Structure, Vol. 13, 413–421, March, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2004.12.015

Maturation Dynamics of Bacteriophage HK97 Capsid

A.J. Rader, Daniel H. Vlad, and Ivet Bahar[1,*]
Department of Computational Biology
School of Medicine
University of Pittsburgh
Pittsburgh, Pennsylvania 15261

Summary

Maturation of the bacteriophage HK97 capsid requires a large conformational change of the virus capsid. Experimental studies have identified several intermediates along this maturation pathway. To gain insights into the molecular mechanisms of capsid maturation, we examined the fluctuation dynamics of the procapsid and mature capsid using a residue-level computational approach. The most cooperative motions of the procapsid are found to be consistent with the observed change in configuration that takes place during maturation. A few dominant modes of motion are sufficient to describe the anisotropic expansion that accompanies maturation. Based upon these modes, maturation is proposed to occur via an overall expansion and reconfiguration of the capsid initiated by puckering of the pentamers, followed by flattening and crosslinking of the hexameric subunits, and finally crosslinking of the pentameric subunits. The highly mobile E loops are stabilized by anchoring to highly stable residues belonging to neighboring subunits.

Introduction

An essential step in the life cycle of many viruses is the maturation of their capsid, the process by which viral precursors become infectious virions. This step usually involves a significant and irreversible structural transformation of the shell of coat proteins from a roughly spherical procapsid (or prohead) into an icosahedral mature capsid (or head) structure (King and Chiu, 1997; Conway et al., 2001). In most bacteriophages, this conformational change is initiated by proteolysis of the coat proteins to create a marginally unstable procapsid structure, followed by an expansion in concert with DNA packaging. This expansion from procapsid to capsid can be controlled in vitro by changing the ambient pH or other chemical stimuli (Conway et al., 1995; Duda et al., 1995; Hendrix and Duda, 1998).

Bacteriophage HK97 capsid has become a model system for experimental study of the maturation pathway of double-stranded DNA (dsDNA) bacteriophages due to availability of structural information on maturation steps. The assembly and maturation of HK97 capsid is a multistep process. The initial step is the assembly of 415 gp5 proteins (of 385 residues each) into a shell (Prohead I) of 60 hexamers and 11 pentamers, together with 12 gp3 portal proteins that form the 12th pentameric face of a T = 7 icosahedron. Approximately 50 HK97 proteases (gp4) are enclosed in the shell, which, after assembly, excise the 103 N-terminal residues of the gp5 proteins (cleavage), thus creating the DNA-packaging envelope structure, Prohead II (Hendrix and Duda, 1998; Wikoff et al., 2000). The latter undergoes a substantial expansion in volume, along with a change in morphology (from spherical into icosahedral shape) and interresidue contact topology upon packaging of phage DNA, as the mature (Head I) structure is formed. An innocuous variant of the HK97 capsid, comprised of 420 copies of the gp5 protein, has been shown to assemble and mature into the Head I structure (Duda et al., 1995; Conway et al., 1995). This variant has been widely studied in vitro to understand the mechanism(s) of transition into the final mature form, Head II. In the Head II structure, side chains (Lys169 and Asn356) from different gp5 proteins crosslink to form a highly catenated protein chainmail (Duda, 1998; Wikoff et al., 2000). Recent studies indicate that the formation of these intersubunit covalent crosslinks occur concurrently with the conformational changes involved in the transition from Prohead II to Head I (Gan et al., 2004).

Low-resolution structures determined by cryo-electron microscopy (cryo-EM) along the in vitro maturation pathway have identified the above-mentioned three intermediate states (Lata et al., 2000). Recently, the cryo-EM structures have been paired with higher resolution X-ray crystallographic data to produce structural models of the viral capsid proteins on an atomic level (Johnson and Reddy, 1998; Conway et al., 2001; Wikoff et al., 1999, 2000; Lee and Johnson, 2003). These structures are formed by 60 icosahedrally arranged asymmetric units, each composed of seven gp5 proteins labeled A–G. The first six (A–F) form the hexamers while the seventh (G) participates with five neighboring G chains in the adjacent pentamer (Figure 1) (Conway et al., 2001; Helgstrand et al., 2003; Wikoff et al., 2000). Residue- and atomic-level data now provide a new framework for elucidating the complex macromolecular mechanics of virus maturation.

Of the three maturation steps—cleavage, expansion, and crosslinking—expansion involves the most pronounced changes in conformation (Conway et al., 1995). Direct comparison of the Prohead II and Head II structures indicates that equivalent chains undergo radial translations of up to 58 Å (G and G′) and rotations up to 39° (B and B′). Additionally, the distance between Lys169 and Asn356 C^{α} atoms decreases from 35 ± 4 Å (Prohead II) to 7 ± 1 Å (Head II) (Conway et al., 2001). These large-scale displacements involve both inter- and intrasubunit rearrangements, such as the rigid-body motions of the structural cores (domains A and P) and the refolding of the extended motifs (N arm and E loop) (Figure 1). The conformational rearrangements of the subunits and domains manifest themselves by a change in the overall architecture of the capsid from spherical to icosahedral, increasing the diameter by

*Correspondence: bahar@pitt.edu
[1]Lab address: http://www.ccbb.pitt.edu/research/bahar_lab/

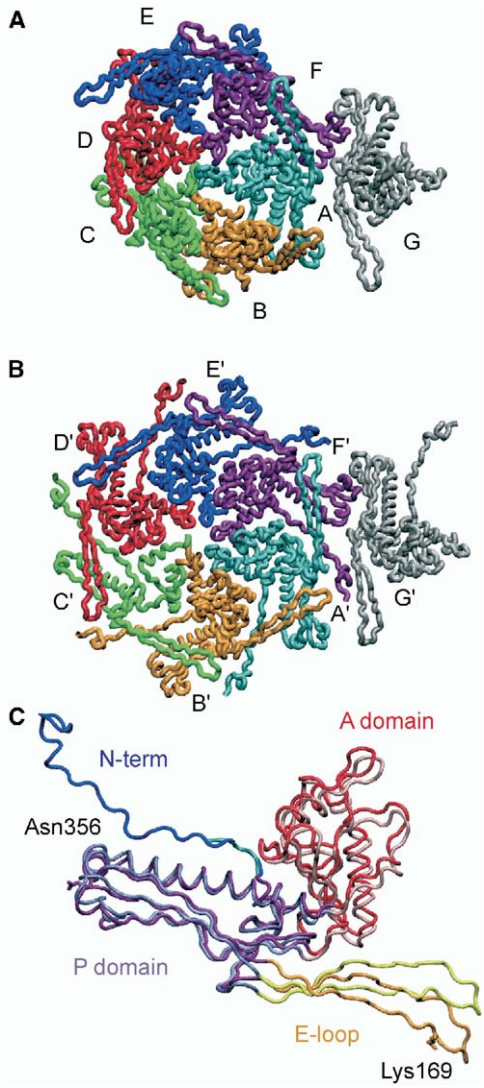

Figure 1. Asymmetric Units Modeling the Prohead II and Head II HK97 Capsid Structures

