Molecular Dynamics Studies of the φ29 Connector–DNA complex

Dissertation
for the award of the degree
"Doctor rerum naturalium" (Dr.rer.nat.)

(alternatively: "Doctor of Philosophy" Ph.D. Division of Mathematics and Natural Sciences)

of the Georg-August-Universität Göttingen

within the doctoral program Biology
of the Georg-August University School of Science (GAUSS)

submitted by

Rajendra Kumar

from Baunsi

Göttingen, 2014

Thesis Committee

Prof. Dr. Helmut Grubmüller
Department of Theoretical and Computational Biophysics
Max Planck Institute for Biophysical Chemistry

Prof. Dr. Holger Stark
3D Electron Cryo-Microscopy Research Group
Max Planck Institute for Biophysical Chemistry

Members of the Examination Board

Reviewer:

Prof. Dr. Helmut Grubmüller
Department of Theoretical and Computational Biophysics
Max Planck Institute for Biophysical Chemistry

Second Reviewer:

Prof. Dr. Holger Stark
3D Electron Cryo-Microscopy Research Group
Max Planck Institute for Biophysical Chemistry

Further members of the Examination Board:

Prof. Dr. Wolfgang Fischle
Chromatin Biochemistry Research Group
Max Planck Institute for Biophysical Chemistry

Prof. Dr. Burkhard Morgenstern
Department of Bioinformatics,
Institute of Microbiology and Genetics
Georg-August-Universität Göttingen

Prof. Dr. Kai Tittmann
Department of Bioanalytics
Albrecht von Haller Institute
Georg-August-Universität Göttingen

Prof. Dr. Ralf Ficner
Department of Molecular Structural Biology
Institute of Microbiology and Genetics
Georg-August-Universität Göttingen

Date of the oral examination: 18.07.2014

To my mother and father.

विद्या ददाति विनयं विनयाद्याति पात्रताम् ।
पात्रत्वाद्धनमाप्नोति धनाद्धर्मं ततः सुखम् ॥

vidyā dadāti vinayam, vinayādyati pātratām ।
pātratvāddhanamāpnoti, dhanāddharma tatā sukham ॥

— Hitopadesh.

Education gives Humility, Humility gives Character,
from character one gets wealth, from wealth one gets
righteousness, in righteousness there is joy.

कर्मणये वाधिकारस्ते मां फलेषु कदाचन ।
मां कर्मफलहेतुर्भूः मां ते संङ्गोस्त्वकर्मणि ॥

karmany evādhikāraste mā phalesu kadācana ।
mā karma-phala-hetur bhūr mā te saṅgo'stv akarmani ॥

— Verse 47, Chapter 2, Bhagavad Gita.

You have a right to perform your prescribed duty,
but you are not entitled to the fruits of action.
Never be motivated by the result of your activities,
and never be attached to not doing your duty.

List of Abbreviations

AAA or AAA+	ATPases Associated with diverse cellular Activities
AFM	Atomic Force Microscopy
ATP	Adenosine triphosphate
Cryo-EM	Cryo-electron Microscopy
DNA	Deoxyribonucleic Acid
FP	Force-probe
fs	femtoseconds
gp	Gene Product
LINCS	Linear Constraint Solver
LJ	Lennard-Jones
MD	Molecular Dynamics
MH	Middle region helix
NPT	Number of Particles, Pressure, Temperature
ns	nanoseconds
NVT	Number of Particles, Volume, Temperature
PC	Principal Component
PCA	Principal Component Analysis
PME	Particle Mesh Ewald
pRNA	Prohead Ribonucleic Acid
ps	picoseconds
QM	Quantum Mechanical
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic Acid
SAMD	Simulated Annealing Molecular Dynamics
US	Umbrella sampling
WHAM	Weighted Histogram Analysis Method

Contents

INTRODUCTION **1**

1.1. DNA PACKAGING MOTOR 3

1.2. COMPONENTS OF THE DNA PACKAGING MOTOR 4

1.2.1. The connector *4*

1.2.2. Prohead RNA *5*

1.2.3. ATPase *6*

1.3. DNA PACKAGING MECHANISM 6

1.3.1. DNA packaging models *7*

1.3.1.1. The rotation model *7*

1.3.1.2. Untwist-twist model *8*

1.3.1.3. Push-roll model *9*

1.3.1.4. One-way-revolution model *9*

1.4. AIMS OF THIS STUDY 10

CONCEPTS AND THEORY **12**

2.1. MOLECULAR DYNAMICS SIMULATIONS 13

2.1.1. Schrödinger equation *13*

2.1.2. Born-Oppenheimer approximation *13*

2.1.3. Molecular mechanics force field *14*

2.1.4. Dynamics of the atoms *15*

2.1.5. Force-probe MD simulations *17*

2.1.5.1. Pulling/Pushing simulations 17

2.1.5.2. Rotation simulations 18

2.2. ELASTIC PROPERTIES FROM THE EQUILIBRIUM FLUCTUATIONS 18

2.2.1. Elastic spring constants *18*

2.2.2. Young modulus of elasticity *19*

2.3. PRINCIPAL COMPONENT ANALYSIS 19

MODELING AND REFINEMENT OF THE CONNECTOR LOOPS **21**

3.1. BACKGROUND 22

3.2. METHODS 22

3.3. RESULTS AND DISCUSSION 24

3.4 SUMMARY 25

ELASTIC PROPERTIES OF THE CONNECTOR **27**

4.1. BACKGROUND 28

4.2. PUBLICATION 29

4.2.1. Abstract *29*

4.2.2. Introduction *29*

4.2.3. Methods *31*

4.2.4. Results and discussion *32*

4.2.5. Conclusion *36*

4.3. SUPPORTING MATERIAL 40

ROLE OF THE CONNECTOR IN THE DNA PACKAGING **61**

5.1. BACKGROUND 62

5.2. METHODS 63

5.2.1. Equilibrium MD simulations *63*
5.2.2. Force-probe simulations *64*
5.3. Results and Discussion 66
5.3.1. The equilibration of the Connector-DNA complex *66*
5.3.2. Impact of the DNA on the connector *67*
5.3.3. Deformation in the DNA on its confinement in the channel *71*
5.3.4. How does the connector prevent the DNA leakage? *74*
5.3.4.1. Does the connector act as a one-way valve? 74
5.3.4.2. How the connector's loops are essential? 75
5.3.4.3. How do the flexible loops minimize DNA leakage? 77
5.3.5. Role of the four electropositive rings in the connector channel *78*
5.3.6. Presence of gap between the channel wall and the DNA helix *80*
5.3.7. Elastic properties of the connector in presence of the DNA *81*
5.4. Summary 84

SUMMARY, CONCLUSION AND OUTLOOK 87

6.1. Summary 88
6.2. Conclusion and Outlook 93

REFERENCES 95

APPENDIX 105

8.1. Derivation of Young's modulus 106

ACKNOWLEDGEMENT 109

CURRICULUM VITAE 110

Introduction

Chapter 1

Viruses can infect almost all type of organisms such as plants, animals and bacteria, and replicate using resources present in the host cell. The viruses infecting bacteria are categorized as bacteriophage. One widely studied bacteriophage is φ29, which is a type of lytic phage, belongs to *Podoviridae* family, and specifically infects the bacterium *Bacillus subtilis* (1). The bacteriophage particle is an assembly of a prolate icosahedral capsid, appendages, a lower collar, a tail knob, and several tails as depicted in Fig. 1.1A (2, 3). The lower collar is similar to a shaft and connects to the capsid via a connector protein and the tails via a tail knob. The capsid serves as a container to store double stranded DNA (dsDNA). The elasticity of capsid was determined to be ~1.8 GPa similar to other proteins such as silk and collagen fibers which are evolved to endure extreme mechanical stress (4, 5).

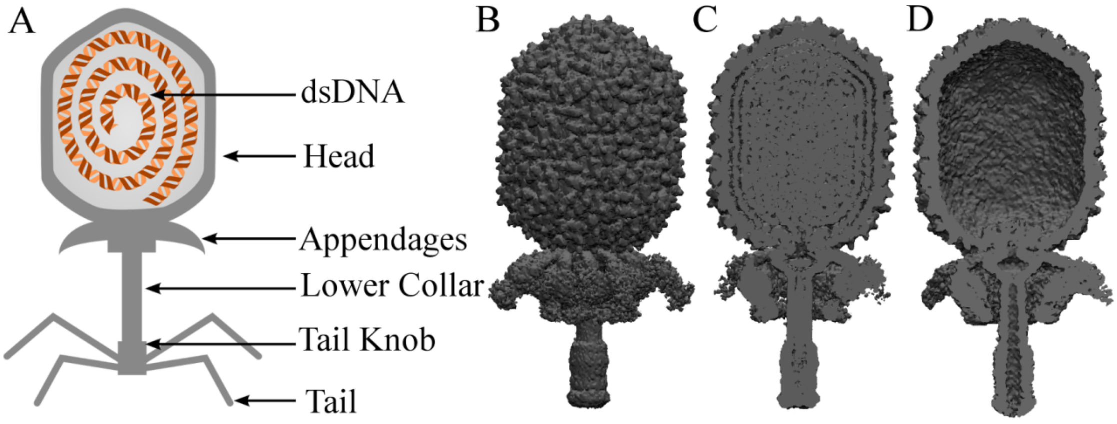


Figure 1.1: (A) Schematic representation of a φ29 phage particle with different regions indicated by arrows. Cryo-EM density maps of a φ29 bacteriophage in (B) full view, (C) cross-section view with filled DNA and (D) cross-section view after the DNA was emptied (3).

φ29 replicates through a lytic cycle, and this complete cycle is illustrated in Fig. 1.2. At first as shown in Fig. 1.2A, it infects a bacterium by injecting the dsDNA into the host cell (1). Subsequently, the dsDNA is replicated using the host enzymes (Fig. 1.2B) and the φ29 structural proteins are synthesized using the molecular machinery of the host bacterium (Fig. 1.2C). Further, as depicted in Figs. 1.2D–F, these newly synthesized proteins and the DNA are assembled to form new mature and infectious phage particles that are released upon lyses of the host cell (1).

During the φ29 assembly, the structural head proteins arrange as a precursor capsid, which is also called as procapsid (Fig. 1.2C). Then, a DNA packaging motor (Fig. 1.3A) is formed to package the newly synthesized viral DNA into the procapsid (6, 7). After completion of the DNA packaging, the motor is dismantled and subsequently, the other phage proteins assemble to form a mature phage particle (Fig. 1.2E).

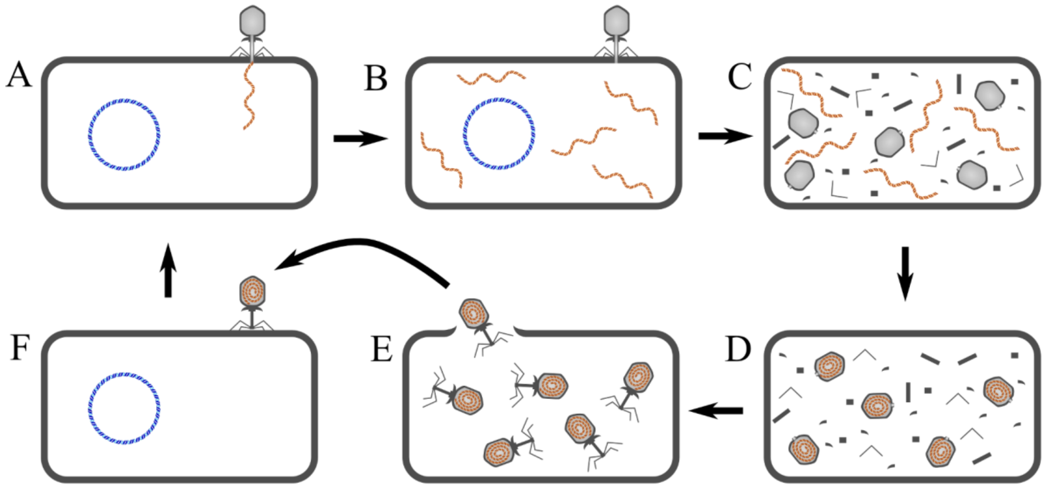


Figure 1.2: Schematic representation of the lytic cycle. (A) Bacteriophage infects the host cell by injecting the viral DNA. (B) Replication of the viral DNA by employing host enzymes. (C) Synthesis of viral structural proteins and assembly of the viral procapsid. (D) Viral DNA is packaged into the procapsid by the DNA packaging motor. (E) The motor dismantles and the complete bacteriophage is formed. Mature phages are released upon lyses of the host cell. (F) The bacteriophage binds to a new host cell for the next lytic cycle.

1.1. DNA packaging motor

The φ29 DNA packaging motor consists of three components, a head-tail connector protein (a product of gene gp10), a prohead RNA (pRNA) (8) and an ATPase enzyme (a product of gene gp16) (8-14). A sketch of the motor with and without a procapsid is illustrated in Fig. 1.3. The connector also being as a channel occupies the vertex of the procapsid and acts like a portal during the viral DNA packaging (Fig. 1.3A). The pRNA binds with the connector, the ATPase and the procapsid. The ATPase, which is located at the bottom of the pRNA, hydrolyzes ATP molecules to release the required energy for the DNA packaging (9, 15, 16) (Fig. 1.3B).

The motor transports and packages the viral DNA by transforming chemical energy into mechanical work. Further, the motor works against a maximum internal to external pressure difference of ~60 atm, which is generated by the negatively charged filled DNA (17). Consequently, it is one of the strongest molecular motors in the biological world. Such extreme properties make it a potential candidate for a motor used in nano-devices in the field of nanotechnology (7). Also, the motor has potential application as a molecular sorter in nano-pore based DNA sequencing devices or as a model system for anti-viral treatment of infections caused by herpes viruses, adenoviruses, and parvoviruses and pox viruses (7). One of the motor components, the prohead-RNA is demonstrated to be a suitable choice for a gene delivery system because it was used as carrier for ribozyme and antisense RNA to inhibit the Hepatitis B virus (18).

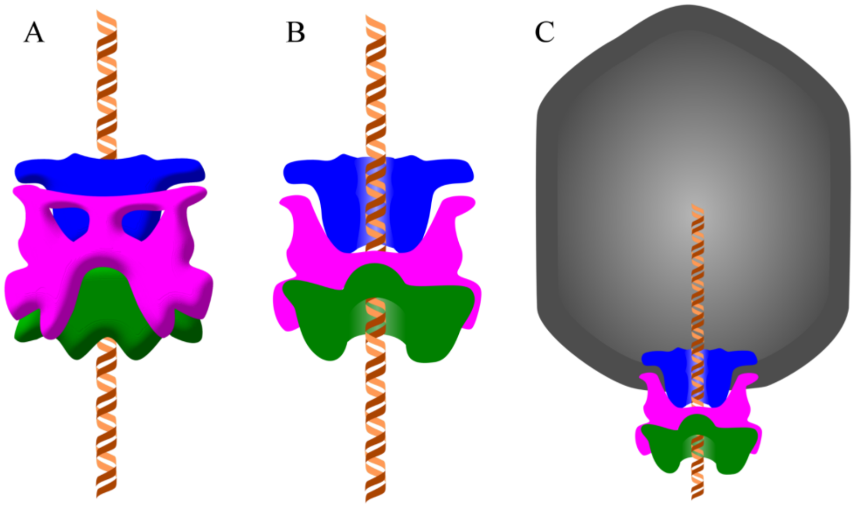


Figure 1.3: Schematic drawing of the DNA packaging motor. The motor's components are drawn in cartoon representation that was adapted from cryo-EM density difference maps (19): (i) connector in blue, (ii) pRNA in magenta, (iii) ATPase in green, (iv) DNA helix in brown, and (v) procapsid in gray. (A) A complete motor assembly including the bacteriophage procapsid. (B) Front view. (C) Longitudinal cross section.

1.2. Components of the DNA packaging motor

1.2.1. The connector

The connector is a truncated cone shaped dodecamer protein with a central channel (Fig. 1.4). Each monomer consists of 309 amino acid residues what sums up to 3708 residues in the whole connector. Four crystal structures of the connector in the absence of the DNA have been determined using X-ray diffraction methods until now (13, 20, 21). The structures of the N–terminal (1 to 15), C–terminal (285 to 309) and loop residues (225 to 239) were not resolved and are therefore missing in all available structures (Fig. 1.4).

The structure of the connector can be divided into five regions namely, upper, loop, middle, hinge, and bottom region (Fig. 1.4C). The upper region is located inside the procapsid. The middle helical region consists of 36 helices and this structural scaffold is conserved in connectors of other head-tail bacteriophage. The bottom region is located outside the procapsid and another motor component the pRNA is attached to this region. The hinge region acts as a bridge between middle and bottom region.

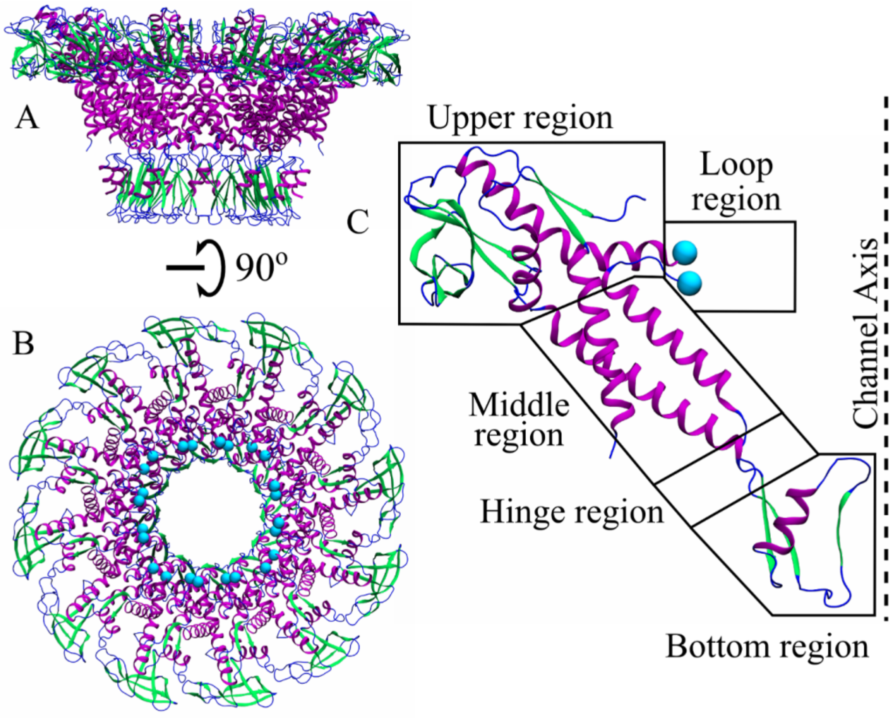


Figure 1.4: The crystal structure of the connector illustrated as cartoon and colored in accordance with the protein's secondary structure (alpha helix: *purple*, beta sheet: *green* and coil: *blue*). (A) Front and (B) top view of the connector. Blue spheres show the location of the missing loops. (C) The enlarged structure of one connector subunit with reference to the channel axis. The five regions of the connector are illustrated by polygonal boxes.

1.2.2. Prohead RNA

Prohead RNA (pRNA) (Fig. 1.5) is an essential component of the viral DNA packaging motor (8, 10, 11). *In vivo*, the pRNA consists of 174 nucleotides whereas *in vitro*, 120 nucleotides were shown to be functional and their sizes may vary depending on the respective purification method (22). The pRNA size variations affect neither the phage assembly nor the DNA packaging process (23). The pRNA forms a cyclic ring at the bottom of the connector in the DNA packaging motor (Fig. 1.2) (6). The number of pRNA molecules in this cyclic ring was controversial on the basis of several experimental studies (2, 6, 10, 13, 19, 24-27). Mutagenesis and ultracentrifugation studies demonstrated that the pRNA is present in the form of a cyclic hexamer (10, 26). Fluorescence and cryo-EM studies also supported the presence of a hexameric ring (6, 25). Recent, AFM imaging results confirmed that the pRNA is present as a hexameric ring *in vitro* (27, 28). In contrast, several cryo-EM reconstruction studies revealed that only a pentameric form can fit into the density maps of the pRNA that were obtained from the completely assembled DNA packaging motor (2, 13, 19, 24). Overall, the number of pRNA molecules present in the cyclic ring is yet to be resolved.

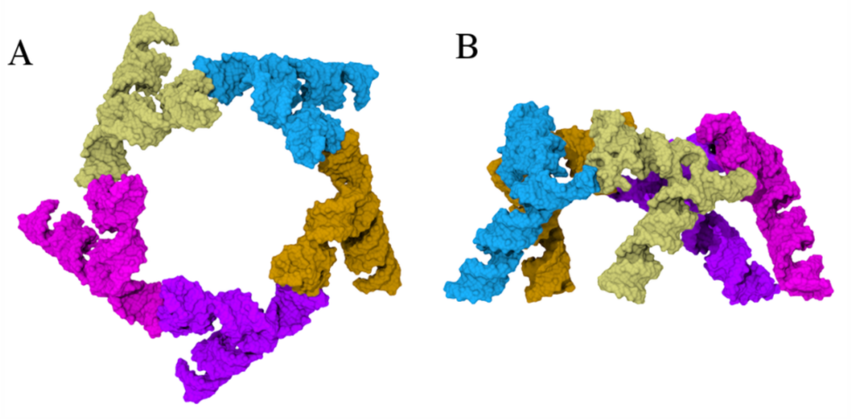


Figure 1.5: Computer model of the pentameric pRNA ring (13). The colors represent the five monomers. (a) Top view and (b) Side view.

1.2.3. ATPase

ATPase, a gene product of gp16, provides the required energy for the DNA packaging via hydrolysis of ATP molecules (9, 15, 16). Comparative genomics studies placed this ATPase in the Ftsk/HerA family of dsDNA translocases. This family itself is a part of the superfamily ASCE (Additional strand, conserved E), which contains functionally diverse proteins including AAA+ and RecA like ATPase proteins (29-31). Cryo-EM density maps revealed that the ATPase has a pentameric cyclic form when attached to the connector and the pRNA (Fig. 1.2) (13, 19). In contrast, several experimental studies showed that the ATPase is present as a hexameric cyclic ring during the DNA packaging process (32-34).

1.3. DNA packaging mechanism

The topologies and the arrangements of the DNA packaging motor's components are discussed in the previous sections. How this motor packages the DNA into the procapsid and how its components contribute to the packaging process? Also, the motor is one of the strongest biological motors, to understand its mechanism would be interesting and eventually helpful to employ in several nano-devices. Therefore, several experimental models have been developed to illustrate the packaging mechanism (6, 7, 16, 34-36). These models mostly consist of the purified procapsid, the pRNA, and the ATPase. Upon adding ATP molecules with initiation factor DNA-gp3, the DNA is packaged up to 90 % into the procapsid (37). After completion of DNA packaging, the infectious bacteriophage is formed in presence of viral structural proteins.

Optical tweezers experiments revealed that the DNA packaging initiates with a rate of ~165 bp/s that gradually decreases as the packaging progresses towards completion (38). Furthermore, the internal force, which resists DNA packaging, is as low as ~7 pN at one third filling, ~14 pN at half filling and it sharply rises to 110 ± 9 pN at the final stage of the packaging (38). In the completely filled procapsid, the pressure difference reaches up to ~60 atm (16, 17). Consequently, the procapsid must possess exceptional elastic properties to withstand such a large counter pressure, and indeed, a Young's modulus of ~1.8 GPa was measured in atomic force microscopy (AFM) experiments (4, 5). The procapsid's elasticity is comparable with that of other structural proteins such as silk fibers and collagen fibrils that also withstand extreme mechanical stress (39-41).

The role of the ATPase during DNA packaging has been studied using optical tweezers experiments and *in vitro* ATPase assays (9, 12, 15, 32-34, 42, 43). The motor component ATPase hydrolyzes ATP molecules to release energy. Early experimental results showed that one ATP is required to transport the two base pairs (BP) of the viral DNA (9). Subsequent optical tweezers studies revealed that 10 DNA BP are packaged under hydrolysis of four ATP molecules (15, 42). Furthermore, the motor specifically pushes the 5′–3′ strand of the dsDNA during the packaging process (43). In conclusion, the motor pushes ~2.5 BP DNA through its 5′–3′ strand with an expense of one ATP molecule and generates the force required to counter the large internal pressure.

The connector's role in the DNA packaging has been studied through mutagenesis, voltage ramping experiments, and sedimentation assays (34, 36, 44-46). The connector can embed into the lipid bi-layer membrane and acts as a channel for transporting the DNA across the membrane along the electrochemical gradient (36). This particular property enables to study the role of the connector in DNA transport via voltage ramping experiments, which revealed that it allows unidirectional transport of DNA across the lipid membrane (46). Further, sedimentation assays have shown that upon mutations of the conserved loop residues K234A.K235A.R237A, the packaged DNA leaked out of the filled procapsid due to the centrifugal force that was generated during the assays (45). Therefore, the connector was suggested to act as a one-way valve preventing DNA leakage caused by internal pressure (34, 44).

1.3.1. DNA packaging models

Since last four decades, several models have been proposed explain the mechanism of the DNA packaging process and reviewed in the publications (7, 47, 48). On the basis of several experiments performed in last two decades, many models eventually ruled out (35, 36, 42). The following four models have been widely got attention and backed by several experimental results (13, 34, 42, 49, 50).

1.3.1.1. The rotation model

The procapsid and the connector structures possess five-fold and six-fold symmetries, respectively, and a free rotation is possible because of this symmetry mismatch (49). The rotation model, which is depicted in Fig. 1.6A, proposes a mechanism as similar to a nut-bolt system, where the connector acts as a nut and rotates by 12° with respect to the procapsid. Subsequently, the DNA, which acts as a bolt, is transported into the procapsid.

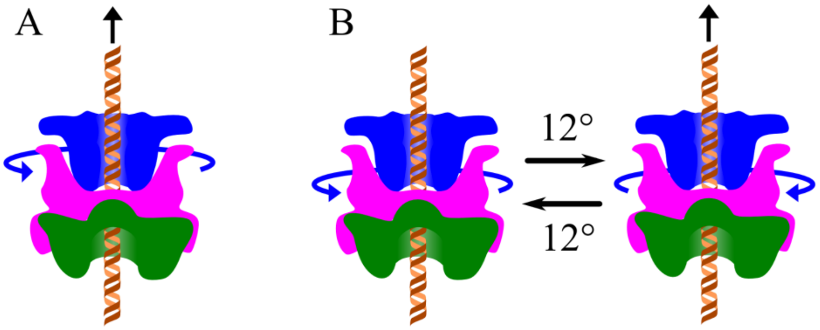


Figure 1.6: Illustration of the rotation and untwist-twist DNA packaging models. Three motor components, the connector (*blue*), the pRNA (*magenta*) and ATPase (*green*) are shown with the DNA (*orange-brown*) in Figs. (A-C). (A) Schematic representation for the rotation model, where blue and black arrows show the rotation of the connector and direction of the DNA transport, respectively. (B) Schematic representation for the untwist-twist model, where blue and black arrows depict the untwist-twist motion of the connector and the DNA transport, respectively.

To test the connector's rotation, single molecule fluorescence spectroscopy was used to study the packaging process (51). In this experiment, one of the connector's subunit was labeled by a fluorescent dye and the polarization of the emission was measured. The observed signals were incompatible with the connector's full rotation by a very high probability (51). Therefore, the full rotation of the connector with respect to the procapsid is ruled out and rendered this model unlikely. However, these experiments could not rule out a partial rotation of the connector because it might push the DNA by rotating and subsequently regaining the original orientation (51).

1.3.1.2. Untwist-twist model

The rotation model was further extended by Simpsons *et al.* in untwist-twist model, according to which, the connector rotates by a spring-like two steps mechanism as depicted in Fig. 1.6B (13). In the first step, the connector expands and untwists by 0.64 nm and 12°, respectively, under the consumption of ~50 $kJ.mol^{-1}$ energy, which is released from the hydrolysis of one ATP molecule. Simultaneously, the motor grasps the next two DNA base-pairs from the outside of the procapsid. In the second step, the connector relaxes by twisting-compression and pushes the two DNA base-pairs into the procapsid (Fig. 1.6B). Because, the connector's full rotation has been ruled out, it might partially rotate through this mechanism as it regains original orientation during the relaxation step (13, 51).

1.3.1.3. Push-roll model

The push-roll model was proposed on the basis of results which were obtained by the optical tweezers experiments (15, 16, 43, 50). Yu *et al.* proposed that the pentameric ATPase pushes 2.5 DNA basepairs by a “lever” into the procapsid via a conformational change that is driven by the energy generated by hydrolysis of one ATP as depicted in Fig. 1.7A (50). The DNA is packaged in two phases, first in a dwell phase; four ATP molecules bind to the four sub-units of the ATPase (15, 42). In the second burst phase, ATP molecules are hydrolyzed to release the required energy that changes ATPase conformations and thereby pushes the DNA into the procapsid (15, 42). During the burst phase, 10 DNA BP are packaged under the hydrolysis of 4 ATP molecules (2.5 base-pairs per step). Simultaneously as illustrated in Fig. 1.7B, the DNA rotates by –30° with respect to the motor and rolls by 18° on the motor’s lumen wall, that results in a total –12° DNA rotation (50, 52).

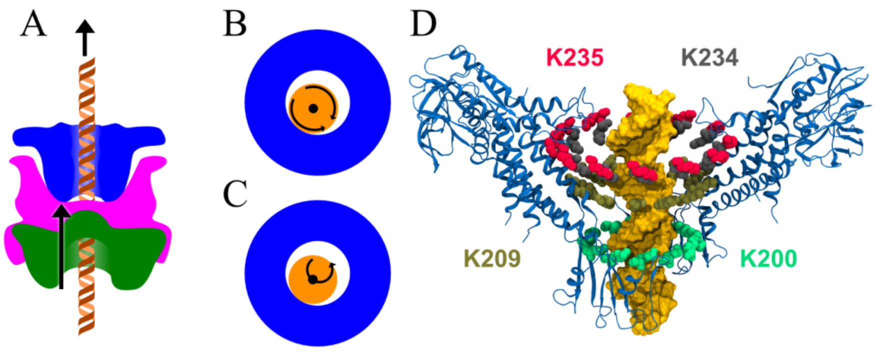


Figure 1.7: Illustration of the push-roll and one-way revolution DNA packaging models. Colors are described in Fig. 1.6. (A) Scheme shows the DNA pushing (*arrow*) performed by the ATPase (*green*) as proposed in both push-roll model and one-way revolution model. (B) According to the push-roll model, the DNA helix (*orange*, *top-view*) rotates and rolls (*arrows*) inside the connector channel (*blue, top-view*) during the DNA packaging. (C) According to the one-way revolution model, the DNA helix (*orange*, *top-view*) revolves (*arrow*) inside the connector channel. (D) Illustration of the four electropositive rings K200, K209, K234 and K235 in the connector (*blue*) which are proposed to be essential for DNA revolution and observed DNA transport step-size of ~2.5 basepairs. The loop residues K234 and K235 were missing in the crystal structures and therefore these residues are modeled in the presented snapshot of the connector.

1.3.1.4. One-way-revolution model

Recently, on the basis of the several fluorescence and voltage ramping experimental results (25, 27, 32, 33, 44, 46), a one-way revolution model was proposed by Zhao *et al.* (34). According to this model, the hexameric ATPase transports viral DNA under hydrolysis of ATP molecules as depicted in Fig. 1.7A. During the transport, the DNA revolves without rotation inside the connector channel (Fig. 1.7C). The connector as a one-way valve allows unidirectional transport of the DNA and prevents DNA leakage what is expected due to the internal pressure. Additionally, the connector’s four electropositive lysine rings K200, K209, K234, and K235 interact with DNA (Fig. 1.7D) and induce a DNA packaging step-size of 2.5 basepairs/step which was observed in optical tweezers experiments (15, 42). Further, the model proposes that these lysine residues are separated by ~9 Å inside the connector channel and consequently, ~2.6 (9/3.4) basepairs are present between the successive rings by considering a DNA helical pitch of 3.4 Å. Moreover, these lysine residues also facilitate the proposed DNA revolution without rotation during the packaging process.

1.4. Aims of this study

Despite of many experimental studies on the DNA packaging motor, several questions about the connector's role in packaging mechanism remain open. The three proposed and previously introduced DNA packaging models discuss the role of the connector in the packaging process. Here, I focused on the connector's role in the DNA packaging process at atomistic scale and examined the proposed packaging models.

The untwist-twist model requires the connector to exhibit an elastic reversible deformation to package the DNA into the procapsid by spring-like oscillations. Additionally, as a part of the procapsid, the connector is expected to exhibit similar mechanical properties by whose procapsid withstand a maximum pressure difference of ~60 atm. Therefore, I determined the mechanical properties of the connector to understand the underlying mechanism of pressure resistance and probed the spring-like motion. This motion that drives the transport of two DNA base-pairs consists of a 12° untwisting and 0.64 nm expansion in the connector under consumption of ~50 kJ mol^{-1} energy released from hydrolysis of one ATP molecule. Therefore, this proposed motion was examined by calculating the required energy to drive this motion.

Both, the one-way revolution and the push-roll model require the presence of a DNA helix in a form of straight rod inside the connector channel during the DNA packaging process. However, structure of the connector in complex with the DNA is not available and the impacts of the DNA on the connector and *vice versa* are ye unknown at atomic resolution. In recent fluorescence spectroscopy experiments performed on the T4 bacteriophage DNA packaging motor, the DNA was observed to be compressed by ~22-24 % inside the motor channel during the packaging process (53). Additionally, in cryo-EM density maps of the connector, the bottom region of the connector appeared to be deformed with respect to the crystal structure after completion of DNA packaging (3). The deformations in either the connector or the DNA would lead to deviation from the required structural features that are proposed in the packaging models. Therefore, I determined the impact of the DNA on the connector and *vice versa* by characterizing their conformational changes.

Next, the one-way revolution model requires the connector to function as a one way valve that only allows unidirectional transport of the DNA. This model was examined by determining required forces to push the DNA into the procaspid and to pull the DNA out of the procapsid through the connector channel. Furthermore, the connector loop's residues were proposed to be essential for this one-way valve function. Roles of these loop's residues in preventing the DNA leakage against a large counter pressure were studied to understand the one-way valve mechanism. The model further proposes that the connector channel's electropositive rings facilitate the transport of the DNA by a revolution mechanism with ~2.5 base-pairs per step; therefore, the role of these rings was studied to validate this model.

Further, the one-way revolution and the push-roll model require revolution and rolling of the DNA inside the connector channel, respectively. These two types of DNA motions are likely to occur in presence of a gap between the DNA helix and the connector channel. Indeed, the gap is present when a modeled B-DNA is placed inside the crystal structure of the connector; however, interaction of the DNA with the channel might deform both structures. Therefore, I probed the gap with in dependence of a likely deformation of both the DNA and the channel to test the proposed revolution or rolling motion of the DNA.

To achieve these aims, I performed all-atom explicit solvent molecular dynamics (MD) simulations, which has been successfully employed to understand bio-chemical processes at atomistic scale (54-70). Equilibrium and force-probe MD simulations have been applied to determine the elastic properties of bio-macromolecules (58, 62, 64, 69, 70) and therefore, these methods were used to study the connector's elastic properties. Equilibrium MD simulations have been used to study large-scale conformational changes in the bio-macromolecules (54, 56, 57, 68) and therefore, simulations are suitable to characterize structural changes in connector-DNA complex. The energy required to deform the connector as proposed in the untwist-twist model was calculated by using umbrella sampling MD simulations combined with the weighted histogram analysis method (61, 63, 67). Force-probe simulations analogous to single-molecule affinity force microscopy or optical tweezers method (59, 60, 62, 64, 69, 70), were utilized to determine the force required to transport the DNA through the connector channel.

Combining above described computational methods, I investigated the connector's role in the DNA packaging process at atomistic scale. The present study is likely to contribute to understanding the connector's role at atomic resolution during DNA packaging process. Particularly, to study the mechanism by which the connector minimizes the DNA leakage against such a large pressure of ~60 atm, would be very interesting. This thesis broadens our knowledge about these types of single-molecule valves and motors, which can be employed in nano-devices for gene-delivery, drug-delivery or transport across lipid membranes. Therefore, this thesis is of general interest in the field of viral assembly and nano-technological applications.

Concepts and Theory

Chapter 2

2.1. Molecular Dynamics Simulations

Molecular dynamics (MD) simulations have emerged as a leading theoretical method to study the microscopical and thermodynamical properties of molecules. The MD method models a real time evolution of molecular configurations by mimicking physical environment using computer processors. The method becomes a choice to understand the bio-chemical processes in atomistic or molecular scale resolutions. The MD simulations have been employed to study the role of bio-macromolecules such as proteins, lipid membrane, DNA, RNA, and small organic molecules in biological processes. The theoretical background of the MD simulations is extensively discussed in literature (71-75). During the MD simulations, the motion of the atoms as a function of the time is computed using the models that are developed on the approximations of the quantum mechanical (QM) methods.

2.1.1. Schrödinger equation

In the QM methods, both nuclei and electrons dynamics can be computed by solving the time dependent Schrödinger equation, which is given as follows,

$$i\hbar\frac{\partial\psi(\mathbf{R},\mathbf{r},t)}{\partial t} = \mathcal{H}\psi(\mathbf{R},\mathbf{r},t) \qquad 2.1$$

where, ψ is the time dependent wave function, **R** and **r** are positions of the nuclei and electrons in space, respectively and t is time. The Hamilton operator $\mathcal{H}$ is related to the electrons and nuclei, and it is the sum of the operators for the kinetic and coulomb electrostatic energy. The Hamilton operator $\mathcal{H}$ further can be separated as follows:

$$\mathcal{H} = T_e + V_{ee} + V_{eN} + V_{NN} + T_N \qquad 2.2$$

where $T_e$and $T_N$is kinetic energy operator for electrons and nuclei. V_{ee}, V_{eN} and V_{NN} is columbic energy operator for the electrons only, between the electrons and the nuclei, and the nuclei only, respectively.

2.1.2. Born-Oppenheimer approximation

Solving above Schrödinger equation (Eq. 2.1) is computationally expensive and not practically feasible for small molecules with more than few atoms. The Born-Oppenheimer approximation alleviates this problem; mass of the electron is much less than mass of the nuclei and consequently, the wave function can be solved separately in steps for electronic motions, nuclear vibrations, and molecular rotations. Using this approximation, one can compute electron and nuclei dynamics separately for a molecule.

Despite this approximation, solving the wave functions is still not practically feasible for macro bio-molecules with more than thousands of atoms. This problem is addressed by two further approximations. First, atoms are assumed to be classical particles that follow Newton's laws of motion. Second, the potential energy surface is approximated by the collection of simple potential functions. These functions combined together form molecular mechanics force field, modeled by approximating the potential energy surface, which is previously calculated by QM methods. These functions are used during the MD simulations to calculate potential energy and subsequently, motions of the atoms are computed using the Newton's second law of motion. Therefore, the usage of force-field significantly speeds up the potential energy calculations for the macromolecules.

2.1.3. Molecular mechanics force field

A force field is a collection of potential energy function coefficients, which are known as the force-field parameters. These coefficients are determined by using QM calculations, and this procedure is called as force-field parameterizations. Force-fields such as MMFF and MM4 have been developed for small organic molecules while AMBER, CHARMM, OPLS and GROMOS were developed for bio-macromolecules such as proteins, lipids, DNA and RNA. These force-fields mainly consist of two sets of potential functions,

$$V_{total} = V_{bonded} + V_{non-bonded} \tag{2.3}$$

where V_{total} is total potential energy of the molecule, V_{bonded} and $V_{non-bonded}$ are potential energies originating from bonded and non-bonded atom-pairs.

