

Article

# Simulations of Phage T7 Capsid Expansion Reveal the Role of Molecular Sterics on Dynamics

Paul C. Whitford<sup>1,2,\*</sup>, Wen Jiang<sup>3</sup>  and Philip Serwer<sup>4</sup> <sup>1</sup> Department of Physics, Northeastern University, Boston, MA 02115, USA<sup>2</sup> Center for Theoretical Biological Physics, Northeastern University, Boston, MA 02115, USA<sup>3</sup> Department of Biological Sciences, Markey Center for Structural Biology, Purdue University, West Lafayette, IN 47906, USA; jiang12@purdue.edu<sup>4</sup> Department of Biochemistry and Structural Biology, The University of Texas Health Center, San Antonio, TX 78229-3900, USA; serwer@uthscsa.edu

\* Correspondence: p.whitford@neu.edu

Received: 26 October 2020; Accepted: 4 November 2020; Published: 7 November 2020



**Abstract:** Molecular dynamics techniques provide numerous strategies for investigating biomolecular energetics, though quantitative analysis is often only accessible for relatively small (frequently monomeric) systems. To address this limit, we use simulations in combination with a simplified energetic model to study complex rearrangements in a large assembly. We use cryo-EM reconstructions to simulate the DNA packaging-associated 3 nm expansion of the protein shell of an initially assembled phage T7 capsid (called procapsid or capsid I). This is accompanied by a disorder–order transition and expansion-associated externalization displacement of the 420 N-terminal tails of the shell proteins. For the simulations, we use an all-atom structure-based model (1.07 million atoms), which is specifically designed to probe the influence of molecular sterics on dynamics. We find that the rate at which the N-terminal tails undergo translocation depends heavily on their position within hexons and pentons. Specifically, trans-shell displacements of the hexon E subunits are the most frequent and hexon A subunits are the least frequent. The simulations also implicate numerous tail translocation intermediates during tail translocation that involve topological traps, as well as sterically induced barriers. The presented study establishes a foundation for understanding the precise relationship between molecular structure and phage maturation.

**Keywords:** bacteriophage T7; disorder; DNA packaging; molecular dynamics simulation; protein dynamics

## 1. Introduction

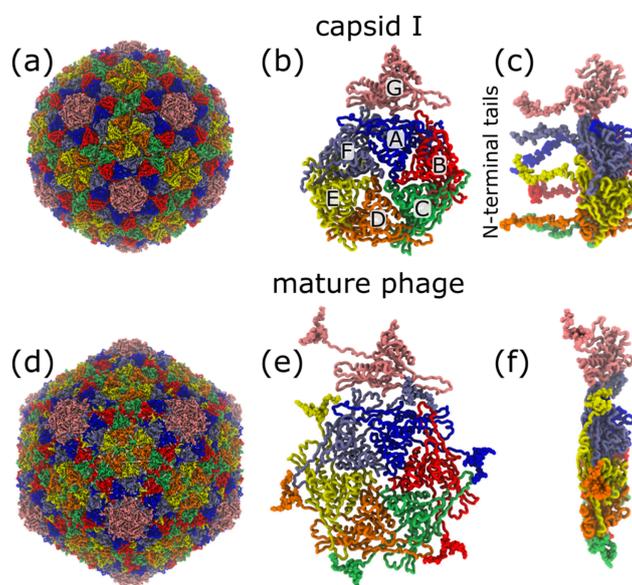
While a major objective in molecular biophysics research is to establish physical descriptions of energy-requiring, motor-like biological processes, distilling the complex nature of the dynamics is a continual challenge. For comparison, physical descriptions of macroscopic motors are based on well-controlled repetitive motions. For such systems, a deterministic view is appropriate to describe the energetics, for example, when determining the maximum efficiency of an ideal motor. However, for biochemical, nanometer-scale motors, the physical description is complicated by the intrinsic stochasticity of microscopic motion. Although cycles repeat, on average, details of each event vary due to Brownian noise that is introduced from thermal energy. This naturally leads to biomolecules exhibiting a significant degree of flexibility [1–3], where individual molecules have even been described as “forever kicking and screaming” [4].

While flexibility of motor components helps define biology, this flexibility also introduces difficulties in the analysis of dynamics. However, these challenges can be reduced when the motor

components are assembled from many identical molecules that arrange symmetrically. In those cases, symmetry can amplify the data available for determining structures, which are necessary for the application of molecular simulations. The nucleic acid-containing protein shells (capsids) of viruses represent one such class of systems [5–7]. Analysis of viral capsid structure, by use of both X-ray crystallography and cryo-electron microscopy (cryo-EM), was a major contribution of the laboratory of Dr. Michael Rossmann [5,8–10], to whom the current symposium is dedicated. Here, we build upon his seminal contributions to begin the detailed analysis of dynamics in these complex systems.

Historically, our philosophical background begins with the projection that new physical principles might be discovered in biology [11]. Within biology, countless challenging physical questions have surfaced through the years. For example, conventional biochemical thought (i.e., Levinthal’s paradox [12]) initially made it difficult to understand the origins of rapid protein folding. However, the adoption of condensed matter physics principles enabled systematic investigation into the many energetic factors that influence folding dynamics [13–15]. Another outstanding challenge is how to understand the duality of power stroke vs. thermal ratchet mechanisms (reviewed in [16,17]). One could also look to larger length scales and inquire about the organizational principles of chromosomes [18,19], or seek to understand the properties that give rise to membrane-less organelles in the cell [20]. Each of these biological phenomena pose distinct challenges, where new physical principles have frequently been proposed and tested, in order to establish precise quantitative insights.

Here, we study phage T7 capsids and explore the structural dynamics (Figure 1) that accompany packaging of DNA (reviewed in [21,22]). Starting with a DNA-free capsid I, packaging is associated with (1) activity of a DNA packaging-driving ATPase and (2) expansion and structural change of the capsid, ultimately to form a structure called capsid II that, with minor changes, becomes the capsid of the phage (Figure 1) [22–24]. When describing the capsid, the proteins are labeled gp (gene product), followed by gene number [25,26]. The capsid I, capsid II and phage shells are composed of 420 gp10 proteins, and capsid structures have been previously determined by cryo-EM to 4.6, 3.5 and 3.6 Å resolution, respectively [23].



**Figure 1.** Expansion during phage maturation involves ordering of the N-terminal tails of gp10. (a) Cryo-EM structure of T7 capsid I (PDB:3J7V) [23]. (b) Exterior-shell view of asymmetric subunit in capsid I conformation. Monomer positions labelled A–F. (c) Side view of panel b. Modeled gp10 N-terminal tails (vdW representation) are on the interior of the shell. (d) Cryo-EM structure of the mature phage (PDB:3J7X) [23]. (e) Exterior-shell view of asymmetric subunit. (f) Side view of panel e. All structural representations are generated with VMD [27].