(A) The Prohead II asymmetric unit consists of seven identical chains (gp5 proteins), A–G, and was determined by cryo-EM (PDB code 1if0; Conway et al., 2001). The first six chains, A–F, form a pair of skewed trimers: AFE and BCD. (B) The Head II asymmetric unit also contains seven gp5 proteins (A′–G′) but was determined by X-ray crystallography (PDB code 1ohg; Helgstrand et al., 2003). Here the first six chains form a fully symmetric hexamer. (C) Superimposition of the seventh chain G (G′) indicates how the domains reorient upon maturation. Each chain is comprised of two core domains, axial (A domain) and peripheral (P domain), relative to the hexamer center, in addition to two extended motifs, the N arm (residues 104–132; unresolved in Prohead II) and the E loop (residues 148–181; a two-stranded β sheet). The E loop contains residue Lys169 which forms an isopeptide bond with Asn356 of a neighboring gp5 protein in the mature form.

about 20% (from 550 Å in the procapsid to 660 Å along the 5-fold axes in Head II) and the internal volume by over 100% (Lata et al., 2000). Only with high cooperativity among interacting subunits can such a large-scale and highly concerted conformational change take place.

Computer simulations of protein dynamics has become a powerful technique for visualizing molecular interactions and assessing functional motions. The immense size (more than 10^5 residues or 9×10^5 heavy atoms) of the HK97 capsid, however, precludes the use of conventional methods such as full atomic molecular dynamics (MD) simulations or normal mode analysis (NMA) (Phelps et al., 2000). One could simplify the problem by using group theory to exploit the symmetry of virus particles (Simonson and Perahia, 1992). Although such symmetry-based calculations have been implemented on an icosahedral dialanine$_{60}$ (van Vlijmen and Karplus, 2001) and the HK97 capsid (Kim et al., 2003), these methods are unable to observe the most cooperative, potentially functional, symmetry-breaking motions.

Alternatively, simplified NMA-based methods have been developed, where structures are represented as elastic networks (ENs) with different levels of details (Bahar et al., 1997; Hinsen et al., 1999; Tama et al., 2000; Atilgan et al., 2001; Doruker et al., 2002). Recent implementations of these low-resolution descriptions along with increased computational power have permitted us to examine the dynamics of increasingly larger biomolecular assemblies (Keskin et al., 2002; Tama and Brooks, 2002; Delarue and Sanejouand, 2002; Li and Cui, 2002; Ming et al., 2003; Wang et al., 2004). These studies take rigorous account of the interresidue contact topology via a connectivity or Kirchhoff matrix (Γ) of interresidue contacts, while overlooking atomic details, based on the premise that the large-scale collective motions are defined by the overall architecture, rather than local interactions.

To further reduce the size and complexity of the problem, one approach is to assume the entire subunits (e.g., coat proteins in viral capsids) form rigid blocks. Such a reduced model has been adopted by Tama and Brooks (2002) in their analysis of cowpea chlorotic mottle virus (CCMV) dynamics. However, the HK97 gp5 proteins are subject to internal (local) changes in conformation that accompany capsid maturation and assist in the autocatalysis of the isopeptide crosslinks. Previous examination of the cryo-EM and X-ray structures suggests the occurrence of hinge motions between the A and P domains, as well as significant changes in the E-loop conformations (Figure 1C). Our goal is to elucidate the collective dynamics of the procapsid at multiple scales, including both the global modes of the entire capsid and the local, residue-level motions, to detect the mechanisms of possible couplings between intramolecular mobilities and entire capsid reconfiguration.

Results obtained with the Gaussian Network Model (GNM; Bahar et al., 1997; Haliloglu et al., 1997) reveal that a small subset of modes (less than 0.1% of all possible GNM modes) effectively drives the transition from the procapsid to the mature capsid. Despite the icosahedral symmetry of the two known structures, a delicate interplay between isocahedrally symmetric and asymmetric modes is revealed, which points to the importance of the high level of cooperativity across the entire capsid to achieve the dramatic change in overall conformation. The GNM also identifies residue Asn356 as a hinge site that plays a key role in coordinating the

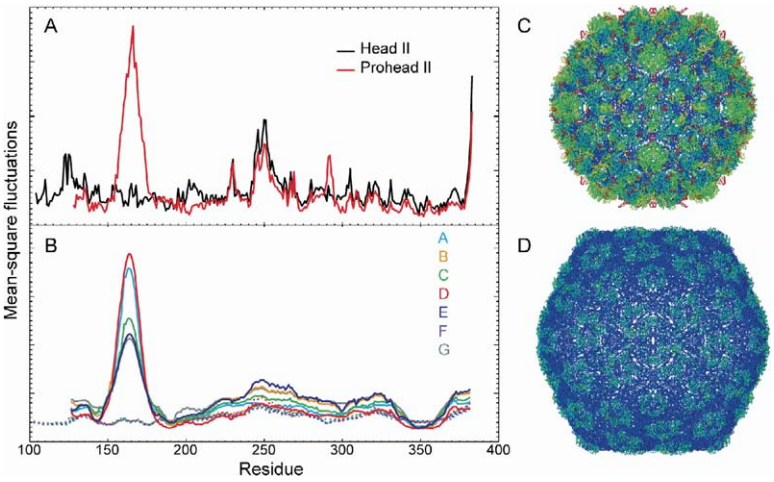

Figure 2. Calculated Mean-Square Fluctuations of Residues for the Prohead II and Head II Structures

(A) The distribution of ms fluctuations is shown for a representative chain G, calculated for the pentamer-centered substructures Pro_5 and $Head_5$. Due to the 5-fold rotational symmetry, the calculated B factors for the other G chains are identical to the one shown. (B) The distributions of ms fluctuations are shown for each chain (A–G) in a representative asymmetric unit of Prohead II (solid) and Head II (dotted). These results refer to the slowest 300 eigenmodes of the entire structures. The residue indices are from the structure deposited in the Protein Data Bank. Significant suppression in the ms fluctuation can be observed for the E loop (residues 148–181). The calculated fluctuations from (B) are mapped onto the three-dimensional structures of (C) Prohead II and (D) Head II, using visual molecular dynamics (VMD; Humphrey et al., 1996). Residues are color coded such that the most mobile are red and the least mobile are blue. Head II (D) is considerably less mobile (i.e., more stable) than Prohead II (C).

cooperative motions, suggesting that its substitution by a bulkier, nonspecific amino acid would potentially disrupt the maturation dynamics of the procapsid.