Bond, angle, proper-dihedral and improper-dihedral functions contribute to V_{bonded} ,

$$V_{bonded} = V_{bond} + V_{angle} + V_{dihedral} + V_{improper} \tag{2.4}$$

Most widely used functions of these four bonded terms are illustrated in Fig. 2.1A-D. van der Waals and electrostatic interactions contribute to $V_{non-bonded}$,

$$V_{non-bonded} = V_{vdW} + V_{electrostatic} \tag{2.5}$$

As can be seen in Fig. 2.1E-F, the Lennard-Jones potential is used for V_{vdW} and the coulomb potential is used for $V_{electrostatic}$. The V_{vdW} values rapidly decrease with increase in distance between atom-pairs (Fig. 2.1E). In contrast, electrostatic interaction decays gradually with the distance (Fig. 2.1F). A cut-off distance is used to reduce the number of non-bonded atom-pairs that are considered for the calculation of the non-bonded interactions. The calculation time decreases because of the reduction in atom-pairs and this speed up the simulation. For the electrostatic interaction, Coulomb equation is used for the atom-pairs that are within the cut-off distance, and particle mesh Ewald method (76) is used for the atom-pairs that fall outside of the cut-off.

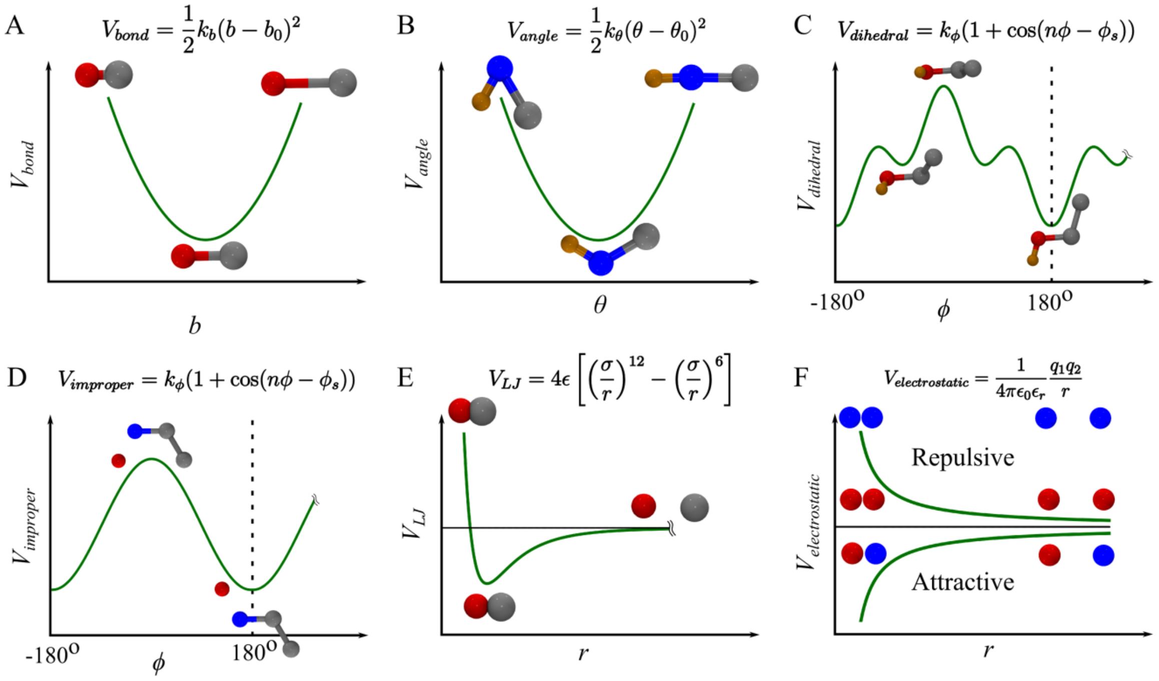


Figure 2.1: Typical potential energy functions used in MD simulations (72). (A) Bond stretching harmonic potential, where k_b is force-constant and b_0 is bond-length with minimum potential. (B) Harmonic angle potential, where k_θ is force-constant and θ_0 is angle with minimum potential. (C) Proper dihedral periodic potential, where k_ϕ is force constant, n is termed as multiplicity and ϕ_s is phase factor. Multiplicity is total number of energy minima, and phase factor is dihedral angle at the respective energy minima within a range of $-\pi$ to π. (D) Improper periodic potential is used to maintain planer conformation of the molecule and is similar to the proper dihedral periodic potential (E) van der Waals interactions V_{vdw} are modeled using the Lennard-Jones potential function V_{LJ}, where σ is the separation between atoms at zero energy and ε is minimum energy value. (F) Electrostatic interactions are modeled by the coulomb potential, where q_1 and q_2 are partial charges of the atoms, ε_0 is vacuum permittivity and ε_r is relative permittivity of the medium.

Force acting on each atom is derivative of these six potential functions,

$$F = -\frac{dV}{dr} \qquad 2.6$$

After calculating the potential energy and forces, next step in simulation is to calculate the motions of atoms with respect to time.

2.1.4. Dynamics of the atoms

Atomic motions are computed using Newton's second law of motions, according to which force F is equal to mass multiplied by rate of change in velocity or acceleration: $F = ma$, where m is mass and a is acceleration. The trajectory of atoms is computed by solving following differential equation,

$$\frac{d^2x_i}{dt} = \frac{Fx_i}{m_i} \tag{2.7}$$

where x is positions of the particles at time t, F is total force acting on the particles and m is the particle mass. However, analytical solution of this equation is difficult to obtain for bio-macromolecules due to many-body problem, which is caused by large number of interacting atoms. To solve this problem, finite difference methods have been developed to integrate equation of the motion (Eq. 2.6). Three most widely used finite difference methods are Verlet (77), Leapfrog (78) and Velocity-Verlet (79) integration schemes. Motions of the atoms are obtained by computing velocity and positions with respect to the time using one of these methods.

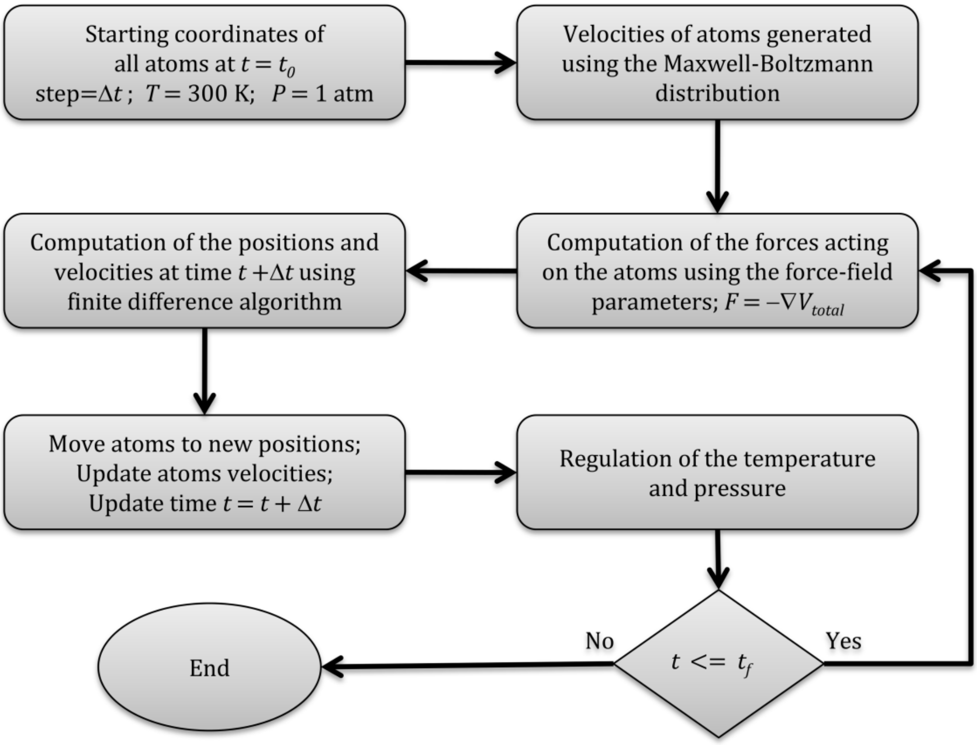

Starting coordinates of all atoms at $t = t_0$ step=Δt; $T = 300$ K; $P = 1$ atm
Velocities of atoms generated using the Maxwell-Boltzmann distribution
Computation of the forces acting on the atoms using the force-field parameters; $F = -\nabla V_{total}$
Computation of the positions and velocities at time $t + \Delta t$ using finite difference algorithm
Move atoms to new positions; Update atoms velocities; Update time $t = t + \Delta t$
Regulation of the temperature and pressure
$t <= t_f$
No
Yes
End

Figure 2.2: A flow-chart illustrating steps in a MD simulation. The step-size of the simulations is denoted by Δt. The starting, current and final time are denoted by t, t_0 and t_f, respectively.

A summary of MD simulations steps are illustrated in a flow-chart (Fig. 2.2). Biological experiments are preformed at constant temperature and pressure; therefore to mimic these conditions, temperature and pressure are regulated during simulations by coupling molecular system to an external bath (55, 80). The step-size of the MD simulation is typically kept at 1 fs, because it is the smallest time scale of vibrations in bonds containing hydrogen atoms. Step-size can be extended to 2 fs by constraining all bonds containing hydrogen atoms using LINCS algorithm (81, 82). The step-size can be further extended to 4 fs by constraining the angular-bond vibrations involving hydrogen atoms.

2.1.5. Force-probe MD simulations

2.1.5.1. Pulling/Pushing simulations

During MD simulations, to pull or push a molecule or group of atoms in specific directions, an external force is applied on the center of mass of the respective groups in the given direction. The applied force could be either constant or vary during the simulations. The most widely used method to apply an external force is via a "virtual spring" with a potential, which is given as follows (60),

$$V(x,t) = \frac{1}{2}k_s(x - x_0 - vt)^2 \qquad 2.8$$

where k_s is spring-constant, x_0 is initial-offset, and v is pulling rate of the spring. As illustrated in Fig. 2.3, one tip of the virtual spring is attached with center of mass and an opposite tip of the spring is pulled in the given direction. The resultant force exerted on atoms or center of mass by this spring is given as follows,

$$F(x,t) = -k_s(x - x_0 - vt) \qquad 2.9$$

Two examples of force-probe simulations that are performed on connector-DNA complex are shown in Fig. 2.3.

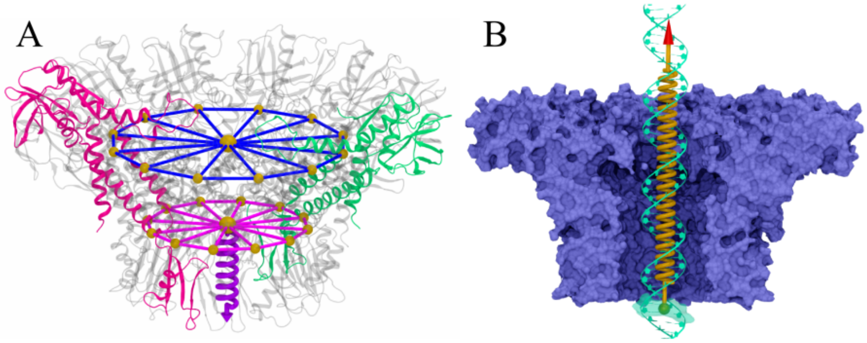


Figure 2.3: Two types of force-probe simulations were performed on the connector and the DNA. (A) Two opposite subunits of the connector are shown in red and green while remaining subunits are kept transparent for visibility. Brown spheres shown at center of both upper and bottom discs depict the center of mass formed by helices residue's Cα atoms (*brown spheres at periphery of discs*) from 12 subunits. The middle helical region of the connector was expanded by pulling center of mass of the lower discs with respect to the upper discs by a virtual 'spring' (*violet*). (B) The connector (*blue*) with the DNA (*green*) in the channel is shown. Center of mass (*green sphere*) of two DNA basepairs was pulled in upward direction by a virtual "spring" (*orange*).

2.1.5.2. Rotation simulations

To rotate a molecule around an axis, an external torque is applied during the MD simulation. In enforced rotation simulations, at first, reference virtual positions of the atoms are rotated with a constant angular velocity around an axis and subsequently, the atoms are allowed to attract towards this virtual positions via a “virtual spring” potential. As a result, the group of atoms rotates around a given rotational axis, which either can be fixed or flexible. Several variants of the rotational potentials have been developed and discussed in a publication by Kutzner *et al.* (64).

2.2. Elastic properties from the equilibrium fluctuations

2.2.1. Elastic spring constants

The elastic constants can be computed under assumptions that the energy landscape is harmonic for the given modes of motions or degrees of freedom. The free energy landscape for a given two degree of freedoms (θ and L) during equilibrium fluctuations under the harmonic approximation is written as follows, $G(\theta, L) = -k_B T \ln[p(\theta, L)]$ where, k_B is and T are the Boltzmann constant and temperature respectively. Using the harmonic approximation, the probability density function $G(\theta, L)$ is given as,

$$p(\theta, L) \propto \exp\left(-\frac{1}{2}[\,(\theta-\bar{\theta})\,, \quad (L-\bar{L})\,]\,\mathbf{C}^{-1}\begin{bmatrix}(\theta-\bar{\theta})\\(L-\bar{L})\end{bmatrix}\right) \tag{2.10}$$

where, $\bar{\theta}$ and $\bar{L}$ are average value of the respective degree of freedoms; **C** is covariance matrix of two degree of freedoms (θ and L). Harmonic free energy for two degree of freedoms can be expressed as follows,

$$G(\theta, L) = \tfrac{1}{2}\, K_\theta (\theta-\bar{\theta})^2 + \tfrac{1}{2}\, K_L (L-\bar{L})^2 + K_c(\theta-\bar{\theta})(L-\bar{L}) \tag{2.11}$$

where K_θ and K_L are two spring constants for respective degree of freedoms, and K_c is coupling constant between the two degree of freedoms. By comparing above three equations, these constants can be written as follows,

$$\begin{bmatrix} K_\theta & K_c \\ K_c & K_L \end{bmatrix} = k_b T\, \mathbf{C}^{-1} \tag{2.12}$$

Therefore, using the harmonic approximation of underlying energy landscape, the elastic properties such as spring constants can be calculated using the above equation from the equilibrium MD simulations. This equation can also be extended for more than two degrees of freedom.

2.2.2. Young modulus of elasticity

Elastic spring constant depends on the geometrical shape of an object. To compare elasticity of two objects with different geometry, a shape independent parameter, Young's modulus of elasticity is calculated. For estimating the connector's Young's modulus (Y), it was considered similar to a truncated hollow cone of homogenous elastic material with varying channel radius (Fig. 2.4). The change in length δ upon applying an axial force F (*arrow* in Fig. 2.4) is given as follows (83),

$$\delta = \int_0^L \frac{F}{YA(x)} dx \tag{2.13}$$

Where, $A(x)$ is cross-sectional area normal to a channel axis. According to Hooke's law for a spring oscillations, force $F = K_L\ \delta$, and therefore, young's modulus Y can be written as follows,

$$Y = \int_0^L \frac{K_L}{A(x)} dx \tag{2.14}$$

which yields, for an assumed hollow truncated cone having variant diameters (Fig. 2.4),

$$Y = \frac{2k_L}{\pi} \frac{L}{D_b d_a - D_a d_b} \ln \left| \frac{(D_a + d_a)(D_b - d_b)}{(D_a - d_a)(D_b + d_b)} \right| \tag{2.15}$$

where D_a and d_a are narrow end exterior and interior diameters, respectively; D_b and d_b are wide end exterior and interior diameters of the truncated cone, respectively (sketch shown in Fig. 2.4). The derivation of the Young's modulus is described in Appendix 8.1.

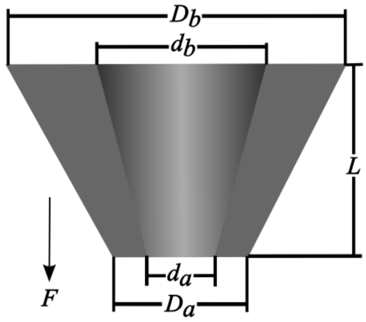


Figure 2.4: Geometrical characterization of the truncated hollow cone model used for calculating the Young's modulus of elasticity. The sketch shows the truncated hollow cone of length L, where D_a and d_a denote narrow end exterior and interior diameter, respectively; D_b and d_b denote wide end exterior and interior diameter, respectively.

2.3. Principal component analysis

Structure of the bio-macromolecules vibrates and changes at the room temperature. Some of these fluctuations are large whereas most of them are small such as bonds, angle and dihedral vibrations. Few of these structural fluctuations and transitions are relevant to the biological functions. Identifying these transitions of interest through trajectory visualization is often difficult and error-prone, and therefore a method is required to filter large conformational changes from the MD trajectories. To discriminate and quantify large and small fluctuations, principal component analysis (PCA) has been employed on the MD trajectories (54, 56, 84-90).

In this method, covariance matrix **C** is calculated for coordinates of the atoms:

$$\mathbf{C} = \langle (\mathbf{x}(t) - \langle \mathbf{x} \rangle)\,(\mathbf{x}(t) - \langle \mathbf{x} \rangle)^{\mathrm{T}} \rangle \qquad 2.16$$

Where, $\langle\ \rangle$ shows an average over complete trajectory. $\mathbf{x}(t)$ are positions of atoms with function of time and **x** is average positions of atoms in the trajectory. The obtained covariance matrix is a symmetric matrix, which is subsequently diagonalized and a set of eigenvectors and eigenvalues are obtained:

$$\mathbf{C} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{\mathrm{T}} \qquad 2.17$$

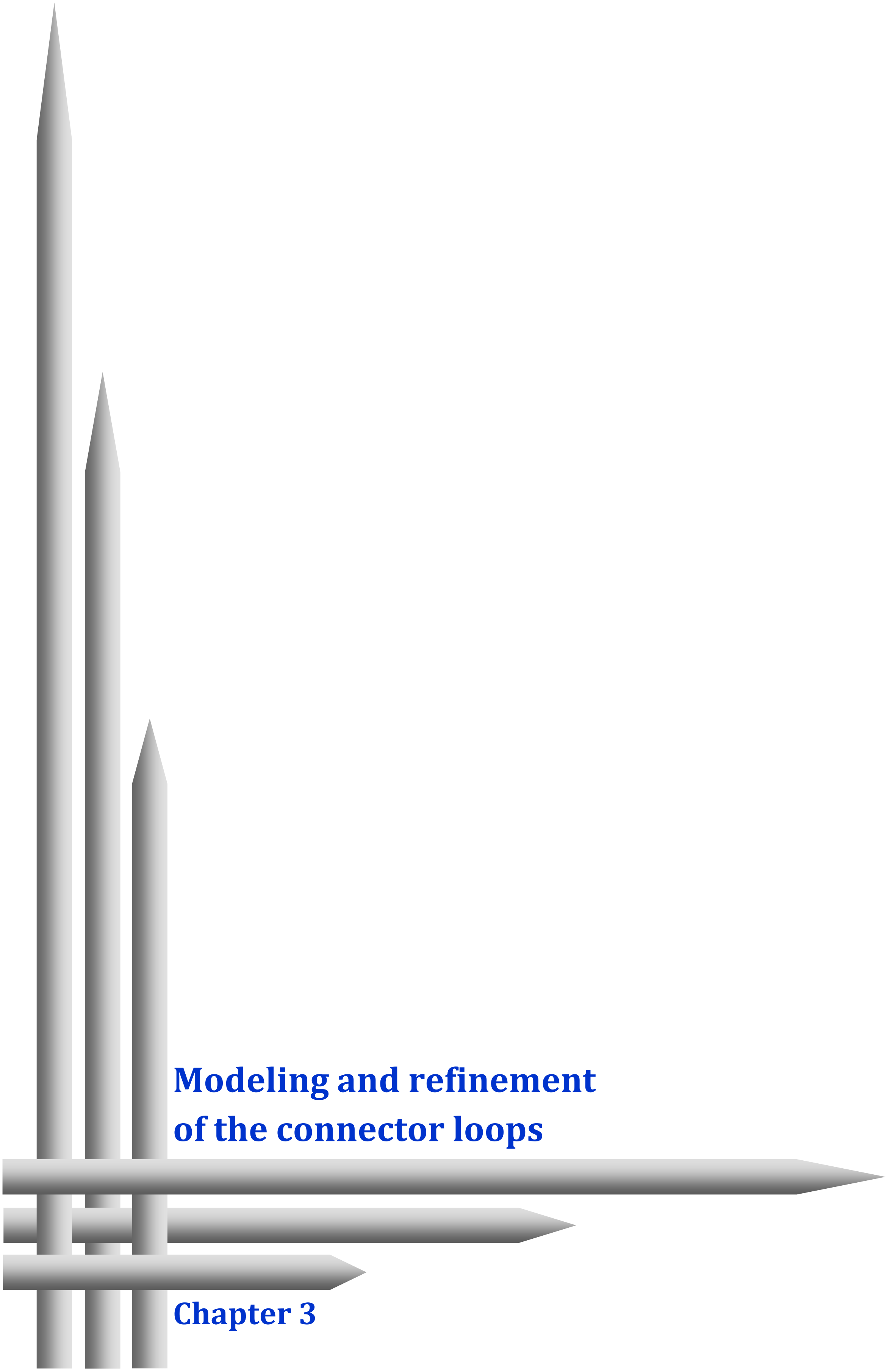
where $\boldsymbol{\Lambda}$ and $\mathbf{T}$ denote eigenvalue and eigenvector matrix. The eigenvectors describe direction of atomic displacement in high dimensional space and corresponding eigenvalues show mean square fluctuation of the total displacement. These eigenvalues are sorted in descending order and therefore, first few eigenvectors describe the largest conformational change obtained from the MD simulation. The trajectory is then projected onto the eigenvectors to obtain principal coordinates $p_i(t)$ that quantify the conformational fluctuations for the respective eigenvector μ_i :

$$p_i(t) = \mu_i . (\mathbf{x}(t) - \langle \mathbf{x} \rangle) \qquad 2.18$$

The MD trajectories can be further filtered to visualize the conformational fluctuations along a given eigenvector:

$$\mathbf{x}_i^{'}(t) = p_i(t) . \mu_i + \langle \mathbf{x} \rangle \qquad 2.19$$

By employing PCA, the slowest and largest conformational transitions are extracted and quantified from the MD trajectories. PCA significantly reduces the number of dimensions to study the conformational fluctuations as first few principal components describe the functionally relevant conformational transitions (87, 88, 90-93).



Modeling and refinement of the connector loops

Chapter 3

3.1. Background

The loop region of the connector is essential for the DNA packaging because these 12 loops restrict the packed DNA inside the procapsid during the packaging process (44, 45). However, the loops' structures A230-S244 are not resolved in the available X-ray crystal structures (Fig. 3.1) due to their high flexibility (13, 20, 21). Moreover, these connector's structures were resolved in the absence of DNA and do not contain the DNA. The loops structure and its interactions with the DNA are crucial to model and to refine for probing functional roles of the loops using the MD simulations. To achieve this aim, 12 loops and DNA were modeled into the connector as illustrated in Fig. 3.2 and further, the loops structure was refined in the presence of the DNA.

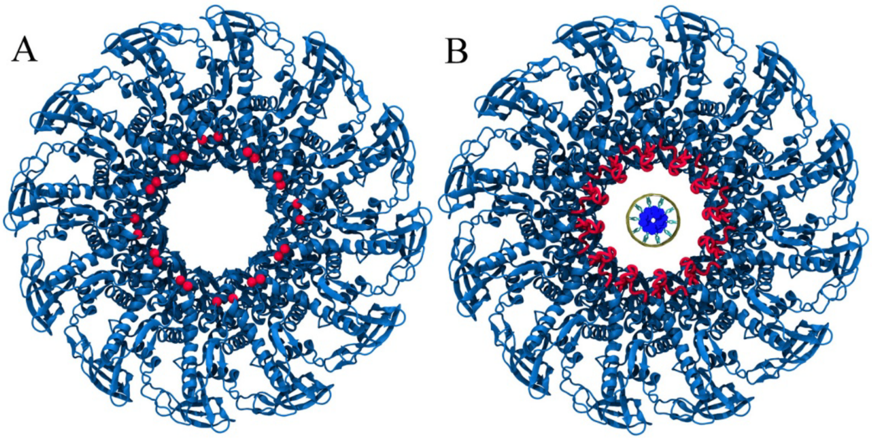


Figure 3.1: Crystal structure of the connector in top view. (A) Red spheres denote the location of the missing loops. (B) Modeled loops (*red*) are not in contact with the DNA (*yellow* and *blue* ring at the center of the channel).

3.2. Methods

An initial structural model of one loop was generated for one of the twelve symmetrical missing loops using the ArchPred server (94). This seed structure was then replicated (according to the 12-fold symmetry of the connector) and manually integrated within all other remaining sub-units after aligning the seed structure (see Fig. 3.1B). In addition, four of the twelve subunits lacked residues (Q166-L169) in the bottom region, and these residues were modeled using Modeller program (95).

Next, first 60 nucleotides (ATG GCA CGT AAA CGC AGT AAC ACA TAC CGA TCT ATC AAT GAG ATA CAG CGT CAA AAA CGG) of a φ29 gp10 gene (NCBI Reference Sequence: NC_011048.1) were modeled in the form of B-DNA by using NAB package (96). Subsequently, the DNA was placed inside center of the connector channel in such a way that channel axis and DNA helical axis aligned on each other.

Table 3.1: Heating-cooling cycles performed in SAMD simulations. The connector loops were heated up and cooled down consecutively 40 times during 10 ns of MD simulations. Each cycle consisted of 252 ps with two and eight steps of heating and cooling, respectively.

Temperature (K)	Time period (ps)	Cumulative time (ps)	Temperature (K)	Time period (ps)	Cumulative time (ps)
300	20	20	600	20	162
300-600	2	22	600-500	2	164
600	22	44	500	20	184
600-1000	2	46	500-450	2	186
1000	50	96	450	20	206
1000-800	2	98	450-400	2	208
800	20	118	400	20	228
800-700	2	120	400-350	2	230
700	20	140	350	20	250
700-600	2	142	350-300	2	252

Next, the initial molecular system was prepared for the Simulated Annealing Molecular Dynamics (SA MD) simulations: The connector-DNA complex was placed at the center of a dodecahedron box, 241 067 water molecules were added in this box, and neutralized by the addition of 202 sodium ions. The system's potential energy was minimized by steepest-descent algorithm, and subsequently, it was heated during a 500 ps constant volume simulation with a 1 fs time step. Then, pressure was equilibrated at 1 atm during a 1 ns NPT simulation with a 2 fs time step. In these two simulations, all heavy atoms were restrained at the starting positions by a force constant of 1000 kJ·mol·nm^{-2}. Afterward, a 10 ns SA MD simulation was performed with 40 annealing cycles by a rate of 252 ps per cycle (Table 3.1). In each cycle, the loops were heated to a temperature of 1000 K in two steps and subsequently cooled down to 300 K in seven steps (Table 3.1). Only loops were allowed to move during the SA MD simulations, whereas heavy atoms of the remaining connector-DNA complex were restrained by the force constant of 1000 kJ·mol·nm^{-2} to preserve the secondary structure at high temperatures.

Next, to quantify the largest conformational transitions during annealing, principal component analysis (PCA) was performed on the last 6 ns of the SAMD trajectory and projections of two largest principal components (PCA subspace) were plotted as shown in Fig. 3.2 (54, 56, 85, 86, 97). ive diverse conformations of 1000 K temperature that are indicated by P1, P2, P3, P4 and P5 in Fig. 3.2A, were chosen on the PCA Subspace. Further, five cooling simulations were performed to cool these conformations from 1000 K to 300 K during 6 ns of time. Subsequently, these conformations were equilibrated for 1 ns at 300 K temperature and five clusters of structures were obtained (indicated by $P1^c$, $P2^c$, $P3^c$, $P4^c$ and $P5^c$ in Fig. 3.2).

The SA MD and the cooling simulations were performed using GROMACS 4.0.7 package (98). BER ff99SB (99) and ParmBSC0 (100) force fields were used for the connector and the DNA, respectively. The temperature was regulated by using the Berendsen temperature coupling during annealing and cooling. The pressure was maintained at 1 atm by using the Berendsen pressure coupling (80). Long range electrostatic interactions were computed by the PME method with a grid spacing of 1.2 Å and a fourth order of cubic interpolation (76). Short range non-bonded interactions were computed for the atom pairs within a distance of 10 Å.

3.3. Results and discussion

The largest conformational transitions were captured via two-dimensional projection (PCA subspace) of the two largest principal components with reference to temperatures as illustrated in Fig. 3.2A. A characteristic semi-circle shape which is depicted as an arrow in Fig. 3.2A was observed on the PCA subspace. This particular shape shows that loops were not able to explore the entire conformational space during the SAMD simulation (101). Moreover, because of the high temperature (1000 K) and a short cooling time (150 ps) during annealing, the conformations are most likely to be present in the high-energy regions of the landscape. Next, I asked a question, are these conformations near to the native conformation?

If the obtained high-energy conformations are near to a native conformation, the loops likely to evolve toward the native conformation during a slow cooling simulation. To this aim, five cooling simulations were performed using five high temperature (1000 K) conformations (P1, P2, P3, P4 and P5 in Fig. 3.2A) and the conformational evolutions are indicated by straight arrows in Fig. 3.2A. As seen in Fig. 3.2A, five diverse loops' conformations appeared to be converging toward the same region on the PCA subspace, but did not attain same conformation due to the time constraint. This result suggests that these loops are likely to be near the native conformation. Further, as indicated by five blue clusters $P1^c$, $P2^c$, $P3^c$, $P4^c$ and $P5^c$ in Fig. 3.2, how are these cooled conformations similar to each other? To check the similarity between these cooled structures, root mean square deviations (RMSD) of these five clusters were calculated with respect to the central structure (Fig. 3.2B). The central structure was earlier determined by clustering RMSD values of these five clusters $P1^c$, $P2^c$, $P3^c$, $P4^c$ and $P5^c$. As seen in Fig. 3.2B, the RMSD values are less than 0.15 nm, and quite small for such a large structure. Therefore, these cooled structures are considered to be very similar to each other, and the obtained central structure is likely to be near of the native conformation.

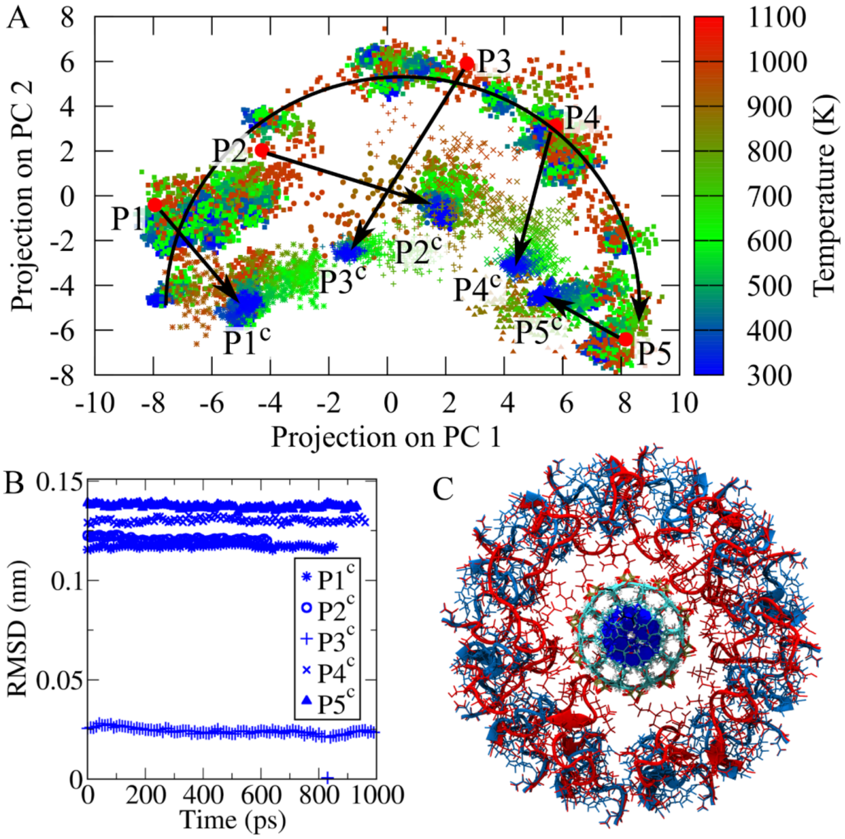


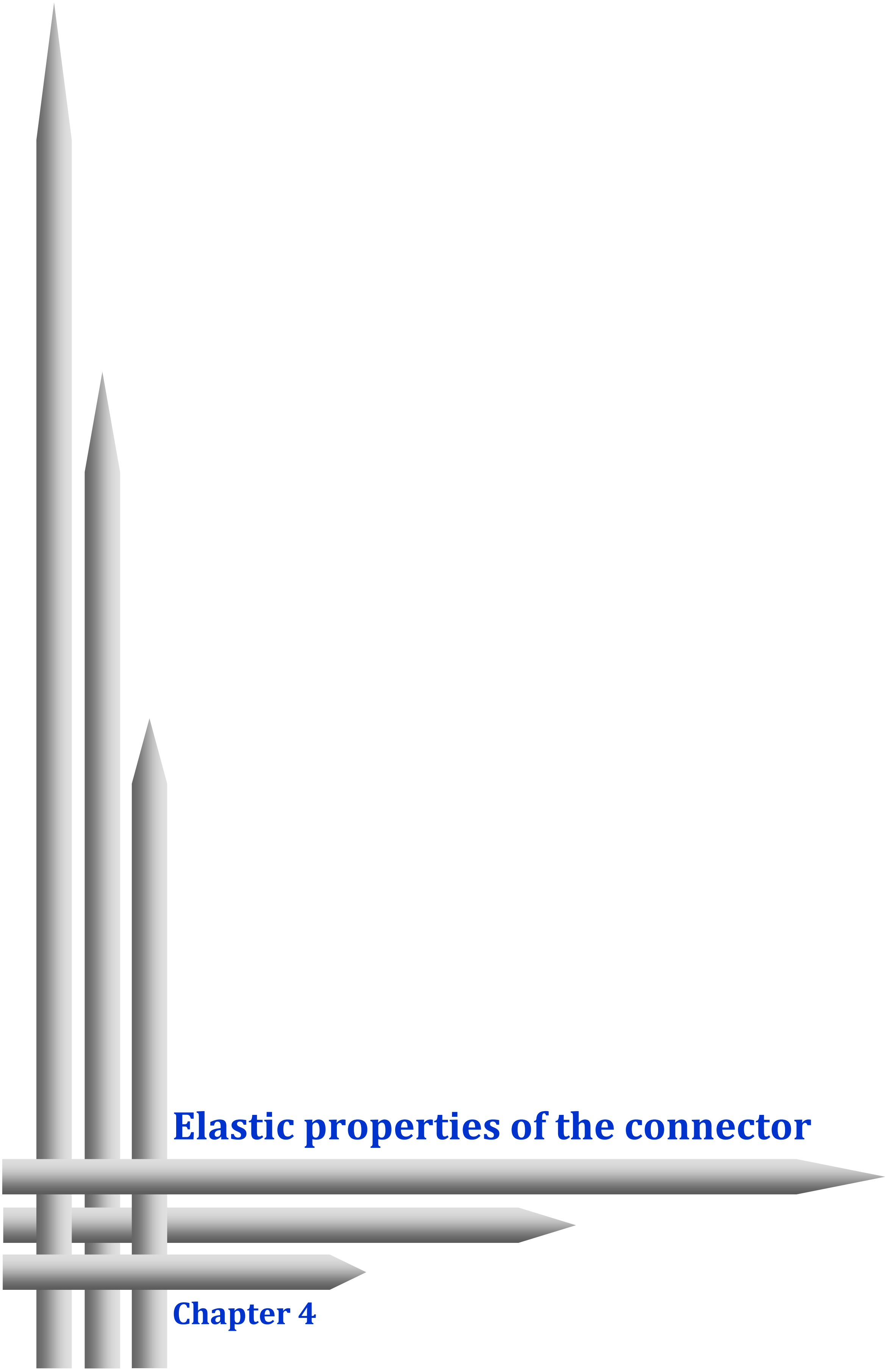
Figure 3.2: Refinement of the connector loops. (A) Plane of projection between first and second principal components (PC1 and PC2) obtained after performing PCA on the SA MD trajectories (*squares*). The semi-circular arrow shows the conformational changes with the time during the SA MD simulation. Five high temperature conformations P1, P2, P3, P4, and P5 (*red circles*) were selected from the projection plane and cooled down to 300 K. Dense clusters (*blue symbols*) labeled with P1^c, P2^c, P3^c, P4^c, and P5^c were obtained after cooling. (B) The structures of these five clusters (*blue symbols*) were compared by computing RMSDs of Cα atoms with reference to the central structure. (C) Comparison of modeled (*blue*) and refined loops (*red*) interacting with the DNA in the channel center.

As proposed in the DNA packaging models, these loops are required to interact with the DNA. Upon closer analysis, the loops are observed to be in contact with the DNA, and particularly the loop residues K234, K235 and R237 are within 0.5 Å distance to the DNA (Fig. 3.2C). Because the central structure is near to the native conformation and the loops interact with the DNA, I considered that the loop-DNA structure is sufficiently modeled to examine the packaging models; consequently, it was used as a starting structure for further equilibrium MD simulations.

3.4 Summary

Twelve missing loops of the connector and DNA were modeled, and the loops were further refined in the presence of the DNA. For the refinement, simulated annealing MD was used, and high temperature enabled the fast exploration of the energy landscape. Further, PCA of

SAMD trajectory shows that the loops were not able to explore the entire conformational space, and the obtained conformations were likely to be from high-energy regions of the landscape. To check closeness of these conformations to the native structure, five diverse conformations were cooled down to 300 K. Subsequent PCA shows that the conformations appear to converge in the same region on the conformational subspace, and this result suggests that all five conformations are approaching the native conformation. These cooled conformations were very similar based on their RMSD values, and a central structure was selected as a representative of these conformations. Additionally, the DNA packaging models require interactions between the loops and the DNA. In the selected structure, the loop residues K234, K235 and R237 are in close contact with the DNA and, therefore, loop-DNA interactions are sufficiently modeled in this structure. Overall, the closeness of the refined structure to the native structure and the obtained interactions between the loops and the DNA are considered to be sufficient for further investigations of the DNA packaging models.



Elastic properties of the connector

Chapter 4

4.1. Background

My first aim is to probe the untwist-twist DNA packaging model by calculating the elastic properties of the connector. This model requires the connector to act as a spring and it should exhibits reversible elastic deformation. During the DNA packaging, the connector untwists and stretches by 12° and 0.68 nm respectively by consuming 50 kJ.mol^{-1} of energy released by hydrolysis of one ATP molecule. Subsequently, it relaxes and pushes two DNA base-pairs by recovering its original orientation.

In the following publication, the untwist-twist model was probed by several explicit solvent MD simulations, which were performed on the connector in the absence of DNA. The publication consists of two major sections. First section discusses elastic properties of the connector, which were obtained by equilibrium fluctuation analysis and force-probe simulations. The elastic properties were computed under the harmonic approximation of energy landscape (see details in Theory chapter). The obtained results suggest that the connector is similar to a spring and has properties of elastic reversible deformation. Additionally, on the basis of simulations results, I proposed a new mechanism by which the connector withstands large counter pressure during the DNA packaging process.