In the current study, we focus simulations on the intermediate states of the disordered 23 amino acid, N-terminal gp10 tail (Figure 1). This tail moves from the inner surface of capsid I to the outer surface when capsid I expands to capsid II and phage capsid [23]. To elucidate the influence of molecular sterics on capsid maturation, we apply simulations to directly study the timing of N-terminal tail translocation and shell expansion. This shows how computational and theoretical techniques can now be leveraged to study global conformational motions of a complete gp10 capsid shell.

In the simulations, we apply a “structure-based” model [28], in which all non-hydrogen atoms are explicitly represented (1.07 million atoms). These models have been used extensively to study global rearrangements in other large-scale assemblies, including the ribosome [29] and HIV capsid [30]. While coarse-grained variants of this model have been applied to investigate structural fluctuations of the HK97 capsid shell [31], to the best of our knowledge, there have not been simulations reported of spontaneous (i.e., without the use of targeting methods) expansion events that include ordering of the terminal tails. Accordingly, the presented simulations demonstrate the feasibility of all-atom molecular dynamics simulations to characterize coordinated rearrangements within a complete capsid. These simulations additionally reveal a complex interplay of expansion, steric interactions and order–disorder phenomena that accompanies the maturation process. Taken together, these observations suggest new strategies by which biochemical and structural techniques may obtain more precise insights into DNA packaging-driven dynamics.

## 2. Materials and Methods

### 2.1. Structure of Phage T7 and its Capsid I and Capsid II

The structures of phage T7, capsid I (procapsid) and capsid II are described in reference [23]. Since the N-terminal tails (residues 1–23) were not resolved in the capsid I state, we generated disordered initial conformations. To generate a random ensemble of disordered conformations for each of the 420 tails, a high-temperature simulation (reduced temperature of 1) was initiated from the capsid I conformation, with non-tail residues constrained to their capsid I positions. Specifically, an all-atom structure-based model [28,30] was defined for the capsid I structure, while all non-tail atoms were constrained to their initial positions. By performing a high-temperature simulation, any bias of the force field is negligible, such that the tails sample a fully disordered range of conformations. Four randomly selected disordered conformations were used to initialize simulations of capsid expansion.

### 2.2. Force Field Description

For the presented simulations, an all-atom structure-based “SMOG” force field [28,30] was defined based on the structure of the mature phage gp10 shell (PDB:3J7X) [23]. As described elsewhere [30], all non-bonded atomic contacts (identified with the Shadow algorithm [32]) and flexible dihedral interactions are set to specifically stabilize the expanded conformation. The interaction distance of each contact was scaled by 0.95, similar to applications of this model to the ribosome [33]. The purpose of this modification is to avoid artificial expansion beyond that found in the mature phage conformation. This ensures that the free energy as a function of the radius of gyration has a minimum that coincides with the value found in the mature phage structure. The only differences from the previous implementation of the model are in the bonded interactions. The bond lengths and angles are assigned the values found in the Amber ff03 force field [34], and rigid dihedral angles (e.g., hybridized, planar) are assigned a cosine potential of periodicity 2 with minima at 0 and 180 degrees.

### 2.3. Simulation Details

All simulations were performed using the GROMACS software package [35,36] (v5.1.4), and the force field was generated using SMOG 2 [30]. Each simulation was performed at a reduced temperature of 0.582 (70 in GROMACS units), which is roughly 0.6–0.7 times the folding–unfolding temperature of a typical single-domain protein with this model [37]. Further, since physiological

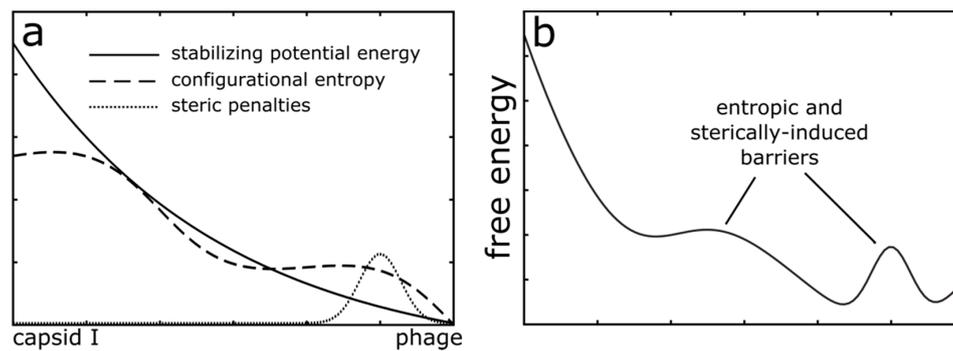
temperature for a given organism is generally  $\sim 0.7$  of the folding temperature of its constituent proteins, this temperature range tends to yield molecular fluctuations that are representative of cellular conditions [37]. Constant temperature was maintained via Langevin Dynamics protocols. A total of 20 simulations were performed for  $1.4\text{--}3.0 \times 10^8$  timesteps of duration 0.002 reduced units ( $3.4 \times 10^9$  timesteps, in aggregate). The duration of each simulation was dependent on available computational resources and shared-resource scheduler policies. Each simulation was performed on 112 compute cores, where an aggregate  $\sim 1.6$  million core-hours was required to perform the set of simulations. To ensure that the results were not dependent on the precise initial conformation, four different randomly selected initial conformations were used, where five simulations were performed for each (20 total simulations). Each simulation was initialized with randomized velocities and Langevin noise. Based on a recent comparison of diffusion coefficients in a SMOG model and an all-atom explicit-solvent model [38], one may estimate the effective simulated time to be approximately 250–500 microseconds, for each simulation.

