Metadata of the chapter that will be visualized online

Chapter Title	Analysis of Protein Conformational Transitions Using Elastic Network Model	
Copyright Year	2013	
Copyright Holder	Springer Science+Business Media New York	
Corresponding Author	Family Name	**Zheng**
	Particle	
	Given Name	**Wenjun**
	Suffix	
	Division	Department of Physics
	Organization	University at Buffalo
	Address	239 Fronczak Hall, Buffalo, NY, 14260, USA
Author	Family Name	**Tekpinar**
	Particle	
	Given Name	**Mustafa**
	Suffix	
Abstract	In this chapter, we demonstrate the usage of a coarse-grained elastic network model to analyze protein conformational transitions in the NS3 helicase (NS3hel) of Hepatitis C virus (HCV). This analysis allows us to identify and visualize collective domain motions involved in the conformational transitions and predict the order of structural events during the transitions. It is highly efficient and applicable to many multi-domain protein structures which undergo large conformational changes to fulfill their functions. This method is made available through a Web server (http://enm.lobos.nih.gov).	
Key words (separated by “-”)	Conformational transition - Coarse-grained model - Elastic network model - Normal mode analysis - Reaction coordinate - Transition pathway - Helicase	

Chapter 9

Analysis of Protein Conformational Transitions Using Elastic Network Model

Wenjun Zheng and Mustafa Tekpinar

Abstract

In this chapter, we demonstrate the usage of a coarse-grained elastic network model to analyze protein conformational transitions in the NS3 helicase (NS3hel) of Hepatitis C virus (HCV). This analysis allows us to identify and visualize collective domain motions involved in the conformational transitions and predict the order of structural events during the transitions. It is highly efficient and applicable to many multi-domain protein structures which undergo large conformational changes to fulfill their functions. This method is made available through a Web server (http://enm.lobos.nih.gov).

Key words Conformational transition, Coarse-grained model, Elastic network model, Normal mode analysis, Reaction coordinate, Transition pathway, Helicase

1 Introduction

Protein conformational dynamics, which is involved in many protein functions, spans a wide range of temporal scales (from femto-seconds to seconds) and spatial scales (from atomic fluctuations to domain motions). The functions of a large number of protein complexes are thought to invoke conformational transitions between a series of biochemical states, some of which were captured by X-ray crystal structures and electron microscopy. However, it remains very challenging to experimentally probe or computationally model the transient intermediates of these transitions, which determine the kinetic mechanism of these protein complexes. To obtain details of protein dynamics, all-atom molecular dynamics (MD) [1] has been widely employed to simulate protein conformational fluctuations and transitions. Nevertheless, such simulations are limited to nanoseconds to microseconds, which fall short of the typical time scale of protein kinetics (milliseconds to seconds). To overcome the time-scale barrier for MD simulations, a variety of coarse-grained models [2] have been developed. Of particular interest is the elastic network model (ENM)

Dennis R. Livesay (ed.), *Protein Dynamics: Methods and Protocols*, Methods in Molecular Biology, vol. 1084,
DOI 10.1007/978-1-62703-658-0_9,

[3–5], which represents a protein structure as a network of C_α atoms with neighboring ones connected by springs with a uniform force constant [6]. The normal mode analysis (NMA) of ENM often yields a handful of low-frequency modes that capture the large-scale conformational changes observed between two protein crystal structures [7, 8]. ENM has formed the basis of several computational methods for modeling conformational transitions between two given protein conformations [9–13].

In this chapter, we will demonstrate ENM-based NMA and a transition pathway modeling method named interpolated-ENM (iENM) [13]. The iENM constructs a pathway by solving the saddle points of a double-well potential built from two ENM potentials based at the beginning and end conformation of a transition [13]. The predicted pathway allows us to deduce the dynamic order of structural events involving various protein parts [13–17]. We will use NS3hel of HCV as an example [18, 19] (*see* **Note 1**).

