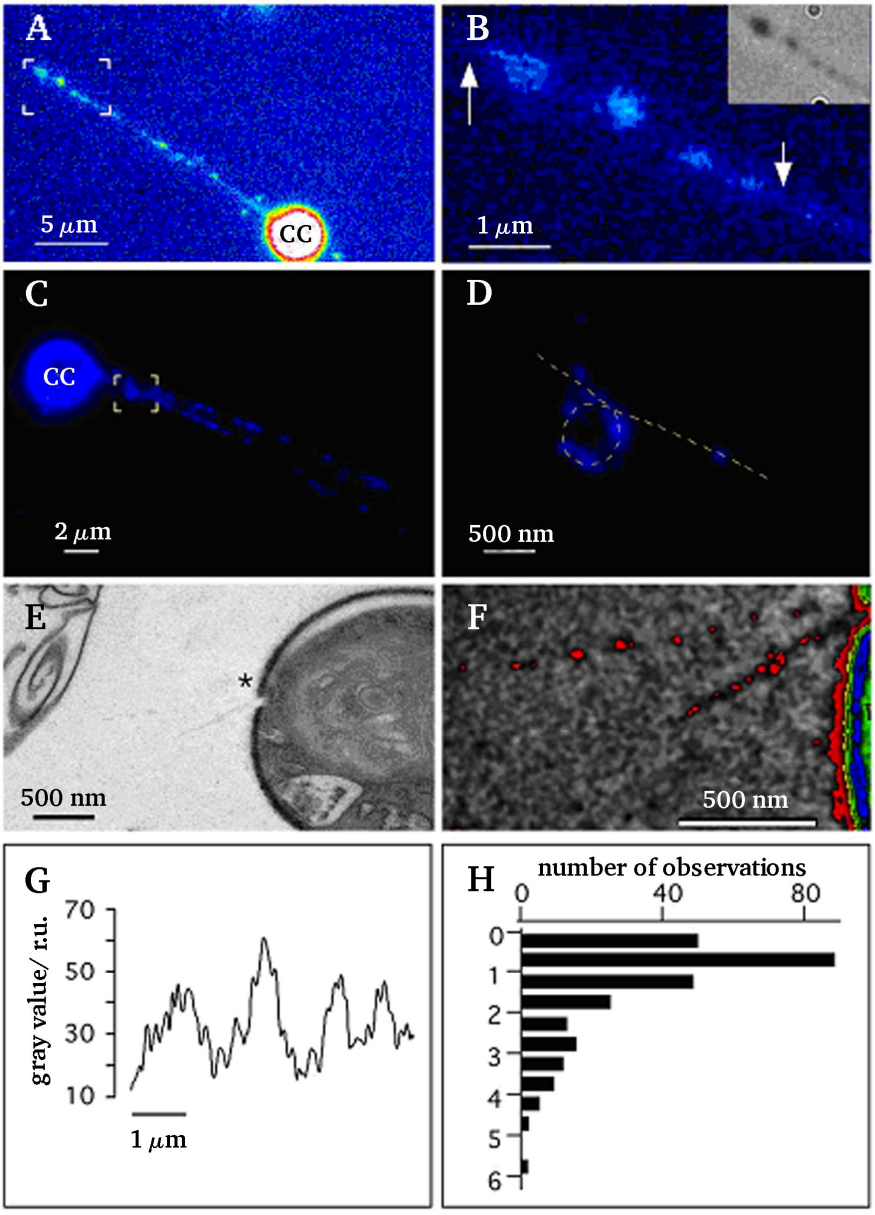

Figure 4.4.1.1: Ejection of viral DNA.

A: Fluorescence images of *C. variabilis* with ejected DNA molecules. The incubation medium contained *C. variabilis* cells and virus PBCV-1 at an m.o.i. of ~100 plus the fluorescent DNA stain DAPI. The image shows a *Chlorella* cell (cc) and the viral DNA molecule, which is propelled away from the alga cell. B: Magnification of the area indicated by the box in A. Inset: same area as in B with conventional light microscopy and phase contrast. C: same as in A but with two DNA bands projecting away from a *Chlorella* cell (cc). D: Magnification of area indicated by box in C with loop like DNA structure. E: Electron micrograph of viral DNA projecting away from host cell wall. The cell wall of the alga exhibits the typical hole (*), which the viruses digest for infection. From this hole two linear structures project towards the left side. The part marked in E is magnified in F and presented in artificial colors in order to highlight the linear structures projecting away from the cell wall hole. G: fluorescence intensity profile along DNA molecule between arrows in B. H: Histogram of distances between individual fluorescence maxima as in E from 30 ejected DNA molecules [49].

4.4.2. Virus PBCV-1 DNA has structure

Close scrutiny of the fluorescently labeled ejected PBCV-1 DNA suggests that it is structured. The images indicate: i) a non-uniform distribution of DNA and ii) loops in the DNA polymer. Images of ejected DNA at higher magnifications indicate that the fluorescence associated with the DNA exhibits distinct maxima. Figure 4.4.1.1 B shows part of an enlarged fluorescent DNA band from Figure 4.4.1.1 A. The fluorescence signal alternates between high and low fluorescence intensity along an imaginary line (Figure 4.4.1.1 B). The locations of the intensity maxima coincide with structures, which occasionally can be seen with phase contrast in a light microscope (inset Figure 4.4.1.1 B). This observation implies that the non-uniform fluorescence of DAPI staining is not caused by a preference of the dye to interact with A-T rich regions in the DNA but instead the intensity maxima are due to local concentrations of DNA. In regions where the fluorescence signal was well resolved the intensity maxima were quasi periodic (Figure 4.4.1.1 G). In this example, the maxima were ca. 1.2 μm apart. The distances between fluorescence peaks measured in 30 individual ejected DNA molecules are summarized in Figure 4.4.1.1 H. The histogram shows a broad distribution of gap sizes. The distribution has a maximum below 1 μm and possibly a second one below 3 μm (Figure 4.4.1.1 H). The structured pattern seen with PBCV-1 DNA does not occur in fully ejected DNA from phages l or T5; the phage DNAs fluoresce homogeneously along the axis of the extended polymer [14, 21]. Collectively, these results suggest that the distribution of fluorescent maxima is an inherent property of PBCV-1 DNA structure. Hence PBCV-1 DNA molecules in the capsid are apparently organized differently than those in the two phages. The PBCV-1 DNA is most likely concentrated in a periodic fashion in submicroscopic coils. In addition to these periodic domains of concentrated DNA, we occasionally detected another higher order organization in the PBCV-1 DNA strands (Figures 4.4.1.1 C, D); in this case, the polymer has a large loop on which several concentrated domains are clustered. The diameter of this loop is approximately 600 nm; similar loops with a mean diameter of 550 $\pm$ 100 were detected in four other images. Collectively, the data suggest that PBCV-1 DNA is stored inside the capsid in an ordered fashion. This structure includes a periodic formation of folds and on a larger scale a formation of loop structures. These loop structures probably open up during the ejection and are hence observed only in rare cases.

4.4.3. Examination of PBCV-1 DNA by AFM

To obtain more information on PBCV-1 DNA structure we viewed the isolated DNA by AFM. For these experiments, PBCV-1 particles were osmotically shocked and subsequently transferred with the emerging content onto fresh mica for imaging. This procedure separates proteins that are tightly bound to DNA from those that are free. Figure 4.4.4.1 A shows a typical ruptured particle and its associated content. Higher magnification shows the emerging DNA together with numerous particles (Figure 4.4.4.1 B, C). These particles are not observed with pure plasmid DNA (results not shown); hence, they are specific to the virus preparation. The particles contain proteins because the particles disappear in the AFM images after treating the preparation for 1 min at 37 °C with protease K (1 mg/ml) (Figure 4.4.4.1 D). This interpretation is supported by the fact that PBCV-1 virions contain many basic proteins (see below). Association of DNA with proteins was reported in a previous AFM study on ruptured PBCV-1 particles [36]. The association of proteins with the DNA is not random. For example, in the Figure 4.4.4.1 C image DNA occupies, 5 % of the total image area but 50 % of the proteins are directly associated with the DNA molecule. A similar bias of DNA and proteins occurred in other images analyzed in the same manner. The intimate association between DNA and proteins is also supported by force measurements. When a protein particle was pulled from the surface with the cantilever, the force was about 100 times higher for particles associated with DNA than for free particles. A comparison between images of PBCV-1 DNA shows that proteins associated with DNA in AFM images produce a different periodic pattern than the fluorescence images. The reason for this difference in DNA/protein association is probably related to the isolation method. The image in Figure 4.4.4.1 B suggests that the osmotic shock method results in a more violent and unorganized release of DNA from the virus capsid than does the ejection method. Consequently, the osmotic forces probably disrupt the more delicate DNA structure. This is consistent with the observation that the typical periodic structure of DAPI-labeled DNA, as in Figure 4.4.1.1, disappears when the DNA is released by an osmotic shock as in Figure 4.4.4.1 (data not shown). To estimate the size of the proteins associated with PBCV-1 DNA, we imaged the volume of a large number of DNA associated protein particles (Figure 4.4.4.1 C, F); this is possible because molecular volumes correlate with the molecular mass of proteins [37, 48]. This aspect was examined in previous experiments with different proteins. Concluding, the higher the molecular mass the higher the volume (data not shown).

Measurements were first made on the 66 kDa bovine serum albumin (BSA) protein (Figure 4.4.4.1 E) and a recombinantly expressed 69 kDa putative PBCV-1 DNA-binding protein (CDS A278L) to calibrate the system (Figure 4.4.4.1 G). The estimated volumes of the BSA and A278L proteins were 57.6 ± 0.13 nm^3 (435 measurements) and 66.4 ± 0.09 nm^3 (652 measurements), respectively. Measurements of 711 randomly chosen protein particles associated with the disrupted virus produced a mean volume of 60.3 ± 0.09 nm^3. These experiments suggest that the PBCV-1 DNA particles are associated with a protein(s) in the range of ca. 60 kDa. The resolution of the images does not allow one to distinguish between monomers of a 60 kDa protein or multimers of smaller proteins. To estimate the ratio of proteins that are associated with DNA from the virus particles, we measured the length of the DNA molecule versus the number of total proteins in 5 images (e.g. Figure 4.4.4.1 B). For this analysis we considered all spherical particles that exceeded background noise by a factor .2 irrespective of whether they were free or associated with DNA. This analysis produced a number of 0.018 ± 0.005 proteins per nm of DNA, which translates into one protein per 55 nm of DNA. This number implies that the entire DNA from a PBCV-1 virion, which is 100 μm long in its extended form, should be associated with > 2000 proteins. This is an absolute minimum estimate because some proteins are probably lost during the preparation of the DNA and small proteins are masked by background noise.

4.4.4. PBCV-1 DNA-binding proteins

Taken together, the results from the two preceding sets of experiments suggest that PBCV-1 DNA is packaged differently than phages and it is probably associated with proteins. Even though PBCV-1 does not contain canonical histones, an association of DNA with proteins is supported by the analysis of the proteins packaged in a related chlorovirus CVK2 [32]. This study reported that the virions contain 7 DNA-binding proteins of which 3, with estimated molecular weights of 63, 42 and 25 kDa, had high affinity for the viral DNA. To extend this previous study to PBCV 1, viral DNA was released by osmotic shock and the DNA with associated proteins was separated from soluble proteins by centrifugation according to [38]. This procedure led to about a 30-fold concentration of DNA in the pellet fraction. The pellet with the DNA and DNA-associated proteins was re-suspended in buffer containing DNAse, incubated for 1 h, and the samples were electrophoresed on SDS-polyacrylamide gels. Several prominent protein bands were detected (Figure 4.4.4.2).

The 6 most prominent protein bands were excised from the gel and subjected to PMF using trypsin. The peptides were analyzed by MALDI-TOF and matches to PBCV-1 proteins were identified with Mascot Server software.

