



ACADEMIC  
PRESS

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Journal of Structural Biology 143 (2003) 107–117

Journal of  
**Structural  
Biology**

[www.elsevier.com/locate/yjsbi](http://www.elsevier.com/locate/yjsbi)

## An elastic network model of HK97 capsid maturation

Moon K. Kim,<sup>a</sup> Robert L. Jernigan,<sup>b</sup> and Gregory S. Chirikjian<sup>a,\*</sup>

<sup>a</sup> Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

<sup>b</sup> Laurence H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, IA 50011-3020, USA

Received 16 December 2002, and in revised form 25 June 2003

### Abstract

The structure of the capsid of bacteriophage HK97 has been solved at various stages of maturity by crystallography and cryo-electron microscopy, and has been reported previously in the literature. Typically the capsid assembles through polymerization and maturation processes. Maturation is composed of proteolytic cleavages to the precursor capsid (called Prohead II), expansion triggered by DNA packaging (in which the largest conformational changes of the capsid appear), and covalent cross-links of neighboring subunits to create the mature capsid called Head II. We apply a coarse-grained elastic network interpolation (ENI) to generate a feasible pathway for conformational change from Prohead II to Head II. The icosahedral symmetry of the capsid structure offers a significant computational advantage because it is not necessary to consider the whole capsid structure but only an asymmetric unit consisting of one hexamer plus an additional subunit from an adjacent pentamer. We also analyze normal modes of the capsid structure using an elastic network model which is also subject to symmetry constraints. Using our model, we can visualize the smooth evolution of capsid expansion and revisit in more detail several interesting geometric changes recognized in early experimental works such as rigid body motion of two compact domains (A and P) with two refolding extensions (N-arm and E-loop) and track the approach of the two particular residues associated with isopeptide bonds that make hexagonal cross-links in Head II. The feasibility of the predicted pathway is also supported by the results of our normal mode analysis.

© 2003 Elsevier Inc. All rights reserved.

**Keywords:** Conformational transitions; Elastic network interpolation; Intermediate structure; Symmetry constraints; Virus capsid

### 1. Introduction

Virus capsid maturation generally involves a major structural change of the procapsid (or prohead). Although the maturation processes vary from system to system, most procapsids have certain common characteristics. For example, the procapsid is round, whereas the mature capsid has polyhedral shape (Fig. 1). During maturation, capsid proteins can be changed by contact with scaffolding proteins that assist in capsid assembly reactions and then are removed (Flint et al., 2000) or newly attached proteins can bind to the capsid. Translocation between inner and outer surfaces (Steven et al., 1991), alternation of secondary structures (Steven et al., 1990), and covalent cross-links between neighboring subunits (Duda, 1998) have been investigated in the

previous literature. These are thought to be irreversible and exothermic reactions. The resulting structural changes may be very large and involve rigid body motions of subunits or motif refolding (Conway et al., 2001).

The double-stranded DNA bacteriophage HK97 is a good example for studying this phenomenon of capsid maturation. Recently the structure of HK97 capsid has been determined with higher resolution than before by using X-ray crystallography and cryo-electron microscopy (Conway et al., 2001; Wikoff et al., 2000). Four discrete stages have been classified as follows: (1) the gp5 capsid protein associates into both pentamers and hexamers to become the earliest precursor with the triangulation number  $T = 7$ , Prohead I; (2) proteolytic cleavage of gp5 associated with the gp4 viral protease produces Prohead II; (3) expansion to Head I is induced by DNA packaging in vivo, while several chemical stimuli may trigger this event in vitro; and (4) covalent cross-links occur between neighboring subunits

\* Corresponding author. Fax: 1-410-516-7254.

E-mail address: [gregc@jhu.edu](mailto:gregc@jhu.edu) (G.S. Chirikjian).

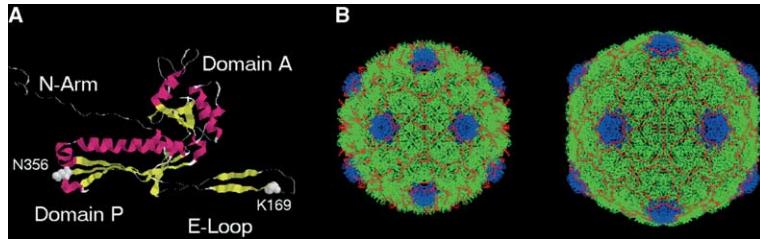


Fig. 1. Capsid structure in Prohead II and Head II. (A) The capsid protein unit consists of the rigid core, A and P domains, and two flexible motifs, N-arm and E-loop. The cross-linked residues, Lys<sup>169</sup> and Asn<sup>356</sup>, are marked with space-filling models. This figure is adopted from Conway et al., 1995. (B) Prohead II (left) and Head II (right) are constructed from C<sub>α</sub> backbones. Pentamers are blue and hexamers are green.

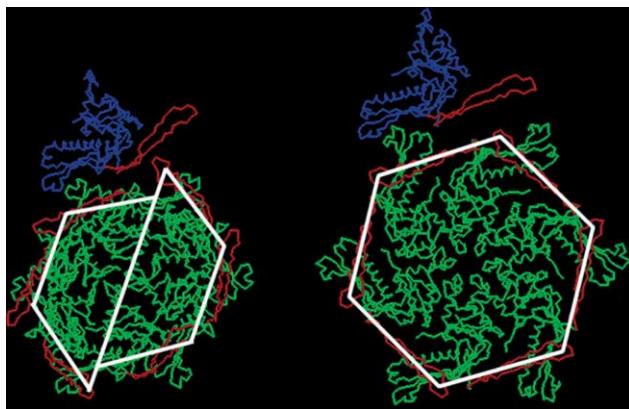


Fig. 2. An asymmetric unit in Prohead II (left) and Head II (right). Pentamers are blue and hexamers are green. All E-loops are red. Skewed trimers are indicated by white lines. During expansion, the hexamer dislocation disappears and becomes perfectly 6-fold symmetric (Conway et al., 2001).

producing the mature capsid, Head II (Conway et al., 1995, 2001; Duda, 1998; Duda et al., 1995; Lata et al., 2000; Wikoff et al., 2000).

During expansion of the capsid, the structural core of subunits (A and P domains) appear to remain rigid, in contrast with two motifs (N-arm and E-loop) which refold (Fig. 1). These changes enable the capsid to be much more stable by increasing the amount of surface area buried at interfaces and bringing the cross-link residues Lys<sup>169</sup> and Asn<sup>356</sup> close together (Conway et al., 2001). 420 isopeptide bonds entangle monomers into covalent rings, which are concatenated by icosahedral symmetry to form protein chainmail (Duda, 1998; Wikoff et al., 2000). In addition, the shear dislocation into two trimers in Prohead states is eliminated and then the hexamers become almost 6-fold symmetric in Fig. 2 (Conway et al., 1995, 2001; Lata et al., 2000).

