1-1-2008

Molecular modeling and Langevin dynamics simulations of viral genome packaging and DS-DNA translocation.

Christopher Forrey
University of Massachusetts Amherst

MOLECULAR MODELING AND LANGEVIN DYNAMICS SIMULATIONS OF VIRAL GENOME PACKAGING AND DS-DNA TRANSLOCATION

A Dissertation Presented

By

CHRISTOPHER FORREY

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2008

Polymer Science and Engineering

MOLECULAR MODELING AND LANGEVIN DYNAMICS SIMULATIONS OF VIRAL GENOME PACKAGING AND DS-DNA TRANSLOCATION

A Dissertation Presented

By

CHRISTOPHER FORREY

Approved as to style and content by:

_____

Murugappan Muthukumar, Chair

_____

Todd S. Emrick, Member

_____

Jonathan Machta, Member

_____

Shaw Ling Hsu, Department Head
Polymer Science and Engineering

MOLECULAR MODELING AND LANGEVIN DYNAMICS

SIMULATIONS OF VIRAL GENOME PACKAGING

AND DS-DNA TRANSLOCATION

A Dissertation Presented

By

CHRISTOPHER FORREY

Approved as to style and content by:

Murugappan Muthukumar, Chair

Todd S. Emrick, Member

Jonathan Machta, Member

Shaw Ling Hsu, Department Head
Polymer Science and Engineering

DEDICATION

This dissertation is dedicated to my mother and father, Dr. Carolyn Broadaway and Dr. Robert Forrey, for their support and encouragement of my pursuit of a doctorate, and to my wife, Dr. Laura Espinal, for everything.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Murugappan Muthukumar, for his many years of unflagging support, as well as for his insistence that my work reflect the highest standard, of which he was always certain I was capable. I would like to express my gratitude for the superlative evaluation letter written for me by Professor Muthukumar, which has in no small way helped me to secure the coveted National Research Council Postdoctoral Fellowship at the National Institute of Standards and Technology. I would like to thank the members of my committee, Professor Todd S. Emrick and Professor Jonathan Machta for their valuable criticism of my research and for their support of my application for my postdoctoral research position. I would also like to thank Professor E. Bryan Coughlin and Professor Daniel Hebert for serving on my committee at various phases.

Finally, I wish to thank the National Science Foundation and the National Institutes of Health for funding this research and the Harvard Nanopore Group, in particular Professors Branton and Golovchenko, for stimulating discussions on DNA translocation.

ABSTRACT

MOLECULAR MODELING AND LANGEVIN DYNAMICS
SIMULATIONS OF VIRAL GENOME PACKAGING
AND DS-DNA TRANSLOCATION

SEPTEMBER 2008

CHRISTOPHER FORREY

B.A. OBERLIN COLLEGE

M.S. UNIVERSITY OF CONNECTICUT

M.S. UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D. UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Murugappan Muthukumar

Over recent decades, molecular biology has been transformed by the progression towards a more physical understanding of biological macromolecules, such that familiar biochemical behavior can be understood in terms of individual macromolecular structure. Such physical insight now guides modern approaches to biotechnology and fundamental biology. Langevin dynamics simulations offer the opportunity to build upon static structural models by treating macromolecules as dynamic entities. Results will be presented of the application of Langevin dynamics simulations to viral genome packaging and voltage-induced polyelectrolyte translocation. An overview will be given demonstrating how Langevin dynamics, taken together with remarkable experimental

efforts, has revealed the central importance of DNA/RNA dynamics in each of these phenomena. Based on the success of these studies, it is suggested that Langevin dynamics, in coordination with experimental efforts, represents a powerful tool for improving the fundamental understanding of various phenomena of biological and biotechnological import.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

MODELING OF DSDNA PACKAGING IN BACTERIOPHAGE

1.A Abstract

We use Langevin dynamics simulations to study the process by which a coarse-grained DNA chain is packaged within an icosahedral container. We focus our inquiry on three areas of interest in viral packing: the evolving structure of the packaged DNA condensate; the packing velocity; and the internal buildup of energy and resultant forces. Each of these areas has been studied experimentally, and we find that we can qualitatively reproduce experimental results. However, our findings also suggest that the phage genome packing process is fundamentally different than that suggested by the inverse spool model. We suggest that packing in general does not proceed in the deterministic fashion of the inverse-spool model, but rather is stochastic in character. As the chain configuration becomes compressed within the capsid, the structure, energy, and packing velocity all become dependent upon polymer dynamics. That many observed features of the packing process are rooted in condensed-phase polymer dynamics suggests that statistical mechanics, rather than mechanics, should serve as the proper theoretical basis for genome packing. Finally we suggest that, as a result of an internal protein unique to bacteriophage T7, the T7 genome may be significantly more ordered than is true for bacteriophage in general.

1.B Introduction

Ubiquitous and intertwined with life throughout human history, viruses are of central importance in biology. In the past century, giant strides have been made in the epidemiology, pathology and biochemistry of viruses, yet because of their sub-cellular size, our knowledge of viruses has necessarily been indirect[1]. With the advent of recent technological innovations, particularly single molecule and atomic force experiments, we are for the first time capable of studying viral properties directly, down to the smallest length scale of the biological spectrum, the individual macromolecule. In the quest to uncover the physical nature of viruses, the stage has been reached where polymer physics and molecular biology merge.

The specific aspect of viruses on which we focus in this paper, the packaging of the viral genome within its rigid protective coat, or capsid, is a remarkable physical process. Linear dimensions of capsids are typically tens of nanometers, while the length of the genome to be packaged is generally three to four orders of magnitude longer[2]. The persistence length - the length scale along the chain contour for which directionality remains correlated - is approximately 50 nm in physiological conditions, the same order of magnitude as the linear dimensions of a typical viral capsid. The bending required of the genome as it is wound tightly within the capsid thus leads to a buildup of energy that is large in comparison with k_BT. Additionally, the presence of phosphate linkages leads to high linear charge density along the DNA backbone and gives rise to large repulsive energy inside of the densely packed capsid. This most simple physical picture of the process of viral genome packing reveals a fundamental conflict of scales, wherein a long

molecule must be compressed within a length scale on which it resists bending and to a density at which it must also overcome strong repulsive forces. The packing process has long been a subject of interest because of the tremendous pressures thought to arise inside the capsid.

We are interested in addressing two fundamental topics regarding the packing of dsDNA into capsids. The first topic is the configuration of the collapsed chain within the capsid, in particular how the structure evolves in the course of the packing process. In free solution, dsDNA is known to collapse in the presence of multi-valent salts into highly compacted doughnut-shaped toroids[3, 4], an example of a process termed 'DNA condensation' [5]. The positively charged multi-valent salt ions mediate an apparent attraction between the negatively charged strands of DNA. The final toroidal structure results from a compromise between attractive interactions and bending energy. The viral packing process involves a similar structural transformation of DNA from a free solution configuration to a highly compressed packaged structure. It is worth exploring how closely the compaction in virus heads parallels the formation of free solution toroids.

For nearly thirty years, wide angle X-ray diffraction (WAXD) data have been available for various bacteriophage (T7, T2, λ) demonstrating that the presence of packaged DNA gives rise to a strong ring at 1/24 Angstroms [6, 7]. Images of cross sections of free-solution toroids obtained from cryo-TEM reveal hexagonal close packing of the DNA strands, indicative of a highly compact structure within the body of the toroid [3]. Most current theoretical studies of the viral packing problem cite TEM images of packed virus heads. One cryo-TEM study in particular presented striking images of order obtained as a composite of individual T7 viruses in cross-section [8]. The virus images

reveal, depending upon the angle of viewing, close-packed or banded structures that have been widely interpreted as arising from some toroidal or highly ordered structure. Other TEM studies have proposed a more modest ordering within viruses, with a local nematic-type order restricted to small regions of the capsid [9-15].

Regardless of how much order can be inferred from TEM studies, packaged ds-DNA structures are not likely to correspond to free solution toroid structures. Despite the many similarities between viral packing and polyelectrolyte condensation, the effect of confinement, unique to the packing process, is likely to prevent the formation of the same structures seen in free-solution toroids within the capsid. The outer diameter of the observed free solution toroidal structures are generally too large to fit inside virus capsids. In addition, a free solution toroidal structure, with its fairly large inner unfilled core (the diameter is set by the relatively large persistence length of dsDNA), could not fill the capsid interior to known packing densities.

The structure now most commonly attributed to packaged dsDNA, and more reasonably in accord with the notion of confinement than a free-solution toroid, is the so-called 'inverse spool' structure [16, 17]. In fact, the inverse spool idea not only indicates a final packaged structure, but also suggests the *process* by which the DNA is wound. Inverse spooling is essentially a prescription for minimizing the bending energy of the chain, where the chain is treated as a mechanical string with bending resistance. The chain winds about and fills the outer bounds of confinement in an orderly fashion, with the most recently packed strands stacked upon previous strands. Eventually a complete new layer is formed which then serves as the outer bounds of confinement. As the process repeats, the chain subsequently winds progressively inwards towards some

central axis. This process is very similar to the inverse of the process by which a length of thread is wrapped about a spool, in this case towards rather than away from a central axis.

However appealing the notion, experimental evidence has not conclusively demonstrated radial ordering within bacteriophage in general. One of the earlier theories of viral packing posited a much less ordered structure akin to a ball of yarn [18] and many other structures have been proposed over the years. How and to what extent the DNA is ordered is not known.

The second question we wish to address involves the interplay of various forces, e.g. bending, electrostatic, and entropic, that arise during translocation[19, 20] and compaction. Trying to deduce the interplay of these forces presents an even larger challenge than trying to deduce the packaged structure itself and is well beyond current experimental capabilities. The study of the interplay of forces has thus far been addressed only by analytical theory [16, 17, 21-24]. Unfortunately, theories to date have required that some final structure be assumed. Additionally, forces in polyelectrolyte solutions are also notoriously difficult to treat with analytical theory, because they arise from both short and long-range potentials. While we can currently infer that forces involved in packaging are very large, we cannot detail which forces are important and how the forces dictate the structure of the condensed viral genome. Progress in the understanding of these two questions would be of great help in understanding this crucial step of the viral replication cycle.

Two recent publications on genome packing in bacteriophage have had great influence, particularly in advancing a more physical approach to the subject. The

publications were the result of remarkable experiments, each providing clues to the physical nature of individual virus particles. The first to appear was the cryo-TEM study of bacteriophage T7 of Cerritelli *et al* [8]. More recently, Smith *et al.* published their now famous single molecule optical tweezer study of the ϕ29 bacteriophage [25]. Taken together, the papers give us information about the packaged structure and internal force buildup of two different viruses, yet the findings do not necessarily represent complementary findings. While the work of Smith *et al.* addresses the buildup of force within the capsid, it gives us no indication of corresponding packaged structure. Conversely, while Cerritelli *et al.* literally give us a picture of the packaged structure, nothing can be said from this picture of internal force.

Further, we do not know that T7 and ϕ29 are representative of bacteriophage in general. The images of pervasive radial order in T7 have not been seen in any other bacteriophage that we are aware of. That order has yet to be detected in other phage may suggest that the radial order in T7 is not characteristic of bacteriophage in general, but rather is the result of a structural feature unique to T7. Additionally, there is some question as to the degree of ordering implied by the cryo-TEM images of Cerritelli *et al.*, as the striking order followed only from a great deal of averaging of individual images.

We feel that current available data does not point conclusively to any general packaged configuration or process whereby it is formed. In light of this uncertaintly, the usefulness of Langevin dynamics simulation becomes evident. Motivated by these considerations, we have preformed Langevin dynamics simulations, where we can simultaneously follow the evolution of packaged DNA structure and the resulting internal

energies, without having to assume a final structure or bias the packing process arbitrarily.

1.C **Simulation Technique**

1.C.1 Simulation of Viral DNA

Because genomes are giant macromolecules composed of an enormous number of individual atoms, all-atom molecular dynamics simulations of entire viral genomes are currently not possible. In order to model ds-DNA, the type of viral genome with which we are concerned, we follow the traditional course-graining approach in which the chemically-detailed double helix is replaced by a wormlike statistical chain of identical beads. Detail of the DNA structure on a length scale smaller than the bead diameter is lost. However, our goal is to describe global, rather than local, properties of the DNA chain. The simulated polymer is parameterized by the bead diameter d, an effective 'bond' length, l_o, the number of beads N, and the effective charge density α. The bead diameter, d, is a measure of the range of excluded volume interaction, which is itself determined by how many physical base pairs are subsumed in each bead. We wish to represent the roughly cylindrical or tubular geometry of the DNA double helix by a linear assembly of spherical beads, which we do in the following way: we fix d to be 2.5 nm, the hydrated diameter of ds-DNA in physiological conditions and we set l_o equal to half of d. Consecutive spherical beads along the chain are thus overlapped to form a more approximately cylindrical structure. To avoid un-physical repulsion, the excluded

volume interaction does not apply to consecutive beads along the chain. The resulting space-filling chain has a diameter that matches ds-DNA. Eight hundred such beads are required to simulate each micrometer of phage genome. Throughout most of this paper, we set α=0. However, we will consider the case of a self-attractive chain, in which we are simulating indirectly the electrostatic bridging of DNA mediated by polyvalent cations. We will also present a brief section using Debye-Huckel interactions to model electrostatics, taking into consideration a range of values of α. For now, ignoring electrostatics greatly simplifies our calculations. We find that our results are qualitatively insensitive to the specific functional forms of the repulsive interactions.

1.C.1 **Simulation of Viral Capsid**

Confinement is imposed on each polymer bead in the simulation by the application of an icosahedrally symmetric force field. Twenty planes are defined, corresponding to the twenty faces of an icosahedron. The interior of the capsid includes the spatial region between these twenty planes, where beads are not subject to any confinement forces. Beads that traverse the confining planes in the outward direction are subjected to an inward-pointing restoring force normal to the confining plane. The magnitude of the restoring force is a function of the distance of encroachment of polymer beads into the confining potential and is given identically by the Lennard-Jones potential (to be given explicitly in the following section) for two polymer beads encountering each other. The capsid is parameterized by a single scalar size parameter (the distance from

capsid center to the center of any face of the icosa) and the Lennard-Jones parameters borrowed from the bead-bead interactions.

Our representation of the capsid as a static force field is appropriate for the focus of this study: the structure of the confined DNA itself. We are interested in the capsid only indirectly, i.e. for the role it plays in determining the packaged DNA structure through confinement. Mechanical properties of the capsid, such as elasticity and ultimate strength, are beyond the scope of this dissertation.

1.C.2 **Force and Energy Calculations**

DNA beads in our simulation are subject to interactions intended to capture the relevant physical forces involved in the packing process. There are two classes of bead potentials in our model: DNA-DNA and DNA-capsid.

First, we describe interactions between beads of our simulated DNA:

1.C.3.a Excluded Volume Interaction

(i) The excluded volume interaction is modeled by a Lennard-Jones potential, U_{LJ}:

$$U_{LJ} = \varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right] \qquad \textbf{(Eq. 1-1)}$$

where r is the center-of-mass distance between two beads, σ is the hard core diameter, and ε is the strength of the potential. The energy minimum is at r=σ. For r < σ, a repulsive force arises, whereas for r > σ, an attractive force is found. For the case of a self-repulsive chain, the Lennard-Jones potential is truncated at r = σ.

1.C.3.b Bond Stretching Interaction

The energy required to extend the bond between two consecutive beads along the chain is given by:

$$U_{stretch} = k_{stretch}(l-l_o)^2 \qquad \textbf{(Eq. 1-2)}$$

where $k_{stretch}$ is the stretch modulus and l_o is the natural bond length.

1.C.3.c Bending Interaction

(iii) Bending interaction. The work required to introduce a bend of angle θ along the backbone of the polymer is given by the three-body interaction:

$$U_{bend} = k_{bend}(cos\theta - cos\theta_o)^2 \qquad \textbf{(Eq. 1-3)}$$

where θ is the bond angle formed by three consecutive beads, θ_0 is the effective equilibrium bond angle, and k_{bend} is the bending modulus. The bending resistance of the genome is of central importance in this work. Therefore, from here on, we refer to k_{bend} simply as k.

Now we describe the DNA-capsid interactions:

1.C.3.d Repulsive Interaction Between the Capsid Wall and DNA

This interaction is modeled by the purely repulsive portion of the potential form used for DNA-DNA excluded volume interactions discussed previously, $U_{LJ}(r;\sigma_{wall},\varepsilon_{wall})$. The distance r refers to the distance from the center-of-mass of an encroaching polymer bead to the nearest restraining wall of the capsid. Parameters are chosen so that the repulsion felt by a polymer bead as it encroaches upon a capsid wall matches that for a polymer bead as it encroaches upon another polymer bead.

1.C.3 Dynamics

The molecular motor that drives the packing is simulated by introducing a force to the portion of the DNA that occupies the position of the opening at the top of the capsid. As packing proceeds, successive portions along the chain contour occupy this position

and are thus subject to the packing force. The simulation algorithm entails solving the Langevin equation for each bead of the polymer chain:

$$m_i \ddot{r}_{ij} = -\vec{\nabla}_j U_i + f_{i^*j}^{motor} - \zeta r_{ij} + f_{ij}^{rand} \qquad \textbf{(Eq. 1-4)}$$

where i is the bead index and j is the component index. Thus, r_{ij} is the j^{th} component of the position vector of the i^{th} bead. Each bead is assigned a mass, m_i. The first term on the right hand side of Eq. 4 represents the sum, for each bead, of the configuration-dependent forces. These forces are derived from the intra-DNA and DNA-capsid interactions described in the previous section. The second term describes the effect of the molecular motor as a downward force on the polymer bead i^* located at capsid mouth. The effect of solvent, acting primarily through frequent collision of the relatively tiny and very fast moving solvent molecules with much larger polymer segments, is described by the third and fourth terms. The third term is the velocity-dependent viscous drag force, where ζ is the coefficient of viscous drag. The fourth term is the random force acting on each bead, which is assumed to obey a Gaussian distribution centered at zero. The square of the standard deviation of the distribution is given by the fluctuation-dissipation theorem, relating the magnitude of the deterministic frictional force to the random thermal bath force in thermal equilibrium,

$$< \boldsymbol{F}_i(t) \cdot \boldsymbol{F}_j(t') > = \delta_{ij} 6 k_B T \zeta \delta(t - t') \qquad \textbf{(Eq. 1-5)}$$

where F(t) is the random force at time t, k_B is the Boltzmann constant, and ζ is the coefficient of viscous drag. We have chosen as characteristic units the thermal energy available at room temperature, k_BT_{298}=4.1 pN nm, the Bjerrum length in water at room temperature, l_B=0.7nm, and $\zeta = 7.5x10^{-13} Ns/m$, for a sphere of diameter 2.5 nm in water at room temperature. The characteristic time scale Δt in our simulation is the minimum interval over which thermal bath forces are uncorrelated, which is determined by the fluctuation-dissipation theorem as $\Delta t = 6\, l_{sim}^2 \zeta / k_B T_{300} \sim 43$ ns. Each Langevin dynamics step corresponds to a time $\delta t = 0.0003\Delta t = 12.9$ ps.

The Langevin dynamics simulation proceeds through the solution of the equations of motion of the N particles in the system, which are integrated using the standard velocity verlet algorithm

$$v_{ij}\left(t+\delta t\right) = v_{ij}\left(t\right) + \frac{1}{2}\delta t\left[\ddot{r}_{ij}\left(t\right) + \ddot{r}_{ij}\left(t+\delta t\right)\right] \qquad \textbf{(Eq. 1-6)}$$

$$r_{ij}\left(t+\delta t\right) = r_{ij}\left(t\right) + \delta t\dot{r}_{ij}\left(t\right) + \frac{1}{2}\delta t^2\ddot{r}_{ij}\left(t\right) \qquad \textbf{(Eq. 1-7)}$$

where r^{ij} is as previously defined and v^{ij} is the j^{th} component of the velocity of the i^{th} bead.

1.C.4 **Varying the Number of Particles in the Simulation**

The simulation does not proceed with a constant number of beads, N. At any given time-step in the simulation, N is comprised only of those beads located within or at the opening of the capsid. Two beads are defined to be in the opening: i*, the bead occupying the location at which the motor force is applied, and the bead directly above i*. The external (*i.e.* un-packaged) length of the chain is simply excluded from consideration. The inherent assumption - that the external portion of the chain does not play a significant role in determining the course of packing – is based on both our initial simulation work that included external chain beads and the experimental work of Smith *et al.*, where the external portion of DNA is attached to a 2.2 μm polystryrene sphere and straightened by tension. Therefore, for the external portion of the chain, we expect that the bending energy is minimal and that the viscosity is dominated by the giant polystyrene sphere and is thus relatively invariant to degree of packing. In our preliminary simulation work, we found that an effective tension, acting through the downward force on the chain at the capsid opening and the opposing viscous drag, maintained the initially vertically aligned configuration of the external beads throughout the course of packing. Thus inclusion of the external beads in our initial simulation work, while demanding in terms of computing time, offered in return very little insight into the packing process. In our current simulations, the initial chain is simply a small (3-4) number of beads aligned vertically at and below the capsid opening. Each time the motor force pushes a bead into the capsid, a new chain-end bead is created at the upper end of the chain, to fill the resulting vacancy created at the capsid opening. N is thus

increased throughout the packing process. At any given time, the terminal bead is given a large frictional coefficient, $\zeta_{terminal} = 75\zeta$, to maintain a large and constant external viscous resistance to packing, a simple approximation of the external chain and PS sphere. This procedure allows us to expend computing resources solely on our stated focus of interest, the evolution of packaged structure and its relationship to the internal buildup of forces.

1.C.5 Determination of Persistence Length

The limited flexibility of ds-DNA plays a crucial role in the current theories of viral packing. A simple statistical mechanical calculation of persistence length can be obtained by assuming each bond angle along the chain to be independent and calculating the ensemble average bond angle as

$$\langle \cos\theta \rangle = \frac{\int_{\phi}\int_{\theta} \cos\theta P(\theta) \sin\theta d\theta d\phi}{\int_{\phi}\int_{\theta} P(\theta) \sin\theta d\theta d\phi} \qquad \textbf{(Eq. 1-8)}$$

where P(θ) is the distribution function. The persistence length is then given as [27],

$$L_p = l_o/(1+\langle\cos\theta\rangle) \qquad \textbf{(Eq. 1-9)}$$

where l_o is the natural bond length and θ is complementary to the bond angle ϕ formed by consecutive bonds. From Eqs. 8 and 9, we have calculated predicted values of persistence length for various values of k, as shown in Fig. 1, a.

In order to make a fair comparison with theory, we must verify that our basic simulated chain ($k_{end} = 750$) has a persistence length of approximately 50 nm. We have performed simulations of our chain in free solution to calculate the persistence length. The calculation involves periodically projecting the end-to-end distance of the chain along the initial bond vector of the chain. The time average of the projected vector (see Fig. 1, b) is the chain persistence length [28]. In addition, we monitor the instantaneous average bond angle throughout the simulation and using Eq. 9 to calculate a persistence length (see Fig. 1, c).

1.D Results and Discussion

In the previous section, we have outlined a number of simplifying assumptions that have allowed us to define our simulations through the selection of a small number of parameters. In all simulations, a constant motor force of 55 pN was applied, approximately the value of the average 'stall' force of the ϕ29 portal motor. As the chain becomes compressed within the capsid, our molecular motor packs against the internal resistive force with the same force as does the ϕ29 molecular motor. The simulated capsid is 0.7 times the height of the ϕ29 capsid, corresponding to approximately 1/3 the volume. $\delta t = 13$ ps for all simulations. For bending constants we have used a series of

values that are integral multiples of the value, k=750, corresponding roughly to a free solution persistence length of 60 nm.

1.D.1 Packing Velocity

Our simulation program outputs the loaded length, or more precisely, the number of beads packaged, as a function of time. The experimentally measurable quantity is the tether length, or the length of viral DNA external to the capsid. For the sake of comparison with experiments, we define for the simulations a tether length: *tether length = [number of beads packed upon completion] – [number of beads currently packed].* The packing velocity can be determined either by taking the slope of the number of beads packed versus time or, equivalently, the negative of the slope of the tether length versus time. We plot the tether length versus simulation time in Fig. 2 and the packing velocity versus loaded length in Fig. 3. In Fig. 4, we plot the loaded length versus simulation time. Chains parameterized with four different bending constants, k = 0, 750, 1500, 2250, have been used. The constants are integer multiples (from of 0 to 3) of the value k=750 corresponding roughly to the persistence length of ds-DNA in physiological conditions. Qualitatively, the curves of simulated tether length (Fig. 2) and packing velocities (Fig. 3) are strikingly similar to those obtained experimentally.

However, by observing packing velocities of individual runs (Fig. 5), we can see that the process of averaging has masked the presence of distinct pauses, which are common in our packing simulations. In addition, the plot of number of beads packaged versus time (Fig. 4) does not fit very closely with current theories of viral packing. The

packing curve (Fig. 4) reveals two qualitatively different modes of packing. For at least the first half of the packing process, the packing curve is highly regular, both in the constancy of packing velocity and in the indistinguishable and uniform behavior from run to run. During the final stage of packing, the packing becomes a highly irregular function of time. Comparing different runs at a given time, we find that one run may be stalled while another is undergoing rapid packing. Similarly, given a single run, the packing can be paused, immediately followed by a burst of packing. Simulations have been run for very long timesf and we find that in most cases additional, albeit miniscule, increases in packing can be observed by simply waiting long enough. This suggests that packing in the later stages depends upon the internal chain configuration undergoing some slow relaxation process, allowing additional space for additional beads. Here we have our first hint of one of our major themes in this paper, that polymer dynamics play a crucial role in the process of chain confinement.

We also note that the magnitudes of packing rates in our simulations are not directly comparable with experiment, as high packing velocities are required in simulations to complete the simulation in a reasonable time. The initial plateau region (Fig. 3) of the velocity curves is similar for all stiffnesses, indicating that the initial rate of packing is not strongly dependent upon chain stiffness. The packing rate is level until some critical degree of packing, at which point there is a decrease in the slope seen as packing nears completion. In the case of k=0, where the chain does not have bending stiffness, the initial velocity of 0.4 simulation units is faster than for cases of finite bending stiffness. However, there does not appear to be significant difference in initial

velocity for the chains with k=750, 1500, or 2250. There is a noticeably more abrupt decrease towards zero packing velocity found in the cases of non-zero bending stiffness.

Fig. 2 shows that the tether length decreases very rapidly in the initial stage of packing and then much more slowly in the final stage of packing. (In the work of Smith *et al.*, the tether length is much longer, as it includes both the portion of the chain remaining to be packed, as well as an additional spacer length.) The tether length curves are in qualitative agreement with experimental results, with one obvious difference: our simulated curves have more prominent plateau regions where packing is temporarily halted. While pauses are also visible in the experimental results, they are not as prominent as those in our simulations. Recall that our simulations proceed with a constant packing force and thus the pauses cannot be related to lack of evenness of the motor. The pauses are then most likely temporary blockages due to crowding of the chain within the capsid, with the duration of the pauses being related to the relaxation dynamics of the confined polymer. Relaxation processes of polymers are quite slow on the time-scales accessible in simulation of polymers. Thus, the relatively exaggerated duration of pauses in simulation results can likely be explained by the small timescale of our simulations. High packing velocity gives a steeper slope in the tether length curve, while on the other hand the chain relaxation process is a function only of chain stiffness, not packing velocity. The high packing velocity thus may explain the exaggerated pause duration as compared to experimental results.

The pauses in our simulations and the pauses seen experimentally may in fact both be the result of chain relaxation phenomena. This idea is supported by further examination of the pauses. The duration of the pauses and the loaded length at which

they arise do not appear to fit any regular pattern, except that no pauses are seen until at least 25% of the genome has been packaged. This behavior is true both in our simulations (Fig. 4) and in experimental results. The fact that pauses are not seen in the early stage of packing supports the idea that the pauses are due to relaxation in the crowded capsid environment. A possible rivaling theory, that experimentally observed pauses are related to specific DNA sequences, is disputed by the experimental results themselves, which indicate no correlation between pauses in the later stages of packing. Smith *et al.* attribute pauses entirely to choppiness of the molecular motor. However, the absence of pauses in the early stage of packing points more plausibly towards internal chain relaxation, which become slower at higher packing density.

Finally, note in Fig. 4 that the final degree of packing, even for identically parameterized simulations, varies from run to run. Thus, the final degree of packing in our simulations is apparently not pre-determined, as it would be in the case of highly ordered packing. This is particularly true for the chains with finite stiffness. Interestingly, in the experiments of Smith *et al.* also, the overall change in tether length varies from run to run. Their interpretation is that the results do not indicate variable degree of final packing, but rather are artifactual, arising from DNA-bead attachment position. On the other hand, we believe that the long polymer is essentially forced into metastable conformations in a short period of time. As the polymer is packed at high velocity, the packaged portion of the chain is likely not attaining a stable equilibrium structure. Instead it constantly searches for lower energy configurations (to be addressed in more detail later) by undergoing various relaxation processes. The end of the pauses

seen in our simulations is likely related to relaxation of chain configuration, which once achieved frees up additional volume for packing.

The explanation of stochastic packing results in terms of polymer relaxation implies that the velocity at which the chain is being fed into the capsid causes the resulting packaged polymer structure to diverge from a corresponding hypothetical equilibrium structure, i.e., the structure that would be obtained from an infinitely slow packing rate. The relaxation processes result from coordinated displacements of chain segments relative to each other and represent incremental steps back towards the 'equilibrium' configuration. The evolution of non-equilibrium chain structure and the punctuated steps back towards equilibrium are what give the packing its stochastic character. The most notable implication of the stochastic packing is that there cannot be a one-to-one correspondence between loaded length and packaged structure. Non-deterministic packaging is a notion quite at odds with current theoretical models of force buildup and the inverse-spool model on which they are based.