### 3. Results

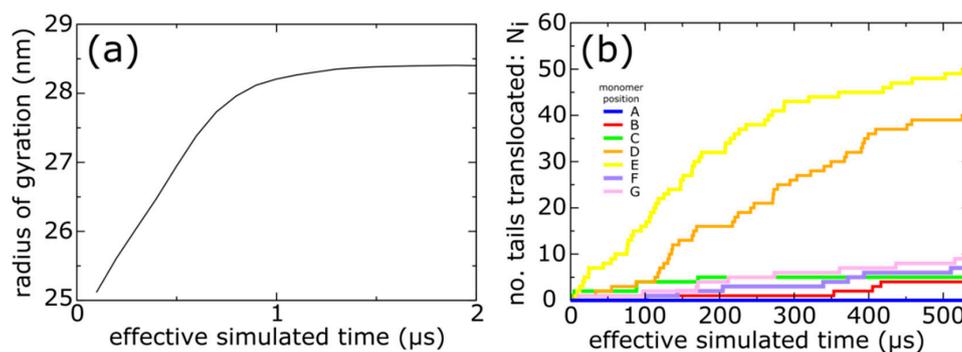
#### 3.1. Molecular Dynamics Simulation of Spontaneous Capsid Expansion

Since our long-term objective is to understand how individual molecular interactions facilitate capsid assembly and maturation, we will begin by asking whether steric effects can limit expansion-associated transitions of the N-terminal tails of gp10. For this, we initiated simulations from the compact capsid I conformation (Figure 1a). Consistent with the lack of density for the N-terminal tails in the cryo-EM reconstruction of the capsid I structure, the N-terminal tails were assigned disordered conformations (see Methods). In each simulation, the system was allowed to move under the influence of a mature phage capsid-specific force field. That is, we applied an all-atom structure-based model [28,30], where every non-bonded interaction in the mature phage conformation was defined to be stabilizing. Based on this construction, the overall potential energy landscape may be considered “downhill” (Figure 2a) in the direction of the mature conformation. Since the expansion process is known to be driven by DNA packaging, the potential energy in this model is intended to describe the *effective energetics* [39–41] of the system. Accordingly, the net effect of DNA packaging is to favor the expanded form, which is implicitly embedded in this model by defining the mature form to be the potential energy minimum. Even though DNA and solvent are not explicitly represented, since the model explicitly represents all non-hydrogen atoms in the capsid (1.07 million atoms), this simplified energetic representation can be used to isolate specific types of factors that can give rise to free-energy barriers. For example, this model can identify barriers that stem from (1) the imperfect structural complementarity of the tail and corresponding cavity, or (2) confinement effects that impose sharp changes in the configurational entropy. Together, these factors can result in a rugged, expansion-favoring free-energy landscape (Figure 2b).

Even though our model explicitly represents every non-hydrogen atom (Figure 1), the simplicity of the structure-based force field allows us to simulate spontaneous (i.e., non-targeted) expansion events (Video S1). During each simulation, the radius of gyration increases from its capsid I value of  $\sim 25$  nm, to  $\sim 28.5$  nm, as the system approaches the mature phage conformation (Figure 3a). If neither sterics nor configurational entropy limit expansion, then simulations with this model would lead to rapid translocation of the N-terminal tails, since the energetics explicitly favor their movement to the shell exterior.



**Figure 2.** Schematic depiction of the modeled energetics. (a) In the current study, MD simulations utilized an all-atom structure-based model [28,30], where the mature phage conformation was defined as the minimum potential energy conformation. Since all stabilizing interactions are defined based on the mature phage conformation, the potential energy is “downhill” in character (solid line). Despite this simplified representation, configurational entropy (large dashed line) and sterically-induced penalties (small dashed line) can arise from the molecular structure and flexibility. (b) Together, the potential energy, sterics and configurational entropy can lead to a complex free-energy landscape that has many barriers.



**Figure 3.** Simulations of spontaneous capsid expansion events reveal that N-terminal tail translocation is position-dependent. (a) Representative time trace of a simulation with an all-atom structure-based model shows that, based on the radius of gyration, the capsid expands during the first few microseconds. (b) The N-terminal tails translocate on longer timescales (hundreds of microseconds). A tail is considered to have completed translocation once 80% of the phage-specific tail contacts are formed. The same simulation is shown as in panel a.

In contrast, we find that the kinetics of trans-shell displacements depend strongly on the location of the gp10 monomer. To monitor the progress of each tail towards the phage conformation, we calculated the fraction of mature-phage contacts of the N-terminal tail (Ala2-Leu25; residue 1 was not resolved in any structure) that are formed as a function of time for each chain:  $Q_i$  ( $i = A-F$  forming the hexon and G is a subunit of the penton). A contact is defined as being “formed” if the distance between the contacting atom pair is less than 1.5 times their distance in the mature phage conformation. The use of “native” (i.e., mature-phage) contacts is motivated by its demonstrated utility for describing protein folding [42]. Based on  $Q_i$ , we then define a chain as having completed shell translocation when at least 80% of the mature-phage-specific intermolecular tail contacts were formed. As a note, since simulations are always performed at finite temperatures, even a fully-folded protein will typically only have 70–90% of the native contacts formed at any given time [28]. In a representative trajectory (Figure 3b), the E monomers were associated with the largest number of translocated tails ( $N_E > 50$ ), and the D position exhibited nearly as many events ( $N_D \sim 40$ ). In stark contrast, we find that shell translocation was less frequent for monomers A–C, F and G. The least likely tails to translocate were those of the A monomers, which lacked any translocation events on the simulated timescale.

With regard to simulated times, the overall expansion process occurred on the microsecond scale (Figure 3a), whereas tail translocation typically required hundreds of microseconds (Figure 3b). Accordingly, the overall picture illustrated by this simulation is that capsid maturation involves an initial global expansion of the shell, followed by trans-shell displacement and ordering of the N-terminal tails. However, due to the simplified character of the model, one should be careful to not overinterpret these timescales. That is, rather than modeling DNA packaging as a time-dependent effect that is proportional to the fraction of packaged DNA, our model simply describes the mature phage as the global potential energy minimum. Accordingly, in solution, T7 expansion will occur on longer timescales that depend on the rate of DNA packaging.

### 3.2. Simulations Implicate Hierarchical Timescales for Trans-Shell Displacements

Inspecting individual simulations, as in Figure 3, can provide anecdotal suggestions into the character of a system. However, biomolecular processes are intrinsically stochastic [43]. Thus, it is necessary to consider statistical properties. To this end, we performed 20 independent simulations of the expansion process, each initialized with randomized velocities (see methods), and then analyzed tail translocation events across this ensemble. The total number of each simulated tail that successfully underwent a trans-shell rearrangement is given in Table 1.