Second section discusses the energetic requirement of the untwist-twist model. The energy required to untwist and stretches the connector by 12° and 0.68 nm, respectively was computed by the umbrella sampling simulations. The obtained results rendered the untwist-twist model unlikely, and therefore, remaining two models, one-way revolution and push-roll model may explain the DNA packaging mechanism.

4.2. Publication

4.2.1. Abstract

4.2.2. Introduction

1338 Biophysical Journal Volume 106 March 2014 1338–1348

Elastic Properties and Heterogeneous Stiffness of the Phi29 Motor Connector Channel

Rajendra Kumar and Helmut Grubmüller*
Max-Planck-Institute for Biophysical Chemistry, Department of Theoretical and Computational Biophysics, Göttingen, Germany

ABSTRACT The DNA packaging motor of the bacteriophage ϕ29, comprising head-tail connector, ATPase, and pRNA, transports the viral DNA inside the procapsid against pressure differences of up to ~60 atm during replication. Several models for the DNA packaging mechanism have been proposed, which attribute different roles to the connector, and require specific mechanical properties of the connector. To characterize these properties at the atomic level, and to understand how the connector withstands this large pressure, we have carried out molecular dynamics simulations of the whole connector both in equilibrium and under mechanical stress. The simulations revealed a quite heterogeneous distribution of stiff and soft regions, resembling that of typical composite materials that are also optimized to resist mechanical stress. In particular, the conserved middle α-helical region is found to be remarkably stiff, similar only to structural proteins forming viral shell, silk, or collagen. In contrast, large parts of the peripheral interface to the ϕ29 procapsid turned out to be rather soft. Force probe and umbrella sampling simulations showed that large connector deformations are remarkably reversible, and served to calculate the free energies required for these deformations. In particular, for an untwisting deformation by 12°, as postulated by the untwist-twist model, more than four times' larger energy is required than is available from hydrolysis of one ATP molecule. Combined with previous experiments, this result is incompatible with the untwist-twist model. In contrast, our simulations support the recently proposed one-way revolution model and suggest in structural terms how the connector blocks DNA leakage. In particular, conserved loops at the rim of the central channel, which are in direct contact with the DNA, are found to be rather flexible and tightly anchored to the rigid central region. These findings suggest a check-valve mechanism, with the flexible loops obstructing the channel by interacting with the viral DNA.

INTRODUCTION

The bacteriophage ϕ29 infects the bacterium *Bacillus subtilis* and uses its molecular machinery for replication (1). During this process, the precursor capsid (procapsid) forms by the incorporation of newly synthesized proteins inside the bacteria. Subsequently, a DNA packaging motor assembles at the base of the procapsid and transports viral DNA into it (1,2). The motor comprises the connector (shown as *red* in Fig. 1 *A*), prohead-RNA (pRNA), and an ATPase (3–9). Both the pRNA as well as ATPase detach from the connector after completion of DNA packaging, and the tail proteins of the phage particle join at the bottom region of the connector, completing the formation of new phage particles.

The DNA packaging motor transports viral DNA against an internal-to-external pressure difference of ~60 atm (10) generated by the densely filled viral DNA inside the procapsid. Accordingly, it is one of the strongest biological molecular motors, which renders it a promising choice for a motor in nanodevices (2). Further potential applications have been suggested, such as a molecular sorter in nanopore-based DNA sequencing devices, or as a model system to develop antiviral drugs to treat infections caused by herpes simplex virus, adenovirus, parvovirus, or pox virus (2). Additionally, motor pRNA is used as a carrier for ribozyme and antisense RNA to inhibit the Hepatitis B virus, which renders it a potential gene delivery system (11).

With a view to these diverse emerging applications, several studies were aimed at elucidating the underlying molecular DNA packaging mechanism, and in particular the properties and the role of the connector (2,12,13). The average viral DNA translocation rate is ~165 basepairs per second at the initial phase, with a gradual decrease as packing progresses (14). The motor generates up to ~110 pN at the final stage of the filling (14). Translocation of 10 basepairs of viral DNA into the procapsid requires hydrolysis of five ATP molecules by the ATPase (13).

Here, we focus on the connector as one of the motor components, for which the crystal structure has been determined (9,15). The connector is a truncated cone-shaped dodecameric protein complex that forms a channel through which viral DNA is transported inside the procapsid (9) (Fig. 1, *A* and *B*, see also Fig. S1 in the Supporting Material). The connector can be divided into three major regions—an upper, a middle, and a bottom region (*blue*, *brown*, and *green regions* in Fig. 1 *B*). The upper and middle regions are in direct contact with the procapsid, whereas the bottom region is connected to the middle region by a hinge region (*red*) and protrudes toward the outside of the procapsid. The central DNA transport channel is mainly formed by the middle and bottom region.

To uncover the mechanism by which the connector withstands the large pressure difference of up to ~60 atm, we focus on its mechanical properties. Because the connector

---

Submitted July 26, 2013, and accepted for publication January 23, 2014.

*Correspondence: hgrubmu@gwdg.de

Editor: Fazoil Ataullakhanov.

© 2014 by the Biophysical Society

0006-3495/14/03/1338/11 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.01.028

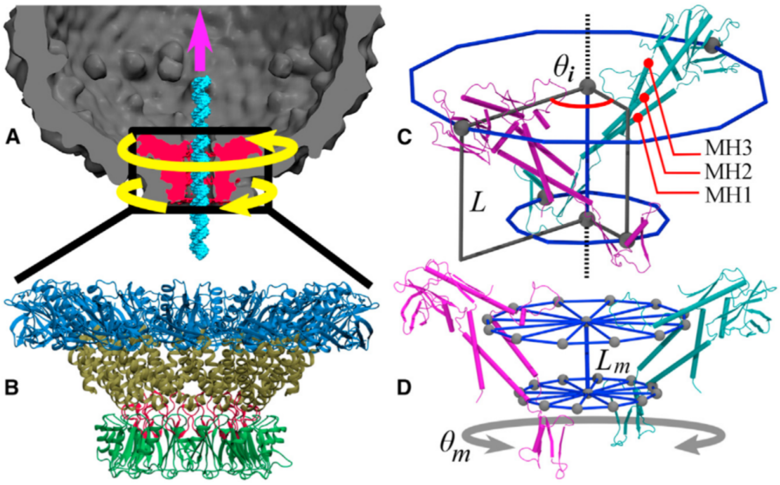


FIGURE 1 Illustration of the untwist-twist DNA packaging model; different regions of the connector (*left*), and definition of length and twist deformations (*right*) used for analysis. (*A*) The connector (*red*) and the DNA (*blue*) are aligned with the cryo-EM density map (obtained from the EM data bank, code EMD-1117 (7,8)) of the ϕ29 procapsid (*gray*). According to the untwist-twist DNA packaging model (9), the connector untwists and twists by 12° (*yellow arrows*) while the DNA is transported in two basepairs steps toward the procapsid interior (*magenta arrow*). (*B*) The structure of the connector reveals four distinct regions. The upper (*blue*) region faces the procapsid interior; the middle (*brown*) and bottom (*green*) regions, which form the channel, are connected by the hinge (*red*) region. (*C* and *D*) Two opposite (*purple*, *cyan*) out of the 12 connector subunits are illustrated in cartoon representations. (*C*) The whole connector's compression-stretching motions are characterized by the length L between the centers of mass (*dark gray spheres*) of the upper and bottom regions along the channel axis (*black dotted line*). Twist deformation is defined via relative rotation of the bottom rim with respect to the top rim (*blue circles*) and determined by the average angle θ of all torsion angles θ_i of the 12 individual subunits. The middle region consists of the three helices MH1, MH2, and MH3, referred to in the text. (*D*) Similarly, compression and twist deformations of the middle region are characterized by the length L_m between the center (*gray spheres*) of the upper and the lower disk, and the respective twist angle θ_m. The disks (*blue*) are defined via the Cα atoms of the MH1 and MH2 terminal residues, respectively (*gray spheres*) and the centers of mass of the Cα atoms of the respective disk (*axial gray spheres*).

is part of the procapsid (Fig. 1 *A*), its mechanical properties should also be similar to those of the procapsid, which have been probed earlier by atomic force microscopy (AFM) (16,17). In particular, a very large elastic modulus of ~1.8 GPa (i.e., similar to that of hard plastic) was measured, which was therefore suggested to be crucial for withstanding such high pressure difference (17). For the procapsid, a spring constant of ~0.07 N/m has been determined, whereas the connector region was found to be softer, with a spring constant of ~0.04 N/m (16). These findings render the question of how such a channel protein resists pressure differences of up to 60 atm even more puzzling.

The role of the connector in the DNA packaging mechanism is also unclear. One model, from Hendrix (18), proposed that the connector performs full rotations with respect to the procapsid, progressing by 12° in each step, thereby translocating two basepairs of the viral DNA inside the procapsid by consuming energy released from the hydrolysis of one ATP. However, more recent single-molecule florescence spectroscopic studies performed by Hugel et al. (19) ruled out such complete rotation of the connector.

A second model, which is compatible with the above results, has been proposed by Simpson et al. (9) and assumes partial rotations, with springlike coupled compression-stretching and twisting-untwisting motions (indicated in Fig. 1 *A*) of the middle helical region. According to this model, which is subsequently referred to as untwist-twist model, the whole connector first untwists by 12° through untwisting of its middle region (*yellow arrows*) under consumption of one ATP molecule; simultaneously, it stretches by 0.68 nm and grasps the next two basepairs of DNA. The connector then relaxes by twisting and contracting into its original configuration, thereby generating the force required to translocate two basepairs of the viral DNA toward the interior of the procapsid. Accordingly, this model requires the untwisting of the connector to be coupled by 17.6°/nm to an increase of its length.

A third model, the push-and-roll model, from Yu et al. (12), Moffitt et al. (13), and Aathavan et al. (20), proposes that a pentameric ATPase pushes the viral DNA by a "lever" into the procapsid in a fashion that the DNA rotates and rolls inside the motor channel during packaging process.

Recently, a fourth, the one-way revolution model, has been proposed by Guo, Schwartz, and co-workers (21–27), according to which the hexa-meric ATPase actively pushes the DNA into the procapsid, and the four electropositive lysine rings of the connector channel facilitate the DNA revolution, albeit without DNA rotation inside the channel during the packaging process. Simultaneously, the connector acts like a one-way valve and prevents leakage of the viral DNA against large counterpressure.

To reveal the molecular determinants of the exceptional mechanical stability and pressure resistance, we first probed the mechanical properties of the connector by equilibrium and nonequilibrium (force probe) molecular dynamics (MD) simulations (28–30). Remarkably, it turned out that the channel formed by the middle region of the connector is one of the stiffest known protein materials. To test which of the above packaging models is compatible with the properties of the connector, we next computed the deformation free energies required for twisting-untwisting as well as for compression-stretching motions both by a fluctuation

4.2.3. Methods

analysis of equilibrium MD simulations and by free energy umbrella sampling simulations. Our results demonstrate that the untwisting of the connector by 12°, as required by the untwist-twist model, would require much more energy than the ~50 kJ/mol provided by hydrolysis of one ATP molecule. Additionally, in recent voltage-ramping experiments (27), the connector facilitated DNA transport across the membrane even in the absence of the ATPase. These results are difficult to reconcile with this model, and thus render either the push-and-roll or the one-way revolution model more likely.

METHODS

Modeling and structure refinement

Four x-ray crystal structures of the connector at different resolutions have been published (9,15,31). We used the highest-resolution structure (PDB:1H5W; 2.10 Å) as a starting structure for all subsequent simulations. In none of the above structures, the residues ranging from A230 to S244 were resolved. These form a loop pointing toward the DNA and most likely interact with the DNA. (These were modeled and further refined as described in Section S1 in the Supporting Material, including Table S1, Fig. S1, and Fig. S2.)

Equilibrium MD simulations

The connector was extracted from the refined connector-DNA complex and taken as starting structure for subsequent equilibrium MD simulations. The simulation system was set up as follows: the connector was placed within a dodecahedron periodic box, solvated by addition of 91,559 water molecules, and neutralized by 84 sodium ions. The complete simulation system comprised 326,925 atoms. This system was energy-minimized and then heated up to 300 K during an NVT simulation with 1 fs time step. During this period, all heavy atoms were restrained with a harmonic force constant of 1000 kJ/(mol nm^2). During a subsequent 3 ns NPT simulation, the restraints for the heavy atoms were gradually removed. Finally, a 200 ns free equilibrium MD simulation was performed using a 2-fs time step.

All simulations were carried out using the GROMACS 4.0.7 package (32) with the AMBER ff99SB force field (33) and TIP3P water model (34). Long-range electrostatic interactions were computed by the particle-mesh Ewald method with a grid spacing of 1.2 Å and fourth-order cubic interpolation (35). Short-range nonbonded interactions were computed for all atom pairs within a 10 Å cutoff. The temperature was maintained at 300 K using the V-rescale algorithm (36) with a coupling time constant of 0.1 ps; during the equilibration and free equilibrium MD simulations, the pressure was maintained at 1 atm using Berendsen and Parrinello-Rahman pressure couplings (37,38) with a 1 ps coupling time constant, respectively. All bonds were constrained using the parallel LINCS implementation (39,40). The VMD program was used for visualization (41).

Umbrella sampling simulations

Deformation free energy profiles for twisting-untwisting and compression-stretching motions of the connector were calculated from umbrella sampling simulations. To obtain the reaction coordinates for umbrella sampling simulations, we performed several force-probe (FP-1, FP-2, FP-3, FP-4, FP-5, and FP-6) and relaxation (Relax-1, Relax-2, Relax-3, and Relax-4) simulations (see Section 3 and Table S2 of the Supporting Material for more details.) For the compression-stretching motion, 14 sampling windows were prepared, with umbrella potential minima between L_m = 2.725 and 3.149 nm (for details, see Table S3). Appropriate starting structures for these simulations were taken from snapshots at various times from the Relax-1, Relax-2, Relax-3, Relax-4, and equilibrium simulations. Each umbrella sampling simulation was performed for 100 ns (total 14 × 100 ns), and otherwise identical to the equilibrium simulations described above. The first 50 ns of each umbrella sampling simulation were discarded as an equilibration phase; the last 50 ns were used for collecting histograms (see Fig. S4), from which deformation free energy was calculated using by weighted histogram analysis method (WHAM) (42–44).

Deformation free energy for the twisting-untwisting motion was obtained in a similar manner, using the twist angle θ_m as a reaction coordinate (see Fig. 1 *E*). Fourteen umbrella windows were defined for which starting structures were taken from the FP-5 and FP-6 simulations. Because the conformations were close to the minimum in the energy landscape during these simulations compared with other force probe simulations, we have selected snapshots at different times from these two simulations as starting structures (for details, see Table S4). The umbrella sampling simulations were preceded by an equilibration simulation for 25 ns to equilibrate the connector structure at each respective umbrella potential minimum (total 14 × 25 ns). In these simulations, the atoms of the reference and rotational group (*upper* and *lower disk* in Fig. 1 *E*, respectively) were harmonically restrained by a 10,000 kJ/(mol nm^2) of force constant at the starting position such that the structures equilibrated at the starting twist angle of each umbrella window. Subsequently, two sets of umbrella sampling simulations were performed (15 ns for each window, total 28 × 15 ns). For the first set, the positions of reference atoms in the upper disk (shown in Fig. 1 *E*) were restrained while those of the bottom disk were subjected to an isotropic pivotal free rotational potential (29), and vice versa for the second set of simulations. The harmonic force constants used in simulations are listed in Table S4. To calculate the deformation free energy of the twisting-untwisting motion, a torsional spring constant for each window was calculated from the applied torque and resulting twist angle (for details, see Fig. S5). Histograms were collected from the umbrella sampling simulations and the deformation free energy was calculated using weighted histogram analysis method (WHAM) (42–44).

Analysis

Structural descriptors for the elastic properties

To probe the elastic properties of the connector, two structural descriptors, namely the twist angle and the length, were considered that describe the twisting-untwisting and compression-stretching motion of the connector, respectively. The connector length L is the distance between the center of masses of the upper and the bottom region (see Fig. 1 *C*). The whole connector's twist angle

$$\theta = \frac{1}{12}\sum_{i=1}^{12}\theta_i$$

has been calculated by averaging over the twist angle θ_i of each subunit (see in Fig. 1 *C*). Similarly, the twist angle θ_m and length L_m of the middle region were calculated to investigate the elasticity of the channel, where θ_m denotes the rotational angle between the upper and the lower disk and L_m the distance between the two disks shown in Fig. 1 *D*. The elastic properties were calculated using the snapshots taken at 2 ps time difference during the last 100 ns of equilibrium MD simulations.

Elastic properties and spring constants from equilibrium simulations

The free energy landscape $G(\theta,L)$ as a function of the connector length L and angle θ as defined above was determined independently from equilibrium simulations in harmonic approximation from the equilibrium fluctuations of the connector via

$$G(\theta,L) = -k_\mathrm{B}T \ln p(\theta,L),$$

4.2.4. Results and discussion

where k_B and T are the Boltzmann constant and temperature, respectively. For a Gaussian approximation of the two degrees of freedom, θ and L, probability density is given as

$$p(\theta, L) \propto \exp\left(-\frac{1}{2}\left[(\theta-\overline{\theta}),(L-\overline{L})\right]\mathbf{C}^{-1}\begin{bmatrix}(\theta-\overline{\theta})\\(L-\overline{L})\end{bmatrix}\right), \quad (1)$$

where $\overline{\theta}$ and $\overline{L}$ are the average twist angle and length, respectively, and $\mathbf{C}$ is the covariance matrix. Because the energy landscape is assumed to be harmonic, its expression is

$$\begin{aligned} G(\theta, L) &= \frac{1}{2}\left[K_\theta\left(\theta-\overline{\theta}\right)^2 + K_L\left(L-\overline{L}\right)^2\right] \\ &\quad + K_c\left(\theta-\overline{\theta}\right)\left(L-\overline{L}\right), \end{aligned}$$

where K_θ is the torsional spring constant, K_L is the stretching spring constant, and K_c is the coupling constant between two motions. By comparing the above three equations, these constants can be written as

$$\begin{bmatrix} K_\theta & K_c \\ K_c & K_L \end{bmatrix} = k_b T\mathbf{C}^{-1}. \quad (2)$$

This equation was used to calculate the spring constants from the equilibrium fluctuation for the whole connector and the middle region-containing channel.

Young's modulus of elasticity

To translate the obtained elastic constant into a Young's modulus of the connector, we described the connector as a homogeneous elastic material with elasticity Y and with the shape of a hollow truncated cone of length L (see Fig. S6). From the observed length change of

$$\delta = \int_0^L \frac{F}{YA(x)}dx$$

upon axial stress force F (45), where $A(x)$ is the cross-sectional area perpendicular to the symmetry axis, the elasticity Y in terms of stretching spring constant K_L is

$$Y = \int_0^L \frac{K_L}{A(x)}dx, \quad (3)$$

which yields, for the assumed hollow truncated cone having variant diameters,

$$Y = \frac{2K_L}{\pi}\frac{L}{D_b d_a - D_a d_b}\ln\left|\frac{(D_a+d_a)(D_b-d_b)}{(D_a-d_a)(D_b+d_b)}\right|, \quad (4)$$

where D_a and d_a are narrow-end exterior and interior diameter, respectively, and D_b and d_b are wide-end exterior and interior diameter of the truncated cone, respectively (*sketch* shown in Fig. S6). The dimensions are calculated for the whole connector and the middle region from equilibrium MD simulations and provided in Table S5.

Calculation of interface area

To quantify inter- and intrasubunit residue packing, we have calculated surface area (SA) accessible by a probe with a radius of 1 Å using the *g_sas* module of the GROMACS package (46). Intrasubunit packing was determined via helices MH1, MH2, and MH3 of the middle region (Fig. 1 *C*),

$$\begin{aligned} SA^{\text{intra}} &= [(SA_{MH2} + SA_{MH3}) - SA_{MH2\text{-}MH3}]/2 \\ &\quad + [(SA_{MH2\text{-}MH3} + SA_{MH1}) - SA_{MH1\text{-}MH2\text{-}MH3})]/2, \end{aligned} \quad (5)$$

where subscripts denote either separate helix or attached pair helices. Similarly, intersubunit packing was calculated via

$$SA^{\text{inter}} = \left[\left(SA_i + SA_j\right) - \left(SA_{ij}\right)\right]/2, \quad (6)$$

where SA_i and SA_j refer to separated adjacent subunits, and SA_{ij} to the attached subunits.

RESULTS AND DISCUSSION

We first monitored the structural stability of the connector during equilibrium MD simulations via root mean-square deviation calculated for all C-α atoms with respect to the crystal structure (see Section S2, Fig. S7, and Fig. S8 in the Supporting Material). Overall, the structural changes during 200-ns equilibration were rather small, with the root mean-square deviation stabilizing at 0.25 nm. The last 100-ns part of the trajectory was used for further analysis.

How does the connector withstand the high pressure difference?

Elasticity of the connector from equilibrium fluctuations

To study the structural determinants that enable the connector and its channel to withstand large relative pressure, we first calculated the elastic properties of the whole connector as well as its middle region (see Table S6). The stiffness opposing twisting-untwisting and compression-stretching motions (see Fig. 1, *C* and *D*), described by torsional (K_θ) and stretching (K_L) spring constants (see Table S6) along with coupling constants (K_c) between motions, was determined using Eq. 2 from the equilibrium fluctuations shown in Fig. 2, *A* and *B*. As can be seen, the calculated torsional and stretching spring constants of the middle region are approximately two and eight times larger than that of the whole connector, respectively; therefore, most structural changes under pressure are expected for the upper and bottom regions.

Using the stretching spring constants and taking into account the geometry of the connector (see Eq. 4), Young's moduli of 0.36 ± 0.06 GPa and 3.4 ± 0.6 GPa for the whole connector and for the middle region, respectively, were obtained. Furthermore, the obtained moduli values were almost unchanged during the last 25 ns of the simulations that suggest these values are sufficiently converged (see Fig. S9). A similar convergence behavior has been seen also for a similar-sized molecular system, i.e., the central shaft γ-subunit of the F1-ATPase motor, for which, additionally, good agreement with experimental results has been shown (47).

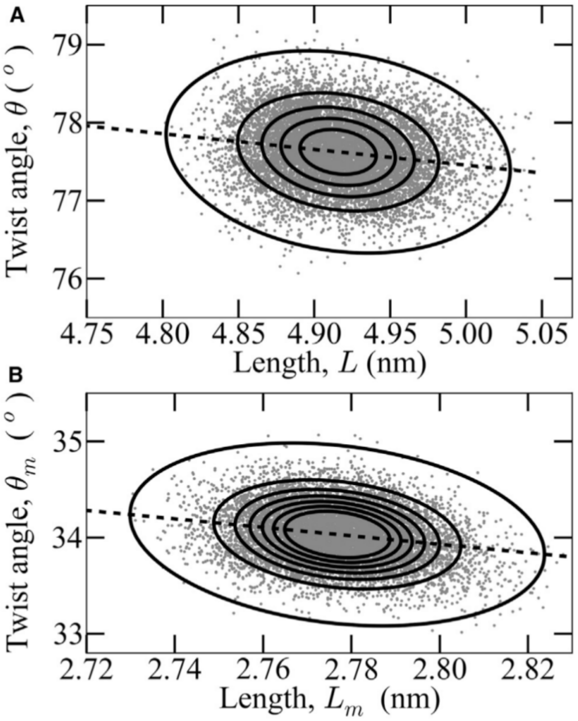


FIGURE 2 Equilibrium fluctuations of twist angles and lengths. (*A*) Fluctuations of twist angle θ and length L of the whole connector, derived from equilibrium simulations (*shaded dots*). The probability density was obtained using Eq. 1 (depicted as *solid lines* ($2^{\circ -1}$ nm^{-1} spacing)). (*B*) Same representation for twist angle θ_m and length L_m and $5^{\circ -1}$ nm^{-1} contour line spacing. Linear least-square fits (*dashed lines*) quantify the coupling between compression-stretching and twisting-untwisting motions.

For the whole ϕ29 procapsid, a Young's modulus of ~1.8 GPa was measured in AFM experiments (16,17); thus the connector as a whole appears to be softer than the procapsid, with a stiffer "core" inside. The elasticity of this middle region is similar to that of other structural proteins that withstand mechanical forces, such as collagen fibrils (0.2–11.5 GPa) (48), single-brin silkworm silk (5–17 GPa) (49), or dragline spider silk (11–13 GPa) (50). We now ask what are the structural determinants and the function of this heterogeneity.

Heterogeneous flexibility of the connector

To characterize the spatial distribution of the observed elastic heterogeneity as well as its relation to molecular fluctuations, we calculated, residuewise, root mean-square fluctuations (RMSF) for each subunit. Fig. 3 *A* (*green line*) depicts the obtained values averaged over the 12 subunits. As can be seen, overall the crystallographic temperature factors (*blue lines*, taken from Guasch et al. (15)) agree well with the RMSF (correlation coefficient: 0.83), which suggests that these fluctuations and the derived elastic properties are captured sufficiently accurately by our simulations.

To characterize the spatial distribution of the elastic properties, we identified rigid regions from RMSF depicted in Fig. 3 *A*. Residues with small fluctuations (RMSF $<$ 0.065 nm) were considered rigid (*red region* in Fig. 3, *B–D*), whereas those residues with larger fluctuations were considered flexible (*blue region*). In the upper region, rigid residues are embedded within flexible parts, very much like in composite materials. The middle region is the stiffest part and mainly consists of rigid residues. In contrast, the residues of the hinge and the bottom regions are flexible (Fig. 3 *B*). This finding suggests that the relatively low stiffness probed by AFM (16) is mainly due to the flexible and elastic bottom region of the connector.

Obviously, the regions defined in Fig. 1 *B* purely based on structural features also reflect different elastic properties of the connector. This mechanic heterogeneity suggests that these regions even fulfill different functions. As visible from Fig. 3 *C*, the pressure acts primarily on the rigid core (*red*), which defines the channel that is in direct contact to the DNA. To maintain a constant diameter of the channel, this region, therefore, has to withstand deformations due to the strong pressure gradient between the capsid interior and exterior as well as mechanical stress via the contacting parts of the procapsid, both in the presence and absence of DNA. Due to the functional relevance of this region, one would expect the structure of this region to be highly conserved. Indeed, very similar rigid helical scaffolds are also seen in other tailed phage connectors such as T7, SPP1, and P22 (51–57). Rigid residues are located inside the procapsid whereas the flexible bottom region is located outside of the procapsid.

Interestingly, the rigid core is directly exposed to the solvent of the capsid interior (*arrows* pointing high pressure in Fig. 3 *C*); here, the exposure to solvent evenly distributes the pressure over the surface of the core region. In contrast, contact of the upper region to the procapsid is buffered by flexible regions, which absorb local force peaks and distribute external forces evenly over the outer surface of the core region. In that respect, and also from the compartmented substructure of the outer rim visible in Fig. 3 *B*, this region resembles a composite material, very similar to the protein nanocrystals that are embedded within flexible unstructured peptide regions in silk (58) or the calcium carbonate (calcite or aragonite) crystals layers within a protein matrix in sea shells (59,60). From closer analysis of the direct contact region to the capsid (sketched in Fig. 3 *D*) we speculate that this structure particularly serves to withstand displacement of the connector due to vertical forces from the capsid interior, while tolerating horizontal motions or forces that the capsid channel rim may exert onto the middle region of the connector.

The flexible bottom region points toward the capsid exterior and, therefore, does not need to withstand large pressure

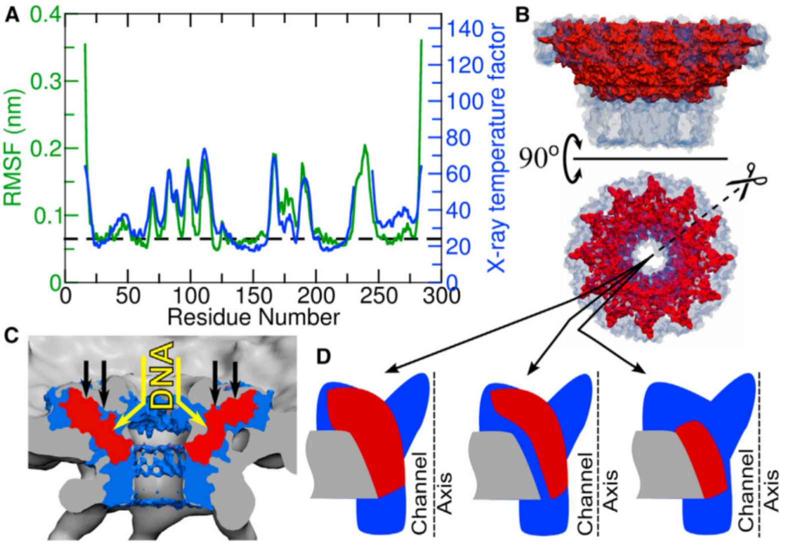


FIGURE 3 Heterogeneous distribution of rigid and flexible connector regions. (*A*) Root mean-square fluctuations (RMSF) (*green*) and temperature factors (*blue*) obtained from crystallography (15). Residues with an RMSF below 0.065 nm (*horizontal dashed line*) are considered rigid, otherwise flexible. (*B*) Spatial distribution of rigid (*red*) and flexible (*blue*) connector regions (*side* and *top view*) and (*C*) shown embedded within the procapsid (*gray*). Mainly rigid residues of the middle region (*red*) are exposed to pressure (*black arrows*) due to compacted DNA. This rigid region also supports loops formed by residues A230–S244 that directly interact with the DNA (*yellow arrows*). (*D*) Three different types of the connector-procapsid interfaces are illustrated by three schematic radial cross-sections (indicated by *three arrows*) of the connector (flexible regions in *blue*, rigid regions in *red*) and the part of the procapsid (*gray*) that is in direct contact with the connector. An alternating pattern of direct contacts is seen between the central rigid region (*red*), the rim of the procapsid (*gray*), and regions where flexible regions (*blue*) act as an interface.

differences or mechanical stress. Accordingly, its high flexibility might facilitate binding of the pRNA and ATPase at the start of the packaging process. Overall, the heterogeneous flexibility pattern seen in our simulations of the ϕ29 connector closely matches the mechanical properties that are required to perform its biological function within the extreme mechanical and pressure conditions of its procapsid environment.

Elastic properties determined from force probe simulations

Above, the elastic properties of the connector were probed via equilibrium simulations. Accordingly, the probed length and twist angle changes only covered the range accessible to thermal fluctuations. Further, by calculating elastic constants, we assumed a harmonic (i.e., Hookean) and reversible behavior of the connector. However, the different models of the packaging mechanism, as discussed in the Introduction, require much larger connector deformations than those observed above; it is thus unclear if the connector also exhibits Hookean properties in this case, which also would require a fully elastic, i.e., reversible behavior. We therefore need to determine its elastic limit, i.e., to which extent it can be deformed reversibly. It is, finally, unclear if the potentially functionally relevant coupling between stretching and twisting of the connector, as calculated from the equilibrium fluctuations shown in Fig. 2 *A*, also extends into deformations that are thermally inaccessible.

To address these questions, we carried out several sets of force probe (FP) simulations, which are presented and discussed in Section S3 of the Supporting Material (see Table S2, Fig. S10, Fig. S11, Movie S1, and Movie S2 in the Supporting Material). In summary, deformations within the range of $L = 4.70$–5.3 nm and $\theta = 70$–$79.5°$ seem to be fully elastic, and the twisting-untwisting motion is linearly coupled to the compression-stretching motion within this elastic range. The connector recovered its equilibrium structure even after structural breakdown within the obtained elastic limit (see Fig. S12 and Fig. S13). The obtained coupling of 18°/nm is remarkably close to the coupling of 17.6°/nm required in the untwist-twist DNA packaging model proposed by Simpson et al. (9). We therefore asked next if the other elastic and energetic properties of the connector are also compatible with this model.

Role of the connector in the DNA packaging process

Twisting-untwisting of the connector

Specifically, according to the untwist-twist model, the coupled untwisting and stretching motion of the connector by 12° and 0.68 nm, respectively, is driven by hydrolysis of one ATP molecule, which under physiological conditions releases an energy of ~50 kJ/mol.

As a first step, to check if this energy suffices to induce the above deformation, we estimated the required free energy in harmonic approximation from the equilibrium fluctuations (see *gray dots* in Fig. S14), as described in the Methods. Extrapolation of this harmonic approximation (depicted as *contour lines*) suggests, however, that a very large amount is actually required for the deformation of 12°/0.68 nm, >10 times larger than the 50 kJ/mol available from ATP hydrolysis.

However, the above force-probe simulations (see Fig. S10, *B* and *C*) suggest that the simple harmonic approximation may not hold for such extreme deformations.

In a second step, we therefore computed the potential of mean force (Fig. 4, *A* and *B*) along both the

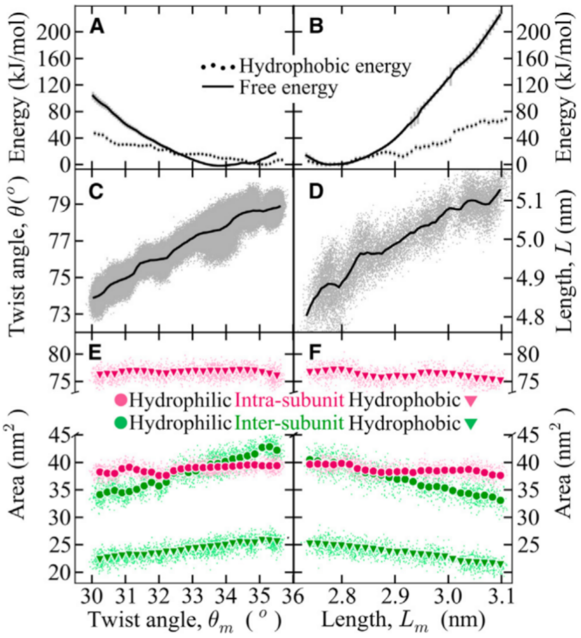


FIGURE 4 Deformation free energies and exposure of hydrophobic and hydrophilic residues of the connector from umbrella sampling simulations. Compressing-stretching deformations (L, L_m) and twisting-untwisting deformations (θ and θ_m) for the whole connector and, respectively, the middle region, are shown. (*A* and *B*) Deformation free energy (*black solid line*) as well as the contributions of hydrophobic solvent-accessible surface area to the free energies (*black dotted lines*) for (*A*) twisting-untwisting and (*B*) compression-stretching deformation of the connector middle region. (*C* and *D*) Coupling between deformations of the whole connector and respective middle region deformations (*small gray dots*), with running averages (*solid lines*). (*E* and *F*) Change of intersubunit (*green*) and intrasubunit (*red*) interface contact areas (*small dots*, as defined in the text) during (*E*) twisting-untwisting and (*F*) compression-stretching deformations of the middle region; decreasing hydrophobic (*triangle*) and hydrophilic (*circle*) contact areas imply increasing solvent exposure. Running averages are shown as larger symbols.

twisting-untwisting (θ_m) as well as the compression-stretching (L_m) coordinate from umbrella sampling simulations using the weighted histogram analysis method (see Methods for details). As can be seen from the first set of umbrella sampling simulations with length (L_m) of the middle region of the connector as reaction coordinate (Fig. 4, *B* and *D*), the available 50 kJ/mol allow for an extension of only 0.1 nm, thus confirming the above extrapolation. Similarly, using the twist angle (θ_m) as a reaction coordinate, already ~100 kJ/mol are required to untwist both the middle region and the whole connector by only 4° (Fig. 4, *A* and *C*), i.e., by one-third of the deformation required by the proposed untwist-twist model. Therefore, our umbrella calculations also strongly suggest that the free energy released by ATP hydrolysis is too small for the required deformations.

Furthermore, to characterize the coupling between the deformations of the middle region and the whole connector that was proposed in the untwist-twist model, we determined the whole connector's deformation, i.e., the change in the twist angle (θ) and the length (L) with respect to that (θ_m and L_m) of the middle region (Fig. 4, *C* and *D*). As can be seen, the deformation of middle region entirely translates to the whole connector, which suggests that the deformations of the whole connector and the middle region are indeed strongly coupled.

We note that our above estimate of the deformation free energy rests on the assumption that it does not decrease with stretching or untwisting beyond L_m =3.1 nm and θ_m =30°, as might happen, e.g., after complete structural breakdown of the connector. However, the reversible recovery of the equilibrium conformation after deformations of up to 5.1 and 5.3 nm, respectively (see Fig. S13, *A* and *B*), strongly speaks against a complete structural breakdown.

To assess the convergence of the umbrella sampling simulations, we compared the resulting untwisting-stretching path from the first set of simulations with the second set (shown in Fig. S15). These two paths are similar to each other (*red* and *black symbols*) and also coincide with the path obtained in FP-5/6 simulations (*green symbols*), underscoring that the free energy profiles are nearly converged.

All potentials of mean force were calculated in absence of the DNA, which may affect the energy required for the proposed untwisting/stretching. However, according to the above results, already half of the proposed deformation requires ~200 kJ/mol. In contrast, the maximum barrier height that can be overcome with the available 50 kJ/mol for an activated process is ~100 kJ/mol, assuming an additional barrier of 50 kJ/mol, estimated from a Kramer's attempt rate of 10^{10}/s and a plausible turnover rate of 165/s (14). We consider it therefore unlikely that the presence of DNA would decrease the required deformation free energy by such a large amount to drive the deformations required by the untwist-twist model. Interestingly, Jing et al. (27) observed in their voltage-ramping experiments that DNA was driven by the electrochemical gradient through the connector even in the absence of pRNA and ATPase. According to the Nernst equation, the applied potential of 75 mV (= 7.24 kJ/mol) translates, for the two basepairs (eight negative charges) by which the DNA is advanced in one cycle, to an energy of ~58 kJ/mol. In these experiments, this is thus the maximal available energy for possible connector deformations—again much smaller than required by the untwist-twist model. Combined with our above results, the observed DNA diffusion in the absence of an ATPase strongly suggests that in these experiments, the connector also facilitates DNA transport without performing the proposed twisting-untwisting or compression-stretching motion.

The connector as a one-way valve

As alternatives, the so-called push-roll and one-way revolution models have recently been proposed (12,21). Both

4.2.5. Conclusion

models postulate that the ATPase interacts with and actively transports the viral DNA. Additionally, the latter model proposes that the connector acts as a one-way valve, restricting leakage of the DNA during the packaging process (21). In fact, in the above-mentioned voltage-ramping experiments, selective one-directional diffusion of DNA across the lipid membrane through the connector was observed; for a reversed potential, no backdiffusion of the DNA was seen (27). Further, the connector's ability to restrict the leakage of DNA was probed through sedimentation assays in which the DNA was retained inside partially filled procapsids in the absence of pRNA/ATPase after ultracentrifugation (27).

Furthermore, in mutagenesis experiments, deletion/mutation of particular connector loop residues (K234–R237; loops depicted in Fig. S1 *B* and Fig. S2 *C*) severely affected the successful production of infectious bacteriophage (61). In sedimentation assays, the mutant K235A.E236A.R237A was unable to retain DNA inside completely filled procapsids in the presence of pRNA/ATPase. These results suggested that the connector loops are essential for the observed unidirectional DNA transport (26) and strongly interact with the DNA via three positively charged residues (K234, K235, and R237) per each subunit (61). One possible mechanism that would involve the connector loops as a key is that the connector acts as a check-valve, with the flexible loops bound to the DNA acting as movable parts that obstruct the channel and arrest the DNA against backmotion toward the outside, due to the strong pressure force. Such mechanism, to function, would require a rigid anchoring of the loops to withstand the backtracking force acting onto the DNA. Indeed, this particular loop region is solidly attached to and supported by the stiff region of the connector that has been identified above (*red region*, Fig. 3, *B* and *C*). Further, recent voltage-ramping experiments revealed that after deletion of these loops, the connector failed to restrict the backtransport of the DNA across the lipid membrane (21). Overall, the obtained heterogeneity in the mechanical properties of the connector seems to be compatible with the one-way revolution model.