Results and Discussion

Equilibrium Fluctuations Identify Conformationally Variable Regions

One measure of the intrinsic potential of residues to undergo or resist a change in conformation is their mean-square (ms) fluctuations, $\langle(\Delta R_i)^2\rangle$, near equilibrium coordinates. X-ray crystallographic B factors (or temperature factors) provide an indirect measure of ms fluctuations (Equation 2 in Experimental Procedures). Temperature factors for the pseudoatomic model of HK97 procapsid were not reported with the PDB-deposited X-ray crystallographic structure file. We computed the ms fluctuations for the Prohead II and Head II structures (PDB code 1if0, Conway et al., 2001, and PDB code 1fh6, Wikoff et al., 2000, respectively) using the GNM. The results are presented in Figure 2. Due to icosahedral symmetry, all 60 gp5 chains of a given type (A–G) exhibit the same distribution of ms fluctuations. Subsequent to our analysis, this 3.60 Å Head II structure was replaced by a further refined, 3.45 Å structure, PDB code 1ohg (Helgstrand et al., 2003), which differ in C^{α} root-mean-square deviation (rmsd) by only 0.779 Å.

Figures 2A and 2B display the residue ms fluctuations obtained using two different levels of analysis. The curves in Figure 2A refer to the complete GNM calculations for pentamer-centered substructures of the capsid and procapsid subject to a mean-field coupling to extrasubstructural elements (see Experimental Procedures). The results are shown for representative G chains in the procasid (red) and capsid (black). Figure 2B displays the fluctuations induced in chains A–G by a dominant set (300) of vibrational modes for the entire procapsid (solid) and capsid (dotted) structures of $\sim 10^5$ residues. Consistent with the inclusion of high-frequency modes, the curves in Figure 2A are much more rugged. The dominant features, however, compare well between the two methods, yielding a correlation coefficient of 0.951 between the Head II G chain fluctuations in the two panels, and 0.997 between the Prohead II G chains. Differences between the Prohead II (red) and Head II (black) curves exemplify the changes in fluctuation dynamics accompanying capsid maturation. Significantly larger fluctuations are calculated in the less stable procapsid than in the mature structure. Because infection proceeds by the injection of genetic material through a tailed structure rather than capsid disassembly for most dsDNA bacteriophages, the mature HK97 capsid is much more stable than other, non-tailed dsDNA bacteriophages, and the observed fluctuation dynamics of the mature capsid is consistent with this functional requirement.

Of great interest is the dramatic decrease in the motions observed for the E loop (residues 148–181) in the mature compared to procapsid structure. The initially exposed E loops are reoriented and bound to the viral shell as the procapsid matures into the capsid structure such that Lys169 residues are able to form chemical crosslinks with Asn356 residues from neighboring chains. The resulting rings of covalently catenated subunits loop through each other to form a very stable protein chainmail (Wikoff et al., 2000). The calculated decrease in the mobilities of the E loops for the mature capsid results from these crosslinks, and demonstrates the effective stabilization of the E loops in the mature form. The opposite change, an increase in fluctuations, is observable in the neighborhood of the P domain helices, $\alpha 2$ and $\alpha 7$ (residues 194–225), upon maturation. This region is spatially adjacent to Asn356 that crosslinks to the highly flexible E loop Lys169 of a neighboring subunit, which can explain the small but perceptible increase in fluctuations induced upon coupling to a very flexible segment.

A second notable, but much weaker, decrease in fluctuation amplitudes is seen in Figure 2A in the immediate vicinity of the 5-fold symmetry axis of the pentamers (residues 285–295). These residues evolve from a relatively flat surface in Prohead II into a protrusion in Head II. Thus, the pentamer center appears to be more

severely constrained (stable) in the mature form than in the less ordered, procapsid form. The fact that this effect is not discernible in Figure 2B implies that the modes suppressed in the mature form are those in the relatively high frequency range.

These two conformational changes are the main mechanisms that impart stability to the HK97 capsid in the mature form. Except for the N-terminal residues (absent from the PDB-deposited Prohead II structure), no other groups of residues exhibit significant changes in their ms fluctuations. The structures of Prohead II (Figure 2C) and Head II (Figure 2D) colored according to their calculated mobilities indicate an overall greater stability in Head II (blue). In the procapsid, the E loops are distinguished by their high mobilities (shown in red).

The Most Cooperative Motions Are Not Icosahedrally Symmetric

The fluctuations shown in Figure 2 describe the likely local motions observable for a given X-ray structure, but cannot be construed as the means by which one structure transforms into another. This is the case especially when there is a large conformational change such as in capsid maturation. While these overall fluctuations suggest regions of mobile residues that may participate in the maturation process, often the combination of the individual modes obscures the most functional motions. Inherent to network models, such as the GNM, is the property that the most cooperative, global motions of the system can be extracted by decomposing the motion into various modes. This set of orthogonal modes defines a basis for describing all potential motions of a given structure. The modes with the lowest frequency have been shown to give insights into functional motions (Tama and Sanejouand, 2001; Xu et al., 2003). Typically, the residues that exhibit the highest translational mobilities in the low-frequency modes act as substrate recognition sites, while the most severely constrained regions act as anchors or hinge sites, critical for coordinating the cooperative movements of the structure (Bahar et al., 1998). For small structures, it has been sufficient to investigate only the first few (about 10) global modes (Haliloglu et al., 1997; Keskin et al., 2002). However, due to its large size, it may be necessary to explore a larger number of modes for the viral capsid, and then examine more closely particular substructures.

Figure 3 illustrates the mobility profiles of the mature capsid (Figures 3A–3C) and procapsid (Figures 3D–3F) induced by the three slowest distinct modes of motion favored by the respective structures. The capsids are color coded in a similar way to Figure 2, with the most mobile regions in red and least mobile in blue. Due to the high degree of symmetry in the structure, it is natural that a large number of collective modes will be degenerate, that is, they have sister modes that share the same frequencies and exhibit symmetrically related motions. The modes illustrated in Figures 3A (or 3D), 3B (or 3E), and 3C (or 3F) are indeed 3-, 5-, and 3-fold degenerate, respectively. Notably, these 11 modes are found (from the eigenvalue distribution of the GNM modes) to account for 41% of the ms fluctuations induced by the 300 dominant modes in Figure 2B. The slowest modes of the procapsid and capsid (Figures 3A and 3D) both have a distinctive polarization of mobile and immobile residues suggesting a breathing dynamics of expanding at opposite poles. The most mobile regions coincide with the pentamer centers. The second set of modes (Figures 3B and 3E) induces a more complex dynamics than the first. In particular, we observe two concentric intersecting, immobile shells instead of a single central banded region that is relatively rigid. Consistent with an increased localization of higher modes compared to slower modes, the general shape for the third set of slow modes (Figures 3C and 3F) is more focused around the highly mobile pentamer-forming G chains. The complementary shape of these dominant modes is in accord with their orthogonal nature.