Because the chain is very long and its motion severely constrained by confinement, relaxation phenomena are quite slow, more particularly for the larger-scale rearrangements. Even with the coarse-graining approach outlined in the previous section, the time span accessible in our simulation is smaller by orders of magnitude than the experimental time scale (~ms to min., respectively). It is an unavoidable consequence that the velocity of packing in simulations is higher than it is in experiment. The packing velocities measured by Smith *et al.* were of the order of 10^{-2} μm/s. Our initial packing velocities were slightly less than 5000 μm/s, similar to the packing rate used in previous simulation work[16]. Thus, it is likely the case for simulations that, should the polymer

configuration fall out of its equilibrium in the course of packing, it may never be able to recover fully in the course of simulations of reasonable time duration. In fact, the same may be true in nature as well. Smith *et al.* measured a characteristic packaging time of the order of a minute. However, in their experiments, the ϕ29 motor is burdened with an unnaturally large load: spliced to the end of the wild type ϕ29 genome is additional piece of DNA approximately equal in length to the entire ϕ29 genome. The resulting strand of DNA is then anchored to a 2.2 μm poly-styrene sphere. Such additional drag surely reduced the packing velocity greatly as compared to wild-type packing, assuming wild-type packing occurs on times of the order of seconds, significantly shorter than the time for large scale polymer relaxations. The packing velocity in nature likely falls somewhere in between experimental and simulation packing velocities. Our results suggest that the pauses and variations in final loaded length, both in simulations and in the experiments of Smith *et al.*, may be attributable to non-equilibrium evolution of packaged structure. This non-equilibrium nature of packaging may be an important, and widely unrecognized, consideration.

1.D.2 Chain Configuration and Order

It is widely believed, based on the bending rigidity of ds-DNA, that a well-ordered ds-DNA structure is formed during confinement within bacteriophage capsids. The ordered structure is thought to result from the minimization of chain bending energy. The idea is supported, partly, by the cryo-TEM images that show radially periodic density in phage T7[8]. Indeed, when bending energy is allowed to act in the absence of

any solvent, it has been shown that energy minimization leads to well-ordered structures [23]. However, at finite temperatures, chaotic thermal motion of solvent molecules constantly agitates the chain towards a state of higher disorder; there is constant competition between disorder-inducing thermal fluctuations and the order-inducing bending potential. The degree of order in the case of viral packing should depend upon the degree to which thermal disordering is overmatched by bending energy, of which the persistence length is a quantitative measure.

It is often assumed, because of the similarity of the persistence length and the capsid dimensions, that the bending forces acting on the packaged DNA dominate thermal forces. However, even on the length scale of the persistence length, a molecule exhibits a good deal of thermal fluctuation. This principle is illustrated in Fig. 6, where we show snapshots from a simulation of a chain undergoing Brownian motion in free solution. The chain was originally perfectly straight, i.e. in its minimum energy configuration, and aligned along the z-axis, to which the viewing axis is perpendicular. The first two beads were held fixed, as is evident in the images. The chain, composed of a non-overlapping string of beads, has been parameterized such that its persistence length is 20 beads. Only the first 20 beads of the chain are shown, so that we are observing chain configuration of a chain segment whose contour length is equal to the persistence length. As can be seen, the chain retains significant flexibility even below the length scale of its persistence length, in obvious contrast with its energy minimized structure, i.e. unbent. Thus, confinement of a chain to a container of linear dimensions of the persistence length does not effectively render the chain inflexible as is commonly presumed.

In light of this insight, we will now proceed by addressing two basic assumptions of ordered packing. First, should we expect order when we apply the mechanical treatment of DNA at finite temperature and second, do we really see such order in cryo-TEM images? Our treatment of DNA, as a mechanical body weakly coupled to a thermal bath, allows us to address the first question, that is, whether an ordered structure can be expected from simple bending and confinement considerations only. In Fig. 7, a and b, periodic snapshots from two simulations, with stiffness parameters, k=750 (Lp = 60 nm) are shown. For each simulation, three snapshots of DNA structure are shown, at the beginning, middle and late stages of packing. The capsid height in both cases, 35 nm, is less than the persistence length of either of the chains. Interactions between different segments of the chain are (a) purely repulsive (σ=1.0, ε=1.0) and (b) attractive (σ=1.5, ε=1.0). Clearly, there is no discernable symmetry in the purely repulsive case (a). In Fig. 7(b), we show a structure formed with chain-chain attraction (σ=1.5, ε=1.0). In the case of the self-attractive chain, we find a structure like the folded toroid proposed by Hud[29], where chain strands are locally aligned in a nematic-type fashion. However, there is no radially symmetric order aligned with the packing axis.

In a related set of experiments, we considered the situation where the persistence length of the chain exceeds the height of the capsid by more than an order of magnitude. In this case, as was also seen by Kindt *et al.*, the chain packed in an inverse-spool fashion initially, wrapping in a helical layer around its outer layer of confinement. However, the inverse spool process was not maintained after this initial layer was formed, with the inner core filling in a disordered fashion. The inability to predict an ordered structure with the mechanical DNA model of our simulations is a significant result because it

indicates that the mechanical approach to modeling DNA is contradicted by, rather than supported by, the experimentally observed order (Cerritelli *et al.*). It is possible that the mechanical model for DNA chain rigidity in bacteriophage is too simplistic. For example the possibility of kinking or melting of short stretches of DNA is not accounted for in our simple mechanical model of DNA. However, these effects would likely lead to less, not more, spool-like order within the capsid.

Another possibility is that the case for interpreting the cryo-TEM images of Cerretelli *et al.* as representing spool-like order within phage has been overstated. A considerable number of other studies, including cryo-TEM, on a variety of bacteriophage have revealed considerably less order, showing for example that where striations exist, they are limited to patches, or domains, of the order of 20 nm[9-15]. Further, these individual domains are not oriented with respect to each other. This more modest ordering agrees more with that find in many of our simulations, see Fig. 7, a and b.

In the case of T7, a considerable amount of disorder could be masked in the procedure by which Cerritelli *et al.* generated images of radial symmetry. Many individual images were averaged into a single composite, which was then subsequently radially averaged. We were interested in determining if our disordered but tightly packed structures give rise to similar bands of radial order when subjected to a similar averaging process. X-ray diffraction studies have demonstrated that, at the high packing density found in a fully packaged phage, chains are on average locally close-packed. With respect to any long axis of an icosahedron, the average radial distance to a confining wall is of fairly constant radius, except near the poles. That is, along the majority of the axis, the chain experiences approximately cylindrical confinement. For a densely packaged

structure, we expect close-packed layers to be pressed against confining walls. Our question was whether, when viewed down the packing axis, circularly symmetric bands of fluctuating particle density might be discernable radiating inwards from the radius of the wall. Diffraction between wall-induced close-packed layers has been speculated to be the cause of the low wavelength ripple seen superposed on the 1/2.4 nm ring in X-ray diffraction data.

To obtain a quantitative measure of circularly symmetric ordering, the bead density of a given configuration from our simulation was projected onto the plane perpendicular to the packing axis. Projected densities obtained in this way have been presented as contour plots, or further averaged azimuthally to obtain a radial distribution function. By recreating contour plots from radial distribution functions and taking the negative of the resulting images, we obtain images that can be compared to the composite TEM images of Cerretelli.

We have found that the radial distribution functions obtained for our disordered packaged structures do not consistently show radially periodic density fluctuations (Fig. 8a). We do however consistently see at least a single faint band at the perimeter of the capsid. This is similar to what has been seen in reconstructions of the T4 virus[14]. Apparently the circular order induced by the wall decays very rapidly and is therefore unlikely to explain the extent of order indicated by the T7 images. In the case of the self-attractive chain, there is clearly more local nematic type packing, although with no alignment to the packing axis (see Fig. 7, b). This local order occasionally, but not consistently, leads to striking radial order.

We thus suspect that there is some genuine radial symmetry to the packaged T7 genome with respect to the packing axis, perhaps more than for other ds-DNA bacteriophage studied. T7 may be unique in possessing an internal structural protein connected to the capsid at the portal vertex and protruding inwards towards the center of the icosahedron. This hollow cylindrical protein within the T7 capsid was first identified by Serwer in 1976[30]. We have performed simulations to determine whether this internal structure might somehow function as an essential order-inducing feature of the T7 capsid. The "Serwer-core" was modeled as a hollow cylindrical tube through which the beads of the chain were extruded into the capsid. We found that, even when a purely repulsive chain was considered, the packaging order was noticeably increased compared to the core-less capsid. When intra-strand attractive interactions were allowed to act, the internal structural core became yet more effective at inducing order.

In Fig. 8, we show density contour plots of projected bead density for a number of different simulations we have performed. Increased radial order can clearly be seen in cases where packing was performed into capsids possessing the internal cylindrical core (the "Serwer core"). The Serwer core seems to function mainly by altering the position at which the chain emerges inside the capsid. The chain is extruded from the bottom of the cylindrical core, located near the center of the icosahedron. Because of limited flexibility, the most recently extruded portion of the chain is directed towards the lower vertex of the icosahedron. The potential of the confining walls in this lower region of the capsid is symmetric about the packing axis, causing the most recently extruded portion of chain to form a small-radius circular loop perpendicular to the packing axis. Subsequent loops are forced underneath previously-formed loops, which consequently rise upwards

along the packing axis. The loops are driven up also in search of a wider cross-sectional radius into which they can expand, available in the region away from the poles of the icosahedron. As this process continues, a uniaxial assembly of stacked hoops forms, with most recent added hoops located at the bottom. This assemblage stabilizes as it is skewered onto the projecting core, which then serves as a structural axis. The structures formed from this procedure are shown in Fig. 7, c and d. The shading gradient along the contour of the chain helps to reveal the history of packing, with recently packaged beads lighter, earlier packaged beads darker. The shading gradient verifies that the Serwer core forces more recently packaged strands of DNA to the bottom of the capsid, as described above.

Finally, we have defined a toroidal order parameter to quantify the degree to which the packaged DNA forms a toroid-like assemblage of stacked hoops aligned with the packing axis. To do this we considered individual arcs of fifteen consecutive beads along the chain. A unit normal vector for each arc was defined by normalizing the cross product of the two vectors connecting the middle bead to each of the end beads. For each arc along the chain contour (neighboring arcs overlap), we calculated the unit arc normal and projected its length along the packing axis. The order parameter was determined by dividing the sum of the magnitude of projected arc normals by the number of arcs. The order parameter approaches unity for a perfect stack of hoops, whereas a group of randomly oriented arc normals gives a value of 0.5. In Fig. 9, we give the order parameter versus loaded length for each of the major classes of simulations in this study – hollow capsid/repulsive chain, hollow capsid/attractive chain, Serwer core/repulsive chain, Serwer core/attractive chain. In the case of the hollow capsid filled by a purely

repulsive chain, the order parameter approaches 0.5 during packing, indicating that the loops of packaged genome are not aligned on average with any particular axis, as is evident also in Fig. 7, a. Packing carried out with a Serwer core present can be seen to give rise to increased toroidal order. The case of self-attractive/Serwer core gave the highest overall toroidal ordering, with order parameter values in excess of 0.9. For the hollow capsid/attractive chain simulations, in the early stages of packing, structures formed appeared toroid-like. However, as the packing proceeded, the toroid-like structures did not maintain alignment with the packing axis and ultimately often lost their toroid-like appearance. This behavior is visible in Fig. 9, where the corresponding value of the order parameter is initially as large as the values corresponding to packaging with a Serwer core.

1.D.3 **Energy, Force, and Pressure**

The energy increase associated with packaging of DNA results from bending of the semi-flexible DNA chain and repulsive interactions that arise during compression, both between the chain and the wall and between different segments of the chain. Knowing these energies as a function of loaded length allows us to calculate a quantitative measure of force resisting further packing. This force is obtained by taking the slope of the internal energy versus number of beads packaged averaged over a number of identically parameterized runs. Note that this force ignores entropic contributions. In averaging energy and force for a group of runs, the range of loaded length considered is determined by the run with the shortest span of total loaded length.

In Fig. 10, the resistive force calculated from average energy difference is plotted versus loaded length of viral genome. The total force versus length packed, as well as the decomposition of the energies and forces into bending and repulsive components, are given for chains with four different bending stiffnesses, k=0,750,1500, and 2250. There are a number of interesting observations to be made of the calculated resistive force. Early in the packaging process, the force is dominated by resistance to bending, which remains fairly constant. The force arising from repulsion is negligible until the later stages of packing, where it increases drastically. These findings are in agreement with earlier simulation studies. The bending and repulsive forces appear to be equal at the termination of packing for k = 750, but not for the less flexible chains.

The idea of non-equilibrium structure may be essential in explaining the next observation, that *in no case does the resistive force increase to a value equal to the packing force*. The simplest explanation for this is that, as the chain is packaged and the conditions inside the capsid become crowded, the chain becomes kinetically frozen in a meta-stable non-equilibrium state. (The neglect of the entropic contribution to the resistive force may also be a factor. However, it has been argued that the entropic contribution is negligible as compared to the bending and repulsive components in genome packing. We have no reason to believe that the present situation is any exception.) Recall that we cannot distinguish for certain between paused and completed packing, i.e. whether packing has truly stopped, or has merely temporarily stalled. By waiting long enough, we find that the chain can eventually relax, so that the packing undergoes a subsequent step-like increase. Because the simulations may have been terminated before packing was absolutely completed, it may be tempting to conclude that

we have ignored some final increment of crucial resistive force. However, evidence arising from our work suggests that the final packaged structure is inherently non-equilibrium or 'glassy'. Termination of packing is apparently not caused by a static force balance, but rather by some combination of resistive force and restricted mobility of the packaged chain. Packing does not occur the same way every time and the dynamics of chain relaxation largely determine chain packing, with energy considerations relegated to secondary importance. In this light, we are not surprised to find that the final resistive force has not reached the same magnitude as the packing force. We propose that *there is no one final packaged structure, pre-determined by orderly packing and energy minimization.* Again, we can point to experimental results to support our ideas in. Smith *et al.* found that the ϕ29 motor can package more DNA than is required by the full ϕ29 genome. Considering the intricate crafting of natural systems, this cushion should not be assumed to be unimportant. This seemingly "excess" packing capacity might not be excess at all, but rather necessary for packing structures that pack to different degrees.

The stochastic nature of packing is further revealed by examining the evolution of bending energy for individual runs. In Fig. 11, we follow bending energy as a function of loaded length for chains with bending modulus k=750. Five identically parameterized runs are shown, each given a unique symbol (squares, circles, down-triangles, left-triangles, pentagon, diamond) and differing only in the value of random seed used. Each data point represents the average bending energy of all chain configurations associated with an incremental loading of 25 nm. Thus, a difference in average bending energy between different runs indicates a difference in average structure, not attributable to

thermal fluctuation. Bending energy is clearly not a single valued function of loaded length.

For each run, the loaded length at which the packaging stalls varies. Horizontal dotted lines point to the internal bending energy at the termination of each of the individual runs; the final configurations do not correspond to any particular loaded length or internal bending energy. The final loaded length appears to be unrelated to bending energy, supporting the non-equilibrium notion of packaged structures. We have also calculated the resistive force directly, by measuring the average force exerted upwards on the bead in the neck. The calculation involves coordinates extracted periodically throughout the course of a packing simulation. Each of these coordinates become the starting configurations for a new set of simulations. The motor force was turned off and the bead corresponding to the neck position was held in place. During each simulation, the beads below in the capsid undergo Brownian motion and exert a force on the bead in the neck, which is measured and averaged over a large number of time steps. We have found that the force attains a constant, pseudo-equilibrium value within reasonable amounts of processing time (see Fig. 12, a).

We have presented numerous pieces of evidence that packing is not driven exclusively by an equilibrium buildup of resistive force. It would be surprising, then, if our direct-force results indicated a smooth increase in internal resistive force, culminating in a final value countering the motor force. Indeed, we do not see such a well-behaved direct force measurements. In Fig. 12, b, we give a plot of direct resistive force versus loaded length for a single packing run, with $k = 750$. At lower loaded lengths, we find a very small and well-behaved monotonic increase in resistive force. As packing proceeds

we find that the direct measurement of resistive force yields erratic values. As new portions of chain are forced through the neck into the capsid, previously packaged strands of the chain near the neck region must locally rearrange in order to accommodate the newly packaged portion; the chain becomes crowded within the capsid and the steric resistance to packing becomes determined by random local motions, which are likely the cause of both the irregular packing and erratic force measurement.

Note that the direct resistive forces display enormous positive fluctuations, up to twice the magnitude of the motor force, prior to the final configuration. The fluctuations in force with very small change in loaded length fits very well with the idea that direct force measurement values are dictated only by local configurations in the neck region. As proof of this idea, we have calculated -- simultaneously with the direct force measurements -- the average pressure over the entire interior capsid walls. Our interpretation of the direct force measurement as a local/non-energetic effect and the energy difference force determination as a global energy effect is strongly supported by measurements of the internal pressure buildup within the capsid. The pressure is determined as the force per unit area exerted by the polymer beads on the capsid walls, averaged over the entire inner surface of the capsid. We expect the pressure to reflect the global energy state within the capsid, i.e. we expect that the pressure behavior will be a relatively well-behaved montonically increasing function of loaded length. In Figure 12, b, we plot on the same axis both the hold force measurements and the corresponding pressure measurements. While direct force values fluctuate wildly with loaded length, the internal pressure inside the capsid is fairly constant and well-behaved.

We find that when the final configuration of any packing run is used as a basis for a hold-force simulation, the resistive force measured is comparable to the motor force. The results of our direct force measurement thus seem to contradict the findings of Kindt *et al.*, both in that our force values become erratic and that they do in fact attain the magnitude of the motor force at the completion of packing. However, we believe the apparent discrepancies can be reconciled by the simple consideration that the packing simulations of Kindt *et al.* were performed with constant velocity conditions. For a constant velocity simulation, the irregular portion of the packing shown in Fig. 4 is inaccessible. Recall that in the irregular phase, pauses occur, for which we must wait for relaxation processes to occur. A constant velocity simulation instead stops at the first pause. Kindt never measures irregular or large forces essentially because his simulations proceeded to lower loaded lengths.

Despite the fact that our direct force measurement values attain the magnitude of the motor force at completion of packing, we do not believe that this either 1) contradicts our previous energy difference results showing that the internal force does not attain the magnitude of the packing force; or 2) supports the idea that the packing process is determined primarily by energetic considerations. We believe that the packing stalls due local steric blockage in the neck region, a kinetically-driven phenomena. The configuration is locally frozen and so with a motor force pushing against it, of course a reaction force equal and opposite to the motor force will arise. The direct resistive force that we measure for our final configurations is then simply a reaction force arising from the application of the packing force to a locally glass-like solid. Our energy difference

force determination, on the other hand, was based on an accounting of the energy of the entire packaged structure.

In Fig. 13, we see that there is no final or equilibrium pressure, as might be erroneously concluded from the direct force measurements. If the direct force measurements were truly showing equilibrium in the final packaged states, the pressure would not display this behavior.

Finally, we show in Fig. 14 the bending energy versus loaded length for the different classes of simulations in this study. It is clear that the presence of the Serwer core during packing leads ultimately to lower bending energy structures. This provides definitive proof that packaged structures formed in the absence of the Serwer core are not dictated by bending energy minimization. It is interesting to note further that the structures formed in the presence or absence of the Serwer core are not distinguishable in the early stages of packing by their bending energies. Thus, the Serwer core acts to guide the chain to a more ordered structure initially, which is essential in the subsequent formation of lower bending energy final structures.

1.D.4. Role of Electrostatics

In the present section we present a brief calculation of electrically charged DNA beads packaged into a smaller capsid and observe that the findings presented in this paper are qualitatively insensitive to electrostatics. The experiments of Smith *et al.* were done at 50mM NaCl and 5mM $MgCl_2$. The Debye screening length, determined predominantly by NaCl under these conditions, is approximately 1.5 nm. The bond

length of our coarse-grained dsDNA is 1.25 nm (bead diameter=2.5 nm; consecutive beads overlap to produce more "tubular" structure). There are approximately 8 phosphate groups per 1.25 nm along the contour of dsDNA. Thus, we assume that there are 8 phosphate groups per bead in our simulations. However, it is not physically realistic to give a full charge per phosphate group, as we know that counterion condensation significantly reduces the effective linear charge density of dsDNA. From previous work in our group(32,33), we know that only a small fraction, $0<\alpha< 0.2$, of phosphate groups remain un-neutralized by counterion (or salt) condensation. We have performed a series of three packing simulations (α=0.0, 0.10, 0.20; namely 0.0, 0.8, and 1.6 charges per bead, respectively) on a capsid with the center-to-vertex distance being 7.5 nm. We use Debye-Huckel interactions with a screening length of 1.5 nm in these simulations. The effect of electrostatics on the time-evolution of loaded length is given in Figure 15. The presence of charge manifests itself only in reducing the final degree of packing (no attractive bridging effects in the presence of Mg^{2+} are considered) as the repulsive energy reduces the final allowed packing density. However, the reduction in the eventual packing density is only modest for the experimentally relevant values of α. More importantly, the generic features of the packing pathway in our simulations (such as the presence of irregularities in the packing rate and packaged order) are independent of electrostatics. These generic features match well with the experimental features and we conclude that their existence is independent of electrostatics but crucially depends on the bending energy, excluded volume, and confinement, under the experimental conditions. The relatively minor role played by electrostatics on the packing process is consistent with the results presented in Figure 10, where the sign and strength of the Lennard-Jones

potential play minor roles. Additionally, we found in Figures 7 and 8 that consideration of a self-attracting chain (simulating the bridging effect of polyvalent cations) led to no qualitative effect on the degree of toroidal ordering of the packaged structure. Therefore, we conclude that, *regardless of the presence or absence of charge, packing proceeds in an irregular manner, leading to disordered packaged structures and pauses due to configurational rearrangement in the packing process.* Given the enormous computational expense of considering detailed electrostatics, we have instead directed our computational resources in this paper towards modeling more accurately the physical dimensions and packing rates in phage genome packaging. Charge may play an interesting and important role in other aspects of viral packaging, which we will explore in future work, but for the findings we present in the current paper, the bending energy, excluded volume, confinement alone lead to the chain configurations becoming non-equilibrium.

1.E Conclusion

We have simulated the packing process of a ds-DNA bacteriophage. The geometry of the confining vessel, the intrinsic bending resistance of the simulated chain, the diameter of the chain, and the applied motor force were all selected based upon physical relevance. In the analysis of simulated packing rate data we have noted a number of stochastic phenomena. The stochastic features do not seem to match with the idea of spool-like packing, yet are in general accord with experimental results of Smith *et al.* In addition, we have found that packaged structure does not evolve as predicted by

spool-like packing. Specifically, stacked hoops do not form and align with the packing axis. The presence of the internal Serwer core does guide the packaging process towards more spool-like ordered final structures. The Serwer core appears to be crucial in the initial stage of packing, selecting for more ordered structures that are indistinguishable in terms of bending energy from disordered arrangement of loops. Simulated TEM images of self-attractive chains packaged in the presence of a Serwer core display the experimentally observed radial ordering. However, the simulated TEM images have proven unreliable in confirming spool-like packing, as radially symmetric structures were also observed in packaged structures showing no discernable orientation with the packing axis. Our future work will explore the implications of packaged structure on the release of the viral genome and the role of electrostatics, specifically how various charge distributions on the interior capsid wall affect the packaging process.

Upon completion of our work, a simulation study closely related to the present work has been published [31].

1.F Acknowledgements

Acknowledgement is made to NIH Grant No. 1R01HG002776-01, NSF Grant No. DMR-0209256, and the MRSEC at the University of Massachusetts, Amherst.

1.G Figures

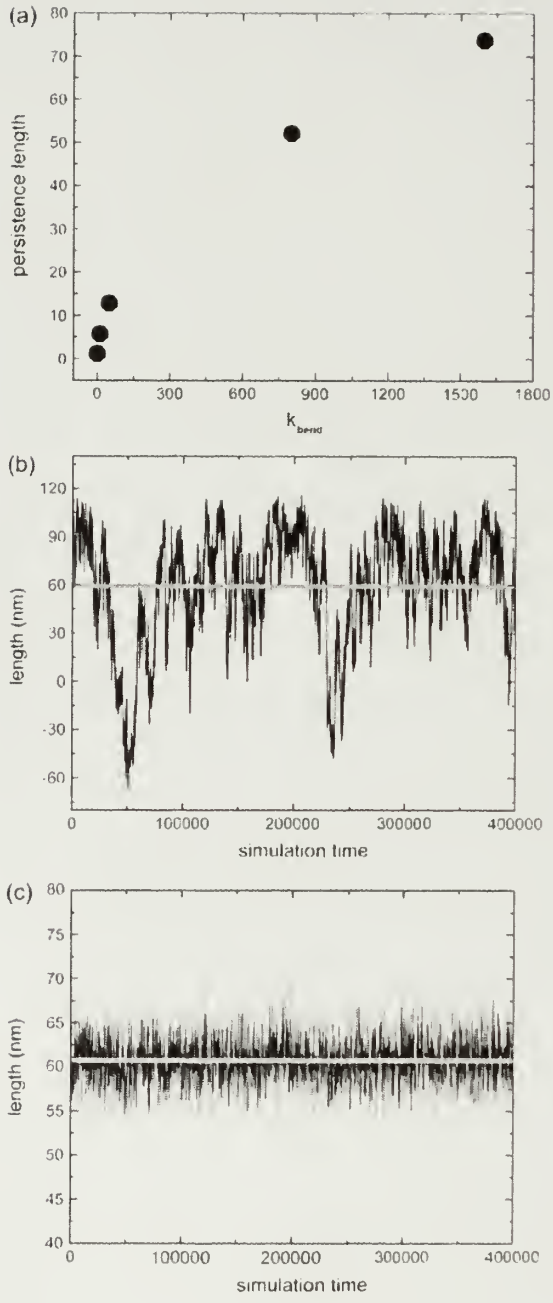

Figure 1-1 Persistence length determination.

(a) Predicted persistence lengths, based upon Eqs. 8 and 9, for a number of different values of k. (b) The projection of end-to-end distance onto the initial bond vector is shown in black for k = 750. The solid horizontal line is the average value of the projection of end-to-end distance onto the initial bond vector, which defines the persistence length. (c) The average of $\cos\theta$ is used at each time step to determine the persistence length, given as $1/(1 + \langle\cos\theta\rangle)$. The average over all time steps is shown as a solid horizontal line.

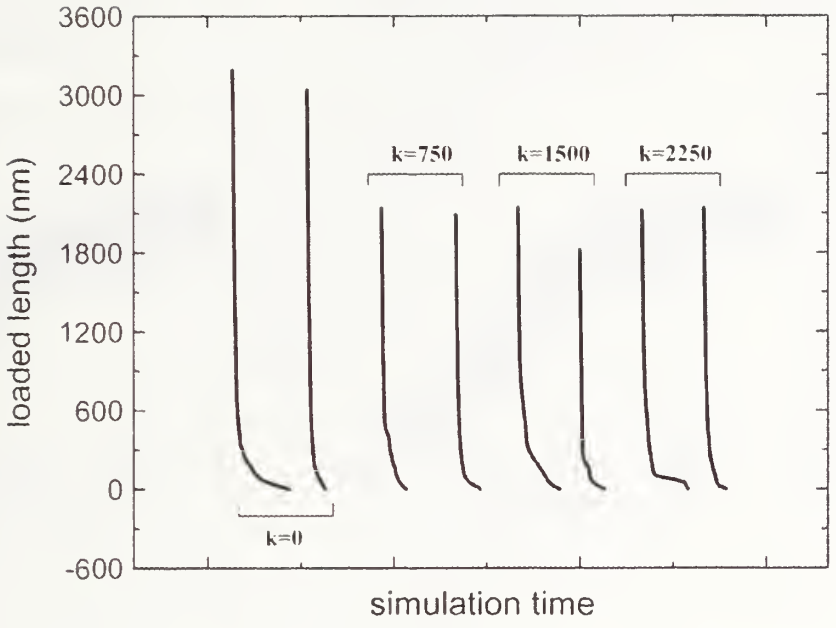

Figure 1-2 Simulated tether length versus time

The tether length for individual runs is plotted as a function of simulation time. The bending stiffness parameters for the runs (k = 0; k = 750; k = 1500; and k = 2250) are indicated in the figure.

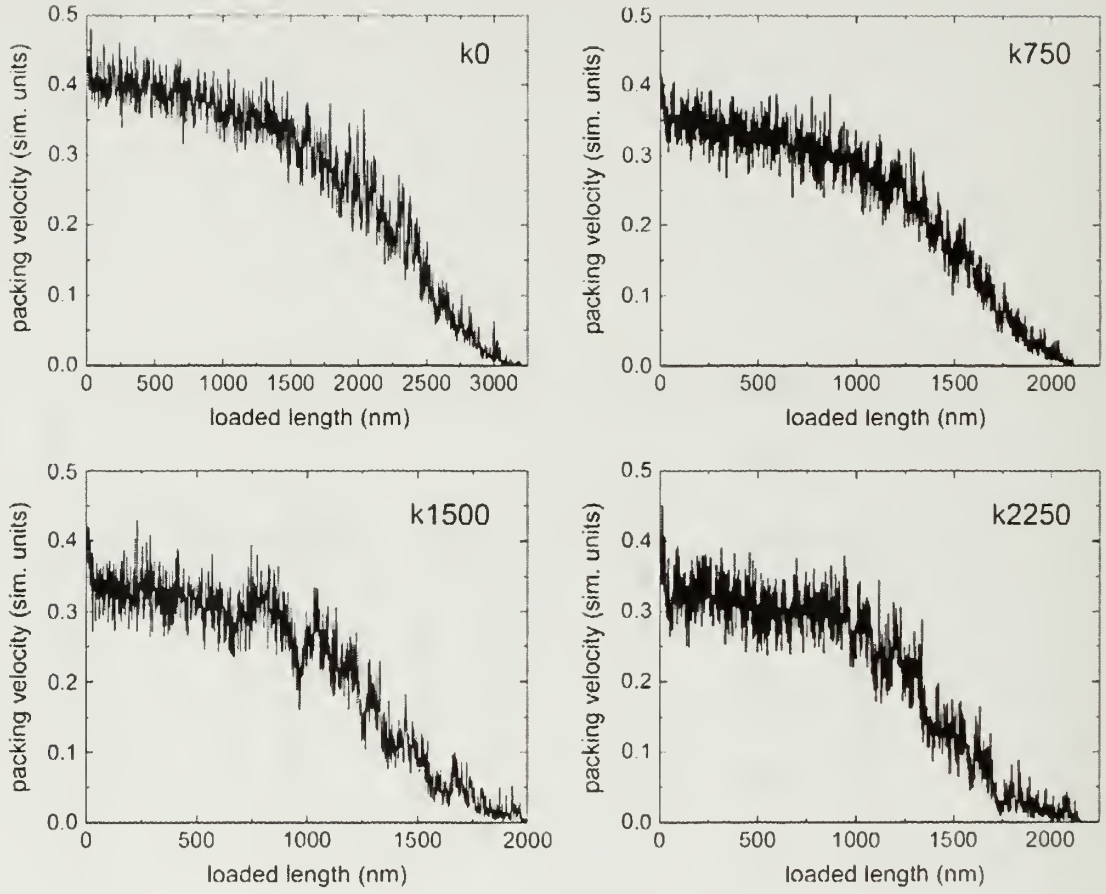

Figure 1-3 Average packing velocity versus loaded length

Averages of packing velocity versus loaded length for a number of identically parameterized runs is shown in black. In each case, the value of the stiffness parameter, k, is indicated in the upper right-hand portion of the figure.