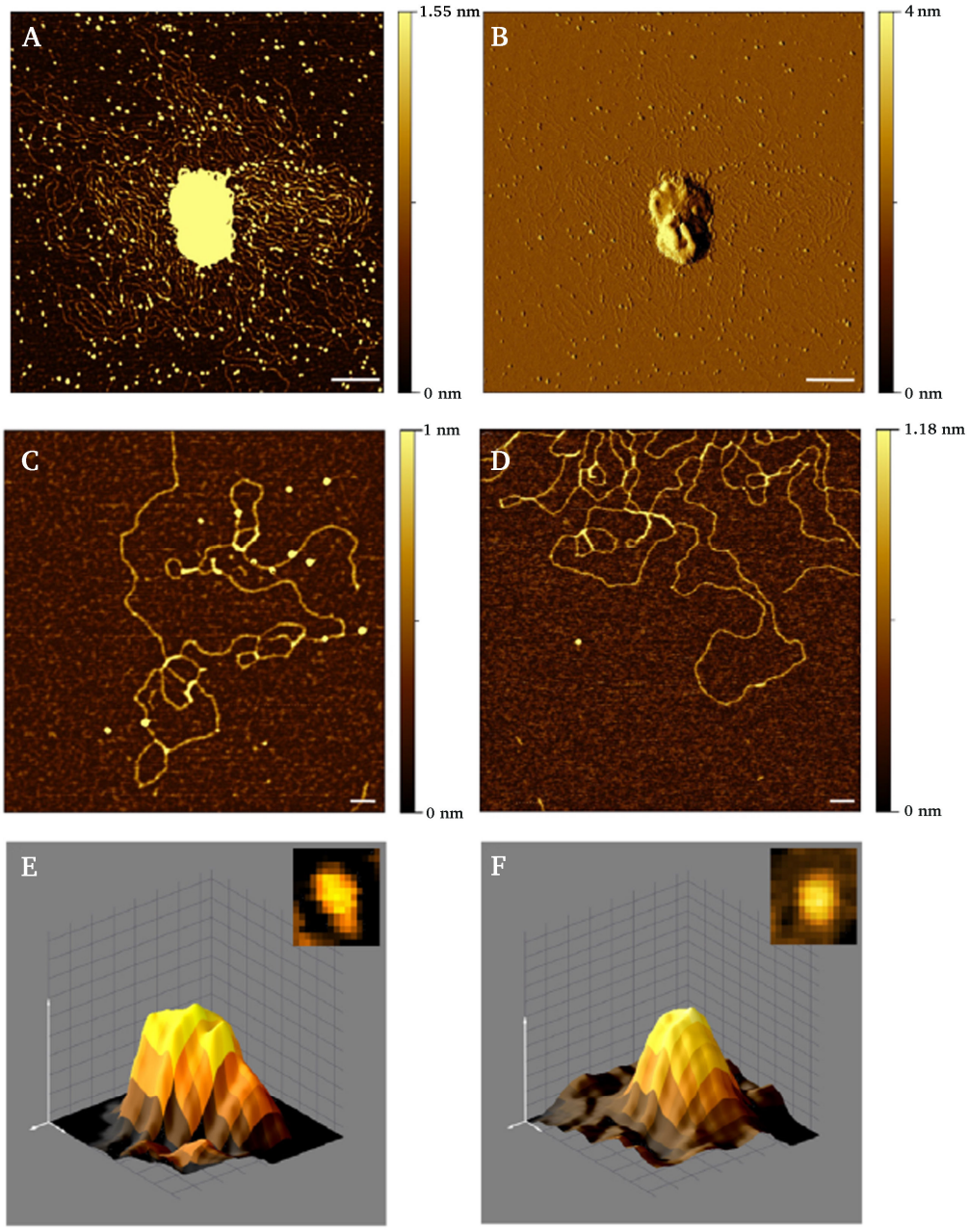

Figure 4.4.4.1: AFM images of viral DNA and associated proteins, which were isolated by osmotic shock. Scan of single PBCV-1 particles after osmotic shock in a height image A and in amplitude image B. The images reveal emerging DNA and protein particles from the disrupted virus. Magnification of DNA from disrupted virus with protein particles C. Proteins are absent after the sample was treated with proteinase K D. 3 dimensional image of individual BSA protein E and of individual purified 70 kDa PBCV-1 protein A278L. The latter is a putative DNA-binding protein coded by virus PBCV-1. Scale bars 100 nm in A – D and 2 nm in E and F [49].

All database searches were also performed against the Mascot Server automatic decoy database. The latter generates a random set of sequences with the same amino acid composition as the authentic database entries. The searches with the decoy database were negative (no significant matches), except for the search with the peptide mass list that identified CDS A523R. In this search, A523R (19 kDa) was the single match with a protein score of 122 (cutoff of 43) (expect value of 6.7 e^{-10}) and one match in the decoy database occurred with a protein score of 48 (cutoff of 43). From the spectra, 9 viral proteins were identified with a significance of $p < 0.05$ (Table 4.4.4.1). Among these nine proteins, there are two abundant PBCV-1 capsid proteins (CDSs A430L and A140/145R), which almost always occur as contaminates when fractionating PBCV-1 proteins; they were eliminated as important DNA-binding candidates.

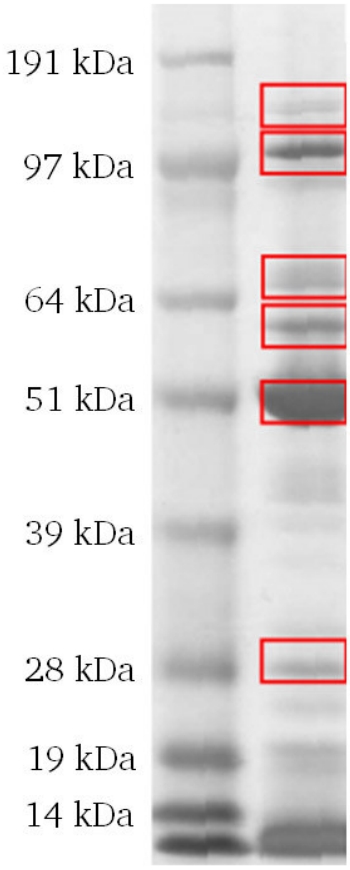

Figure 4.4.4.2: SDS-PAGE pattern of proteins associated with PBCV- 1 DNA.

DNA was released from capsids by osmotic shock and separated from soluble proteins by centrifugation. The DNA-containing pellet was treated with DNase to release DNA-bound proteins. The framed bands were excised and used for MALDI TOF analysis [49]. Lane 1: weight marker, molecular weight in kDa, lane 2: proteins obtained after DNAse treatment.

Of the remaining 7 proteins, 6 have high isoelectric points, which are well suited for binding and neutralizing DNA. A bioinformatics prediction of DNA-binding sites in these proteins revealed several putative interactive sites between the proteins and DNA (Table 4.4.4.1). These experimental results are consistent with the finding that the PBCV-1 virion contains many basic proteins which are suitable for binding/neutralizing the viral DNA and which are associated with the DNA, even after isolation. Of these 6 possible PBCV-1 DNA-binding proteins, two have molecular weights of 70 kDa [A278L (69 kDa) and A282L (63 kDa)]. In addition, a dimer of A284L (60 kDa) and a trimer or tetramer of A523R (57 or 78 kDa) could produce a protein of 70 kDa. It is interesting, that both A278L and A282L have protein kinase activity [47]. Also, one of the virus DNA-binding proteins identified by [32] had protein kinase activity. This putative dual function of the DNA-binding proteins is interesting in the context of small proteins from *Baculoviridae*, which bind and dissociate from the viral DNA depending on their phosphorylation state [39]. The inherent kinase activity of the putative DNA-binding proteins in the chloroviruses could have a similar regulatory importance for DNA condensation. In a separate experiment, we determined the proteome of the entire virus particle [46] and estimated the most abundant proteins in the particle using the exponentially modified protein abundance index (emPAI) algorithm [40]. Based on the assumption that the major capsid protein (A430L) is present in 1440 copies per virion, we estimated the abundances of the major proteins in the virion (Table 4.4.4.1). The data show that some of the proteins are present with copy numbers in the range of several hundreds up to 2000. Interesting to note is, that some of the most abundant proteins again are very basic; some of these abundant proteins were also detected in association with the DNA, including A278L, A282L, A284L and A523R (Table 4.4.4.1). In addition, one host encoded histone-like protein was detected in the virion that appears to be present in low amounts.

Table 4.4.4.1: Identification of DNA-binding proteins and most abundant proteins in PBCV-1 virion from MALDI-TOF PMF analysis.

ORFs	EmPAI Exponentially-modified protein abundance index	Copy number per virion	MW/kDa	% DNA binding	IP/pH	annotation
A548L	0.06	7	57.4	19	9.5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5
A189/192R	0.5	55	143.6	39	11.4	Chromosome remodelling complex
A561L	1.44	159	71	n.d.	9.9	unknown
A363R	2.01	222	128.4	36	10.9	Similar to D6/D11-like helicase
A284L	2.41	266	30.8	20	9.2	Choloylglycine hydrolase
A035L	2.48	273	65.6	n.d.	8.9	unknown
A565R	2.49	275	73.2	n.d.	7.3	unknown
A140/145R	3	331	120.9	35	11	Located at the unique vertex of the virus particle
A278L	3.6	397	69.2	36	10.8	Protein kinase domain/PBCV-1 specific basic adaptordomain
A282L	4.83	533	63.4	35	10.8	Protein kinase domain/PBCV-1 specific basic adaptor domain
A430L	13.06	1440	48.2	48	7.5	Large eukaryotic DNA virus major capsid protein
A437L	19.75	2178	10.9	n.d.	11	Non-histone chromosomal protein MC1

The table lists (in bold) the PBCV-1 CDSs from which peptides were detected in MALDI-TOF spectra. The peptides correspond to DNA associated proteins described in Figure 4.4.4.2. Only database entries with a protein score with a significance $p < 0.05$ are presented. The identity of the A278L protein was confirmed by the MS/MS sequencing of 3 peptides. Also included are the copy numbers per virion of the most abundant proteins in PBCV-1 virion. The copy numbers were estimated using the exponentially modified protein abundance index (emPAI) algorithm [40]. Based on the knowledge that the major capsid protein (A430L) is present in 1440 copies per virion, we calculated the abundances of the major proteins in the virion. For these proteins, parameters such as the molecular weight (MW), the isoelectric point (IP) and the functional annotation from the Greengene database) are provided. Putative DNA-binding sites were further identified by the BindN algorithm. The data are presented as percentage of putative DNA-binding sites relative to total protein (% DNA-binding) [49].

4.4.5. Identification of potential protein binding domains in PBCV-1 DNA

The periodic pattern of isolated DNA bands reported in Figure 4.4.1.1 prompted us to develop a Fast-Fourier-Transformation (FFT) protocol in order to identify potential periodicities of binding motifs in the PBCV-1 sequence. We chose the Hamming distance with respect to a given motif and averaged over all motifs of a given length. This procedure revealed Fourier components spanning lengths of 9935 bp, 2138 bp, and 17020 bp that were the three major contributors to the Fourier series expansion of the Hamming distance data set (Figure 4.4.5.1). These periodic bp patterns translate into distances of 3.2 μm, 0.7 μm and 5.4 μm. Hence, the PBCV-1 genome exhibits a pattern, which roughly resembles the distribution of fluorescence maxima along the isolated DNA polymer.