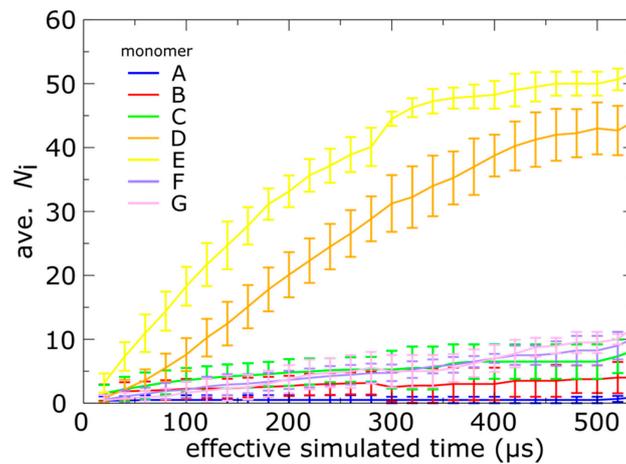
**Table 1.** Total number of N-terminal tails observed to undergo trans-shell displacements. A total of 20 independent simulations were performed (1200 tails in total, per monomer position).

Monomer	Number of Events
A	10
B	70
C	112
D	658
E	845
F	117
G	144

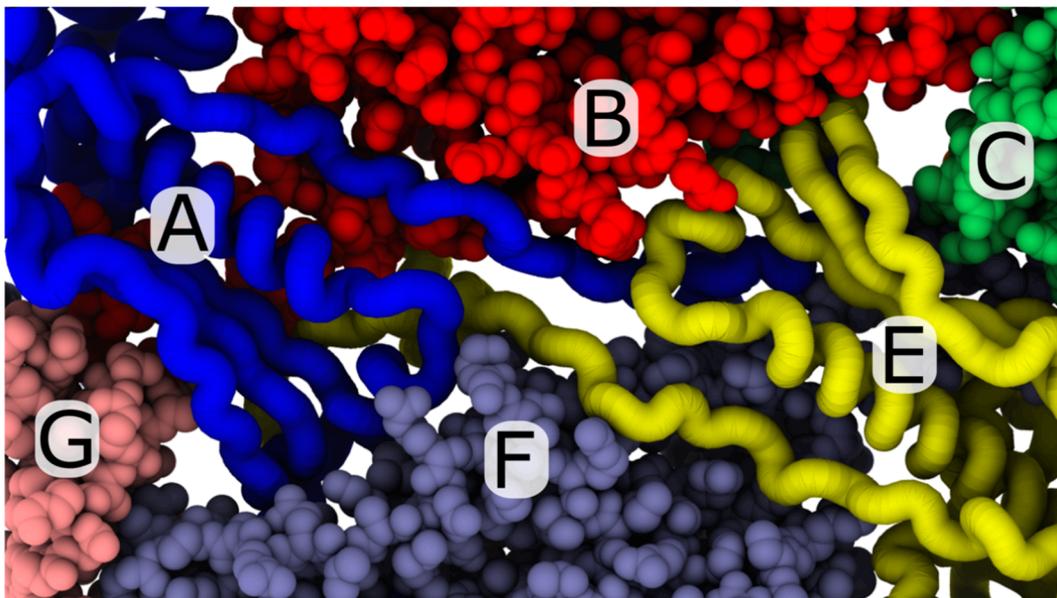
Consistent with the representative time traces in Figure 3b, the tails of monomers E and D were found to undergo trans-shell displacements most frequently, whereas tails B, C, F and G appear to be slower and translocate on comparable timescales (Figure 4). We found that not a single simulation yielded more than two translocated tails in the A position, which is also consistent with the slow dynamics of the A monomers. To identify whether the timescale differences are statistically significant, we also calculated the standard deviation as a function of time (error bars in Figure 4). Consistent with the stochasticity of biomolecular events, the standard deviation is non-negligible, and it ranges from ~1 to 5 tails. However, the differences between the D and E positions can exceed 10 tails, which shows that the position-dependence of the kinetics is preserved across the set of simulations. For monomers B, C, F and G, there is substantial overlap in the sampled  $N_i$  values, indicating that these chains translocate on similar timescales. However, monomer A is consistently the least likely to reach the phage conformation, even when the uncertainty is taken into account.

The difference in timescales between E and A monomers is surprising, because they share a common cavity (Figure 5). To better understand this observation, we calculated the average time for monomers to reach the exterior shell; the average only included successful crossing events. That is, we only considered the 10 and 845 successful trans-shell displacements of A and E monomers. While A monomers were less likely to reach the exterior of the shell, the successful events were found to occur on shorter timescales (1/10 the time) than successful E-monomer transitions. Since only successful events were considered, this differential timescale cannot be used to infer a rate. Instead, one should

interpret it as suggesting the A monomers occasionally reach the exterior rapidly, but conformational changes in the capsid impede motion of tails that do not successfully cross in the initial stage of the simulation.



**Figure 4.** Position-dependent N-terminal dynamics is robust. Average number of translocated tails,  $N_i$ , as a function of simulated time, where the average is taken over 20 independent simulations. Error bars represent one standard deviation, and monomers A–G are shown in different colors. Consistent with the representative trace in Figure 3b, monomers D and E are consistently faster than the remaining tails. Further, monomer A is consistently slowest, and it rarely reaches the exterior of the capsid shell in the simulations.

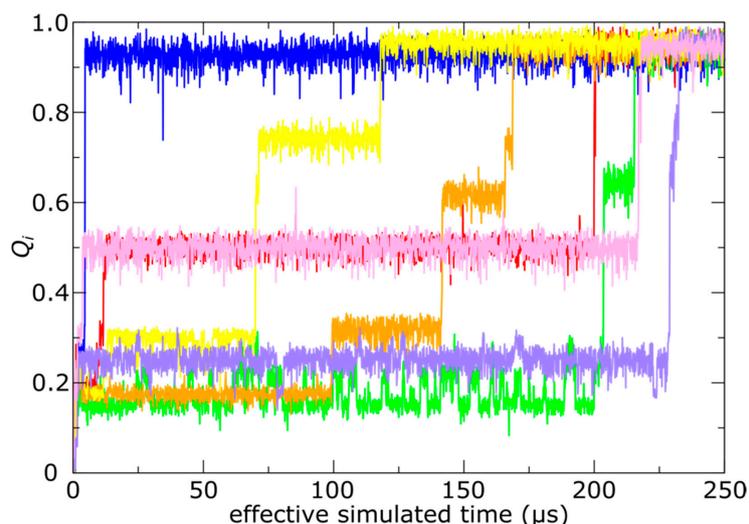


**Figure 5.** View of the trans-shell cavity associated with A- and E-monomer tails. In the mature phage conformation, the A (blue) and E monomers (yellow) adopt a nearly-rotationally-symmetric arrangement, where the N-terminal tails are positioned between B and F monomers. The perspective shown is from the interior of the capsid shell. Despite the fact that the A and E monomers share a common trans-shell cavity, the A monomer is far less likely to undergo trans-shell displacements in the simulations. A possible explanation of the asymmetric dynamics may be that the A monomer is adjacent to a penton (monomer G), while the E monomer is associated with a hexon (monomer C). Differences in the dynamics of expansion around the pentons and hexons have the potential to alter the positions of each N-terminal tail, which may then influence the dynamics of trans-shell displacements.