Structure and energetic determinants for connector stiffness

To answer the question of which features of the connector determine its remarkable mechanical properties, we calculated and analyzed the extent to which selected structural and energetic quantities changed during the deformations induced by the above-described umbrella sampling simulations. As one possible source of the observed elastic restoring forces, we focused at the hydrophobic core (62–64) of the middle region of the connector. Should this core become partly exposed to the solvent during untwisting or stretching motions, the corresponding free energy cost might explain these forces. To test if this is actually the case, we calculated the hydrophobic solvent-accessible surface areas of the middle region (see Fig. S16) and estimated the resulting hydrophobic free energy from an energy cost of 18 cal/mol per unit surface area ($Å^{-2}$) (65) (*dotted lines* in Fig. 4, *A* and *B*). As can be seen, the hydrophobic energy indeed contributes approximately one-third to the total free energy, but clearly does not fully explain the connector's elastic properties.

We then asked if a possible loosening of the packing within the connector, and the resulting loss of intra- and intersubunit interactions within the middle regions, might also contribute to the mechanical properties. To this aim, we have quantified the packing within each subunit by the intrasubunit packing (Eq. 5), defined by the interface area (Fig. 4, *E* and *F*, *red symbols*) between each of the three helices (MH1, MH2, and MH3 shown in Fig. 1 *C*), as well as the packing between adjacent subunits (Eq. 6), defined by the interface area between two adjacent subunits (*green symbols*). Further, Fig. 4, *E* and *F*, distinguishes between hydrophobic and hydrophilic interface surface.

For the intrasubunit packing, the hydrophobic area was approximately two times larger than the hydrophilic area, suggesting the presence of an internal hydrophobic "core" within each subunit. However, no significant changes of the respective areas are present during both untwisting and stretching, such that their contribution to the deformation free energy is small. In contrast, both intersubunit hydrophilic and hydrophobic interface areas decreased during stretching and untwisting, indicating a loosening of their mutual packing. That this loosening actually implies significant interaction enthalpy changes is demonstrated in Fig. S17, which shows the respective van der Waals and electrostatic energies during deformation. Because, due to the size of the simulation system, the corresponding entropic part could not be quantified, we cannot decide whether this second free energy contribution, together with the hydrophobic surface changes quantified above, suffices to fully explain the deformation free energies derived from our umbrella sampling simulations. However, we consider it unlikely that the respective enthalpy/entropy reduces the resulting free energy to such a large extent as to render the intersubunit packing changes unimportant.

CONCLUSIONS

During assembly of many bacteriophages, strong motors transport the viral DNA inside the phage procapsid through a connector against pressure difference of up to ~60 atm. Here we have investigated the mechanical elastic properties of the ϕ29 head-tail connector (a component of the ϕ29 bacteriophage DNA packaging motor), and its possible role in the DNA packaging mechanism. A fluctuation analysis of equilibrium MD simulations revealed an exceptionally stiff α-helical channel region, with an elastic modulus that is only approached by proteins that have evolved to withstand strong mechanical forces, such as crystalline silk domains. The functional relevance of this middle region is underscored by its high conservation level in head-tail connectors of other phages such as SPP1 and P22 (52,54–56). The high stiffness may serve to stabilize the channel and in particular to maintain its inner diameter in absence of DNA. Remarkably, soft regions were also seen within the connector, such that, overall, the connector displays an unexpected mechanical heterogeneity, resembling that of typical composite materials such as silk and sea shells—not only structurally but also, most likely, functionally.

Our force probe simulations revealed a large elastic regime of the connector, and showed that quite large untwisting-twisting and compression-stretching deformations are in fact reversible on very short timescales. The extent of fully reversible deformations is markedly larger than that typically seen for proteins such as immunoglobulin domains or ubiquitin (66,67), but comparable to proteins with functionally important elasticity such as importin-β (68,69) or viral capsid proteins (30,70). It will be interesting to study how mechanical properties of the connector might change in the presence of DNA.

Subsequent closer analysis of umbrella sampling trajectories revealed that the observed pronounced exposure of hydrophobic residues to solvent during deformation markedly contributes to the total free energy required to deform the connector, as also been observed for viral capsid proteins (70). Upon deformation of the connector, its hydrophobic residues core is exposed to the solvent, particularly at the intersubunit interface, which suggests that this core is a structural feature that contributes to the exceptional stiffness of the connector. Further, the interaction enthalpies between adjacent subunits were observed to weaken during the deformations, which suggests an additional, mainly enthalpic, contribution to the overall elasticity.

Our results also shine light on the four possible different packaging mechanisms that have been proposed (9,12,13,18–20,26,27,61). In particular, the free energy required to deform the connector to an extent required for the untwist-twist mechanism is much larger than the energy provided by ATP hydrolysis, which—combined with results from voltage-ramping experiments (27)—speak against that mechanism. Our results neither rule out nor support the first proposed mechanism, involving full rotations of the connector (18); however, this mechanism has already been ruled out experimentally (19).

The heterogeneous stiffness pattern of the connector and its elastic properties, as obtained from our simulations, are compatible with the fourth, recently proposed one-way revolution mechanism. According to this mechanism, the connector acts as a one-way valve and thereby restricts leakage of the viral DNA during and after the DNA packaging process. In particular, strong interactions are expected between flexible loops located at the inner rim of the channel formed by residues K234–R237, which can obstruct the channel similar to a check-valve mechanism, whereas the stiff regions of the connector provide a solid anchor for these flexible loops. We note that, because no structural information is available on the bound DNA, it has not been included in our simulations and, therefore, our results neither rule out nor support the DNA revolution mechanism that is proposed in the one-way revolution model. Similarly, our results neither rule out nor support the DNA rotation-roll mechanism that is proposed in the third proposed model, the push-roll model. To that aim, it remains to be tested how the connector interacts with the pRNA and ATPase to facilitate DNA translocation and how the viral DNA moves into the procapsid during the packaging process.

SUPPORTING MATERIAL

Six tables, 17 figures, two movies and References (71–80) are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00126-X.

This work has been supported by the European Commission under grant Nos. 211800 SBMPS and FP7/2007-2013.

REFERENCES

1. Meijer, W. J., J. A. Horcajadas, and M. Salas. 2001. ϕ29 family of phages. *Microbiol. Mol. Biol. Rev.* 65:261–287.
2. Guo, P. 2005. Bacterial virus ϕ29 DNA-packaging motor and its potential applications in gene therapy and nanotechnology. *Methods Mol. Biol.* 300:285–324.
3. Guo, P. X., S. Bailey, …, D. Anderson. 1987. Characterization of the small RNA of the bacteriophage ϕ29 DNA packaging machine. *Nucleic Acids Res.* 15:7081–7090.
4. Guo, P. X., S. Erickson, and D. Anderson. 1987. A small viral RNA is required for in vitro packaging of bacteriophage ϕ29 DNA. *Science.* 236:690–694.
5. Lee, T. J., and P. Guo. 2006. Interaction of gp16 with pRNA and DNA for genome packaging by the motor of bacterial virus ϕ29. *J. Mol. Biol.* 356:589–599.
6. Valpuesta, J. M., and J. L. Carrascosa. 1994. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Q. Rev. Biophys.* 27:107–155.
7. Lawson, C. L., M. L. Baker, …, W. Chiu. 2011. EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res.* 39 (Database issue):D456–D464.
8. Morais, M. C., K. H. Choi, …, M. G. Rossmann. 2005. Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of ϕ29. *Mol. Cell.* 18:149–159.
9. Simpson, A. A., Y. Tao, …, M. G. Rossmann. 2000. Structure of the bacteriophage ϕ29 DNA packaging motor. *Nature.* 408:745–750.
10. Smith, D. E., S. J. Tans, …, C. Bustamante. 2001. The bacteriophage straight ϕ29 portal motor can package DNA against a large internal force. *Nature.* 413:748–752.
11. Hoeprich, S., Q. Zhou, …, P. Guo. 2003. Bacterial virus ϕ29 pRNA as a hammerhead ribozyme escort to destroy hepatitis B virus. *Gene Ther.* 10:1258–1267.
12. Yu, J., J. Moffitt, …, G. Oster. 2010. Mechanochemistry of a viral DNA packaging motor. *J. Mol. Biol.* 400:186–203.
13. Moffitt, J. R., Y. R. Chemla, …, C. Bustamante. 2009. Intersubunit coordination in a homomeric ring ATPase. *Nature.* 457:446–450.

14. Rickgauer, J. P., D. N. Fuller, …, D. E. Smith. 2008. Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage ϕ29. *Biophys. J.* 94:159–167.
15. Guasch, A., J. Pous, …, M. Coll. 2002. Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage ϕ29 connector particle. *J. Mol. Biol.* 315:663–676.
16. Carrasco, C., A. Luque, …, P. J. de Pablo. 2011. Built-in mechanical stress in viral shells. *Biophys. J.* 100:1100–1108.
17. Ivanovska, I. L., P. J. de Pablo, …, G. J. Wuite. 2004. Bacteriophage capsids: tough nanoshells with complex elastic properties. *Proc. Natl. Acad. Sci. USA.* 101:7600–7605.
18. Hendrix, R. W. 1978. Symmetry mismatch and DNA packaging in large bacteriophages. *Proc. Natl. Acad. Sci. USA.* 75:4779–4783.
19. Hugel, T., J. Michaelis, …, C. Bustamante. 2007. Experimental test of connector rotation during DNA packaging into bacteriophage ϕ29 capsids. *PLoS Biol.* 5:e59.
20. Aathavan, K., A. T. Politzer, …, C. Bustamante. 2009. Substrate interactions and promiscuity in a viral DNA packaging motor. *Nature.* 461:669–673.
21. Zhao, Z., E. Khisamutdinov, …, P. Guo. 2013. Mechanism of one-way traffic of hexameric ϕ29 DNA packaging motor with four electropositive relaying layers facilitating antiparallel revolution. *ACS Nano.* 7:4082–4092.
22. Schwartz, C., and P. Guo. 2013. Ultrastable pRNA hexameric ring gearing hexameric ϕ29 DNA-packaging motor by revolving without rotating and coiling. *Curr. Opin. Biotechnol.* 24:581–590.
23. Schwartz, C., G. M. De Donatis, …, P. Guo. 2013. Revolution rather than rotation of AAA+ hexameric ϕ29 nanomotor for viral dsDNA packaging without coiling. *Virology.* 443:28–39.
24. Schwartz, C., G. M. De Donatis, …, P. Guo. 2013. The ATPase of the ϕ29 DNA packaging motor is a member of the hexameric AAA+ superfamily. *Virology.* 443:20–27.
25. Zhang, H., C. Schwartz, …, P. Guo. 2012. Chapt. 9. "Push through one-way valve" mechanism of viral DNA packaging. *In* Advances in Virus Research. Ł. Małgorzata and S. Wacław, editors. Academic Press, New York, pp. 415–465.
26. Fang, H., P. Jing, …, P. Guo. 2012. Role of channel lysines and the "push through a one-way valve" mechanism of the viral DNA packaging motor. *Biophys. J.* 102:127–135.
27. Jing, P., F. Haque, …, P. Guo. 2010. One-way traffic of a viral motor channel for double-stranded DNA translocation. *Nano Lett.* 10:3620–3627.
28. Grubmüller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science.* 271:997–999.
29. Kutzner, C., J. Czub, and H. Grubmüller. 2011. Keep it flexible: driving macromolecular rotary motions in atomistic simulations with GROMACS. *J. Chem. Theory Comput.* 7:1381–1393.
30. Zink, M., and H. Grubmüller. 2009. Mechanical properties of the icosahedral shell of Southern bean mosaic virus: a molecular dynamics study. *Biophys. J.* 96:1350–1363.
31. Simpson, A. A., P. G. Leiman, …, M. G. Rossmann. 2001. Structure determination of the head-tail connector of bacteriophage ϕ29. *Acta Crystallogr. D Biol. Crystallogr.* 57:1260–1269.
32. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.
33. Hornak, V., R. Abel, …, C. Simmerling. 2006. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725.
34. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
35. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald—an $N.\log(N)$ method for Ewald sums in large systems. *j. Chem. Phys.* 98:10089–10092.
36. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
37. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
38. Nose, S., and M. L. Klein. 1983. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50:1055–1076.
39. Hess, B. 2008. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4:116–122.
40. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
41. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:27–38.
42. Hub, J. S., B. L. de Groot, and D. van der Spoel. 2010. G_WHAM-A free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* 6:3713–3720.
43. Kumar, S., D. Bouzida, …, J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* 13:1011–1021.
44. Roux, B. 1995. The calculation of the potential of mean force using computer-simulations. *Comput. Phys. Commun.* 91:275–282.
45. Gere, J. M. 2004. Mechanics of Materials. Brooks/Cole-Thomson/Cengage Learning, Stamford, CT, pp. 78–83.
46. Eisenhaber, F., P. Lijnzaad, …, M. Scharf. 1995. The double cubic lattice method—efficient approaches to numerical-integration of surface-area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* 16:273–284.
47. Czub, J., and H. Grubmüller. 2011. Torsional elasticity and energetics of F_1-ATPase. *Proc. Natl. Acad. Sci. USA.* 108:7408–7413.
48. Wenger, M. P. E., L. Bozec, …, P. Mesquida. 2007. Mechanical properties of collagen fibrils. *Biophys. J.* 93:1255–1263.
49. Perez-Rigueiro, J., C. Viney, …, M. Elices. 2000. Mechanical properties of single-brin silkworm silk. *J. Appl. Polym. Sci.* 75:1270–1277.
50. Cunniff, P. M., S. A. Fossey, …, D. L. Vezie. 1994. Mechanical and thermal properties of dragline silk from the spider *Nephila clavipes*. *Polym. Adv. Technol.* 5:401–410.
51. Agirrezabala, X., J. Martín-Benito, …, J. L. Carrascosa. 2005. Structure of the connector of bacteriophage T7 at 8Å resolution: structural homologies of a basic component of a DNA translocating machinery. *J. Mol. Biol.* 347:895–902.
52. Cuervo, A., and J. L. Carrascosa. 2012. Viral connectors for DNA encapsulation. *Curr. Opin. Biotechnol.* 23:529–536.
53. Cuervo, A., M. C. Vaney, …, L. Oliveira. 2007. Structural rearrangements between portal protein subunits are essential for viral DNA translocation. *J. Biol. Chem.* 282:18907–18913.
54. Lebedev, A. A., M. H. Krause, …, A. A. Antson. 2007. Structural framework for DNA translocation via the viral portal protein. *EMBO J.* 26:1984–1994.
55. Lhuillier, S., M. Gallopin, …, S. Zinn-Justin. 2009. Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. *Proc. Natl. Acad. Sci. USA.* 106:8507–8512.
56. Olia, A. S., P. E. Prevelige, Jr., …, G. Cingolani. 2011. Three-dimensional structure of a viral genome-delivery portal vertex. *Nat. Struct. Mol. Biol.* 18:597–603.
57. Veesler, D., and C. Cambillau. 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiol. Mol. Biol. Rev.* 75:423–433.
58. Römer, L., and T. Scheibel. 2008. The elaborate structure of spider silk: structure and function of a natural high performance fiber. *Prion.* 2:154–161.
59. Addadi, L., and S. Weiner. 1992. Control and design principles in biological mineralization. *Angew. Chem. Int. Ed. Engl.* 31:153–169.

60. Luz, G. M., and J. F. Mano. 2010. Mineralized structures in nature: examples and inspirations for the design of new composite materials and biomaterials. *Compos. Sci. Technol.* 70:1777–1788.
61. Grimes, S., S. Ma, …, P. J. Jardine. 2011. Role of ϕ29 connector channel loops in late-stage DNA packaging. *J. Mol. Biol.* 410:50–59.
62. Honig, B., and A. S. Yang. 1995. Free energy balance in protein folding. *Adv. Protein Chem.* 46:27–58.
63. Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.
64. Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29:7133–7155.
65. Eisenhaber, F. 1996. Hydrophobic regions on protein surfaces. Derivation of the solvation energy from their area distribution in crystallographic protein structures. *Protein Sci.* 5:1676–1686.
66. Gräter, F., and H. Grubmüller. 2007. Fluctuations of primary ubiquitin folding intermediates in a force clamp. *J. Struct. Biol.* 157:557–569.
67. Fernandez, J. M., and H. Li. 2004. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science.* 303:1674–1678.
68. Kappel, C., U. Zachariae, …, H. Grubmüller. 2010. An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. *Biophys. J.* 99:1596–1603.
69. Kappel, C., N. Dölker, …, H. Grubmüller. 2012. Universal relaxation governs the nonequilibrium elasticity of biomolecules. *Phys. Rev. Lett.* 109:118304.
70. Zink, M., and H. Grubmüller. 2010. Primary changes of the mechanical properties of Southern Bean Mosaic Virus upon calcium removal. *Biophys. J.* 98:687–695.
71. Amadei, A., A. B. Linssen, and H. J. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.
72. Berendsen, H. J., and S. Hayward. 2000. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* 10:165–169.
73. Eswar, N., B. Webb, …, A. Sali. 2007. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein. Sci.* Chapter 2:Unit 2.9. http://dx.doi.org/10.1002/0471140864.ps0209s50.
74. Fernandez-Fuentes, N., J. Zhai, and A. Fiser. 2006. ArchPRED: a template based loop structure prediction server. *Nucleic Acids Res.* 34 (Web Server issue):W173–W176.
75. García, A. E. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699.
76. Hayward, S., A. Kitao, …, N. Go. 1993. Effect of solvent on collective motions in globular protein. *J. Mol. Biol.* 234:1207–1217.
77. Hess, B. 2000. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E.* 62 (6 Pt B):8438–8448.
78. Kitao, A., F. Hirata, and N. Gō. 1991. The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.* 158:447–472.
79. Macke Thomas, J., and A. Case David. 1997. Modeling unusual nucleic acid structures. *In* Molecular Modeling of Nucleic Acids American Chemical Society, Washington, DC, pp. 379–393.
80. Pérez, A., I. Marchán, …, M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* 92:3817–3829.

4.3. Supporting Material

Contents

Sections

1. Description of connector loops modeling and refinement 41
2. Structural and conformational equilibration of the connector 42
3. Elastic properties determined from force probe simulations 43

Tables

Table S1: Heating-cooling cycles performed in SA MD simulations 47
Table S2: Force rates and torque/force constants applied in FP simulations 47
Table S3: Details of compression-stretching umbrella sampling simulations 48
Table S4: Details of twisting-untwisting umbrella sampling simulations 49
Table S5: Dimensions of the connector and its middle region used to calculate the Young's modulus of elasticity 50
Table S6: Mechanical properties of the whole connector and its middle region 50

Figures

Figure S1: Crystal structure of the connector 51
Figure S2: Refinement of the connector loops 51
Figure S3: Starting non-equilibrium conformations used to characterize relaxation behaviors 52
Figure S4: Collected histograms from compression-stretching umbrella sampling simulations 52
Figure S5: Torsional harmonic constants and collected histograms from twisting-untwisting umbrella sampling simulations 53
Figure S6: Geometrical characterization of the truncated hollow cone model used for calculating the Young's modulus of elasticity 53
Figure S7: Root Mean Square Deviations of the connector from equilibrium simulations 54
Figure S8: Deviations in twist angle during equilibrium simulations 54
Figure S9: Convergence in the Young's modulus during the simulations 54
Figure S10: Elastic properties of the whole connector derived from FP simulations 55
Figure S11: Elastic properties of the middle region derived from FP simulations 56
Figure S12: Convergence of relaxation paths towards an equilibrium 56
Figure S13: Reversible recovery of the equilibrium conformation 57
Figure S14: Free energy landscape from equilibrium fluctuations 57
Figure S15: Convergence of deformation paths in umbrella sampling simulations 58
Figure S16: Deformation dependent changes in hydrophobic solvent accessible surface areas 58
Figure S17: Residue packing dependent changes in interaction energies 59
References 59

1. Description of connector loops modeling and refinement

The connector loops A230-S244 were unresolved in the X-ray crystal structure (Fig. S1A) and therefore modeled as follows. First, an initial structure model was obtained for one of the twelve symmetry-related missing loops using the ArchPred server (1). This seed structure was then replicated (according to the 12-fold symmetry of the connector) and manually integrated within all other remaining sub-units after aligning the seed structure (see Fig. S1B). In addition, four of the twelve subunits of the connector lacked residues (Q166-L169) in the bottom region. These residues were modeled using Modeller program (2) by taking a template subunit in which these residues are present.

Also, in a structure of the connector in complex with DNA, the diversity in DNA-loop interactions will cause the 12 connector loop structures to slightly differ from each other because of the lack of 12-fold symmetry in the DNA helical grooves. To model this structural heterogeneity, all 12 connector loops were refined by a combined simulated annealing and molecular dynamics (SA/MD) protocol as described in following.

First, the viral DNA of the φ29 bacteriophage gp10 gene (NCBI Reference Sequence: NC_011048.1) was modeled into the connector channel center. The first 60 nucleotides (ATG GCA CGT AAA CGC AGT AAC ACA TAC CGA TCT ATC AAT GAG ATA CAG CGT CAA AAA CGG) were modeled as B-DNA strand using the Nucleic Acids Builder NAB (3) and is depicted as *yellow* and *blue* ring (top view) in Fig. S1B.

Next, the initial molecular system was prepared for the Simulated Annealing Molecular Dynamics (SA MD) simulations: The connector-DNA complex was set in the center of a dodecahedron periodic box, solvated with 241,067 water molecules, and neutralized by addition of a sufficient number of Sodium ions (in total 202). The system was energy-minimized and subsequently heated during a NVT simulation of 500 ps with a 1 fs time step. Pressure was equilibrated during a subsequent NPT simulation of 1 ns with a 2 fs time step. In both simulations, the positions of all heavy atoms were restrained by an harmonic force constant of 1000 kJ/(mol nm^2). 40 SA cycles of 252 ps length were performed for 10 ns during the SA MD simulations. In each SA cycle, the loops were heated up in two steps to a temperature of 1000 K and subsequently cooled down in seven steps to 300 K (Table S1). Only the loops were free to move during the SA MD simulations, whereas other heavy atoms of the connector-DNA complex were restrained at the starting position by a force constant of 1000 kJ/(mol nm^2) to preserve the structure of the complex at high temperatures.

The SA MD and below described cooling simulations were performed using the GROMACS 4.0.7 package (4), in which AMBER ff99SB (5) and parmbsc0 (6) force fields were used for connector and DNA, respectively. The temperature was regulated during heating and cooling by Berendsen temperature coupling, whereas the pressure was maintained at 1 atm by the Berendsen pressure coupling (7). Long range electrostatic interactions were computed by applying the PME method with a grid spacing of 1.2 Å and a 4th order of cubic interpolation (8). Short range non-bonded interactions were computed for the atom pairs within a distance of 10 Å. The temperature used in SA MD simulations flattened the energy landscape and facilitated random changes in loop conformations that were collected during 10 ns. To assess the convergence of loop conformations during MD simulations, principal component analyses (PCA) were performed (9-13) on the last 6 ns of each SA MD trajectory. The first two principal components (PC) were used because these have the largest variance and hence show the largest conformation changes.

Figure S2A shows the two-dimensional plane of the first PC with respect to the second PC. The semi-circle shape of the plane reveals randomness in conformations along the first two PCs (14). The obtained high temperature conformations (red dots in Fig. S2A) are most likely located at a high energy region on the free energy landscape and were not used for MD simulations. However, cooling of these conformations is expected to allow achieving nearest free energy minimum loop conformations. To this aim, five different high temperature conformations were chosen from the projection planes of PC1 and PC2 (A, B, C, D and E in Fig. S2A). These conformations were cooled from 1000 K down to 300 K in 6 ns and subsequently equilibrated for 1 ns at 300 K in a NPT simulation (cooling simulations). Finally, five clusters of structures were obtained (A′, B′, C′, D′ and E′ in Fig. S2A) at 300 K. These were compared by computing root mean square deviations (RMSD) of C-alpha atoms of the loops with respect to the initial structure model (five different blue symbols in Fig. S2B) and found to be within 0.2 nm difference. The deviation is not significant for these highly flexible loops. Consequently, the central structure from this cluster was used for further studies. The final loop conformations and DNA-loop interactions are illustrated in Fig. S2C.

2. Structural and conformational equilibration of the connector

We monitored the structural stability of the connector during equilibrium MD simulations via root mean square deviation (RMSD) calculated for C-alpha atoms with respect to the crystal structure (Fig. S7). After an initial sharp rise during the first 10 ns, the RMSD-values stabilize at about 0.25 nm during the subsequent 190 ns. Averaged over the last 190 ns of the equilibration phase, a twist angle of $\theta = 77.7°$ and a length of $L = 4.89$ nm was obtained, with a standard deviation of 0.4° and 0.04 nm, respectively (Fig. 2A). Because of the RMSD drift observed during the first 100 ns, average values were also computed for the last 100 ns, and a similar twist angle ($\theta = 77.6°$) and length ($L = 4.91$ nm) were obtained. Whereas the connector length remained at that of the crystal structure ($L = 4.91$ nm), the twist angle increased by ca. 3.5° with respect to the crystal structure ($\theta = 74.2°$) during the first 10 ns. Closer inspection showed that the deviation was actually due to an increased tilt of the helices as depicted in Fig. S8, and in particular those helices which are in contact with symmetry mates in the crystal structure. This increased tilt of those helices slightly shifted the center of mass of the bottom region of each subunit that was used to calculate the twist angle (see Figs. 1C and D). We therefore assume that this tilt in the crystal structure is due to crystal contacts between these alpha helices and the upper region. The structures of the middle and upper regions are largely unaffected. Given the size of the connector, we consider the observed RMSD of ca. 0.25 nm and the underlying structural changes during equilibration rather small. Because the overall structure and the RMSD remained unchanged during the last 100 ns, this part of the trajectory was used for further analysis.

3. Elastic properties determined from force-probe simulations

3.1. Methods

3.1.1. Force probe MD simulations

To study the mechanical properties of the connector by force probe simulations, the structure was subjected to external forces via two structural descriptors (see Fig. 1D), the twist angle θ_m which describes the twisting-untwisting motion of the middle region of the connector, and its length L_m, which quantifies the compression-stretching motion of the middle region. The upper and lower disk in Fig. 1E represents Cα atom from the upper- and lower-terminus of MH1 and MH2 helices of each subunit, respectively. We will refer to the atoms of the lower and upper disk as pull or rotational group and reference group in the subsequent sections, respectively. Rotation between two disks defines the twisting-untwisting motion while motion of disks along channel axis shows the compression-stretching motion.

Six force-probe MD simulations (FP-1, FP-2, FP-3, FP-4, FP-5 and FP-6) were performed to investigate the twisting-untwisting and compression-stretching motions of the connector beyond the equilibrium fluctuations (see details in Table. S2). All force probe simulations started from the equilibration trajectory after 25 ns. To drive the compression and stretching motion, the center of mass of the pull group (lower disk in Fig. 1D) was moved away (stretching) and towards (compression) the center of mass of the reference group (upper disk in Fig. 1D). To drive the untwisting and twisting motions, the atoms of the rotational group (lower disk in Fig. 1D) were forced to move along circles centered at the z-axis by applying a torque potential in both anti-clock wise (untwist) and clock wise (twist) direction, respectively (grey arrow in Fig. 1D). At the same time, the center of mass of the reference group was kept fixed by a harmonic restraint of 1000 kJ.mol^{-1}.nm^{-2}). As defined in Ref. (15) the variant *rm2-pf* was chosen as rotational potential, which allowed the radial motion of the rotational group atoms (15). All simulations were carried out using a modified version of GROMACS in which rotational pulling was implemented (15). The total simulation time was ~160 ns.

3.1.2. Relaxation simulations

To study within which regime the observed deformation is reversible and hence elastic, two intermediate structures of 5.1 and 5.3 nm length, were chosen from the FP-6 simulation (see Figs. S3B and C) for subsequent relaxation simulations (Relax-1 and Relax-2, respectively).

To equilibrate the structure at the starting twist angle and length, all atoms of pull and reference group were restrained by force constant of 1000 kJ/(mol nm^2) for the first 4.4 ns. The force was then removed, and the connector was allowed to relax freely, during which period the approach to its equilibrium length and twist angle was monitored. Furthermore, to check the structural reversibility of the compressed connector, an intermediate representative structure of the connector with a middle region's length L_m=2.7 nm was taken from the 7.52 ns of the FP-5 simulation (see Fig. S3A) for a subsequent relaxation simulation (Relax-3). At this specific snapshot of the FP-5 trajectory, the length L_m=2.699 nm was nearest to the required value of L_m=2.7 nm. In this relaxation simulation, the connector was kept at its enforced twist angle and length for 3.3 ns before the pulling and restraint potentials were removed in order to allow relaxation towards its equilibrium length. An additional relaxation simulation (Relax-4) was performed from FP-6 simulation to obtain relaxed structures of the connector which was later used for the umbrella sampling simulations.

3.2. Results and discussion

In the first set of simulations (FP-1 and FP-2), the connector was compressed and stretched by applying linear pulling forces as described in the methods section (Fig. 1D and Table S2). The required forces as well as the resulting connector twist angle θ as a function of connector length L was observed (Fig. S10A, red symbols). Indeed, for the range accessible to equilibrium fluctuations (black ellipse) as well as for larger compressions beyond the equilibrium fluctuations (left branch of the curve), the connector's twist angle changes by 2°/nm (Table S6), and is herewith similar to the one determined above (dashed line in Fig. S10A). Up to the maximum exerted force of 5000 pN, and at a compression by ca. 4%, the force extension curve (Fig. S10B, red line) is linear, indicating Hookean behavior within this range.

For stretching beyond the equilibrium range (outside of ellipse in Fig. S10A), an unexpectedly large untwisting is observed, by 13.5°/nm up to 5.1 nm length, with a continued linear (Hookean) force. Beyond that critical force, the structure started to break down, which is also reflected in the decreased slope of the force extension curve and may indicate non-elastic behavior. Interestingly, the coupling between extension and untwisting remains linear also beyond the critical force. Within the Hookean regime, a spring constant of ~27200 pN/nm (red line in Fig. S10B) is observed.

The large untwisting motion beyond L=5.05 nm is unexpected because non-equilibrium relaxation would likely make the untwisting lag behind its equilibrium pathway (dashed curve in Fig. S10A), in contrast to the observed increased slope. Note, however, that this additional untwisting is solely due to structural re-arrangements of the bottom region of the connector, which is not subjected to the pulling force (cf. inset in Fig. S10A). Indeed, the untwisting motion of the middle region alone displays linear coupling up to full extension (see red line Fig. S11A). It remains to be tested whether or not this linear pathway resembles the equilibrium path.

In a second set of FP simulations (FP-3 and FP-4), the connector was twisted and untwisted by applying a torque (see Methods Section), and the required torque as well as the length change as a function of twist angle was monitored (blue symbols, Figs. S10A and C). Beyond the equilibrium fluctuations, the connector twist angle θ changed linearly with length L by 27.9°/nm, markedly larger than that calculated from equilibrium fluctuations (Table S6). By exceeding an untwisting of 70^{o} the structure started to break down. The torque-angle curve (blue line, Fig. S10C) is Hookean between 75^{o} and 78^{o}, with a torsional spring constant of ~3500 pN nm/^{o}.

In all of the above simulations (FP-1 to FP-4), the obtained stretching and torsional spring constants were larger than those derived from equilibrium simulations. We assume this discrepancies to be caused by non-equilibrium effects in force probe simulations, which are absent in equilibrium simulations. Specifically, during the relatively short time of about ten nanoseconds for each simulation, relaxation is likely to be incomplete for several of the internal degrees of freedom of the connector, as previously described and quantified in terms of a two-dimensional time-dependent transition state theory also for the connector (16). In this scenario, as a result of this incomplete relaxation, the twist angle θ should 'lag behind' (i.e., larger θ values) a fully equilibrated pathway along the minimum ('valley') of the underlying untwisting-stretching free energy landscape. Vice versa, when untwisting the connector (blue symbols, Fig. S10A), the length expansion is expected to lag behind, which explains why this untwisting-stretching curve falls below that of the stretching simulations. Accordingly, the equilibrium pathway is expected in between these two extremes.

We tested this hypothesis by performing several relaxation simulations, starting from different points along the stretching and untwisting simulations, in which the pulling force or torque, respectively, was removed. If relaxation effects dominate the observed differences between stretching and untwisting, one would expect the relaxation trajectories to deviate from the force probe trajectories and to converge towards their common equilibrium pathway. Indeed, Fig. S12 clearly shows such convergence. Both twist angle and length changed rapidly towards an intermediate pathway, and almost reach an equilibrium region within ca. 10 to 20 ns. Notably, the slope of the converged line is larger than that of the equilibrium coupling, which suggests that the coupling mechanism between length change and untwisting outside the equilibrium region differs from that in thermal equilibrium.

To approximate this equilibrium path, which will also be used for subsequent deformation free energy calculations, two further force probe simulations (FP-5 and FP-6) were performed in which both stretching forces and torque were applied simultaneously. Because the above relaxation simulations suggest that the equilibrium path is located approximately right in between the paths probed by simulations FP 1-2 and 3-4, respectively, the same ratio between pulling and rotation speed was chosen here. To allow for more complete relaxation, a ten times reduced pulling force and torque rate (Table S2) was used.

The obtained length and twist angle changes are shown in Fig. S10A (green symbol). As expected, the resulting deformation pathway falls right in between the paths obtained in the stretching-only, untwisting-only and relaxation simulations. As an example, movie S1 shows an animation of trajectory FP-6. Linear coupling is observed, with a coupling coefficient of 18°/nm throughout the whole stretching/untwisting process (green line in Fig. S10A). Also in these simulations, structural break-down occurred at $L = 5.15$ nm (shown in movie S2). No linear coupling between compressions and twisting is observed; rather, the twist angle saturates at 79.5° during compression. Within the regime of equilibrium fluctuations, Hookean behavior of both the stretching and untwist is observed (Figs. S10B and C, green symbols), with stretching and torsional spring constants similar (Table S6) to those obtained from simulations FP-1/2 and FP-3/4, respectively. For extensions exceeding 5.0 nm length and twist angle below 76°, respectively, non-Hookean behavior sets in. Remarkably, the change of the connector's mechanical properties is very abrupt at this critical deformation, with no additional torque being required for further large connector untwisting.

To test whether the observed non-Hookean behavior also implies non-elastic behavior for both the untwisting-stretching motion as well as the observed structural break-down, we have carried out further relaxation simulations (see methods section), starting from non-equilibrium conformations extracted from simulations FP-5 and FP-6 at lengths L of 4.75, 5.1, and 5.3 nm, respectively (the corresponding structures are shown in Fig. S3). Figure S13A shows the obtained relaxation motions in terms of both length L and twist angle θ during the final part of the restrained simulation (left of the dashed line) and after removal of the restraining potential (right of dashed line). The equilibrium values are regained in all three simulations. Even the extreme non-Hookean deformation of up to $L = 5.3$ nm length approaches the equilibrium value after 100 ns, despite the structural deformations shown in Fig. S3C (shaded box) and Movie S2, which demonstrates that the connector is able to recover its equilibrium structure even after structural break down. As can be seen in Figs. S13B and C, all three relaxation trajectories remain close to the untwisting/stretching paths of simulations FP-5 and FP-6, which underscores that these simulations remained close to the 'valley' of minimal perturbation.

Overall, deformations within the range of L = 4.70-5.3 nm and θ = 70°-79.5° seem to be fully elastic, and the twisting-untwisting motion is linearly coupled to the compression-stretching motion within this elastic range. The obtained coupling of 18°/nm is remarkably close to the coupling of 17.6°/nm required in the untwist-twist DNA packaging model proposed by Simpson et al. (17). We therefore asked next if also the other elastic and energetic properties of the connector are compatible with this model.

Tables

Table S1: Heating-cooling cycles performed in SA MD simulations. The connector loops were heated up and cooled down consecutively 40 times during 10 ns of MD simulations. Each cycle consisted of 252 ps with two and eight steps of heating and cooling, respectively.

Temperature (K)	Time period (ps)	Cumulative time (ps)	Temperature (K)	Time period (ps)	Cumulative time (ps)
300	20	20	600	20	162
300-600	2	22	600-500	2	164
600	22	44	500	20	184
600-1000	2	46	500-450	2	186
1000	50	96	450	20	206
1000-800	2	98	450-400	2	208
800	20	118	400	20	228
800-700	2	120	400-350	2	230
700	20	140	350	20	250
700-600	2	142	350-300	2	252

Table S2: Force rates and torque/force constants applied in FP simulations. Corresponding simulation names are used throughout the main article and the supplementary text.

Name	Force probe simulations (along motion type)	Rate		Force constant [kJ/(mol nm^2)]	
		Rotational (°/ps)	Pulling (nm/ps)	Rotational	Pulling
FP-1	Compression*	—	0.010	—	100
FP-2	Stretching*	—	0.010	—	100
FP-3	Twisting*	0.0040	—	500	—
FP-4	Untwisting*	0.0040	—	500	—
FP-5	Twisting and compression	0.0004	0.001	500	100
FP-6	Untwisting and stretching	0.0004	0.001	500	100

*These simulations were performed three times in parallel

Table S3: Details of compression-stretching umbrella sampling simulations. The simulation type from which the reference starting conformation for each sampling window was extracted is denoted accordingly. The frame time corresponds to the time of the reference simulation from which the reference structure was taken. The relaxation simulations Relax-1, Relax-2, Relax-3, and Relax-4 were performed using the non-equilibrium starting structures from simulations FP-5 and FP-6.

Sampling window	Reference simulation type	Frame time (ps)	Reference reaction coordinate, L_m (nm)	Force constant [kJ/(mol nm^2)]
1	Relax-3	3300	2.725	2500
2	Equilibrium	107000	2.750	500
3	Equilibrium	74200	2.800	500
4	Relax-1	16940	2.850	2500
5	Relax-1	4460	2.876	5000
6	Relax-2	16420	2.900	7500
7	Relax-2	8000	2.950	7500
8	Relax-2	5320	2.997	10000
9	Relax-2	4720	3.025	15000
10	Relax-2	4580	3.052	15000
11	Relax-4	5200	3.075	17500
12	Relax-4	5140	3.100	20000
13	Relax-4	5060	3.125	20000
14	Relax-4	5040	3.149	20000

Table S4: Details of twisting-untwisting umbrella sampling simulations. The simulation type from which the reference starting conformation for each sampling window was extracted is denoted accordingly. The frame time corresponds to the time of the reference simulation from which the reference structure was taken.

Sampling window	Reference simulation type	Frame time (ps)	Reference reaction coordinate, θ_m°	Force constant [kJ/(mol.nm^2)]
1	FP-6	24700	29.55	10000
2	FP-6	24090	30.02	7500
3	FP-6	22880	30.50	6000
4	FP-6	21570	31.01	5000
5	FP-6	18520	31.57	5000
6	FP-6	18310	32.02	2500
7	FP-6	12522	32.44	2500
8	FP-6	11362	33.13	1000
9	FP-6	8772	33.45	500
10	FP-6	7042	33.90	500
11	FP-5	1576	34.50	1000
12	FP-5	5566	35.00	6500
13	FP-5	9748	35.50	6500
14	FP-5	11030	36.00	10000

Table S5: Dimensions of the connector and its middle region used to calculate the Young's modulus of elasticity. Values were obtained from equilibrium MD simulations using Eq. 4. Notations are in accordance with labels used in the truncated cone sketch shown in Fig. S6. Given error represents the SE obtained using the block-averaging method (18).