The most frequent spanning lengths of 2138 bp and 9935 bp coincides with the two maxima in the distribution of the gaps between fluorescent maxima below 1 and 3 mm (Figure 4.4.5.1). The fact that the measured amplitudes at corresponding frequencies between fluorescent maxima does not correlate with the calculated frequency of spanning length distances between putative binding motives in the genome, may suggest that the different motives have different binding affinities or binding specificity for DNA-binding proteins (Figure 4.4.5.1). According to the distribution of Fourier-amplitudes (Bioinformatics Analysis S1), the most pronounced sequence motifs are significantly distant from a random, average sequence motif. We note that the results were similar using six and eight bp-size motifs, thus robust under motif length change. This protocol has an additional advantage because it can conceptually reveal periodicities of much larger motifs; e.g., consider the simple case of two motifs separated by some variable genomic region. Separately, these two motifs would each exhibit the same periodicity. If we analyze the data averaged over all motifs, however, we also reveal the “synchronization” of the motifs along the chain and therefore the correlated periodicity. To verify our results and to demonstrate the robustness of our protocol, we repeated the analysis using a strict criterion: an exact match between the motif without wildcard characters and the genomic fragment. As mentioned in the Materials and Methods section, this method is not as sensitive with smaller motifs; therefore, we restricted our motif length to six bps. We then repeated the FFT on this data set using the strict criterion. The experiment produced conclusive results with inverse frequencies of 9571 bp, 17300 bp and 16801 bp; in addition, we also obtained a signal at 6835 bp. Details on the procedure and on the most important motifs for the Hamming distance FFT are listed in Bioinformatics Analysis S1. Interestingly, all possible CG-combinations contributed to a large extent, while those with A or T nucleotides were less significant, that is periodic. The data also show that there is a large ratio of the FFT-coefficients, thus indicating the significant different pattern of periodicity of CG- vs. “some-AT”-regions.

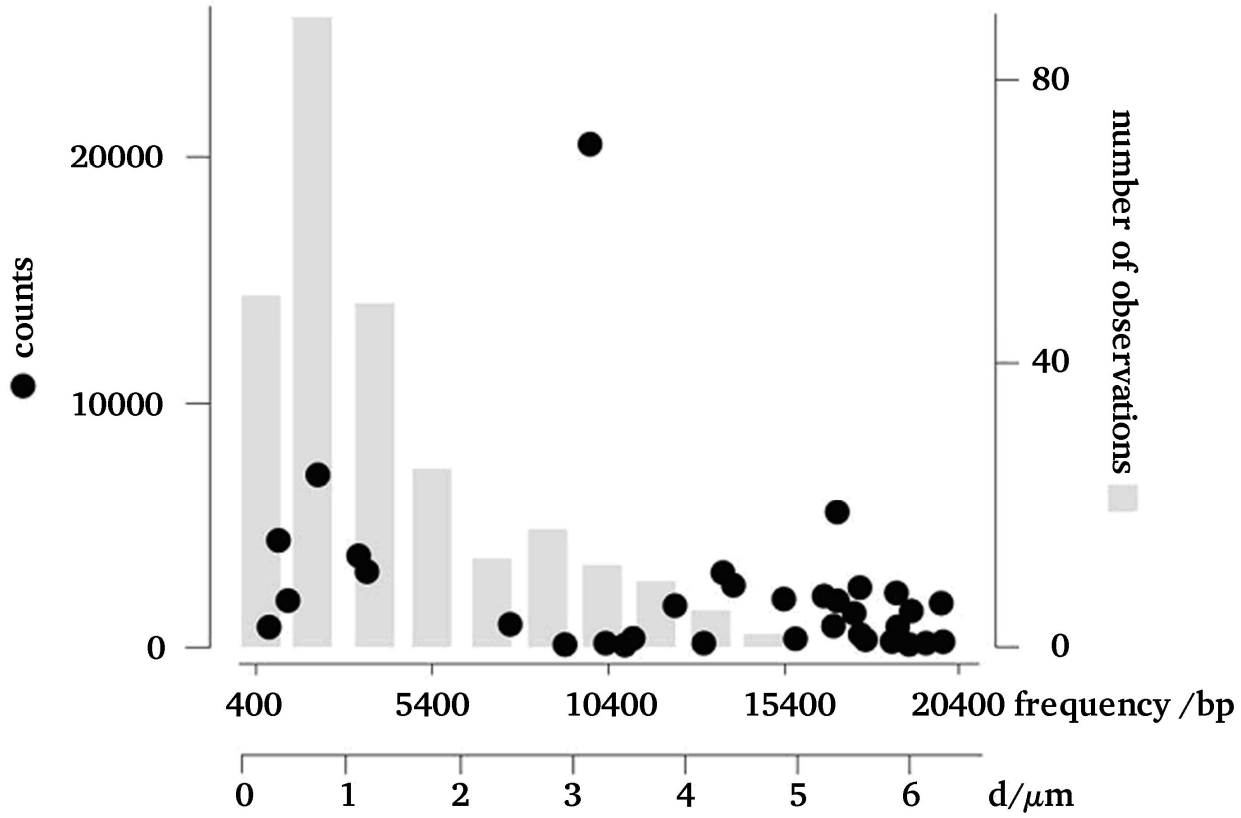

Figure 4.4.5.1: Histogram of the leading inverse Fourier-frequency (in bp) for all sequence motifs of length eight (round symbols). The most pronounced peaks occur at 9935 bp, 2138 bp, and 17020 bp. For comparison we also illustrate the distribution (grey bars) of distances between individual fluorescence maxima from Figure 4.4.1.1 H. We assume that 1 kb DNA is in the extended form 0.323 μm long [49].

4.5. Conclusion

The present results suggest that chlorovirus PBCV-1, like their eukaryotic hosts, neutralize their DNA with DNA-binding proteins. The data are consistent with a model in which the viral genome has an inherent pattern with periodically spaced GC rich regions, which provide interactive sites for DNA-binding proteins. DNA is presumably wound around the respective basic proteins for neutralization and packaging. The interaction results in the isolated DNA containing periodic thickenings; a higher order of organization may also involve small loops, which contribute to packaging DNA and/or gene regulation. The robust aggregation of the DNA with proteins also favors a stable structure for the virus DNA when it is ejected into the cytosol of the host, where the cation concentration is reduced [11].

The current results do not distinguish if this organization of DNA is contributing to the neutralization of the entire DNA polymer or if it is just a component of its meta-organization with the goal to achieve a crystalline-like or ordered structure inside the virion. The data however, indicate that the *Chlorella* viruses and possibly other large DNA viruses have developed a DNA packaging strategy, which involves proteins and hence shares similarities to that of chromatin in eukaryotes.

4.6. References

[1] **Iyer L. M., Balaji S., Koonin E. V., Aravind L. (2006)**: Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res* **117**, 156–184

[2] **Raoult D., Audic S., Robert C., Abergel C., Renesto P. (2004)**: The 1.2- megabase genome sequence of mimivirus. *Science* **306**, 1344–1350

[3] **Villarreal L. P., DeFilippis V. R. (2000)**: A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J Virol* **74**, 7079–7084

[4] **Iyer L. M., Aravind L., Koonin E. V. (2001)**: Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol* **75**, 11720–11734

[5] **Koonin E. V., Yutin N. (2010)**: Origin and evolution of eukaryotic large nucleocytoplasmic DNA viruses. *Intervirology* **53**, 284–292

[6] **Yan X., Olson N. H., Van Etten J. L., Bergoin M., Rossmann M. G., et al. (2000)**: Structure and assembly of large lipid-containing dsDNA viruses. *Nat Struct Biol* **7**, 101–103

[7] **Cherrier M. V., Kostyuchenko V. A., Xiao C., Bowman V. D., Battisti A. J. (2009)**: An icosahedral alga virus has a complex unique vertex decorated by a spike. *Proc Natl Sci USA* **106,** 11085–11089

[8] **Zhang X., Xiang Y., Dunigan D. D., Klose T., Chipman P. R., et al. (2011)**: The structure and function of the Paramecium bursaria chlorella virus capsid. *Proc Natl Acad Sci USA* **108**, 14837–14842

[9] **Meints R. H., Lee K., Burbank D. E., Van Etten J. L. (1984)**: Infection of a Chlorellalike alga with the virus PBCV-1: ultrastructural studies. Virology 138: 341–346

[10] **Frohns F., Käsmann A., Kramer D., Schäfer B., Mehmel M. (2006)**: Potassium ion channels of chlorella viruses cause rapid depolarization of host cells during infection. *J Virol* **80,** 2437–2444

[11] **Neupärtl M., Meyer C., Woll I., Frohns F., Kang M., (2008)**: Chlorella viruses evoke a rapid release of K+ from host cells during early phase of infection. *Virology* **372**, 340–348

[12] **Agarkova I., Dunigan D. D., Gurnon J., Greiner T., Barres J., et al. (2008)**: Chlorovirus-mediated membrane depolarization of chlorella alters secondary active transport of solutes. *J Virol* **82**, 12181–12190

[13] **Thiel G., Moroni A., Dunigan D., Van Etten J. L. (2010)**: Initial Events Associated with Virus PBCV-1 Infection of Chlorella NC64A. *Prog Bot* **71**, 169–183

[14] **Grayson P., Han L., Winther T., Philips R. (2007)**: Real-time observations of single bacteriophage lambda DNA ejection in vitro. *Proc Natl Acad Sci USA* **104**, 14652–14657

[15] **Cordova A., Deserno M., Gelbart W. M., Ben-Shaul A. (2003)**: Osmotic shock and the strength of viral capsids. *Biophys J* **85** 70–74

[16] **Earnshaw W. C., Casjens S. R. (1980)**: DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**, 319–331

[17] **Kuznetsov Y., Gershon P. D., McPherson A. (2008)**: Atomic force microscopy investigation of vaccinia virus structure. *J Virol* **82**, 7551–7566

[18] **Kuznetsov Y. G., Xiao C., Sun S., Raoult D., Rossmann M. (2010)**: Atomic force microscopy investigation of the giant mimivirus. *Virology* **404**, 127–137

[19] **Li Z., Wu J., Wang Z. G. (2008)**: Osmotic pressure and packing structure of caged DNA. *Biophys J* **94**, 737–746

[20] **Kindt J., Tzlil S., Ben-Shaul A., Gelbart W. M. (2001)**: DNA packaging and ejection forces in bacteriophage. *Proc Natl Acad Sci USA* **98,** 13671–13674

[21] **Mangenot S., Hochrein M., Räder J., Letellier L. (2005)**: Real-time imaging of DNA ejection from single phage particles. *Curr Biol* **15**, 430–435

[22] **Grayson P., Molineux I. J. (2007)**: Is phage DNA 'injected' into cells - biologists and physicists can agree. *Current Opin Microbiol* **10**, 401–409

[23] **Bloomfield V.A. (1997)**: DNA condensation by multivalent cations. *Biopolymers* **44**, 269–282.