Physically realistic simulation of the maturation process can provide new insights. Molecular Dynamics (MD) simulation, a powerful method for the study of details of molecular motion, and normal mode analysis (NMA) using all-atom empirical potentials, are often used to follow the dynamics of proteins (Brooks and

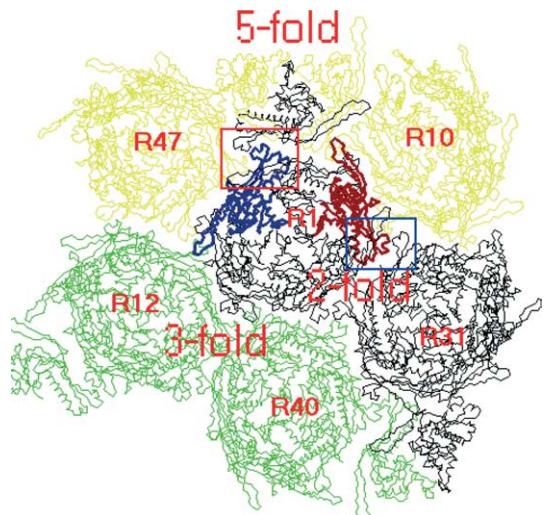


Fig. 3. The assembly of asymmetric units. The center unit, R1, is completely surrounded by 5 identical ones. Yellow units, R10 and R47, and two other units (not displayed here) collectively form a 5-fold cluster with the center unit. Likewise, green units, R12 and R40, together with the center unit form a 3-fold cluster. These surrounding units impose symmetry constraints on the deformation of the center unit during maturation. Each numbered symbol, R, indicates a specific rotation matrix listed in Table 1. Chains B (red) and F (blue) are painted to compare with each other. The tail of Chain B heads for a 5-fold vertex, but that of Chain F points out the bottom of an asymmetric unit. In addition, observing the boxed regions, Chain F (red box) is connected to Chain G which is a component of a pentamer. However, Chain B (blue box) just lies on a 2-fold line. Each chain has its own set of inter-residue interactions caused by different surroundings. This explains why mode shapes of the 7 chains that constitute each asymmetric unit are somewhat different from each other in Fig. 8.

Karplus, 1983). However, the use of atomic approaches becomes computationally inefficient with the increased size of a system. To reduce such a computational burden, many recent papers have demonstrated the utility of coarse-graining protein models by including, for example, only C<sub>α</sub> atoms as point masses representing residues and using a simplified potential for considering internal interactions between neighboring residues. Such models are suitable to describe the global motions of complex systems composed of multiple small proteins such as virus capsids (Atilgan et al., 2001; Bahar and

Jernigan, 1998). More recently, the swelling process of the icosahedral virus, cowpea chlorotic mottle virus (CCMV), was studied by NMA (Tama and Brooks, 2002). In that work, a series of intermediate structures of CCMV were produced from displacement along a linear combination of normal modes with an energy minimization process. In other works, NMA of large icosahedral systems has been made more efficient by using group theory. The use of symmetry constraints block-diagonalizes the Hessian (i.e., stiffness) matrix to more efficiently compute all the normal modes of icosahedral structures, regardless of whether these motions are symmetric or not (Vlijmen and Karplus, 2001). The substructure synthesis method (SSM) is also reported (Ming et al., 2003) in which the normal modes for the assembled structure can be synthesized from the substructure modes by applying the Rayleigh–Ritz principle.

In our previous publications, we developed a new interpolation method (called ENI) for generating feasible pathways for conformational transitions using coarse-grained protein models. The basic concept is to interpolate uniformly the distances between spatially proximal residues in both conformations within the context of an elastic network model. NMA is not necessary for producing intermediate structures in this method. Since we interpolate relative distance between spatially close residues, unrealistic conformations and steric clashes become far less likely (Kim et al., 2002a,b).

The HK97 capsid has on the order of one hundred thousand residues. Hence, it is still very difficult to handle such a huge data set on a PC even when a coarse-grained C<sub>x</sub> model is used. To overcome this computational limitation, we use the icosahedral symmetry of the capsid to derive a modified version of ENI for symmetric structures in which it is not necessary to keep the whole data set to generate intermediates, but only a repeated unit is required with symmetry constraints. This idea is also applied here to NMA of the HK97 capsid. In both our symmetry-constrained ENI method and NMA, we restrict our attention to symmetric motions, thus efficiently reducing the size of our system by a factor of 60. Since there is currently no direct method to measure the maturation process and no stable intermediate structures is known yet, our predicted pathway may bring new insights into the maturation mechanism of the HK97 capsid. In addition, NMA results for the Prohead II structure provide a way to verify the realism of pathway generated by ENI.

In the future, our methodology might be applied to another  $T = 7$  isometric bacteriophage P22. The maturation transition of P22 has been revealed by cryo-electron microscopy and image reconstruction (Zhang et al., 2000), but high-resolution atomic coordinates have not yet been reported.

## 2. Methods: ENI with symmetry constraints

### 2.1. NMA using a coarse-grained elastic network model

Once structures have been obtained from the Protein Data Bank (Berman et al., 2000), we can define a set of  $N$  representative atoms (e.g., C<sub>x</sub> atoms in the present case). The position of the  $i$ th atom at time  $t$  is denoted

$$\mathbf{x}_i(t) = [x_i(t), y_i(t), z_i(t)]^T. \quad (1)$$

When defining the vector norm as  $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + v_3^2}$  for any  $\mathbf{v}$ , the total kinetic energy in a network of  $N$  point masses has the form

$$T = \frac{1}{2} \sum_{i=1}^N m_i \|\dot{\mathbf{x}}_i(t)\|^2 = \frac{1}{2} \dot{\boldsymbol{\delta}}^T M \dot{\boldsymbol{\delta}}, \quad (2)$$

where we define a  $3N$ -dimensional vector  $\boldsymbol{\delta}$  in which  $\boldsymbol{\delta}_i(t)$  is the vector describing small displacements of the  $i$ th point mass,

$$\mathbf{x}_i(t) = \mathbf{x}_i(0) + \boldsymbol{\delta}_i(t), \quad (3)$$

$$\boldsymbol{\delta} = [\boldsymbol{\delta}_1^T, \dots, \boldsymbol{\delta}_N^T]^T, \quad (4)$$

and the matrix  $M$  is the global mass matrix for the whole network (Kim et al., 2002a,b).

The total potential energy in a network of connected springs has the form

$$V = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N k_{i,j} \{ \|\mathbf{x}_i(t) - \mathbf{x}_j(t)\| - \|\mathbf{x}_i(0) - \mathbf{x}_j(0)\| \}^2, \quad (5)$$

where  $k_{i,j}$  is the  $(i,j)$  element of the “linking matrix” or “contact matrix”, which is assumed to be unity for all contacting pairs and zero for pairs not in contact, regardless of the residue types concerned. Springs represent close residues, all interacting in identical ways, and the elastic potential energy follows a harmonic potential, appropriate for small deviations from equilibrium. In general, Eq. (5) is a non-linear function of the deformations even though the springs are linear. However, when we assume that the deformations are small,  $V$  can be approximated as a classical quadratic potential energy function

$$V = \frac{1}{2} \boldsymbol{\delta}^T K \boldsymbol{\delta}, \quad (6)$$

where the matrix  $K$  is the stiffness matrix for the whole network (Kim et al., 2002a,b).