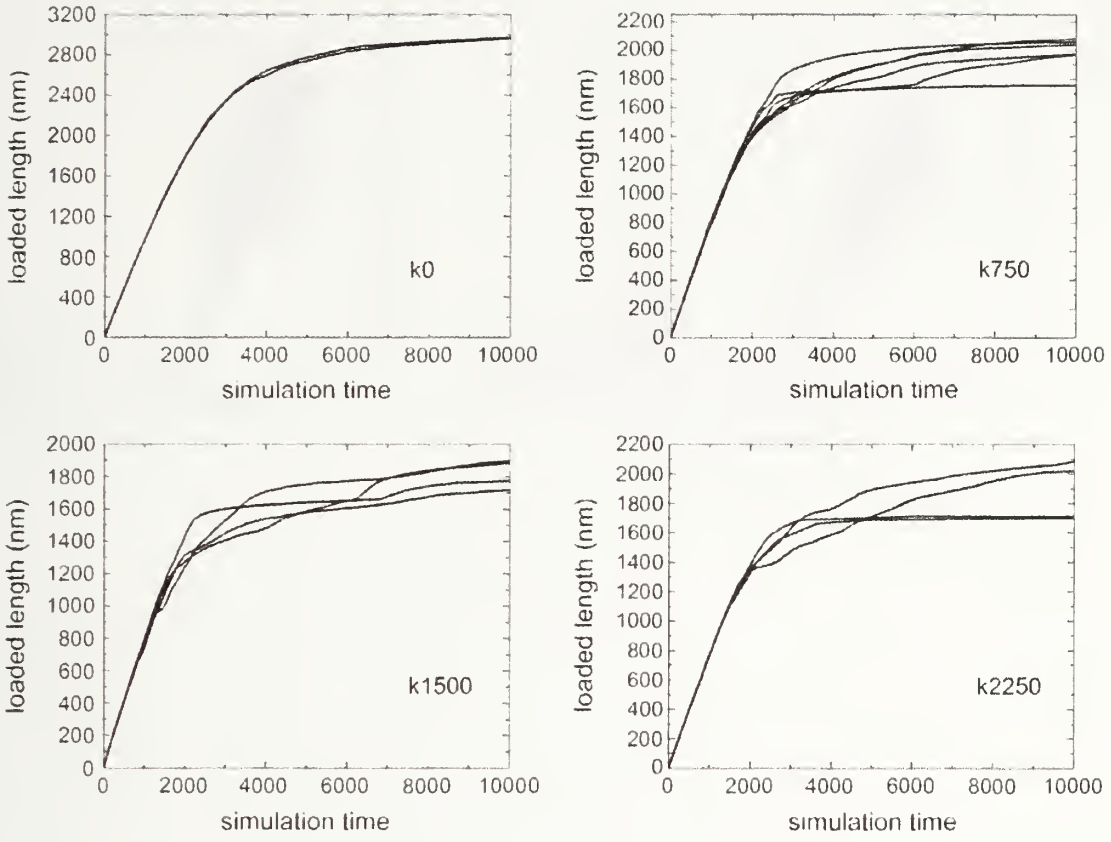

Figure 1-4 Loaded length versus simulation time

Results of loaded length versus simulation time for a number of individual runs are shown for each k used in our simulations. In the cases where the chain has finite bending stiffness: i.e., for $k > 0$, characteristics of stochastic packing can be seen.

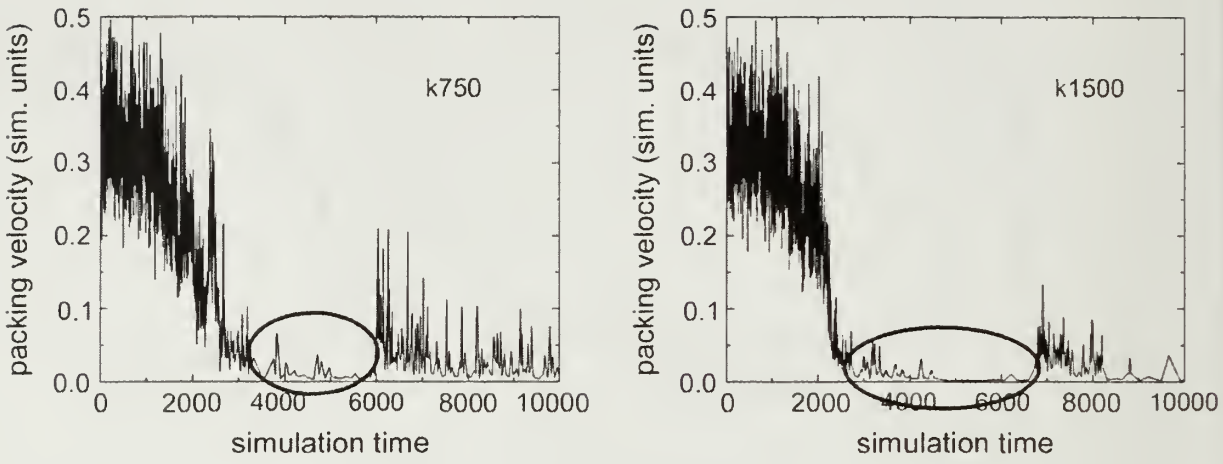

Figure 1-5 Packing velocity versus simulation time for individual runs

k = 750 in both cases. The presence of pauses, as indicated by circles, is clearly visible in individual runs.

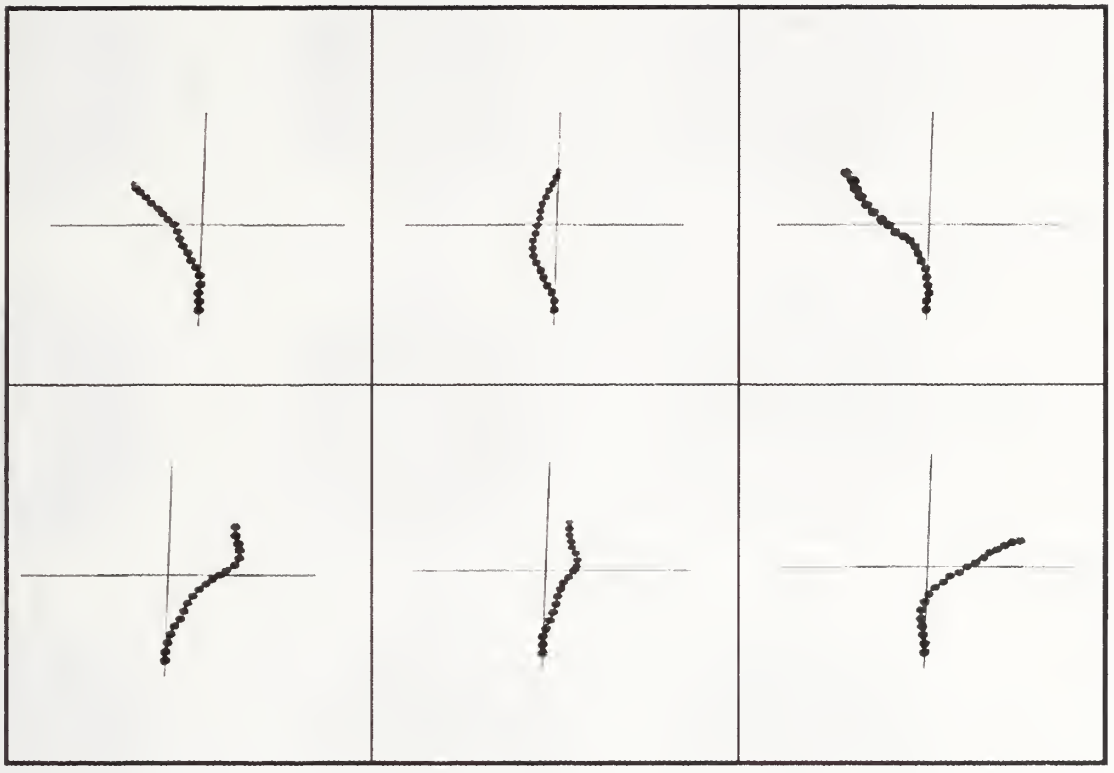

Figure 1-6 Snapshots of DNA

The images show the gyrations of a chain with stiffness constant k = 750. The contour length of the chain shown is equal to its persistence length. The figure demonstrates the flexibility a polymer exhibits, even when considering only a short segment of contour length, equal to the persistence length.

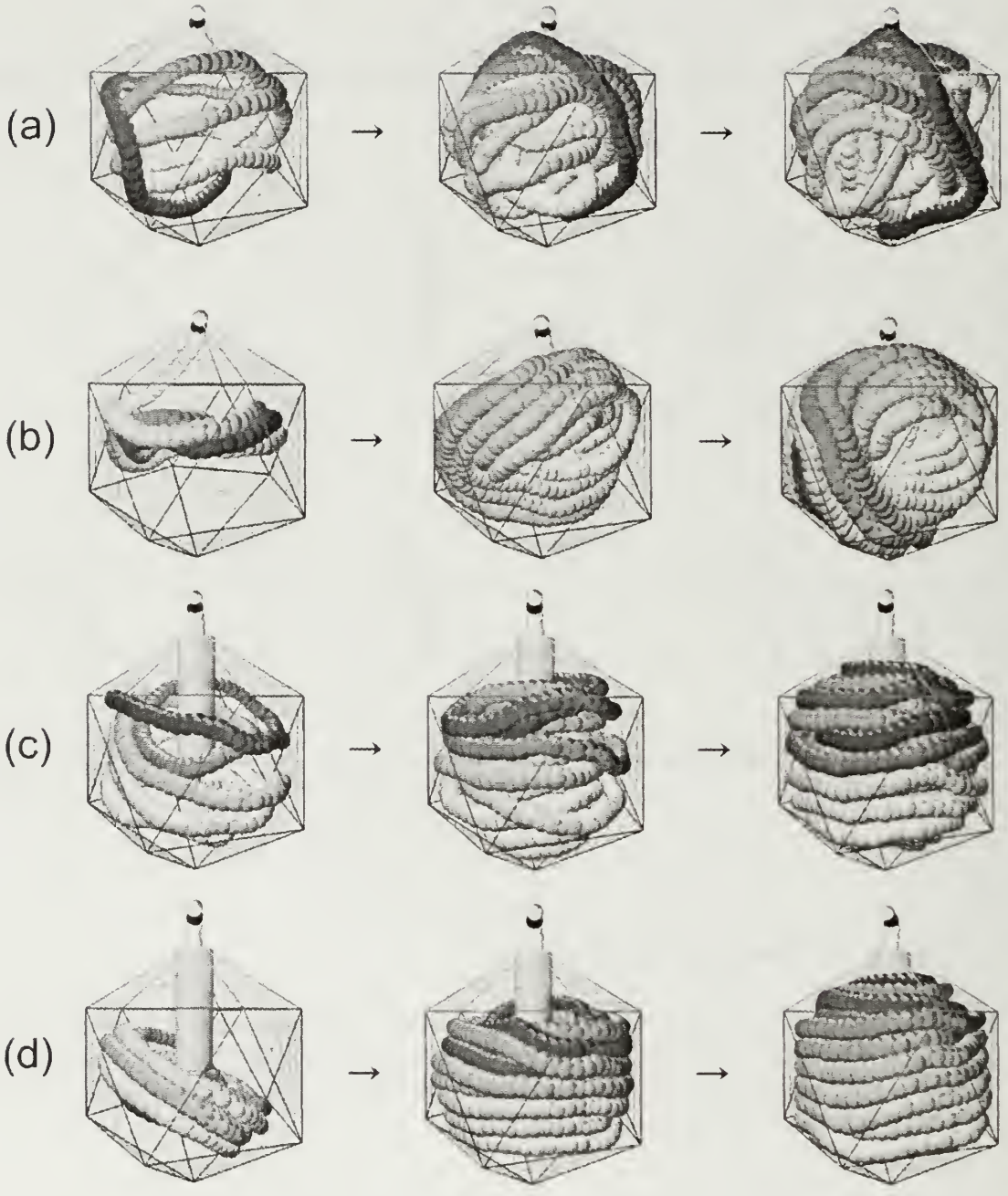

Figure 1-7 Evolution of packaged structure throughout the course of packing

Simulation snapshots demonstrate the progression of structures taken by the packaged genome. Each row corresponds to one of the packing conditions considered in this study. Hollow capsids (no Serwer core) are shown in panels a and b. Capsids with Serwer cores present are shown in panels c and d. (a and c) Interactions between DNA strands are purely repulsive; (b and d) DNA strands are self-attracting. In all cases, the bending stiffness is k = 750, corresponding to the bending stiffness of dsDNA in physiological conditions. The effect of the Serwer core on the order of the packaged structure is evident from these pictures.

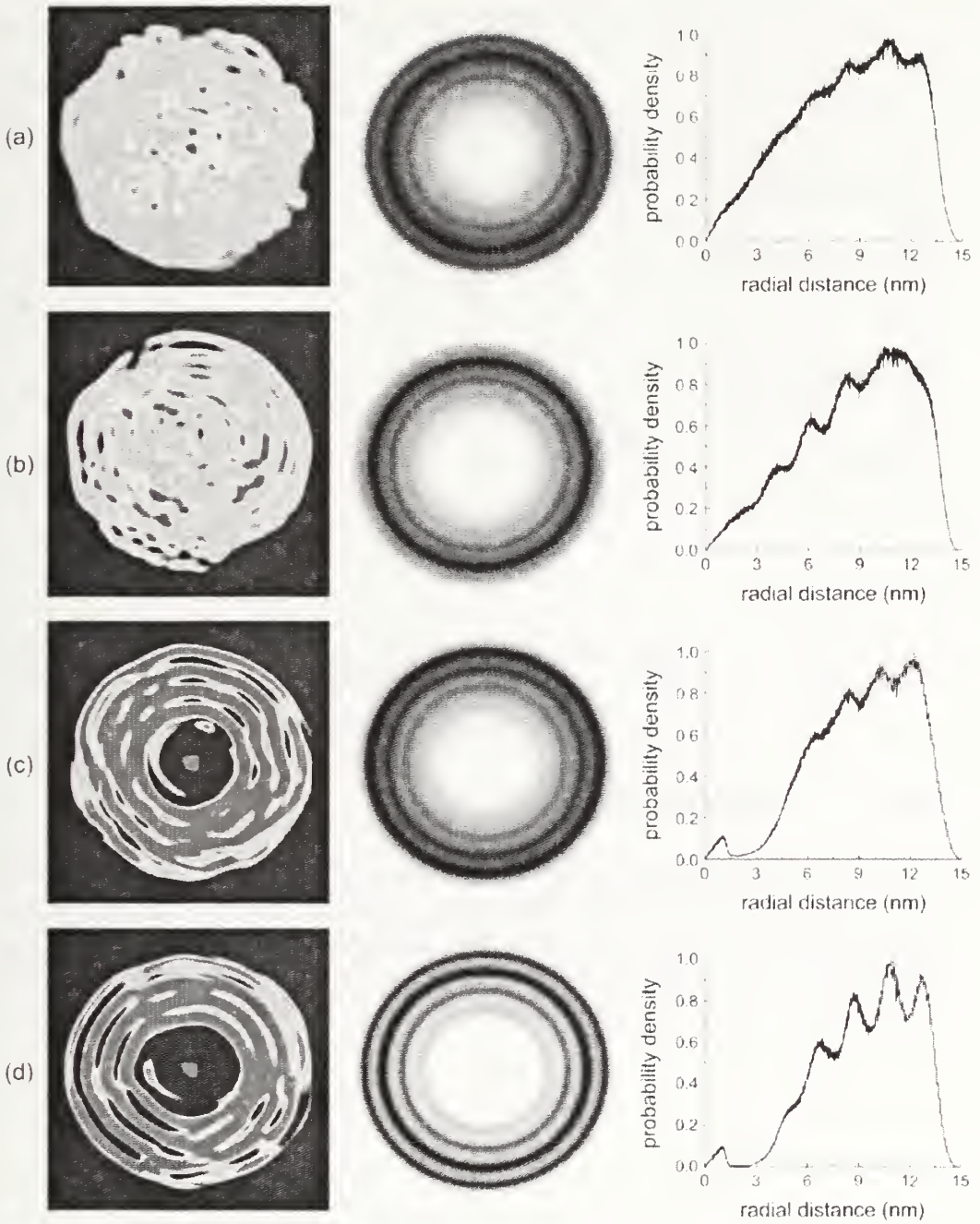

Figure 1-8 Probability density of packaged structures

Projected bead density on the xy plane (perpendicular to packing axis) is shown in the first column. Projected bead density is averaged over polar angle and presented as radial probability density in columns 2 and 3. (a–d) These correspond to the same packing conditions as in Fig. 7.

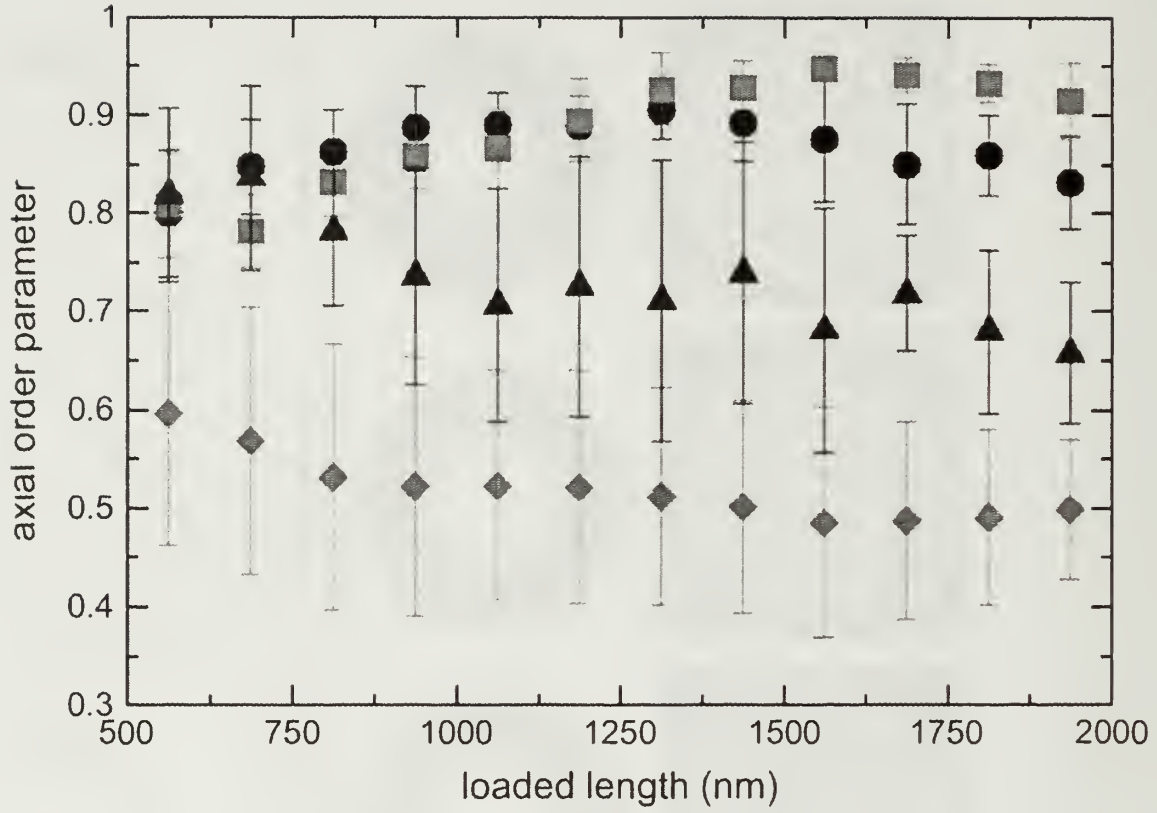

Figure 1-9 Toroidal order parameter

The ordering of packaged structures into hoops aligned with the packing z axis is given by the toroidal order parameter. The order parameter is followed throughout the course of packing for the major packing conditions considered in this work. (■, Serwer core/self-attractive chain; •, Serwer core/repulsive chain; ▲, no Serwer core/ self-attractive chain; and ◆, no Serwer core/repulsive chain.

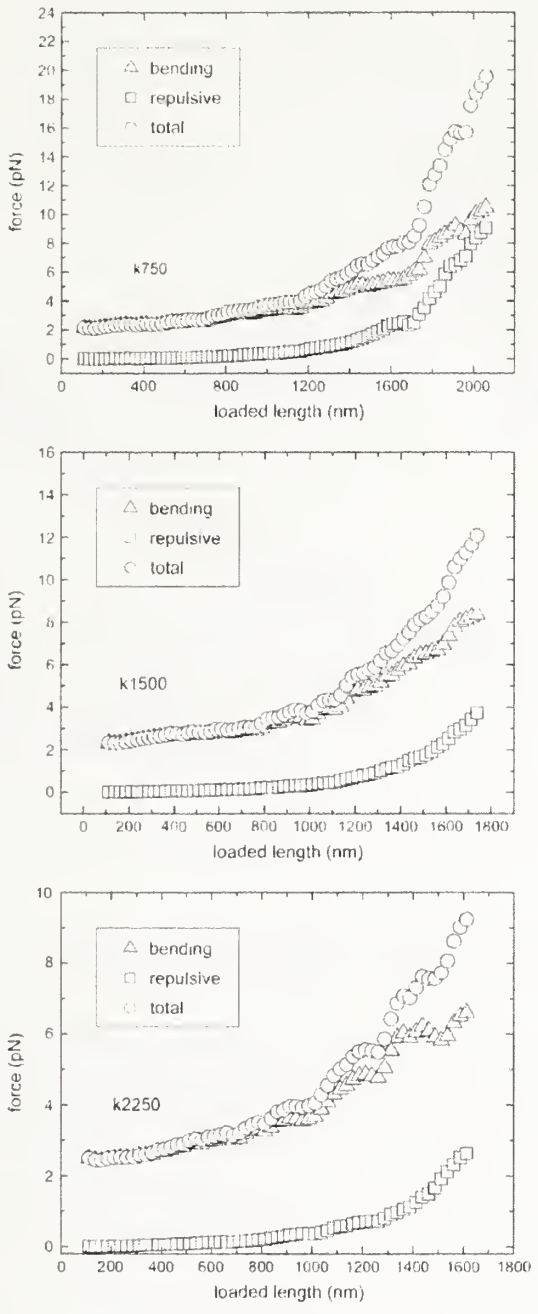

Figure 1-10 Force versus loaded length

Repulsive, bending, and total energy are shown as a function of loaded length. Chains of three stiffness parameters were used: (top) k = 750; (middle) k = 1500; (bottom) k = 2250. For each k considered, the force versus loaded length curves represent averages over a number of identically parameterized runs.

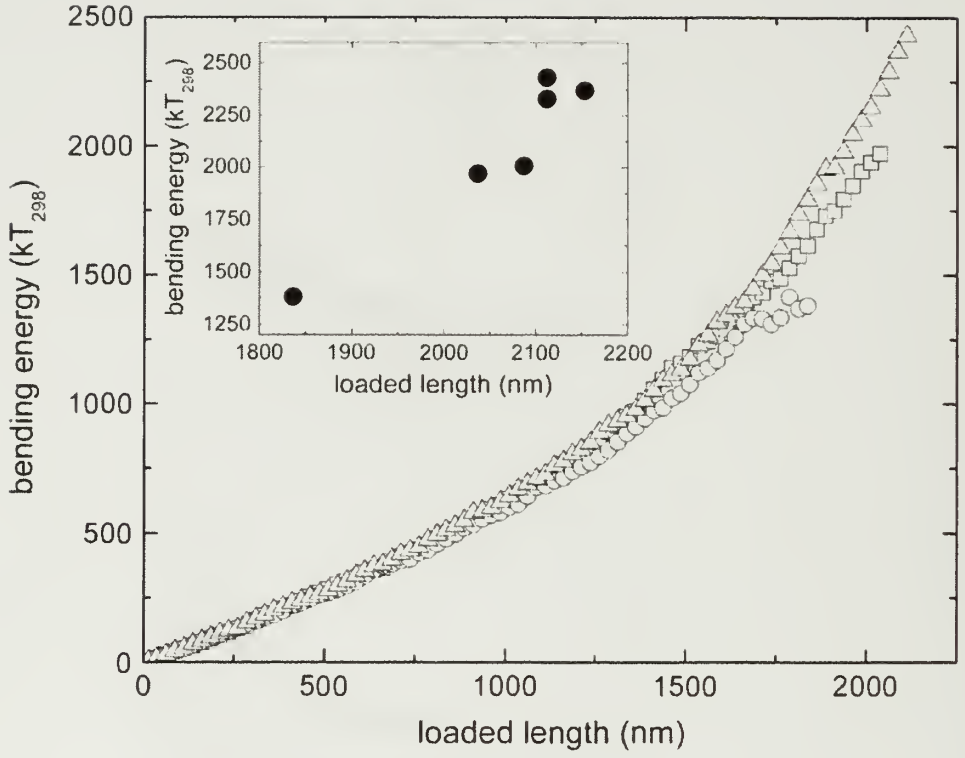

Figure 1-11 Bending energy versus packaged length

The bending energy of packaged structures is given as a function of loaded length. In all cases k = 750. Three identically parameterized runs are shown, with each run given a unique symbol. Each data point represents the average bending energy of the enormous number of configurations associated with an incremental loading of several tens of nanometers. For a given loaded length, average bending energy varies from run to run. The loaded length and associated bending energy at which the packaging stalls for all six runs we have performed is plotted in the inset figure. The final configurations do not correspond to any particular loaded length or internal bending energy; the termination of packing is evidently unrelated to the bending energy at which it occurs.

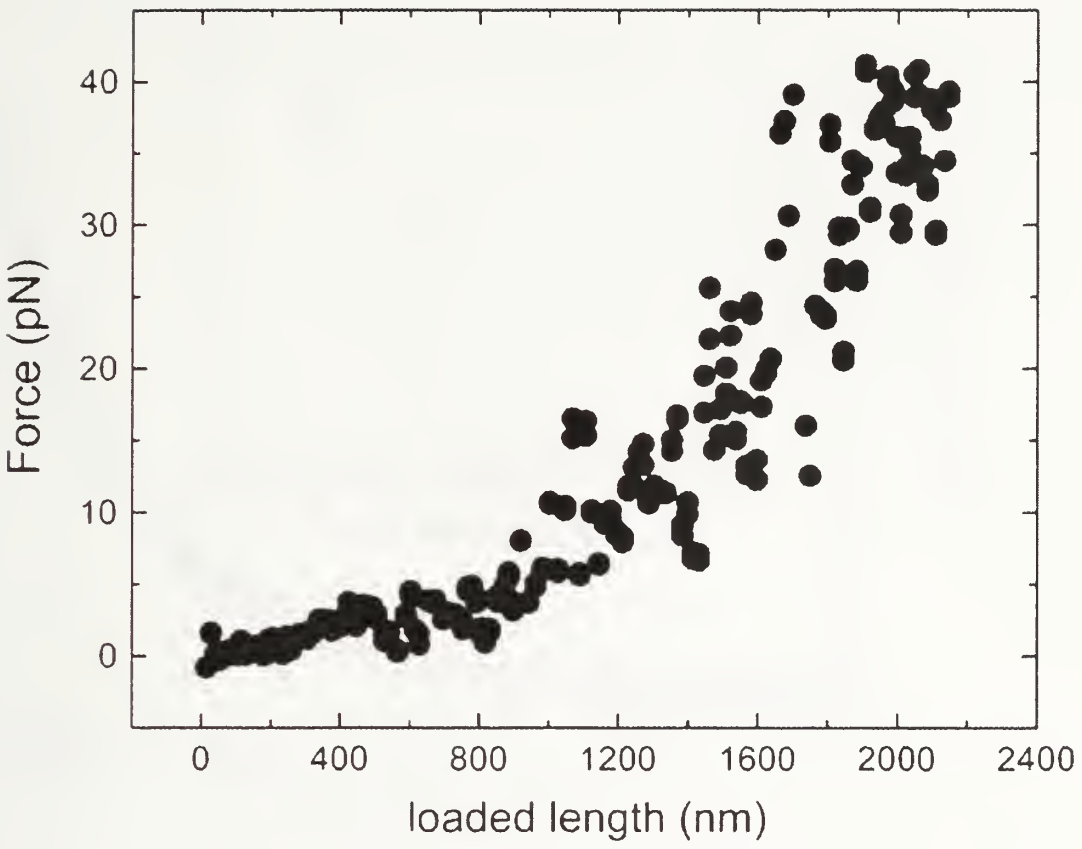

Figure 1-12 Direct force measurements of the bead in the neck

Each point represents the pseudoequilibrium force due to interactions with packaged beads. The noisiness of the force with incremental change in loaded length is not due to uncertainty in the measurement, but rather is a genuine demonstration of nonequilibrium behavior, particularly the local conformational fluctuations near the neck, of the packaged structures.