[24] **Ames B. N., Dubin D. T. (1960)**: The role of polyamines in the neutralization of bacteriophage deoxyribonucleic acid. *J Biol Chem* **235**, 769–775

[25] **Hass R., Murphy R. F., Cantor C. R. (1982)**: Testing models of the arrangement of DNA inside bacteriophage lamda by crosslinking the packaged DNA. *J Mol Biol* **159**, 71–92

[26] **Serwer P. (1986)**: Arrangement of double-stranded DNA packaged in bacteriophage capsids. An alternative model. *J Mol Biol* **190**, 509–512

[27] **Casjens S. R. (2011)**: The DNA-packaging nanomotor of tailed bacteriophages. *Nature Reviews* **9**, 647–657

[28] **Tan K. B. (1977)**: Histones: metabolism in simian virus 40-infected cells and incorporation into virions. *Proc. Natl Acad Sci USA* **74**, 2805–2809

[29] **Pfister H., zur Hausen H. (1978)**: Characterization of proteins of human papilloma viruses (HPV) and antibody response to HPV 1. *Med. Microbiol Immunol* **166**, 13–19

[30] **Tweeten K. A., Bulla L. A., Consigli R. A. (1980)**: Characterization of an extremely basic protein derived from granulosis virus nucleocapsids. *J Virol* **33**, 866–876

[31] **Wang M., Tuladhar E., Shen S., Wang H., van Oers M. M., et al. (2010)**: Specificity of Baculovirus P6.9 basic DNA-binding proteins and critical role oft he c terminus in virion formation. *J Virol* **84,** 8821–8828

[32] **Yamada T., Furukawa S., Hamazaki T., Songsri P. (1996)**: Characterization of DNA-binding proteins and protein kinase activities in Chlorella virus CVK2. *Virology* **219**, 395–406

[33] **Meints R. H., Burbank D. E., Van Etten J. L., Lamport D. T. (1988)**: Properties of the Chlorella receptor for the virus PBCV-1. *Virology* **164**, 15–21

[34] **Greiner T., Frohns F., Kang M., Van Etten J. L., Käsmann A., et al. (2009)**: Chlorella viruses prevent multiple infections by depolarizing the host membranes. *J Gen Virol* **90**, 2033–2039

[35] **Ore A., Pollard E. (1956)**: Physical mechanism of bacteriophage injection. *Science* **124,** 430–432

[36] **Kuznetsov Y. G., Gurnon J. R., Van Etten J. L., McPherson A. (2005)**: Atomic force microscopy investigation of a chlorella virus, PBCV-1. *J Struc Biol* **149**, 256–263

[37] **Schneider S. W., Lärmer J, Henderson R. M., Oberleithner H. (1998)**: Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy. *Pflügers Arch - Eur J Physiol* **435,** 362–367

[38] **Cavalcanti M. C., Rizgalla M., Geyer J., Failing K., Litzke L. F., et al. (2009)**: Expression of histone 1 (H1) and testis-specific histone 1 (H1t) genes during stallion spermatogenesis. *Animal Reprod Sci* **111**, 220–234

[39] **Funk C. J, Consigli R. A. (1993)**: Phosphate cycling on the basic protein of Plodia interpunctella granulosis virus. *Virology* **193**, 396–402

[40] **Ishihama Y., Oda Y., Tabata T., Sato T., Nagasu T., et al. (2005)**: Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol & Cell Proteomics* **4**, 1265–1272

[41] **Hoshina R., Iwataki M., Imamura N. (2010)**: Chlorella variabilis and Micractinium reisseri sp. nov. (Chlorellaceae, Trebouxiophyceae): Redescription of the endosymbiotic green algae of Paramecium bursaria (Peniculia, Oligohymenophorea) in the 120th year. *Phycol Res* **58**, 188–210

[42] **Van Etten J. L., Burbank D. E., Kuczmarski D., Meints R. H. (1983)**: virus infection of culturable chlorella-like algae and development of a plaque assay. *Science* **219**, 994–996

[43] **Hamacher K. (2008)**: Relating Sequence Evolution of HIV1-Protease to Its Underlying Molecular Mechanics. *Gene* **422**, 30–36

[44] **Bremm S., Schreck T., Boba P., Held S., Hamacher K. (2010)**: Computing and Visually Analyzing Mutual Information in Molecular Co-Evolution. *BMC Bioinform* **11**, 330

[45] **Cock P. J., Antao T., Chang J. T., Chapman B. A., Cox C. J., et al. (2009)**: Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423

[46] **Dunigan D. D., Cerny R. L., Bauman A. T., Van Etten J. L., (2012)**: *Paramecium bursaria* Chlorella Virus 1 Proteome Reveals Novel Architectural and Regulatory Features of a Giant Virus. *J. Virol*.. **86**, 8821-8834

[47] **Yanai G. M., (2009)**: Transcription analysis of the chlorovirus Paramecium bursaria chlorella virus-1. *Dissertations and Theses in Biological Sciences*. **Paper 5**

[48] **Neaves K., Cooper L. (2009)**: Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerization and DNA looping. *Nucl. Acids Res.***37***, 2053-2063*

[49] **Wulfmeyer T., Polzer C., Hiepler G., Hamacher K., Shoeman R., Dunigan D. D., Van Etten J. L., Lolicato M., Moroni A., Thiel G., Meckel T. (2012)**: Structural Organization of DNA in Chlorella Viruses. *PLoS ONE* **7**, 2

5. Chapter 4 – The putative DNA binding protein A278L aggregates DNA

5.1. Abstract

DNA aggregation is from a thermodynamic point of view an elaborate process due to the size and the charge of the DNA molecule. To guarantee a successful packaging of dsDNA genomes, a variety of different processes have developed throughout evolution. In case of the *Chlorella* virus, PBCV-1, it is not yet fully understood how this virus can condensate a ca. 100 μm long dsDNA genome in a particle with a diameter of less than 200 nm. In previous experiments only small amounts of molecules with a neutralizing potential, such as polyamines and cations were found in the virus particle, rendering an exclusive neutralization by small molecules unlikely. Proteomic studies of the PBCV-1 particle revealed the abundant presence of small cationic proteins, which resemble in their physical properties DNA binding proteins. This suggests that one of these putative DNA binding proteins is coded by ORF A278L. When the recombinantly expressed and purified A278L protein was incubated with viral DNA the protein caused a higher degree of density of viral DNA than the neutral protein BSA. These data indicate an interaction between the A278L proteins and viral DNA. These results were supported by AFM force measurements. The force, which is required to pull a native viral protein from the DNA is in the same range as that used to pull the recombinant A278L protein from the viral DNA. These forces, which are in the range of nN are specific for a protein/DNA interaction because a several orders of magnitude lower force is required to lift the recombinant A278L protein from the mica surface. The results of these experiments imply that DNA associated proteins are in the virus PBCV-1 involved in DNA aggregation; they may favour the organisation of the DNA in the viral capsid and contribute to the neutralization of the DNA polymer.

5.2. Introduction

A tight packaging of long DNA molecules poses a large energetic barrier due to the necessary bending of the stiff double helix and electrostatic repulsion of the negatively charged DNA phosphates.

Throughout evolution different proteins evolved which interact with DNA and which help to overcome these energetic obstacles [13]. Eukaryotes for example use specific DNA associated proteins, which are responsible for DNA condensation. The most common DNA binding proteins are histones [14]. They are characterized by their small size and their physical properties; e.g. a high isoelectric point (IP) and a high content of positively charged amino acids. This composition causes a high ionic attraction to the negatively charged phosphate groups and guarantees its compensation. In eukaryotes the 14 histone–DNA contacts on the nucleosome involve mainly the sugar–phosphate backbone of the DNA.
The nucleosome might be considered as a cylindrical protein disk that is wrapped by 146 bp of DNA in ~ 1.7 superhelical turns, and functions as a large, loaded spring that is held in place by contacts with the histone octamer. The binding energy between DNA and protein is approximately 59 kJ/mol; each histone–DNA contact provides ~ 4.2 kJ/mol of energy [10].

Also, large dsDNA viruses are able to store their genome with a high density in a small volume of a virus particle [11]. The mechanism with which they package their DNA at this high density is not yet understood [11] In the context to the aforementioned situation in eukaryotes it is interesting to note, that the large dsDNA virus PBCV-1 contains in its virion a large number of proteins [15]. If some of these viral proteins associate with the DNA and if they have a histone like function e.g. if they bind and condensate DNA it should be possible to measure the adhesion energies of these proteins with the DNA. It has already been mentioned that several ORFs from virus PBCV-1 encode for proteins with physical properties suitable for DNA binding [11]. One putative DNA-binding protein, A278L was heterologously expressed in *E. coli* and purified; the purified protein was then used to examine its binding to isolated viral DNA.

As a method of choice for analyzing protein/DNA interactions we employed Atomic force microscopy (AFM). This sensitive technique is a useful tool for visualising single molecules and for measuring binding forces between molecules [16]. A molecule of choice, in this case the protein, can be attached via a gold-thiol interaction to the tip of an AFM probe upon physical contact. The binding of the protein to the tip of the cantilever is in the order of 100 kJ/mol. This strong binding guarantees that the protein remains attached to the tip of the cantilever when it is retracted from its position on the mica surface or from the DNA [9].

If the protein is interacting with DNA a specific force will be required to lift the protein away from the DNA. The precise strength of the DNA/protein interaction is given by the maximal force, which is measured before the complex ruptures.

In the present study, we analyzed the interaction between the recombinant A278L protein and isolated viral DNA with the AFM method. The data show that the viral protein binds to the DNA with a strength that is 2 orders of magnitude stronger than its binding to the mica surface. Hence, the small cationic protein has the hallmarks of a DNA binding protein and may serve in the virus particle for DNA condensation.

5.3. Material and Methods

5.3.1. Amplification and ligation of the gene construct A278L

The gene fragment of the putative viral DNA binding protein was amplified using ligation-independent cloning (LIC) of PCR products containing LIC overlaps. The sense primer was: CAAGGACCGAGCAGCCCCTCCATBTCAACGACACCTGAGAGA and the anti-sense primer was: ACCACGGGGAACCAACCCTTATTATACCGCTTTTGGTGTGAATAG. Due to the higher GC content of the LIC overlabs a two step-PCR was performed. The PCR product was purified and digested with T4-DNA-Polymerase, generating sticky ends. The vector was linearized with *Sma* I and treated with T4-DNA-Polymerase. The product was directly cloned into a vector, which contains a complementary LIC cassette [1, 2]. In addition the pET-based LIC vector also encodes a human Rhinovirus protease cleavage site (HRV3C) followed by a C-terminal Maltose-binding protein and a His6-tag (Dan Minor, UCSF, San Francisco, USA).

5.3.2. Overexpression of protein A278L

Two different cell strains of *Escherichia coli* were used. XL1-blue competent cells were used for screening. BL21 (DE3) pLysS candidates were used for expression optimization and protein production. The modified *E. coli* BL21 (DE3) pLysS cultures were grown in 500 ml Luria-Bertani (LB) media at 37 °C until the cell reached the logarithmic growth phase (OD_{600} nm 0.5 – 0.8.).

Induction was done under sterile conditions at 37 °C using 1 mM IPTG (Isopropyl-β-D-thiogalactopyranosid). After 24 h of growth post-induction, the cells were centrifuged (2 min. at 13000 rpm) and the pellet was resuspended in 100 mL of 500 mM NaCl, 30mM Hepes, 10 % Glycerol, pH 7.5 salt solution. For examination of a successful expression of the protein, the samples were subjected to a SDS-PAGE according to Laemmli and a Western-blot analysis [3, 4]. All samples were mixed 1:1 with 2 x protein loading buffer (20 mM Tris–HCl [pH 8.0], 2 % SDS, 100 mM DTT, 20 % Glycerol, 0.016 % Bromphenol Blue), disrupted by boiling for 5 min and separated in 12 % SDS-PAGE by electrophoresis. Staining was performed with Coomassie brilliant blue R-250 (Sigma-Aldrich, USA). For immunoblotting, proteins were transferred from the gel to a PVDF (polyvinylidene-fluoride) membrane by semi-dry transfer. After blocking with 5 % non-fat dry milk, a mouse anti-His-tag antibody was used as the primary antibody to detect the heterologously expressed protein in the samples.
The secondary antibody was an alkaline phosphatase-conjugated goat anti-mouse immunoglobulin G (IgG) antibody. All the samples were observed by developing with NBT/BCIP solution.

5.3.3. Purification of protein A278L

The successive use of a sonic lance enables the isolation of the viral proteins. Solubilized proteins were separated from insoluble material by ultracentrifugation for 45 min. at 55000 rpm at 4 °C in an ultracentrifuge (Beckman L7 - 65). Following recovery, the supernatant was sterile filtered and brought to a pre-equilibrated Ni-NTA-column. The protein-bound column was washed with the appropriate salt solution containing 30 mM imidazole. Samples were eluted with the same buffer containing 300 mM imidazole. After a successful production of protein, which was examined by SDS-PAGE, the sample was brought to an amylose-column to separate the maltose-binding protein and the candidate protein A278L. The flow through was collected and the volume was reduced to a total volume of 250 μl using a spin column. The protein concentration was determined by absorbance at 280 nm [4]. The subsequent size exclusion chromatography was kindly done by Marco Lolicato (University of Milan, Italy).