Finally we can obtain the equation describing harmonic motions of the capsid protein composed of  $N$  residues as

$$M \ddot{\boldsymbol{\delta}} + K \boldsymbol{\delta} = 0. \quad (7)$$

Since we use Cartesian coordinates, the elements of  $M$  are of the form  $M_{i,j} = m_i \delta_{i,j}$  (i.e.,  $M$  is diagonal). If  $m_i = m$ , then normal modes are the eigenvectors of  $K$ .

Normal modes generated using this coarse-grained model can be used to evaluate potential motions about a single equilibrium conformation of a large capsid protein with relatively little computational cost.

## 2.2. Elastic network interpolation (ENI)

The main idea of ENI is to interpolate between two values for the distances between spatially proximal  $C_\alpha$  atoms. Since this interpolation must resolve conflicting goals, it is formulated as the minimization of a quadratic cost function. One can generate intermediate conformations of a virus capsid by finding small changes in positions of  $C_\alpha$  atoms that result from inducing correspondingly small changes in inter-residue distances.

Suppose that we have sets of  $C_\alpha$  coordinates for Prohead II and Head II denoted as  $\{\mathbf{x}_i\}$  and  $\{\mathbf{z}_i\}$ , respectively. We introduce a cost function as follows

$$C(\boldsymbol{\delta}) = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N k_{i,j} \{ \|\mathbf{x}_i + \boldsymbol{\delta}_i - \mathbf{x}_j - \boldsymbol{\delta}_j\| - l_{i,j} \}^2. \quad (8)$$

Here  $\boldsymbol{\delta} = [\boldsymbol{\delta}_1^\top, \dots, \boldsymbol{\delta}_N^\top]^\top$  is a  $3N$ -dimensional vector of displacements with  $N$  being the number of residues. An intermediate conformation is defined by the value of  $\boldsymbol{\delta}$  that minimizes this cost when all other parameters are held constant. Here the linking matrix is formed as the “union” of the two linking matrices for  $\{\mathbf{x}_i\}$  and  $\{\mathbf{z}_i\}$  so that  $k_{i,j}$  can have non-zero value whenever residues  $i$  and  $j$  are judged to be in contact in either conformation. The value  $l_{i,j}$  is the desired distance between  $i$  and  $j$ , which can be chosen as

$$l_{i,j} = (1 - \alpha) \|\mathbf{x}_i - \mathbf{x}_j\| + \alpha \|\mathbf{z}_i - \mathbf{z}_j\|, \quad (9)$$

where  $\alpha$  is the coefficient specifying how far a given state is along the transition from  $\{\mathbf{x}_i\}$  to  $\{\mathbf{z}_i\}$ . Using the “union” linking matrices confines the intermediate conformations to the interval between the two end conformations. This guarantees all the resultant intermediates do not have spatially very close residue pairs which are physically unfavorable during the maturation process. This statement is quantitatively verified later.

Our goal is to find values of  $\boldsymbol{\delta}$  that minimize Eq. (8), which itself can be approximated for small values of  $\|\boldsymbol{\delta}_i\|$  and  $\|\boldsymbol{\delta}_j\|$  with the Taylor series approximation

$$C(\boldsymbol{\delta}) \approx \frac{1}{2} \boldsymbol{\delta}^\top \boldsymbol{\Gamma} \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\gamma}^\top \boldsymbol{\delta} + B, \quad (10)$$

where  $\boldsymbol{\Gamma}$  is a  $3N \times 3N$  matrix,  $\boldsymbol{\gamma}$  is a  $3N$ -dimensional row vector, and  $B$  is a constant (Kim et al., 2002a,b). The reduction of Eqs. (8)–(10) is analogous to that of Eqs. (5) and (6). Only now the goal is not the determination of normal modes, but rather finding displacements. We minimize  $C(\boldsymbol{\delta})$  with respect to  $\boldsymbol{\delta}$ , which results in the following constraint equation

$$\frac{\partial C(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = \boldsymbol{\Gamma} \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\gamma}^\top = 0. \quad (11)$$

In general,  $\boldsymbol{\Gamma}$  is a singular matrix because of the translational invariance of ENI. To get the unique displacement, one can add the constraint of linear momentum conservation (Kim et al., 2002a,b). However, the symmetry-constrained ENI introduced later in this paper is not invariant under translation because icosahedral symmetry does not allow net rigid-body translation of subunits during the maturation. Hence the solution is deterministic.

In the implementation, we calculate  $\boldsymbol{\delta}$  to be the solution of Eq. (11) when  $\alpha = 0.01$ . Then we obtain the first intermediate conformation denoted by  $\{\mathbf{x}_i^1\}$ , which is 1% along the path from Prohead II to Head II. That is

$$\mathbf{x}_i^1 = \mathbf{x}_i + \boldsymbol{\delta}_i, \quad (12)$$

where  $\mathbf{x}_i^1$  is the position of the  $i$ th residue out of the set  $\{\mathbf{x}_i^1\}$ . Likewise for the next incremental conformation  $\{\mathbf{x}_i^2\}$ ,

$$\mathbf{x}_i^2 = \mathbf{x}_i^1 + \boldsymbol{\delta}_i, \quad (13)$$

where  $\boldsymbol{\delta}$  is the solution of Eq. (11) when  $\alpha = 0.02$  in the next incremental step. The remaining conformations are then obtained in this iterative way.

## 2.3. Topological review of HK97 capsid

(Pro)Head II consists of 60 identical asymmetric units. Each of them is composed of 7 polypeptide chains with identical secondary structures as shown in Fig. 2. The chains A–F form a hexamer and the chain G is in part of a pentamer. The PDB entry “1IF0” (Conway et al., 2001) is a quasi-atomic model of Prohead II which was produced by homology modeling from the 3.6 Å Head II structure (i.e., 7 identical chains are fitted into an electron microscopy map as rigid bodies). Hence, there is no surprise that the secondary structures in A and P domains are preserved during maturation as a result of artificial mapping. The E-loop appears to act as a hinge relative to the domains A and P. It brings cross-link residues close together in the Head II state. The sequence of Prohead II is a truncated version of PDB entry “1FH6” of Head II (Wikoff et al., 2000). 1IF0 contains only 256  $C_\alpha$  coordinates from Gly<sup>128</sup> to Ser<sup>383</sup>, while 1FH6 has all atomic coordinates of 280 residues from Ser<sup>104</sup> to Ser<sup>383</sup>. Hence the reconstruction of the Prohead II capsid includes  $60 \times 7 \times 256 = 107\,520$   $C_\alpha$  atoms as Head II needs  $60 \times 7 \times 280 = 117\,600$   $C_\alpha$  atoms. To visually compare secondary structures of Prohead II and Head II we disregard residues Ser<sup>104</sup> to Pro<sup>127</sup> on the N-arm of Head II which are not seen in Prohead II (Fig. 1).