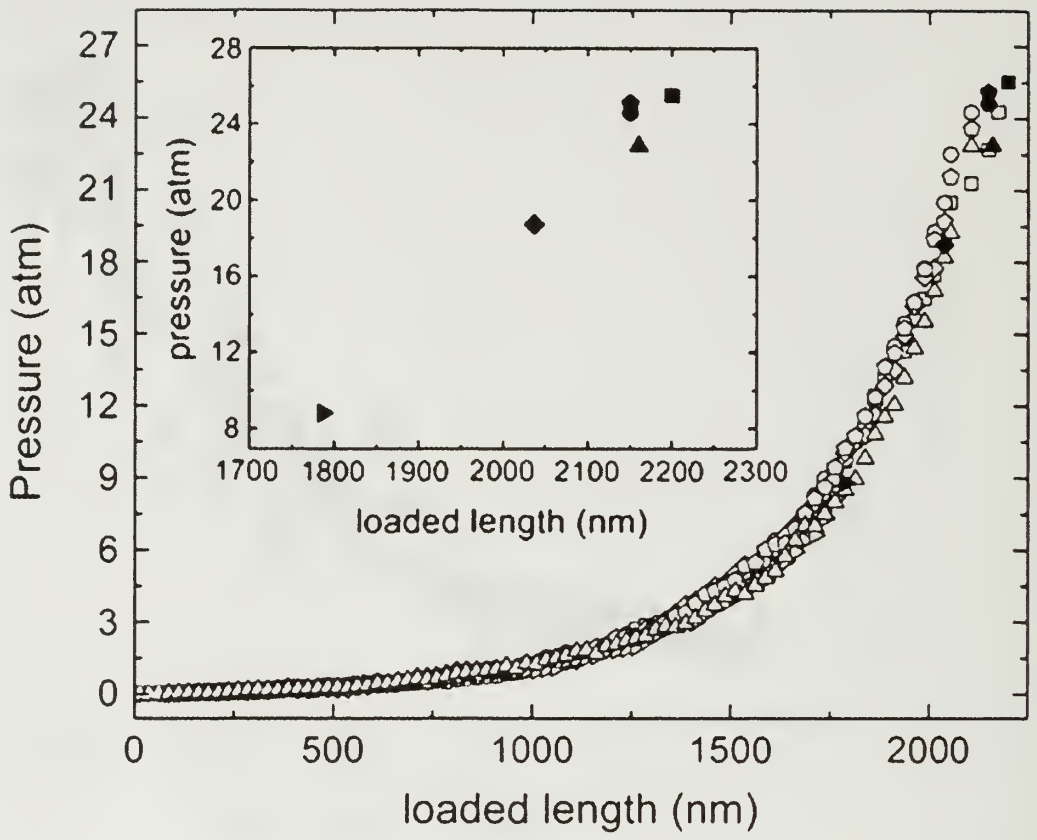

Figure 1-13 Pressure versus loaded length

The average measured pressures versus loaded lengths are shown for six identically parameterized runs. In all cases k = 750 and there is no Serwer core present. For a given loaded length, the pressure does not display the noisiness from run to run as seen for the direct measurements of resistive force (Fig. 12). The measured pressures, in tens of atmospheres, are in agreement with previous work. The terminal values of pressure and loaded length for each individual run is shown in the inset. As was the case with bending energy, the termination of packing does not correlate with the buildup of internal pressure to any specific terminal value.

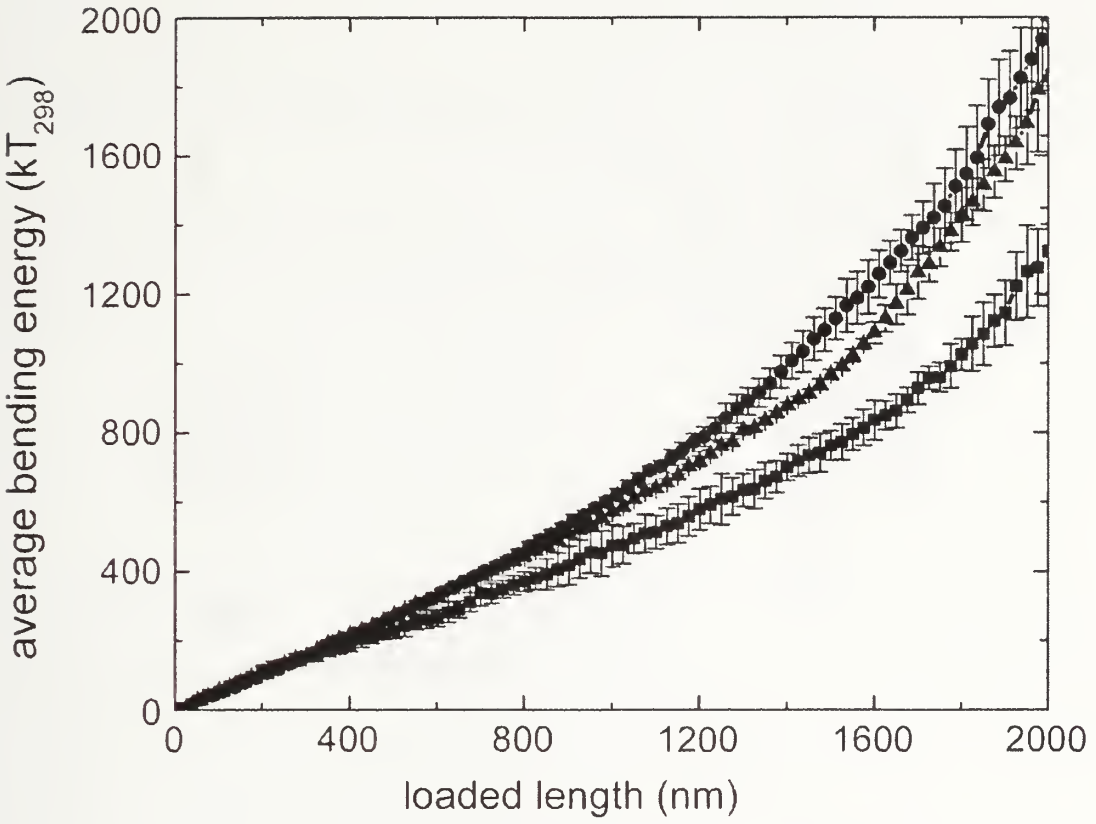

Figure 1-14 Average bending energy versus loaded length—effect of Serwer Core

No Serwer core/self-repulsive chain (•), Serwer core/self-repulsive chain (▲); and Serwer core capsid/attractive chain (■) are shown. In all cases, k = 750. The presence of the Serwer core guides the chain to lower energy configurations, particularly at higher degrees of packing. Two points are interesting: 1), bending energy in absence of Serwer core is clearly not minimized; and 2), the structures during early stages of packing are not distinguishable in terms of bending energy, although we know that the structures formed in the presence of the Serwer core are more ordered. Without an energy difference to select for the more ordered initial structures, a more disordered initial structure arises. At higher degrees of packing, the more ordered structures are clearly of lower bending energy. However, the initially unordered structures are unable to rearrange in the crowded capsid interior to form lower energy structures and so we find that the final configurations in the absence of the Serwer core remain stuck in a state of relatively high energy.

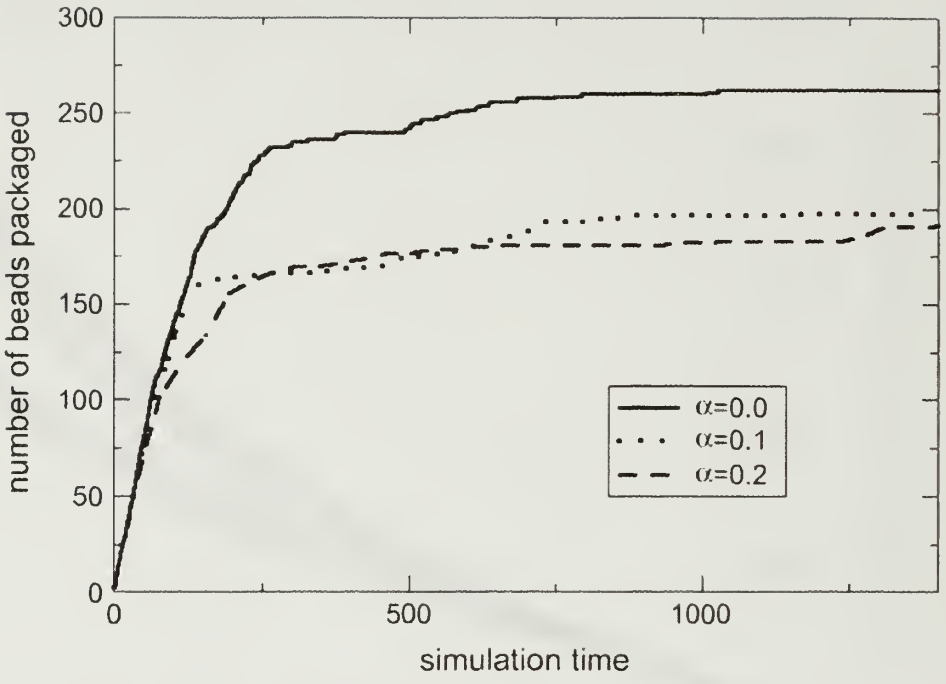

Figure 1-15 Influence of electrostatics on loaded length versus simulation time

A smaller scale system was used to allow for inclusion of electrostatics: the vertex-to-vertex size of the capsid was 15 nm, the bending stiffness constant, k, was 350, and the packing force was 38 pN. The values used for α, the fraction of bare charge, are given in the legend. Loaded length is given per bead (bond length = 1.25 nm). The units of simulation time for this reduced scale system do not match those for other simulations presented in this article. Pauses, the fingerprint of irregular packing, are clearly present in each case. The only feature distinguishing the different runs is the final degree of packing, which for the physically relevant values of α shown, was reduced, likely a consequence of the introduction of additional repulsive energy within the capsid.

1.H References

1. Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson, *Molecular Biology of the Cell*. 4th ed.; Garland Science: New York, 2002.

2. david m. knipe, p. m. h., *fundamental virology*. 4 ed.; lippincott williams and wilkins: 2001; p 1395.

3. Hud, N. V.; Downing, K. H., Cryoelectron microscopy of lambda phage DNA condensates in vitreous ice: The fine structure of DNA toroids. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **2001,** 98, (26), 14925-14930.

4. Hud, N. V.; Downing, K. H.; Balhorn, R., A Constant Radius Of Curvature Model For The Organization Of Dna In Toroidal Condensates. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **1995,** 92, (8), 3581-3585.

5. Bloomfield, V. A., DNA condensation. *Current Opinion In Structural Biology* **1996,** 6, (3), 334-341.

6. Earnshaw, W.; Casjens, S.; Harrison, S. C., Assembly Of Head Of Bacteriophage P22 - X-Ray-Diffraction From Heads, Proheads And Related Structures. *Journal Of Molecular Biology* **1976,** 104, (2), 387-410.

7. Earnshaw, W. C.; King, J.; Harrison, S. C.; Eiserling, F. A., Structural Organization Of Dna Packaged Within Heads Of T4 Wild-Type, Isometric And Giant Bacteriophages. *Cell* **1978,** 14, (3), 559-568.

8. Cerritelli, M. E.; Cheng, N. Q.; Rosenberg, A. H.; McPherson, C. E.; Booy, F. P.; Steven, A. C., Encapsidated conformation of bacteriophage T7 DNA. *Cell* **1997,** 91, (2), 271-280.

9. Booy, F. P.; Newcomb, W. W.; Trus, B. L.; Brown, J. C.; Baker, T. S.; Steven, A. C., Liquid-Crystalline, Phage-Like Packing Of Encapsidated Dna In Herpes-Simplex Virus. *Cell* **1991,** 64, (5), 1007-1015.

10. Richards, K. E.; Williams, R. C.; Calendar, R., Mode Of Dna Packing Within Bacteriophage Heads. *Journal Of Molecular Biology* **1973,** 78, (2), 255-&.

11. Lepault, J.; Dubochet, J.; Baschong, W.; Kellenberger, E., Organization Of Double-Stranded Dna In Bacteriophages - A Study By Cryoelectron Microscopy Of Vitrified Samples. *Embo Journal* **1987,** 6, (5), 1507-1512.

12. Serwer, P.; Hayes, S. J.; Watson, R. H., Conformation Of Dna Packaged In Bacteriophage-T7 - Analysis By Use Of Ultraviolet Light-Induced Dna-Capsid Cross-Linking. *Journal Of Molecular Biology* **1992,** 223, (4), 999-1011.

13. Aubrey, K. L.; Casjens, S. R.; Thomas, G. J., Studies Of Virus Structure By Laser Raman-Spectroscopy .38. Secondary Structure And Interactions Of The Packaged Dsdna Genome Of Bacteriophage-P22 Investigated By Raman Difference Spectroscopy. *Biochemistry* **1992,** 31, (47), 11835-11842.

14. Olson, N. H.; Gingery, M.; Eiserling, F. A.; Baker, T. S., The structure of isometric capsids of bacteriophage T4. *Virology* **2001,** 279, (2), 385-391.

15. Poy, S., *Molecular Biology of Bacteriophage T4* American Society for Microbiology: Washington, DC, 1994.

16. Kindt, J.; Tzlil, S.; Ben-Shaul, A.; Gelbart, W. M., DNA packaging and ejection forces in bacteriophage. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **2001,** 98, (24), 13671-13674.

17. Purohit, P. K.; Kondev, J.; Phillips, R., Mechanics of DNA packaging in viruses. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **2003,** 100, (6), 3173-3178.

18. Black, L. W., Dna Packaging In Dsdna Bacteriophages. *Annual Review Of Microbiology* **1989,** 43, 267-292.

19. Muthukumar, M., Polymer translocation through a hole. *Journal Of Chemical Physics* **1999,** 111, (22), 10371-10374.

20. Kong, C. Y.; Muthukumar, M., Polymer translocation through a nanopore. II. Excluded volume effect. *Journal Of Chemical Physics* **2004,** 120, (7), 3460-3466.

21. Odijk, T., Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. *Biophysical Journal* **1998,** 75, (3), 1223-1227.

22. Tzlil, S.; Kindt, J. T.; Gelbart, W. M.; Ben-Shaul, A., Forces and pressures in DNA packaging and release from viral capsids. *Biophysical Journal* **2003,** 84, (3), 1616-1627.

23. Arsuaga, J.; Tan, R. K. Z.; Vazquez, M.; Sumners, D. W.; Harvey, S. C., Investigation of viral DNA packaging using molecular mechanics models. *Biophysical Chemistry* **2002,** 101, 475-484.

24. Purohit, P. K.; Inamdar, M. M.; Grayson, P. D.; Squires, T. M.; Kondev, J.; Phillips, R., Forces During Bacteriophage DNA Packaging and Ejection. *Biophysical Journal* **2005,** 88, 851-866.

25. Smith, D. E.; Tans, S. J.; Smith, S. B.; Grimes, S.; Anderson, D. L.; Bustamante, C., The bacteriophage phi 29 portal motor can package DNA against a large internal force. *Nature* **2001,** 413, (6857), 748-752.

26. Tildesley, M. P. A. a. D. J., *Computer simulation of liquids*. Clarendon Press: oxford, 1987.

27. Yamakawa, H., *Modern theory of polymer solutions*. Harper & Row: New York, 1971.

28. Stevens, M. J.; Kremer, K., The Nature Of Flexible Linear Polyelectrolytes In Salt-Free Solution - A Molecular-Dynamics Study. *Journal Of Chemical Physics* **1995,** 103, (4), 1669-1690.

29. Hud, N. V., Double-Stranded Dna Organization In Bacteriophage Heads - An Alternative Toroid-Based Model. *Biophysical Journal* **1995,** 69, (4), 1355-1362.

30. Serwer, P., Internal Proteins Of Bacteriophage-T7. *Journal Of Molecular Biology* **1976,** 107, (3), 271-291.

31. Spakowitz, A. J.; Wang, Z. G., DNA packaging in bacteriophage: Is twist important? *Biophysical Journal* **2005,** 88, (6), 3912-3923.

32. Liu, S., and M. Muthukumar. **2002**. Langevin dynamics simulation of counterion distribution around isolated flexible polyelectrolyte chains. *J. Chem. Phys.*, **2002**, 116:9975–9982.

33. Liu, S. Ghosh, K. and Muthukumar, M. Polyelectrolyte solutions with added salt: A simulation study. *Journal Of Chemical Physics*, **2003,** 119: 1813-1823.

CHAPTER 2

LANGEVIN DYNAMICS SIMULATIONS OF dsDNA TRNANSLOCATION

2.A **Abstract**

We have implemented a coarse-grained model to study voltage-driven as-DNA translocation through nanopores located in synthetic membranes. The simulated trajectory of the DNA through the nanopores was calculated using Langevin dynamics. We present the results based on more than 120 000 individual translocations. We are particularly interested in this work in probing the physical basis of various experimentally observed—yet poorly understood—phenomena. Notably, we observe in our simulations the formation of ds-DNA hairpins, widely suspected to be the basis for quantized blockage. We study the translocation time, a measurable quantity crucially important in polyelectrolyte characterization, as a function of hairpin vertex location along the polymer backbone, finding that this behavior can be tuned to some degree by simulation parameters. We also study the voltage dependence of the tendency of hairpins to serve as the initiators of translocation events. Surprisingly, we find that the resulting probability depends vitally upon whether the events counted are ultimately successful or not. Further details lead us to propose that failed attempts in experimental translocation studies may be more common—and deceptive—than is generally recognized. We find the time taken by successful single file translocations to be directly proportional to the ratio of chain length to the applied voltage. Finally, we address a common yet puzzling phenomenon in

translocation experiments: translocation events in which the current through the pore is highly, yet incompletely, blocked. We present the findings that offer a new explanation for such events.

2.B **Introduction**

The past decade has seen the inception and rapid growth of the field of polyelectrolyte translocation. Translocation experiments involve the electrically-driven translation of charged polymers through pores, whose typical diameters range from nanometers to tens of nanometers. The pores themselves are located in and provide a means of passage through otherwise impermeable membranes. The membranes divide the solution chamber into a donor (cis) compartment and a recipient (trans) compartment, into which polymers are delivered by translocation. A potential difference is applied across the membrane, leading to a large electric field within the pore, which provides the driving force for translocation. During translocation, the pathway of the current through the pore becomes obstructed in a time dependent manner, providing a rich electrical signature of the translocating molecule. Previous work has reported the ability to distinguish between polymers of different molecular weight and monomer sequence[1-7]. Our goal is to uncover the detailed physical behavior of ds-DNA during translocation and how it gives rise to certain experimentally observable phenomena.

The technique of polyelectrolyte translocation was pioneered[1, 8, 9] using the alpha hemolysin toxin (αHL), a proteinaceous pore of the bacterium Staphylococcus aureus. The toxin is composed of seven identical protein monomers that self-assemble in

phospho-lipid membranes to form a heptameric mushroom-shaped structure, featuring a 1.4 nm diameter cylindrical stem. Because this diameter is too narrow to allow for passage of ds-DNA, analyte detection in αHL translocation experiments has been restricted to ss-RNA, ss-DNA, or more recently synthetic polyelectrolytes[10].

More recently, there have been tremendous advances in the ability to sculpt nano-pores in synthetic membranes[11-17]. Synthetic pores, being less vulnerable to extremes of voltage and pH, are more robust than their biological counterparts and translocation data is available over a much wider range of voltages. In addition, synthetic pores can be sculpted with slightly larger diameters (10-20 nm), allowing the passage of ds-DNA. Gaining the ability to translocate ds-DNA eliminated the burden of maintaining polynucleotides in their denatured state - as was required for single stranded RNA or DNA. However, an additional complication has been introduced: instead of a single characteristic drop in current associated with translocation, blockage events now gave rise to a range of quantized blockage values[16]. A common view is that ds-DNA may pass in either single file or folded configuration through the wider synthetic pores, with each strand contributing more or less independently to the current drop.

It has been reported experimentally that the fraction of single file translocation events increases with the potential difference used to drive the translocation[16]. This is of great practical significance because it suggests that at lower potential differences, the benefits of slower translocation – *i.e.* more detailed blockage signatures - may be offset by the increasingly complex signals associated with non-linear translocation. However, the idea that larger driving potentials would reduce the relative likelihood of hairpin translocation is counterintuitive when one considers that the energy barrier required to

bend ds-DNA into a hairpin configuration should be more, not less, easy to overcome at high voltages. We will examine this important issue closely in the results section.

The previous decade has seen remarkable advances in translocation technology, yet much of the observed phenomena are confusing and contradictory. As an example, the translocation time has been found to be proportional to N and $N^{1.27}$ in Ref. [16] and Ref 18, respectively, for similar experimental conditions. While there have been significant theoretical contributions towards understanding the fundamental physics of translocation[19-29], we feel that dynamics simulations of translocation, based on a simple physical model, can address some of the outstanding questions raised by the experimental observations.

We employ Langevin dynamics to explore the process of ds-DNA translocation. Langevin dynamics has an outstanding track record, having successfully contributed to our understanding of numerous phenomena in which polymer dynamics plays a central role. We will attempt to harness the physical detail offered by simulation to provide physical insight into some of the issues that have perplexed the translocation community.

2.C **Simulation Model and Technique**

Our model membrane is a 30 nm thick rigid and impermeable from which dynamic DNA beads are excluded, extending infinitely in the x and y directions and bounded at +/- z_o. The membrane is surrounded on either side by solvent, manifested by the stochastic term in the Langevin equation and the dielectric constant (ε=80) used in calculations of electric fields. Translocation of ds-DNA occurs through a hollow

cylindrical pore, in which the conducting solution is also present. Data have been collected for two different pore diameters: d=10,15 nm. The inner walls of the pore are lined with beads, each acting as a point source for purely repulsive Lennard-Jones interactions with dynamic DNA beads. By employing beads in this way, we leave open the possibility of exploring various charge patterns along the internal surface of the pore.

The ds-DNA is modeled following the traditional course-graining approach, in which the chemically-detailed double helix is replaced by a wormlike statistical chain of identical beads. The simulated polymer is parameterized by the bead diameter d, an effective 'bond' length, l_o, the number of beads N, and the effective charge density. The bead diameter, d, is a measure of the range of excluded volume interaction, which is determined the number of base pairs subsumed by each bead. d is taken to be 2.5 nm, the hydrated diameter of ds-DNA, and l_o is set slightly larger than d. The diameter of the resulting space-filling chain matches ds-DNA. The bending energy was parameterized in such a way as to give the chain a persistence length of ~60 nm in free solution, an appropriate value for ds-DNA. Details of this parameterization can be found in a previous paper[30].

The applied potential difference is modeled as a continuous linear decrease occurring entirely across the width of the membrane; therefore the electric field is constant within the pore and zero in either chamber. We have not included any hydrodynamic flow fields in our model. In a separate work, we have studied the effect of electro-osmotically induced flow, in which analyte molecules brought within a certain capture radius are drawn towards the pore[31]. To focus on translocation itself, we have

replaced this function of the electro-osmotic flow field by pre-positioning the chains in the spherical cis-side container described previously.

To explore the molecular weight dependent behavior of translocation, we have used three chain lengths, N=40,80,120 beads. While every effort has been made to select our parameters in correspondence with experimental conditions, the nature of course-graining is such that a strictly literal translation to physical units is not appropriate or useful. Throughout the remainder of the chapter, we will give results in terms of simulation units.

2.C.1 **Force and Energy Calculations**

DNA beads in our simulation are subject to interactions intended to capture the relevant physical forces involved in the translocation process:

2.C.1.a Excluded Volume Interaction

The excluded volume interaction is modeled by a Lennard-Jones potential, U_{LJ}:

$$U_{LJ} = \varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right] \qquad \textbf{(Eq. 2-1)}$$

where r is the center-of-mass distance between two beads, σ is the hard core diameter, and ε is the strength of the potential. The energy minimum is at $r=\sigma$. The Lennard-Jones potential is truncated at $r = \sigma$ and thus provides purely repulsive interactions between beads.

2.C.1.b Bond Stretching Interaction

The energy required to extend the bond between two consecutive beads along the chain is given by:

$$U_{stretch} = k_{stretch}(l-l_o)^2 \qquad \textbf{(Eq. 2-2)}$$

where $k_{stretch}$ is the stretch modulus and l_o is the natural bond length.

2.C.1.c Bending Interaction

The work required to introduce a bend of angle θ along the backbone of the polymer is given by the three-body interaction:

$$U_{bend} = k_{bend}(cos\theta - cos\theta_o)^2 \qquad \textbf{(Eq. 2-3)}$$

where θ is the bond angle formed by three consecutive beads, θ_o is the effective equilibrium bond angle, and k_{bend} is the bending modulus. There are two harmonic bending potentials in common use, the cos(θ) that we employ and a form harmonic with respect to θ. The potentials are empirical and - with appropriate parameterization to impart the chain with the correct persistence length - are both considered suitable potentials. We work with the cos(θ) form, as it insures that $U_{bend}(\theta) = U_{bend}(-\theta)$ for any choice of θ_o.

2.C.1.d Electrostatic Interaction

(i) Bead-bead electrostatic interaction.

Each of our ds-DNA beads subsumes approximately 16 negatively charged phosphate groups. However, it is not realistic to give a full charge per phosphate group, as counterion condensation significantly reduces the linear charge density of ds-DNA[32]. Only a small fraction (α), $0 < \alpha < 0.2$, of phosphate groups remain unneutralized by counterion (or salt) condensation[33, 34]. We have simply chosen to give each bead a single negative charge. (Recall, however, from section (*iii*) that bending energy of our chain is calibrated by a free solution simulation to a give our chain a persistence length appropriate for ds-DNA. Thus, most of the real electrostatic contribution to stiffness is parameterized into the bending energy term.) The electrical interaction between beads is calculated using the pairwise Debye-Hückel potential:

$$U_{DH} = \frac{\ell_B k_B T}{r} \exp[-\kappa r] \qquad \textbf{(Eq. 2-4)}$$

where l_B is the Bjerrum length, k_BT is the thermal energy, κ^{-1} is the Debye length and r is the separation distance between the centers of two beads. We have chosen to work with κ^{-1} = 1nm, l_B = 0.7nm, and T = 298K.

(ii) Bead-field electrostatic interaction.

The applied potential difference gives rise to an electric field

$$\overline{E} = -\nabla\Psi \qquad \textbf{(Eq. 2-5)}$$

where Ψ is the externally applied potential. We assume that the entire potential drop occurs evenly along the axis of the pore and thus the resulting electric field is constant within the pore and zero outside. As translocation proceeds, successive portions along the chain contour occupy the pore and become subject to the force field. We have carried out simulations using a range of ΔV, from 22.5 to 360 mV.

(iii) Repulsive interaction between the membrane/spherical wall and DNA.

This interaction is modeled by a Lennard-Jones potential (with force being purely repulsive) truncated at r=σ:

$$U_{LJ} = \varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right] \qquad \textbf{(Eq. 2-6)}$$

where in this case σ is the bead diameter and r is the distance from the wall to the bead center-of-mass. A bead thus feels no force until making contact with a wall, at which point the repulsive force pushes it back into solution.

2.C.2 **Dynamics**

Once the force on each bead is calculated, the simulation algorithm entails solving the Langevin equation for each bead of the polymer chain:

$$m_i \ddot{r}_{ij} = -\vec{\nabla}_j U_i + f_{i^* j}^{external} - \zeta r_{ij} + f_{ij}^{rand} \qquad \textbf{(Eq. 2-7)}$$

where i is the bead index and j is the component index. Thus, r_{ij} is the j^{th} component of the position vector of the i^{th} bead. Each bead is assigned a mass, m_i. The first term on the right hand side of Eq. 2.7 represents the sum of the configuration-dependent forces described in the previous section, minus the applied field. The applied field affects only to beads i^*, located in the pore, and is described by the second term. The effect of

solvent, acting primarily through frequent collision of the relatively tiny and fast moving solvent molecules with much larger polymer segments, is described by the third and fourth terms. The third term is the velocity-dependent viscous drag force, where ζ is the coefficient of viscous drag. The fourth term is the random force acting on each bead, which is assumed to obey a Gaussian distribution centered at zero. The square of the standard deviation of the distribution is given by the fluctuation-dissipation theorem, relating the magnitude of the deterministic frictional force to the random thermal bath force in thermal equilibrium,

$$\left\langle F_i(t) F_j(t') \right\rangle = \delta_{ij} 6 k_B T \zeta \delta(t - t') \qquad \textbf{(Eq. 2-8)}$$

where F(t) is the random force at time t, k_B is the Boltzmann constant, and ζ is the coefficient of viscous drag. We have chosen as characteristic units the thermal energy available at room temperature, $k_B T_{298}$=4.1 pN nm, the Bjerrum length in water at room temperature, l_B=0.7nm.

Subsequent configurations are determined through the solution of the equations of motion of the N particles in the system, which are integrated using the standard velocity verlet algorithm [35]

$$v_{ij}(t + \delta t) = v_{ij}(t) + \frac{1}{2}\delta t\left[\ddot{r}_{ij}(t) + \ddot{r}_{ij}(t + \delta t)\right] \qquad \textbf{(Eq. 2-9)}$$

$$r_{ij}(t + \delta t) = r_{ij}(t) + \delta t \dot{r}_{ij}(t) + \frac{1}{2}\delta t^2 \ddot{r}_{ij}(t) \qquad \textbf{(Eq. 2-10)}$$

where r^{ij} is as previously defined and v^{ij} is the j^{th} component of the velocity of the i^{th} bead.