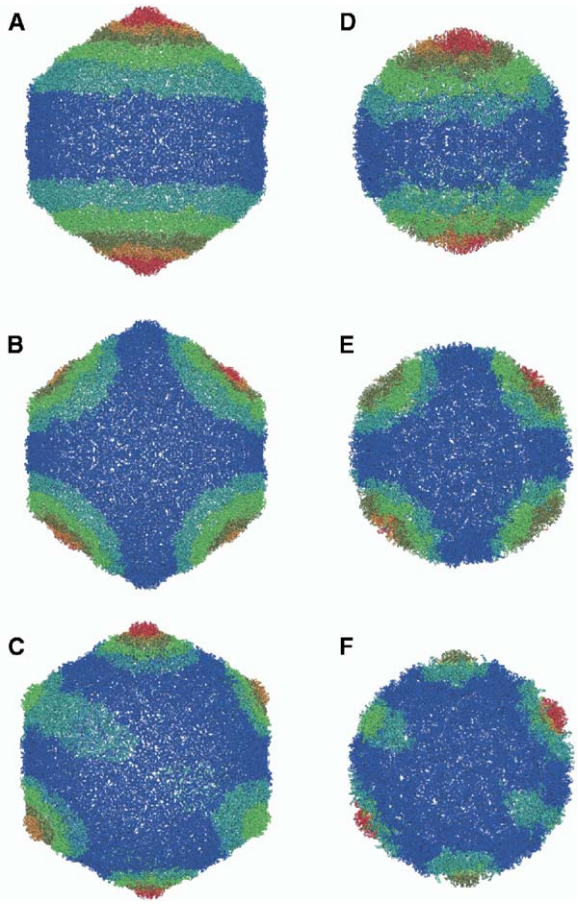

Figure 3. The Three Slowest Distinct Modes Calculated for the Entire HK97 (Pro)Capsid Structures

(A) The slowest mode from Head II is 3-fold degenerate and shows two distinct mobile regions centered at the pentamers at opposing poles. (B) The next slowest mode for Head II is 5-fold degenerate and displays four evenly spaced islands of highly mobile residues. (C) The next slowest mode for Head II is 3-fold degenerate and distinguishes the pentamers as more mobile than the surrounding hexamers. (D–F) The results for the Prohead II structure indicate similar distributions of mobilities in these slowest (degenerate) modes. The structures are color coded such that the most mobile residues are red and the least mobile residues are blue, and were generated using VMD (Humphrey et al., 1996).

It is important to note that none of these most prob-

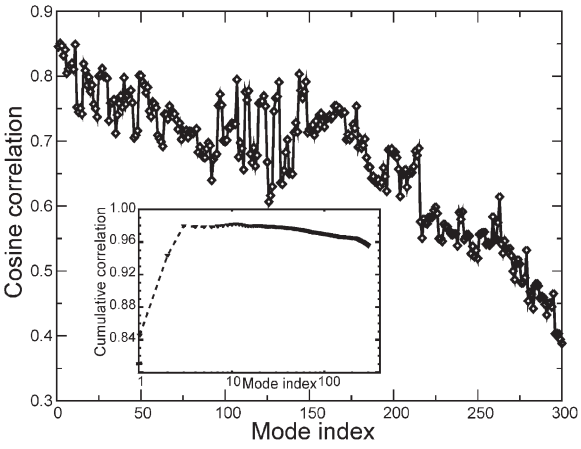

Figure 4. Correlation Cosine (Equation 3) between Mobilities and Net Displacements of Residues for Each Mode Individually Plotted as a Function of Mode Index

The correlation for any given mode varies between 0.85 and 0.38, gradually decreasing as the mode number increases. This decreasing trend reflects that higher modes refer to more localized and disordered motions with higher frequency while the lower modes account for the highly cooperative, global motions relevant for maturation. The cumulative correlation cosine found from the weighted sum over individual modes is shown (on a logarithmic scale, for clarity) in the inset. The cumulative correlation for the first three (degenerate) modes is 0.974 and reaches a maximum of 0.982 at mode 11.

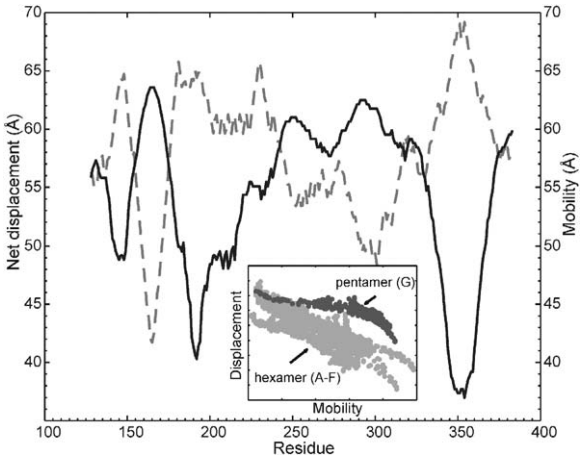

Figure 5. Comparison of Calculated Mobilities from Slowest 100 Modes (Solid) and Displacement Vector (Dashed) Describing the Expansion from Prohead II to Head II, for One Representative G Chain

Note that the calculated mobilities would show a remarkable agreement if subtracted from an isotropic expansion of about 110 Å uniformly distributed over all residues. The remarkable anticorrelation indicates how well this simple model can explain the local changes in conformation accompanying the isotropic expansion of the procapsid. The two sets plotted against each other (inset) give an overall correlation coefficient (Equation 4) of –0.582. The anticorrelations increase to –0.824 (chains A–F) and –0.842 (chain G) when the predictions for hexameric and pentameric chains are separately compared to their experimental counterparts.

able collective modes is icosahedrally symmetric on an individual basis. How does the capsid undergo a transition from an icosahedrally symmetric structure (Prohead II) to another (Head II) while the most important (slowest) modes of reconfiguration are *not* icosahedrally symmetric? This intriguing issue will be investigated next.

How Do Global Fluctuations Correlate with Observed Displacements?

One means of quantifying the large conformational change that occurs during maturation is to calculate the net displacement of each residue between the two known endpoints, Prohead II and Head II. These net displacements include the overall expansion as well as the intra- and intersubunit rearrangements accompanying the expansion. Of interest is to assess to what extent the individual modes predicted by the GNM contribute to, or correlate with, these actual displacements.