In summary, we find that translocation of the D- and E-monomer tails is frequent, while B-, C-, F- and G-monomer translocation appears to be substantially slower. Further, it appears that translocation of the A-monomer tails is impeded by prior passage of E-monomer tails. This apparent hierarchy of timescales is caused by differences in the monomer-proximal steric composition, which changes as the capsid expands. As described below, the subtle differences in local steric composition can also give rise to numerous intermediate conformations during maturation.

### 3.3. Capsid Shell Sterics Lead to Multiple Intermediates during N-tail Translocation

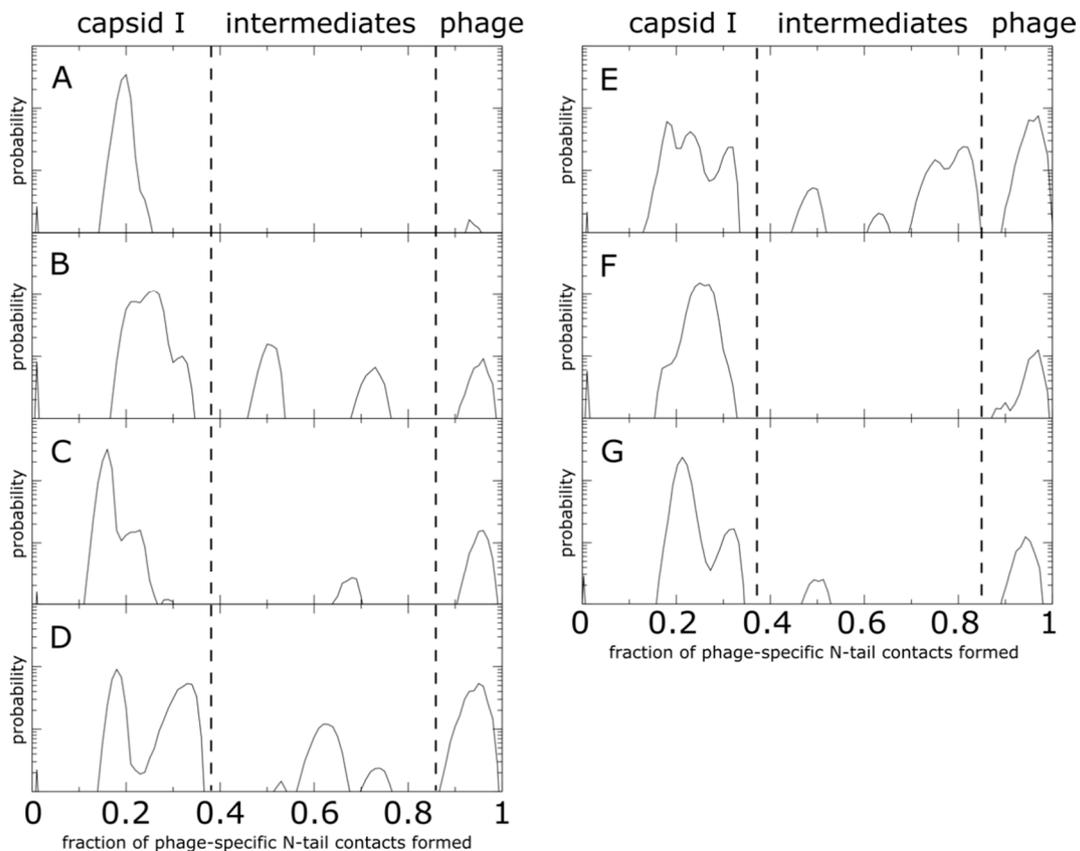
In addition to facilitating position-dependent timescales, the steric composition of the capsid shell also leads to the adoption of various intermediate ensembles (intermediates) during N-terminal tail translocation. To identify intermediates, we inspected the number of mature phage-specific contacts that are formed, as a function of time. As described above, we considered the fraction of these contacts that are formed between the N-terminal tail and the capsid shell for each frame:  $Q_i$ , where  $i$  denotes the monomer position. A value of zero indicates that the tail is on the interior of the capsid shell, whereas a value of 1 indicates that the tail has fully translocated and adopted its mature phage conformation. For illustrative purposes, we show  $Q_i$  for individual monomer tails in a single simulation (Figure 6). These representative time traces have plateaus, each of which reveals the presence of a relatively stable intermediate conformation during trans-shell displacements of the tail. As described below, these intermediates may be attributed to the dense steric environment of the shell.



**Figure 6.** N-terminal tails sample intermediate ensembles. During trans-shell displacements, many tails transiently adopt relatively stable conformations in which only some of the phage-specific contacts are formed. Time traces show the fraction of phage-specific contacts formed ( $Q_i$ ) for representative monomers at each position (1, of 60, per capsid). The color of each trace signifies a different monomer position, where color assignments are consistent with Figures 3 and 4. With the exception of the A (blue) and F (purple) monomers, all other positions display clear pauses at intermediate values of  $Q_i$ .

As done for our analysis of kinetics, we evaluated the statistics of tail dynamics. To describe the distribution of sampled conformations, we calculated the probability as a function of  $Q_i$  for each monomer position (Figure 7). We find that the distributions for monomers A and F only display dominant peaks at high (exterior positions) and low (interior positions) values of  $Q_i$ . In contrast, monomers B–E and G all display peaks at intermediate values of  $Q_i$ , indicating that these positions are likely to transiently halt progress towards the phage conformation. As with the above arguments, these features arise from the imperfect spatial complementarity of gp10 monomers. One may find examples of similar behavior in other systems, for example, where structure-based models have

predicted a range of sterically-induced intermediate conformations of tRNA on the ribosome [29], even when the energetic direction is downhill.