2.C.3 **Prepositioning of DSDNA**

The translocation process can be divided into three parts: 1) the translational diffusion of the chain to the opening of the pore; 2) initiation of translocation, which can involve a number of unsuccessful translocation attempts, in which the chain is partially inserted into the pore, yet is unable to surmount the free energy barrier and is ultimately rejected back into the cis chameber; and 3) successful translocation, which is terminated when the entire chain resides in the receiving chamber. The sequence of steps is not necessarily one-directional: after a failed attempt, part of step 2, the chain may drift away from the opening of the pore, leading to a repeat of step 1. To avoid the computational expense of simulating the diffusion-based approach of the polymer to the pore opening (step 1), we preposition our ds-DNA molecule so that its center of mass is located one radius of gyration of the chain away from the pore opening. To prevent subsequent translational drift away from this position as the chain attempts to translocate, we confine the chain to a spherical container whose radius, R, is equal to the free solution radius of gyration of the chain. In this way, only activation and eventual translocation through the nano-pore are considered. A schematic of the simulation model is given in Figure 1.

2.C.4 **Batch Run Preparation**

Before translocation simulations are performed, we run a series of free solution simulations, one for each chain length considered, to determine the equilibrium radius of

gyration for each chain length, Rg_N. We then take a representative chain configuration from the free solution simulation and place it so that its center of mass is one radius of gyration away from the pore opening. In this way, for each chain length, we have established a template configuration. For a given N, the spherically-symmetric confining potential (of radius Rg_N) is also centered at the same position. For all individual simulation runs involving that specific N, we perform a pre-equilibration step with a unique random seed, in the absence of any applied potential difference, to obtain a unique initial chain configuration for each individual run. At that time, the potential field differences are applied and the translocation behavior observed. The simulation preparation procedure is outlined in Figure 2.

2.D **Results and Discussion**

2.D.1 **Hairpin Formation and Translocation Time**

Hairpin formation during translocation is a regular feature in our simulations. In Figure 3, simulation snapshots are presented, illustrating both single-file (3a) and hairpin (3b) events. In Figure 4, a hairpin-type initiation event is followed in time-elapse images from the perspective of the cis chamber. Figure 4a is a snapshot of the chain gyrating near the pore opening. Figure 4b) shows chain capture. As was common in our simulations, the captured portion of the chain is not a chain end. In Figure 4c), the movement of the lead bead towards the trans-end of the pore induces the formation of a hairpin and results in two parallel strands being threaded through the pore. These two

parallel strands corroborate the view that hairpin formation is basis for quantized current blockage. However, in our simulations we have not witnessed the formation of multiple hairpin configurations, hypothesized to give rise to experimentally observed higher order quantized blockades. Multiple hairpin formation would require that a translocating chain fold back on itself multiple times, resulting in three or more parallel strands of DNA spanning the length of the pore. Although hairpin formation occurred frequently in our simulations, our short chain lengths likely prevented multiple hairpin formation; hairpins, once formed, translocate too quickly for a free chain end to bend back and find the pore. It is also possible that our pore diameters were too small to allow for three or more stands to simultaneously fill the pore.

2.D.2 **Translocation Time**

We now address the issue of how hairpins affect an important measurable quantity, translocation time. Translocation time alone has been used in αHL experiments to distinguish various chain lengths and sequences from each other. Unfortunately, the presence of hairpins introduces another parameter that may complicate the efforts to characterize polyelectrolytes based on translocation time alone. We are interested in two issues: 1) how is the translocation time affected by the presence of a hairpin vertex at a specific location along the translocating chain; and 2) how likely is a hairpin to form at that given location along the chain?

The vertex can be identified as the foreward-most bead of the chain during translocation (e.g., Figure 3b). Hairpin formation leads to multiple strands

simultaneously spanning the pore (see Fig. 3b,4c). There are three possible effects on translocation time: 1) the more strands passing simultaneously through the pore, the more beads are subject to the electrical field and the larger is the electrostatic force pulling the chain through the pore; 2) a simple shape effect – the presence of a hairpin shortens the length of the object to be translocated; 3) friction – the more strands that are simultaneously present in the pore, the more crowded the beads are, thus increasing the number of collisions of the beads with themselves and the wall of the pore, leading to a frictional effect. Additionally, the polymer bead bombardment on the planar membrane will lead to a reaction force resisting translocation. Such a force would be twice as large for a translocating hairpin as for a single file chain. Nevertheless, such an effect would likely be minimal, as chain stiffness leads to extended chain configurations normal to the planar membranes, minimizing bombardment. Both effects 1 and 2 will tend to decrease the translocation time, while effect 3 will tend to increase translocation time. The frictional effect should be more pronounced as the pore diameter narrows. Because the effects are counterbalancing, it may be possible to "tune" the translocation velocity by favoring one of the effects over the others. The strategic placement of charges along the interior of the pore could also provide additional tuning of the translocation time.

In Figure 5, the average translocation time through the larger pore is plotted as a function of the location of the hairpin vertex along the chain. Each curve represents a different applied potential difference. The larger the applied potential difference, the shorter is the translocation time. As can be seen, the greater the proximity of the hairpin vertex to the center of the chain, the shorter the translocation time. The combination of geometric and pulling force effects dominate over any frictional effect, leading to

reduced translocation time for all vertex locations as compared to linear translocation times. However, we find that the frictional effect can indeed be enhanced by narrowing the pore diameter. In Figure 6, translocation time is compared for pores of two different diameters. For a given potential difference and vertex location, the translocation time is longer in the case of the smaller pore. In addition, the minimum in translocation time for centered-vertex hairpins seen for the larger pore disappears. Instead, the translocation time appears to be nearly independent of vertex location. Presumably, the friction effect is neutralizing the geometric shortening effect.

The translocation time can vary substantially as a function of vertex position and potential difference under certain conditions. If experiments are carried out under such conditions, the resulting translocation time data, even for a single chain length, will comprise a range of values, undoubtedly complicating the task of interpreting the data. However, for a given potential difference and pore diameter, the translocation time as a function of vertex position (Figure 5) is fairly simple. One could use a simulation run of the types shown in Figure 5 and Figure 6 as a "calibration" to assist in interpretation of experimental translocation time data. Figure 6b) suggests that the experimental conditions might even be "tuned" to remove the dependence of translocation time on the location of the hairpin vertex.

The second issue we raised, of the tendency of hairpins to form at various positions along the chain, cannot be addressed without refining our question further. Specifically, we must ask not only of the tendency of hairpins to form, but further whether those hairpins are ultimately successful and contribute in a meaningful way to

the statistics of translocation time, or are ultimately rejected from the pore, leading to complicated results. We address the issue of success versus failure in the next section.

2.D.3 Translocation Failure and its Consequences

Despite the trans-directed field within the pore, not every instance of chain penetration leads to successful translocation. In fact, most contact between polymer and pore in our simulations results in an *unsuccessful* attempt, in which the chain returns back to the cis chamber. This distinction between successful and unsuccessful translocation can be drawn clearly only in simulations; while failure undoubtedly occurs experimentally, the resulting current drop may be difficult to distinguish from that due to translocation. As we will demonstrate, analysis performed without a clear distinction between success and failure can lead to misleading results.

To quantify attempt failure, we define the success ratio, the ratio of the number of successful translocations to the number of total attempts. Success ratios for all potential differences and chain lengths considered in this study are plotted in Figure 7. The results are for the larger-diameter pore (d=15 nm). Notice that even at the highest potential difference considered (2.4 simulation units), at most only half of the translocation attempts were successful. The primary reason for failure is the well-known entropic barrier, in which the loss of configurational entropy associated with entrance into the pore gives rise to a free energy barrier. As seen in Figure 7, increasing the potential difference leads to a higher success ratio, as the resulting force fields help to drive the chain over the free energy barrier. The success ratio is also seen to increase for shorter chain lengths at

all potential differences. The shorter the chain, the higher is the fraction of free ends per chain, meaning that shorter chains will make a higher proportion of single file (chain end) attempts. Single file attempts do not suffer the energetic penalty hairpin formation and thus push success ratios higher.

2.D.4 **Translocation Type: Single File Versus Hairpin**

We now turn to an intriguing experimental claim, that the propensity for single file translocation increases with potential difference. Such a trend with potential difference is counterintuitive: larger potential differences should assist in surmounting the energetic barrier required to form hairpins in ds-DNA, making hairpin formation more, not less, likely at high potential differences. The authors proposed that long-ranged and powerful field lines emanate from the pore on the cis side, elongating and pre-aligning ds-DNA into a single-file configuration[16]. Because the electric field in our model is entirely restricted to the nano-pore, translocation trends at higher potential differences in our simulations certainly cannot be attributed to electric field induced pre-alignment of the chain within the cis chamber.

In Figure 8 and Figure 9, we plot f_s, the fraction of polymer/pore encounters that occur in single file fashion. f_s is plotted as a function of potential difference for all three chain lengths. However, in Figure 8, we exclude from consideration failed attempts, whereas in Figure 9, we include all attempts. As can be seen in Figure 8, the fraction of single file events increases with potential difference, even in the absence of a cis-side field; conversely, excluding failed attempts from the

calculation leads to the opposite trend. An alternative to field-induced pre-alignment hypothesis is apparent: the counterintuitive trend observed experimentally simply comes from the inability to properly distinguish between successful and failed attempts.

In Figure 10, histograms of vertex position are shown for both cases, i.e. all events considered (dashed curve) and successful translocation only considered (solid curve). The histograms are normalized by number of occurrences counted, so that the relative proportion of free end versus hairpin occurrences can be compared. The data correspond to the larger pore (d=15 nm) at the lowest potential difference considered in our study (ΔV=0.15). It is clear that hairpin events, in which the indices of the initiating beads are at interior positions, is much more represented when all events are counted and much less represented when consideration is restricted to successful translocation events. This discrepancy indicates that unsuccessful hairpin attempts are a significant contributor to the total number of events at this low potential difference.

We have already seen in Figure 7 that, at this potential difference, the number of failed attempts is much larger than the number of successful attempts. We can see explicitly in Figure 11 that, of this large fraction of failed attempts at low voltage, a majority are of the hairpin type. Figure 11 is a histogram of failed attempts, in which the abscissa is the maximum depth of penetration attained during the event and the ordinate is the number of events. Both single file (solid line) and hairpin (dashed line) failures are analyzed. It is evident that hairpin attempts are more common and penetrate further into the pore than single file attempts.

As potential difference is increased, failed attempts become many times less likely (Figure 7), and so the miscounting of failed events as actual translocations becomes

less significant. This is evident in Figure 10b, in which the distinction between all events and translocation events is relatively negligible.

Finally, in both Figure 8 and Figure 9, the fraction of single file translocation events is also seen to increase with decreasing chain length. For the shortest chain length considered, N=40, single file translocation is favored for all potential differences. Conversely, for the longest chain length considered, N=120, hairpin translocations are predominant at all potential differences. This trend is a consequence of the higher fraction of free ends larger for shorter chains, thus increasing the probability that a chain end finds the mouth of the pore. Experimental data are in agreement regarding this trend with chain length[16].

2.D.5 **Translocation Time and Scaling Behavior**

We now focus on the single file chain length dependence of translocation time. In Figure 12, the single file translocation time is plotted as a function of chain length. Data are included for five different potential differences (see figure caption). There is a linear relationship between translocation time and chain length. Inset in Figure 12 is a plot of translocation time, τ, as a function of the reduced variable, $N/\Delta V$. The data points are for N=80 and N=120, the larger two of our chain lengths, so as to ensure that the asymptotic regime of large N is considered. For sufficiently large chain lengths, the translocation time is directly proportional to $N/\Delta V$.

2.D.6 **Implications for αHL and Multiple Blockade Levels**

We now wish to analyze our results in a way meaningful to the case of αHL translocation. Focus is thus limited to the subset of our translocation results that occur only in single file fashion. We can then construct what is commonly referred to as an event diagram (Figure 13). Each point on the event diagram gives two pieces of information regarding an event, the maximum penetration depth attained during the event (ordinate) and the total duration of the event (abscissa). The data points at short times and shallow penetration depths in our simulations form a sloped "shoulder"-type region and represent failed attempts. This cluster of points is labeled in Figure 13 as partial insertions. Successful translocations are found in a separate cluster at longer times and at a penetration depth corresponding to the length of the pore. (y-axis noise, generated as a Gaussian deviate with a standard deviation of 0.08, was added to the successful translocation events for clarity – the penetration depth for all successful translocation events register a reading of exactly 5.0, the length of the pore). These event diagram features resemble those reported recently in our group[10]. The experimental event diagrams also display both a clear short-time shoulder and a high-blockage and long-time cluster. We propose that the physical bases for these experimentally obtained features are the same as in our simulations, i.e. partial insertions and successful translocations, respectively. There is an additional and conspicuous feature in our experimental event diagrams that is absent in our simulation work, that of "banding" at fairly partial but deep blockade level and persisting over many orders of magnitude in time. Similar clusters of events, characterized by deep but incomplete blockade levels, have been reported

frequently and appear to be a regular feature of translocation data using αHL[2, 9, 36-39]. However, there is controversy regarding their physical origin. We now suggest that the source of the banding, or deep but partial blockage, can be attributed to the charge distribution along the β-barrel of αHL[40]. The β-barrel is the cylindrical portion of αHL that spans the lipid membrane and as such is composed predominantly of uncharged amino acids. However, there is circular cross section at the trans-side of the cylinder along which is distributed a net charge of -7. The diameter of the β-barrel is only 1.5 nm, meaning that even single stranded nucleic acids must translate in close proximity to the barrel wall. The large negative charge at the trans-end of the barrel must present an additional energy barrier to negatively charged polyelectrolytes. The force from the applied potential would then push the chain into the pore until it reached the barrier of negative charges, effectively stalling the chain at a deep blockade level. In Figure 14, we show an event diagram obtained by making two changes to our model: the diameter was reduced to impede hairpin formation and 12 negative charges were placed in a ring near the trans end of the pore. With these modifications, we observe a clear banding region, analogous to those reported in numerous experimental studies.

The banding apparent in Figure 14 is not found at all potential differences. At higher potential differences, the banding is much less pronounced, whereas at lower potential differences, no translocation is possible. We have previously found that a minimum "threshold" level of applied potential difference exists, below which translocation does not occur. The αHL/lipid membrane system makes measurements impossible at larger applied voltages.

Banding may present a serious obstacle in designing sensing and sequencing devices on the αHL pore because the bands occur at fairly high blockade levels and are difficult to resolve from the useful translocation data points. The synthetic nano-pores, on the other hand, are burdened with the complication of hairpin events.

2.E Conclusions

A model has been created to study the dynamics of a ds-DNA chain undergoing voltage-induced translocation through a nanoscopic pore. When non-chain-end portions of the chain are inserted into the pore, a sharp bend in the chain configuration, or a "hairpin", is formed. At lower field strengths, the hairpin vertex often penetrates significantly into the pore before ultimately being rejected back to the cis compartment. When such failed events are not distinguished from successful translocation events – and unfortunately, such a distinction is not trivial experimentally – the number of hairpin translocations can be overcounted at low voltage. When failed attempts are included in statistics, the fraction of events occurring in single file fashion increases with voltage. Conversely, when only successful translocations are considered, the fraction of single file events is found to decrease with increasing voltage.

Translocation time was found under certain conditions to decrease the closer the hairpin vertex is to the chain center. While such a variance in translocation time for a single chain length introduces a complicating factor in using translocation data for characterization purposes, we have also found that under specific conditions, the

translocation time can be made to be invariant with respect to the identity of the chain vertex.

Finally, the banding phenomenon observed in experimental event diagrams can be produced in simulations with a simple modification of the model pore, borrowing two features from the structure of αHL: a narrower pore diameter and a trans-side negatively charged residues. The negative charges present an additional barrier at the trans-side of the pore, in effect stalling the chain in a partially inserted state for long periods of time. The stalled states lead directly to familiar horizontal bands along the time axis of event diagrams.

2.F Acknowledgements

It is a pleasure to thank the members of the Harvard Nanopore Group, in particular Professors Branton and Golovchenko, for stimulating discussions. Acknowledgement is made to the National Institutes of Health grant No. 1R01HG002776-01, National Science Foundation grant No. 0605833 and the Materials Research Science and Engineering Center at the University of Massachusetts Amherst.

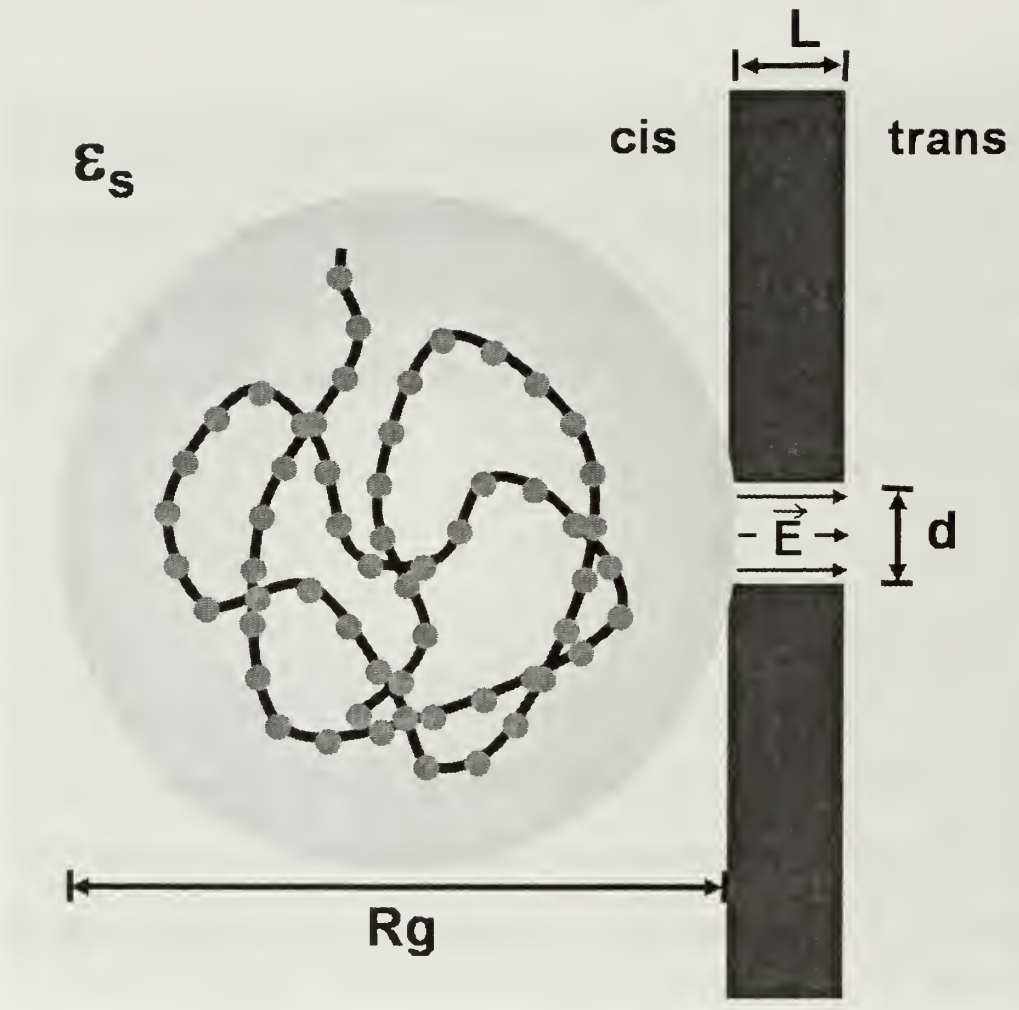

Figure 2-1 Schematic of simulation model

The impermeable membrane, shown in black, separates the cis (donor) side from the trans (recipient) side. A fluid of dielectric constant εS fills the cis chamber, the trans chamber, and the pore providing a means of passage through the membrane. On the cis side, a spherical container of radius Rg prevents the diffusion of the polymer away from the channel opening, which provides the only means of escape. The applied potential drop is assumed to occur entirely across the length L of the membrane. Thus the electric field E is constant within the channel.

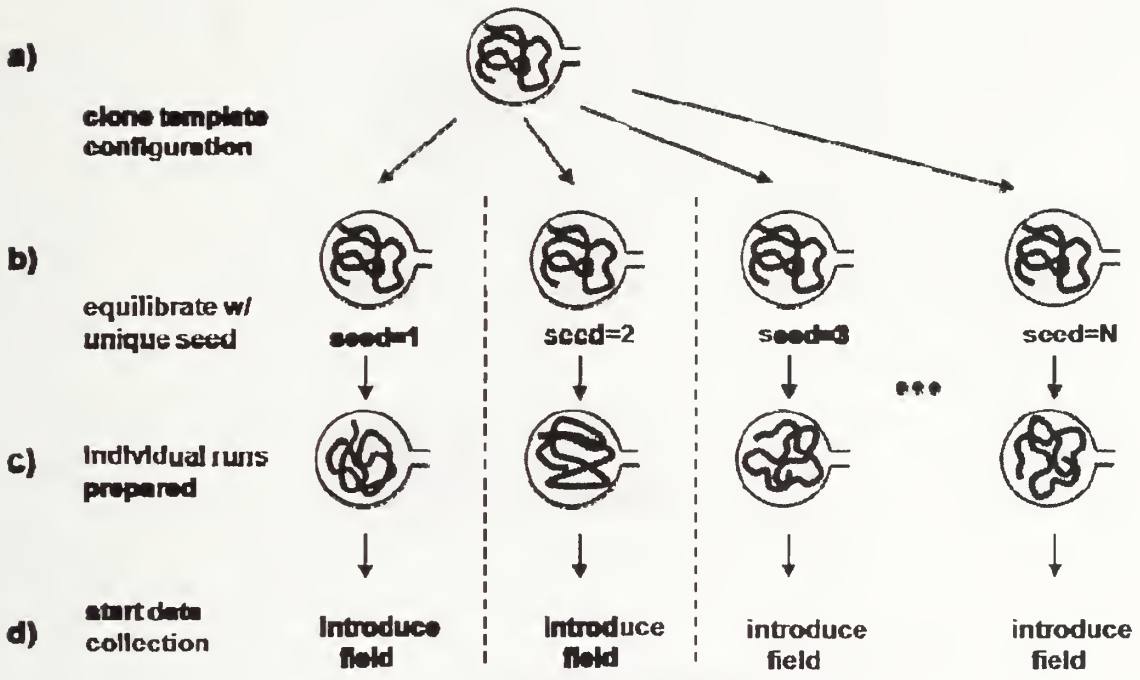

Figure 2-2 Simulation preparation procedure

For each N, a free solution Rg is determined and used as the radius of the confining sphere to be used in subsequent steps (a)–(d). (a) A single free solution configuration, selected so as not to extend beyond the walls of the confining sphere, is placed inside the confining sphere. This configuration then serves as the template configuration for all simulations with chains of length N. (b) Clones of the template configuration are generated and equilibrated with a unique random seed (no field), generating (c) a large number of uncorrelated initial configurations. (d) Simulation runs: The electric potentials are applied to each of the initial configurations and data are collected.

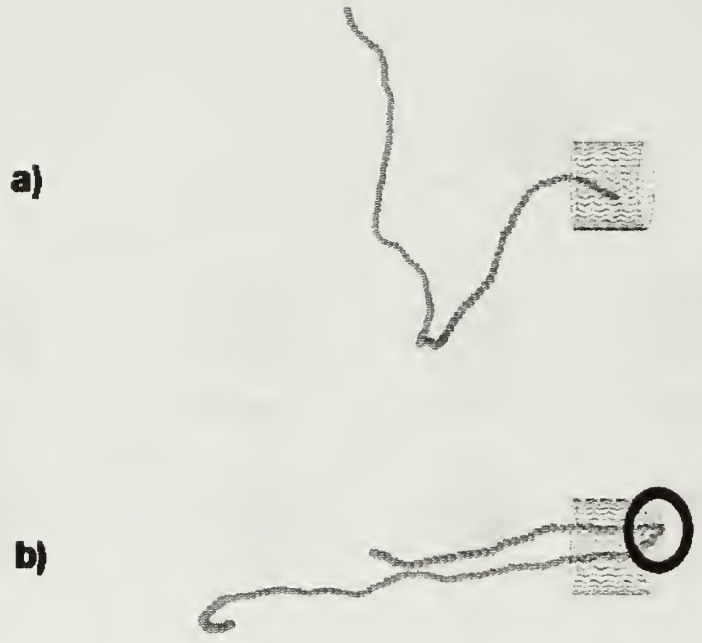

Figure 2-3 Sample translocation snapshots

The emphasis is on the ds-DNA configurations. The pore is outlined by the black wire frame. The membrane is not shown. (a) Single-file-initiated translocation attempt. (b) Hairpin translocation attempt. The position of the hairpin vertex is illustrated with a circle.

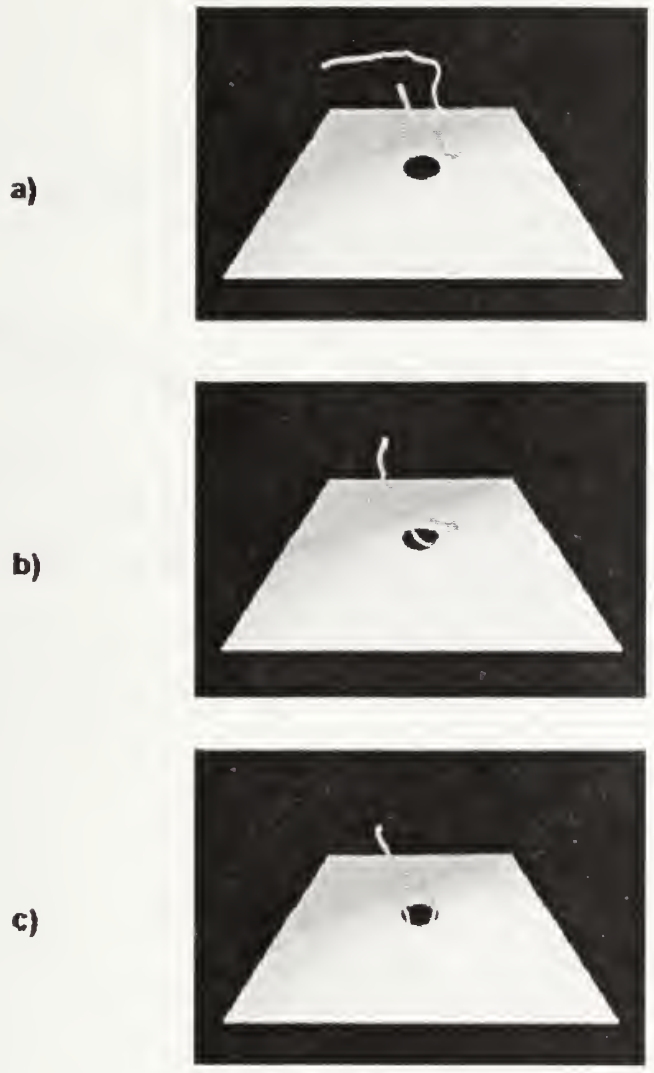

Figure 2-4 Snapshot of hairpin initiation

A cis-side view of a "hairpin" event. (a) The chain gyrates near the entrance of the channel; (b) a non-chain-end bead is presented at the opening and is drawn in, forcing the formation of a hairpin; (c) the vertex of the hairpin pushes trans ward into the channel, with two strands trailing behind. While such a configuration should lead to measurable drop in current flux through the channel, we have found that the chain may still ultimately be rejected by the channel. Such rejection is more prominent at lower potential differences and when αHL-like charge distributions are included in the channel.

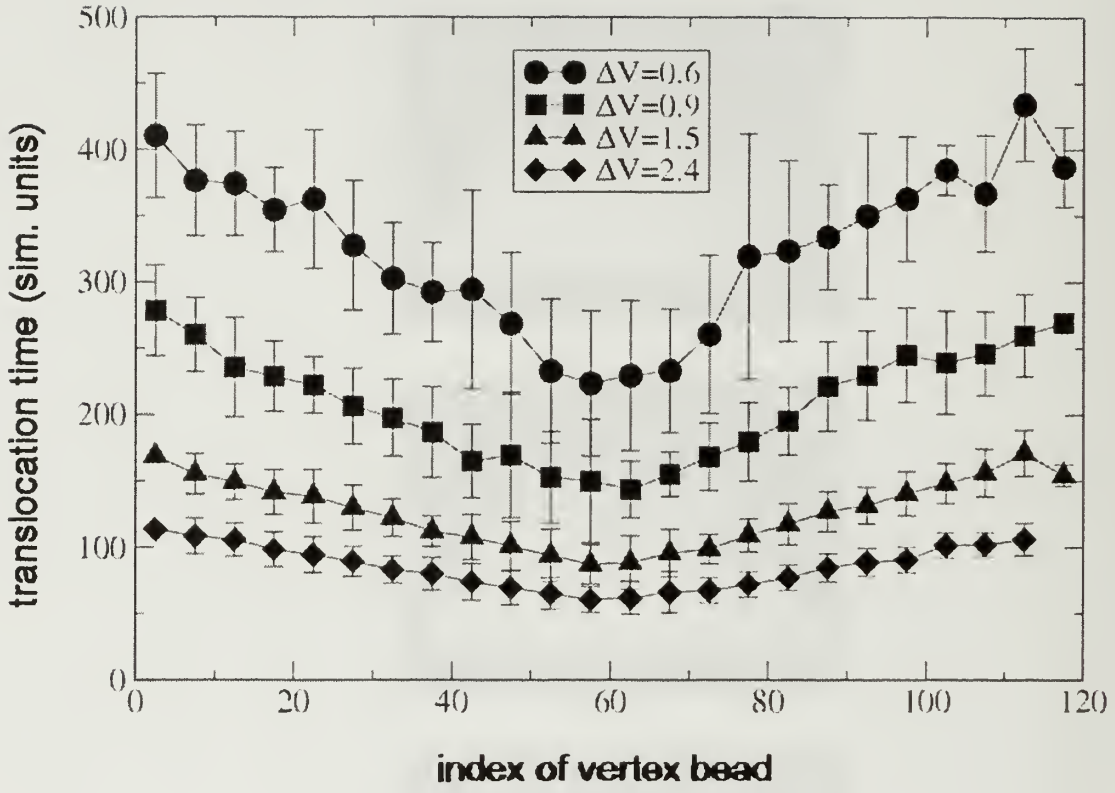

Figure 2-5 Translocation time as a function of vertex position for larger pore

Translocation time data for chains of length N=120 through the 15 nm diameter channel is plotted. Curves for five different potential differences are given. Larger potential differences lead to shorter translocation times. For all potential differences, the minimum in translocation time is found when the hairpin vertex is located directly in the middle of the chain.