To this aim, we examined the correlation cosine between the observed displacements of procapsid residues (organized as an N-dimensional vector $\boldsymbol{d}$, where N = 107,520) and the mobilities, $\boldsymbol{m}$, induced by the individual modes (given by the elements of the eigenvectors of Γ; see Experimental Procedures). This quantity reflects the correlation between the experimentally observed sizes of residue fluctuations, and those induced by individual modes. Notably, the mobilities from the single slowest mode (Figure 3D) exhibit a cosine correlation of 0.846 with the net displacements they undergo during maturation. Therefore, the slowest GNM mode tends to reconfigure the procapsid pentamers toward the conformational state they assume in the mature form Head II. Repeating this analysis for all the N residues' fluctuations in each of the 300 slowest modes led to correlation cosines presented in Figure 4. The inset shows that the cumulative correlation cosine reaches a maximum of 0.982 at the 11th mode, and then slightly weakens due to the superimposition of less strongly correlated higher modes. This analysis demonstrates (1) the importance of extracting the most cooperative modes and filtering out the higher (noisy) modes for elucidating functional motions, and (2) the relevance of the mobilities induced in the slowest modes to the actual displacements of residues during maturation, despite the lack of icosahedral symmetry in the individual slow modes.

Superposition of 100 Slowest Modes Suggests a Two-Phase Maturation Mechanics

The dashed curve in Figure 5 displays the net displacements $\boldsymbol{d}$ of residues (between the Prohead II and Head II forms) for a representative gp5 monomer (chain G), and the solid curve represents the cumulative mobilities, $\boldsymbol{M}^{(100)}$, induced by the slowest 100 modes. The correlation cosine between the two sets of data is 0.970. The corresponding correlation coefficient (see Experimental Procedures) is –0.5821 (Figure 5, inset). This anticorrelation strengthens to –0.8421 (black) and –0.8241 (gray) when one considers the respective pentamer-forming residues (chain G) and hexamer-forming residues (chains A–F) separately.

The experimental displacements of residues and

their predicted mobilities show surprisingly correlated patterns. What is intriguing is that a strong anticorrelation (rather than correlation) is observed. The theory predicts, for example, that the E loops are subject to very large mobilities, and thereby tend to reconfigure in the procapsid, which is intuitively consistent with their solvent exposure and observed reconfiguration to form the crosslinks. On the other hand, the displacement of the E loop between Prohead II and Head II appears to be the relatively small compared to other residues. These contradictory features can only be explained by a two-phase maturation dynamics, an isotropic expansion and overall surface flattening, leading to a relatively small net displacement of the E loop despite its high mobility. A superimposition of a uniform swelling of 110 Å on the predicted mobilities is indeed sufficient to explain the observed net displacement. These two processes are not necessarily sequential, but most likely occur concurrently, leading to the Head II structure.

A Few Icosahedrally Symmetric Modes Contribute to the Observed Reconfiguration

Inspired by the ability of a few modes to capture the expansion, we investigated each of the slowest 300 modes independently and calculated their contribution to this maturation process. To this aim, we calculated the correlation coefficient between the displacements vector $\boldsymbol{d}$ and the eigenvectors $\mathbf{u}_k$ of Γ, which are representative of the shape of the individual modes (see Experimental Procedures). The results are presented in Figure 6A in the form of cumulative squared correlation coefficients. It is clear that just a few modes show a distinguishable correlation with the observed displacements. As pointed out above, due to the highly symmetric nature of the viral capsid, most of the predicted modes are degenerate, which obscures the interpretation for individual modes. However, the six step changes in Figure 6A all refer to nondegenerate modes involving icosahedrally symmetric motions over the entire viral capsid. The slowest of these motions is produced by mode 31. Mode 31 involves only the residues of the pentamer-forming G chains (Figure 6B). Using similar coloring as in previous figures where blue signifies the most immobile regions and red the most mobile, combined modal contributions are shown for the next two modes (k = 107 and 144) that have a distinctive correlation with the observed displacements (Figure 6C).

Dynamics of Maturation

Our analysis sheds light on the most probable conformational changes that the Prohead II structure will make to reach the mature, Head II form. The first observation with regard to the dynamics of the procapsid is that the pentamers are the most mobile structural elements and thus have the highest potential to reconfigure during maturation. Because the frequency and hence time scale of each mode of motion is proportional to its corresponding eigenvalue (see Experimental Procedures), it is possible to construct a probable maturation pathway purely from the analysis of different modes of motion. The slower modes (lower mode numbers) may indeed be viewed as the softer modes that are likely to operate during the initial reconfiguration.

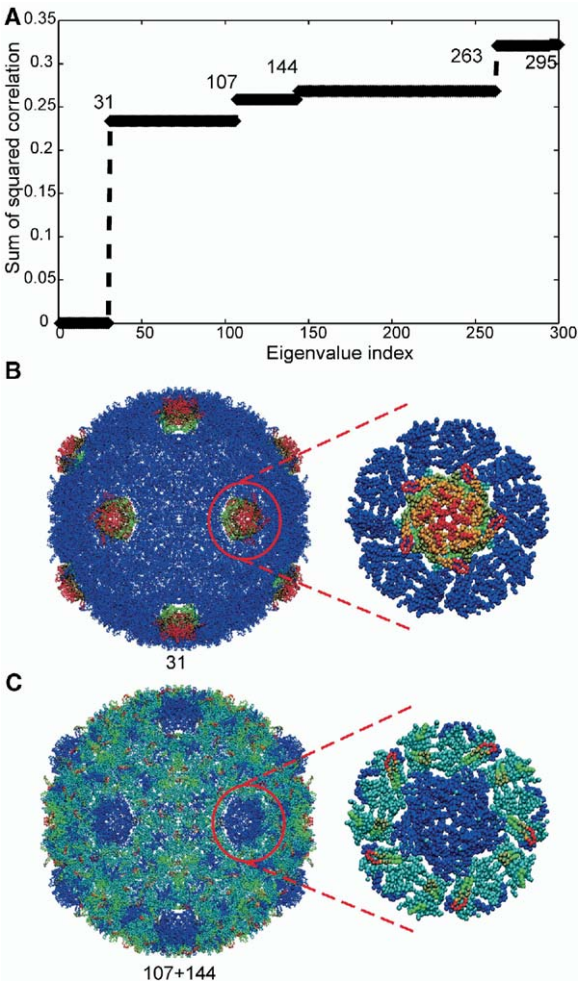

Figure 6. Contribution of Icosahedrally Symmetric Modes to Expansion from Procapsid to Capsid

(A) Cumulative sum squared of the calculated correlation coefficient (Equation 4) between single modes' mobilities and the net displacement of residues between Prohead II and Head II plotted versus mode index. Only five modes (31, 107, 144, 263, and 295) have nonnegligible correlation. (B) Graphical representation of mode 31, colored by relative fluctuation, indicates a cooperative, high degree of flexibility for each of the 12 pentamers (red). During expansion, these 12 pentamers pucker to reach the mature conformation. An individual pentamer-centered substructure (Pro_5) is magnified to indicate the stark difference between fluctuations in pentamers and hexamers in mode 31. The next two significant modes (107 and 144) have similar mode frequencies and hence describe concurrent fluctuations. (C) The combination of these modes suggests that mobility propagates away from the 12 pentamers upon maturation. The magnified pentamer-centered region for these modes suggests mobile hexamers encage an immobile pentamer, implicating portions of the A and P domains that participate in the flattening and straightening of the hexamers. Modes 263 and 295 (not shown) display similar, very localized dynamics for E-loop residues of three chains in each hexamer. These fluctuations are necessary for the crosslinking of Lys169 to Asn356 and final stabilization of the mature structure.