To generate a conformational transition using ENI with symmetry constraints, we need to know how a capsid asymmetric unit is surrounded by other identical units. Among the given 60 rotation matrices in the PDB entries, 5 particular rotation matrices correspond to

Table 1  
Axis-angle parameterization of rotation matrices relative to the identity rotation R1

Rotation matrix	Angle (deg) <sup>a</sup>	Components of axis vector <sup>b</sup>		
		x	y	z
R10	72	0.8507	0.5257	0
R12	120	0.9342	0	0.3578
R31	180	1	0	0
R40	-120	0.9342	0	0.3578
R47	-72	0.8507	0.5257	0

<sup>a</sup> Magnitude of rotation angle with respect to its rotation axis.

<sup>b</sup> Cartesian coordinates of the normalized rotation axis

neighboring units of each unit as shown in Fig. 3. We note that an asymmetric unit can be surrounded by these 5 identical ones. Table 1 presents the axis–angle parameterization of those rotation matrices. R10 and R47 duplicate the original coordinates of a unit through rotation about the 5-fold axis by  $\pm 72^\circ$ , respectively. R31 represents the rotation about the 2-fold axis by  $180^\circ$ . R12 and R40 make a rotation about the 3-fold axis by  $\pm 120^\circ$ , respectively.

The minimum distance between  $C_\alpha$  atoms within each of these six asymmetric units in Prohead II (including the original one and five surrounding neighbors) is  $1.8 \text{ \AA}$ , corresponding to Gly<sup>128</sup> in chain E and Gly<sup>328</sup> in chain F within a subunit. Similarly the minimum value encountered in Head II is  $3.3 \text{ \AA}$  between Ala<sup>272</sup> and Gly<sup>328</sup> in the same chain A. It is noted that copying and juxtaposing units does not cause steric clashes with the original unit at all because the resulting minimal intercapsomer distance is not smaller than the minimum intracapsomer distance mentioned above. In addition, capsid swelling can be inferred from the fact that the magnitude of the minimum intracapsomer distance increases by about a factor of two upon maturation.

#### 2.4. Linking matrix construction

The behavior of the elastic network model of a protein depends on the cutoff values for defining contacts. Large cutoff values give rise to an increase in the number of interacting residue pairs and a greater cohesiveness. Consequently, systems become stiffer and motions are usually more cooperative. Also for relatively short cutoff values, there can be large amplitude fluctuations along particular directions for particular residues. The present interpolation model is likewise sensitive to cutoff values. Short cutoff values force residues to be in contact with only local neighbors. This can sometimes cause unrealistic results that lead to discontinuous motions. On the other hand, larger cutoff values make a denser linking matrix which tremendously increases computation time for generating intermediate transitions in large protein models such as the HK97 capsid. To assure having uniformly sparse linking matrices we can impose a cutoff

on the number of residue contacts, instead of using a cutoff distance. Namely, we can connect residues to their neighbors in order from the closest one, increasing the distance gradually until the fixed limiting number is reached. This method reduces computational costs for the whole interpolation process and also guarantees realistic results (Kim et al., 2002a,b).

The connectivity between  $C_\alpha$  atoms is defined both within each asymmetric unit and between every pair of asymmetric units in contact as seen in Fig. 3. Even though this reduces the dimensions by a factor of 10 relative to the size of the whole capsid, it is still so huge (i.e., 6 asymmetric units  $\times$  7 chains  $\times$  256 residues = 10 752). Therefore, we need to adopt sparse matrix methods to overcome memory limitations, then can divide such a huge matrix into small pieces to handle it with ease. Fig. 4 shows the pattern of sparseness of the linking matrix. The marks on off-diagonal blocks indicate the intercapsomer connections. To save computing time and consider local packing effect, we actually take a cubic box the center of which is the position of a particular residue. The dimensions of the box are  $\pm 15 \text{ \AA}$  in each direction. We then make connections between the center residue and others located inside the box until 20 connections are achieved. We use two different spring constants 1 and 10. The spring constant of 1 is used for inter-connections such as domain–domain interactions or domain–loop interactions. The spring constant of 10 is used for intra-connections within the same domains or loops, as well as for the backbone. This reflects rigid-body motions of A and P domains in ENI model.

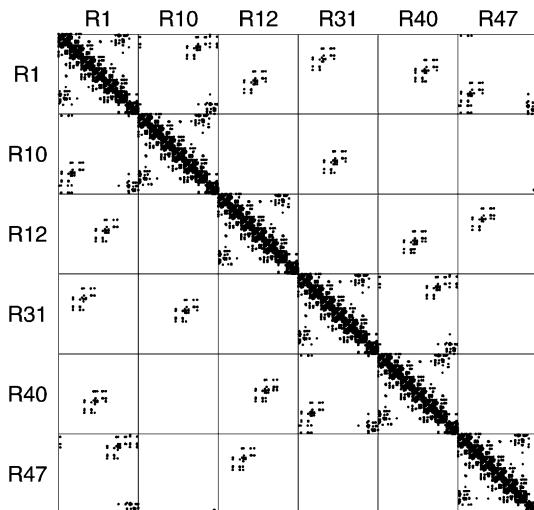


Fig. 4. The pattern of sparseness of the linking matrix with the constant cutoff number of 20. The intracapsomer interactions are marked on diagonal blocks, while the intercapsomer interactions are represented on off-diagonal blocks. R1–R47 indicate the original unit and its surrounding ones as seen in Fig. 3.

## 2.5. Symmetry-constrained ENI model

Since both the Prohead II and Head II structures of the HK97 capsid have icosahedral symmetry, the conformational change of each asymmetric unit is assumed to be identical and compatible with those of neighboring units. That is, we impose constraints on our model such that all intermediates also have icosahedral symmetry. We apply this hypothesis to ENI in order to get a reduced form of Eq. (10). This enables us to generate icosahedrally symmetric intermediate conformations using only a single asymmetric unit with symmetry constraints on behalf of the whole capsid structure composed of 60 identical asymmetric units.

If there are 60  $3 \times 3$  rotation matrices  $R_i$  ( $i = 1, 2, \dots, 60$ ) such that  $R_i R_j = R_k$  for any  $i, j \in [1, 60]$  and some  $k \in [1, 60]$ , and if there is a total of  $N = 60n$   $C_\alpha$  atoms, then using unconstrained ENI,

$$C(\delta) = \frac{1}{2} \delta^T \Gamma \delta + \frac{1}{2} \gamma \delta + B, \quad (14)$$

where  $\Gamma$  is a  $3N \times 3N$  matrix,  $\gamma$  is a  $3N$ -dimensional row vector such that  $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_{60}]$ , and  $B$  is a constant. Here  $\delta$  can be defined as

$$\delta = [\Delta_1^T, \dots, \Delta_{60}^T]^T, \quad (15)$$

and  $\Delta_i$  is a  $3n$ -dimensional vector consisting of all the displacements for one asymmetric unit. The symmetry constraint means that we can divide  $\Gamma$  into  $60 \times 60$  blocks  $\Gamma_{i,j}$  in which each one is a  $3n \times 3n$  matrix and express  $\Delta_i$  ( $i = 1, 2, \dots, 60$ ) as a function of  $\Delta_1$  such that