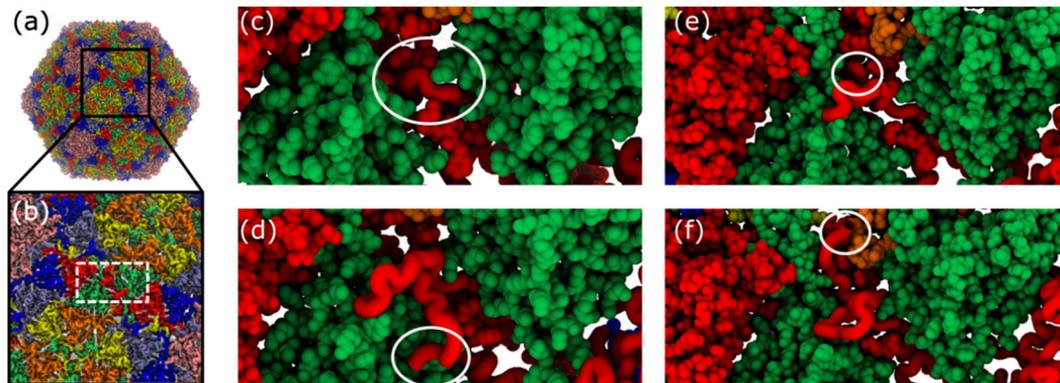


**Figure 7.** N-terminal tail intermediate populations during shell translocation. Each distribution represents the probability as a function of the fraction of mature-phage-specific tail–shell contacts formed ( $Q_i$ ), averaged over all monomers of the same position. The corresponding monomer position (A–G) is listed in the top left of each plot. Distributions were calculated from 20 independent simulations. Probability is shown on a log-linear plot, where the y axis spans three orders of magnitude. For all monomers, there are pronounced peaks at low values of  $Q_i$ , consistent with the presence of an initial relaxation stage during which the tail samples conformations on the inner surface of the shell. There are also peaks at high  $Q_i$  values, which signify that tails in each position do reach the exterior surface of the shell. In addition to the endpoints, monomers B–E have peaks at intermediate values of  $Q_i$ . This shows that the local steric environment introduced by each cavity can impede translocation at various points during the trans-shell displacement. Interestingly, there is no common pattern to the positions of the intermediate peaks, which further illustrates how differences in the steric composition of each cavity can lead to position-specific tail dynamics.

Inspecting the structural properties of intermediate snapshots can highlight steric and topological features that impede capsid maturation in the simulations. For this discussion, we will consider the B monomer and associated cavity, since the B monomer exhibited two clear intermediate peaks (Figure 7), as well as relatively slow trans-shell displacement kinetics (Figure 4). Unlike the other cavities, the B-monomer cavity is associated with N-terminal tails that are from two monomers of the same position within the asymmetric subunit, which arrange around the 2-fold axis of the capsid. Here, we will consider structural snapshots for which  $Q_B$  is approximately 0.3, 0.5, 0.7 and 1.0.

When  $Q_B \sim 0.3$ , the majority of the N-terminal tail is on the interior of the capsid shell, while a sharp loop-like arrangement has entered the trans-shell cavity (Figure 8c). Interestingly, a “loop-first” mode of cavity passage is reminiscent of the “slip-knot” mechanism seen during the folding of knotted

proteins [44]. In those systems, it has been shown that the slip-knot mechanism is entropically favored over threading the terminal residue first, as the latter would require an abrupt reduction in configuration entropy. Based on this, we anticipate that slip-knotting will be a robust feature of trans-shell displacements in T7, irrespective of simulation details.



**Figure 8.** Structural representation of trans-shell displacement intermediates. During trans-shell displacement of the B monomer, various intermediate conformations are sampled (Figure 7). (a) Structure of mature phage structure. (b) Close-up view of shell exterior, centered about the B monomer cavity (white dashed box). (c) During B-monomer (red) translocation, the tail initially ( $Q_i \sim 0.3$ ) forms a tight turn (circled). (d) On occasion, after reaching the exterior surface, only a low number of phage-specific contacts are formed ( $Q_i \sim 0.5$ ), where the tail adopts a twisted conformation (circled), relative to the phage conformation (panel d). (e) In many simulations, the B monomer tail region is displaced through the shell, except for the final few terminal residues (circled). These terminal residues remain trapped within the dense steric environment imposed by the cavity. (f) Phage conformation of the B monomer tail. Note, in all images, the B monomer of interest is shown in tube representation, while the other B monomer associated with the same cavity is shown in vdW representation.

For larger values of  $Q_B$  (0.5 and 0.7), the tail is nominally on the exterior of the capsid shell. However, when  $Q_B \sim 0.5$  (Figure 8d), the tail appears twisted, relative to the phage conformation (Figure 8f). While this is a relatively minor difference, in terms of atomic distances, the dense network of steric interactions limits the mobility of the tail, resulting in slow reorganization to the phage conformation. Interestingly, this phenomenon, where the order of native-contact formation introduces energetic traps, is similar to the observation of “topological frustration” in protein folding [45]. For  $Q_B \sim 0.7$ , the N-terminal tail is confined by the complementary B-monomer with which it shares the cavity (Figure 8e). These conformations may result from an incomplete slip-knot displacement of the N-terminal loop. Again, the densely packed environment introduces substantial steric obstacles that can significantly reduce the rate of adopting the phage conformation. Drawing on analogies with protein folding, it would appear that this caging effect may introduce glassy-like dynamics [14], despite the fact that the model is energetically downhill.

The presented simulations illustrate how molecular sterics and configurational entropy may contribute to the order and timing of N-terminal tail dynamics during phage maturation. This helps establish a baseline description, which opens a range of new areas of investigation. Building upon the simplified models described here, one could dissect the relative contributions of individual steric factors by introducing perturbations and comparing kinetics/mechanisms. As a motivation for such efforts, similar perturbative strategies have been extensively applied to reveal the influence of structure on dynamics in the ribosome [29]. With the demonstrated complexity of capsid maturation, there are now many variations to consider, in order to delineate the precise contributions of intermolecular interactions on large-scale collective dynamics.

## 4. Discussion

### 4.1. Non-Uniformity of Subunit Structure during Expansion

Although quasi-equivalence has been the foundation for thinking about the mature structure of virus capsids [6–8], non-equivalence of expansion intermediates emerges from the simulations performed here. This non-equivalence is derived from sterics-promoted, position-dependent heterogeneous kinetics. In capsid I, the initially-skewed hexon arrangement later adopts a nearly 6-fold symmetric structure in the mature phage. This subtle asymmetry may contribute to a large variability in the expansion dynamics of the various gp10 subunits. Thus, the intermediates are expected to have a structure with less equivalence than the mature structure.

Furthermore, the possibility exists that the non-equivalent dynamics of the tails are, in some conditions, a source of more global variability. For example, in the case of DNA packaging, DNA interaction with the slow tails, if it occurred before expansion completion, might cause a packaged DNA length-dependent variability of global gp10 shell structure. Variability of this type might be part of a proposed back-up motor [46] that is relatively simple from a physical perspective. The variability would eventually be lost when packaging was completed.