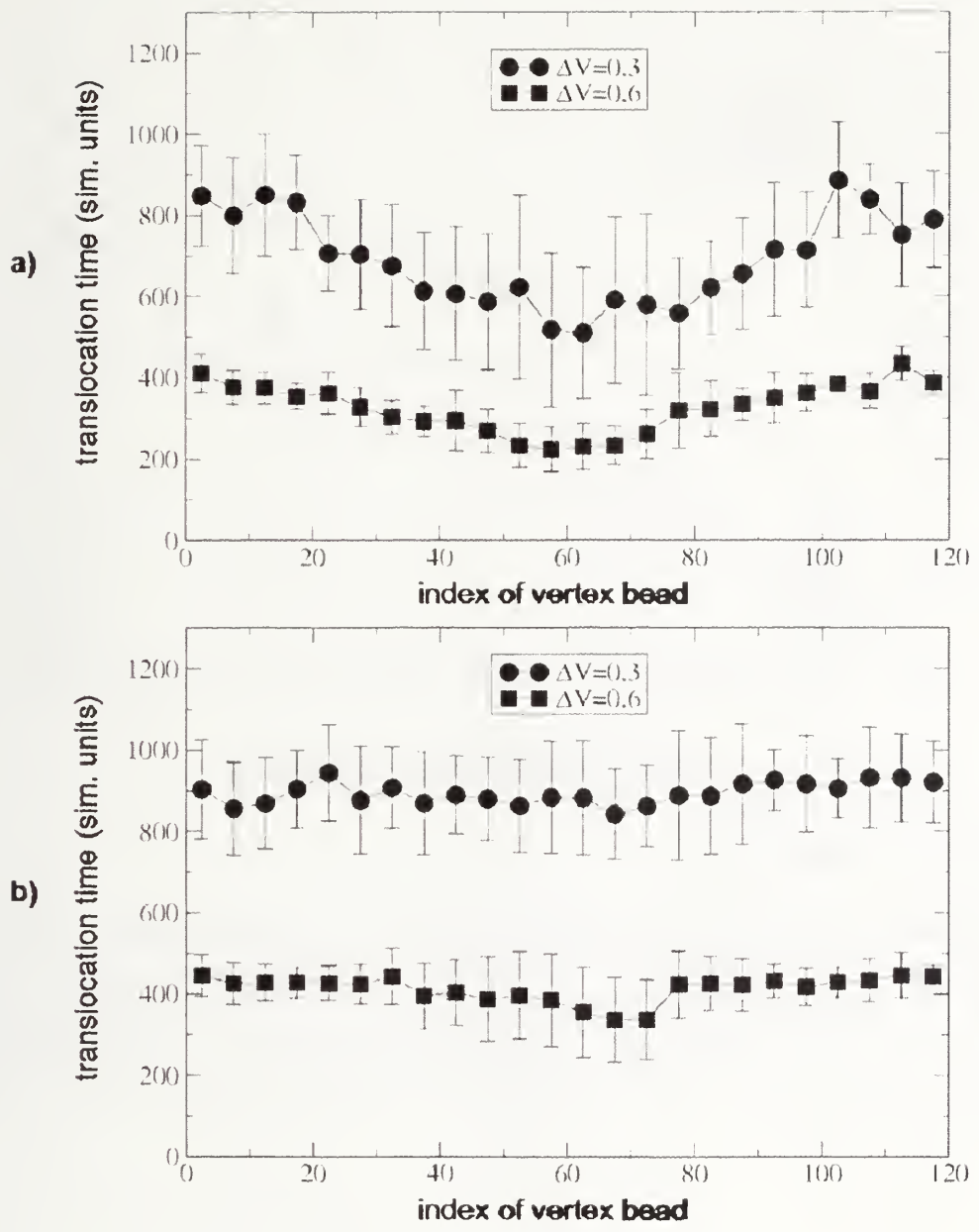

Figure 2-6 Comparison of larger and smaller channels—translocation time

Translocation time data for chains of length N=120 through both (a) 10 nm and (b) 15 nm diameter channels are plotted. Curves for two different potential differences are given. In both cases, larger potential differences lead to shorter translocation times. However, translocation times are longer through the 10 nm channel. In addition, the vertex position dependence of the translocation time is not evident in the smaller.

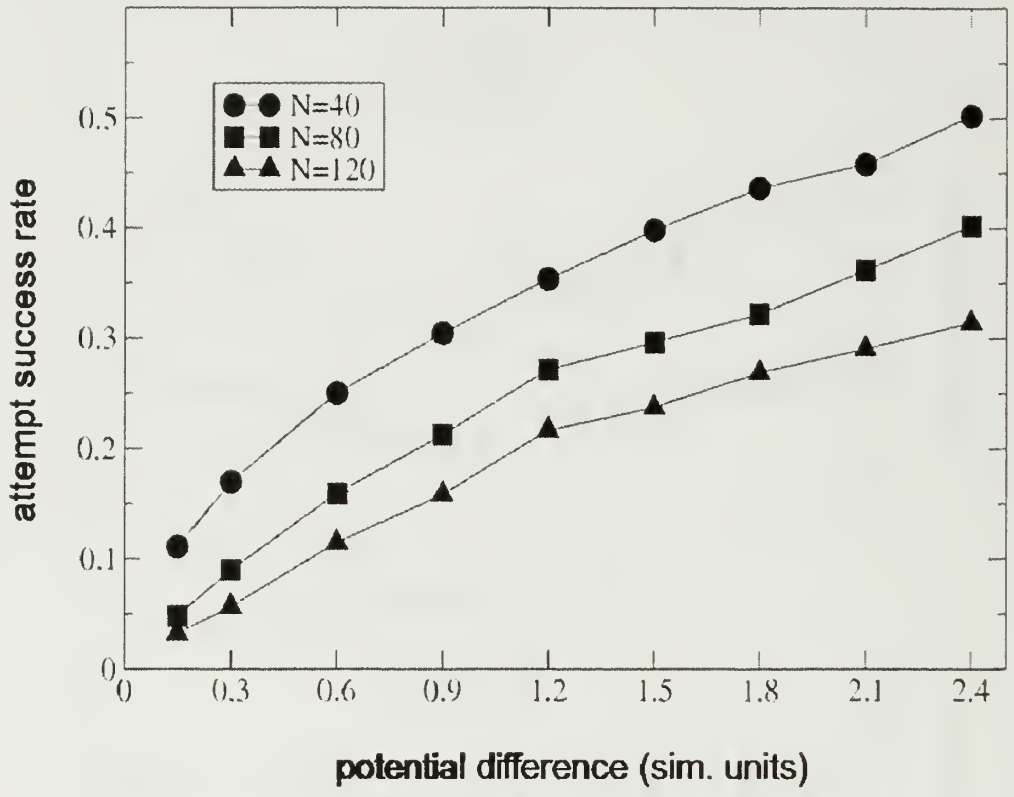

Figure 2-7 Success ratio

Success ratio is plotted for each chain length and potential difference. For all chain lengths, success ratio increases monotonically with increasing applied potential differences. At all potential differences, success ratio increases with decreasing chain length. This increase follows from the increasing fraction of free ends with decreasing molecular weight: More free ends means more single file attempts, which in turn are more likely to succeed.

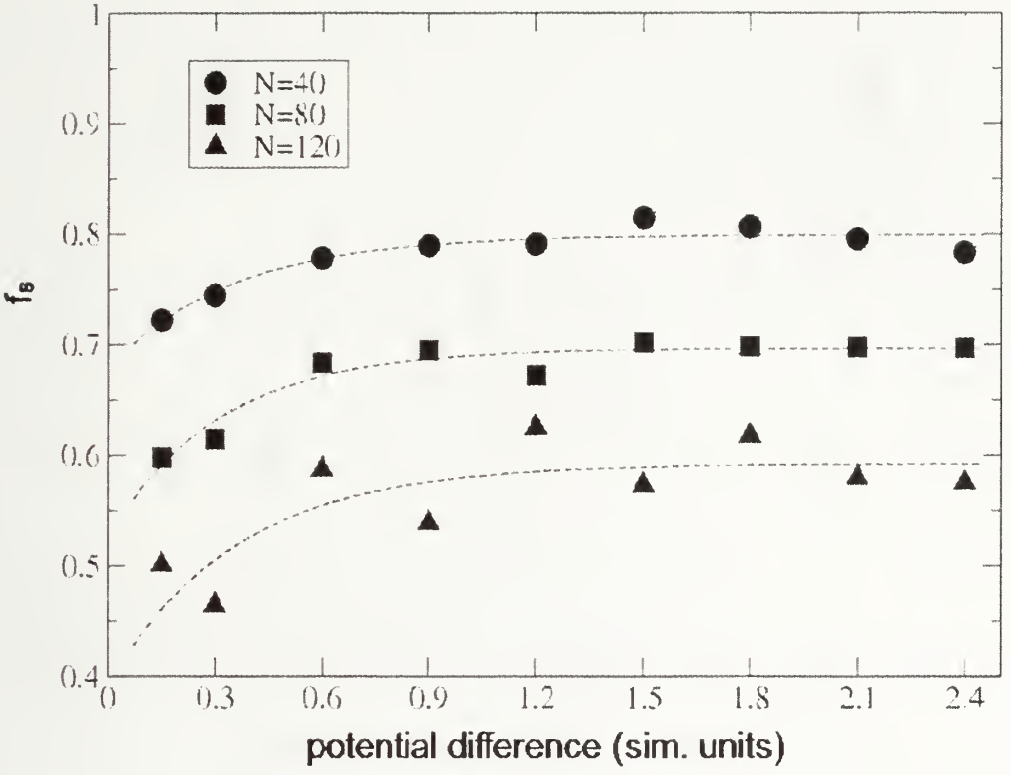

Figure 2-8 Fraction of single file events

For each chain length (N=40,80,120) and potential difference (0.15–2.4) considered in this study, the fraction fS of total events in which chain penetration occurred in a single file fashion is plotted. Dashed lines are exponential fits. All events, regardless of whether they lead to successful translocation, were considered. At all potential differences, single file events are more prevalent for shorter chain lengths, due to the increased fraction of free ends for shorter chains. As has been reported experimentally, larger potential differences lead to higher probabilities of single file events.

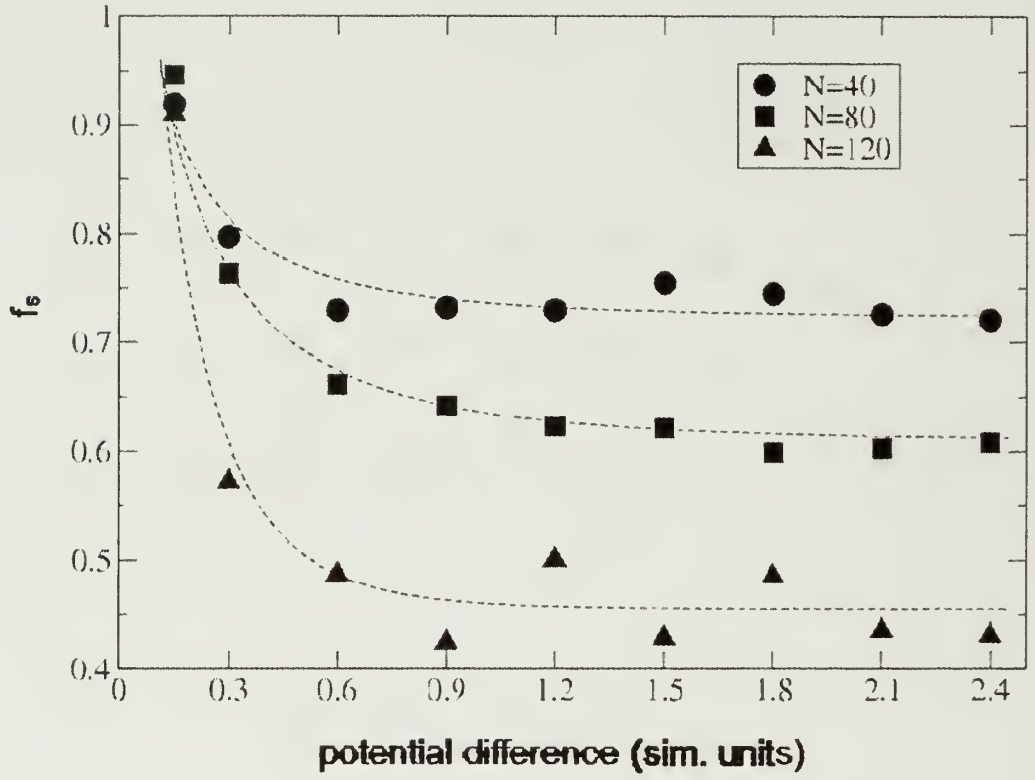

Figure 2-9 Fraction of single file translocations

This figure is as Fig. 8, except that failed translocation events are excluded from consideration. The fraction of single-file successful translocations can be seen to decrease at higher potential differences, as the energetic barrier to hairpin formation becomes less significant at higher potentials.

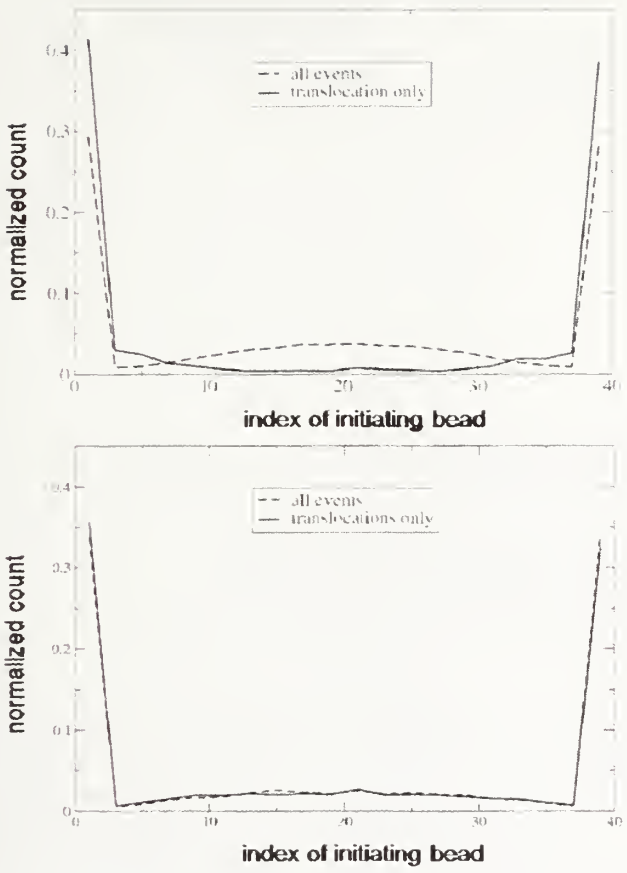

Figure 2-10 Histogram of index of initiating bead N=40

Histograms are plotted for all events (dashed line) and translocations only (solid line). For each type, the histogram is normalized by the number of attempts for that type (i.e., the total area under each curve is 1). The bead index of the initiating bead identifies an event as being either single file (indices 1–5 and 36–40) or hairpin (indices 6–35). (a) Δ V=0.15. As can be seen from the dotted curve, a reasonable fraction of all attempts are initiated by hairpins. However, successful translocations (solid curve) are dominated by single file configurations (see solid curve). Therefore, at this low driving potential difference, the population of failed attempts must be dominated by hairpin configurations. (b) ΔV=2.4. Histograms are similar for all events and translocations only.

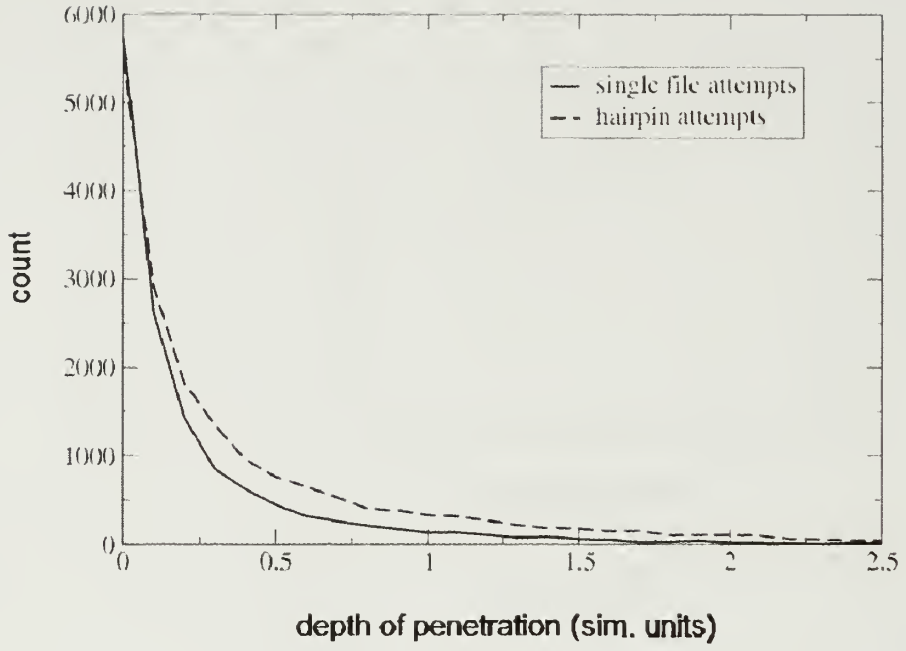

Figure 2-11 Failed event depth histogram N=40, ΔV=0.15

The abscissa is the maximum distance of penetration attained during a failed translocation attempt; the ordinate is the relative number of occurrences. The failed attempts quantified in this figure occurred in the course of what eventually became 3000 successful translocations. Two features of this graph are particularly significant: Firstly, hairpin attempts can be seen to comprise the majority of failed attempts at this low potential. Secondly, hairpins penetrate on average more deeply into the channel before being rejected than do chain ends. As potential difference is increased, these features become progressively indiscernible.

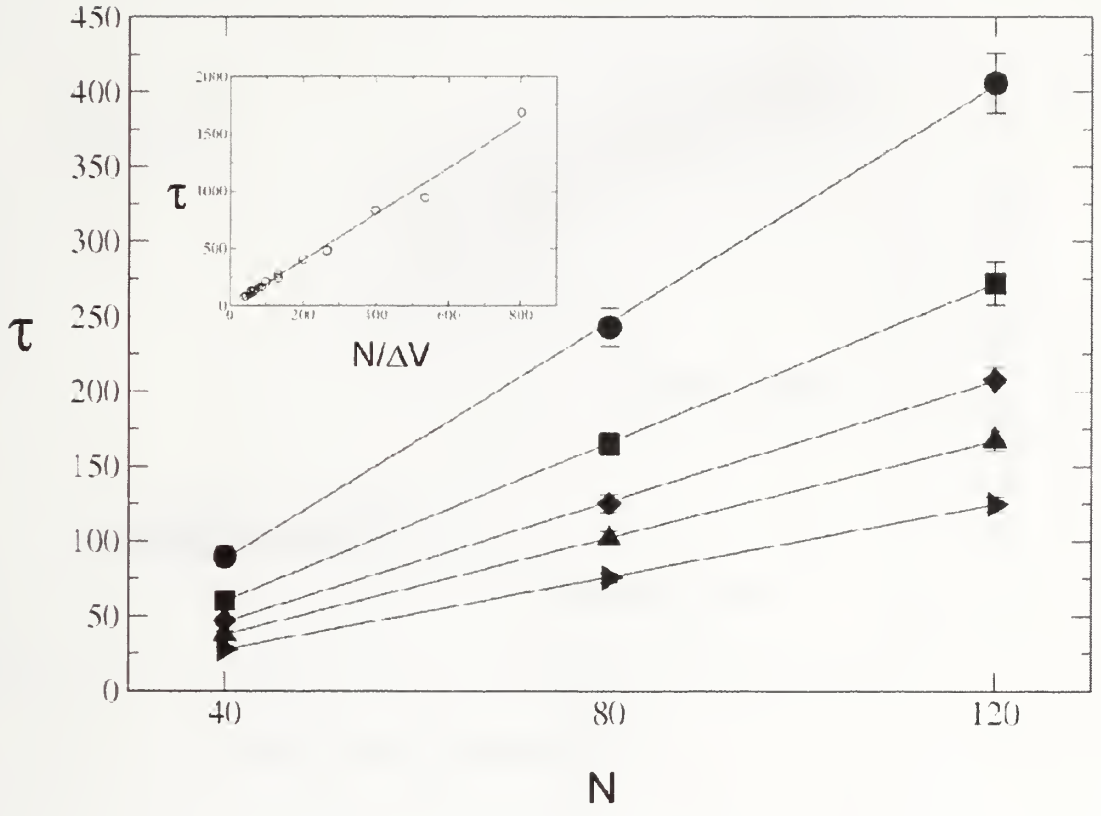

Figure 2-12 Translocation time τ as a function of chain length N

A linear relationship is found for translocation time as a function of chain length. Five different voltages are shown: Circles ΔV=0.6; squares ΔV=0.9; diamonds ΔV=1.2; up triangles ΔV=1.5; sideway triangles ΔV=2.1. Solid lines are linear fits. Inset: translocation time as a function of N/ΔV for N=80 and N=120.

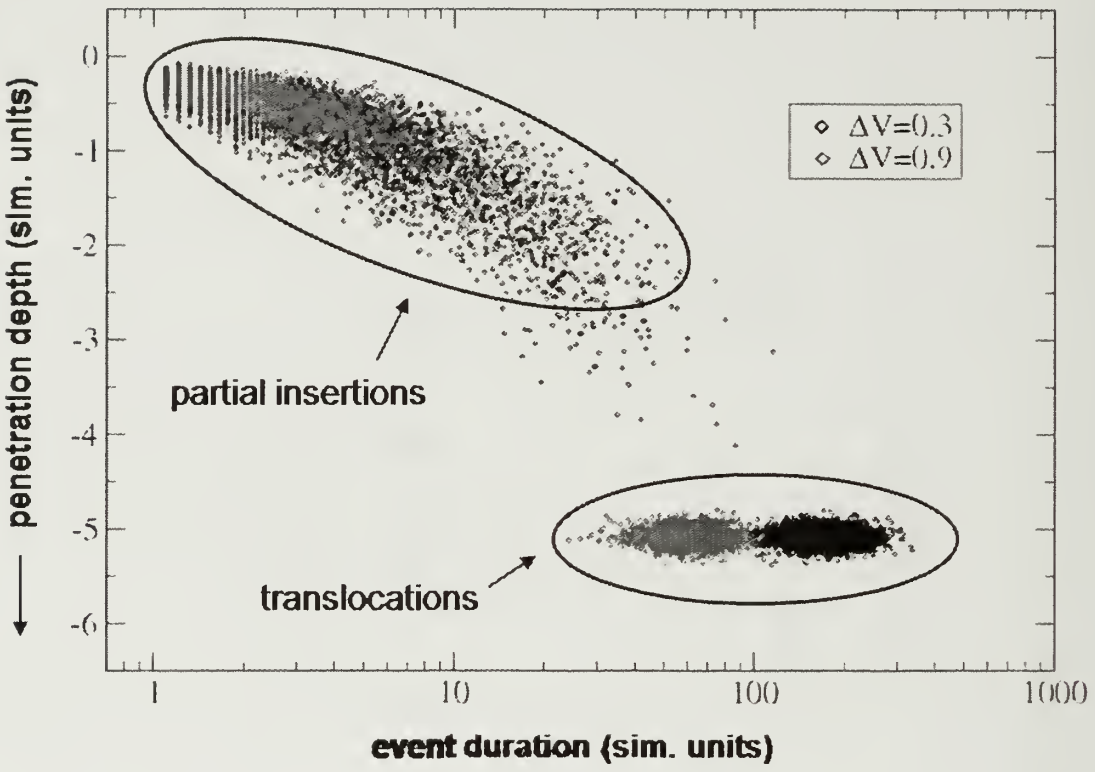

Figure 2-13 Event Diagram: Nanopore

Data points correspond to individual events (see text). Data are plotted for chain length N=40 and two potential differences, 0.3 and 0.9. The penetration depth refers to the position of maximum penetration attained by the leading bead of the polymer chain during an event. The pore spans from y=0 (cis end) to y=−5 (trans end). Successful translocation penetration depths are found in the region of y=−5. Translocation times decrease with an increase in potential difference. Short-time data points form a shoulder-shaped region, also seen in experimental translocation data. The banding phenomenon prevalent in experiments is not present.

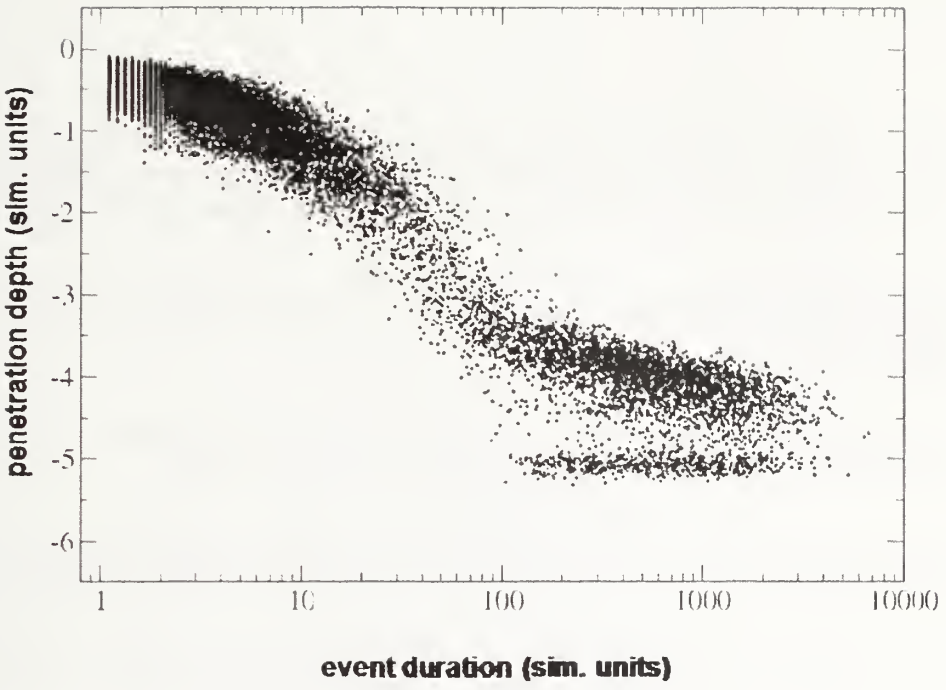

Figure 2-14 Event diagram for N=40, Δ V=0.6 with modified nanochannel

The data for the event diagram were obtained using a modified nanochannel meant to mimic the αHL channel through the inclusion of two key features: A narrower channel and the presence of a ringlike distribution of negative charges along the channel wall near the trans opening (see text). The small diameter allows only single file access to the channel, whereas the charges act as an additional barrier, temporarily "stalling" the chain within the channel. The stalling yields long-time events of intermediate penetration depth, which are visible as the "band" overlying the translocation events.

2.H References

1. Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D., Rapid nanopore discrimination between single polynucleotide molecules. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, 97, (3), 1079-1084.

2. Kasianowicz, J. J.; Henrickson, S. E.; Weetall, H. H.; Robertson, B., Simultaneous multianalyte detection with a nanometer-scale pore. *Analytical Chemistry* **2001**, 73, (10), 2268-2272.

3. Meller, A.; Nivon, L.; Branton, D., Voltage-driven DNA translocations through a nanopore. *Physical Review Letters* **2001**, 86, (15), 3435-3438.

4. Howorka, S.; Cheley, S.; Bayley, H., Sequence-specific detection of individual DNA strands using engineered nanopores. *Nature Biotechnology* **2001**, 19, (7), 636-639.

5. Deamer, D. W.; Branton, D., Characterization of nucleic acids by nanopore analysis. *Accounts of Chemical Research* **2002**, 35, (10), 817-825.

6. Sauer-Budge, A. F.; Nyamwanda, J. A.; Lubensky, D. K.; Branton, D., Unzipping kinetics of double-stranded DNA in a nanopore. *Physical Review Letters* **2003**, 90, 23.

7. Mathe, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Meller, A., Nanopore unzipping of individual DNA hairpin molecules. *Biophysical Journal* **2004**, 87, (5), 3205-3212.

8. Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W., Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, 93, (24), 13770-13773.

9. Akeson, M.; Branton, D.; Kasianowicz, J. J.; Brandin, E.; Deamer, D. W., Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophysical Journal* **1999**, 77, (6), 3227-3233.

10. Murphy, R. J.; Muthukumar, M., Threading synthetic polyelectrolytes through protein pores *Journal of Chemical Physics* **2007**, 126, (5), 1-4.

11. Li, J.; Stein, D.; McMullan, C.; Branton, D.; Aziz, M. J.; Golovchenko, J. A., Ion-beam sculpting at nanometre length scales. *Nature* **2001**, 412, (6843), 166-169.

12. Siwy, Z.; Fulinski, A., Fabrication of a synthetic nanopore ion pump. *Physical Review Letters* **2002**, 89, 198103.

13. Li, J. L.; Gershow, M.; Stein, D.; Brandin, E.; Golovchenko, J. A., DNA molecules and configurations in a solid-state nanopore microscope. *Nature Materials* **2003**, 2, (9), 611-615.