$$\Delta_i = \widehat{R}_i \Delta_1, \quad (16)$$

where  $\widehat{R}_i$  is the  $3n \times 3n$  matrix which has  $R_i$  repeated  $n$  times as diagonal blocks. Hence,

$$\delta^T \Gamma \delta = \sum_{i=1}^{60} \sum_{j=1}^{60} \Delta_1^T \widehat{R}_i^T \Gamma_{i,j} \widehat{R}_j \Delta_1 = 60 \Delta_1^T \Gamma' \Delta_1, \quad (17)$$

where

$$\begin{aligned} \Gamma' = & \widehat{R}_1^T \Gamma_{1,1} \widehat{R}_1 + \widehat{R}_1^T \Gamma_{1,10} \widehat{R}_{10} + \widehat{R}_1^T \Gamma_{1,12} \widehat{R}_{12} \\ & + \widehat{R}_1^T \Gamma_{1,31} \widehat{R}_{31} + \widehat{R}_1^T \Gamma_{1,40} \widehat{R}_{40} + \widehat{R}_1^T \Gamma_{1,47} \widehat{R}_{47}. \end{aligned} \quad (18)$$

The matrix  $\Gamma'$  can be generated using only the coordinates of 6 asymmetric units (Fig. 3) and their linking matrix (Fig. 4). This is because every asymmetric unit is surrounded by neighbors in exactly the same way as the representative we have chosen. Likewise,

$$\gamma \delta = \sum_{i=1}^{60} \gamma_i \widehat{R}_i \Delta_1 = 60 \gamma' \Delta_1, \quad (19)$$

where  $\gamma' = \gamma_1 \widehat{R}_1$ . Substitution into Eq. (14) yields the reduced form of the cost function such that

$$C(\Delta_1) = \frac{60}{2} \Delta_1^T \Gamma' \Delta_1 + \frac{60}{2} \gamma' \Delta_1 + B, \quad (20)$$

where  $\Gamma'$  is a  $3n \times 3n$  non-singular matrix and  $\gamma'$  is a  $3n$ -dimensional row vector. The solution  $\Delta_1$  which minimizes the reduced cost function uniquely determines the intermediate conformation of an asymmetric unit. We can reconstruct the whole capsid by copying this unit resulting from the symmetry-constrained ENI model.

The stiffness matrix in NMA (Eq. (6)) also can be reduced in this manner as

$$V(\Delta_1) = \frac{60}{2} \Delta_1^T K' \Delta_1, \quad (21)$$

where  $K'$  is a  $3n \times 3n$  non-singular matrix. Obviously the normal modes calculated from  $K'$  represent only the subset of the total that preserve icosahedral symmetry of the capsid. If it were necessary to obtain all the normal modes of the capsid, regardless of whether they preserve capsid symmetry or not, methods such as those in (Vlijmen and Karplus, 2001) would be required. However, this is not the case here.

## 3. Results

### 3.1. Maturation pathway

Using the symmetry-constrained ENI, we generate 99 intermediate conformations between Prohead II and Head II. Fig. 5 shows conformational transitions of an asymmetric unit. The expanded capsid, Head II, is apparently different from Prohead II being a polyhedron shape and larger in diameter. The displacements vary among capsid chains (A–G) but each chain moves radially so that the concave hexamer becomes much flatter (Conway et al., 2001). The shear dislocation of about 25 Å in the hexamer is also eliminated (Conway et al., 1995).

A measurement of the realism of the simulation results is the fluctuations observed in the virtual bond angles. Since we only take  $C_\alpha$  atoms in the coarse-grained network modeling, the bond angle can be defined “virtually” as the angle formed by three consecutive  $C_\alpha$  atoms. As discussed earlier, homology modeling apparently preserves secondary structures of the capsid core (A and P domains) during the transition. Hence their virtual bond angles change little between Prohead II and Head II. However, major changes appear near the E-loop which moves 30–40 Å to bring two isopeptide bond sites close to each other. For example, we can see from the simulation result that the distance between Lys<sup>169</sup> of chain G of the original asymmetric unit and Asn<sup>356</sup> of chain F of the unit marked as R10 (Fig. 3) decreases monotonically from 33 to 9 Å. Such a large deformation is apparent even using ENI with the 12 Å-resolution Prohead II structure. However, it is limited to capture small local changes which may occur during the real maturation process. Because Prohead II and Head II are assumed to have the same rigid-body

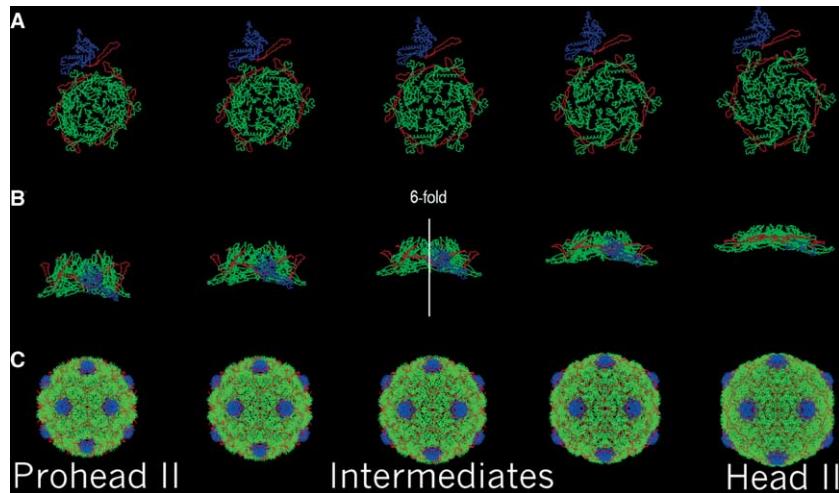


Fig. 5. Simulation of computed intermediate conformations between Prohead II and Head II of the HK97 capsid. (A) An asymmetric unit swells and then the shear dislocation of the hexamer disappears. (B) The side view of an asymmetric unit shows the position change of chains causing this unit to become thinner. (C) The rounder capsid expands to a more polyhedral conformation. The whole capsid is reconstructed with 60 copies of identical asymmetric units.

domains by homology modeling and ENI is subject to those two end structures. In Fig. 6, simulation results show that virtual bond angles change smoothly and continuously when following a pathway generated using ENI from Prohead II to Head II. The striking angle changes are viewed only near the junctions between the E-loop and other domains. The lower boundary indicates the set of minimum bond angles observed in each intermediate conformation while the upper boundary corresponds to the set of maximum bond angles. The middle line shows a smooth change in the virtual bond angle of Asn<sup>146</sup> where the biggest angle change appears. Virtual bond angles of other residues also vary smoothly within these boundaries.