5.3.4. Viral DNA isolation

Viruses were isolated as described previously [5].

5.3.5. Force measurement using an AFM

In this study the AFM-method (JPK, NanoWizard ®) was used to image and quantify the interaction between proteins and DNA [6]. For topographical images with high resolution, samples were scanned with a small probe (cantilever) at a distance where cantilever and sample are in physical contact. The cantilever has on one edge a sharp tip with a radius of approximately < 10 nm. This cantilever tip is moved by a piezo crystal with a desired frequency. A double gold-coated, tetrahedral cantilever with a resonance frequency of approximately 21 kHz and a spring constant of 0.011 N/m (APPNANO, HYDRA2R100NGG) was used for the topographical overview scan as well as for the subsequent force measurements.

Any deflection of the cantilever, which is caused by obstacles (i.e. molecules) during the scanning process is sensed with a laser beam, which is reflected from the cantilevers back onto a position-sensitive detector. Hence, any deviation of the laser beam position corresponds to a deflection of the cantilever. This in turn, can be translated into the physical size of the investigated structure and also provides information on adherent forces between object and cantilever tip.

Heterologously expressed and purified A278L protein was diluted in buffer (500 mM NaCl, 30mM Hepes, 10 % Glycerol, pH 7.5) to a final concentration of 100 nM and then incubated under physiological conditions (22 °C, 30 min) with freshly isolated DNA (1 ng/μl) [7, 8]. Further, also the DNA of osmotically shocked PBCV-1 was analyzed. Two ml of DNA-protein solution were transferred onto a smooth mica surface, which provides plain physical qualities. The measurement was performed in aqueous conditions. The gold coated tip of the cantilever was brought close to the mica surface with a contact to the region of interest. By approaching the target structure, the tip applied a force of 50 nN on the surface.

5.4. Results and Discussion

5.4.1. Synthesis of DNA binding protein

The putative DNA-binding protein A278L has physicochemical properties which are similar to histones; it has a small size and a cationic isoelectric point.

Therefore, it seems interesting to analyse the binding properties of the protein to DNA by AFM. The protein was therefore heterologously expressed in *E. coli* and purified on a Ni-NTA column (Figure 5.4.1). The His6-tag and the maltose-binding protein (MBP) were removed and separated via an amylose-column. Further impurities were removed by size exclusion chromatography. The volume and the diameter of the protein were analyzed and compared with BSA; a protein which resembles A278L in size and volume, but not in its physical properties; BSA has a IP of 5.64).

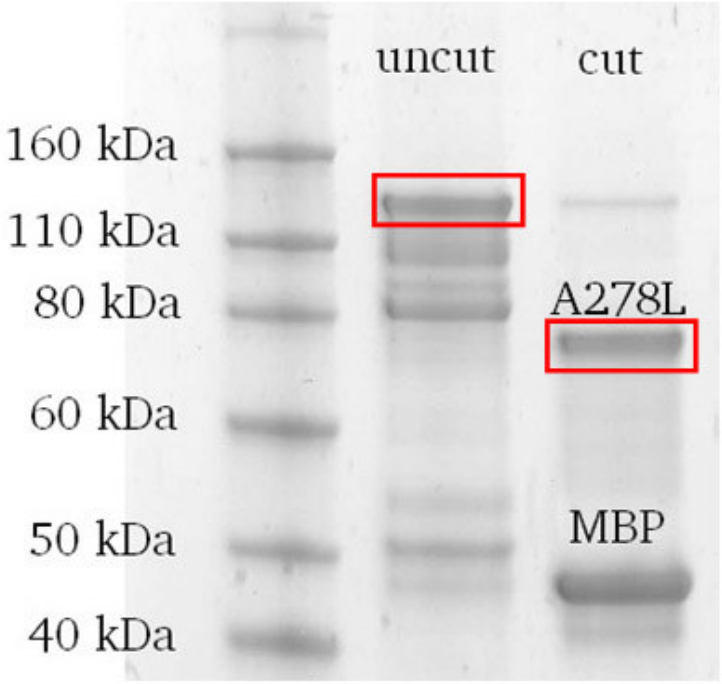

Figure 5.4.1: SDS-PAGE pattern of protein putative DNA binding protein A278L.
The A278L protein was heterologously expressed in *E. coli* and purified on a Ni-NTA column. The His6-tag and the maltose-binding protein (MBP) were removed. The successful protein expression of the protein with His6-tag and MBP (un-cut) and after removal of tags (cut) was examined by a SDS-PAGE. A clear band at 69 kDA is detectable which corresponds to the molecular weight of A278L. Lane 1: weight marker, molecular weight in kDa, lane 2: uncut protein, lane 3: cut protein

The highly purified protein was used for further imaging and force measurements by AFM. Under physiological conditions it was possible to determine the binding affinity of the protein to DNA.

5.4.2. A278L aggregates DNA

Topographical AFM scans were performed with DNA in the absence and presence of either A278L or BSA. Figures 5.4.2 A and B show representative images of viral DNA in the presence of the protein A278L or BSA.

Inspection of many scans implies that the DNA appears more aggregated in the presence of the viral protein than in its absence. We assume that the viral protein leads to a higher degree of condensation due to the higher content of positively charged amino acids.
These are able to bind to the negatively charged phosphate backbone and induce an unspecific packaging. In the topographical AFM overview scans it is recognizable that the recombinant protein A278L, in comparison to BSA, revealed in a qualitatively way, a higher degree of condensation. To quantify the qualitative impression we analyzed standard size areas around DNA molecules by measuring gray values in the selected image sections. Since the gray value of DNA is higher than the background in these images, the value is low when the DNA is not aggregated and high when the DNA is aggregated.

Figure 5.4.2 shows the selection of square fields of interest around DNA molecules; the fields where placed is such a way that the DNA was filling the diagonal of the filed of interest. This procedure favours the comparability of different images. For a quantitative analysis 35 randomly selected DNA molecules with and without A278L protein were selected and analyzed as described. The data in Figure 5.4.2 C show that the putative DNA-binding protein A278L yields, according to this analysis, indeed in a higher degree of condensation.

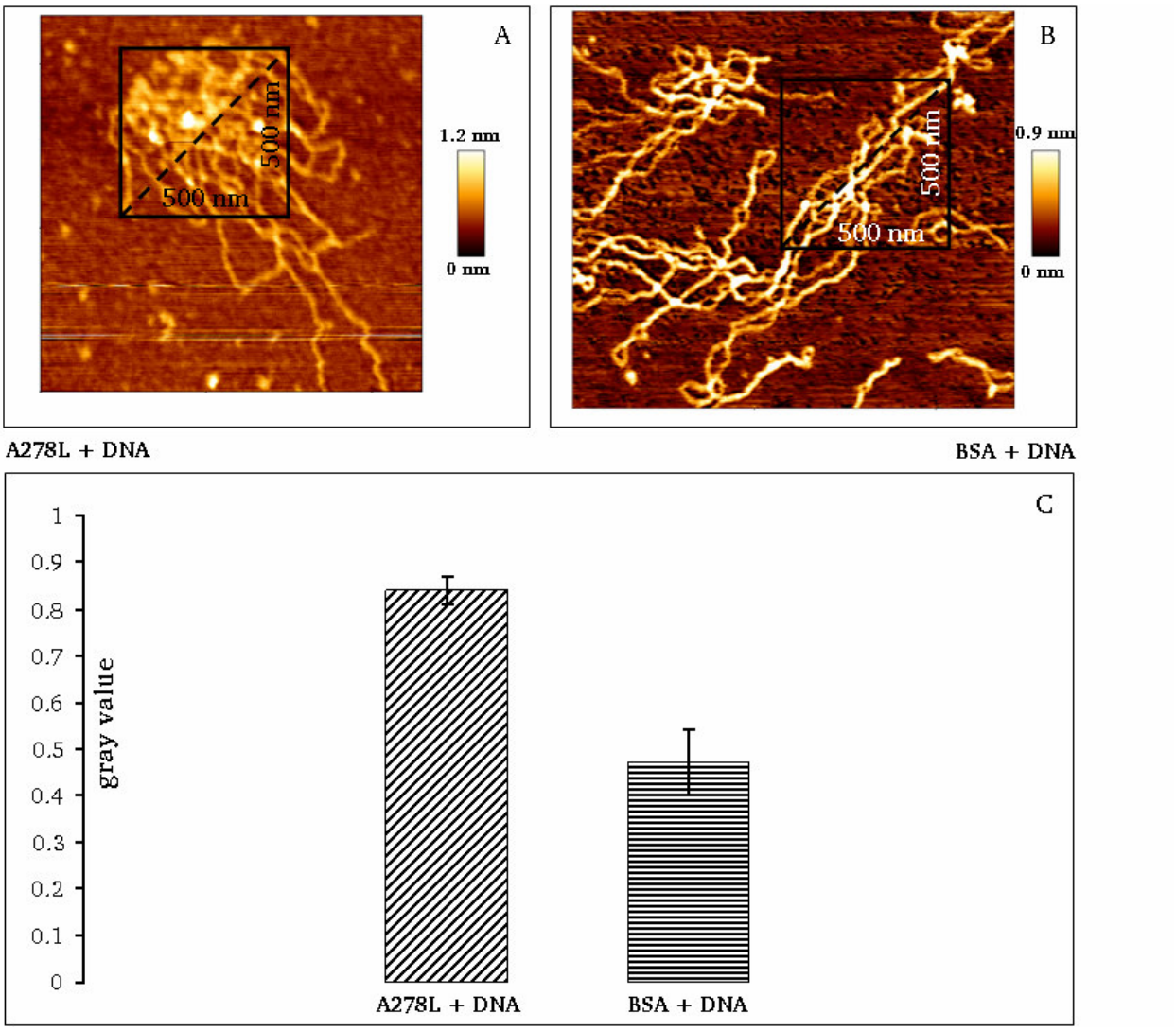

Figure 5.4.2: Degree of DNA condensation visualized in topographical images of DNA-protein complexes. The degree of aggregation increases with the measured gray values. The genomic condensation behaves in a proportional way to the gray value. The higher the condensation, the higher the gray value. A: After adding the viral protein A278L, the DNA seems to aggregate in a higher fashion which is not seen in the presence of BSA (B). C: The aggregation with the putative DNA-binding protein A278L is twice that high.

Collectively, the results of these experiments suggest that the small cationic protein A278L from virus PBCV-1 is indeed a DNA binding protein. This binding favours the condensation of the DNA. The presence of the protein in the virus particle, can in this way contribute to the packaging of the DNA in the virus capsid. To examine the mode of interaction between DNA and A278L further, we performed force measurements using the AFM method.

5.4.3. Viral particles are associated with the DNA

The curvature radius of the AFM tip is below 10 nm, which makes it easy to bring the tip in contact with proteins. Upon contact the gold surface of the AFM tip interacts with the thiol groups of the target protein and thereby forms a stable bond. In case, that the target protein binds DNA, the protein can be pulled away from the DNA by retracting the AFM tip. The rupturing of the protein/DNA interaction can be measured by the AFM and the force value required to pull protein and DNA apart is an indirect measure for the forces with which DNA and protein interact.

In the present experiments, we used synthesized protein A278L and isolated proteins from the virus particles with a radius of approximately 10 nm. This size is sufficiently large to guarantee a binding between tip and protein.