Connector region	D_a [nm]	D_b [nm]	d_a [nm]	d_b [nm]	L [nm]
Whole connector	6.259 ± 0.015	14.720 ± 0.012	4.229 ± 0.103	5.866 ± 0.011	6.578 ± 0.025
Middle region	5.194 ± 0.044	9.268 ± 0.012	3.668 ± 0.028	7.188 ± 0.005	2.861 ± 0.002

Table S6: Mechanical properties of the whole connector and its middle region. Equilibrium mechanical properties were obtained using Eq. 2. Best-fit lines within the linear regime are depicted in Figs. S10B, S10C, S11B and S11C. Couplings between untwisting-stretching motions were obtained in a similar manner and are shown in Figs. 2, S10A and S11A.

Simulation type	Connector region	K_θ [(pN nm)/Deg^2]	K_L [pN/nm]	K_c [pN/Deg]	Coupling (Deg/nm)
Equilibrium	Whole connector	1347	3073	2715	2.0
	Middle region	3374	24109	14848	4.4
FP 1 and 2	Whole connector	—	27204	—	2.0 and 13.5
	Middle region	—	33437	—	4.2
FP 3 and 4	Whole connector	3558	—	—	27.9
	Middle region	3404	—	—	51.3
FP 5 and 6	Whole connector	3033	27747	—	18.1
	Middle region	3428	28282	—	17.7

Figures

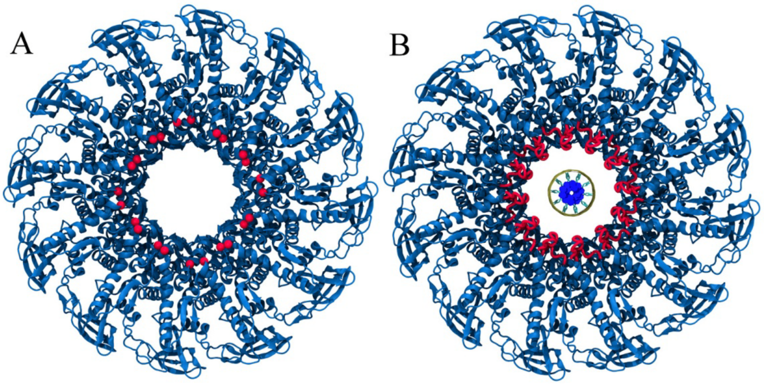


Figure S1: Crystal structure of the connector in top view. (A) Red spheres denote the location of the missing loops. (B) Modeled loops (*red*) are not in contact with the DNA (*yellow* and *blue* ring at the center of channel).

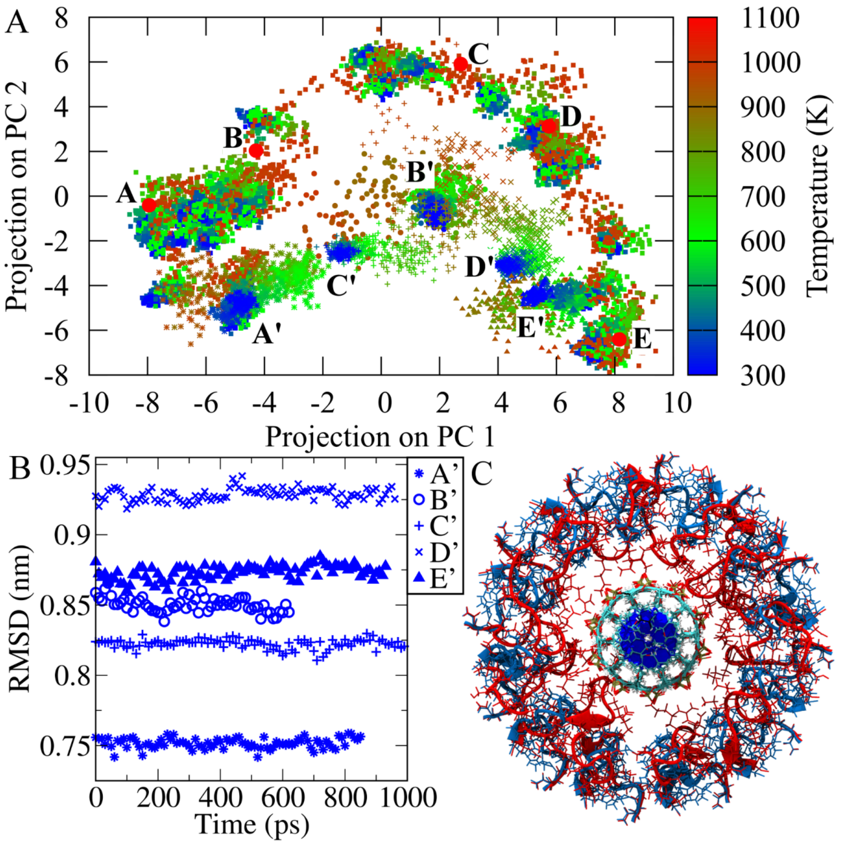


Figure S2: Refinement of the connector loops. (A) Plane of projection between first and second principal components (PC1 and PC2) obtained after performing PCA on the SA MD trajectories (*squares*). Five high temperature conformations A, B, C, D, and E (*red circles*) were selected from the projection plane and cooled down to 300 K. Dense clusters (*blue symbols*) labeled with A', B', C', D', and E' were obtained after cooling. (B) The structures of these five clusters (*blue symbols*) were compared by computing RMSDs of Cα atoms with reference to the starting loop structure models. (C) Comparison of modeled (*blue*) and refined loops (*red*) interacting with the DNA in the channel center.

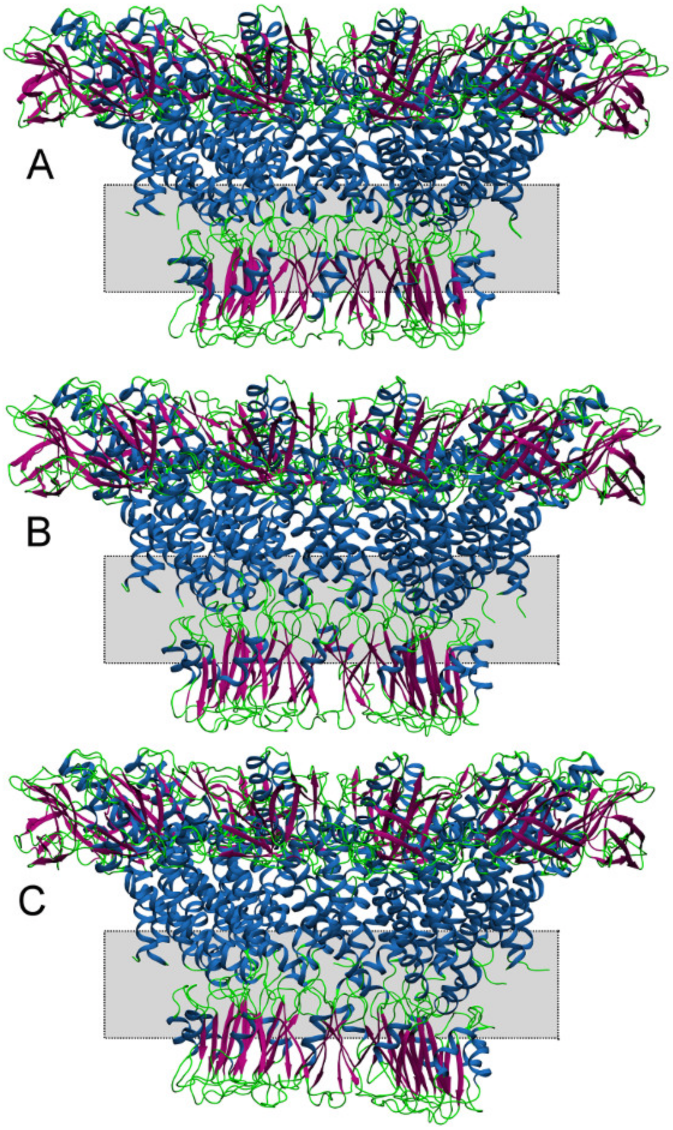


Figure S3: Starting non-equilibrium conformations used to characterize relaxation behaviors. Three non-equilibrium connector conformations of (A) 4.72 nm, (B) 5.1 nm, and (C) 5.3 nm length were selected for relaxation simulations. Region of structural deformations is shown as *shaded area*.

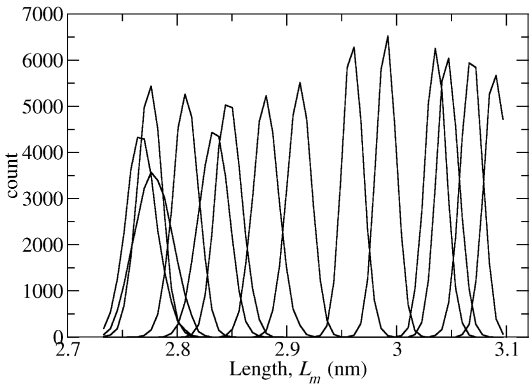


Figure S4: Collected histograms from compression-stretching umbrella sampling simulations. These histograms show the population of connector conformations along the reaction coordinate L_m during umbrella sampling simulations and were used to calculate the deformation free energy using WHAM (19-21).

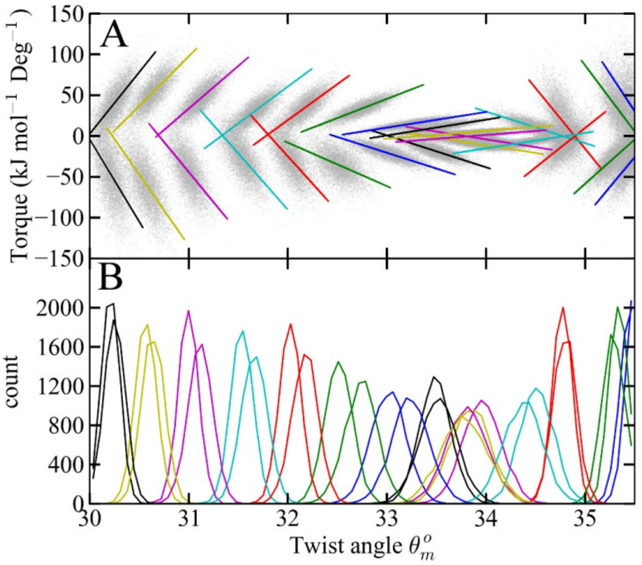


Figure S5: Torsional harmonic constants and collected histograms from twisting-untwisting umbrella sampling simulations. The group of atoms was harmonically restraint by applying a tangential harmonic force constant (kJ mol^{-1} nm^{-2}) during umbrella sampling simulations (15). Colors represent different sampling windows. The torsional harmonic force constants (kJ mol^{-1} deg^{-2}) were calculated from best line fits of the torque-twist angle values and were used to derive the deformation free energies for twisting-untwisting motions using WHAM (19-21). (A) Torque with respect to the twist angle θ_m during umbrella sampling simulations. Colored lines are the best fit lines of the respective sampling window. (B) Collected histograms showing the population of connector conformations along the reaction coordinate θ_m^o from each sampling window.

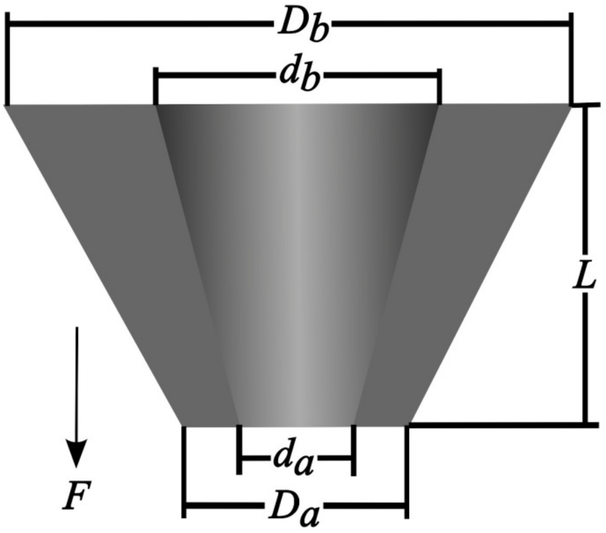


Figure S6: Geometrical characterization of the truncated hollow cone model used for calculating the Young's modulus of elasticity. The sketch shows the truncated hollow cone of length L, where D_a and d_a denote the narrow end exterior and interior diameter, respectively; D_b and d_b denote the wide end exterior and interior diameter, respectively. Equation 4 was derived on the basis of this model by assuming that the force F is acting in arrow direction.

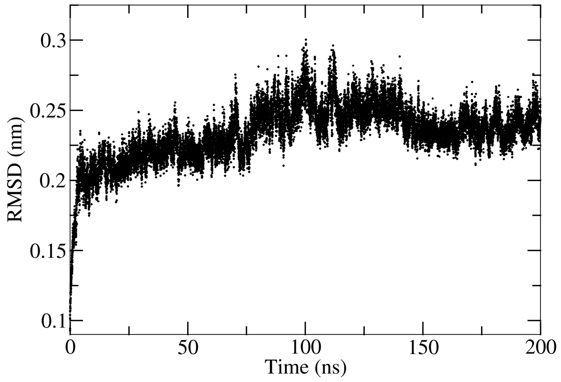


Figure S7: Root Mean Square Deviations of the connector from equilibrium simulations. RMSDs were calculated for Cα atoms of the connector with reference to the X-ray crystal structure (PDB ID: 1H5W).

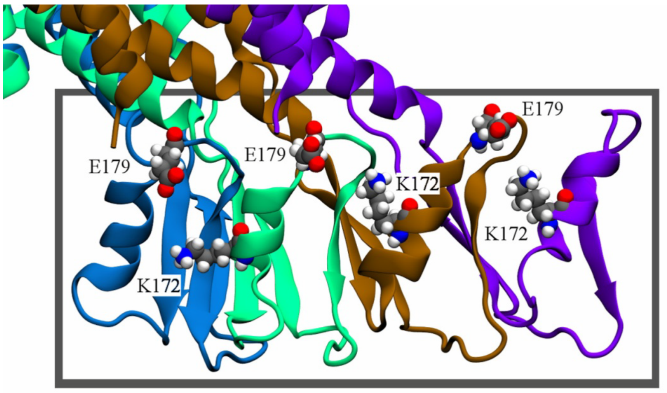


Figure S8: Deviations in twist angle during equilibrium simulations. During MD simulations, the twist angle of the connector deviated by 3.5° with respect to the crystal structure (PDB ID: 1H5W). The bottom connector region rotated with respect to the remaining part of the connector during the first 10 ns of the simulation. The grey box highlights the bottom region of four subunits that are depicted as colored cartoon representations. Residues E179 and K172 (*space-fill*) are mostly responsible for tilting of encasing α-helix.

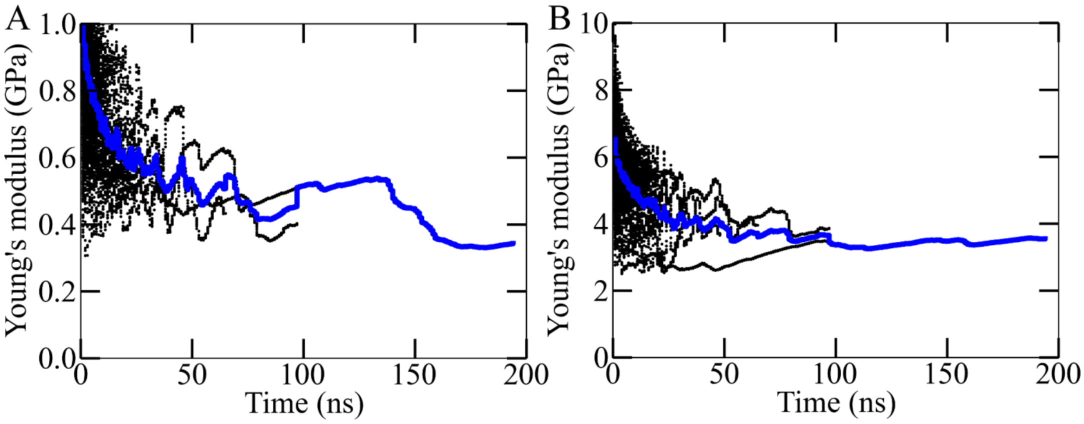


Figure S9: Convergence in the Young's modulus during the simulations. Both the time-blocks (black dots) as well as average (*blue line*) values of the Young's moduli with respect to the time are shown for (A) the whole connector and (B) the middle region. The moduli were calculated using Eq. 4 and, the required average of the dimensions (Fig. S6) over time and the stretching spring constant were obtained from non-overlapping time-blocks of the MD trajectory. The obtained modulus from each block is shown as a black dot. The average Young's modulus over the respective block size is shown as a blue solid line.

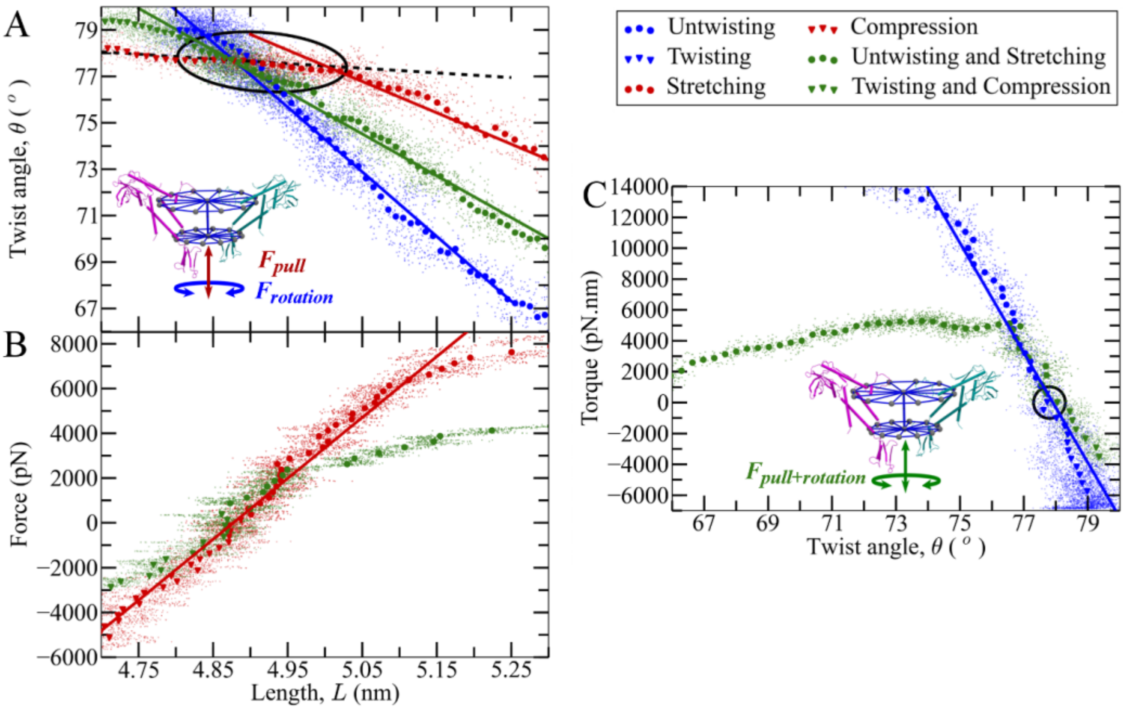


Figure S10: Elastic properties of the whole connector derived from force probe (FP) simulations; fluctuating instantaneous lengths and twist angles are shown as small dots, solid symbols represent averages over intervals in *x*-direction. The insets indicate pulling forces (F_{pull}) and torques ($F_{rotation}$) applied for different simulation sets FP-1/2 (*red arrow*), FP-3/4 (*blue arrow*), and FP-5/6 (*green arrows*) referred to in the text. (A) Enforced length change and/or twist: change in twist angle θ during enforced change of length L (*red*), length change during enforced twist/untwist (*blue*), twist angle and length change during simultaneous enforced twisting-compression and untwisting-stretching (*green*). The coupling between compression-stretching and untwisting-twisting motions during FP-1/2 (*red line*), FP-3/4 (*blue line*), and FP-5/6 (*green line*) is compared to the coupling (*black dashed line*) and the fluctuation range (*black ellipse*) obtained from equilibrium simulations. (B, C) Same data and coloring as (A), shown with applied force and torque and linear least square fits to the respective linear regime (*lines*). In (C), the black circle denotes the initial twist angle.

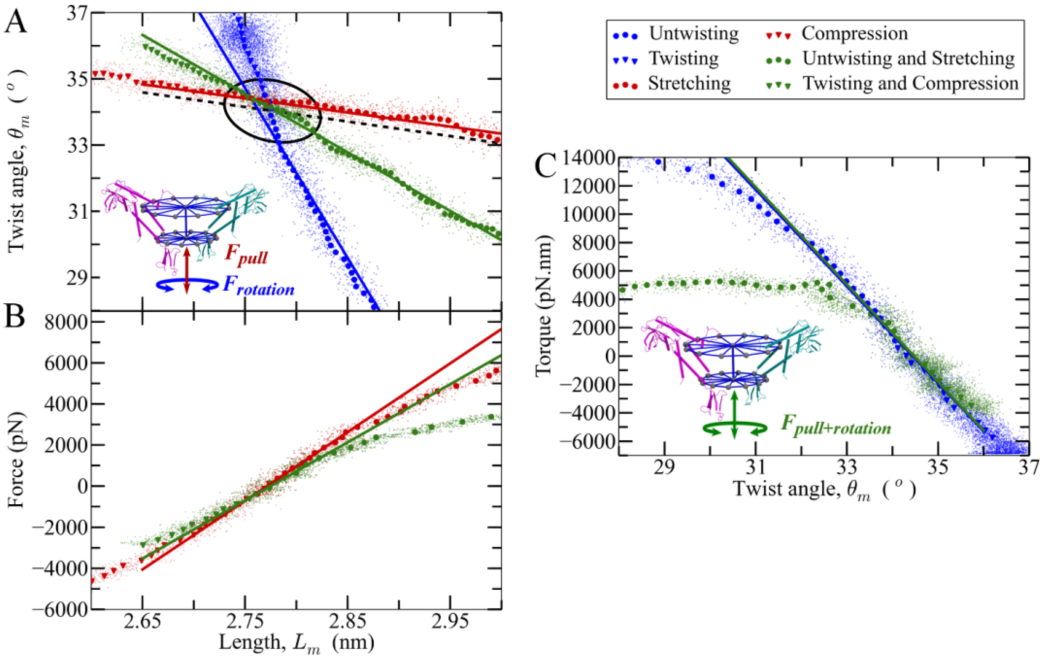


Figure S11: Elastic properties of the middle region derived from FP simulations; fluctuating instantaneous lengths and twist angles are shown as small dots, solid symbols represent averages over intervals in *x*-direction. Insets indicate applied pulling forces (F_{pull}) and torques ($F_{rotation}$) used in different simulation sets FP-1/2 (*red arrow*), FP-3/4 (*blue arrow*), and FP-5/6 (*green arrows*), as listed in Table S2. (A) Enforced length change and/or twist: change in twist angle θ during enforced change of length L (*red*), length change during enforced twist/untwist (*blue*), twist angle and length change during simultaneous enforced twisting-compression and untwisting-stretching (*green*). The coupling between compression-stretching and untwisting-twisting motions during FP-1/2 (*red line*), FP-3/4 (*blue line*), and FP-5/6 (*green line*) is compared to the coupling (*black dashed line*) and the fluctuation range (*black ellipse*) obtained from equilibrium simulations. (B, C) Same data and coloring as (A), shown with applied force or torque and linear least square fits to the respective linear regime (*lines*).

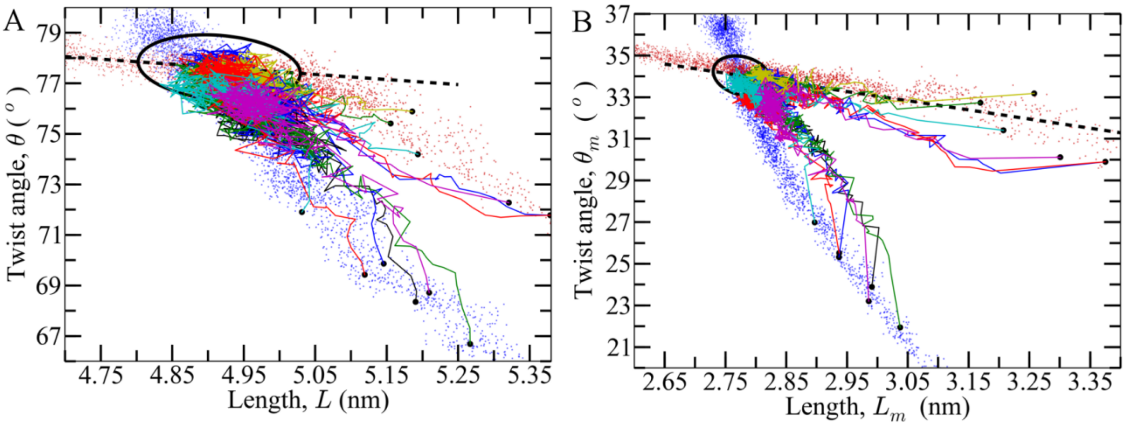


Figure S12: Convergence of relaxation paths towards an equilibrium. The comparison of relaxation paths (*colored lines*) with the obtained paths from enforced untwisting-only (*blue dots*) and stretching-only (*red dots*) motions for the (A) whole connector and (B) the middle region. Black dots denote starting non-equilibrium conformations taken from stretching simulation FP-2 (*red dots*) and untwisting simulation FP-4 (*blue dots*). To compare with equilibrium simulations, the range of fluctuations is denoted as black ellipse. Untwisting-stretching coupling is given by the slope of the best-fit line (*black dashed line*).

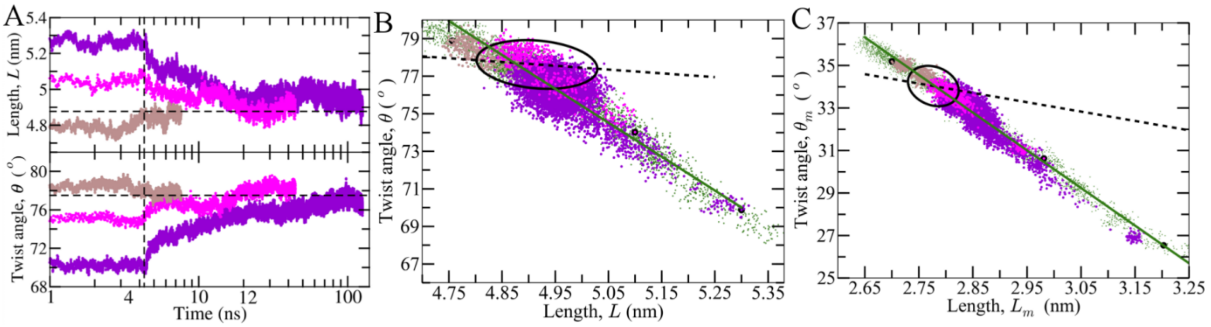


Figure S13: Reversible recovery of the equilibrium conformation. For three deformed connector structures with L=4.72, 5.1, and 5.3 nm (Figs. S3A-C), selected from simulations FP-5/6, their relaxation behavior (*purple, magenta,* and *light brown dots*) are illustrated. (A) The relaxation of connector's twist angle and length along the time towards the equilibrium structure (*horizontal dashed lines*) was observed for 120 ns (note the logarithmic scale). At 4.4 ns (*vertical dashed black line*) the restraint force was removed for complete relaxation. Relaxation pathways for (B) the whole connector and (C) the middle region towards the equilibrium structural fluctuations (*black ellipse*). Also, the comparisons of the relaxation paths with the obtained paths (green symbols and line) of untwisting-stretching motions from the same simulation set are shown. The coupling between compression-stretching and twisting-untwisting motions during FP-5/6 (*green line*) is compared to the coupling (*black dashed line*) obtained from the equilibrium simulations.

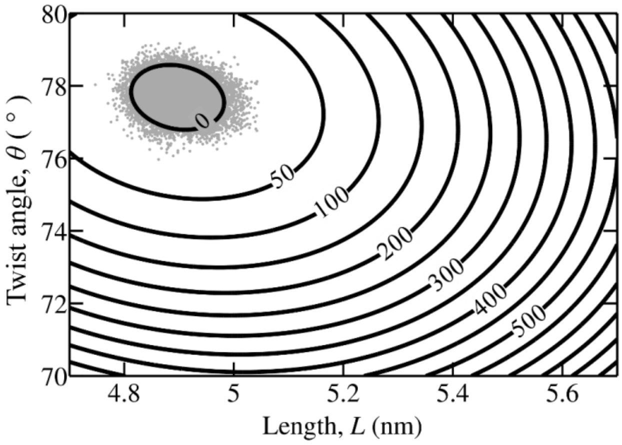


Figure S14: Free energy landscape from equilibrium fluctuations. The free energy of deformations for twisting-untwisting and compression-stretching motions was computed from equilibrium simulations under harmonic approximation of the energy landscape. The free energy (kJ/mol) was calculated from the probability density using the following equation $G(\theta,L) = -k_B T \ln[p(\theta,L)]$, where θ denotes the twist angle, L the length of the connector, k_B the Boltzmann constant, and T the temperature, respectively. The landscape was extrapolated beyond the equilibrium region to test the proposed untwist-twist DNA packaging mechanism.

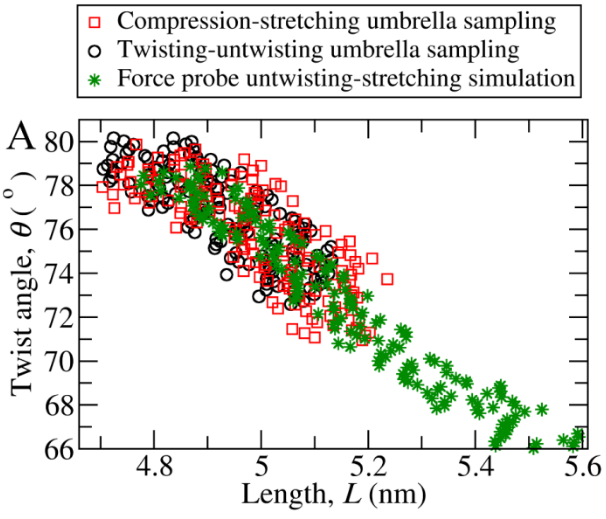


Figure S15: Convergence of deformation paths in umbrella sampling simulations. The pathways of untwisting-stretching motions were obtained from compression-stretching (*red squares*) and twisting-untwisting (*black circles*) umbrella sampling simulations and are compared to each other and to force probe simulation FP-6 (*green asterisks*) for (A) the whole connector and (B) the middle region.

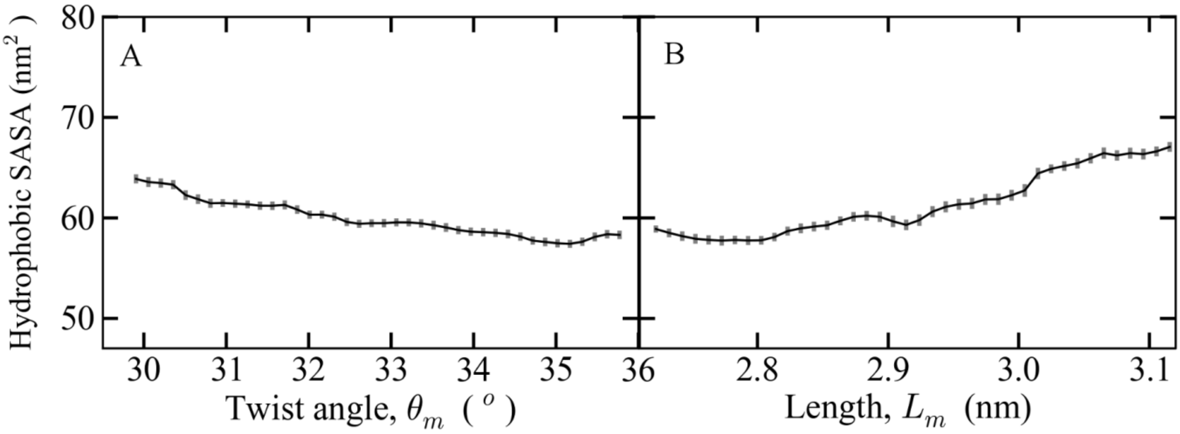


Figure S16: Deformation dependent changes in hydrophobic solvent accessible surface areas (SASAs) of the middle connector region derived from umbrella sampling simulations. Change in hydrophobic SASAs during (A) untwisting-stretching (twist angle θ_m) and (B) compression-stretching (length L_m) deformations. Hydrophobic SASAs were used to calculate the respective free energy. Error bars represent SE obtained from boot-strapping.

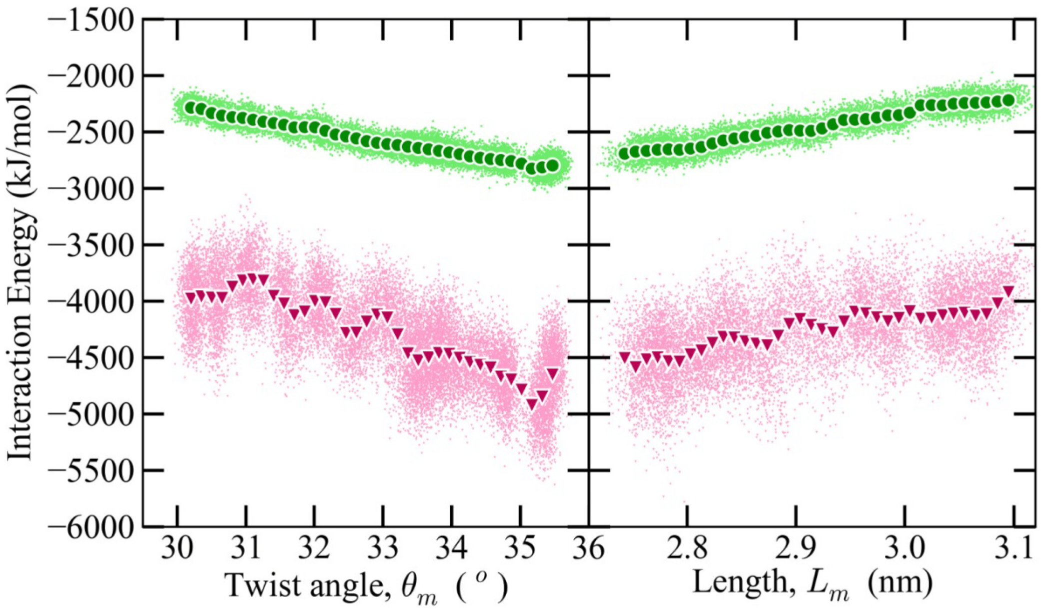


Figure S17: Residue packing dependent changes in interaction energies between deformed subunits of the middle connector region derived from umbrella sampling simulations. Fluctuating instantaneous interaction energies are shown as small dots; solid symbols represent averages over intervals in *x*-direction. Calculated van der Waals (*green circles*) and electrostatic (*red triangles*) interaction energies with respect to the deformation in (A) twist angle θ_m and (B) length L_m are depicted.

References

1. Fernandez-Fuentes, N., J. Zhai, and A. Fiser. 2006. ArchPRED: a template based loop structure prediction server. Nucleic. Acids. Res. 34:W173-176.

2. Eswar, N., B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, and A. Sali. 2007. Comparative protein structure modeling using MODELLER. Curr. Protoc. Protein. Sci. Chapter 2:Unit 2 9.

3. Macke Thomas, J. and A. Case David. 1997. Modeling Unusual Nucleic Acid Structures. In Molecular Modeling of Nucleic Acids. American Chemical Society. 379-393.

4. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory. Comput. 4:435-447.

5. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712-725.

6. Perez, A., I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, 3rd, C. A. Laughton, and M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys. J. 92:3817-3829.

7. Berendsen, H. J. C., J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. 1984. Molecular-Dynamics with Coupling to an External Bath. J. Chem. Phys. 81:3684-3690.

8. Darden, T., D. York, and L. Pedersen. 1993. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. J. Chem. Phys. 98:10089-10092.

9. Amadei, A., A. B. Linssen, and H. J. Berendsen. 1993. Essential dynamics of proteins. Proteins 17:412-425.

10. Berendsen, H. J. and S. Hayward. 2000. Collective protein dynamics in relation to function. Curr. Opin. Struct. Biol. 10:165-169.

11. Garcia, A. E. 1992. Large-amplitude nonlinear motions in proteins. Phys. Rev. Lett. 68:2696-2699.

12. Hayward, S., A. Kitao, F. Hirata, and N. Go. 1993. Effect of solvent on collective motions in globular protein. J. Mol. Biol. 234:1207-1217.

13. Kitao, A., F. Hirata, and N. Gō. 1991. The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. Chem. Phys. 158:447-472.

14. Hess, B. 2000. Similarities between principal components of protein dynamics and random diffusion. Phys. Rev. E Stat. Phys. Plasmas. Fluids. Relat. Interdiscip. Topics 62:8438-8448.

15. Kutzner, C., J. Czub, and H. Grubmuller. 2011. Keep It Flexible: Driving Macromolecular Rotary Motions in Atomistic Simulations with GROMACS. J. Chem. Theory. Comput. 7:1381-1393.

16. Kappel, C., N. Dolker, R. Kumar, M. Zink, U. Zachariae, and H. Grubmuller. 2012. Universal relaxation governs the nonequilibrium elasticity of biomolecules. Phys. Rev. Lett. 109:118304.

17. Simpson, A. A., Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2000. Structure of the bacteriophage phi29 DNA packaging motor. Nature 408:745-750.

18. Hess, B. 2002. Determining the shear viscosity of model liquids from molecular dynamics simulations. J. Chem. Phys. 116:209-217.

19. Hub, J. S., B. L. de Groot, and D. van der Spoel. 2010. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J. Chem. Theory. Comput. 6:3713-3720.

20. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. 1992. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. J. Comput. Chem. 13:1011-1021.

21. Roux, B. 1995. The Calculation of the Potential of Mean Force Using Computer-Simulations. Comput. Phys. Commun. 91:275-282.

Role of the Connector in the DNA Packaging

Chapter 5

5.1. Background

In the previous chapter, the untwist-twist DNA packaging model was examined, and the simulation results render this model unlikely. Two additional models, the one-way revolution and push-roll models are examined in this chapter. There are several following open questions concerning the connector (Fig. 5.1) and its proposed role in the DNA packaging mechanism, which is discussed in the two models. To address these questions, I performed several equilibrium (3x370 ns and 3x110ns) and force-probe MD simulations (12x32 ns and 12x120 ns) of the connector-DNA complex, and subsequently, analyzed the obtained trajectories. The starting structure of the connector-DNA complex, which was used for the simulations, is shown in Fig. 5.1.

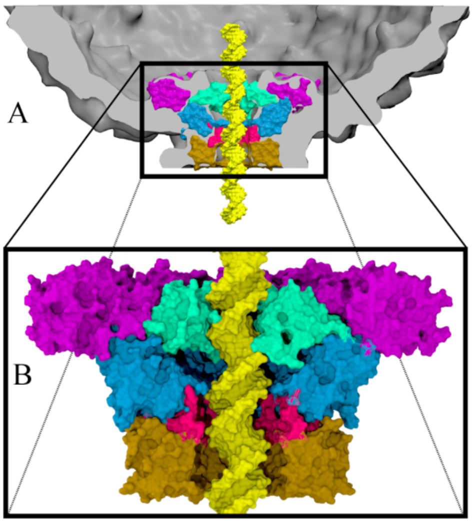


Figure 5.1: The connector-DNA complex with the viral procapsid. (A) Location of the connector with its five regions (*magenta*, *green*, *blue*, *red* and *brown*) in the procapsid (*grey*). The DNA (*yellow*) is present inside the connector channel. The connector's upper region (magenta) is inside the procapsid. The loop (*green*), middle (*blue*), hinge (*red*) and bottom (*brown*) regions form a channel, which contains the viral DNA (*yellow*). (B) Enlarged view of the connector with five regions and the DNA.