### 4.2. Interplay of N-Terminal Tails at Different Positions

One may envision at least two structural explanations for the observation of infrequent, yet fast, translocation events of A monomers (described in Section 3.2). First, it is possible that once the E monomer enters the cavity, the steric limitations are altered, resulting in nearly absent trans-shell displacements of the A monomer. Accordingly, if the vacant-pore rates of the E monomer are slightly faster than for the A monomer, there would be a greater likelihood that the E monomer will enter first and halt trans-shell displacement of the A monomer. This is a physically-intuitive rationalization, since translocation of the E-monomer tail will necessarily lead to a more constricted channel through which the A-monomer tail must pass.

A second explanation is that expansion leads to opening of the cavity, which introduces a time-dependent steric environment that differentially impacts the A monomer. In that scenario, the A-monomer tail could translocate during the initial stages of the simulation, though upon shell expansion, the cavity would disfavor trans-shell displacements of the A monomer. This type of asymmetry may be expected based on the structure around the A/E pore. Specifically, each A monomer is proximal to a penton, while the E monomer is proximal to a monomer of the hexon. Consistent with previous simulations that predict transient symmetry breaking during expansion [31], it is possible that similar transient asymmetries may contribute to differential N-terminal tail dynamics. While isolating the precise origins of asymmetry is beyond the scope of the current study, future variants of the model may be designed to distinguish between these potential contributors to tail dynamics.

### 4.3. The Future of Modeling Capsid Dynamics

With the ever-increasing capacity and accessibility of high-performance computing resources, combined with the development of novel theoretical models, the field of computational molecular biophysics is reaching a point where one can rapidly propose and test theoretical concepts pertaining to large assemblies. That is, while simulations of full capsids are still computationally demanding, it is becoming possible to iteratively introduce changes to the model and test the resulting dynamics. By starting with a simple energetic model, as done here, one may gradually increase the model complexity and carefully assess the effects of each additional factor.

In the case of T7, we are now ready to extend the single-basin structure-based model, in order to account for additional competing conformations. For example, one may introduce interactions that stabilize the capsid I, capsid II and mature-phage conformations. The relative stability of each conformation could then be modulated, in order to determine which aspects of maturation are

robust to the energetic details. This general strategy has been widely used to study conformational rearrangements in multi-domain proteins [47], as well as in assemblies [29,33,48].

Thus, from a technical standpoint, the presented simulations provide a foundation, upon which future studies may explore the balance of intermediate populations and localized motions of individual monomers. An additional potential avenue of investigation would be to increase the model complexity by explicitly representing the interactions with DNA [49], which could separately treat the steric/entropic and electrostatic effects. Through this systematic extension of the models, and comparison with experiments, we anticipate that the rich features of phage maturation may be fully understood.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1999-4915/12/11/1273/s1>, Video S1: Simulation of capsid expansion.

**Author Contributions:** Conceptualization, P.C.W., W.J., P.S.; methodology, P.C.W.; formal analysis, P.C.W.; writing—original draft preparation, P.C.W., W.J., P.S.; writing—review and editing, P.C.W., W.J., P.S.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the NSF (Grants MCB-1915843). Work at the Center for Theoretical Biological Physics is supported by the NSF (Grant PHY-2019745). PS acknowledges support from the Morrison Trust (grant no. 44040-10009383).

**Acknowledgments:** We would also like to acknowledge generous computational and technical support provided by the Northeastern University Discovery Cluster and Research Computing department.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Whitford, P.C.; Onuchic, J.N.; Wolynes, P.G. Energy landscape along an enzymatic reaction trajectory: Hinges or cracks? *HFSP J.* **2008**, *2*, 61–64. [[CrossRef](#)] [[PubMed](#)]
- Howard, J. Motor Proteins as Nanomachines: The Roles of Thermal Fluctuations in Generating Force and Motion. In *Biological Physics*; Seminaire Poincare XII; Birkhauser Basel: Basel, Switzerland, 2011; pp. 33–44.
- Milescu, L.S.; Yildiz, A.; Selvin, P.R.; Sachs, F. Maximum likelihood estimation of molecular motor kinetics from staircase dwell-time sequences. *Biophys. J.* **2006**, *91*, 1156–1168. [[CrossRef](#)] [[PubMed](#)]
- Weber, G. Energetics of ligand binding to proteins. *Adv. Protein Chem.* **1975**, *29*, 1–83. [[PubMed](#)]
- Rossmann, M.G. Structure of viruses: A short history. *Q. Rev. Biophys.* **2013**, *46*, 133–180. [[CrossRef](#)] [[PubMed](#)]
- Harrison, S.C. Protein tentacles. *J. Struct. Biol.* **2017**, *200*, 244–247. [[CrossRef](#)] [[PubMed](#)]
- Twarock, R.; Luque, A. Structural puzzles in virology solved with an overarching icosahedral design principle. *Nat. Commun.* **2019**, *10*, 4414. [[CrossRef](#)]
- Liu, Y.; Hill, M.G.; Klose, T.; Chen, Z.; Watters, K.; Bochkov, Y.A.; Jiang, W.; Palmenberg, A.C.; Rossmann, M.G. Atomic structure of a rhinovirus C, a virus species linked to severe childhood asthma. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8997–9002. [[CrossRef](#)]
- Sirohi, D.; Chen, Z.; Sun, L.; Klose, T.; Pierson, T.C.; Rossmann, M.G.; Kuhn, R.J. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* **2016**, *352*, 467–470. [[CrossRef](#)]
- Rossmann, M.G. My happy years of collaboration with Fumio Arisaka. *Biophys. Rev.* **2018**, *10*, 139–140. [[CrossRef](#)]
- Delbrück, M.A. A physicist looks at biology. In *Phage and the Origins of Molecular Biology*; Cairns, J., Stent, G.S., Watson, J.D., Eds.; Cold Spring Laboratory of Quantitative Biology: Cold Spring Harbor, NY, USA, 1966; pp. 9–22.
- Levinthal, C. Mossbauer Spectroscopy in Biological Systems. *Proceedings* **1969**, *67*, 22–24.
- Bryngelson, J.D.; Wolynes, P.G. A Simple Statistical Field-Theory of Heteropolymer Collapse with Application to Protein Folding. *Biopolymers* **1990**, *30*, 177–188. [[CrossRef](#)]
- Bryngelson, J.D.; Wolynes, P.G. Intermediates and Barrier Crossing in A Random Energy-Model (with Applications to Protein Folding). *J. Phys. Chem.* **1989**, *93*, 6902–6915. [[CrossRef](#)]
- Bryngelson, J.D.; Onuchic, J.N.; Socci, N.D.; Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **1995**, *21*, 167–195. [[CrossRef](#)]