Figure 5.4.3 shows the average binding affinity of particles, which are associated to the DNA. The complex of viral DNA with proteins was isolated by the osmotic shock method [11]. The isolated viral DNA revealed repetitive structures, which seem to be attached to the DNA. To estimate the strength of DNA/protein interaction, we used AFM-force measurements. Figure 5.4.3 shows trace (red) and retrace (blue) curves obtained in force spectroscopy measurements. The trace curve reflects the approaching of the AFM tip on the target protein and the retrace curve shows the force required to pull the tip plus associated protein away from the probe. The increase in energy required to retract the tip from the probe reflects the breakage of the protein/DNA interaction.
Similar binding forces were measured in experiments performed with the protein particles associated with DNA. The maximal force recorded in this experiment was 60 nN.

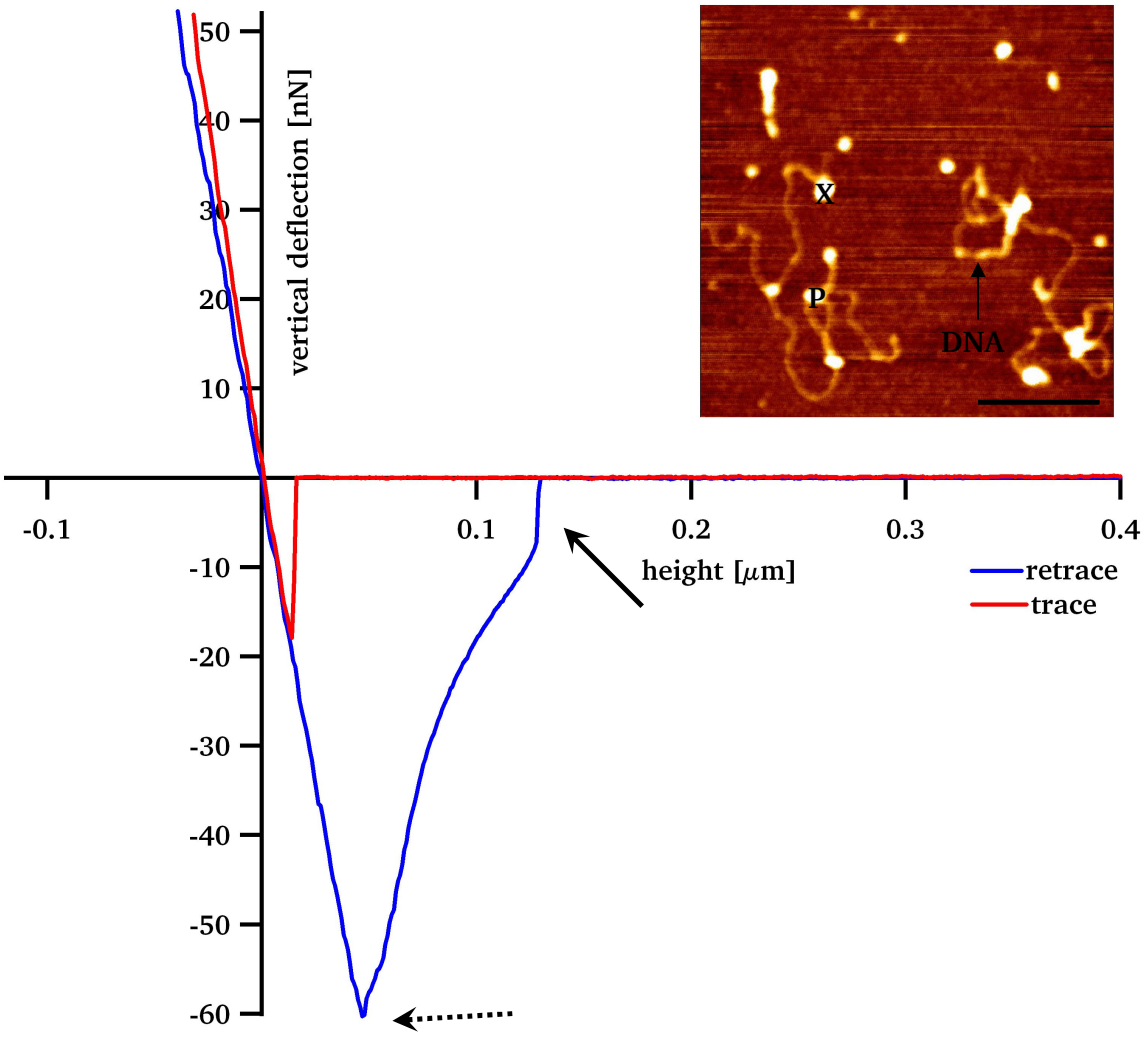

Figure 5.4.3: Force-distance curve of a protein, which is associated with viral DNA.

Following an osmotic shock, viral particle released DNA and proteins. The AFM image on the upper right shows DNA fragments (DNA) with proteins (P) (Inset scale bar: 500 nm). The proteins are either associated with the DNA or free. The binding forces between a protein and the DNA (marked by X) were analyzed by force measurement experiments. The AFM cantilever was therefore advanced on the target structure until the tip generated a force of 50 nN. By retracting the tip, two unfolding processes were detected (dashed arrow, arrow). The relaxation of the force is probably to the breaking up of the protein/DNA interaction).

Interestingly, in Figure 5.4.3 in both spectra an adhesion of the tip is detectable. The increase of the force in the trace curve (red) is due the pulling effect of the particle.

By getting in contact with the sample, the attractive forces of the overlapping van der waals forces yield to a snap in effect of the tip to the surface which results in a shift of the measured force.

This artefact was measured in all spectra and does not play a major role in the following force curves. In comparison, in the retrace curve it is visualized that by retracting the tip, the binding between protein and DNA breaks, which results in a measured force of 60 nN. The curve revealed a characteristic kinetic. The progression of the curve shows a shoulder followed by a steep part of the curve (Figure 5.4.3 arrow). We assume that these enfolding events are due to the breakage of the protein/DNA complex. The steep event at the backmost part of the retrace curve is interpreted as a rapid enfolding event of the protein by breaking a binding between protein and DNA (Figure 5.4.3, arrow).

5.4.4. The viral particles bind DNA in a specific fashion

It has been mentioned above that the un-binding between protein and DNA shows a complex kinetic. After pulling the tip away from the probe the force reaches a maximum at a distance of ca. 20 to 50 nm from the probe. This maximum force value can be interpreted as the binding force for protein and DNA. The average force value, estimated from Figure 5.4.4, is 64 nN. After passing this force maximum the retrace curve relaxes in the same fashion before the force rapidly jumps to the zero line (Figure 5.4.4). The reason for this complex curve remains rather unknown. However, it is reasonable to assume that the retrace curve, after an initial first breaking of protein/DNA binding, shows some further enfolding of the protein/DNA aggregate. The final jump with a mean value of 7 nN may then reflect a second binding between DNA and protein.

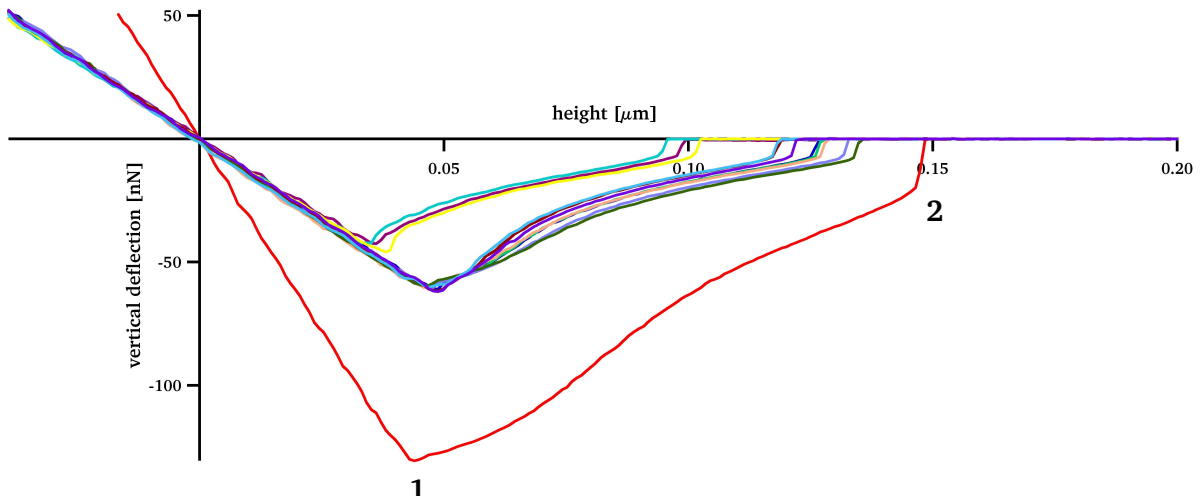

Figure 5.4.4: Force-distance curves from protein/DNA interactions.

Force distance curves were recorded as in Figure 5.4.3 for 12 individual proteins in direct contact with DNA. All force spectra show the same kinetics. The force relaxes over a distance of >100 nm in a complex fashion after passing an initial maximum (1); in a final step the force drops to zero (2). The mean maximal force measured at the onset of the relaxation is 64 nN; the mean force of the final relaxation is 7 nN.

We assume that the protein binds to DNA in a more intensive fashion. In addition to the two breakages of the complex (Figure 5.4.4, 1.2), we conclude that the DNA is twice wrapped around the protein. By pulling with the tip at the aggregate, the binding breaks and the protein dissociates from the DNA.

5.4.5. The viral protein A278L binds DNA in the same order

To determine whether the putative DNA binding protein A278L is indeed binding DNA, we examined the interaction of the recombinant A278L protein with DNA. The experiments were done under the same conditions as those reported before on the interaction between viral DNA and viral proteins. Thereby, it is possible to compare the data also on a quantitative level. Figure 5.4.5.1 shows a typical experiment in which we incubated viral DNA and the purified A278L protein together. In AFM images, it is possible to find single proteins, which spontaneously associate with the DNA polymer. To measure the force of protein/DNA association, we advanced the AFM cantilever until the tip applied a force of 50 nN. The average retrace curve of these experiments is blotted in Figure 5.4.5.1. In 8 experiments a mean maximal force of 22 nN was measured in case of the protein/DNA aggregate (dashed line). This force reflects the interaction between protein and DNA.

Interestingly, the force spectrum reveals an abrupt jump to the zero line. Worth noting is, that the retrace curve shows in the case of the A278L protein, a much more simple kinetic, than that recorded with native viral protein. The single jump in the retrace curve measured with the A278L protein implies a single binding interaction between protein and DNA.