Both one-way revolution and push-role models were proposed by considering the crystal structure of the connector and B-form DNA helix. The structure of the connector with the DNA is not yet resolved and impact of the DNA on the connector structure and *vice versa* is not known. The bound connector-DNA structure may change with respect to the available structures because of the potential interactions, and eventually packaging models could be affected by these structural deviations. Therefore, first question I addressed here, what is the impact of the DNA on the connector structure and *vice versa* during the DNA packaging process? One-way revolution model proposes that the connector acts as a one-way valve and prevents the DNA leakage under high internal pressure. Here the question arises, how does the connector prevent the DNA leakage during the packaging process against a large counter pressure?

Also, do the four electropositive channel lysine rings (Figs. 1.7D and 5.2) facilitate the experimentally observed DNA transport of ~2.5 base-pairs per step (15) as proposed in the one-way revolution model? Further, the one-way-revolution model assumes DNA revolution while the push-roll model assumes DNA rolling. Is there a sufficient gap between the connector channel and the DNA helix to facilitate either the DNA revolution or rolling during the packaging process?

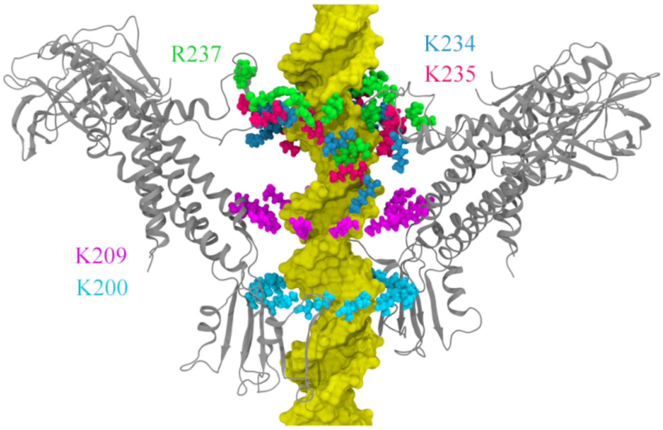


Figure 5.2: Five electropositive residues of the connector channel (*grey*) with the DNA (*yellow*). Three residues, K234 (*blue*), K235 (*red*) and R237 (*green*) are part of the loop region and two residues K200 and K209 form rings inside the channel. Only four opposite subunits of the connector are shown for the visibility.

Furthermore, in the previous chapter, the connector's elastic properties were determined in the absence of the DNA. How does the DNA influence the elastic properties of the connector?

5.2. Methods

5.2.1. Equilibrium MD simulations

The refined connector-DNA complex was used to start equilibrium MD simulations. The complex was solvated with 253 571 water molecules in a dodecahedron box, excess charge was neutralized by counter-ions, and ionic strength was adjusted to 0.150 M NaCl, which was equivalent to 815 and 740 sodium and chloride ions, respectively. The total number of atoms in the prepared molecular system was 822 519 atoms. Similarly, a starting structure of the free DNA was prepared in a dodecahedron box, and the final molecular system consisted of 834 668 atoms that include 276 386 water molecules, 874 sodium and 756 chloride ions.

Subsequently, both molecular systems were energy-minimized to remove bad contacts between the atoms. Then, systems were heated from 0 to 300 K in 100 ps constant volume simulations. In the heating simulations, all heavy atoms were restrained at the starting positions with a force constant of 1000 kJ·mol^{-1}·nm^{-2} while water molecules were allowed to move freely. In the next phase of simulations, the force constant was gradually reduced to 100 kJ·mol^{-1}·nm^{-2} during a 750 ps NPT simulation, in which a constant 300 K temperature and 1 atm pressure were maintained. The position restraints were gradually removed in subsequent 750 ps NPT simulation. In these three simulations, pressure and temperature were regulated by Berendsen pressure and temperature coupling, respectively.

Subsequently, a total of 3x370 ns (referred to as SimA, SimB and SimC in the subsequent text) and 3x130 ns of equilibrium simulations were carried out for the connector-DNA complex and free DNA, respectively with 4 fs time-step. The temperature and pressure were maintained at 300 K and 1 atm by v-rescale (102) and Parrinello-Rahman (103) algorithms, respectively. The coupling time constant for temperature and pressure were kept at 0.1 and 1 ps, respectively.

For all simulations, GROMACS 4.5.x package (98) was used, and a maximum speed of ~2.8 ns/day was obtained on 64 processors of the Intel Xeon "Harpertown". AMBER ff99SB (99) and ParmBSC0 (100) force fields were used for the connector and the DNA, respectively. The force field parameters for water molecules were taken from the TIP3P model (104). Short range electrostatic and van der Waals interactions were computed for atom pairs that fall within the cut-off distance of 14 Å. The PME method with 1.2 Å grid spacing and fourth order cubic interpolation was used to compute long range electrostatic interactions (76). All bonds were constrained using a parallel LINCS algorithm with sixth order expansion of the constraint coupling matrix (81, 82). VMD was used to visualize molecules in the obtained MD trajectories (105).

5.2.2. Force-probe simulations

To perform force-probe simulations, three conformations of the connector-DNA complex were extracted at 213.3, 308.9 and 212.9 ns time from the three trajectories SimA, SimB and SimC, respectively. Subsequently, as shown in Fig. 5.3, two virtual non-interacting particles were placed above and below to the principal channel axis outside the connector. These two particles were used as reference positions to pull or push the DNA through the channel.

For all the following force-probe simulations, three starting molecular systems were prepared from three extracted connector-DNA complexes by the similar procedure, which is above described for the equilibrium MD simulations. Subsequently, heating and equilibrations simulations were performed and further these three equilibrated molecular systems were taken for the force-probe simulations. The settings for the force-probe simulations were similar to those of the equilibrium MD simulations. To prevent the translational and rotational motions of the connector with the reference particles, C-alpha atoms of two opposite helices and reference particles were restrained at starting positions with a force-constant of 10,000 kJ·mol^{-1}·nm^{-2}. The positions of reference particles and two helices are illustrated in Fig 5.3. During the equilibrium simulations, these two helices were observed to be most rigid of the 36 helices in the middle region. Therefore, these were assumed to be rigid even during the force-probe simulations.

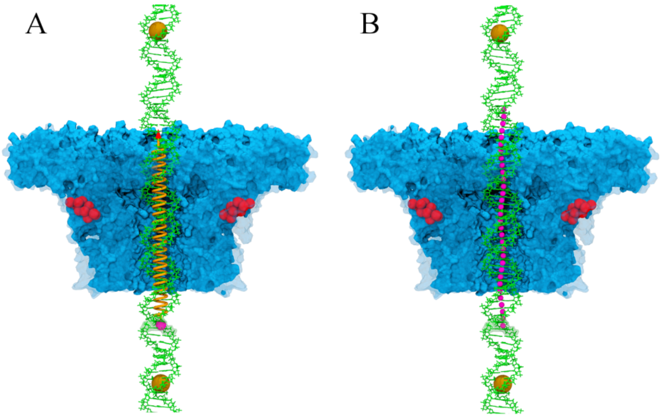


Figure 5.3: Two types of force-probe simulations that were performed to move DNA (*green*) through the connector channel (*blue*). Brown spheres depict the location of reference virtual particles. Red spheres represent C-alpha atoms, which were restrained during simulations to restrict translational and rotational motion of the connector. (A) Center of mass (*magenta sphere*) of one base-step was pulled toward upper virtual particles with a “virtual” spring (*brown*). (B) Center of masses (*magenta spheres*) of 32 base-steps were pulled simultaneously toward upper and bottom virtual particle with “virtual” springs.

Two types of force-probe simulations were performed to pull or push the DNA through the connector channel. In the first category of force-probe simulations (illustrated in Fig 5.3A), the center of mass of two base-pairs was pulled by 0.015 $m\cdot s^{-1}$ along the channel-axis with a virtual “spring” of force-constant 500 $kJ.mol^{-1}.nm^{-2}$ in a direction to the upper reference particle. Three force-probe simulations were performed for 3x170 ns by using the three different structures that were prepared as discussed in the previous paragraphs.

In the second category of force-probe simulations (Fig 5.3B), 32 virtual “springs” of force-constant 50 $kJ\cdot mol^{-1}\cdot nm^{-2}$ were attached to centers of mass of the 32 base-steps and pulled or pushed toward the reference particles by rates of 0.15 and 0.015 $m\cdot s^{-1}$. This scheme was used by assuming that the forces were uniformly applied on the DNA helix during the simulations. Additionally, similar force-probe simulations were performed for the connector-DNA complex with three mutations K234A, K235A and R237A to compare the required force to pull or push the DNA through the wild-type and mutated channel. The three mutant connector-DNA complexes were first equilibrated for ~32 ns, and subsequently force-probe simulations were carried out using the equilibrated molecular systems. A total of 2x2x3x~20ns (0.15 $m\cdot s^{-1}$) and 2x2x3x~80ns (0.015 $m\cdot s^{-1}$) simulations were performed, these involved pulling and pushing of the DNA at two different rates through three wild-type and mutant connector-DNA complexes.

5.3. Results and Discussion

5.3.1. The equilibration of the Connector-DNA complex

To test the proposed DNA packaging models, three independent equilibrium simulations, SimA, SimB and SimC were carried out on the connector-DNA complex by using identical settings. Although, the loops refinement process allowed to model the interactions at the loop region of the connector, interactions were still missing in the remaining channel regions due to the applied restraints (Fig. 5.1B). To monitor the potential change in the structure of the connector-DNA complex because of these missing interactions, its structural equilibration was studied during SimA, SimB, and SimC simulations. Further, interaction between the connector and the DNA was monitored to study the energetic equilibration during these simulations. The structural and energetic equilibration was quantified by the variation in RMSD and interaction energies, respectively along the simulation time (linear and logarithmic scale) as shown in Figs. 5.4.

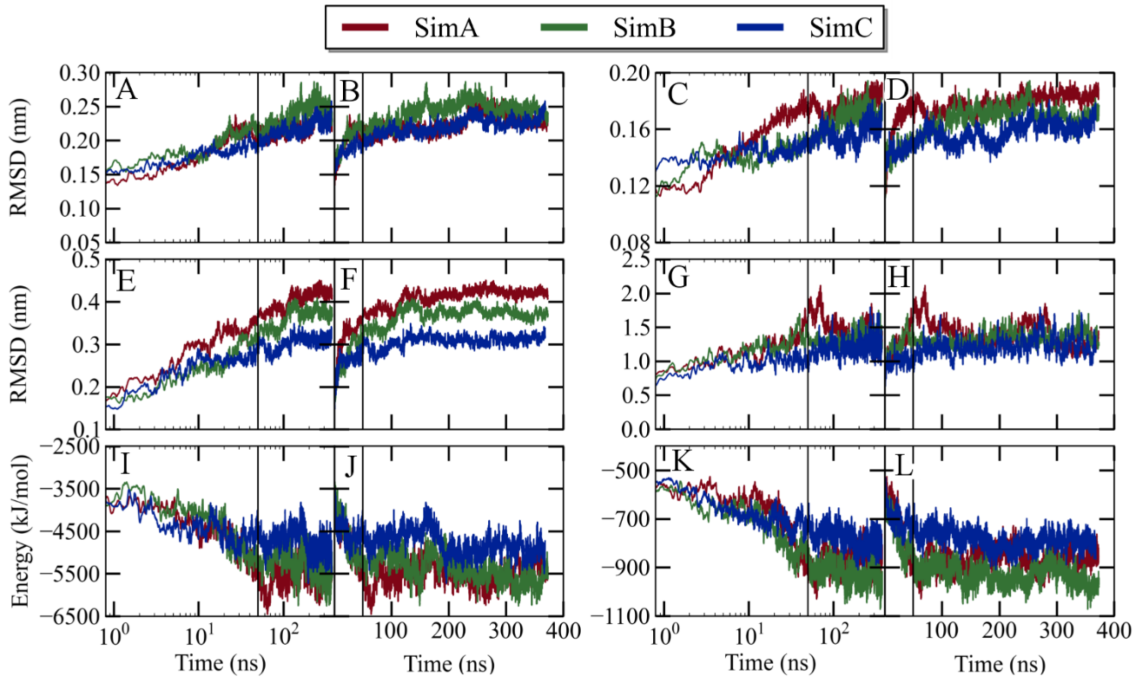


Figure 5.4: RMSDs and interaction energies during three equilibrium simulations SimA (*red*), SimB (*green*) and SimC (*blue*). Variations in RMSD with respect to logarithmic and linear time are shown for (A-B) the whole connector, (C-D) the channel, (E-F) the loop-DNA region, and (G-H) the DNA. Variations in (I-J) electrostatic and (K-L) van der Waals interaction energy between the connector and the DNA with respect to the logarithmic and the linear time are shown. Vertical line represents 50 ns time of simulations in all plots.

As seen in Figure 5.4A–B, the whole connector's RMSD rapidly increases to maximum 0.24 nm during initial 50 ns and subsequently, remains less than 0.27 nm during all three simulations. Similarly, the RMSD of the channel comprising middle, hinge and bottom regions rapidly increases to a maximum of 0.18 nm during the first 50 ns and subsequently, remains less than 0.2 nm (Fig. 5.4C–D). The RMSD of the loop-DNA region fluctuates around values of 0.3, 0.38, and 0.42 nm after a gradual increase during the initial ~120 ns of SimA, SimB and SimC simulations (Fig. 5.4E–F), respectively. The RMSD of the DNA fluctuates in a range of 1.0 to 2.0 nm after a sharp increase during the initial 10 ns (Fig. 5.4G–H). Such a large RMSD arises from the large structural changes in the DNA with respect to the starting B-form structure.

These results show that the whole connector and its channel region sharply deviate from the starting structure during the first 50 ns, and subsequently, the rate of deviations was markedly slowed down. During the next 320 ns of the simulations, the obtained maximum RMSD of 0.27 and 0.2 nm for the whole connector and the channel structure, respectively suggest a quite small structural deviation for such a large structure. The loop-DNA structures diverged (0.3 to 0.42 nm) from its earlier refined conformation during the three simulations. However, its conformation is not crucial for the packaging process because of the extreme flexibility of the loops.

Next, to monitor the variation in the interaction between the connector and the DNA, I calculated the electrostatic (Fig. 5.4I–J) and van der Waals (Fig. 5.4K–L) interaction energies between the connector and the DNA during the three simulations. After rapid decrease in energies during the initial ~50 ns, the electrostatic and van der Waals energies either slowly decreased or fluctuated in ranges of -6000 to -4500 and -1100 to -700 kJ·mol^{-1}, respectively. The obtained energy values suggest that the DNA interacts with the connector primarily through the attractive electrostatic forces.

Overall, for the channel region, which is crucial for DNA packaging, the obtained RMSDs were quite small. Additionally, the interactions between DNA and connector were strengthened during the respective simulations (Figs. 5.4J and L). The simulations is considered to be nearly equilibrated to study the DNA packaging models because of the very small structural changes and strong interactions between the connector and the DNA during the last 320 ns. Therefore, all subsequent analyses were done using the last 320 ns MD trajectories and the initial 50 ns were discarded.

5.3.2. Impact of the DNA on the connector

The impact of the DNA on the connector structure during the packaging process is not known at the molecular level and not even a single crystal structure with the DNA is available. Further, in the available cryo-EM density maps, the connector's bottom region appears to be deviating from its crystal structure after completion of the DNA packaging (3). Moreover, the one-way revolution model requires a particular arrangement of four electropositive lysine rings that are present in the channel lumen. The proposed arrangement based on the crystal structure may change in the presence of the DNA. To study the potential structural changes in the connector, I calculated two structural descriptors, twist angle and length of whole connector (as discussed in Figs. 1C and D of the previous chapter) that quantify untwisting-twisting and stretching-compression motions, respectively. Additionally, previous studies suggested that the middle region is one of the stiffest protein regions (106), and to study the DNA's impact on this region, these descriptors were calculated for this region also.

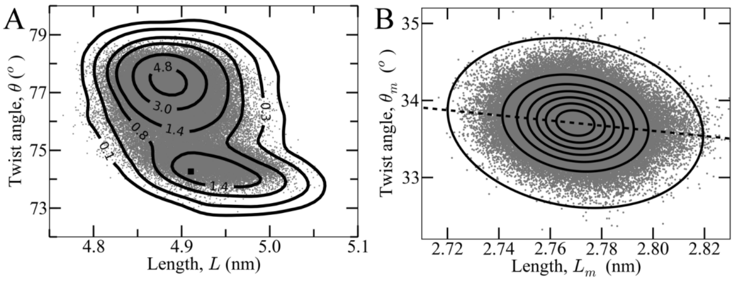


Figure 5.5: Twist angle and length fluctuation analysis. (A) Twist angle and length of the whole connector (*grey dots*). Square dot shows a value, which is obtained from X-ray crystal structure. Contour lines illustrate joint probability densities for angle and length. (B) Twist angle and length of the connector's middle region (*grey dots*). Contour lines with a spacing of 5 $\text{deg}^{-1}\text{nm}^{-1}$ show joint probability densities for angle and length. Dashed straight line shows the coupling between twist angle and length.

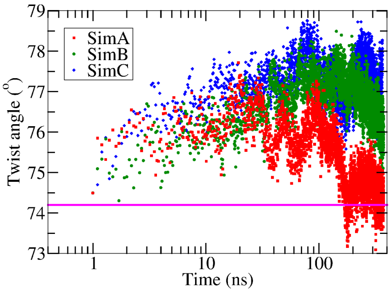


Figure 5.6: Deviations in whole connector's twist angle (*red*, *green* and *blue dots*) with respect to crystal structure value (*magenta line*) during initial 120 ns of the three simulations. Note that the time is shown in logarithmic scale.

For the whole connector, fluctuations in its twist angle were calculated as a function of its length (Fig. 5.5A). As seen in Fig. 5.5A. a bimodal distribution was obtained instead of the single-peak distribution previously obtained for the connector without the DNA (106). This bimodality is only observed for the twist angle. For further analysis, the change in twist angle with the simulation time is plotted and shown in Fig. 5.6. The twist angle changes by 2.3° from the starting crystal structure's value of 74.2° during the first 25 ns (Fig. 5.6) as previously observed during the simulations of the connector without the DNA (chapter 4). However, after 180 ns of the SimA simulation, the connector's twist angle recovered to ~74.5°, and remains within the range of 74° to 75° during course of the subsequent simulation (Fig. 5.6). The observed bimodal distribution (Fig. 5.5A) was caused by this recovery of twist angle during the SimA simulation. This result suggests that the connector untwists and recovers its starting twist angle value in the presence of the DNA.

Further as seen in Fig. 5.5A, the obtained fluctuations in the length of the whole connector are similar to the previously observed for the connector without the DNA (106). This result suggests that the DNA did not affect the length of the whole connector (Fig. 5.5A). Next, fluctuations in twist angle and length of the middle region were calculated, and the obtained distribution is presented in Fig. 5.5B. The middle region's angle and length are comparable in presence and absence of the DNA (106). Therefore, this region was unaffected by the DNA during the simulations.

The above results show that the whole connector's twist angle regains its starting value in the course of one simulation. However, the two descriptors twist angle and length are not sufficient to capture the full conformational fluctuations of the connector. Consequently, the observed change in twist angle may not correspond to full recovery of the crystal's conformation during the simulations. To probe and characterize the conformational fluctuations during simulations, PCA was performed on the middle and the bottom connector regions with reference to the two most rigid helices of the middle region (Fig. 5.1). For the middle region, projections of the first two principal components (PC) that capture the largest conformational fluctuations were calculated for the free and the bound connector. These fluctuations are compared in Fig. 5.7A.

As shown in Fig. 5.7A, the middle region conformations were similar in presence and absence of the DNA during the SimB simulation, and differ along PC-1 and PC-2 during the SimA and SimC simulations, respectively. Further, to compare these deviations, three average structures were obtained from the filtered trajectories containing conformational fluctuations along PC-1 and PC-2 (Fig. 5.7B). These average structures appear well aligned to each other upon visual inspection. The pair-wise RMSDs between the three structures were further calculated to quantify this alignment (Table 5.1). The obtained RMSD values are less than 0.8 Å implying that the observed conformational deviations are rather minuscule (Table 5.1).

Similarly, for the bottom region, projections of the first two PCs were calculated for the free and bound connectors and compared to each other (Figure 5.7C). As seen in Figure 5.7C, the bottom region's conformations were similar to each other in presence and absence of the DNA with the exception of PC-1 during the SimA simulation. To compare structural deviations, three average structures were obtained from the three trajectories that were filtered for PC-1 and PC-2 (Fig. 5.7D). These structures appeared to be aligned to each other except for the six subunits during SimA simulation. Further, pair-wise RMSDs were calculated for these structures (Table 5.1). A very small RMSD of 0.3 Å was obtained between the average structure of the SimB and SimC simulations. In contrast, the RMSD of SimA with respect to SimB and SimC are comparatively large (2.2 and 2.1 Å, respectively).

These results show that the conformation of the bottom region deviated from its crystal structure despite the above observed recovery of the twist angle during the SimA simulations (Fig. 5.6). The angle was recovered by the slight shift in the subunits' centers of mass used to calculate the angle. In summary, the conformations of the both middle and bottom regions were independent of the DNA in two out of three simulations.

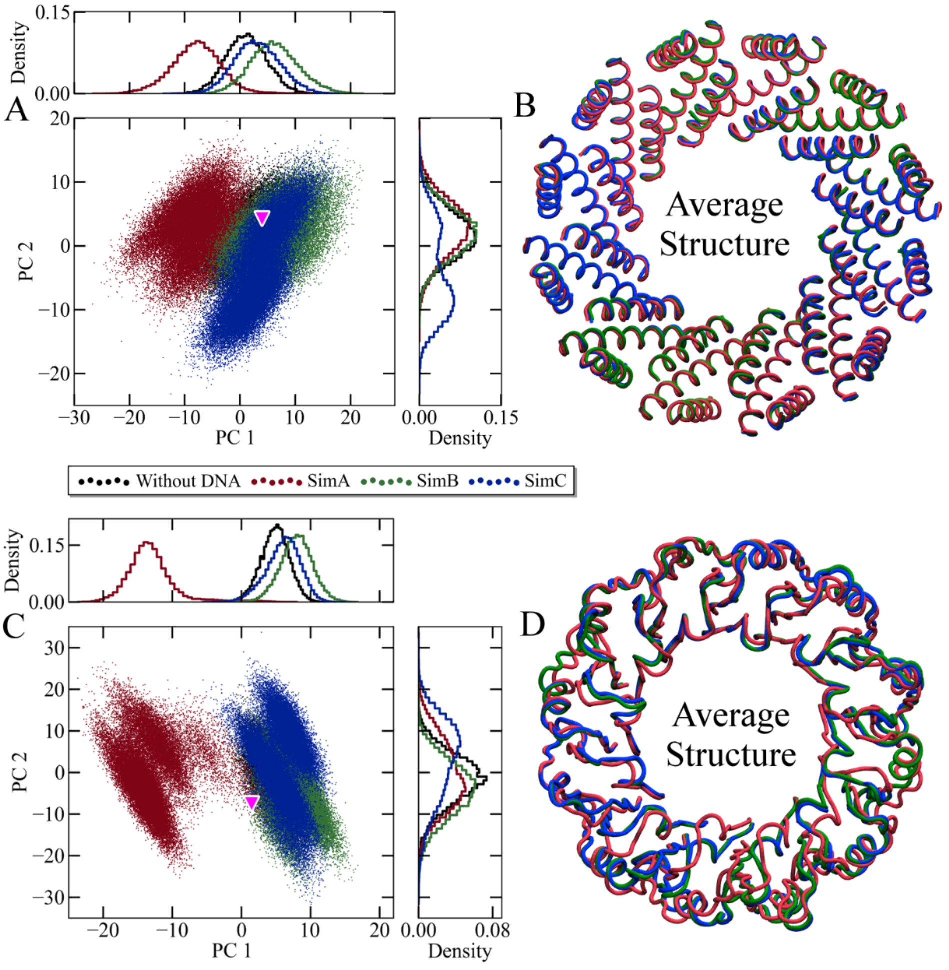


Figure 5.7: Comparison of conformational sampling during three and one equilibrium simulations of the connector with (*red*, *green* and *blue*) and without (*black*) the DNA, respectively using principal component analysis. For reference, projections for the crystal structure (*magenta triangle*) are also shown. (A) Two-dimensional projection of first and second PCs that are obtained for the connector's middle region. (B) Illustrations of the middle region's average conformations that were obtained from respective three trajectories after filtering for first and second PCs. (C) Projection plane for the connector's bottom region. (D) Illustrations of the bottom region's average conformations that were obtained from respective three trajectories after filtering for first and second PCs.

The structural deviation in the middle region is quite small in presence of the DNA (Table 5.1). In contrast, the conformations of the bottom region change from the x-ray structures in the presence of the DNA. However, note that simulations were carried out in the absence of other motor components, the pRNA and the ATPase, which may also affect conformational flexibility of the bottom region. These results corroborate the observations from cryo-EM experiments that the structure of the bottom region significantly changes after the completion of the DNA packaging process (3). Overall, the connector's middle region is hardly affected, and the bottom region's conformations may change in the presence of the DNA.

Table 5.1: RMSD in Å between the average structures of the middle and the bottom regions, which are shown in Figs. 5.7 B and D.

	SimA	SimB	SimC	
SimA		0.8	0.8	Middle region
SimB	2.2		0.4	
SimC	2.1	0.3		
	Bottom region			

5.3.3. Deformation in the DNA on its confinement in the channel

The one-way revolution and push-roll models proposed the presence of a B-form DNA as a straight double helix inside the connector channel (13, 34, 50). The DNA helix may deform because of its confinement in the channel, which contains several charged and polar residues. Moreover, the external forces generated by either the ATPase or the internal pressure may deform the DNA during the DNA packaging process. Indeed, the average atomic density maps which were obtained from the three equilibrium simulations suggest a considerable DNA deformation (Fig. 5.8). To characterize the structural deformation in the DNA in the presence and the absence of external force, I calculated deviations in base-step descriptors of the confined DNA with respect to its free-form structure. These descriptors include helical-rise, helical-twist, inclination and X-displacement that quantify helical length, helical twisting, helical bending and local helical displacement, respectively (107, 108).

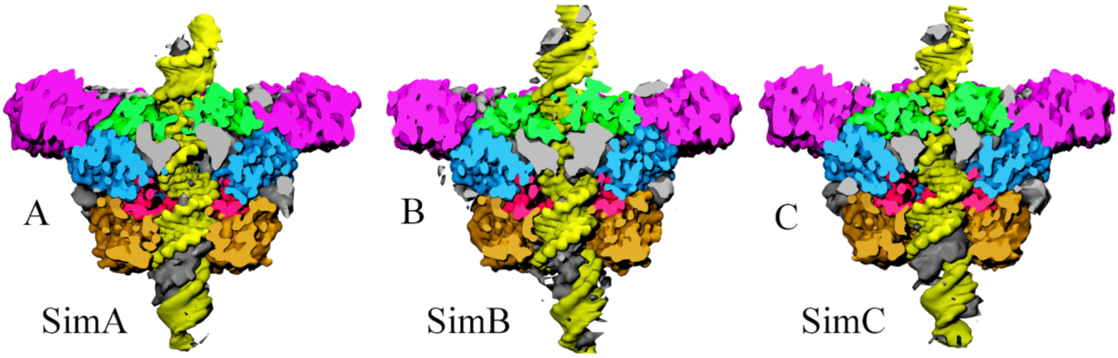


Figure 5.8: Average atomic densities of the connector (*five colors* except *yellow* and *grey*) with the DNA (*yellow*) and ions (*grey*) from three independent simulations, (A) SimA, (B) SimB and (C) SimC. The five colors of the connector correspond to the five regions as described in the Fig. 5.1B.

In the absence of an external force (equilibrium simulations), the obtained deviations in the descriptors along the channel-axis are shown in Fig. 5.9. The figure also depicts locations of different connector regions along the channel-axis such as the loop region denoted by

“LOOP” (*green*) is in a range of –40 to 0 Å. The bottom region indicated by the positions of D194-E197-K200 residues is in a range of 20 to 40 Å. As seen in Fig. 5.9, all four descriptors deviated locally with respect to those of the free DNA at the loop and the bottom regions. The helical-twist increased and decreased by ~10° at the loop and the bottom regions, respectively. Further, the local helical-rise was decreased by ~1 and ~2 Å at both the loop and the bottom regions, respectively (Fig. 5.9). For a 14 base-pairs DNA segment, the helical-length was reduced by maximum 10 % in the bottom region of the channel. Similarly, the inclination and X-displacement of the base-steps either decreased or increased along the channel-axis.

The observed deviations in the base-step descriptors suggest that the confined DNA deforms, particularly at the loop and the bottom regions of the connector. The DNA un-twists and over-twists at the bottom and the loop region, respectively. Furthermore, the decrease in the helical-rise shows that the DNA helix is compressed by a maximum of ~10 % in the channel. The change in the inclination together with the X-displacement shows that the DNA helix locally bends with respect to its free-form in the channel (Fig. 5.9). In summary, the DNA deform from its proposed B-form conformation inside the connector channel in the absence of external force.

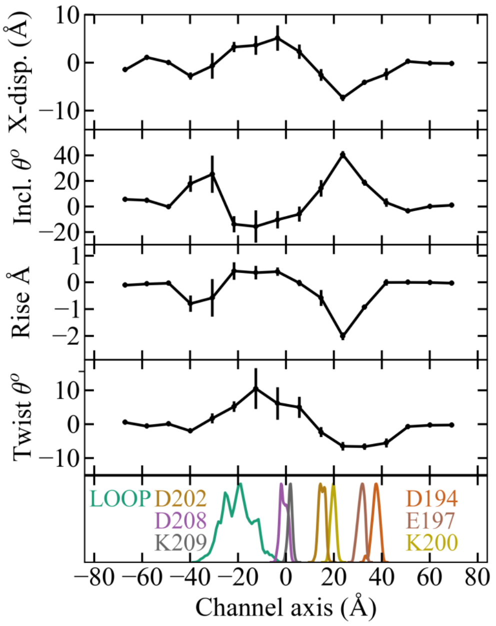


Figure 5.9: Deformation in DNA helix inside the connector channel during equilibrium simulations. The deviations in base-step parameters X-displacement, inclination, helical-rise and helical-twist with respect to that of the free DNA are illustrated along the channel. Except the lowest panel, solid-line and error-bar represent the average and standard error from three simulations, respectively. Lowermost panel depicts positions distributions of charged residues and loops that are located inside the channel. “LOOP” shows the combine distribution of K234, K235, E236 and R237. The position of D194 to K200 marks the bottom region and “LOOP” marks the loop region of the connector. D202, D208 and K209 are located in the middle region.

The above observed DNA deformation is likely to be enhanced in the presence of an external force generated during the packaging process. To characterize this potential deformation, I performed several force-probe simulations and pushed the DNA into the channel by moving center of mass of two base-pairs with a harmonic potential with a rate of 0.015 m.s^{-1} (see methods for details).

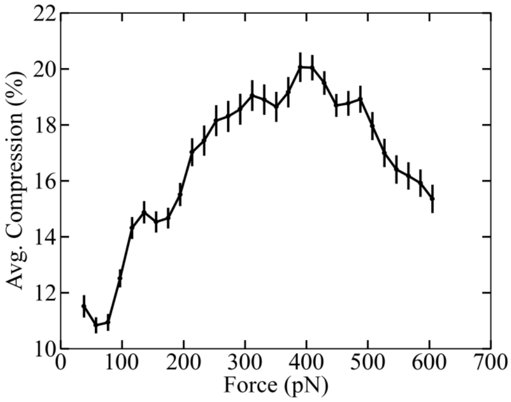


Figure 5.10: Compression of the DNA inside the connector channel under the external forces. The DNA was pulled in direction of the procapsid through channel during the force-probe simulations as explained in Fig. 5.3A. Solid-line shows maximum average compression in 14 basepairs with respect to the applied external force. Error-bar represents standard error of each block from three independent simulations.

Further, the average compression was calculated as a function of the applied force for the 14 base-pair DNA segment (Fig. 5.10). The chosen DNA segment was the most compressed segment during the equilibrium simulations. As illustrated in Fig. 5.10, the compression in DNA increased with the applied force up to a maximum of ~20 % at ~400 pN force. Subsequently, the DNA helix expanded towards its original length. The DNA was compressed during the starting phase of the simulations because the DNA segment located at the inner channel opening was held at the original position by the interactions with the loops whereas the DNA segment located at the outer opening was pushed into the channel. However, after ~20 % compression, these interactions break down and the DNA begins to relax in the subsequent phase of the simulations.

The above results suggest that the DNA compression increase by at least two-fold in presence of an external force applied to the DNA. However, note that the motor can generate a maximum force of ~110 pN (38), i.e. four times less than the obtained 400 pN force. This discrepancy between the experiments and simulations is likely due to the insufficient relaxation time during non-equilibrium simulations of the DNA compression (109). The obtained DNA compression of ~20 % for 14 base-pairs is almost similar to that of (~22 %) measured by fluorescence spectroscopy experiments performed on the bacteriophage T4 DNA packaging motor (53).

The DNA is likely to be deformed by the channel confinement in the absence of force, particular by the connector's loop and bottom regions. During the DNA packaging process, the DNA deformation is further enhanced in the presence of the forces generated by either the ATPase or the internal pressure. The DNA helix compresses similarly in both φ29 and T4 (53) DNA packaging motor channel; therefore, DNA compression appears to be a common process during the packaging process of the head-tail bacteriophage. However, the mechanistic role of DNA compression in the packaging process still remains open, and is not discussed in the current φ29 DNA packaging models.

5.3.4. How does the connector prevent the DNA leakage?

The one-way revolution model proposes that the connector acts as a one-way valve, preventing potential DNA leakage caused by a large internal pressure during the packaging process (32-34). In voltage ramping experiments, the connector was found to block the DNA transport in the reverse direction across the lipid membrane, allowing unidirectional transport of the DNA (34, 44). It is natural to ask if the connector acts as a one-way valve during the MD simulations.

5.3.4.1. Does the connector act as a one-way valve?

To address this question, several force-probe simulations were performed to move the DNA along both inward (forward) and outward (backward) directions of the procapsid with rates of 0.15 and 0.015 $m.s^{-1}$ (see details in methods). As presented in Figs. 5.11A and B, average forces required to displace a DNA base-pair forward (*red*) and backward (*blue*) from starting position were calculated at two pulling rates. When the DNA was moved by the rate of 0.15 $m.s^{-1}$, the required force to displace the DNA increased linearly for the forward motion and non-linearly for the backward motion. In the case of slower rate 0.015 $m.s^{-1}$, the required forces to displace the DNA in both directions were increasing linearly for up to a displacement of ~0.35 nm and subsequently, the force was saturated during both types of motions. These results suggest that the connector resist both the forward as well as backward transport of the DNA during the packaging process, and external forces are necessary to overcome the resistance.

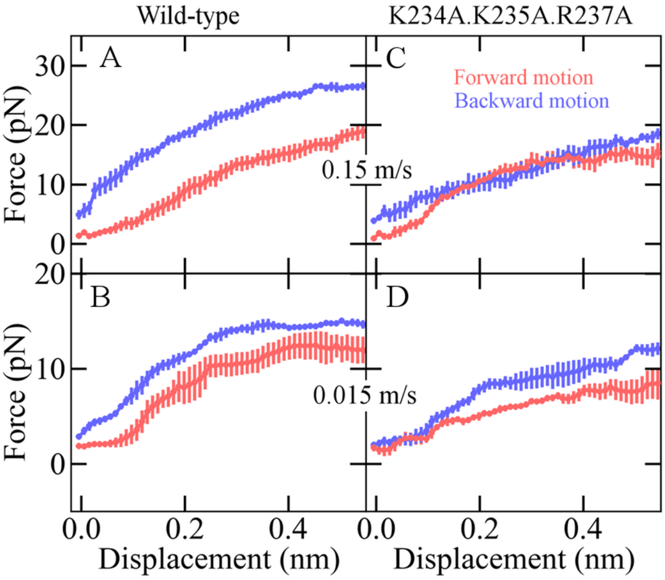


Figure 5.11: Forces applied to pull the DNA through wild-type (*left panel*) and modified (*right panel*) connector during the force-probe simulations. Solid-line shows average applied force and resulting displacement in the DNA during forward and backward motions. Error-bars show standard error of each block from three independent simulations. The forward (*red*) and backward (*blue*) motions denote directions of the DNA pulling to inward and outward of the procapsid, respectively as described in Fig. 5.3B. The modified connector contained K234A, K235A, and R237A mutations. The DNA was pulled by rate of (A) 0.15 and (B) 0.015 $m.s^{-1}$ through wild-type, and (C) 0.15 and, (D) 0.015 $m.s^{-1}$ through mutant connector during the force-probe simulations as explained in Fig. 5.3B.

For the one-way valve function, the force required to push the DNA forward into the procapsid needs to be less than the force necessary to pull it backward outside of the procapsid through the channel. Indeed, as seen in Fig. 5.11A, when DNA was moved with a rate of 0.15 m.s^{-1}, the average force to move a base-pair by 0.34 nm forward (one base-step length) was ~10 pN less than the average force to move it by same distance backward. At the slower pulling rate of 0.015 m.s^{-1} (Fig. 5.11B), the average force for the forward transport was ~5 pN less than that of the backward transport. Differences between the forces required for the backward and forward transport, were ~10 pN and ~5 pN when 10 to 25 pN and 5 to 15 pN forces were applied on the DNA, respectively. These results suggest that the force difference between forward and backward transport likely to increase with the rising external forces during the packaging process. During the late stage of the DNA packaging, the force reaches extreme of up to ~110 pN inside the procapsid (38) and, the force difference between forward and backward DNA transport is likely to be larger than the ~10 pN obtained in the force-probe simulations. The connector considerably favors the forward transport of the DNA into the procapsid whereas it disfavors DNA movement out of the procapsid, particularly at the late stage of the packaging process.

The connector allowed the DNA movement in both directions during the force-probe simulations (Figs. 5.11A and B). In contrast, it completely blocked the backward transport of DNA across the lipid membrane during voltage ramping experiments (34, 44). This discrepancy is likely to be caused by the different initial conditions; the DNA was outside of the channel during the experiments, whereas the DNA was already present inside the channel during simulations. Indeed, when the DNA was already present inside the connector channel during the packaging process, the packed DNA was found to slip out of the procapsid, and subsequently, this DNA segment is repackaged by the motor (16, 17). These observations corroborate the simulation results, and further suggest that instead of a perfect one-way valve, the connector allows a certain amount of the DNA leakage during the packaging process. Next, how does the connector minimize the DNA leakage?

5.3.4.2. How the connector's loops are essential?

In voltage ramping experiments, the connector's one-way valve function was completely abolished upon the deletion of loop residues 229-246 (34, 44). These loops from 12 subunit were proposed to prevent the DNA leakage by strongly interacting with the DNA through conserved charged residues (45). To probe these loop-DNA interactions, I calculated the average interaction enthalpy of the DNA with both an intact and connectors in which one of the loop residues E233, K234, K235, E235 or R237 were deleted (Fig. 5.3).