As a simpler alternative to ENI, one could try to simulate a conformational transition from Prohead II to Head II using Cartesian (linear) interpolation of an asymmetric unit. In this approach, Cartesian coordinates of both conformations are linearly interpolated to generate intermediate conformations. The whole capsid structure in each intermediate stage would still be made by duplicating the interpolation result of one unit and assembling these units using the 60 rotational matrices provided in the PDB. Since the interactions among neighbor units would not be modelled in this case, one couldn't ensure that each intermediate would observe steric constraints well. However, the advantage of the symmetry-constrained methodology is that it observes intercapsomer steric constraints. Fig. 7A shows the minimum distances between C<sub>z</sub> atoms during transitions in both cases. Symmetry-constrained ENI prevents the minimum distance from going below its initial value, whereas Cartesian interpolation results in lower values in some ranges. It follows that ENI results should be more realistic in the sense of sterics. Fig. 7B displays

RMSD with respect to Prohead II. The ENI pathway is somewhat different from the Cartesian pathway by 2 Å RMSD. Fig. 7C draws the “virtual” bond length between C<sub>z</sub> atoms of Ser<sup>145</sup> and Asn<sup>146</sup>. This is the hinge junction between the N-arm and E-loop. Cartesian interpolation results show unrealistically small distance values when compared with those of ENI. In addition, rigid-body domains of the HK97 capsid are preserved in ENI better than in Cartesian interpolation as shown in

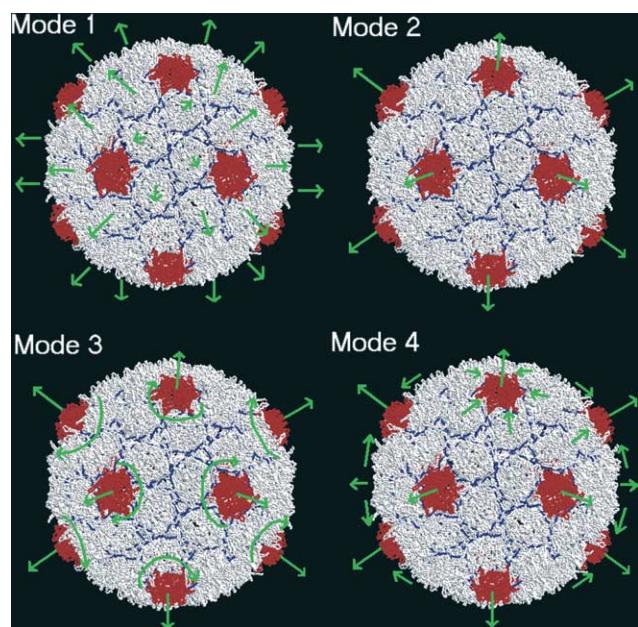


Fig. 9. Normal mode shapes of the HK97 capsid. These cartoons illustrate non-rigid-body normal modes 1–4. Mode 1 shows the expansion of hexamers. In contrast, mode 2 shows the expansion of pentamers. Mode 3 contains two motions such as hexamer rotation and pentamer expansion. Mode 4 displays hexamers pushing on pentamers, which are lifted radially.

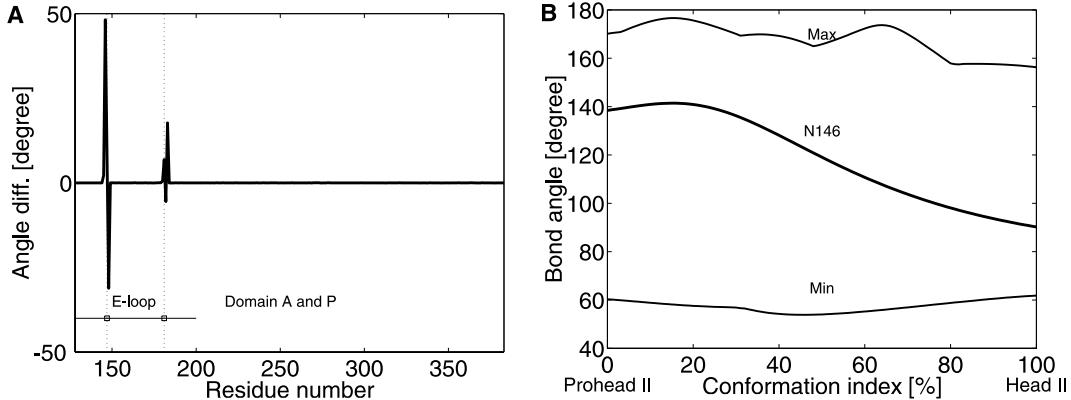


Fig. 6. Virtual bond angle changes between Prohead II and Head II. (A) The virtual bond angle difference between Prohead II and Head II structure is shown for chain A. Most residues do not change their virtual bond angles except for a few residues located near the junctions between the E-loop and other domains. This implies that a hinge-like motion is favored at the E-loop. Likewise, Chains B–G have similar angle changes at both ends of each loop (not displayed here). (B) The lower line presents the set of minimum bond angles, while the upper limit corresponds to the set of maximum bond angles. For example, the middle line shows the maximum virtual bond angle change of Asn<sup>146</sup> which varies continuously over about a 50° range.

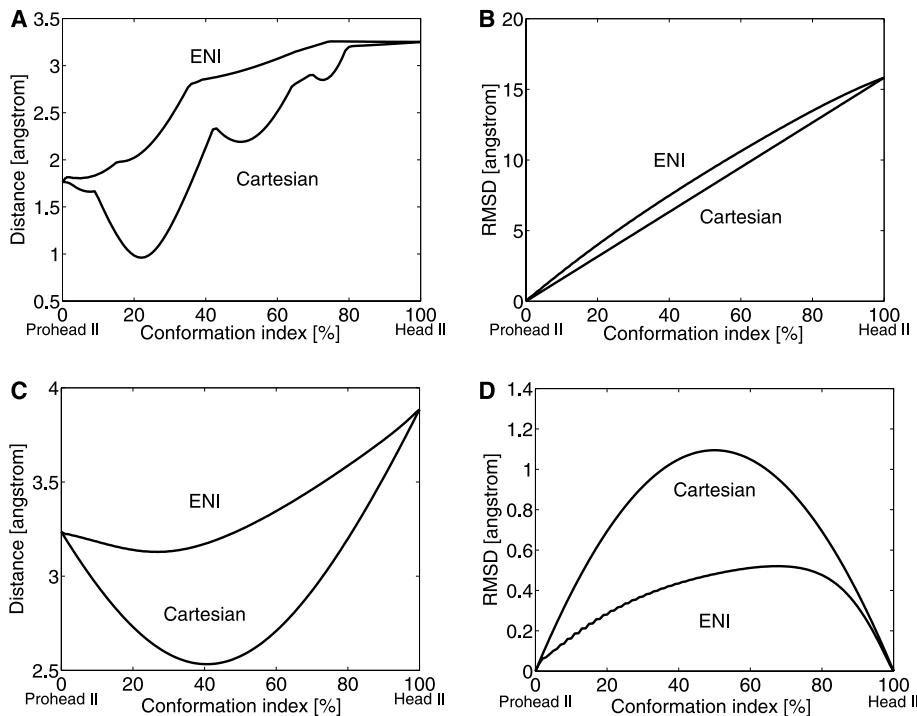


Fig. 7. Comparison between ENI and Cartesian interpolation. (A) the minimum distances between every two  $C_x$  pairs in the capsid shows that ENI with symmetry constraints efficiently generates a putative pathway without steric clash problems. (B) RMSD of intermediate conformations with respect to the initial conformation of Prohead II increases monotonically up to 15 Å in both cases. (C) the virtual bond length between Ser<sup>145</sup> and Asn<sup>146</sup> are displayed. It corresponds to the junction of the N-arm and E-loop. Cartesian interpolation results are not favorable. (D) RMSD of the A and P domains is shown. During the transition, ENI preserves the rigidity of these domains better than Cartesian interpolation.