2 Methods

2.1 Normal Mode Analysis of Elastic Network Model

Given the C_α atomic coordinates of a protein crystal structure from Protein Data Bank (http://www.rcsb.org), we build an elastic network model by connecting all pairs of C_α atoms that are within a cutoff distance R_c (chosen to be 10 Å by default) by using harmonic spring. The total ENM energy is

$$E_{\mathrm{ENM}} = \frac{1}{2} \sum_{d_{ij}^0 < R_c} C(d_{ij} - d_{ij}^0)^2, \tag{1}$$

where C is the spring force constant which is set to 1 by default, although it can be determined by fitting the crystallographic B factors if needed [20], d_{ij} is the distance between the C_α atoms i and j, and d_{ij}^0 is the value of d_{ij} as given by the crystal structure.

We expand the ENM energy to second order:

$$E_{\mathrm{ENM}} \approx \frac{1}{2} \delta X^T H \delta X = \frac{1}{2} C \sum_{d_{ij}^0 < R_c} \delta X^T H_{ij} \delta X, \tag{2}$$

where $\delta X = X - X_0$, X is a $3N$-dimensional vector representing the C_α atomic coordinates, X_0 gives the equilibrium C_α coordinates in the crystal structure, $H = C \sum_{d_{ij}^0 < R_c} H_{ij}$ is the $3N \times 3N$ Hessian matrix (second derivatives of E_{ENM}), where $H_{ij} = \frac{1}{2} \nabla^2 \left[(d_{ij} - d_{ij}^0)^2 \right]$.

For the Hessian matrix H, we can perform the normal mode analysis (NMA) to obtain $3N$ normal modes. Each mode m has an eigenvalue λ_m and a $3N$-dimensional eigenvector V_m which satisfy

$HV_m = \lambda_m V_m$. The normal modes can be solved using the dsyevr subroutine of a linear algebra package named LAPACK (http://www.netlib.org/lapack/). The lowest six zero modes, corresponding to three translations and three rotations, are removed from the spectrum (mode numbering starts from #1 for the lowest non-zero mode).

To validate NMA, we need to use two different crystal structures of a protein. We first perform NMA using the first structure (see above). Then, we superimpose the second structure on top of the first structure using the PROFIT program (*see* http://www.bioinf.org.uk/software/profit/). Finally, we compare each mode (mode m) with the observed structural changes between the two superimposed structures (represented by a $3N$-dimensional vector δX_{obs}) by calculating the following overlap:

$$I_m = \frac{\left|\delta X_{\text{obs}}^T V_m\right|}{|\delta X_{\text{obs}}| \cdot |V_m|}, \tag{3}$$

where $\delta X_{\text{obs}}^T V_m$ is the dot product between vectors δX_{obs} and V_m, $|\delta X_{\text{obs}}|$ and $|V_m|$ represent their magnitude. I_m ranges from 0 to 1, and higher I_m indicates greater involvement of mode m in the observed structural changes δX_{obs}. In addition, the following cumulative overlap is calculated to assess how well the lowest ten modes describe δX_{obs}:

$$C_{10} = \sqrt{\sum_{1 \le m \le 10} I_m^2}. \tag{4}$$

Because $\sum_{1 \le m \le 3N-6} I_m^2 = 1$, C_{10}^2 gives the percentage of the observed structural changes captured by the lowest ten modes.

For an example of the normal mode analysis, *see* **Note 2**.

2.2 Coarse-Grained Modeling of Protein-DNA System

To study how NS3hel translocates along a single-stranded DNA (ssDNA), we need to refine the modeling of protein–DNA interactions and intra-ssDNA interactions. To this end, we modify the ENM as follows:

1. For residue–residue interactions within NS3hel, we use the C_α-based ENM with $R_c = 10$ Å (*see* Eq. 1).
2. For interactions within ssDNA, we use a modified ENM which represents each nucleic acid by a bead located at the C4′ atom, and adds springs between first, second, and third nearest-neighbor (NN) beads with the same force constant k_{DNA}:

$$E_{\text{DNA}} = \frac{1}{2} \sum_{1 \le |j-i| \le 3} k_{\text{DNA}} (d_{ij} - d_{ij,0})^2, \tag{5}$$

where d_{ij} is the distance between the C4′ atom i and j, d_{ij}^0 is the value of d_{ij} as given by an NS3hel-ssDNA structure.