16. Linke, H.; Hocker, B.; Furuta, K.; Forde, N.R.; Curmi, P.M.G. Synthetic biology approaches to dissecting linear motor protein function: Towards the design and synthesis of artificial autonomous protein walkers. *Biophys. Rev.* **2020**, *12*, 1041–1054. [[CrossRef](#)]
17. Hwang, W.; Karplus, M. Structural basis for power stroke vs. Brownian ratchet mechanisms of motor proteins. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19777–19785. [[CrossRef](#)] [[PubMed](#)]
18. Zhang, B.; Wolynes, P.G. Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6062–6067. [[CrossRef](#)]
19. Di Pierro, M.; Zhang, B.; Aiden, E.L.; Wolynes, P.G.; Onuchic, J.N. Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12168–12173. [[CrossRef](#)]
20. Lin, Y.H.; Forman-Kay, J.D.; Chan, H.S. Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles. *Phys. Rev. Lett.* **2016**, *117*, 178101. [[CrossRef](#)]
21. Fujisawa, H.; Morita, M. Phage DNA packaging. *Genes Cells* **1997**, *2*, 537–545. [[CrossRef](#)]
22. Serwer, P.; Jiang, W. Dualities in the analysis of phage DNA packaging motors. *Bacteriophage* **2012**, *2*, 239–255. [[CrossRef](#)]
23. Guo, F.; Liu, Z.; Fang, P.A.; Zhang, Q.; Wright, E.T.; Wu, W.; Zhang, C.; Vago, F.; Ren, Y.; Jakana, J.; et al. Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4606–E4614. [[CrossRef](#)] [[PubMed](#)]
24. Guo, F.; Liu, Z.; Vago, F.; Ren, Y.; Wu, W.; Wright, E.T.; Serwer, P.; Jiang, W. Visualization of uncorrelated, tandem symmetry mismatches in the internal genome packaging apparatus of bacteriophage T7. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6811–6816. [[CrossRef](#)]
25. Pajunen, M.I.; Elizondo, M.R.; Skurnik, M.; Kieleczawa, J.; Molineux, I.J. Complete Nucleotide Sequence and Likely Recombinatorial Origin of Bacteriophage T3. *J. Mol. Biol.* **2002**, *319*, 1115–1132. [[CrossRef](#)]
26. Studier, F.W.; Dunn, J.J. Organization and expression of bacteriophage T7 DNA. *Cold Spring Harb. Symp. Quant. Biol.* **1983**, *47*, 999–1007. [[CrossRef](#)] [[PubMed](#)]
27. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Gr.* **1996**, *14*, 33–38. [[CrossRef](#)]
28. Whitford, P.C.; Noel, J.K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K.Y.; Onuchic, J.N. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins* **2009**, *75*, 430–441. [[CrossRef](#)]
29. Levi, M.; Noel, J.K.; Whitford, P.C. Studying ribosome dynamics with simplified models. *Methods* **2019**, *162*, 128–140. [[CrossRef](#)]
30. Noel, J.K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R.L.; Onuchic, J.N.; Whitford, P.C. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput. Biol.* **2016**, *12*, e1004794. [[CrossRef](#)]
31. May, E.R.; Feng, J.; Brooks, C.L., 3rd. Exploring the symmetry and mechanism of virus capsid maturation via an ensemble of pathways. *Biophys. J.* **2012**, *102*, 606–612. [[CrossRef](#)]
32. Noel, J.K.; Whitford, P.C.; Onuchic, J.N. The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B* **2012**, *116*, 8692–8702. [[CrossRef](#)]
33. Nguyen, K.; Whitford, P.C. Steric interactions lead to collective tilting motion in the ribosome during mRNA-tRNA translocation. *Nat. Commun.* **2016**, *7*, 10586. [[CrossRef](#)]
34. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012. [[CrossRef](#)]
35. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. [[CrossRef](#)]
36. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Mol. Model. Ann.* **2001**, *7*, 306–317. [[CrossRef](#)]
37. Jackson, J.; Nguyen, K.; Whitford, P.C. Exploring the balance between folding and functional dynamics in proteins and RNA. *Int. J. Mol. Sci.* **2015**, *16*, 6868–6889. [[CrossRef](#)]
38. Yang, H.; Bandarkar, P.; Horne, R.; Leite, V.B.P.; Chahine, J.; Whitford, P.C. Diffusion of tRNA inside the ribosome is position-dependent. *J. Chem. Phys.* **2019**, *151*, 085102. [[CrossRef](#)]

39. Di Pierro, M.; Potoyan, D.A.; Wolynes, P.G.; Onuchic, J.N. Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 7753–7758. [[CrossRef](#)]
40. Hyeon, C.; Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011**, *2*, 487. [[CrossRef](#)] [[PubMed](#)]
41. Chan, H.S.; Zhang, Z.; Wallin, S.; Liu, Z. Cooperativity, local-nonlocal coupling, and nonnative interactions: Principles of protein folding from coarse-grained models. *Annu. Rev. Phys. Chem.* **2011**, *62*, 301–326. [[CrossRef](#)]
42. Cho, S.S.; Levy, Y.; Wolynes, P.G. P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 586–591. [[CrossRef](#)] [[PubMed](#)]
43. Frauenfelder, H.; Sligar, S.G.; Wolynes, P.G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598–1603. [[CrossRef](#)] [[PubMed](#)]
44. Sulkowska, J.I. On folding of entangled proteins: Knots, lassos, links and theta-curves. *Curr. Opin. Struct. Biol.* **2020**, *60*, 131–141. [[CrossRef](#)]
45. Gosavi, S.; Chavez, L.L.; Jennings, P.A.; Onuchic, J.N. Topological frustration and the folding of interleukin-1 beta. *J. Mol. Biol.* **2006**, *357*, 986–996. [[CrossRef](#)]
46. Serwer, P. Proposed ancestors of phage nucleic acid packaging motors (and cells). *Viruses* **2011**, *3*, 1249–1280. [[CrossRef](#)]
47. Whitford, P.C.; Miyashita, O.; Levy, Y.; Onuchic, J.N. Conformational transitions of adenylate kinase: Switching by cracking. *J. Mol. Biol.* **2007**, *366*, 1661–1671. [[CrossRef](#)]
48. Hyeon, C.; Onuchic, J.N. Mechanical control of the directional stepping dynamics of the kinesin motor. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17382–17387. [[CrossRef](#)]
49. Givaty, O.; Levy, Y. Protein Sliding along DNA: Dynamics and Structural Characterization. *J. Mol. Biol.* **2009**, *385*, 1087–1097. [[CrossRef](#)]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).