A second important observation is that a small subset of modes (Figure 6) governs the icosahedrally symmetric expansion of residues that takes place during

maturation. This does not imply, however, that the most cooperative but not icosahedrally symmetric modes (Figure 3) do *not* contribute. On the contrary, the rms displacements, or mobilities associated with these modes, exhibit a remarkable (anti)correlation with the actual displacements (Figures 4 and 5), suggesting that the reconfiguration that accompanies the icosahedrally symmetric expansion is essentially controlled by these modes. So, the observed displacements are explained by two countermechanisms: icosahedrally symmetric expansion and highly cooperative reconfigurations leading to an overall flattening and stabilization of the capsid shell in the mature form. The combination of these two mechanisms explains, for example, the observed smaller net displacements of the E loops between the two endpoints, although the E loops enjoy an extremely high mobility.

Among the icosahedrally symmetric modes, the first one (mode 31) points to the cooperative fluctuations of the pentamers that presumably play a critical role in precipitating the conformational changes that lead to Head II structure. The next nondegenerate motion, mode 107, augments this postulated reconfiguration by identifying domains from hexamer chains located at the interfaces between pentamers and hexamers as the most mobile regions. In view of their comparable eigenvalues (λ_{107} = 0.0361 and λ_{144} = 0.0425), mode 144 should operate on a similar time scale as mode 107 distinct from mode 31. The combination of these two modes (Figure 6C) illustrates how the 12 pentamers are encaged by more mobile regions that facilitate the change in conformations within the pentamers. These modes also identify active fluctuations involving domains of individual gp5 protein chains, which can be understood as a means to explore the flattening of hexamers required to adopt the Head II structure. This progression in mobile residues from pentamer to surrounding hexamers suggests that expansion propagates from the initial puckering of 12 icosahedrally symmetric pentameric faces to the rest of the capsid surface.

The two remaining nondegenerate modes (k = 263 and 295) are at least 1.5 times faster than the previous pair and identify localized motions of the highly flexible E loops in the hexamer-forming chains. These modes may be instrumental in the crosslinking of Lys169 from mobile E loops with neighboring gp5 proteins. It should be noted that in these modes, the pentamer-forming G chains are *not* predicted to be flexible, suggesting G-chain crosslinking on a different time scale, that is, after the cooperative crosslinking and flattening of hexameric chains. Recent investigations at low pH have shown that covalent crosslinking may occur concurrently with global expansion and local refolding, rather than in a purely sequential process (Gan et al., 2004), consistent with our interpretation of Figure 5.

Identification of Critical Interactions

Motivated by the ability to obtain a detailed description of fluctuation dynamics in Figure 2A, we made a closer analysis of the dynamics at the pentamer-centered substructures. Figure 7A displays the residue mobilities for a representative G chain from the pentamer-centered Prohead II (solid gray) and Head II (solid black) substructures indicated in Figures 7B and 7C, respectively. Interestingly, this mobility profile (solid gray), consistent with Figure 5 (solid black), is captured by the slowest (single) mode of the substructure. Additionally, this substructure slowest mode has a correlation coefficient of 0.969 with mode 31 (dashed gray) from the full procapsid structure analysis. The most striking difference between these curves is the significantly higher mobility of Prohead II compared to Head II. This difference in the overall mobility of the same chain in the two forms can be directly deduced from the ratio of the eigenvalues: $\lambda_1^{prohead}$ to λ_1^{head}, which is 0.363. Because the amplitudes of these motions scale inversely with their eigenvalues, the procapsid motions are 2.75 times larger than those of the mature capsid.

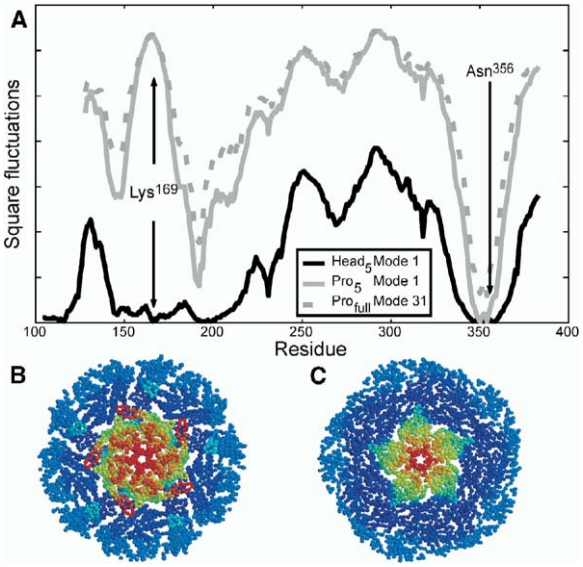

Figure 7. Distribution of Residue Fluctuations in the Slowest Icosahedrally Symmetric Mode for Chain G

The calculated mobilities refer to the slowest (mode 1) motions of the substructures Pro_5 (gray, solid) and $Head_5$ (black, solid), and to mode 31 of the intact procapsid (gray, dashed). (A) Modal mobility plotted against residue indicates larger amplitude motions in Prohead II than in Head II. Residues Lys169 and Asn356, which are responsible for covalent crosslinking, exhibit very different mobilities in the two structural conformations. (B) Pro_5 substructure for Prohead II is color coded according to the residue fluctuations in the slowest mode. (C) $Head_5$ substructure for Head II colored according to residue fluctuations indicates a much more constrained and stable structure.

Maxima in the procapsid curve reveal the key residues that are engaged in large fluctuations. Minima in the mature capsid curve, on the other hand, point to the key sites engaged in highly constrained, and consequently cooperative, interactions that stabilize the mature form. Several key interactions can thus be identified, including those among the residues in the vicinity of the 5-fold symmetry axis that undergo large fluctuations in both the Prohead II and Head II substructures. In vivo, one pentamer is replaced by a tail spike and portal protein complex that controls the entry and re-

lease of viral DNA. The high mobility of the pentamers may be relevant to this inherent biological function.