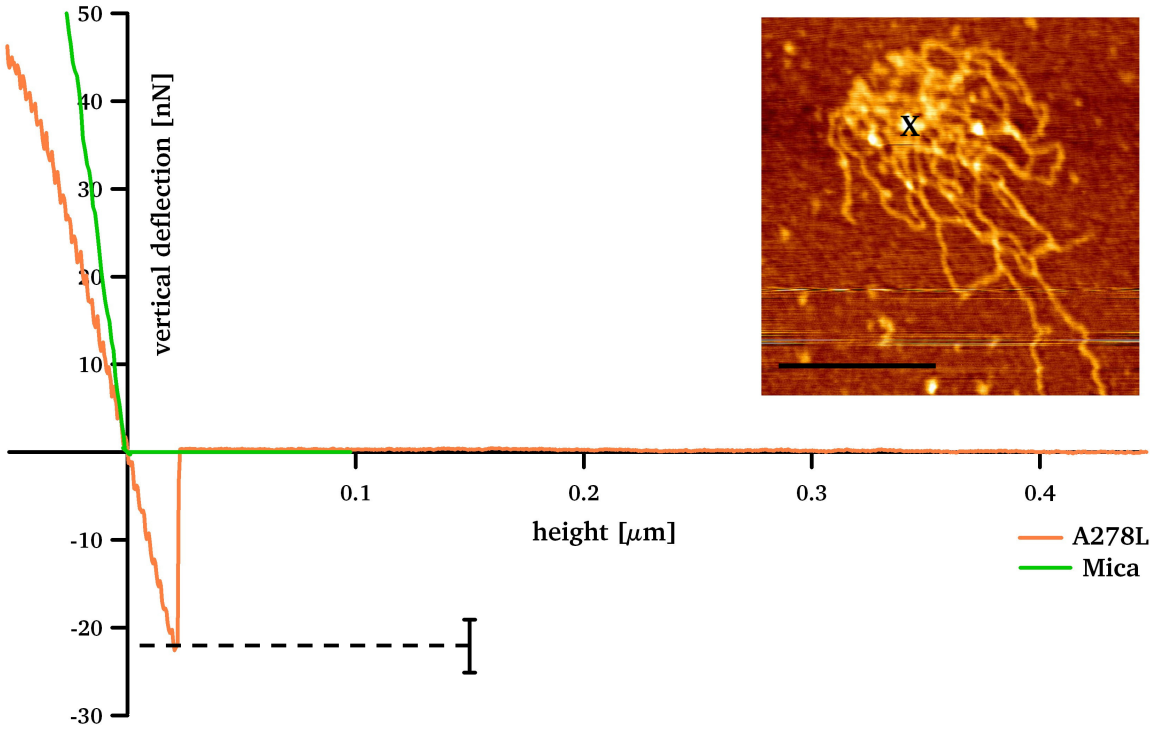

Figure 5.4.5.1: Force-distance curve for the purified viral protein A278L in association with viral DNA.
The AFM image shows isolated viral DNA incubated together with purified A278L proteins. The X marks the position of a protein in contact with the DNA (Inset scale bar: 500 nm). A force-distance curve was done as in Figure 5.4.3 with the cantilever tip targeted to the bare mica (green curve) or to the protein marked (orange curve). The average retracting forces from the mica are very small (0.27 nN) and negligible. When the cantilever was retracted from the protein, an additional force occurred which relaxed abruptly at about 22 nN. The dashed line shows the measured mean maximal force of n: 8 measurements. This force is proportional to the interaction between DNA and protein. Only the retrace curves are blotted. An oscillation on the force trace is probably an artefact due to external noise.

Another difference is the flickering event after the tip approached the surface. We assume that this is supposed to be an artefact due to external noise, which does not correlate to the binding affinity of the protein. Concluding, the curvature shows a different kinetic. But the artificial viral protein bind DNA in the same range as the isolated particle does. A mean force of approximately 22 nN was analyzed.

The data show, that the A278L protein is binding to DNA; the strength of binding is in the same order of magnitude, albeit smaller than that recorded with the association of the viral structures and DNA. In order to test if the binding of the protein to the DNA is specific we tested in the next step the forces, which are required to pull a purified A278L protein from mica surface. Figure 5.4.5.2 shows the average retract traces of experiments in which the cantilever was pushed against isolated viral structures (blue) and protein/DNA complexes (orange). Further, measurements were performed against mica surface (green), or against a protein on mica surface (purple). The retrace curves recorded under the two conditions (mica, A278L – DNA) show no appreciable forces. This implies that the tip of the cantilever does not bind to the mica surface. The data also show, that the isolated proteins can be lifted from the mica without a large force. This implies that the protein is not interacting with the surface. The data further underscore, that the forces measured with proteins in contact with DNA are indeed reporting the interaction between DNA and protein. The dashed line in Figure 5.4.5.2 shows the measured mean maximal force of n = 8 in case of the protein/DNA complex and n: 12 in case of the viral structures.

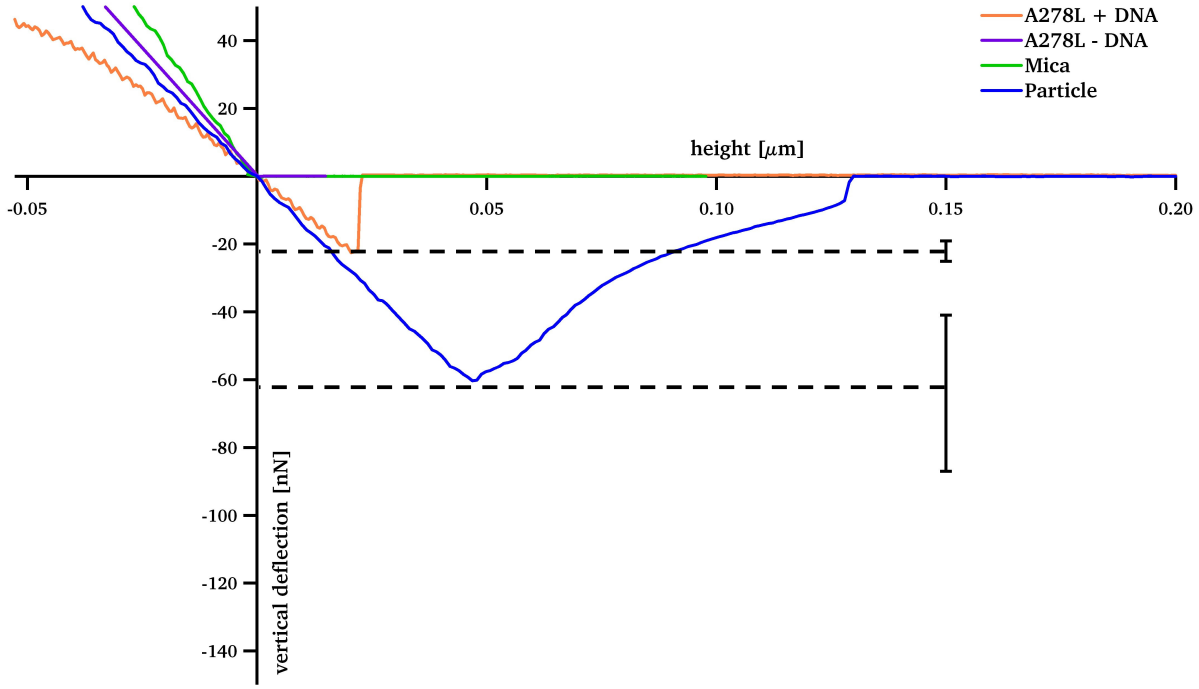

Figure 5.4.5.2: Force-distance curves of different aggregates.

Comparison of average force-distance curves for different targets. The AFM cantilever was targeted as in Figure 5.4.3 either to the mica surface (green), to a DNA/A278L protein complex (orange) to a single A278L protein detached from DNA (purple) or to a native viral protein attached to DNA (blue). The dashed line shows the measured mean maximal force of n = 8 in case of the protein/DNA complex and n = 12 in case of the viral structures.

5.5. Conclusion

The present results suggest that chlorovirus PBCV-1 neutralizes and structures its DNA with DNA-binding proteins. The data support the hypothesis that PBCV-1 presumably encodes small cationic proteins, which have a high affinity for the negatively charged DNA and which contribute with this to the neutralization of the viral genome. We found in the genome of PBCV-1 one ORF, A278L, which presents an interesting candidate for DNA binding. The gene product is small, strongly cationic and present at a high copy number in the virus particle [11]. We found in a qualitatively assay that the recombinant protein A278L causes a condensation of the viral DNA, which is twice as much as what the neutral protein BSA generates in DNA condensation. AFM force spectroscopic measurements reveal that the viral proteins are tightly associated with the viral DNA.

Proteins, which are associated with the isolated viral DNA, exhibit on average a binding force of 64 ± 23 nN with the DNA. Also the putative DNA-binding protein A278L interacts with the DNA; for recombinant A278L an average interaction force of 22 ± 3 nN with viral DNA was recorded. These protein/DNA interactions are specific because the viral protein binds to the DNA with a strength that is two orders of magnitude stronger than its binding to the mica surface. Also, the aforementioned high forces were not detected in case of individual proteins, which were in the preparation not directly associated with the DNA. In the latter case a mean force of 70 pN was required to detach a native isolated protein of the recombinant A278L protein from the mica surface. This force is in the range of values (44 pN to 277 pN [12]) reported for the typical rupture force of single proteins e.g. lysozyme, ferritin, bovine serum albumin and myoglobin, which are not specifically connected to a substrate. In case of specific interactions, e.g. for protein/protein aggregation of avidin/biotin, the rupture force goes up to 20 nN [16]. These values are in the same range as those measured here for the protein/DNA interaction. The comparison with published data supports the view that the viral proteins bind tightly to the viral DNA. This binding may contribute to the packaging of the DNA in the small viral particle by neutralizing negative charges and by organizing the DNA inside the particle.

5.6. References

[1] **Aslanidis C., de Jong P. J. (1990):** Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research* **18**, 069-6074

[2] **Hammon J., Palanivelu D., Chen J., Patel C., Minor D. (2008)**: A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms. *Protein Sci*. **18**, 121–133

[3] **Laemmli U.K. (1970):** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, S. 680-685

[4] **Yan L., Guo H., Sun X., Shao L., Fang Q. (2012)**: Characterization of grass carp reovirus minor core protein Vp54. *Virol J*. **9**, 89

[5] **Howarth A. (1988):** Methods for rapid isolation and analysis of algal virus DNA. *Journal of Virological Methods* **19**, 3–4

[6] **Lohr D., Bash R., Wang H., Yodh J., Lindsay S. (2007)**: Using atomic force microscopy to study chromatin structure and nucleosome remodelling. *Methods* **41**, 333-341

[7] **Schneider S. W., Lärmer J., Henderson R. M., Oberleithner H. (1998)**: Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy. *Eur J Physiol* **435**, 362–367

[8] **Hansma H. G., Revenko I., Kim K., Laney D. E. (1996)**: Atomic force microscopy of long and short double-stranded, single-stranded and triple-stranded nucleic acids. *Nucleic Acids Research* **24**, 713–720

[9] **Forster R. J., Keyes T. E., Vos J. G. (2003)**: Interfacial Supramolecular Assemblies. *Wiley*. 88–94.

[10] **Saha A., Wittmeyer J., Cairns B. R. (2006)**: Chromatin remodelling: the industrial revolution of DNA around histones. *Nature Reviews Molecular Cell Biology* **7**, 437-447

[11] **Wulfmeyer T., Polzer C., Hiepler G., Hamacher K., Shoeman R., Dunigan D. D., Van Etten J. L., Lolicato M., Moroni A., Thiel G., Meckel T. (2012)**: Structural Organization of DNA in Chlorella Viruses. *PLoS ONE* **7**, 2

[12] **Sagvolden G (1999)**: Protein adhesion force dynamics and single adhesion events. *Biophys J* **77**, 526–532

[13] **Richmond T. J., Davey C. A. (2003)**: The structure of DNA in the nucleosome core. *Nature* **423**, 145-150

[14] **Hayes J. J., Tulliust T. D., Wolffe A. P. (1990)**: The structure of DNA in a nucleosome. *Proc. Nati. Acad. Sci. USA* **87**, 7405-7409

[15] **Zhang X., Xiang Y., Dunigan D. D., Klose T., Chipman P. R., Van Etten J. L., Rossmann M. G. (2011)**: Three-dimensional structure and function of the Paramecium bursaria chlorella virus capsid. *PNAS* **108,** 14837-14842

[16] **Lin S., Chen J. L., Huang L. S., Lin H. W. (2005)**: Measurements of the Forces in Protein Interactions with Atomic Force Microscopy. *Current Proteomics* **55,** 81 55

6. Summery

DNA condensation is in all living organisms a difficult process due to the large size of a genome, in which as much genetic information as possible has to be packaged into a small volume. Frequently, the storing compartment has a diameter, which is thousand times smaller than the length of the DNA molecule. Evolution has solved this problem in different ways. Overlapping genes in viruses are able to increase the genomic information without compromising the size of the DNA molecule. In other viruses positively charged molecules and cations guarantee a neutralization of the DNA polymer and hence dense packaging. In higher organisms proteins are employed in tightly packaging DNA aggregates. An interesting example for successful condensation and packaging of a long dsDNA genome in a small volume is presented by the *Chlorella* viruses. In the case of virus PBCV-1 (*Paramecium bursaria Chlorella* virus-1), the prototype of *Chlorella* viruses, a ca. 100 μm long DNA molecule is packaged into a volume of only 1.7 * 10^{-7} m^3. The small amounts of polyamines, which are present in the viral particle, are insufficient for charge compensation. To test whether virus PBCV-1 uses, like phages, cations for charge compensation, we measured the content of cations in the virus particles with Energy dispersive X-ray spectroscopy and Inductively coupled plasma-mass spectrometry. The data reveal that cations can be detected, but the concentration in the particle is not sufficient to accomplish the whole neutralization of the DNA molecule; only one fifth of the total phosphor concentration can be neutralized by Mg^{2+} or K^{+} or Ca^{2+}. By summing up all cations, detected in the particle, we can estimate, that 58 % of the total phosphate can be neutralized in this way.