14. Saleh, O. A.; Sohn, L. L., An artificial nanopore for molecular sensing. *Nano Letters* **2003**, 3, (1), 37-38.

15. Storm, A. J.; Chen, J. H.; Ling, X. S.; Zandbergen, H. W.; Dekker, C., Fabrication of solid-state nanopores with single-nanometre precision. *Nature Materials* **2003**, 2, (8), 537-540.

16. Chen, P.; Gu, J. J.; Brandin, E.; Kim, Y. R.; Wang, Q.; Branton, D., Probing single DNA molecule transport using fabricated nanopores. *Nano Letters* **2004**, 4, (11), 2293-2298.

17. Heng, J. B.; Ho, C.; Kim, T.; Timp, R.; Aksimentiev, A.; Grinkova, Y. V.; Sligar, S.; Schulten, K.; Timp, G., Sizing DNA using a nanometer-diameter pore. *Biophysical Journal* **2004**, 87, (4), 2905-2911.

18. Storm, A. J.; Storm, C.; Chen, J. H.; Zandbergen, H.; Joanny, J. F.; Dekker, C., Fast DNA translocation through a solid-state nanopore. *Nano Letters* **2005**, 5, (7), 1193-1197.

19. Sung, W.; Park, P. J., Polymer translocation through a pore in a membrane. *Physical Review Letters* **1996**, 77, (4), 783-786.

20. DiMarzio, E. A.; Mandell, A. J., Phase transition behavior of a linear macromolecule threading a membrane. *Journal of Chemical Physics* **1997**, 107, (14), 5510-5514.

21. Lubensky, D. K.; Nelson, D. R., Driven polymer translocation through a narrow pore. *Biophysical Journal* **1999**, 77, (4), 1824-1838.

22. Muthukumar, M., Polymer translocation through a hole. *Journal of Chemical Physics* **1999**, 111, (22), 10371-10374.

23. Muthukumar, M., Translocation of a confined polymer through a hole. *Physical Review Letters* **2001**, 86, (14), 3188-3191.

24. Ambjornsson, T.; Apell, S. P.; Konkoli, Z.; Di Marzio, E. A.; Kasianowicz, J. J., Charged polymer membrane translocation. *Journal of Chemical Physics* **2002**, 117, (8), 4063-4073.

25. Berezhkovskii, A. M.; Gopich, I. V., Translocation of rodlike polymers through membrane channels. *Biophysical Journal* **2003**, 84, (2), 787-793.

26. Slonkina, E.; Kolomeisky, A. B., Polymer translocation through a long nanopore. *Journal of Chemical Physics* **2003**, 118, (15), 7112-7118.

27. Kong, C. Y.; Muthukumar, M., Polymer translocation through a nanopore. II. Excluded volume effect. *Journal of Chemical Physics* **2004**, 120, (7), 3460-3466.

28. Aksimentiev, A.; Schulten, K., Imaging alpha-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophysical Journal* **2005**, 88, (6), 3745-3761.

29. Fyta, M. G.; Melchionna, S.; Kaxiras, E.; Succi, S., Multiscale coupling of molecular dynamics and hydrodynamics: Application to DNA translocation through a nanopore. *Multiscale Modeling & Simulation* **2006**, 5, (4), 1156-1173.

30. Forrey, C.; Muthukumar, M., Langevin dynamics simulations of genome packing in bacteriophage. *Biophysical Journal* **2006**, 91, (1), 25-41.

31. Wong, C. T. A. M., M., Polymer capture by electro-osmotic flow of oppositely charged nanopores. *Journal of Chemical Physics* **2007**, 126, 164903 -164908.

32. Manning, G. S., Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions .I. Colligative Properties. *Journal of Chemical Physics* **1969**, 51, (3), 924.

33. Liu, S.; Muthukumar, M., Langevin dynamics simulation of counterion distribution around isolated flexible polyelectrolyte chains. *Journal of Chemical Physics* **2002**, 116, (22), 9975-9982.

34. Liu, S.; Ghosh, K.; Muthukumar, M., Polyelectrolyte solutions with added salt: A simulation study. *Journal of Chemical Physics* **2003**, 119, (3), 1813-1823.

35. Tildesley, M. P. A. a. D. J., *Computer simulation of liquids*. Clarendon Press: oxford, **1987**.

36. Kasianowicz, J. J., Mechanism of ionic current blockades during polymer transport through pores of nanometer dimensions. In *Structure and Dynamics of Confined Polymers*, Kasianowicz, J. J. K., M.; Deamer, David W. , Ed. Springer: **1999**; Vol. 87.

37. Meller, A., Dynamics of polynucleotide transport through nanometre-scale pores. *Journal of Physics-Condensed Matter* **2003**, 15, (17), R581-R607.

38. Wang, H.; Dunning, J. E.; Huang, A. P. H.; Nyamwanda, J. A.; Branton, D., DNA heterogeneity and phosphorylation unveiled by single-molecule electrophoresis. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, 101, (37), 13472-13477.

39. Butler, T. Z.; Gundlach, J. H.; Troll, M. A., Determination of RNA orientation during translocation through a biological nanopore. *Biophysical Journal* **2006**, 90, (1), 190-199.

40. Song, L. Z.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E., Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **1996**, 274, (5294), 1859-1866.

CHAPTER 3

MINIMAL MODEL OF THE EQUILIBRIUM MODEL OF THE EQUILIBRIUM STRUCTURE OF AN RNA VIRUS GENOME

3.1 Abstract

We have used Langevin dynamics to explore the packaged structure of RNA in pariacoto viruses. Our course grained model was based on the published crystal structure of the rigid portion of the pariacoto capsid and also included flexible N-terminal protein arms, attached to the rigid capsid at the appropriate locations. The inclusion of charged residues in our model was dictated solely by the location of charges inherent in the pariacoto sequence itself. Our model of the pariacoto capsid and its N-terminal arms, while course grained, was a straightforward mapping from the known physical structure and amino acid sequence of the pariacoto virus. The intent of our study was to explore the role of electrostatics on the structure of RNA packaged within the pariacoto virus. It has been reported that experimentally discernable features of RNA structure is independent of the specific sequence of encapsidated RNA. Two of these experimentally observed features of RNA structure within pariacoto viruses are reproduced in our simulations: 1) a dodecahedral cage of RNA is formed through complexation with charged residues on the inner capsid surface; 2) the radial distribution of disordered RNA (uncomplexed with the capsid) shows a bimodal distribution. Results of our work can be interpreted primarily as a consequence of electrostatics, as consideration of base-pairing

has been omitted. We propose that our work supports the growing body of evidence that electrostatic interactions play a crucial role in RNA viral assembly and structure.

3.2 Introduction

Over the past few decades, stunningly detailed crystal structures and three-dimensional reconstructions of icosahedral viral particles have become commonplace, yielding an enormous wealth of information about the structure of viruses. Cryo-transmission electron microscopy[1-3] and X-ray crystallography[4-6] reveal, at Angstrom-level resolution, highly symmetrical supra-molecular assemblages. The structural information corresponds mainly to the rigid protein shell, or capsid, which encapsulates the viral genome. As data have appeared for more viruses, a general understanding of capsid structure has emerged. Capsids are composed of a small number - often only one - of viral proteins, which are associated in a basic building block, called the asymmetric unit. The capsid is composed of multiple copies of this building block, whose relative positions obey icosahedral symmetry[7, 8]. This high degree of symmetry is responsible for the success of diffraction techniques[9].

For the case of single stranded RNA (ssRNA) viruses, encapsidated genome structures buried within the surrounding capsid have proven more difficult to study[10]. Because the RNA does not in general conform to the overall icosahedral symmetry of the capsid, the genomes are largely invisible in typical x-ray diffraction and cryo-TEM studies. For this reason, computer simulation has been increasingly applied as a means of exploring the packaging/assembly kinetics and equilibrium genome structure of ssRNA viruses[11-13]. Our approach is to model the environment inside viral capsids, employing

Langevin dynamics to simulate the dynamics of encapsidated viral genomes. We have previously used this approach to explore the packaging and ordering of ds-DNA within icosahedral capsids[14]. In the present work, we address the question of the structure inside single-stranded RNA containing viruses. In future work, we will address the kinetics of assembly of ss-RNA viruses.

A key issue of packaged genome modeling centers on identifying the interactions that most influence the genome structure. We found in our modeling work of ds-DNA-containing viruses that the bending energy of the DNA double helix and its response to confinement played a central role in the packaged structure and the dynamics of packaging[14]. This is unlikely to be the case for ss-RNA viruses as ss-RNA is much more flexible than ds-DNA. There are likely four classes of interactions that will influence RNA structure: sequence-dependent specific interactions (e.g. interactions giving rise to specific secondary structures); electrostatic interactions, acting strictly on the charges of the nucleic acids and amino acids and therefore independent of their sequences; excluded volume interactions between the various chemical constituents in the crowded capsid interior; and bond connectivity linking together the individual mers of the RNA and the protein arms.

Previous analytical work has underscored the importance of electrostatics in encapsidated genome structure and assembly[15, 16]. We are particularly intrigued by experimental work that suggests that, despite the presence of sequence dependent secondary structures, electrostatic interaction independently shapes some elements of the encapsidated genome structure. While ss-RNA genomes are always mostly invisible, there have been a growing number of X-ray diffraction and cryo-TEM studies showing

regular and symmetrical ordering of the part of the viral genome associated with specific regions along the inner capsid surface[17-21]. These regions on the inner surface of the capsid contain an excess of fixed positive charges and are believed to function as a sequence-independent template for RNA. The notion of the capsid serving as a non-specific template for RNA has been strongly supported by a remarkable cryo-TEM study demonstrating that in Flock-House Virus the RNA/capsid complex forms a dodecahedral cage structure in both the *in-vivo* virus as well as in virus constructs assembled with exogenous RNA[17]. Apparently, the dodecahedral structure of the RNA is "pre-programmed" by the capsid, independent of the details of the secondary structure of the RNA from which it is formed.

As diffraction data reveal only the averaged density of the ordered RNA at given icosahedrally symmetric positions along the capsid wall, a crucial piece of information is unavailable: the specific nucleotide sequences involved in the structure. What is known is that the crystallized RNA portions of these putatively single stranded RNA viruses are comprised of A-type double stranded helical RNA[17]. Thus, while the complexation of RNA with the inner capsid surface is independent of RNA sequence, the complexed RNA is comprised of a helical secondary structure associated with sequence-matched base pairing. It is not understood how exogenous RNA sequences form base paired helical secondary structures which are then formed into a larger dodecahedral structure. One possibility is that portions of base paired helices, thought to arise for any given RNA sequence, are simply preferentially complexed with capsid charges as a consequence of the high linear charge density of double stranded nucleic acid. Another suggested possibility is that the double helix is "non-canonical", meaning that the double-helical

portions are not formed from perfectly complementary strands of RNA. The first possibility involves helix formation prior to complexation, whereas the second would involve independent complexation of two strands in close proximity, followed by local helix formation. It is interesting to note that dodecahedral RNA structures have also been found in dsRNA viruses[22, 23], which would suggest helix formation prior to complexation. We will return to the topic of the dodecahedral structure and its double stranded RNA in the discussion section.

To explore the role of electrostatics in partially ordered RNA genomes, we have constructed a basic pariacotovirus (PaV) model, in which the capsid proteins and RNA interact through non-specific electrostatic interactions. While we are interested strictly in the structure of the packaged genome, it must be pointed out that ss-RNA capsid assembly and genome packaging occur as part of the same process. In the present work, we will take a simplified approach, modeling the capsid as a completed structure based on its final crystallized structure. Dynamic RNA chains were attached to crystallized residues at the appropriate sites. By starting with a preformed capsid, we are neglecting the kinetic aspects of the particle assembly, focusing on the issue of equilibrium genome structure. In future work, we will treat the entire spontaneous assembly of RNA virus-like particles. In this work, we will describe the RNA density within the pariacoto capsid, which we will compare to experimental results.

3.3 Simulation Method

3.3.1 Model of Pariacoto Capsid (Minus Flexible N-terminal Arms)

The spatial coordinates of a single asymmetric unit of the pariacoto capsid were obtained from a protein data bank. Each amino acid residue was modeled as a single

stationary bead, whose center-of-mass was assigned to the coordinates of the α carbon of the corresponding amino acid in the crystal structure (see Figure 1 a,b). The asymmetric unit consists of three chemically equivalent copies of the coat protein. The copies each have a unique orientation and therefore are typically designated as subunits A,B, and C. A cross section of the asymmetric unit is roughly triangular in shape. By subjecting each of the beads in the asymmetric unit to the set of sixty icosahedral symmetry operators, a model capsid – minus the flexible N-terminal tails - was generated (see Figure 1c). The capsid geometry displays T=3 symmetry and corresponds to a thirty-sided polygon called a rhombic triacontahedron.

We have used a solid-walled rhombic triacontahedron container, whose center to face distance is 13.8 nm, as the outer physical barrier for all dynamic beads in our simulation (see Figure 2a). To model the internal environment of the capsid, amino acid residues with proximity to the inner capsid surface are essential. These inner capsid residues serve both as the primary physical barrier containing the genomic RNA and as point sources of electrostatic potential. The impenetrable container is lined on its inner surface with stationary beads duplicating the topological details of the inner capsid surface (see Figure 2b). Capsid beads representing charged residues were assigned the appropriate charge. Any charged residues outside of the rhombic triacontahedron, but within 2.5 nm were also included in our model, as they may exert electrostatic forces on beads within the capsid interior (see Figure 2c). A total of 16740 topological capsid beads are used in our simulations.

3.3.2 Model of Flexible N-Terminal Arms

Residues 7-48 of the B subunit and 7-50 of the C subunit retain free mobility following assembly and are therefore are not present in the crystal structure of pariacoto virus. In total there are 120 flexible N-terminal amino acid arms that project in to the capsid interior. The arms, rich in positively charged lysine and arginine residues, interact electrostatically with the negatively charged RNA genome. We model the arms as a string of beads mounted to the stationary portion of the capsid at the appropriate residue (residue 49 for B subunit arms and residue 51 for C subunit arms). Each model bead subsumes two amino acids and is given a bond length of 0.75 nm. The beads are dynamic, in that they respond to the various forces present in the simulation. The amino acids sequences of the arms are known, allowing us to assign appropriate charge to the corresponding model beads. Figure 3 gives a pictoral representation of our course graining approach to N-terminal arms.

3.3.3 Model of Genome RNA

The genome of pariacotovirus consists of two messenger-sense strands, RNA1 and RNA2, of single stranded RNA. RNA1, which encodes the viral RNA replicase, is the longer strand, with 3,011 nucleotides. RNA2, encoding the coat protein precursor α, is 1,311 nucleotides in length. We model the pariacoto genome as two separate chains of beads. Each bead, with a diameter of 0.75 nm, corresponds roughly to two nucleotide repeat units and thus subsumes approximately two charged phosphate units. However, counterion condensation limits the linear charge density of a polyelectrolyte in the presence of counterions. Only a fraction, α, of the charged groups along the chain remain unpaired with counterions. As the diameter of our beads is approximately equal

to the Bjerrum length, we assign to each RNA bead a single charge, corresponding to α=0.5. The number of beads in our model for RNA1 and RNA2 are 1755 and 585, respectively. With these values, the 3:1 ratio between the lengths of the viral RNA is maintained. The total number of RNA beads was selected to preserve net electroneutrality in our model system.

3.4 Results and Discussion

3.4.1 Capsid Serves as a Template for Dodecahedral RNA Structure

The density of encapsidated RNA in our simulation runs was first averaged icosahedrally for each stored time step and then averaged over all time steps. The composite density field was visualized as an isosurface-value plot using the openDX data visualization package. We considered two cases corresponding to the experimental work of Tihova *et al.*[17]: one where the inward-pointing and flexible N-terminal tails were present, as in the full pariacoto virus; and the case where the N-terminal tails were absent, as in the mutant pariacoto virus-like particles. The RNA-density isosurface images shown in Figure 5 clearly take the form of a dodecahedral cage. The edges of the dodecahedral cage are cradled within the positive charge-rich topological valleys connecting three-fold axes of symmetry (Figure 2d). Figure 2c shows how two abutting asymmetric groups form a valley, which is richly lined with positive charge. The dodecahedral cage portion of the structure is not noticeably different than in the case where the N-terminal arms are omitted, which indicates that the flexible N-terminal arms are not obstructing the process of RNA/capsid templating.

Examination of simulation movies demonstrates that, while the dodecahedral cage appears to be continuous structure in the averaged density field, it is

not formed from a continuous routing of RNA. Instead, portions of RNA are seen to complex with the capsid for short distances, separate from the capsid and submerge into the disordered interior, from which a different portion of the RNA emerges to continue the dodecahedral cage. Upon averaging over time and icosahedrally symmetric positions, the piecemeal nature of the dodecahedral cage appears to be continuous.

Thus, it may well be that the icosahedrally averaged structure obtained experimentally are also deceptively suggestive of an entirely connected path. The present results show that a continuous path of RNA is not necessary to form an icosahedrally-averaged continuous structure. This may be an important conclusion given that it is known that the ds-RNA dodecahedral structure cannot be formed from a continuous uninterrupted pathway of ds-RNA.

We can suggest two possible mechanisms by which the double stranded nature of the dodecahedral cage may arise. The first recognizes that the formation of double-stranded stretches may be entirely unrelated to the interaction of the RNA with the capsid. Rather, in free solution the putatively single-stranded RNA is thought to be folded back upon itself in such a way that there are in fact many double-stranded stretches along the chain contour. These double-stranded stretches possess large negative linear charge density, much like ds-DNA. Such highly charged helical stretches would be particularly prone to electrostatic complexation with the capsid and would provide the ds-RNA building blocks for the dodecahedral cage. The complexation would not consist of continuous stretches of RNA, but rather would occur in the piecemeal fashion that we have just proposed for ssRNA segments. A similar mechanism has been proposed by Tihova *et al.*[17], who suggested that base pairing between RNA segments in free solution

is reversible and that there are a sufficient number of secondary structures of RNA from which to select one that is particularly well-suited to formation of a dodecahedral cage.

Our second proposed mechanism treats ss-RNA as being truly single-stranded and involves the sequential electrostatic complexation of two individual strands of RNA along each templating-zone on the capsid surface. Indeed, the positively charged topological valley serving as the template location on the inner capsid surface is suited to hold two strands. Once the second segment has been immobilized along the capsid, the two aligned strands could further reduce the energy of the system by undergoing base-pair/base-pair hydrogen bonding with its neighboring strand of RNA. Because the two partnering strands were recruited in a non-specific way, through electrostatic complexation with the capsid charges, their sequences would be somewhat non-complementary. However, once two strands were aligned alongside each other, even non-ideal base-pairing would lower the free energy of the complex. Indeed, experimental results have suggested that sequences of RNA involved in helix formation are "non-canonical" or "non-Watson-Crick" in nature. Regardless of how the dodecahedral RNA comes to be double stranded, evidence is mounting that electrostatic interactions are primarily responsible for formation of the of the dodecahedral order.

3.4.2 Radial Density Profile of RNA

While the majority of packaged RNA retains sufficient mobility to elude detection by crystallographic methods, electron microscopy has been effective in determining averaged structural information about the RNA, in particular the radial density profiles. Tihova *et al.*[17] have reported two distinct peaks in the radial density distribution of RNA in Flock House virus, a close relative of pariacoto virus. This bi-modal radial density

distribution was also found for a virus-like construct encapsidating non-viral RNA. However, in the virus-like particles in which the N-terminal arms were omitted, only a single peak of density was reported. The presence of multiple peaks in the radial density profile is thus evidently due to the interaction of the RNA with the N-terminal arms. The experimental finding that the bi-modal RNA radial density distribution appears to be RNA sequence independent suggests, as was the case with the dodecahedral cage of RNA discussed previously, that sequence-independent electrostatics forces must play a central role. In Figure 6a, we plot the radial density of RNA in our simulations of the pariacoto virus. We also find that there are two prominent peaks of RNA density, at about 6 and 10 nm. The topology of the inner capsid surface is manifested in the RNA density profile as a shoulder of density at approximately 12 nm. This shoulder corresponds to the dodecahedral structure partially buried within the capsid valley, which overlaps with the density of the capsid itself. The outmost prominent peak of RNA, centered at 10 nm, is sharply spiked due to the high density at the point of attachment of the N-terminal arms to the capsid, at the apex of the topographical mountains of the inner capsid surface. The amino acids comprising these mounting points are positively charged and immobile, leading to the corresponding spike in RNA density. Such a spike would not be resolvable experimental profiles.

It is interesting to note that the linear distribution of charges along the flexible N-terminal tails of pariacoto virus, which project inward and interact with the encapsidated genome, and are also bi-modal (Figure 3). Specifically, positive charges, in the form of lysine and arginine residues, are clustered in two regions along each of the N-terminal arms, separated by an uncharged spacer. Should the N-terminal arms be elongated within

the virus, pointing inward away from the capsid wall like a surface grafted brush, a bi-modal radial charge distribution of positive protein charge would result. In this case, the negatively charged RNA would minimize electrostatic energy by conforming to the bi-modal radial density set by the N-terminal arms. However, examination of the radial density distribution of brush charge in our simulation reveals (Figure 6b) that the linear charge bi-modality along the contour of the brush is not manifested into two distinct peaks of radial density. This is in agreement with the observation that the N-terminal arms are not rigidly extended in our simulation movies. However, there is a shoulder of positive charge extending inward from the prominent (outer) peak, indicating that the flexible arms are somewhat, if not highly, extended. It is fair then to consider the N-terminal arms as constituting a brush, which forms a protein rich, and positive charge-rich, region extending inwards from the outer capsid wall.

The fact that the radial RNA density has a prominent inner peak, rather than the more subtle shoulder of protein arm charge, indicates that electrostatics are not acting alone in sculpting the RNA density. When viewing superimposed RNA and N-terminal arm charge densities, it is evident that the inner-most peak of RNA is displaced towards the inner side of the shoulder of brush charge density, indicating some degree of exclusion of RNA from the brush region. The polymer response to crowding is different when brushes are grafted on the inner surface of a sphere than when grafted to a flat surface. In the latter case, the chain can extend to alleviate crowding, whereas in the former case, extending leads to higher density in the center of the sphere, where there is less volume available. The grafted chains are thus restricted from extending fully to the center. We thus propose that the location of the second (inner) RNA peak is such as to

satisfy an equilibrium between two countering potentials: 1) the favorable electrostatic interactions with the brush arms in the shoulder region, tending to drive the RNA towards the shoulder region; and 2) excluded volume repulsion from the brush region, which works to expel RNA away from the brush region.

Even the simplest viruses are extraordinarily elaborate entities, displaying hierarchical levels of organization, generally built up from highly intricate secondary structures. Electrostatics, on the other hand, are inherently non-specific, seeing only positive and negative charge and blind to the specific sequences of RNA or amino acids. It therefore seems reasonable to question why nature, the sculptor of exquisitely detailed and precise systems, should employ such a "crude" tool. On the other hand, from a physical standpoint, the complexation of positively charged amino acids and negatively charged nucleic acid molecules must unavoidably be describable in some fundamental way by electrostatics.

We believe that the message that is emerging from experiments and our work is that specific interactions and electrostatic interactions are both of central importance to viruses. It will be important to identify phenomena for which electrostatics can provide sufficient explanation and those for which the complexity of specific interactions is irreducible. However as nature apparently has chosen in certain instances to employ the non-specific tool of electrostatics, it stands to reason that there be an adaptive benefit to doing so. Take for example the ordered dodecahedral cage of RNA in Flock-House virus, which has been suggested to impart enhanced structural stability and resistance to proteolytic degradation. Clearly, the non-specific nature of electrostatic interactions means that random mutation to a portion of the RNA that will end up as part of the

dodecahedral cage will not inhibit cage formation. Thus, the probability of non-catastrophic mutation is enhanced. The specificity of assembly, while highly effective, seems to be accounted for by localized specific signal sequences, not wholesale specificity between RNA and capsid proteins. Indeed, if all components of assembly were delicately dependent upon the entire RNA sequence, the ability of viruses to adapt might be profoundly compromised.

3.5 Conclusions

We have modeled the interior environment of the pariacoto virus and used Langevin dynamics simulation to study the structure of the encapsidated RNA. We have incorporated details of the crystal structure of the inner surface of the pariacoto virus (PaV) capsid. In addition, we include the dynamic N-terminal amino acid tails, which extend inward from the inner capsid surface. While the tails maintain their flexibility and are mostly invisible in electron microscopy and X-ray crystallography studies, we present findings suggesting that they are nonetheless essential to RNA virus assembly and that their influence is directly and indirectly manifested in experimental results. Both the N-terminal tails and numerous amino acid residues on the inner surface of the crystallized portion of the capsid bear charge and interact with the negatively charged RNA genome. While the encapsidated genome structure of RNA-containing viruses has been difficult to study experimentally, there have been a number of reports that the RNA in close proximity to the inner capsid surface displays icosahedrally symmetric ordering. Our simulations results reveal a dodecahedral cage of encapsidated RNA in pariacoto virus, formed through electrostatic complexation with specific charged regions on the inner

capsid surface. This dodecahedral structure resembles those observed experimentally in both Pariacoto virus and its close relative Flock-House virus. Additionally, we find that the radial density distribution of encapsidated RNA has two peaks, which is in agreement with experimental results. We propose that electrostatic complexation of viral RNA with charged capsid residues - both flexible and crystallized - are primarily responsible for these phenomena and an essential component of RNA virus assembly.

3.6 Acknowledgements

It is a pleasure to acknowledge Dr. Vladimir Belyi for stimulating discussions. We acknowledge financial support from NIH (Grant No. 2R01HG002776-04A1), the National Science Foundation (Grant No. DMR-0605833), and the Materials Research Science and Engineering Centre at the University of Massachusetts, Amherst.

3.7 Figures

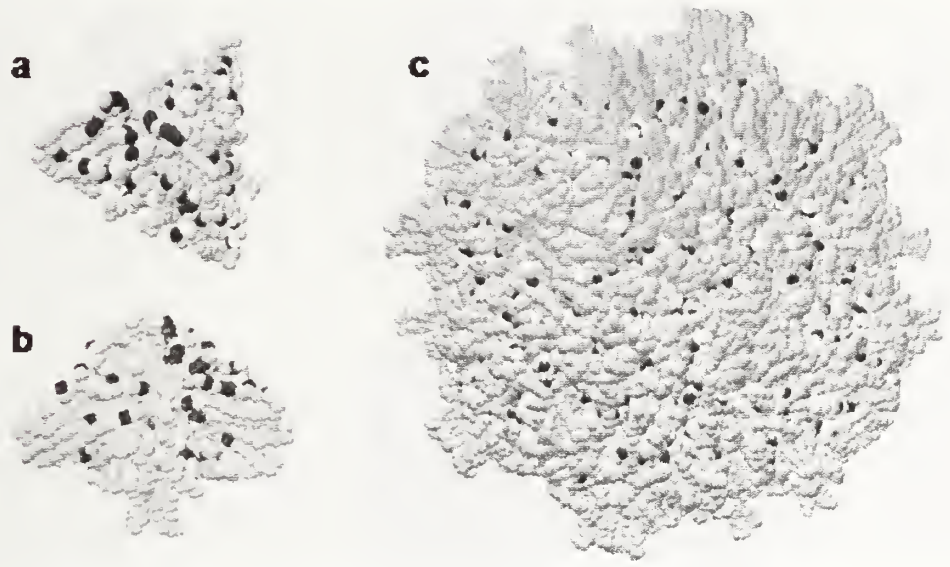

Figure 2-1 Model of crystallized portion of capsid.
Each bead represents a single amino acid residue. Uncharged residues are grey, positively charged residues black, and negatively charged residues white. The asymmetric unit is shown from two perspectives revealing a) the triangular aspect is apparent and b) the stem-like protuberence . c) External view of the full capsid structure obtained through icosahedral transformations to the single asymmetric unit.

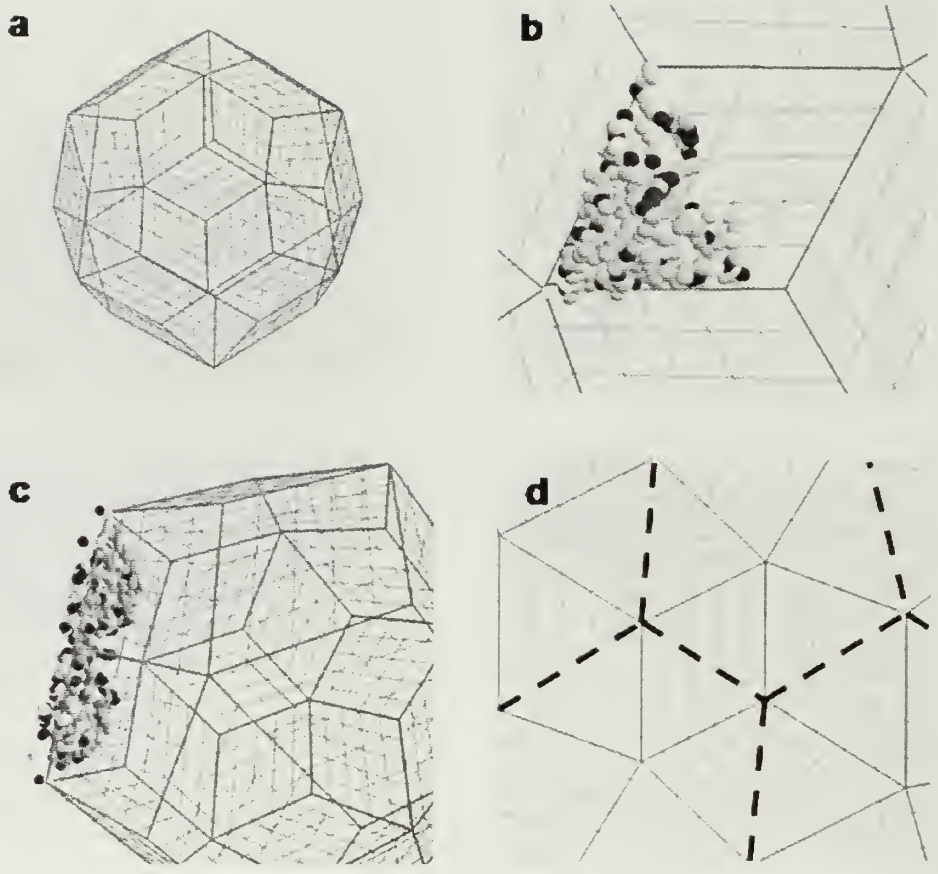

Figure 2-2 Rhombic triacontahedron with topographical capsid beads.
a) the polygon structure associated with T=3 quasi-equivalence. 30 diamond-shaped facets are visible. b)The relationship of a single asymmetric unit to the T=3 symmetry. c) Exterior view of icontahedron with 2 abutting asymmetric units. Charged beads exterior to the container are included for their electrostatic interaction with packaged beads. A topographical valley is visible between the two abutting asymmetric units, connecting two 3-fold axes of symmetry. The valley is rich in positive charge. d) Dotted lines represent the topographical valleys. The lines form the edges of a dodecahedron.