In Prohead II, the largest fluctuations are found in the E loop (residues 148–181; Figure 7A). Lys169 is located near this peak. Due to additional degrees of freedom from its long side chain, this exposed residue is likely to exhibit even greater mobility than indicated by the GNM for its C^{α} atom. The high mobility of Lys169 allows it to readily sample a large region in space and come into contact with other residues. This is a nontrivial motion involving a change in distance of nearly 30 Å between the two conformations (Conway et al., 2001). Complementing this high degree of mobility is Asn356, a severely constrained (immobile) amino acid located at a global hinge-bending site (minimum in the global mode shape). Upon crosslinking, both Lys169 and Asn356 are located near a minimum of the first mode severely constraining the mobility of Lys169 (Figure 7A). Thus, the catenation into a protein chainmail is achieved by coupling the highly flexible Lys169 to the highly stable Asn356. By allowing a flexible residue the freedom to sample conformations and recruit a highly immobilized partner residue, a very effective mechanism of stabilization is utilized. As such, Asn356 becomes a reliable, static target for covalent binding to the wandering Lys169 residue, efficiently arresting Lys169 into a highly confined state. We postulate that the isopeptide bond formed between these two residues would not be as stabilizing if Asn356 did not possess such a distinguishable stability before maturation. Additionally, we note that two spatial neighbors of Asn356, Gln191 and Ala192, are also located in the same hinge site in the procapsid to assist in stabilizing this residue as a reliable crosslinking target.

In conclusion, it is significant that a few global, icosahedrally asymmetric modes induce mobilities that correlate so strongly with the structural reconfigurations accompanying HK97 capsid maturation. As the most cooperative motions, these global modes describe the background fluctuations on which icosahedrally symmetric modes, such as mode 31, operate and are important to capture the details of residue configuration required for maturation. Both in the overall global modes (Figure 3) and the icosahedrally symmetric modes (Figure 6), the procapsid pentamers are distinguished as the most mobile regions. This suggests that the conformational mobility of the pentamers plays a crucial role in promoting capsid maturation. Although any of these vibrations are reversible individually, the cooperative superposition of many modes allows the observed crosslinking residues to come into contact. The formation of isopeptide bonds significantly alters the available motions, effectively locking the protein into the Head II structure.

Experimental Procedures

Building on the theory of elasticity in polymer networks (Flory, 1976), the equilibrium dynamics is fully described by the connectivity of the GNM that represents the capsid, or by the Kirchhoff matrix of contacts, Γ (Bahar et al., 1997; Haliloglu et al., 1997), based on a 7.3 Å cutoff distance between α carbons. A significant aspect of the GNM is the existence of an analytical solution. For a protein with N residues (nodes), diagonalization of Γ identifies a collection of *N-1* orthogonal modes of motion that fully describes the system dynamics.

The mobility of residue *i* induced by the *k*th mode is defined in terms of the nonzero eigenvalues (λ_k) and eigenvectors ($\mathbf{u}_k$) of Γ as

$$\boldsymbol{m}_{ii}^{(k)} = \sqrt{\lambda_k^{-1}[\boldsymbol{u}_k \boldsymbol{u}_k^T]_{ii}}.$$

The overall ms fluctuations $<(\Delta R_i)^2>$ of residue *i* are given by the summation

$$<\Delta \mathbf{R}_i^2> = (3k_BT/\gamma)\sum_k \lambda_k^{-1}[\mathbf{u}_k\mathbf{u}_k^T]_{ii} = (3k_BT/\gamma)\sum (\boldsymbol{m}_{ii}^{(k)})^2. \quad (1)$$

The elements of the *k*th eigenvector, $\boldsymbol{u}_k$, describe the displacements of the *N* sites along the *k*th mode (Haliloglu et al., 1997). Because each mode contributes by a factor inversely proportional to λ_k, the slowest modes (those with the smallest eigenvalues) contribute the most to the global dynamics and hence have the largest impact in functional motions (Bahar et al., 1998). The weighted sum of all modes scale with experimental B factors as

$$\boldsymbol{B}_i = (8\pi^2/3) < (\Delta \boldsymbol{R}_i)^2 > . \quad (2)$$

The force constant, γ, is the only adjustable parameter in the model. Its choice does not affect the relative values for the residue fluctuations, but uniformly rescales them. Based on a recent detailed study of 113 proteins, the value for k_bT/γ = 0.87 ± 0.46 Å^2 was used to calculate the fluctuations (Kundu et al., 2002). The cumulative mobility, $\boldsymbol{M}^{(n)}$, induced by a subset of modes (e.g., n = 100 modes in Figure 5, solid curve), is found by performing the weighted summation in Equation 1 over the subset of interest, and taking the square root of the computed squared fluctuations; that is, for a given residue *i*,

$$\mathbf{M}_i^{(n)} = \sqrt{\sum_{k=1}^{n} (\mathbf{m}_{ii}^{(k)})^2}.$$

Two measures of correlation were used to compare the computational and experimental data, $\vec{x}$ and $\vec{y}$. The first is the correlation cosine

$$\cos(\theta_{x,y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}||\vec{y}|}. \quad (3)$$

The correlation cosine in Figure 4 is calculated by replacing $\vec{x}$ and $\vec{y}$ by the net displacements $\boldsymbol{d}$ of residues and their mobilities $\boldsymbol{m}^{(k)}$ in the modes k = 1–300.

A second, more stringent measure is the correlation coefficient, ρ_{xy}, calculated by

$$\rho_{xy} = \frac{\Sigma (x_i - \mu_x)(y_i - \mu_u)}{\sqrt{\Sigma (x_i - \mu_x)^2 \Sigma (y_i - \mu_y)^2}}, \quad (4)$$

where μ_x is the mean value for the variable *x*. Figure 6 displays the cumulative sum of squared ρ_{xy} values where $\vec{x}$ and $\vec{y}$ have been replaced by $\boldsymbol{d}$ and $\boldsymbol{m}^{(k)}$.

In the application to supramolecular assemblies, the diagonalization of Γ can exceed the memory size of most computers and/or be prohibitively time consuming. HK97 procapsid and capsid contain N = 107,520 and 117,600 residues, respectively, which are excessively large for standard residue-level analysis. To circumvent this problem, we exploited the high degree of sparsity (0.000098) of Γ and analyzed the HK97 dynamics from two related perspectives. The first analysis involved calculating a subset of slow modes using the Lanczos method implemented in the BLZPCK algorithm (Marques, 2001), consistent with previous studies of large structures (Marques and Sanejouand, 1995; Tama and Brooks, 2002; Wang et al., 2004). This approach provides the dominant (slow mode) fluctuations for the entire (pro)capsid structure with residue-level detail. The second analysis relied upon the abstraction of the assembly as a substructure and a bath surrounding the substructure. This is achieved by expanding the size of Γ to *M + 1* for a substructure composed of *M* residues, the additional site corresponding to the mean-field approximation of the bath. Two different pentamer-centered substructures, comprised of all residues within 100 Å from the center, were examined: Pro_5 (*M* = 4,240) and $Head_5$ (*M* = 3,565). The remarkable agreement between the results pre-

dicted by the two analyses (illustrated in Figures 6 and 7) lends support to the robustness of the results.