Imaging of ejected viral DNA indicates that it is intimately associated with proteins in a periodic fashion. A combination of fluorescence images of ejected DNA and a bioinformatics analysis of the DNA reveal periodic patterns in the viral DNA. The periodic distribution of GC rich regions in the genome provides potential binding sites for basic proteins. Collectively the data indicate that the large *Chlorella* viruses have a DNA packaging strategy that differs from that of bacteriophages; it involves proteins and shares similarities to that of chromatin structure in eukaryotes.

In the genome of PBCV-1 an ORF, A278L, encodes for a protein, with kinase activity, which resembles in its physical properties, i.e. small size and basic isoelectric point (IP: 10.8), histones. Further, the protein is present with a high copy number (397) in the virus particle. Altogether, this makes the gene product of A278L an interesting candidate for DNA condensation. We find in a qualitatively way that the recombinantly produced and purified protein A278L indeed causes a higher degree of genomic condensation. In comparison with the neutral protein BSA, A278L causes a roughly two fold higher degree of DNA clustering. Atomic force microscopy force spectroscopy was used to determine the DNA binding capacity of the A278L protein. An analysis of binding of native viral proteins, which were in contact with the viral DNA after a release by osmotical shock from the PBCV-1 virions, revealed a characteristic binding behaviour. The data revealed an average force of 64 ± 23 nN for the binding between native proteins and viral DNA. The recombinant protein A278L showed a similar binding strength to isolated viral DNA; on average 22 ± 3 nN were required to separate the protein from the DNA. These measured forces for separating proteins from DNA are specific for a protein/DNA interaction because proteins could be pulled away from the mica surface with a 100 times smaller force. The results of these experiments imply that virus PBCV-1 employs small molecules such as cations and polymaines in combination with proteins for packaging its large dsDNA genome in the virus particle. The proteins may support the organization of meta-structures and the consequent achievement of a crystalline-like order inside the virion. In addition, they may also provide substantial contribution to the neutralization of the DNA. The present results provide some new understanding of viral DNA packaging. This may help to create artificial devices for DNA shuttles in the field of gene therapy.

7. Zusammenfassung

Der Prozess der DNA-Kondensation stellt in allen lebenden Organismen einen komplizierten Prozess dar. Der Größenunterschied zwischen DNA-Molekül und Zelle bzw. Viruskapsid, sowie die negative Ladung des Phosphatrückgrates der DNA sind die größte Schwierigkeit einer erfolgreichen DNA-Aggregation.

Im Laufe der Evolution haben sich unterschiedliche Prozesse entwickelt, um diese Aufgabe erfolgreich durchzuführen. So kodieren z.B. eukaryotische Zellen für Proteine, so genannte Histone, die wie in einer Spule von der DNA eng umschlungen werden. Dadurch ist es möglich die DNA zu kondensieren und eng zu packen. Dieser Prozess kann ebenfalls von positiv geladenen Molekülen, Polyaminen oder Kationen durchgeführt werden. Ebenso wurden Prozesse etabliert, die es ermöglichen Geninformation auf geringem Raum zu kodieren, die so genannten overlapping genes, um die Größe des Genoms zu reduzieren. Chlorellaviren sind ein positives Beispiel für eine erfolgreiche DNA-Kondensation wie am Prototyp dieser Familie, PBCV-1 zu erkennen ist. Die Herausforderung für diese Virengruppe besteht darin, ein Genom in ihrem Kapisd zu verpacken, welches 1000 mal länger ist.

In topographischen Atomic force microscopy - und fluoreszenzmikroskopischen Messungen wurde festgestellt, dass die isolierte, virale DNA repetitive Strukturen aufweist. Des Weiteren zeigen Sequenzanalysen, dass eine periodische Verteilung GC-reicher Sequenzen vorliegt, die mögliche Bindungsstellen für basische Proteine sein können. Biochemische Experimente zeigen, dass keine signifikanten Konzentrationen an Polyaminen vorliegen und diese Menge lediglich 0.02 % der DNA kondensieren könnte. Ebenso liegt keine höhere Konzentration eines einzelnen Kations, z.B. Mg^{2+}, vor, wie Energy dispersive X-ray spectroscopy und Inductively coupled plasma-mass spectrometry belegen. Es ist allerdings möglich, dass die Summe aller Kationen für die Kondensierung teilweise verantwortlich ist, da bereits 58 % der DNA neutralisiert werden könnte. Diese Erkenntnis steht im Gegensatz zu den Ergebnissen der Phage λ, die ausschließlich Kationen und Polyamine zur Aggregation nutzt.

PBCV-1 kodiert für ein Protein, A278L, welches Kinaseaktivität aufweist. Des Weiteren ähnelt A278L in den physikalischen Eigenschaften Histonen und ist somit ein wichtiger Kandidat für die Aggregation der viralen DNA.

Das rekombinante Protein zeigt in topographischen AFM-Messungen, im Vergleich zum neutralen Protein BSA, eine Verdopplung der Aggregation der viralen DNA.

Kraftspektroskopische Messungen zeigen, dass das rekombinante Protein A278L (25 nN) eine ähnliche Bindungskraft zur DNA aufweist wie die an der DNA assoziierten Partikel (64 nN). Dieses Protein könnte, neben Kationen, für das Aggregieren der DNA verantwortlich sein, so dass eine Metastruktur und eine Organisation der viralen DNA gewährleistet werden kann.

8. Danksagung

Zum Abschluss dieser Arbeit möchte ich mich bei all denjenigen bedanken, die auf die ein oder andere Art und Weise zum Gelingen dieser Arbeit beigetragen haben. Dabei gilt mein Dank vor allem:

Prof. Dr. Gerhard Thiel, für die Betreuung dieser sehr spannenden und interessanten Arbeit. Für die stetige Hilfe und Unterstützung bei allen schwierigen Fragestellungen.

Prof. Dr. Felicitas Pfeifer, für alle freundlichen Ratschläge und guten Hinweise, sowie für die Übernahme des Koreferats.

Ich möchte mich für die wissenschaftliche und freundschaftliche Zusammenarbeit bedanken bei **Dr. Tobias Meckel, Dr. Marek Janko, Dr. Manuela Gebhardt, Dr. Timo Greiner, Dr. Vera Bandmann, Dr. Brigitte Hertel, Charlotte von Chappuis, Alice Kress, Fenja Siotto, Silvia Haase, Barbara Reinhards und Sylvia Lenz**.

Des weiteren gilt besonderer Dank meinen Freunden **Bastian Roth**, **Christian Polzer Christian Braun**, **Sebastian Vögler**, **Florian Ludwig, Maic Fredersdorf, Tim Küsel und Thomas Guthmann.** Vielen Dank für die tolle Unterstützung.

Ebenso möchte ich der gesamten **AG Thiel, AG Moroni, AG Bertl** für eine schöne Zusammenarbeit in Darmstadt und Mailand und für die viele Unterstützung bedanken.

Weiterer Dank gilt meinen Eltern **Angelika** und **Ulrich Wulfmeyer,** meinem Burder **Hendrik Wulfmeyer** und meiner Schwägerin **Julia Wulfmeyer**, die immer an mich geglaubt und mich stets unterstützt haben.

Besonderer Dank gilt auch meiner Freundin **Sina Marquardt**, für all die Liebe, Freundschaft und Unterstützung während der gesamten Zeit.

9. Curriculum vitae

Personal data

Name	Timo Wulfmeyer
Date of birth	28.10.1982

Academic Career

Since 12/2009	Dissertation: Plant Membrane Biophysics at Institute of Botany, Technische Universität Darmstadt, Germany Title: "Interaction of DNA and proteins in virus PBCV-1 particles"

Diplomathesis

01/2009-09/2009	Diplomathesis: Plant Membrane Biophysics at Institute of Botany, Technische Universität Darmstadt, Germany Title: "Mechanism of DNA charge compensation in *Chlorella* Virus PBCV-1"

Academic studies

10/2004-09/2009	Study of Biology, Technische Universität Darmstadt, Germany

School Education

1993-2003	Lichtenberggymnasium, Darmstadt Graduation: general qualification for university entrance (Allgemeine Hochschulreife)
1989-1993	Erich-Kästner-Schule, Pfungstadt primary school

Workshops

01/2010	AFM in Biology Class, Asylum Research, Mannheim, Germany
09/2010-10/2010 12/2010-12/2010	Advanced internship about protein expression systems and purification processes at the University of Milan, Italy

Presentations on scientific conferences

10/2010	Biophysiktagung, Bochum, Germany
08/2010	AFM Biomed Conference, Paris, France

Publications

02/2012	Wulfmeyer T., Polzer C., Hiepler G., Hamacher K., Shoeman R., et al. (2012): Structural Organization of DNA in Chlorella Viruses. PLoS ONE 7(2)

10. Affidavit

Herewith, I declare, that I prepared the present Doctoral thesis “Interaction of DNA and proteins in virus PBCV-1 particles” on my own and with no other sources and aids than quoted. This thesis has not been submitted to any other examination authority in its current or an altered form, and it has not been published.

Darmstadt, 05.10.2012

_____

Dipl.-Biol. Timo Wulfmeyer

11. Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.
Ich habe bisher noch keinen Promotionsversuch unternommen.

Darmstadt, 05.10.2012

_____

Dipl.-Biol. Timo Wulfmeyer

12. List of Abbreviations

- AFM — atomic force microscopy
- BCIP — 5-Brom-4-chlor-3-indolylphosphat
- bp — base pairs
- BSA — bovine serum albumin
- CDS — protein-encoding genes
- DAPI — 4',6-diamidino-2-phenylindole
- DNA — deoxyribonucleic acid
- DTT — dithiothreitol
- EDX — energy dispersive X-ray spectroscopy
- FFT — Fast-Fourier-Transformations
- H1 — histone 1
- H2A — histone 2A
- H2B — histone 2B
- H3 — histone 3
- H4 — histone 4
- HRV3C — human Rhinovirus protease cleavage site
- ICP-MS — Inductively coupled plasma-mass spectrometry
- IP — isoelectric point
- IPTG — isopropyl-β-D-thiogalactopyranosid
- Kcv — K^+ Channel *Chlorella* Virus
- LIC — ligation-independent cloning
- MBP — maltose-binding protein
- MBBM — Modified Bold´s Basal Medium
- m.o.i. — multiplicity of infection
- NBT — nitro blue tetrazolium
- NCLDV — nucleocytoplasmic large DNA virus
- OD_{600} — optical density at 600 nm
- p.i. — post infection
- PBCV-1 — *Paramecium bursaria Chlorella* virus -1
- PAGE — polyacrylamide gel electrophoresis
- PFU — plaque forming unit

- ppm	parts per million
- RMS	root mean square
- rpm	revolutions per minute