Fig. 7D. Therefore, we conclude that ENI generates more feasible conformational transition of the HK97 capsid rather than intuitively simple Cartesian interpolation.

### 3.2. Normal mode analysis

As a means of validation for ENI results, normal mode analysis using symmetry constraints is performed.

We take a closer look at the first four lowest modes and characterize their motions. Due to the strategy of using symmetry constraints, the resultant modes for the whole capsid are obviously icosahedrally symmetric. However, it is not necessary that mode shapes of the 7 chains within a single asymmetric unit should be identical with each other because each one has its own set of inter-residue interactions caused by different surroundings (Figs. 3 and 8).

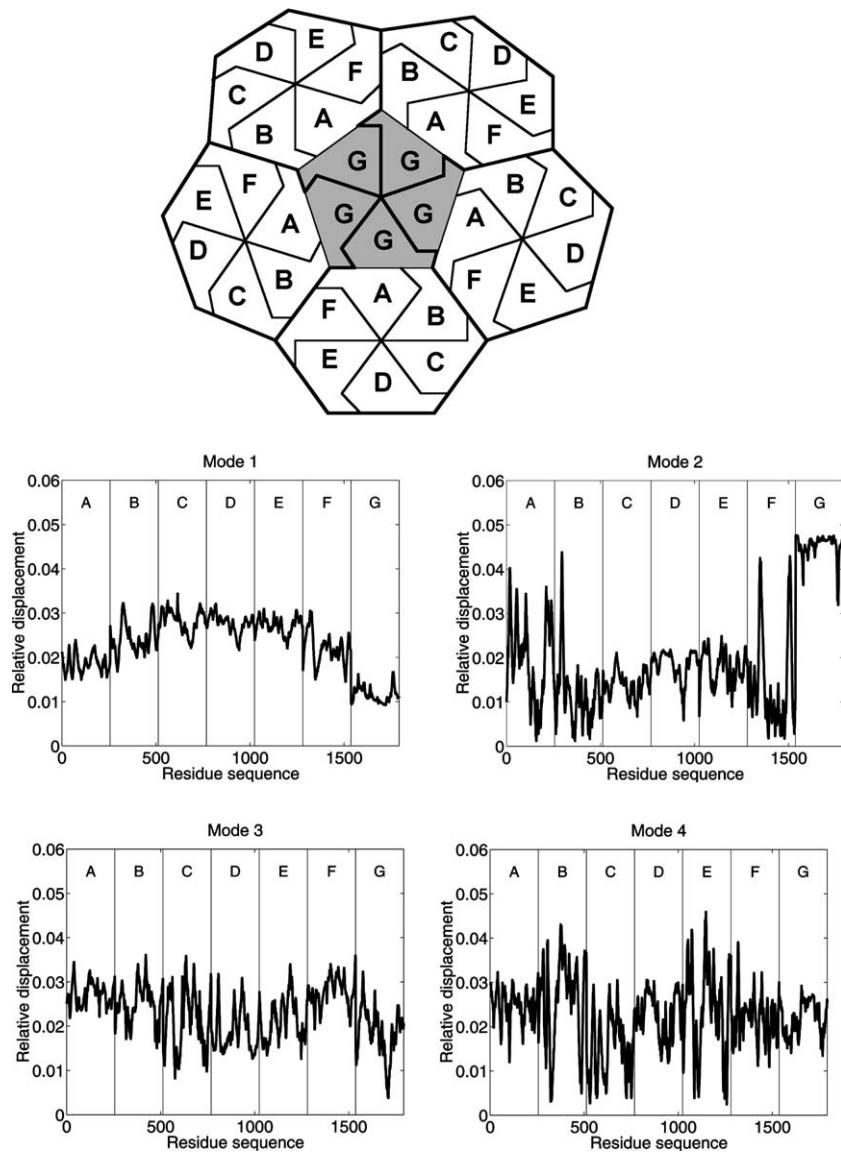


Fig. 8. Symmetry-constrained normal mode analysis. The diagram indicating the chain location around the pentamer is shown at the top. "A–G" represent the different chain names, respectively. The fluctuation scale of residues associated with the first 4 lowest modes is also presented. As discussed in Fig. 3, it is not necessary that mode shapes of these 7 chains of a single asymmetric unit be identical with each other.

The structure can fluctuate back and forth along a normal mode vector so that there exists two possible directions in each mode. We just mention here only the direction that is associated with capsid swelling. Fig. 8 shows the magnitude of residue fluctuations of modes 1–4. On the other hand, Fig. 9 presents cartoons illustrating the first four modes. In this figure, only the direction of each mode that is favorable to swelling is displayed. In the first mode, the residue motions are concentrated on the hexamers, whereas the pentamers have relatively small fluctuations. In the global motion it appears that the icosahedral capsid could expand radially. The expansion rate is proportional to the distance away from vertices. The second mode is opposite to the first mode. The residue motions are highly concentrated on the pentamers denoted by chain "G". The 12 vertices

of the icosahedral capsid could expand radially. In the third mode, the shear motion of hexamer could make the pentamer rotate and lift outward. As reviewed earlier, the major structural changes of the capsid are in the capsid size and the position of subchains. The first three modes represent well the potential (favorable) motions of the Prohead II structure such as capsid expansion and relative position change between subchains. In the fourth mode, the motion appears to be a squeezing outward of the center pentamer by the surrounding hexamers. 3-D movies for these mode shapes of Prohead II are posted on the web (<http://www.custer.me.jhu.edu/proteins/vmode.htm>).