Figure 2-3 Image of capsid beads and flexible N-terminal arms.
This simulation snapshot represents the environment of the genome RNA. The half of the capsid nearest the viewer is not shown to allow for this interior perspective. The flexible nature of the protein arms is apparent.

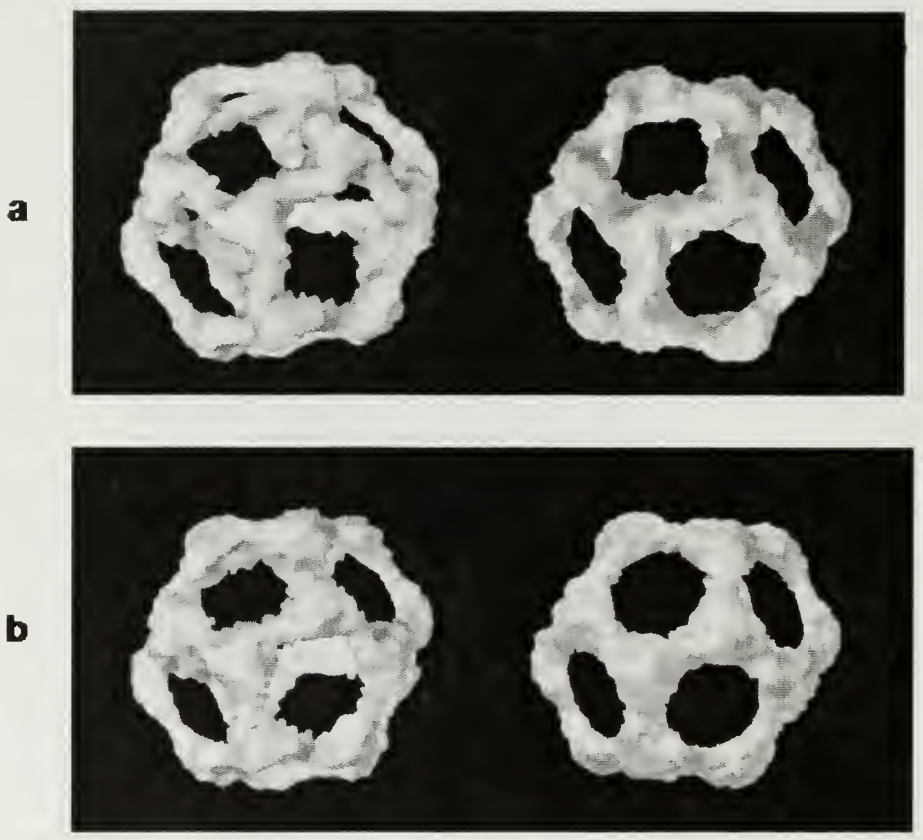

Figure 2-4 RNA density isosurface plots.
RNA density was averaged over time and icosahedrally-symmetric positions. The openDX data viewing program was used to visualize the 3-D averaged-density fields. The images in a) were obtained from a system in which the N-terminal flexible capsid tails were absent. Images in b) were obtained from the full system, including flexible tails. In both cases, similar dodecahedral cages were obtained.

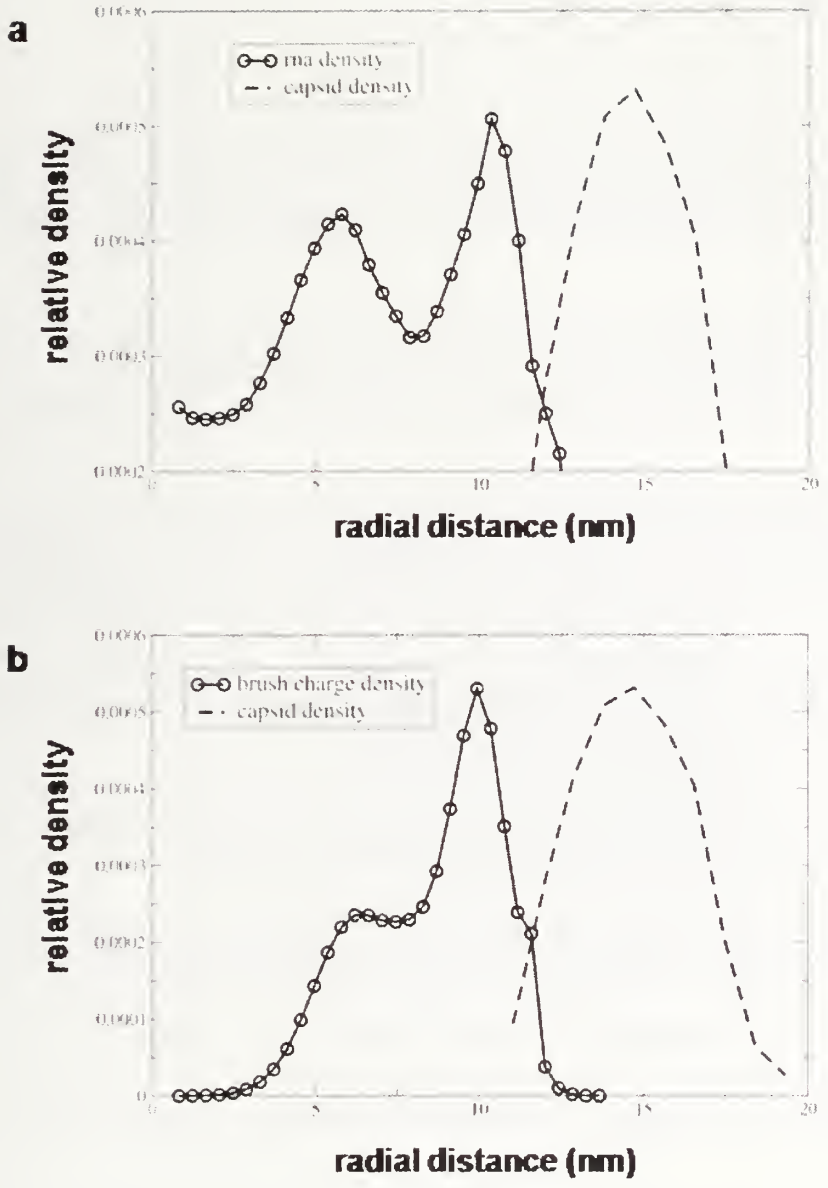

Figure 2-5 Radial density profiles.
Averaged radial density of encapsidated RNA (a) and charges on flexible N-terminal arms (b). Distance is given in nm from the capsid center. Capsid density is shown in both (a) and (b) as a dotted line. Two peaks are evident in RNA density, whereas brush charge density features a prominent peak and a secondary shoulder of density towards the capsid center.

3.8 References

1. van den Worm, S.; Koning, R.; Warmenhoven, H.; Koerten, H.; van Duin, J., Cryo electron microscopy reconstructions of the Leviviridae unveil the densest icosahedral RNA packing possible. *Journal of Molecular Biology* **2006,** 363, (4), 558-565.

2. Opalka, N.; Tihova, M.; Brugidou, C.; Kumar, A.; Beachy, R.; Fauquet, C.; Yeager, M., Structure of native and expanded sobemoviruses by electron cryo-microscopy and image reconstruction. *Journal of Molecular Biology* **2000,** 303, (2), 197-211.

3. Koning, R.; van den Worm, S.; Plaisier, J.; van Duin, J.; Abrahams, J.; Koerten, H., Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2. *Journal of Molecular Biology* **2003,** 332, (2), 415-422.

4. Chen, Z.; Stauffacher, C.; Li, Y.; Schmidt, T.; Wu, B.; Kamer, G.; Shanks, M.; Lomonossoff, G.; Johnson, J., Protein-Rna Interactions In An Icosahedral Virus At 3.0-A Resolution. *Science* **1989,** 245, (4914), 154-159.

5. Fisher, A.; Johnson, J., Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature* **1993,** 361, (6408), 176-179.

6. Larson, S.; Lucas, R.; Greenwood, A.; McPherson, A., The RNA of turnip yellow mosaic virus exhibits icosahedral order. *Virology* **2005,** 334, (2), 245-254.

7. Cann, A., *Principles of Molecular Virology*. 4 ed.; Academic Press: 2005.

8. Knipe, D.; Howley, P., Fundamental Virology. 4 ed.; Lippincott Williams & Wilkins: 2001.

9. Rossman, M.; Arisaka, F.; Battisti, A.; Bowman, V.; Chipman, P., From structure of the complex to understanding of the biology. *Acta Crystallographica Section D Biological Crystallography* **2007,** D63, 9-16.

10. Schneemann, A., The structural and functional role of RNA in icosahedral virus assembly. *Annual Review of Microbology* **2006,** 60, 51-67.

11. Freddolino, P.; Arkhipov, A.; Larson, S.; McPherson, A.; Schulten, K., Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **2006,** 14, (3), 437-449.

12. Zhang, D.; Konecny, R.; Baker, N.; McCammon, J., Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. *Biopolymers***2004,** 75, (4), 325-337.

13. Konecny, R.; Trylska, J.; Tama, F.; Zhang, D.; Baker, N.; Brooks, C.; McCammon, J., Electrostatic properties of cowpea chlorotic mottle virus and cucumber mosaic virus capsids. *Biopolymers* **2006,** 82, (2), 106-120.

14. Forrey, C.; Muthukumar, M., Langevin dynamics simulations of genome packing in bacteriophage. *Biophysical Journal* **2006,** 91, (1), 25-41.

15. Belyi, V.; Muthukumar, M., Electrostatic origin of the genome packing in viruses. *Proceedings of the National Academies of Sciences of the United States of America* **2006,** 103, (46), 17174-17178.

16. van der Schoot, P.; Bruinsma, R., Electrostatics and the assembly of an RNA virus. *PHYSICAL REVIEW E* **2005,** 71, (6), -.

17. Tihova, M.; Dryden, K.; Le, T.; Harvey, S.; Johnson, J.; Yeager, M.; Schneemann, A., Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *Journal of Virology* **2004,** 78, (6), 2897-2905.

18. Larson, S.; Koszelak, S.; Day, J.; Greenwood, A.; Dodds, J.; Mcpherson, A., 3-Dimensional Structure Of Satellite Tobacco Mosaic-Virus At 2.9 Angstrom Resolution. *Journal of Molecular Biology* **1993,** 231, (2), 375-391.

19. Larson, S.; Koszelak, S.; Day, J.; Greenwood, A.; Dodds, J.; Mcpherson, A., Double-Helical Rna In Satellite Tobacco Mosaic-Virus. *NATURE* **1993,** 361, (6408), 179-182.

20. Johnson, K.; Tang, L.; Johnson, J.; Ball, L., Heterologous RNA encapsidated in pariacoto virus-like particles forms a dodecahedral cage similar to genomic RNA in wild-type virions. *Journal of Virology* **2004,** 78, (20), 11371-11378.

21. Tang, L.; Johnson, K.; Ball, L.; Lin, T.; Yeager, M.; Johnson, J., The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nature Structural Materials* **2001,** 8, (1), 77-83.

22. Xia, Q.; Jakana, J.; Zhang, J.; Zhou, Z., Structural comparisons of empty and full cytoplasmic polyhedrosis virus - Protein-RNA interactions and implications for endogenous RNA transcription mechanism. *Journal of Biological Chemistry* **2003,** 278, (2), 1094-1100.

23. Prasad, B.; Rothnagel, R.; Zeng, C.; Jakana, J.; Lawton, J.; Chiu, W.; Estes, M., Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. *Nature* **1996,** 382, (6590), 471-473.

BIBLIOGRAPHY

Akeson, M.; Branton, D.; Kasianowicz, J. J.; Brandin, E.; Deamer, D. W., Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophysical Journal* **1999**, 77, (6), 3227-3233.

Aksimentiev, A.; Schulten, K., Imaging alpha-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophysical Journal* **2005**, 88, (6), 3745-3761.

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson, *Molecular Biology of the Cell*. 4th ed.; Garland Science: New York, 2002.

Ambjornsson, T.; Apell, S. P.; Konkoli, Z.; Di Marzio, E. A.; Kasianowicz, J. J., Charged polymer membrane translocation. *Journal of Chemical Physics* **2002**, 117, (8), 4063-4073.

Arsuaga, J.; Tan, R. K. Z.; Vazquez, M.; Sumners, D. W.; Harvey, S. C., Investigation of viral DNA packaging using molecular mechanics models. *Biophysical Chemistry* **2002,** 101, 475-484.

Aubrey, K. L.; Casjens, S. R.; Thomas, G. J., Studies Of Virus Structure By Laser Raman-Spectroscopy .38. Secondary Structure And Interactions Of The Packaged Dsdna Genome Of Bacteriophage-P22 Investigated By Raman Difference Spectroscopy. *Biochemistry* **1992,** 31, (47), 11835-11842.

Belyi, V.; Muthukumar, M., Electrostatic origin of the genome packing in viruses. *Proceedings of the National Academies of Sciences of the United States of America* **2006,** 103, (46), 17174-17178.

Berezhkovskii, A. M.; Gopich, I. V., Translocation of rodlike polymers through membrane channels. *Biophysical Journal* **2003**, 84, (2), 787-793.

Black, L. W., Dna Packaging In Dsdna Bacteriophages. *Annual Review Of Microbiology* **1989,** 43, 267-292.

Bloomfield, V. A., DNA condensation. *Current Opinion In Structural Biology* **1996,** 6, (3), 334-341.

Booy, F. P.; Newcomb, W. W.; Trus, B. L.; Brown, J. C.; Baker, T. S.; Steven, A. C., Liquid-Crystalline, Phage-Like Packing Of Encapsidated Dna In Herpes-Simplex Virus. *Cell* **1991,** 64, (5), 1007-1015.

Butler, T. Z.; Gundlach, J. H.; Troll, M. A., Determination of RNA orientation during translocation through a biological nanopore. *Biophysical Journal* **2006**, 90, (1), 190-199.

Cann, A., *Principles of Molecular Virology*. 4 ed.; Academic Press: 2005.

Cerritelli, M. E.; Cheng, N. Q.; Rosenberg, A. H.; McPherson, C. E.; Booy, F. P.; Steven, A. C., Encapsidated conformation of bacteriophage T7 DNA. *Cell* **1997,** 91, (2), 271-280.

Chen, P.; Gu, J. J.; Brandin, E.; Kim, Y. R.; Wang, Q.; Branton, D., Probing single DNA molecule transport using fabricated nanopores. *Nano Letters* **2004**, 4, (11), 2293-2298.

Chen, Z.; Stauffacher, C.; Li, Y.; Schmidt, T.; Wu, B.; Kamer, G.; Shanks, M.; Lomonossoff, G.; Johnson, J., Protein-Rna Interactions In An Icosahedral Virus At 3.0-A Resolution. *Science* **1989,** 245, (4914), 154-159.

Deamer, D. W.; Branton, D., Characterization of nucleic acids by nanopore analysis. *Accounts of Chemical Research* **2002**, 35, (10), 817-825.

DiMarzio, E. A.; Mandell, A. J., Phase transition behavior of a linear macromolecule threading a membrane. *Journal of Chemical Physics* **1997**, 107, (14), 5510-5514.

Earnshaw, W.; Casjens, S.; Harrison, S. C., Assembly Of Head Of Bacteriophage P22 - X-Ray-Diffraction From Heads, Proheads And Related Structures. *Journal Of Molecular Biology* **1976,** 104, (2), 387-410.

Earnshaw, W. C.; King, J.; Harrison, S. C.; Eiserling, F. A., Structural Organization Of Dna Packaged Within Heads Of T4 Wild-Type, Isometric And Giant Bacteriophages. *Cell* **1978,** 14, (3), 559-568.

Fisher, A.; Johnson, J., Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature* **1993,** 361, (6408), 176-179.

Forrey, C.; Muthukumar, M., Langevin dynamics simulations of genome packing in bacteriophage. *Biophysical Journal* **2006**, 91, (1), 25-41.

Freddolino, P.; Arkhipov, A.; Larson, S.; McPherson, A.; Schulten, K., Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **2006,** 14, (3), 437-449.

Fyta, M. G.; Melchionna, S.; Kaxiras, E.; Succi, S., Multiscale coupling of molecular dynamics and hydrodynamics: Application to DNA translocation through a nanopore. *Multiscale Modeling & Simulation* **2006**, 5, (4), 1156-1173.

Heng, J. B.; Ho, C.; Kim, T.; Timp, R.; Aksimentiev, A.; Grinkova, Y. V.; Sligar, S.; Schulten, K.; Timp, G., Sizing DNA using a nanometer-diameter pore. *Biophysical Journal* **2004**, 87, (4), 2905-2911.

Howorka, S.; Cheley, S.; Bayley, H., Sequence-specific detection of individual DNA strands using engineered nanopores. *Nature Biotechnology* **2001**, 19, (7), 636-639.

Hud, N. V., Double-Stranded Dna Organization In Bacteriophage Heads - An Alternative Toroid-Based Model. *Biophysical Journal* **1995,** 69, (4), 1355-1362.

Hud, N. V.; Downing, K. H., Cryoelectron microscopy of lambda phage DNA condensates in vitreous ice: The fine structure of DNA toroids. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **2001,** 98, (26), 14925-14930.

Hud, N. V.; Downing, K. H.; Balhorn, R., A Constant Radius Of Curvature Model For The Organization Of Dna In Toroidal Condensates. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **1995,** 92, (8), 3581-3585.

Johnson, K.; Tang, L.; Johnson, J.; Ball, L., Heterologous RNA encapsidated in pariacoto virus-like particles forms a dodecahedral cage similar to genomic RNA in wild-type virions. *Journal of Virology* **2004,** 78, (20), 11371-11378.

Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W., Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, 93, (24), 13770-13773.

Kasianowicz, J. J.; Henrickson, S. E.; Weetall, H. H.; Robertson, B., Simultaneous multianalyte detection with a nanometer-scale pore. *Analytical Chemistry* **2001**, 73, (10), 2268-2272.

Kasianowicz, J. J., Mechanism of ionic current blockades during polymer transport through pores of nanometer dimensions. In *Structure and Dynamics of Confined Polymers*, Ed. Springer: 1999; Vol. 87.

Kindt, J.; Tzlil, S.; Ben-Shaul, A.; Gelbart, W. M., DNA packaging and ejection forces in bacteriophage. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **2001,** 98, (24), 13671-13674.

Knipe, D.; Howley, P., Fundamental Virology. 4 ed.; Lippincott Williams & Wilkins: 2001.

Konecny, R.; Trylska, J.; Tama, F.; Zhang, D.; Baker, N.; Brooks, C.; McCammon, J., Electrostatic properties of cowpea chlorotic mottle virus and cucumber mosaic virus capsids. *Biopolymers* **2006,** 82, (2), 106-120.

Kong, C. Y.; Muthukumar, M., Polymer translocation through a nanopore. II. Excluded volume effect. *Journal of Chemical Physics* **2004**, 120, (7), 3460-3466.

Koning, R.; van den Worm, S.; Plaisier, J.; van Duin, J.; Abrahams, J.; Koerten, H., Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2. *Journal of Molecular Biology* **2003,** 332, (2), 415-422.

Larson, S.; Koszelak, S.; Day, J.; Greenwood, A.; Dodds, J.; Mcpherson, A., 3-Dimensional Structure Of Satellite Tobacco Mosaic-Virus At 2.9 Angstrom Resolution. *Journal of Molecular Biology* **1993,** 231, (2), 375-391.

Larson, S.; Koszelak, S.; Day, J.; Greenwood, A.; Dodds, J.; Mcpherson, A., Double-Helical RNA In Satellite Tobacco Mosaic-Virus. *NATURE* **1993,** 361, (6408), 179-182.

Larson, S.; Lucas, R.; Greenwood, A.; McPherson, A., The RNA of turnip yellow mosaic virus exhibits icosahedral order. *Virology* **2005,** 334, (2), 245-254.

Lepault, J.; Dubochet, J.; Baschong, W.; Kellenberger, E., Organization Of Double-Stranded Dna In Bacteriophages - A Study By Cryoelectron Microscopy Of Vitrified Samples. *Embo Journal* **1987,** 6, (5), 1507-1512.

Li, J.; Stein, D.; McMullan, C.; Branton, D.; Aziz, M. J.; Golovchenko, J. A., Ion-beam sculpting at nanometre length scales. *Nature* **2001**, 412, (6843), 166-169.

Li, J. L.; Gershow, M.; Stein, D.; Brandin, E.; Golovchenko, J. A., DNA molecules and configurations in a solid-state nanopore microscope. *Nature Materials* **2003**, 2, (9), 611-615.

Liu, S.; Ghosh, K.; Muthukumar, M., Polyelectrolyte solutions with added salt: A simulation study. *Journal of Chemical Physics* **2003**, 119, (3), 1813-1823.

Liu, S.; Muthukumar, M., Langevin dynamics simulation of counterion distribution around isolated flexible polyelectrolyte chains. *Journal of Chemical Physics* **2002**, 116, (22), 9975-9982.

Lubensky, D. K.; Nelson, D. R., Driven polymer translocation through a narrow pore. *Biophysical Journal* **1999**, 77, (4), 1824-1838.

Manning, G. S., Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions .I. Colligative Properties. *Journal of Chemical Physics* **1969**, 51, (3), 924.

Mathe, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Meller, A., Nanopore unzipping of individual DNA hairpin molecules. *Biophysical Journal* **2004**, 87, (5), 3205-3212.

Meller, A., Dynamics of polynucleotide transport through nanometre-scale pores. *Journal of Physics-Condensed Matter* **2003**, 15, (17), R581-R607.

Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D., Rapid nanopore discrimination between single polynucleotide molecules. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, 97, (3), 1079-1084.

Meller, A.; Nivon, L.; Branton, D., Voltage-driven DNA translocations through a nanopore. *Physical Review Letters* **2001**, 86, (15), 3435-3438.

Murphy, R. J.; Muthukumar, M., Threading synthetic polyelectrolytes through protein pores *Journal of Chemical Physics* **2007**, 126, (5), 1-4.

Muthukumar, M., Polymer translocation through a hole. *Journal Of Chemical Physics* **1999**, 111, (22), 10371-10374.

Muthukumar, M., Polymer translocation through a hole. *Journal of Chemical Physics* **1999**, 111, (22), 10371-10374.

Muthukumar, M., Translocation of a confined polymer through a hole. *Physical Review Letters* **2001**, 86, (14), 3188-3191.

Odijk, T., Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. *Biophysical Journal* **1998**, 75, (3), 1223-1227.

Olson, N. H.; Gingery, M.; Eiserling, F. A.; Baker, T. S., The structure of isometric capsids of bacteriophage T4. *Virology* **2001**, 279, (2), 385-391.

Opalka, N.; Tihova, M.; Brugidou, C.; Kumar, A.; Beachy, R.; Fauquet, C.; Yeager, M., Structure of native and expanded sobemoviruses by electron cryo-microscopy and image reconstruction. *Journal of Molecular Biology* **2000**, 303, (2), 197-211.

Poy, S., *Molecular Biology of Bacteriophage T4* American Society for Microbiology: Washington, DC, 1994.

Prasad, B.; Rothnagel, R.; Zeng, C.; Jakana, J.; Lawton, J.; Chiu, W.; Estes, M., Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. *Nature* **1996**, 382, (6590), 471-473.

Purohit, P. K.; Inamdar, M. M.; Grayson, P. D.; Squires, T. M.; Kondev, J.; Phillips, R., Forces During Bacteriophage DNA Packaging and Ejection. *Biophysical Journal* **2005**, 88, 851-866.

Purohit, P. K.; Kondev, J.; Phillips, R., Mechanics of DNA packaging in viruses. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **2003**, 100, (6), 3173-3178.

Richards, K. E.; Williams, R. C.; Calendar, R., Mode Of Dna Packing Within Bacteriophage Heads. *Journal Of Molecular Biology* **1973**, 78, (2), 255-&.

Rossman, M.; Arisaka, F.; Battisti, A.; Bowman, V.; Chipman, P., From structure of the complex to understanding of the biology. *Acta Crystallographica Section D Biological Crystallography* **2007,** D63, 9-16.

Saleh, O. A.; Sohn, L. L., An artificial nanopore for molecular sensing. *Nano Letters* **2003**, 3, (1), 37-38.

Sauer-Budge, A. F.; Nyamwanda, J. A.; Lubensky, D. K.; Branton, D., Unzipping kinetics of double-stranded DNA in a nanopore. *Physical Review Letters* **2003**, 90, 23.

Schneemann, A., The structural and functional role of RNA in icosahedral virus assembly. *Annual Review of Microbology* **2006,** 60, 51-67.

Serwer, P., Internal Proteins Of Bacteriophage-T7. *Journal Of Molecular Biology* **1976,** 107, (3), 271-291.

Serwer, P.; Hayes, S. J.; Watson, R. H., Conformation Of Dna Packaged In Bacteriophage-T7 - Analysis By Use Of Ultraviolet Light-Induced Dna-Capsid Cross-Linking. *Journal Of Molecular Biology* **1992,** 223, (4), 999-1011.

Siwy, Z.; Fulinski, A., Fabrication of a synthetic nanopore ion pump. *Physical Review Letters* **2002**, 89, 198103.

Slonkina, E.; Kolomeisky, A. B., Polymer translocation through a long nanopore. *Journal of Chemical Physics* **2003**, 118, (15), 7112-7118.

Smith, D. E.; Tans, S. J.; Smith, S. B.; Grimes, S.; Anderson, D. L.; Bustamante, C., The bacteriophage phi 29 portal motor can package DNA against a large internal force. *Nature* **2001,** 413, (6857), 748-752.

Song, L. Z.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E., Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **1996**, 274, (5294), 1859-1866.

Spakowitz, A. J.; Wang, Z. G., DNA packaging in bacteriophage: Is twist important? *Biophysical Journal* **2005,** 88, (6), 3912-3923.

Stevens, M. J.; Kremer, K., The Nature Of Flexible Linear Polyelectrolytes In Salt-Free Solution - A Molecular-Dynamics Study. *Journal Of Chemical Physics* **1995,** 103, (4), 1669-1690.

Storm, A. J.; Chen, J. H.; Ling, X. S.; Zandbergen, H. W.; Dekker, C., Fabrication of solid-state nanopores with single-nanometre precision. *Nature Materials* **2003**, 2, (8), 537-540.

Storm, A. J.; Storm, C.; Chen, J. H.; Zandbergen, H.; Joanny, J. F.; Dekker, C., Fast DNA translocation through a solid-state nanopore. *Nano Letters* **2005**, 5, (7), 1193-1197.

Sung, W.; Park, P. J., Polymer translocation through a pore in a membrane. *Physical Review Letters* **1996**, 77, (4), 783-786.

Tang, L.; Johnson, K.; Ball, L.; Lin, T.; Yeager, M.; Johnson, J., The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nature Structural Materials* **2001,** 8, (1), 77-83.

Tihova, M.; Dryden, K.; Le, T.; Harvey, S.; Johnson, J.; Yeager, M.; Schneemann, A., Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *Journal of Virology* **2004,** 78, (6), 2897-2905.

Tildesley, M. P. A. a. D. J., *Computer simulation of liquids*. Clarendon Press: oxford, **1987**.

Tzlil, S.; Kindt, J. T.; Gelbart, W. M.; Ben-Shaul, A., Forces and pressures in DNA packaging and release from viral capsids. *Biophysical Journal* **2003,** 84, (3), 1616-1627.

van der Schoot, P.; Bruinsma, R., Electrostatics and the assembly of an RNA virus. *PHYSICAL REVIEW E* **2005,** 71, (6), -.

van den Worm, S.; Koning, R.; Warmenhoven, H.; Koerten, H.; van Duin, J., Cryo electron microscopy reconstructions of the Leviviridae unveil the densest icosahedral RNA packing possible. *Journal of Molecular Biology* **2006,** 363, (4), 558-565.

Wang, H.; Dunning, J. E.; Huang, A. P. H.; Nyamwanda, J. A.; Branton, D., DNA heterogeneity and phosphorylation unveiled by single-molecule electrophoresis. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, 101, (37), 13472-13477.

Wong, C. T. A. M., M., Polymer capture by electro-osmotic flow of oppositely charged nanopores. *Journal of Chemical Physics* **2007**, 126, 164903 -164908.

Xia, Q.; Jakana, J.; Zhang, J.; Zhou, Z., Structural comparisons of empty and full cytoplasmic polyhedrosis virus - Protein-RNA interactions and implications for endogenous RNA transcription mechanism. *Journal of Biological Chemistry* **2003,** 278, (2), 1094-1100.

Yamakawa, H., *Modern theory of polymer solutions*. Harper & Row: New York, 1971.

Zhang, D.; Konecny, R.; Baker, N.; McCammon, J., Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. *Biopolymers***2004,** 75, (4), 325-337.