Acknowledgments

Useful discussions with Roger Hendrix, Robert Jernigan, and Yongmei Wang are gratefully acknowledged. This work was partly supported by Pre-NPEBC NIH grant 1P20GM065805-01A1 to I.B.

Received: October 5, 2004
Revised: December 2, 2004
Accepted: December 3, 2004
Published: March 8, 2005

References

Atilgan, A.R., Durrell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys. J. *80*, 505–515.

Bahar, I., Atilgan, A.R., and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold. Des. *2*, 173–181.

Bahar, I., Atilgan, A.R., Demirel, M.C., and Erman, B. (1998). Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. Phys. Rev. Lett. *80*, 2733–2736.

Conway, J.F., Duda, R.L., Cheng, N., Hendrix, R.W., and Steven, A.C. (1995). Protolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. J. Mol. Biol. *253*, 86–99.

Conway, J.F., Wikoff, W.R., Cheng, N., Duda, R.L., Hendrix, R.W., Johnson, J.E., and Steven, A.C. (2001). Virus maturation involving large subunit rotations and local refolding. Science *292*, 744–748.

Delarue, M., and Sanejouand, Y.H. (2002). Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. J. Mol. Biol. *320*, 1011–1024.

Doruker, P., Jernigan, R.L., and Bahar, I. (2002). Dynamics of large proteins through hierarchical levels of coarse-grained structures. J. Comput. Chem. *23*, 119–127.

Duda, R.L. (1998). Protein chainmail: catenated protein in viral capsids. Cell *94*, 55–60.

Duda, R.L., Hempel, J., Michel, H., Shabanowitz, J., Hunt, D., and Hendrix, R.W. (1995). Structural transitions during bacteriophage HK97 head assembly. J. Mol. Biol. *247*, 618–635.

Flory, P.J. (1976). Statistical thermodynamics of random networks. Proc. R. Soc. Lond. A *351*, 351–380.

Gan, L., Conway, J.F., Firek, B.A., Cheng, N., Hendrix, R.W., Steven, A.C., Johnson, J.E., and Duda, R.L. (2004). Control of crosslinking by quaternary structure changes during bacteriophage HK97 maturation. Mol. Cell *14*, 559–569.

Haliloglu, T., Bahar, I., and Erman, B. (1997). Gaussian dynamics of folded proteins. Phys. Rev. Lett. *79*, 3090–3093.

Helgstrand, C., Wikoff, W.R., Duda, R.L., Hendrix, R.W., Johnson, J.E., and Liljas, L. (2003). The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. J. Mol. Biol. *334*, 885–899.

Hendrix, R.W., and Duda, R.L. (1998). Bacteriophage HK97 head assembly: a protein ballet. Adv. Virus Res. *50*, 235–288.

Hinsen, K., Thomas, A., and Field, M.J. (1999). Analysis of domain motions in large proteins. Proteins *34*, 369–382.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38.

Johnson, J.E., and Reddy, V.S. (1998). Biggest virus molecular structure yet! Nat. Struct. Biol. *5*, 849–854.

Keskin, O., Bahar, I., Flatow, D., Covell, D.G., and Jernigan, R.L. (2002). Molecular mechanisms of chaperonin GroEL-GroES function. Biochemistry *41*, 491–501.

Kim, M.K., Jernigan, R.L., and Chirikjian, G.S. (2003). An elastic network model of HK97 capsid maturation. J. Struct. Biol. *143*, 107–117.

King, J., and Chiu, W. (1997). The procapsid-to-capsid transition in double-stranded DNA bacteriophages. In Structural Biology of Viruses, W. Chiu, R.M. Burnett, and R. Garcea, eds. (New York: Oxford University Press), pp. 288–311.

Kundu, S., Melton, J.S., Sorensen, D.C., and Phillips, G.N., Jr. (2002). Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys. J. *83*, 723–732.

Lata, R., Conway, J.F., Cheng, N., Duda, R.L., Hendrix, R.W., Wikoff, W.R., Johnson, J.E., Tsuruta, H., and Steven, A.C. (2000). Maturation dynamics of a viral capsid: visualization of transitional intermediate states. Cell *100*, 253–263.

Lee, K.K., and Johnson, J.E. (2003). Complementary approaches to structure determination of icosahedral viruses. Curr. Opin. Struct. Biol. *13*, 558–569.

Li, G., and Cui, Q. (2002). A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca^{2+}-ATPase. Biophys. J. *83*, 2457–2474.

Marques, O. (2001). BLZPACK: Description and User's Guide. http://crd.lbl.gov/~osni/#Software.

Marques, O., and Sanejouand, Y.H. (1995). Hinge-bending motion in citrate synthase arising from normal mode calculations. Proteins *23*, 557–560.

Ming, D., Kong, Y., Wu, Y., and Ma, J. (2003). Substructure synthesis method for simulating large molecular complexes. Proc. Natl. Acad. Sci. USA *100*, 104–109.

Phelps, D.K., Speelman, B., and Post, C.B. (2000). Theoretical studies of viral capsid proteins. Curr. Opin. Struct. Biol. *10*, 170–173.

Simonson, T., and Perahia, D. (1992). Normal-modes of symmetrical protein assemblies: application to tobacco mosaic-virus protein disk. Biophys. J. *61*, 410–427.

Tama, F., and Brooks, C.L., III. (2002). The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. J. Mol. Biol. *318*, 733–747.

Tama, F., and Sanejouand, Y.H. (2001). Conformational change of proteins arising from normal mode calculations. Protein Eng. *14*, 1–6.

Tama, F., Gadea, F.X., Osni, M., and Yves-Henri, S. (2000). Building-block approach for determining low-frequency normal modes of macromolecules. Proteins *41*, 1–7.

van Vlijmen, H.W.T., and Karplus, M. (2001). Normal mode analysis of large systems with icosahedral symmetry: application to $(dialanine)_{60}$ in full and reduced basis set implementations. J. Chem. Phys. *115*, 691–698.

Wang, Y., Rader, A.J., Bahar, I., and Jernigan, R.L. (2004). Global ribosome motions revealed with elastic network model. J. Struct. Biol. *147*, 302–314.

Wikoff, W.R., Duda, R.L., Hendrix, R.W., and Johnson, J.E. (1999). Crystallographic analysis of the dsDNA bacteriophage HK97 mature empty capsid. Acta Crystallogr. D Biol. Crystallogr. *55*, 763–771.

Wikoff, W.R., Liljas, L., Duda, R.L., Tsuruta, H., Hendrix, R.W., and Johnson, J.E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. Science *289*, 2129–2133.

Xu, C., Tobi, D., and Bahar, I. (2003). Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin TR2 transition. J. Mol. Biol. *333*, 153–168.