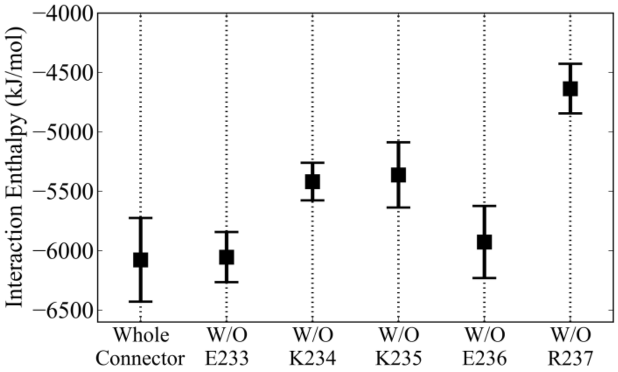


Figure 5.12: Interaction of the DNA with intact and connectors in which one of the loop residues E233, K234, K235 E236, or R237 were deleted.

As seen in Fig. 5.12, upon removal of the conserved residues K234, K235 and R237, the interaction between connector and DNA was significantly reduced (Fig. 5.12). Further, to examine the roles of K234, K235 and R237 in minimizing DNA leakage, several force-probe simulations were performed using a modified connector similar to the above discussed force-probe simulations (Figs. 5.11C and D). The forces required to displace the DNA in both directions increased as a function of the displacement during the simulations, although they were lower than those of the respective simulations of the wild-type connector. This result was observed for both pulling rates of 0.15 and 0.015 $m.s^{-1}$. Therefore, the mutated connector as the wild-type form resists both the forward and the backward transport of the DNA but with reduced strength in comparison to the wild-type.

Upon mutations of loop residues K234, K235 and R237 to Alanine, the loop-DNA interactions are expected to be reduced, and consequently, the force necessary to pull the DNA forward and backward is likely to be lower than that of the wild-type connector. Indeed, at a pulling rate of 0.15 $m.s^{-1}$ (Figs. 5.11C and A), the average force required to move one DNA base-pair backward by 0.34 nm through the mutated connector was ~10 pN less than through the wild-type form. Further, the average force required to move one DNA basepair forward by 0.34 nm was unchanged upon the mutation. With a ten times slower pulling rate, ~5 pN less force was required to move the DNA backward through mutated connector whereas ~2 pN less force is required to move it forward. These simulation results show that the loop-DNA interactions are essential to restrict the DNA movement outward of the procapsid during and after the packaging process. Furthermore, in the mutagenesis experiments, individual R237A, K234A and K235A mutations reduced the virus production rate by ~2500, ~1500 and ~100 times, respectively, and combined mutations further lowered this rate by ~4000 times (45). Additionally, upon mutating these three loop residues, the DNA leaked from the completely packed procapsid during the sedimentation assays due to generated centrifugal force (44, 45); therefore, these observations corroborate the simulation results. How does the strong interaction between highly mobile loops and the DNA minimize the DNA leakage under large counter pressure?

5.3.4.3. How do the flexible loops minimize DNA leakage?

In the previous chapter, a mechanism was proposed that the connector act similar to a check-valve and its movable component, the flexible loops, constrict the inner channel opening during forward (packaging) and widen it during backward (leakage) movement of the DNA (106). Because of the strong loop-DNA interactions, these loops move alongside the DNA during both the forward and the backward transport. During the forward transport, the loops move out of the channel towards the procapsid until the interactions break down asymmetrically because of the symmetry mismatch between the double-helix DNA and the 12-fold connector loops (Fig. 5.13B). In contrast, during backward transport, the loops move into the channel, constrict the channel opening and grasp the DNA firmly via the strong interactions at the inner channel rim (Fig. 5.13A). To probe this proposed mechanism, the average movement of three loops residues K234–R237 was computed with respect to their starting positions (Figs. 5.13C and D) in the course of the above discussed force-probe simulations (Figs. 5.11A and B).

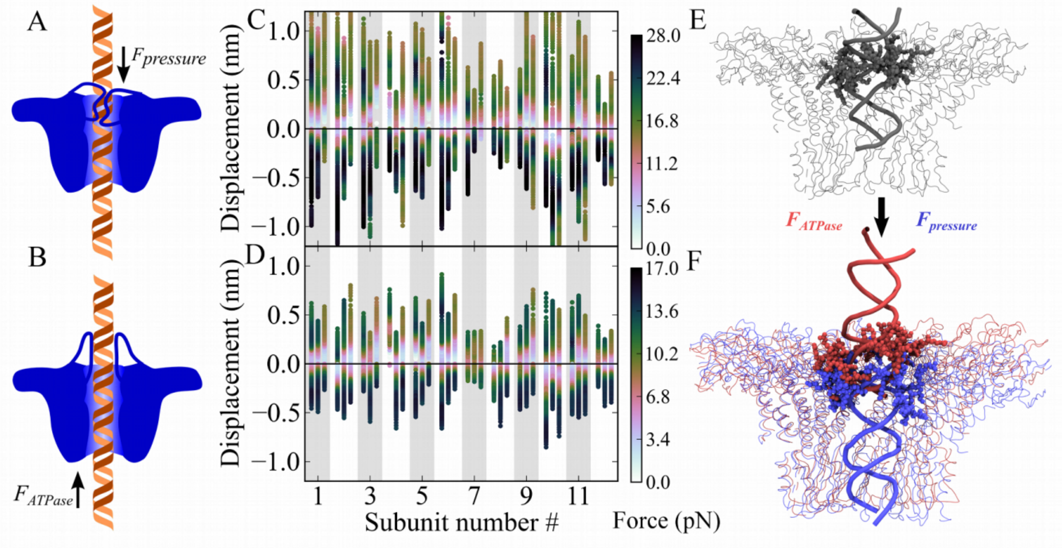


Figure 5.13: Schematic illustration of check-valve mechanism and movement of loops residues K234–R237 with the DNA. (A) Check-valve mechanism: when internal pressure ($F_{pressure}$) pushes the DNA backward, the loops constrict the inner opening of the channel by strongly interacting with the DNA. (B) During the DNA packaging, when ATPase (F_{ATPase}) pushes the DNA, the loops move outside of the channel, allowing the DNA to move inside the procapsid. Average displacements of 12 loops segment K234–R237 (*center of mass*) when the DNA was pulled by rate of (C) 0.15 m.s^{-1} and (D) 0.015 m.s^{-1} during three independent force-probe simulations (*three vertical dotted bars*). Colors represent the applied external force during the simulations. Positive and negative displacements show the forward and backward motion of the loops from starting positions, respectively. (E) Starting structure of the connector (*thin grey lines*) with the DNA (*thick grey helix*) used for one of the force-probe simulations. The loops residues K234-R237 are depicted as grey spheres. (F) A representative final positions of the loops (*spheres*) and a small segment of the DNA (*thick helix*) with respect to the connector (*thin lines*) at the end of force-probe simulations, when force was applied on the DNA inward (*red,* ~F_{ATPase}) and outward (*blue,* ~$F_{pressure}$) of procapsid as explained in Fig. 5.3B.

Figures 5.13C and D show displacement of center of mass of K234–R237 from the starting positions in 12 subunits along forward and backward DNA motions with the applied forces (color-bar) during three force-probe simulations (three bars) at pulling rates of 0.15 and 0.015 $m.s^{-1}$, respectively. Several loops were displaced by ~1 and ~0.75 nm in both the directions at 0.15 and 0.015 $m.s^{-1}$ pulling rates, respectively. Furthermore, as seen in Figs. 5.13C and D, for both pulling rates, the forces required to move the DNA backward (negative displacement) through the channel was large compared to those required to move it forward (positive displacement). Additionally, two representative snapshots of loop-DNA displacement from the starting position (grey) in forward (red) and backward (blue) direction are illustrated in Figs. 5.13E and F. As visible in these figures, the loops were moved in forward and backward direction alongside the DNA, even though the forces were applied only on the DNA. These results suggest that the residues K234–R237 are likely to move back and forth by at least 2 nm with the DNA during the transport process. The obtained differences in forces during forward and backward motions suggest that the loops minimize the DNA leakage by constricting the inner channel opening through strong interactions (Fig. 5.13A). Moreover, the loops flexibility allows widening of the inner channel opening during the DNA packaging process (Fig. 5.13B). Therefore, the observations of force-probe simulations support the check-valve functioning of the connector, which was previously proposed on the basis of the connector's elastic properties (106). The connector is likely to act as a check-valve and favors the forward over the backward DNA transport through the flexible loops, which are solidly anchored at the conserved stiff middle helical region (106).

In summary, the above discussed simulations and previous experimental results suggest that the connector favors DNA packaging whereas it minimizes DNA leakage via its conserved charged loop residues likely by a check-valve mechanism.

5.3.5. Role of the four electropositive rings in the connector channel

The one-way-revolution model proposes that transport of ~2.6 DNA base-pairs per step is facilitated by four lysine rings that are formed inside the channel lumen by respective four residues K200, K209, K234, and K235 from each of the 12 subunits (Fig. 1.7D) (34). These four rings are required to be ~9 Å apart from each other and thereby, by considering a DNA helical pitch of 3.4 Å per base-step, ~2.6 DNA base-pairs are present between two consecutive rings (Fig. 1.7D). According to the model, this particular arrangement of these four rings and their strong interactions with the DNA facilitates the DNA revolution without rotation during the packaging process. To test this proposed mechanism, at first, I calculated the average interaction enthalpy of the DNA with both an intact and connectors in which acidic and basic residues of the channel lumen were deleted (Figs. 5.12 and 5.14). In addition, the arrangement of these four electropositive rings was determined in the presence of the DNA during the equilibrium simulations (Figs. 5.15).

As seen in Figs. 5.12 and 5.14, upon deleting the residues K200, K234 and K235 from all 12 subunits, the interaction between the connector and the DNA decreases by ~500 to 1000 kJ.mol^{-1}. However, upon deletion of other channel residues, R162, D194, E197 D208, K209, E233 and E236 from all 12 subunits, the losses in interactions were negligible. As a result, three residues K200, K234 and K235 strongly interact with the DNA as required for the one-way revolution model, whereas fourth residue K209 barely interacts with the DNA. Therefore, the simulation results do not entirely support the proposed requirement of interactions between the four lysine rings and the DNA as only three rings strongly interact with the DNA.

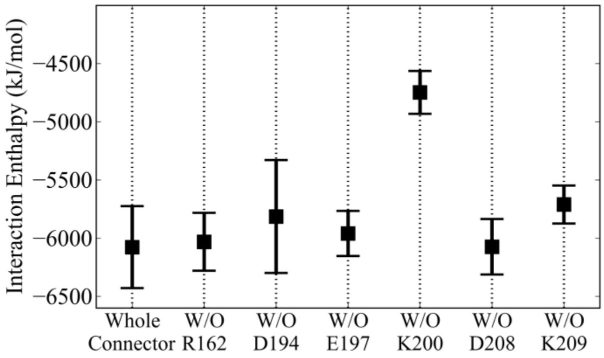


Figure 5.14: Interaction of the DNA with intact and truncated connector, in which six channel residues R162, D194, E197, K200, D208, and K209 were deleted. Error-bar shows the standard error from three independent simulations

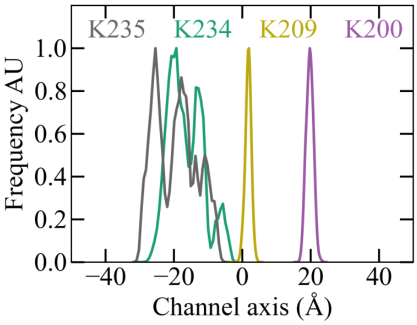


Figure 5.15: Positions of four lysine residues from combined 12 subunits along the channel during the equilibrium simulations. Distributions show the fluctuations in positions during the three simulations

Next, to probe the arrangement of the four lysine rings inside the channel, fluctuations in the positions of these rings were analyzed along the channel axis during the equilibrium MD simulations (Fig. 5.15). As seen in Fig. 5.15, the positions of residues K200 and K209 from the 12 subunits narrowly fluctuated within ~5 Å distance along the channel axis. However, the positions of the other two residues K234 and K235 from the 12 subunits overlapped with each other and broadly fluctuated within a ~30 Å distance along the channel axis (Fig. 5.15). Moreover, the distance between the K200 and K209 rings was ~20 Å (Fig. 5.15). These results suggest that residues K234 and K235 do not form rings in the channel lumen as depicted in Fig. 5.3, and remain very near to each other (Fig. 5.15), and not at the required ~9 Å distance. Furthermore, residues K200 and K209 do form rings inside the channel but these rings are separated by twice of the required distance. Additionally, considering a DNA helical pitch of 3.4 Å per base-step, the channel region between the K200 and K209 rings is expected to contain ~5.8 base-pairs. In contrast, the DNA was compressed relative to its free form (Fig. 5.10), and consequently the respective channel region contains 7±1 DNA base-pairs as compared to the ~2.6 base-pairs required by the models.

The above simulation results do not support two requirements of the one-way-revolution model: the channel electropositive rings are ~9 Å apart, and the ~2.6 DNA base-pairs are present between the two consecutive rings. Also, the expected interactions between channel residue K209 and the DNA were not observed. Moreover, mutagenesis experiments have ruled out the essential role of the two electropositive rings K200 and K209 in the packaging process because the mutations did not significantly affect the phage production in comparison to the K234A and K235A mutant (44). These observations suggest that the proposed mechanisms for the observed ~2.5 basepairs per step of DNA transport and the proposed DNA revolution are unlikely. However, these results cannot rule out DNA revolution by other unknown mechanism.

5.3.6. Presence of gap between the channel wall and the DNA helix

The one-way-revolution and the push-roll model propose the rolling and the revolution of the viral DNA during the DNA packaging, respectively (34, 50). Both, the DNA revolution and the DNA rolling require a gap between the DNA helix and the channel wall of the connector. According to the push-roll model, this gap is defined by the ratio between the radius of the channel wall (R_p) and the DNA helix (R_n) (50). To estimate the gap between the channel wall and the DNA helix, the radius of the channel as well as that of the DNA helix were calculated from the equilibrium simulations. The radius ratio ($g = R_p/R_n$) along the channel axis is illustrated in Fig. 5.16.

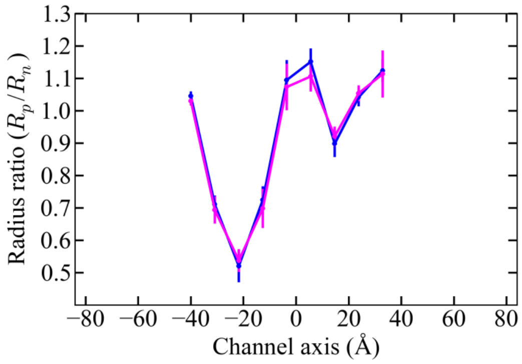


Figure 5.16: Gap ($g = R_p/R_n$) between the connector channel wall and the DNA helix. Solid lines represent average radius ratio g along the channel axis from three equilibrium simulations. The two colors (*blue* and *magenta*) depict the ratio for two separate DNA helical strands. Error-bar represents the standard error from three independent simulations.

As seen in the figure, the maximum radius ratio of 1.2 was obtained at the middle and both ends of the channel, whereas $g < 1$ is obtained at the loop and the bottom regions of the connector (Fig. 5.16). According to Yu *et al.*, with a $g \leq 1$, the DNA cannot roll inside the motor channel (50). Therefore, DNA rolling is unlikely inside the connector channel. Also, the gap ($g \leq 1$) is not enough for the DNA revolution even though the one-way-revolution model does not define the role of the gap in the revolution mechanism. These results render the both DNA revolution and the DNA rolling unlikely inside the connector channel. However note that these results neither rule out nor support either the revolution or rolling of the DNA in the other two motor components pRNA and ATPase as these DNA motions are believed to be initiated by the ATPase (15, 33, 34, 44, 50).

5.3.7. Elastic properties of the connector in presence of the DNA

The procapsid's elasticity (elastic modulus ~1.8 GPa) is similar to that of polymers such as hard plastic polypropylene (4, 5). The connector as part of the procapsid is expected to have similar elastic properties as above to withstand a maximum pressure difference ~60 atm. Moreover, these properties are required for the connector to function as a check-valve against a counter pressure during the packaging process. In chapter 5, the elastic properties were determined for the whole connector and the middle region without the DNA using the MD simulations (106). These results revealed that the middle region is one of the stiffest regions among the protein material. It provides a stiff foundation to highly flexible loops, which are essential for the check-valve function. Furthermore, the connector is expected to withstand the internal pressure through a heterogeneous distribution of elasticity in a manner similar to composite materials such as silk and sea-shell (106, 110-112). These elastic properties of the connector were determined in the absence of DNA. How does the DNA affect these elastic properties of the connector? To address this question, the elastic properties were estimated for both the connector and the middle region in the presence of the DNA along their untwisting-twisting and compression-stretching motions. These motions were quantified by two structural descriptors: twist angles and lengths, described in Figs. 1C and D of chapter 4 (106).

Table 5.3: Elastic constants of the whole connector and the middle region. These constants are calculated using three equilibrium MD trajectories SimA, SimB and SimC. Standard errors were calculated using block-averaging method with non-overlapping blocks.

Elastic properties	Whole connector	Middle region
Stretching spring constant	~2909 pN.nm^{-1}	20056 ± 3200 pN.nm^{-1}
Torsional spring constant	–	2333 ± 150 pN.nm.Deg^{-2}
Coupling constant	–	7832 ± 4000 pN. Deg^{-1}
Young's modulus elasticity	0.4±0.08 GPa	3.6 ± 0.6 GPa

Under harmonic approximation of the free energy landscape, the stretching and torsional spring constants were calculated (Table 5.3) using probability densities of length and twist angle (Fig. 5.5). The stretching spring constant of the whole connector in the presence of DNA is ~2909 pN.nm^{-1} and comparable to that of obtained for without the DNA (~3073 pN.nm^{-1}). However, both the torsional and stretching spring constants of the middle region were reduced in presence of the DNA from (3374±600) to (2333±150) pN.nm.Deg^{-2} and (24109±1500) to (20056±3200) pN.nm^{-1}, respectively (106). These results suggest that the connector in the presence of the DNA undergoes twisting-untwisting and compression-stretching motions with larger amplitude as compared to the connector without DNA under similar mechanical stress.

Next, to compare the connector's stiffness with that of other structural proteins, Young's modulus of elasticity was calculated for it in the presence of the DNA. The obtained modulus of 0.4±0.08 GPa and 3.6±0.6 GPa for the connector and middle region, respectively are similar to that without DNA (106) and thereby the connector's elasticity is unaffected by the presence of the DNA. The connector's middle region stiffness is similar to that of other structural protein such as procapsid (~1.8 GPa), collagen fibrils (0.2–11.5 GPa) and single-brin silkworm silk (5–17 GPa) whose primary function is to withstand mechanical stress (4, 5, 39-41). The connector appears to have a "stiff core" with a soft peripheral protein region, suggesting the presence of heterogeneity in the stiffness inside the connector independent of the DNA.

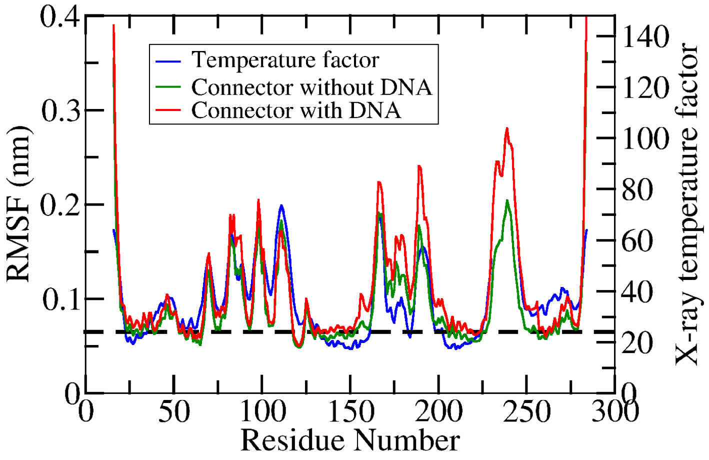


Figure 5.17: Distribution of flexible and rigid regions in the connector. Root mean square fluctuations (RMSF) of the connector with (*red*) and without DNA (*green*), their comparison with temperature factors (*blue*) obtained from crystallography (21). Residues with an RMSF below 0.065 nm (*horizontal dashed line*) are considered rigid in absence of the DNA, otherwise flexible. However, maximum RMSF of these rigid residues were about 0.08 nm in presence of the DNA.

Because the residue's stiffness is inversely proportional to the mean square fluctuation, the RMSF was used to probe the impact of the DNA on the stiffness distribution inside the connector. Residues' RMSFs were calculated and averaged for 12 subunits from three SimA, SimB, and SimC simulations (Fig. 5.17). Further, to validate the obtained flexibility with the experimental measurements, the RMSF of the connector in the presence (*red*) and the absence (*green*) of the DNA were compared to the X-ray crystallographic temperature factors (*blue*) as depicted in Fig. 5.17.

As seen in Fig. 5.17, the obtained RMSF values for connector in presence of the DNA are in agreement with the temperature factors with the correlation of $r = 0.78$, which is less compared to the correlation ($r = 0.83$) for the connector without the DNA (106). This reduced correlation is caused by additional flexibility of the bottom region as shown for residues 165–190 in Fig. 5.17. The connector's RMSF in presence of the DNA is comparable to that of the without DNA except for the bottom and loop regions. As discussed in above PCA (Figs. 5.7C and D), the bottom region undergoes larger conformational fluctuations in the presence of the DNA that lead to the larger RMSF values. Moreover, the loop's RMSF increased due to the heterogeneous interactions of the loops residues with the DNA in three independent simulations. The obtained correlations between the RMSFs and temperature factors suggest that the flexibility of the residues was accurately captured during the simulations.

In the absence of the DNA, the connector's residues with the RMSF ≤ 0.065 nm have been considered to be significantly stiff as compared to the remaining residues as discussed in the previous chapter (106). However, in the presence of the DNA, the RMSF of these stiff residues increased to the maximum value of 0.08 nm (Fig. 5.17). This minor increase is assumed to be negligible and, the distribution of the stiff residues in the connector is unaffected by the presence of the DNA. These stiff and flexible regions are arranged in an alternate fashion around the channel rim as depicted in Fig. 3 of the previous chapter (106). This particular arrangement is independent of the DNA and shows the similarity of the connector to the composite materials such as the crystalline silk and sea shells. Therefore, the previously proposed mechanism by which the connector withstands a large counter pressure (106), remains viable even with the DNA present in the channel. According to this mechanism, the stiff regions that are exposed to the internal pressure inside the procapsid interior distribute the pressure force below the stiff middle region. Moreover, the flexible regions that are present at the interface of connector and procapsid absorb and uniformly distribute the local force peaks on the stiff middle region. Thereby, the connector channel avoids large-scale deformation and withstands the mechanical stress generated by both the packed DNA and the procapsid.

The elastic properties of the connector and middle region are hardly affected by the presence of the DNA, and the middle region is observed to be as stiff as other mechanical stress handling proteins. The α-helical scaffold of this middle region is conserved in other head-tail connectors of bacteriophage such as T7, SPP1 and P22 despite residues sequence dissimilarities (113-119). The α-helical middle region appears to be an essential structural feature, which is required to withstand large mechanical stress and to function as a check-valve during the packaging process. Moreover, the connector's similarity to composite materials is expected to prevent the collapse of the channel that may be induced by both longitudinal and lateral pressure generated by both the packed DNA and the procapsid.

5.4. Summary

In this chapter, I studied the one-way revolution and push-roll models by performing explicit solvent MD simulations of the connector-DNA complex. Both models were proposed on the basis of known characteristic features of the connector's crystal structure and of B-form DNA. The structure of the DNA with the connector is not available, and the impact of the DNA on the connector and *vice versa* is unknown; potential deviations from the crystal structure may affect the proposed mechanisms in these models. Specifically, the DNA revolution proposed in one-way revolution model and the DNA rotation–rolling proposed in push-roll model (34, 50), are sensitive to the structural feature of the connector channel and the DNA. To probe the potential conformational changes in the connector in presence of the DNA, PCA was performed using the equilibrium MD trajectories. The obtained results revealed that the conformational changes in the middle region are quite small and suggest that this region is unaffected by the presence of the DNA. In contrast, the bottom region's structure was slightly deformed and suggests that this region is sensitive to the presence of the DNA.

The impact of the connector on the DNA was characterized in presence and absence of the external forces by computing the deviations in base-step parameters (helical-twist, helical-rise, X-displacement and inclination) with respect to free-form DNA. In the absence of the force, the DNA over-twists and un-twists at the loop and bottom regions of the channel and it compresses in the entire channel by a maximum of 10 % at the bottom region. Also, the DNA helix was locally bended at the loops and bottom region of the channel. Moreover, when the DNA was pushed into the channel by applying an external force during the force-probe simulations, the DNA was compressed by a maximum of ~20 %. A similar DNA compression (~22 %) was observed during the DNA packaging process of the T4 bacteriophage in previous fluorescent experiments (53). Therefore, this compression appears to be a common step during the packaging process in the head-tail bacteriophage. The observed deformation in the DNA with respect to the required B-form DNA is likely to affect the DNA packaging models.

Next, the connector's function as a one-way valve, which is required by one-way revolution model, is probed by the several force-probe simulations. The DNA was moved through the connector channel by applying external forces inward and outward of the procapsid during these simulations. The forces required to move the DNA inward were smaller compared to that those required to move it outward. Therefore, the connector resists DNA transport in both directions with unequal strength. It favors inward DNA transport over the outward transport, and this observation is compatible with the one-way valve mechanism. Additionally, instead of a complete DNA blockage for outward transport, the connector only partially blocks this transport. This observation is compatible with previous observations of DNA slippage during the packaging process in optical tweezers experiments (16, 17).

The connector was previously proposed to prevent the DNA leakage by strongly interacting with the DNA via charged loops residues K234, K235 and R237 (34, 44, 45). Indeed, analysis of the interaction between the connector and the DNA during the simulations revealed that these three loop residues strongly interact with the DNA. Moreover, when force-probe simulations were performed on the modified connector-DNA complex with K234A, K235A and R237A mutations; the forces required to move the DNA inward and outward were similar to each other and less than those required in the wild-type simulations. Therefore, these three mutations considerably reduce the connector's resistance to DNA transport in both directions during the packaging process and suggest their essentiality for the DNA packaging process. Indeed, during previous sedimentation assays, the bacteriophage production rate was reduced by ~4000 times upon mutation of these three residues (45). These loops are highly flexible, in this case, how do the loops minimize the DNA leakage against a large internal pressure?

Here, new mechanism, by which the loops deter the DNA leakage, is proposed based on the observations from the equilibrium and force-probe MD simulations. According to this mechanism, the connector acts as a check-valve, and its flexible loops similar by to movable valve parts widen and constrict the inner channel opening during inward and outward DNA transport, respectively. Particularly, the interactions between loops and DNA are strongest when the DNA begins to move outward, and thereby the inner channel opening constricted and this minimizes DNA leakage. Moreover, these loops are anchored to the stiffest middle region, and the potential severe deformation, which may be caused by a large pressure, is likely to be prevented during the DNA leakage. Overall, the connector acts as a check-valve under a large counter pressure because of its heterogeneous stiffness and its strong interaction with the DNA.

The one-way revolution model (34) further proposes that a particular arrangement of the four lysine residues K200, K209, K234, and K235 that form four rings at distances of ~9 Å inside the channel is required to facilitate the DNA transport by ~2.5 base-pairs per step (15, 42). In contrast, during the MD simulations, K200 and K209 rings are separated by twice the required distance, and two other residues K234 and K235 do not form rings and are broadly located in ~25 Å length of the channel axis. Furthermore, more than two times of the DNA base-pairs required by the model are present between the K200 and K209 rings inside the channel. Therefore, these four residues are unlikely to facilitate the observed DNA transport by ~2.5 base-pairs per step and results do not support the proposed mechanism for this particular step-size. The observed step-size is most likely to be driven by the ATPase as proposed in the push-roll model.

To drive the DNA revolution or rolling which is proposed in the one-way revolution model and push-roll models, respectively a gap should be present between the DNA helix and the channel wall. In contrast, during the equilibrium simulations, the required gap is not available due to the observed DNA deformation, and these motions are unlikely in the connector channel. However, these results cannot rule out either the DNA revolution or the DNA rolling in other two motor components pRNA and ATPase.

The connector acts as a portal for the DNA transport against a large counter pressure, in this case, how does it withstand such a large mechanical stress? The elastic properties of the connector in absence of the DNA are discussed in the previous chapter. In the presence of the DNA, the spring constants were reduced by 5 to 25 % as compared to that of the without DNA, and despite this reduction, available external forces are not sufficient to deform the connector outside of its elastic regime. Furthermore, Young's modulus of the whole connector and the middle region are similar to their respective values that were obtained in the absence of the DNA. Therefore, presence of the DNA does not appear to influence the connector's elasticity. The obtained exceptional stiffness of the middle region suggests its functional role in the DNA packaging because its α-helical scaffold is conserved in the head-tail connectors of the other bacteriophage. Further, residues' fluctuation analysis shows that the distributions of the stiff residues were similar inside the connector in both the presence and the absence of the DNA. The observed arrangement of the stiff and flexible residues inside the connector is appeared to be similar to that of the composite materials. The stiff region which faces the procapsid interior distributes the pressure to underneath stiff middle region. The flexible regions which are present at the connector and procapsid interface act as a "shock absorber" and distributes local force peak uniformly on the stiff middle region. Thereby, the connector likely to withstand the counter pressure through the heterogeneity in stiffness and the channel is prevented from the potential collapse under the large mechanical stress.

Summary, Conclusion and Outlook

Chapter 6

6.1. Summary

Bacteriophage infect bacteria and replicate by exploiting molecular resources of the host cell. The DNA packaging process is an essential step for the assembly of a mature bacteriophage inside the host cell (1). A motor packages the viral DNA into the viral procapsid against a maximum pressure difference of ~60 atm generated by the already packed DNA (17). This motor is one of the strongest biological motors known and therefore, the idea of using it as nano-devices is emerging in the field of nanotechnology (7, 18). The φ29 DNA packaging motor is one of the most studied biological motors and several experimental approaches such as fluorescence spectroscopy, optical-tweezers or membrane based voltage ramping were used to understand its mechanism of action (6, 7, 16, 34-36). This motor consists of three components, the connector channel, the pRNA, and the ATPase that assemble at the vertex of icosahedron procapsid (7). The connector acts as a portal for the DNA transport, the ATPase hydrolyzes ATP molecules to release required energy, and the pRNA bridges the connector and the ATPase (6, 7).

The aim of my study was to understand the role of the connector in the packaging process in atomistic detail. To this aim, I examined three DNA packaging models, the untwist-twist, the one-way revolution, and the push-roll DNA packaging model (13, 34, 50). The untwist-twist model proposes that the connector untwists and stretches simultaneously by 12° and 0.64 nm, respectively at the expense of one ATP molecule (13). Subsequently, the connector relaxes, regains the original orientation, and pushes two basepairs of the viral DNA into the procapsid. These untwisting-stretching and relaxation motions are proposed to be driven by the connector's elastic properties with a reversible deformation. Additionally, the elastic properties are also essential for the connector to withstand a large counter pressure. These properties were studied using equilibrium and force-probe simulations of the connector in the presence and the absence of DNA.

Equilibrium fluctuation analysis of the connector with and without DNA revealed an exceptionally stiff middle helical region with a Young's modulus of elasticity (~3.6 GPa), which is similar to that of other mechanical stress handling proteins such as viral procapsids, silk, and collagen fibers (4, 5, 39-41). Conservation of the middle region's α-helical scaffold in connectors of other head-tail bacteriophage such as T7, SPP1 and P22 (113-119) suggests that its stiffness is essential for the function. The middle region can resist the large irreversible deformation and indeed, its structure appears to be similar in packed and empty procapsid (3). Additionally, the whole connector (~0.4 GPa) is softer than its middle region (~3.6 GPa), and this difference in stiffness can be caused by the heterogeneous distribution of stiff and flexible residues inside the connector.

Individual residue fluctuations from equilibrium simulations were used to compute the spatial distribution of stiff residues, and the obtained fluctuations were in good agreement with X-ray temperature factors (21). Therefore, equilibrium simulations appear to accurately capture the flexibility of the connector. Further, stiff and flexible regions were discriminated by considering a threshold maximum fluctuation for the stiff residues. The stiff region is composed of middle and partial upper regions of the connector, whereas the entire bottom region is flexible. The stiff and flexible regions are arranged in an alternating pattern similar to the composite materials such as sea shells and silk fibers (110-112). This stiff region is also exposed to the procapsid interior, and it distributes internal pressure on underlying stiff middle α-helical scaffold. At the interface of the connector and the procapsid, the flexible layers act as the "shock absorber", which uniformly distributes the local force peaks on the stiff middle region. Thereby, the connector can withstand the counter pressure through the heterogeneous distribution of the stiffness that prevents the collapse of both empty and filled channel under mechanical stress.

Equilibrium fluctuation analyses were not sufficient to determine the connector's elastic regime with its reversible deformations that are required by the untwist-twist DNA packaging model. To probe the elastic limit, force probe simulations were performed, and the connector without the DNA was deformed by applying external force/torque. These simulations revealed that the connector possesses an elastic regime of ~6 Å extension and ~6° twist. Upon subsequent removal of the force, the connector recovered the equilibrium configuration within ~120 ns time. Therefore, the connector exhibits the property of reversible elastic deformation as required by the untwist-twist model. This model further proposes that ~50 kJ mol^{-1} of energy released from the hydrolysis of one ATP molecule, drives the untwisting-stretching motions of the connector. To test this model, umbrella sampling simulations of the connector without the DNA were performed to compute the energy required for the proposed motion. Energy of ~200 kJ mol^{-1} was required to drive half of the proposed motion, and therefore, is much larger than the proposed energy. Further, the impact of DNA on the connector's elastic properties was quite small and consequently, energy required to deform the connector is likely to be comparable in the presence and the absence of the DNA. Moreover, voltage ramping experiments revealed that the connector facilitates the DNA transport along electrochemical gradient without the ATP hydrolysis (25, 27, 32, 33, 44, 46) and the energy generated by this gradient is not sufficient to drive the proposed untwist-stretch motion. Therefore, the connector most likely transports the DNA without undergoing the proposed spring-like motion, and the untwist-twist model appears to be implausible.

The above results lead to the question of what structural and energetic features contribute to the connector's stiffness. Analysis of the umbrella sampling trajectories revealed that the hydrophobic core residues are exposed to the solvent during the connector deformations. Moreover, interactions present at the inter-subunit interface are disrupted during the deformations. Therefore, the presence of the hydrophobic core and the strong inter-subunit interactions contribute to the stiffness of the middle region.

The next two packaging models, one-way valve revolution and push-roll models require the presence of the B-form DNA in the connector channel (34, 50). However, crystal structures of the connector with the DNA are not available, and impact of the DNA on the connector and *vice versa* at atomic resolution is not yet known. Motions of the DNA proposed in both models could be influenced by the potential deformation of the connector channel or the DNA. To quantify the structural changes in the connector in the presence of the DNA, principal component analysis was performed on equilibrium MD trajectories. This analysis revealed that the connector's middle region, which is also the stiffest region, hardly deforms in the presence of the DNA. The 12-fold symmetry of the bottom region slightly deforms without the disruption of the subunit's secondary structure in the presence of the DNA. Note that the deformation was observed in the absence of other motor components, namely the pRNA and the ATPase, and their interactions may also influence the bottom region's structure. However, these observations are not sufficient to relate the bottom region's deformation with the proposed DNA motions and thus require further investigations.

The deformation of the DNA inside the connector channel with respect to free DNA was determined by quantifying the deviations in base-step parameters such as helical rise and helical twist. The DNA locally deforms by bending, un(over)-twisting, and compressing at the loop and the bottom region of the connector channel. These deformations may be aggravated in the presence of external forces that are generated by either the ATPase or the packed DNA. Indeed, when DNA was pushed through the connector by applying external force, a maximum compression of ~20 % was observed for the 14 base-pairs DNA segment. A similar DNA compression was previously observed inside the DNA packaging motor of T4 bacteriophage during fluorescent spectroscopy experiments (53). Therefore, this DNA compression appears to be a common stage of the packaging process in head-tail bacteriophage, yet its functional role is not proposed in the one-way revolution and push-roll packaging models. The DNA confined by the motor channel, considerably deviates with respect to its B-form conformation, which is proposed in both one-way revolution and push-roll models.

The one-way revolution model further proposes that the connector acts as a one-way valve and prevents DNA leakage during the packaging process (34). To probe this function, force-probe simulations were performed to determine the required force for displacing the DNA through the connector channel. The forces required to displace the DNA in a direction of the procapsid were smaller by ~10 pN and ~5 pN compared to that of the opposite direction when forces in the range of 10–25 pN and 5–15 pN were applied on the DNA via harmonic potentials, respectively. Moreover, the DNA was effectively displaced in both directions during the simulations in contrast to the previously observed unidirectional DNA transport during voltage ramping experiments (34, 44). These two observations are different because of the starting location of the DNA, which was already present inside the channel during the simulations whereas it was outside the channel during the experiments. These observations suggest that the DNA can move out of the procapsid under the influence of internal pressure during the packaging process. Previous optical-tweezers experiments revealed that the packed DNA sporadically slips out of the procapsid during the packaging process (16, 17). My simulation results suggest that the connector favors inward DNA transport over the outward transport, and thus the DNA leakage is minimized during the packaging process. This leads to the next interesting question of how the connector reduces DNA leakage caused by the large pressure from the packed DNA during the packaging process?

The connector was proposed to act as a one-way valve by strongly interacting with the negatively charged DNA via its loops that contain positively charged residues K234, K235, and R237 (34, 44, 45). Interaction analysis of the equilibrium MD trajectories showed that the three residues, K234, K235 and R237 strongly interact with the DNA and the interactions between the connector and the DNA were reduced by ~45 % upon deletion of these residues. Moreover, to probe the role of these residues, I modified the connector by mutating these residues to alanine, and performed force-probe simulations to determine the force required to displace the DNA through the mutated connector. These simulations revealed that the forces required to displace the DNA were similar for both the directions and less than that of the respective wild-type connector. This result suggests that the mutated connector offers considerably smaller resistance to the DNA displacement as compared to the wild-type connector. Additionally, previous mutagenesis experiments combined with sedimentation assays revealed that the virus production reduces by ~4000 fold, and the packed DNA leaks from the filled procapsid upon centrifugation when these three residues are mutated in the connector (45). Therefore, the three residues K234, K235 and R237 are essential for the DNA packaging and hinder the DNA leakage by interacting with the DNA during the packaging process. But when the loops containing K234, K235 and R237 residues are highly flexible, then how do the loops minimize the DNA leakage against a large counter pressure?

To explain the mechanism by which the loops minimize the DNA leakage, I propose a new check-valve model based on the equilibrium and the force-probe simulations. According to this model, the connector acts as a check-valve and its loops being a movable part, widen or constrict the channel's inner opening during the packaging and leaking, respectively. During the DNA packaging, both the loops and the DNA move into the procapsid because of their interactions with each other and subsequently the inner channel opening widens due to the symmetry mismatch between flexible 12-folds loops and DNA-helix. Whereas, during the DNA leakage, the loops move with the DNA out of the procapsid, and eventually these loops obstruct the channel's inner opening by interacting with the DNA because of the limited space present inside the channel. Additionally, the loops are solidly anchored on the stiff middle region, which resists potential large deformation caused by the large internal pressure during the packaging process. Thereby, the connector is likely to act as a check-valve to minimize the DNA leakage even under large counter pressure.