The global slow motions of capsid structures are dominantly ruled by a few of the lowest modes. Statistical mechanics predicts that the contribution to the

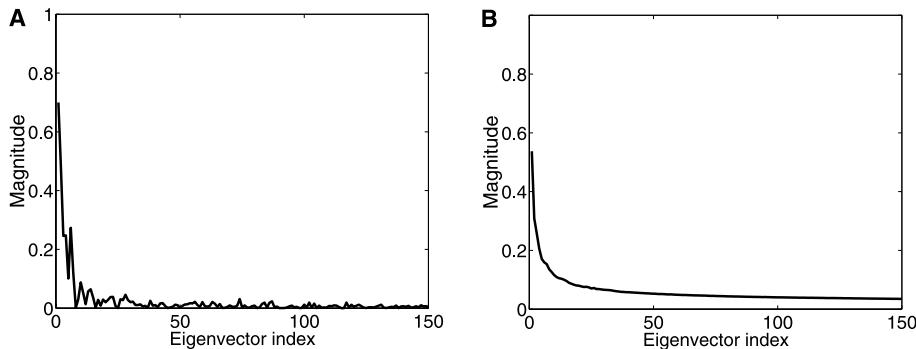


Fig. 10. Decomposition of the normalized displacement from the symmetry-constrained ENI over eigenvectors from the symmetry-constrained coarse-grained NMA. (A) The normalized displacement which corresponds to the intermediate conformation of 1% deviation from Prohead II to Head II is decomposed over the set of orthonormal eigenvectors obtained by NMA of Prohead II. Only the first 150 non-rigid lower modes are displayed. The strong concentration near the lowest modes shows that the initial part of the large motions predicted by the network interpolation model are consistent with small harmonic motions from NMA. (B) The normalized plot of  $1/\omega_i$  represents the contribution of each normal mode to the partition function computed for the elastic network model.

partition function due to motion along a normal mode is inversely proportional to the square root of the corresponding eigenvalue (Chirikjian, 2003). That means the low-frequency mode is naturally favorable. We compare NMA results with those of the ENI in Fig. 10. Let  $\{\mathbf{v}_i\}$  be the normalized eigenvectors of  $K'$  in Eq. (21) such that  $K'\mathbf{v}_i = \lambda_i \mathbf{v}_i$ . Assuming the harmonic motion  $\delta(t) = e^{i\omega_i t} \mathbf{v}_i$ , then  $\lambda_i = \omega_i^2$ . The small displacement generated using the ENI for  $\alpha = 0.01$  from Prohead II is denoted here as  $\delta$ . Fig. 10A plots the function

$$F_1(i) = (\delta \cdot \mathbf{v}_i) / \|\delta\|. \quad (22)$$

This shows the correlation between the behavior of conformational transition and non-rigid modes of Prohead II structure. Fig. 10B displays the function

$$F_2(i) = \omega_i^{-1} \sqrt{\sum_{i=1}^{3n} \omega_i^{-2}}, \quad (23)$$

where  $\omega_i = \sqrt{\lambda_i}$ . This normalized plot represents the contribution of each mode to a complex motion predicted by equilibrium statistical mechanics. As can be seen, small harmonic motions about these equilibria are consistent with the large motions predicted by the ENI.

#### 4. Conclusions

The procapsid (Prohead II) of the icosahedrally symmetric bacteriophage HK97 has been observed to undergo a large conformational change to the mature (“swollen”) state (Head II) in both vivo and in vitro (Duda et al., 1995). To study this maturation process, ENI based on a coarse-grained network model has been used. Icosahedral symmetry of the capsid structure reduces the computational cost because only an asymmetric unit (which is one sixtieth of the whole structure) and its contact information with neighboring units are

required. Using the symmetry-constrained ENI, a series of intermediate coordinates have been calculated. These intermediate frames are converted to PDB format and rendered as 3D pictures with Rasmol software. These static pictures are processed to create animations. Several different measurements such as virtual bond angles, the distance between two cross-linked residues, and the minimum distance between  $C_\alpha$  atoms are performed. The results indicate that the ENI method presented here reliably generates sequences of feasible intermediates without steric clashes. 3D movies showing the HK97 maturation process are posted on the web (<http://custer.me.jhu.edu/proteins/virus.html>). Our results may enable us to infer the folding mechanism of the HK97 capsid during maturation process without any known intermediate structures. In addition, normal mode analysis has been performed efficiently with symmetry constraints. Small harmonic motions captured from NMA are consistent with the initial part of the large motions predicted by ENI. This provides a way to verify our predicted pathway.

#### References

- Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., Bahar, I., 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80, 505–515.
- Bahar, I., Jernigan, R.L., 1998. Vibrational dynamics of transfer RNAs: comparison of the free and synthetase bound forms. *J. Mol. Biol.* 281, 871–884.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. *Nucl. Acids Res.* 28, 235–242.
- Brooks, B., Karplus, M., 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA* 80, 6571–6575.
- Chirikjian, G.S., 2003. A methodology for determining mechanical properties of macromolecules from ensemble motion data. *Trends Anal. Chem.*, to appear.

- Conway, J.F., Duda, R.L., Cheng, N., Hendrix, R.W., Steven, A.C., 1995. Proteolytic 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., Jonson, J.E., Steven, A.C., 2001. Virus maturation involving large subunit rotations and local refolding. *Science* 292, 744–748.
- 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., Hendrix, R.W., 1995. Structural transitions during bacteriophage HK97 head assembly. *J. Mol. Biol.* 247, 618–635.
- Flint, S.J., Enquist, L.W., Krug, R.M., Racaniello, V.R., Skalka, A.M., 2000. Principles of virology: molecular biology, pathogenesis, and control. ASM Press, Washington, DC.
- Kim, M.K., Chirikjian, G.S., Jernigan, R.L., 2002b. Elastic models of conformational transitions in macromolecules. *J. Mol. Graph. Model.* 21, 151–160.
- Kim, M.K., Jernigan, R.L., Chirikjian, G.S., 2002a. Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.* 83, 1620–1630.
- Lata, R., Conway, J.F., Cheng, N., Duda, R.L., Hendrix, R.W., Wikoff, W.R., Jonson, J.E., Tsuruta, H., Steven, A.C., 2000. Maturation dynamics of a viral capsid: visualization of transitional intermediate states. *Cell* 100, 253–263.
- Ming, D., Kong, Y., Wu, Y., Ma, J., 2003. Substructure synthesis method for simulating large molecular complexes. *Proc. Natl. Acad. Sci. USA* 100, 104–109.
- Steven, A.C., Bauer, A.C., Bisher, M.E., Robey, F.A., Black, L.W., 1991. The maturation-dependent conformational change of phage-T4 capsid involves the translocation of specific epitopes between the inner and the outer capsid surfaces. *J. Struct. Biol.* 106, 221–236.
- Steven, A.C., Greenstone, H., Bauer, A.C., Williams, R.W., 1990. The maturation-dependent conformational change of the major capsid protein of bacteriophage T4 involves a substantial change in secondary structure. *Biochemistry* 29, 5556–5561.
- Tama, F., Brooks III, C.L., 2002. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. *J. Mol. Biol.* 318, 733–747.
- Vlijmen, H., Karplus, M., 2001. Normal mode analysis of large systems with icosahedral symmetry: application to (Dialanine)<sub>60</sub> in full and reduced basis set implementations. *J. Chem. Phys.* 115, 691–698.
- Wikoff, W.R., Liljas, L., Duda, R.L., Tsuruta, H., Hendrix, R.W., Johnson, J.E., 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* 289, 2129–2133.
- Zhang, Z., Greene, B., Thuman-Commike, P.A., Jakana, J., Prevelige Jr., P.E., King, J., Chiu, W., 2000. Visualization of the maturation transition in bacteriophage P22 by electron cryomicroscopy. *J. Mol. Biol.* 297, 615–626.