The φ29 motor transports ~2.5 DNA base-pairs per step during the packaging process (15, 42). However, the packaging models propose different mechanisms for this particular packging step-size with the consequential DNA motions. The push-roll model proposes that the DNA is pushed and negatively rotated by the ATPase, and simultaneously the DNA positively rolls on the channel wall due to the presence of a gap between the channel and DNA-helix (50). In contrast, according to the one-way revolution model, the DNA revolves inside the channel without the rotation at a packaging rate of ~2.5 DNA base-pairs per step (34). This particular step-size is facilitated by the interaction between the viral DNA and the connector's four electropositive Lysine rings that are formed by residues K200, K209, K234, and K235 at the inner wall of the dodecameric channel. These rings should be separated by ~9 Å distance to accommodate ~2.6 base-pairs of a B-form DNA helix with 3.4 Å helical pitch.

Next, to probe the role of the connector in the observed packaging step-size, I determined the spatial distribution of these electropositive residues along the connector channel axis from the equilibrium simulations. Residues K200 and K209 were ~20 Å apart and 7±1 DNA base-pairs were found to be present between these two rings. Moreover, instead of being rings, K234 and K235 were broadly distributed, and their positions overlapped with each other, causing interactions with the same DNA base-pairs. These results suggest that the spatial distribution of the electropositive rings in the channel is incompatible with the required distribution, and consequently the channel's regions between two successive rings accommodate either more or less than ~2.5 DNA base-pairs. Therefore, the simulation results do not support the mechanism proposed for both the observed packaging step-size as well as the DNA revolution without rotation. However, the DNA may revolve without rotation with an unknown mechanism because my simulation results are not sufficient to completely rule out this DNA motions. The observed DNA packaging step-size is most likely to be driven by the ATPase, and is independent of both the connector and the DNA which is present inside the connector channel.

Both the DNA revolution and the DNA rolling motions proposed during the packaging process, require the presence of a gap between the DNA helix and the channel wall. The gap was quantified as the ratio of the radii between the channel-lumen and the DNA helix, respectively. During equilibrium simulations, this ratio was observed to be either equal to or less than one. This result suggests that the gap between the channel and the DNA is not sufficient to facilitate the proposed DNA revolution or DNA rolling. Note that these DNA motions were proposed to be driven by the ATPase, and, the simulation results neither support nor rule out the DNA motions inside pRNA and ATPase.

6.2. Conclusion and Outlook

During replication of the φ29 bacteriophage, a DNA packaging motor packages the viral DNA into the procapsid against a maximum pressure difference of ~60 atm, which is generated by the already packed DNA. Several models have been proposed to explain the DNA packaging mechanism and some of these models were ruled out by the experimental studies (7, 47, 48). Here, three remaining DNA packaging models, untwist-twist (20), one-way revolution (34), and push-roll model (50) were studied to elucidate the role of the connector in the packaging process using MD simulations.

In contrast to the untwist-twist model, current simulations and previous experimental studies (46) suggest that the connector does not actively push the viral DNA during the packaging process. Further, the available energy from ATP hydrolysis is too little to carry out the large-scale spring-like motions in the connector as proposed by this model. The required energy for these motions is very large because the connector's middle region is as stiff as other mechanical stress handling proteins such as silk and collagen fibers. Moreover, this stiffness appears to be essential for the connector to withstand the large internal pressure. The similar stiff structural scaffold of the middle region is conserved in the connectors of other bacteriophages such as T7, SPP1 and P22 (114).

My simulations results also suggest that instead of the connector, the ATPase actively pushes the viral DNA into the procapsid transport as proposed in remaining two models; i.e. one-way revolution and push-roll models. The connector's essential role is then to minimize the DNA leakage by acting as a check-valve as proposed in the one-way revolution model. Moreover, the potential irreversible deformations in the connector caused by the large internal pressure, is prevented by the heterogeneous stiffness distribution.

During the packaging process, the viral DNA is packaged at a rate of ~2.5 base-pairs per step (15, 42). As opposed to the one-way revolution model, this particular packaging rate appears to be independent of the connector's structure on the basis of the current study. The packaging rate might be determined by the ATPase as proposed in the push-roll model.

Recently, the motions of the DNA during the packaging process are controversially discussed in the one-way revolution and the push-roll models (34, 50). My study suggests that the DNA revolution and rolling are implausible inside the connector because the gap between the DNA helix and the channel is not enough to carry out these motions. However, the occurrence of these motions in the pRNA and the ATPase has to be investigated. Furthermore, the current study neither supports nor opposes the proposed DNA rotation because the ATPase was not present in the MD simulations. Additionally, DNA is compressed by an external force during the packaging process as previously observed for the T4 bacteriophage (53). The mechanistic role of DNA compression in the packaging process remains to be understood.

This study was focused on the connector's roles in the packaging process. Functional roles of the two motor components, the pRNA and the ATPase remain to be studied at atomic detail by assuming the crystal structure will be available. Understanding the coordination of the connector with the pRNA and the ATPase to transport the DNA is necessary to understand the mechanism of the DNA packaging process.

References

Chapter 7

1. Meijer, W. J., J. A. Horcajadas, and M. Salas. 2001. Phi29 family of phages. Microbiol. Mol. Biol. Rev. 65:261-287.
2. Tao, Y., N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann, and T. S. Baker. 1998. Assembly of a tailed bacterial virus and its genome release studied in three dimensions. Cell 95:431-437.
3. Xiang, Y., M. C. Morais, A. J. Battisti, S. Grimes, P. J. Jardine, D. L. Anderson, and M. G. Rossmann. 2006. Structural changes of bacteriophage phi29 upon DNA packaging and release. EMBO J 25:5229-5239.
4. Ivanovska, I. L., P. J. de Pablo, B. Ibarra, G. Sgalari, F. C. MacKintosh, J. L. Carrascosa, C. F. Schmidt, and G. J. Wuite. 2004. Bacteriophage capsids: tough nanoshells with complex elastic properties. Proc. Natl. Acad. Sci. U S A 101:7600-7605.
5. Carrasco, C., A. Luque, M. Hernando-Perez, R. Miranda, J. L. Carrascosa, P. A. Serena, M. de Ridder, A. Raman, J. Gomez-Herrero, I. A. Schaap, D. Reguera, and P. J. de Pablo. 2011. Built-in mechanical stress in viral shells. Biophys. J. 100:1100-1108.
6. Ibarra, B., J. R. Caston, O. Llorca, M. Valle, J. M. Valpuesta, and J. L. Carrascosa. 2000. Topology of the components of the DNA packaging machinery in the phage phi29 prohead. J Mol Biol 298:807-815.
7. Guo, P. 2005. Bacterial virus phi29 DNA-packaging motor and its potential applications in gene therapy and nanotechnology. Methods. Mol. Biol. 300:285-324.
8. Guo, P. X., S. Erickson, and D. Anderson. 1987. A small viral RNA is required for in vitro packaging of bacteriophage phi 29 DNA. Science 236:690-694.
9. Guo, P., C. Peterson, and D. Anderson. 1987. Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage phi 29. J Mol Biol 197:229-236.
10. Guo, P., C. Zhang, C. Chen, K. Garver, and M. Trottier. 1998. Inter-RNA interaction of phage phi29 pRNA to form a hexameric complex for viral DNA transportation. Mol Cell 2:149-155.
11. Guo, P. X., S. Bailey, J. W. Bodley, and D. Anderson. 1987. Characterization of the small RNA of the bacteriophage phi 29 DNA packaging machine. Nucleic. Acids. Res. 15:7081-7090.
12. Lee, T. J., and P. Guo. 2006. Interaction of gp16 with pRNA and DNA for genome packaging by the motor of bacterial virus phi29. J. Mol. Biol. 356:589-599.
13. Simpson, A. A., Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2000. Structure of the bacteriophage phi29 DNA packaging motor. Nature 408:745-750.

14. Valpuesta, J. M., and J. L. Carrascosa. 1994. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. Q. Rev. Biophys. 27:107-155.

15. Moffitt, J. R., Y. R. Chemla, K. Aathavan, S. Grimes, P. J. Jardine, D. L. Anderson, and C. Bustamante. 2009. Intersubunit coordination in a homomeric ring ATPase. Nature 457:446-450.

16. Chemla, Y. R., K. Aathavan, J. Michaelis, S. Grimes, P. J. Jardine, D. L. Anderson, and C. Bustamante. 2005. Mechanism of force generation of a viral DNA packaging motor. Cell 122:683-692.

17. Smith, D. E., S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante. 2001. The bacteriophage straight phi29 portal motor can package DNA against a large internal force. Nature 413:748-752.

18. Hoeprich, S., Q. Zhou, S. Guo, D. Shu, G. Qi, Y. Wang, and P. Guo. 2003. Bacterial virus phi29 pRNA as a hammerhead ribozyme escort to destroy hepatitis B virus. Gene Ther. 10:1258-1267.

19. Morais, M. C., J. S. Koti, V. D. Bowman, E. Reyes-Aldrete, D. L. Anderson, and M. G. Rossmann. 2008. Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor. Structure 16:1267-1274.

20. Simpson, A. A., P. G. Leiman, Y. Tao, Y. He, M. O. Badasso, P. J. Jardine, D. L. Anderson, and M. G. Rossmann. 2001. Structure determination of the head-tail connector of bacteriophage phi29. Acta. Crystallogr. D Biol. Crystallogr. 57:1260-1269.

21. Guasch, A., J. Pous, B. Ibarra, F. X. Gomis-Ruth, J. M. Valpuesta, N. Sousa, J. L. Carrascosa, and M. Coll. 2002. Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle. J. Mol. Biol. 315:663-676.

22. Bourassa, N., and F. Major. 2002. Implication of the prohead RNA in phage phi29 DNA packaging. Biochimie 84:945-951.

23. Wichitwechkarn, J., S. Bailey, J. W. Bodley, and D. Anderson. 1989. Prohead RNA of bacteriophage phi 29: size, stoichiometry and biological activity. Nucleic Acids Res 17:3459-3468.

24. Morais, M. C., Y. Tao, N. H. Olson, S. Grimes, P. J. Jardine, D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2001. Cryoelectron-microscopy image reconstruction of symmetry mismatches in bacteriophage phi29. J Struct Biol 135:38-46.

25. Shu, D., H. Zhang, J. Jin, and P. Guo. 2007. Counting of six pRNAs of phi29 DNA-packaging motor with customized single-molecule dual-view system. EMBO J 26:527-537.

26. Zhang, F., S. Lemieux, X. Wu, D. St-Arnaud, C. T. McMurray, F. Major, and D. Anderson. 1998. Function of hexameric RNA in packaging of bacteriophage phi 29 DNA in vitro. Mol Cell 2:141-147.

27. Zhang, H., J. A. Endrizzi, Y. Shu, F. Haque, C. Sauter, L. S. Shlyakhtenko, Y. Lyubchenko, P. Guo, and Y. I. Chi. 2013. Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. RNA 19:1226-1237.

28. Schwartz, C., and P. Guo. 2013. Ultrastable pRNA hexameric ring gearing hexameric phi29 DNA-packaging motor by revolving without rotating and coiling. Current Opinion in Biotechnology 24:581-590.

29. Erzberger, J. P., and J. M. Berger. 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct 35:93-114.

30. Iyer, L. M., K. S. Makarova, E. V. Koonin, and L. Aravind. 2004. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res 32:5260-5279.

31. Kainov, D. E., R. Tuma, and E. J. Mancini. 2006. Hexameric molecular motors: P4 packaging ATPase unravels the mechanism. Cell Mol Life Sci 63:1095-1105.

32. Schwartz, C., G. M. De Donatis, H. Fang, and P. Guo. 2013. The ATPase of the phi29 DNA packaging motor is a member of the hexameric AAA+ superfamily. Virology 443:20-27.

33. Schwartz, C., G. M. De Donatis, H. Zhang, H. Fang, and P. Guo. 2013. Revolution rather than rotation of AAA+ hexameric phi29 nanomotor for viral dsDNA packaging without coiling. Virology 443:28-39.

34. Zhao, Z., E. Khisamutdinov, C. Schwartz, and P. Guo. 2013. Mechanism of one-way traffic of hexameric phi29 DNA packaging motor with four electropositive relaying layers facilitating antiparallel revolution. ACS Nano 7:4082-4092.

35. Guo, P., C. Schwartz, J. Haak, and Z. Zhao. 2013. Discovery of a new motion mechanism of biomotors similar to the earth revolving around the sun without rotation. Virology 446:133-143.

36. Wendell, D., P. Jing, J. Geng, V. Subramaniam, T. J. Lee, C. Montemagno, and P. Guo. 2009. Translocation of double-stranded DNA through membrane-adapted phi29 motor protein nanopores. Nat Nanotechnol 4:765-772.

37. Guo, P., S. Grimes, and D. Anderson. 1986. A defined system for in vitro packaging of DNA-gp3 of the Bacillus subtilis bacteriophage phi 29. Proc Natl Acad Sci U S A 83:3505-3509.

38. Rickgauer, J. P., D. N. Fuller, S. Grimes, P. J. Jardine, D. L. Anderson, and D. E. Smith. 2008. Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage phi29. Biophys. J. 94:159-167.

39. Cunniff, P. M., S. A. Fossey, M. A. Auerbach, J. W. Song, D. L. Kaplan, W. W. Adams, R. K. Eby, D. Mahoney, and D. L. Vezie. 1994. Mechanical and thermal properties of dragline silk from the spider Nephila clavipes. Polym. Adv. Technol. 5:401-410.

40. Perez-Rigueiro, J., C. Viney, J. Llorca, and M. Elices. 2000. Mechanical properties of single-brin silkworm silk. J. Appl. Polym. Sci. 75:1270-1277.

41. Wenger, M. P. E., L. Bozec, M. A. Horton, and P. Mesquida. 2007. Mechanical properties of collagen fibrils. Biophys. J. 93:1255-1263.

42. Chistol, G., S. Liu, C. L. Hetherington, J. R. Moffitt, S. Grimes, P. J. Jardine, and C. Bustamante. 2012. High degree of coordination and division of labor among subunits in a homomeric ring ATPase. Cell 151:1017-1028.

43. Aathavan, K., A. T. Politzer, A. Kaplan, J. R. Moffitt, Y. R. Chemla, S. Grimes, P. J. Jardine, D. L. Anderson, and C. Bustamante. 2009. Substrate interactions and promiscuity in a viral DNA packaging motor. Nature 461:669-673.

44. Fang, H., P. Jing, F. Haque, and P. Guo. 2012. Role of channel lysines and the "push through a one-way valve" mechanism of the viral DNA packaging motor. Biophys. J. 102:127-135.

45. Grimes, S., S. Ma, J. Gao, R. Atz, and P. J. Jardine. 2011. Role of phi29 connector channel loops in late-stage DNA packaging. J. Mol. Biol. 410:50-59.

46. Jing, P., F. Haque, D. Shu, C. Montemagno, and P. Guo. 2010. One-way traffic of a viral motor channel for double-stranded DNA translocation. Nano Lett. 10:3620-3627.

47. Rao, V. B., and M. Feiss. 2008. The bacteriophage DNA packaging motor. Annu Rev Genet 42:647-681.

48. Sun, S., V. B. Rao, and M. G. Rossmann. 2010. Genome packaging in viruses. Curr Opin Struct Biol 20:114-120.

49. Hendrix, R. W. 1978. Symmetry mismatch and DNA packaging in large bacteriophages. Proc. Natl. Acad. Sci. U S A 75:4779-4783.

50. Yu, J., J. Moffitt, C. L. Hetherington, C. Bustamante, and G. Oster. 2010. Mechanochemistry of a viral DNA packaging motor. J. Mol. Biol. 400:186-203.

51. Hugel, T., J. Michaelis, C. L. Hetherington, P. J. Jardine, S. Grimes, J. M. Walter, W. Falk, D. L. Anderson, and C. Bustamante. 2007. Experimental test of connector rotation during DNA packaging into bacteriophage phi29 capsids. PLoS Biol. 5:e59.

52. Hetherington, C. L., A. Karunakaran, J. Schnitzbauer, P. Jardine, S. Grimes, D. Anderson, and C. Bustamante. 2009. Bacteriophage Phi29 Negatively Twists DNA During Packaging. Biophys J 96:416a-416a.

53. Ray, K., C. R. Sabanayagam, J. R. Lakowicz, and L. W. Black. 2010. DNA crunching by a viral packaging motor: Compression of a procapsid-portal stalled Y-DNA substrate. Virology 398:224-232.

54. Amadei, A., A. B. Linssen, and H. J. Berendsen. 1993. Essential dynamics of proteins. Proteins 17:412-425.

55. Andersen, H. C. 1980. Molecular-Dynamics Simulations at Constant Pressure and-or Temperature. J Chem Phys 72:2384-2393.

56. Berendsen, H. J., and S. Hayward. 2000. Collective protein dynamics in relation to function. Curr. Opin. Struct. Biol. 10:165-169.

57. Bock, L. V., C. Blau, G. F. Schroder, I. I. Davydov, N. Fischer, H. Stark, M. V. Rodnina, A. C. Vaiana, and H. Grubmuler. 2013. Energy barriers and driving forces in tRNA translocation through the ribosome. Nature Structural & Molecular Biology 20:1390-1396.

58. Czub, J., and H. Grubmuller. 2011. Torsional elasticity and energetics of F1-ATPase. Proc Natl Acad Sci U S A 108:7408-7413.

59. Grater, F., J. Shen, H. Jiang, M. Gautel, and H. Grubmuller. 2005. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. Biophys J 88:790-804.

60. Grubmuller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 271:997-999.

61. Hub, J. S., B. L. de Groot, and D. van der Spoel. 2010. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J. Chem. Theory. Comput. 6:3713-3720.

62. Kappel, C., U. Zachariae, N. Dolker, and H. Grubmuller. 2010. An Unusual Hydrophobic Core Confers Extreme Flexibility to HEAT Repeat Proteins. Biophys. J. 99:1596-1603.

63. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. 1992. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. J. Comput. Chem. 13:1011-1021.

64. Kutzner, C., J. Czub, and H. Grubmuller. 2011. Keep It Flexible: Driving Macromolecular Rotary Motions in Atomistic Simulations with GROMACS. J. Chem. Theory. Comput. 7:1381-1393.

65. Kutzner, C., H. Grubmuller, B. L. de Groot, and U. Zachariae. 2011. Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. Biophys J 101:809-817.

66. Levitt, M., and A. Warshel. 1975. Computer simulation of protein folding. Nature 253:694-698.

67. Roux, B. 1995. The Calculation of the Potential of Mean Force Using Computer-Simulations. Comput. Phys. Commun. 91:275-282.

68. Shaw, D. E., P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers. 2010. Atomic-level characterization of the structural dynamics of proteins. Science 330:341-346.

69. Zink, M., and H. Grubmuller. 2009. Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study. Biophys. J. 96:1350-1363.

70. Zink, M., and H. Grubmuller. 2010. Primary Changes of the Mechanical Properties of Southern Bean Mosaic Virus upon Calcium Removal. Biophys. J. 98:687-695.

71. Warshel, A. 1997. Computer Modeling of Chemical Reactions in Enzymes and Solutions. John Wiely & Sons, Inc., New York, Chichester, Weinheim, Brisbane, Singapore, Toronto.

72. Leach, A. R. 2001. Molecular modelling: Principles and Applications. Prentice Hall, London.

73. Griebel, M., S. Knapek, and G. Zumbusch. 2007. Numerical Simulation in Molecular Dynamics: Numerics, Algorithms, Parallelization, Applications. Springer, Berlin Heidelberg.

74. Frenkel, D., and B. Smit. 2002. Understanding Molecular Simulation: From Algorithms to Applications. Acadmic Press, New York.

75. Haile, J. M. 1997. Molecular Dynamics Simulation: Elementary Methods. John Wiely & Sons, Inc., New York, Chichester, Weinheim, Brisbane, Singapore, Toronto.

76. Darden, T., D. York, and L. Pedersen. 1993. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. J. Chem. Phys. 98:10089-10092.

77. Verlet, L. 1967. Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. Phys Rev 159:98-&.

78. Hockney, R. W. 1970. Potential calculation and some applications. Academic Press, New York.

79. Swope, W. C., H. C. Andersen, P. H. Berens, and K. R. Wilson. 1982. A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters. J Chem Phys 76:637-649.

80. Berendsen, H. J. C., J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. 1984. Molecular-Dynamics with Coupling to an External Bath. J. Chem. Phys. 81:3684-3690.

81. Hess, B. 2008. P-LINCS: A parallel linear constraint solver for molecular simulation. J. Chem. Theory. Comput. 4:116-122.

82. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 18:1463-1472.

83. Gere, J. M. 2004. Mechanics of materials. Brooks/Cole - Thomson learning Inc.

84. Ichiye, T., and M. Karplus. 1991. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 11:205-217.

85. Kitao, A., F. Hirata, and N. Gō. 1991. The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. Chem. Phys. 158:447-472.

86. Hayward, S., A. Kitao, F. Hirata, and N. Go. 1993. Effect of solvent on collective motions in globular protein. J. Mol. Biol. 234:1207-1217.

87. van Aalten, D. M., J. B. Findlay, A. Amadei, and H. J. Berendsen. 1995. Essential dynamics of the cellular retinol-binding protein--evidence for ligand-induced conformational changes. Protein Eng 8:1129-1135.

88. Balsera, M. A., W. Wriggers, Y. Oono, and K. Schulten. 1996. Principal component analysis and long time protein dynamics. J Phys Chem-Us 100:2567-2572.

89. Hayward, S., and B. L. de Groot. 2008. Normal modes and essential dynamics. Methods Mol Biol 443:89-106.

90. Maisuradze, G. G., A. Liwo, and H. A. Scheraga. 2009. Principal component analysis for protein folding dynamics. J Mol Biol 385:312-329.

91. Mu, Y., P. H. Nguyen, and G. Stock. 2005. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. Proteins 58:45-52.

92. de Groot, B. L., G. Vriend, and H. J. Berendsen. 1999. Conformational changes in the chaperonin GroEL: new insights into the allosteric mechanism. J Mol Biol 286:1241-1249.

93. de Groot, B. L., S. Hayward, D. M. van Aalten, A. Amadei, and H. J. Berendsen. 1998. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. Proteins 31:116-127.

94. Fernandez-Fuentes, N., J. Zhai, and A. Fiser. 2006. ArchPRED: a template based loop structure prediction server. Nucleic. Acids. Res. 34:W173-176.

95. Eswar, N., B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, and A. Sali. 2007. Comparative protein structure modeling using MODELLER. Curr. Protoc. Protein. Sci. Chapter 2:Unit 2 9.

96. Macke Thomas, J., and A. Case David. 1997. Modeling Unusual Nucleic Acid Structures. In Molecular Modeling of Nucleic Acids. American Chemical Society. 379-393.

97. Garcia, A. E. 1992. Large-amplitude nonlinear motions in proteins. Phys. Rev. Lett. 68:2696-2699.

98. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory. Comput. 4:435-447.

99. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712-725.

100. Perez, A., I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, 3rd, C. A. Laughton, and M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys. J. 92:3817-3829.

101. Hess, B. 2000. Similarities between principal components of protein dynamics and random diffusion. Phys. Rev. E Stat. Phys. Plasmas. Fluids. Relat. Interdiscip. Topics 62:8438-8448.

102. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126.

103. Nose, S., and M. L. Klein. 1983. Constant Pressure Molecular-Dynamics for Molecular-Systems. Mol. Phys. 50:1055-1076.

104. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 79:926-935.

105. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14:33-38, 27-38.

106. Kumar, R., and H. Grubmüller. 2014. Elastic Properties and Heterogeneous Stiffness of the Phi29 Motor Connector Channel. Biophysical Journal 106:1338-1348.

107. Lu, X. J., and W. K. Olson. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res 31:5108-5121.

108. Lu, X. J., and W. K. Olson. 2008. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat Protoc 3:1213-1227.

109. Kappel, C., N. Dolker, R. Kumar, M. Zink, U. Zachariae, and H. Grubmuller. 2012. Universal relaxation governs the nonequilibrium elasticity of biomolecules. Phys. Rev. Lett. 109:118304.

110. Addadi, L., and S. Weiner. 1992. Control and Design Principles in Biological Mineralization. Angew. Chem. Int. Edit. 31:153-169.

111. Luz, G. M., and J. F. Mano. 2010. Mineralized structures in nature: Examples and inspirations for the design of new composite materials and biomaterials. Compos. Sci. Technol. 70:1777-1788.

112. Romer, L., and T. Scheibel. 2008. The elaborate structure of spider silk: structure and function of a natural high performance fiber. Prion 2:154-161.

113. Agirrezabala, X., J. Martin-Benito, M. Valle, J. M. Gonzalez, A. Valencia, J. M. Valpuesta, and J. L. Carrascosa. 2005. Structure of the connector of bacteriophage T7 at 8A resolution: structural homologies of a basic component of a DNA translocating machinery. J. Mol. Biol. 347:895-902.

114. Cuervo, A., and J. L. Carrascosa. 2012. Viral connectors for DNA encapsulation. Curr. Opin. Biotechnol. 23:529-536.

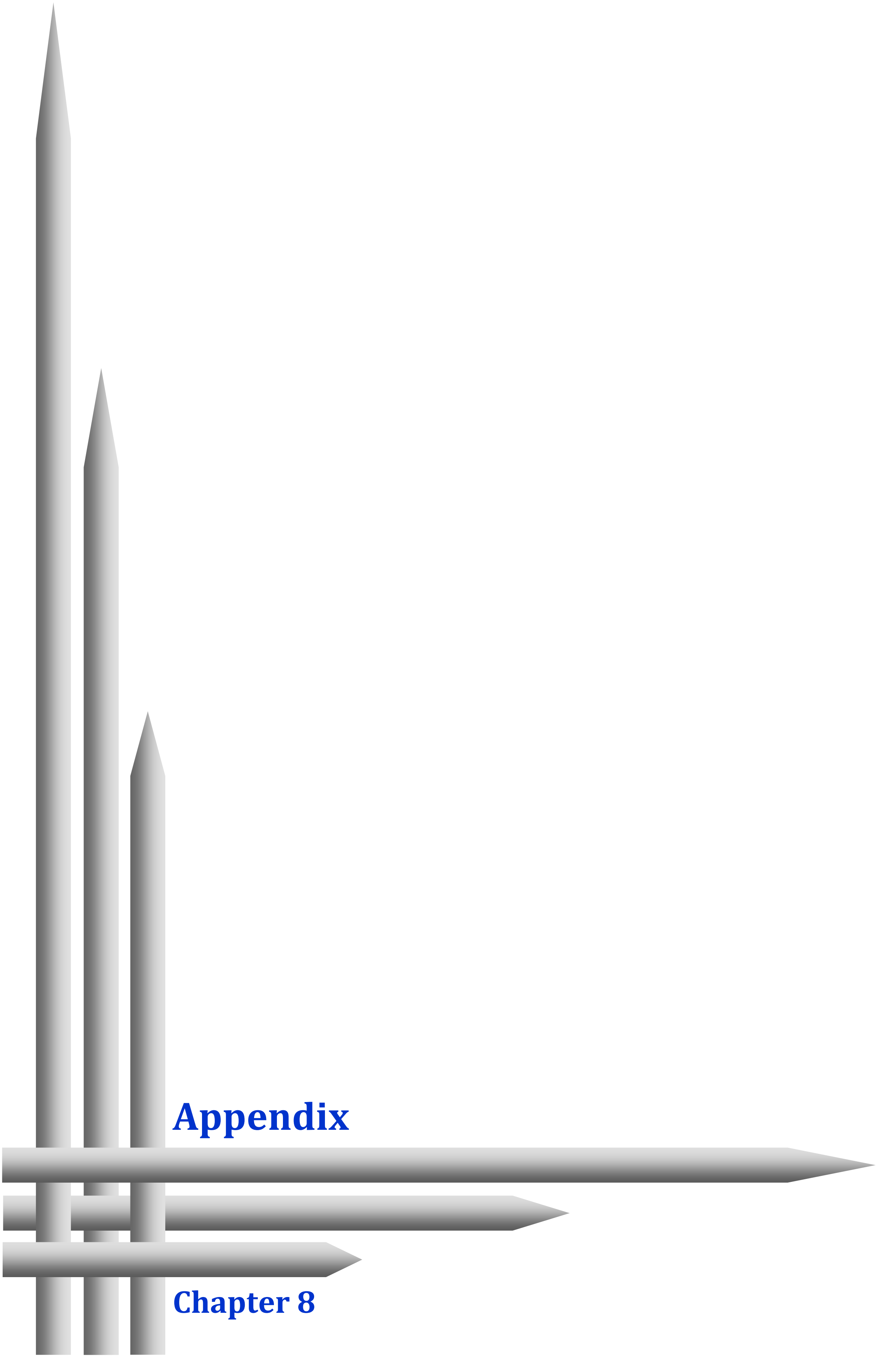
115. Cuervo, A., M. C. Vaney, A. A. Antson, P. Tavares, and L. Oliveira. 2007. Structural rearrangements between portal protein subunits are essential for viral DNA translocation. J. Biol. Chem. 282:18907-18913.

116. Lebedev, A. A., M. H. Krause, A. L. Isidro, A. A. Vagin, E. V. Orlova, J. Turner, E. J. Dodson, P. Tavares, and A. A. Antson. 2007. Structural framework for DNA translocation via the viral portal protein. EMBO J. 26:1984-1994.

117. Lhuillier, S., M. Gallopin, B. Gilquin, S. Brasiles, N. Lancelot, G. Letellier, M. Gilles, G. Dethan, E. V. Orlova, J. Couprie, P. Tavares, and S. Zinn-Justin. 2009. Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. Proc. Natl. Acad. Sci. U S A 106:8507-8512.

118. Olia, A. S., P. E. Prevelige, Jr., J. E. Johnson, and G. Cingolani. 2011. Three-dimensional structure of a viral genome-delivery portal vertex. Nat. Struct. Mol. Biol. 18:597-603.

119. Veesler, D., and C. Cambillau. 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol. Mol. Biol. Rev. 75:423-433.



Appendix

Chapter 8

8.1. Derivation of Young's modulus

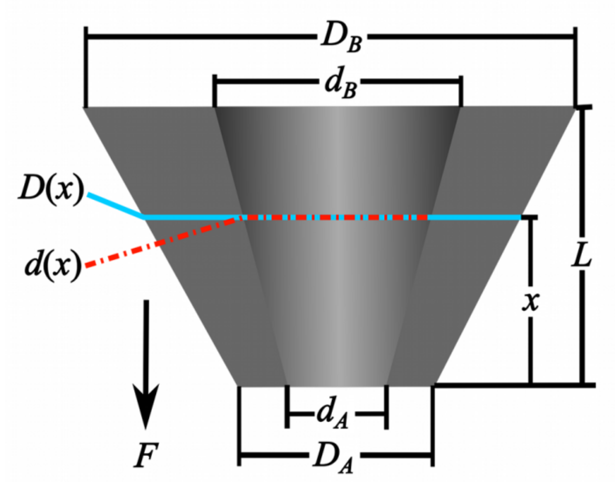


Figure A1.1: Sketch of the hollow truncated cone with the dimensions which was used in derivation of the Young's modulus elasticity.

Diameters vary with length L of the cone, therefore, at any length x diameters are given as follows,

$$d(x) = d_A + \frac{x}{L}(d_B - d_A) \tag{A1}$$

$$D(x) = D_A + \frac{x}{L}(D_B - D_A) \tag{A2}$$

The Young's modulus Y is given as follows,

$$Y = \int_0^L \frac{K_L}{A(x)} dx \tag{A3}$$

Where, K_L is spring constant and $A(x)$ is cross-section area normal to the channel axis. The $A(x)$ can be written as follows,

$$A(x) = \frac{\pi}{4}[\, D^2(x) - d^2(x)\,] \tag{A4}$$

After replacing $A(x)$ in Eq. A3, Young's modulus is given as follows,

$$Y = \int_0^L \frac{K_L}{\frac{\pi}{4}[\, D^2(x) - d^2(x)\,]} dx \tag{A5}$$

Equation A5 can be written as follows after replacing $D(x)$ and $d(x)$,

$$Y=\frac{4K_L}{\pi}\int_0^L \frac{dx}{\left[D_A+\frac{x}{L}(D_B-D_A)\right]^2-\left[d_A+\frac{x}{L}(d_B-d_A)\right]^2}$$

$$=\frac{4K_L}{\pi}\int_0^L \frac{dx}{\left[\frac{(D_B-D_A)^2}{L^2}-\frac{(d_B-d_A)^2}{L^2}\right]x^2+\left[\frac{2D_A(D_B-D_A)}{L}-\frac{2d_A(d_B-d_A)}{L}\right]x+[D_A^2-d_A^2]} \tag{A6}$$

Diameters and length of the truncated cone are constant and can be removed by the replacing the following equations,

$$a=\frac{(D_B-D_A)^2}{L^2}-\frac{(d_B-d_A)^2}{L^2} \tag{A7}$$

$$b=\frac{2D_A(D_B-D_A)}{L}-\frac{2d_A(d_B-d_A)}{L} \tag{A8}$$

$$c=\ D_A^2-d_A^2 \tag{A9}$$

Now, Eq. A6 can be written as follows,

$$Y=\frac{4K_L}{\pi}\int_0^L \frac{dx}{ax^2+bx+c} \tag{A10}$$

Solution for Eq. A10 depends on the value of ($4ac-b^2$), which can be written using Eq. A7, A8 and A9 as follows,

$$4ac-b^2=-\frac{4}{L^2}(d_AD_B-D_Ad_B)^2 \tag{A11}$$

Also,

$$\sqrt{b^2-4ac}=\frac{2}{L}(d_AD_B-D_Ad_B) \tag{A12}$$

Since, $4\mathrm{ac}-\mathrm{b}^2<0$; the solution of Eq. A10 can be written as follows,

$$Y=\frac{4K_L}{\pi}\frac{1}{\sqrt{b^2-4ac}}\left[\ln\left|\frac{2aL+b-\sqrt{b^2-4ac}}{2aL+b+\sqrt{b^2-4ac}}\right|-\ln\left|\frac{b-\sqrt{b^2-4ac}}{b+\sqrt{b^2-4ac}}\right|\right] \tag{A13}$$

The obtained Eq. A13 is required to be further simplify by substituting the expression for a, b and c from Eq. A7, A8 and A9, respectively. Therefore, several terms of Eq. A13 were simplified in separate steps which are shown in the following paragraphs.

Simplification of the term ($2aL + b$) using Eq. A7 and A8:

$$2aL + b == \frac{2}{L}[D_B(D_B - D_A) - d_B(d_B - d_A)] \qquad \text{(A14)}$$

Simplification of the term $\left(2aL + b - \sqrt{b^2 - 4ac}\right)$ using Eq. A12 and A14:

$$2aL + b - \sqrt{b^2 - 4ac} = = \frac{2}{L}[(D_A + d_A)(d_B - D_B) + D_B^2 - d_B^2] \qquad \text{(A15)}$$

Simplification of the term $\left(2aL + b + \sqrt{b^2 - 4ac}\right)$ using Eq. A12 and A14:

$$2aL + b + \sqrt{b^2 - 4ac} = \frac{2}{L}[(d_A - D_A)(d_B + D_B) + D_B^2 - d_B^2] \qquad \text{(A16)}$$

Simplification of the term $\left(b - \sqrt{b^2 - 4ac}\right)$:

$$b - \sqrt{b^2 - 4ac} = \frac{2}{L}[(D_A - d_A)(D_B + d_B) - (D_A^2 - d_A^2)] \qquad \text{(A17)}$$

Simplification of the term $\left(b + \sqrt{b^2 - 4ac}\right)$:

$$b - \sqrt{b^2 - 4ac} = \frac{2}{L}[(D_A + d_A)(D_B - d_B) - (D_A^2 - d_A^2)] \qquad \text{(A18)}$$

Now, after substituting and simplifying the terms derived in Eqs. A14, A15, A16, A17 and A18 into Eq. A13, Young's modulus can be written as follows,

$$\boxed{Y \;= \frac{2K_L}{\pi}\frac{L}{(D_B d_A - D_A d_B)} \ln\left|\frac{(D_B - d_B)}{(D_B + d_B)}\frac{(D_A + d_A)}{(D_A - d_A)}\right| \qquad \text{(A19)}}$$

Unit of Young's modulus: Y = MPa; where If K_L = pn/nm, D_i = nm and L=nm.

Acknowledgement

First, I would like to thank my supervisor ***Prof. Dr. Helmut Grubmüller*** *for providing me the opportunity to work in his research group, for his constant support and guidance throughout my Ph.D. research. I would also like to thank my thesis committee member* ***Prof. Dr. Holger Stark*** *for his participations in the committee meetings, and his valuable suggestions for the Ph.D. research.*

I would also like to thank ***Sarath C. Dantu, Andrea Vaiana, Christian Blau, Christian Kappel, Lars Bock*** *and* ***Shreyas Kaptan*** *for several helpful discussions regarding my research. I am thankful to* ***Carsten Kutzner*** *for his support and suggestions regarding the computer programming. I am also thankful to* ***Petra Keller*** *for her valuable suggestions on ways to improve my manuscript and the PhD thesis. Special thanks to* ***Mrs. Eveline Heinemann*** *for all the administrative support to work fluently in this department. I would like to thank all my colleagues from the department for providing a very friendly and motivational environment.*

I am grateful to my B. Pharm. professor ***Dr. Arijit Basu*** *for introducing this fascinating research field, and my M.S.(Pharm.) professors* ***Dr. M. E. Sobhia, Dr. P. V. Bharatam, Dr. C. Gopimohan*** *and* ***Dr. Prabha Garg*** *for seeding the foundation of my current research field.*

I am thankful to all my friends for their support, motivation and inspiration for pursuing research and for achieving this academic goal. My special thanks to my friend ***Nishant*** *for motivating and encouraging me. My special thanks also goes to my B. Pharm. friends,* ***Purushotam, Ajitabh, Pankaj, Shailesh, Chandan, Ashish, Rakesh, Amit, Shweta, Sabita*** *and* ***Shuchi*** *for their motivating talks and encouraging words. I would also like to extend my sincere thanks to* ***Hitesh, Hirdesh, Prakash, Vivek, Mihir, Vaibhav, Sikhar, Ashish, Sarvanan, Pramod, Akanksha,*** *and* ***Divya*** *for their motivational talks and valuable suggestions during my M.S.(Pharm.) course that helps me for building the foundation of my current research field.*

Words have no power to express my gratitude to my family members; ***mother, father, brother, sister-in-law, sister,*** *and* ***Brother-in-law*** *who blessed me with everything throughout my career. Their constant love, psychological support, and words have always been a source of inspiration to me and helped me in pursuit of my goals. Very special and heart-felt thanks to* ***Rashmi*** *for being with me in all good and bad times, and for all love and support she gave me, when I was lonely and gloomy in a foreign country thousands of miles away from my home.*

Curriculum Vitae

Personal Data

Name	Rajendra Kumar
Address	Augustinerstr. 26, Göttingen, 37077
Date of Birth	07.01.1985
Nationality	Indian

Education

2010 – Current	Pursuing Ph.D. in "Molecular Dynamics Studies of the φ29 Connector-DNA complex" at Max Planck Institute for Biophysical Chemistry, Göttingen, Germany.
07.2009 – 10.2009	Worked as Junior Research Fellow in a Research Project at School of Information Technology, Jawaharlal Nehru University, New Delhi, India.
2007 – 2009	Masters in Science (Pharm.) in Pharmacoinformatics from National Institute of Pharmaceutical Education and Research, S.A.S. Nagar, Punjab, India.
2003 – 2007	Bachelors in Pharmaceutical Science from Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India
1999 – 2001	AISSCE (12th) from Patna Central School, Patna, Bihar, India
1999	H.S.C. (10th) from R.S.B. High School, Samastipur, Bihar, India