3. To accurately represent protein–DNA interactions, we use a structure-based Leonard-Jones 6–12 potential to allow protein-DNA contacts to form/break readily during a transition:

$$E_{\text{prot-DNA}} = \frac{1}{2} \sum_{\substack{li \in \text{prot} \\ j \in \text{DNA}}} k_{\text{prot-DNA}} \frac{d_{i,\min}^2}{36} \left(1 - \frac{d_{i,\min}^6}{d_{ij}^6}\right)^2, \qquad (6)$$

where the summation is over residues which form heavy-atom contacts (within 4 Å) with DNA backbone in a NS3hel-ssDNA structure, and $d_{i,\min}$ is the minimal C_α–C4′ distance for residue i, and $k_{\text{prot-DNA}}$ is the force constant. We only consider contacts between protein and DNA backbone in our modeling because functional and structural data suggested that these contacts are sufficient for ensuring that NS3hel maintains a grip on the ssDNA track and undergoes continuous translocation [21]. For NMA, we replace Eq. 9.6 with its harmonic counterpart $E'_{\text{prot-DNA}} = \frac{1}{2} \sum_{\substack{li \in \text{prot} \\ j \in \text{DNA}}} k_{\text{prot-DNA}} (d_{ij} - d_{ij,0})^2$.

For NS3hel, we choose $k_{\text{DNA}} = 1$ and $k_{\text{prot-DNA}} = 1.3$ based on the fitting of crystallographic B factors. One should check a range of parameter values to make sure that the modeling results are not sensitive to the particular choice of these parameters.

2.3 Interpolated Elastic Network Model (iENM)

We consider an *arbitrary* double-well potential function $F(E_1, E_2)$ with two minima at the beginning and end conformation of a transition. It satisfies: $F(E_1, E_2) \approx E_1$ if $E_1 \ll E_2$, and $F(E_1, E_2) \approx E_2$ if $E_2 \ll E_1$, where E_1 and E_2 are two single-well potentials. The saddle points (SP) of $F(E_1, E_2)$ are solved as follows

$$0 = \nabla F(E_1, E_2) = \frac{\partial F}{\partial E_1} \nabla E_1 + \frac{\partial F}{\partial E_2} \nabla E_2, \qquad (7)$$

which is equivalent to solving the following equation (after setting $\lambda = \frac{\partial F}{\partial E_1} \Big/ \left(\frac{\partial F}{\partial E_1} + \frac{\partial F}{\partial E_2}\right)$)

$$0 = \lambda \nabla E_1 + (1 - \lambda) \nabla E_2, \qquad (8)$$

where λ is a parameter of interpolation that varies from 1 to 0 (assuming $\frac{\partial F}{\partial E_1} \geq 0$ and $\frac{\partial F}{\partial E_2} \geq 0$). Therefore, the problem of solving SP for the double-well potential function $F(E_1, E_2)$ is converted to the problem of finding the minima of a linearly interpolated potential function $\lambda E_1 + (1 - \lambda) E_2$. Equation 9.8 gives a set of minimal-energy crossing points between E_1 and E_2 where $E_1 = E_2$ is at minimum.

Based on the above general formulation, we have proposed an iENM protocol [13] using a double-well potential $F(E_{\text{ENM1}} +$

$E_{\text{col}}, E_{\text{ENM2}} + E_{\text{col}}$), where E_{ENM1} and E_{ENM2} are two ENM potential functions (*see* Eq. 1) based at the beginning and end conformation of a transition, and E_{col} is a steric collision energy defined as follows:

$$E_{\text{col}} = \frac{1}{2}\sum_{i=3}^{N}\sum_{j=1}^{i-2} C_{\text{col}}\theta(R_{\text{col}} - d_{ij})(d_{ij} - R_{\text{col}})^2, \qquad (9)$$

where $R_{\text{col}} = 4$ Å, $C_{\text{col}} = 10$, and chemically bonded residue pairs ($j = i \pm 1$) are excluded. The addition of E_{col} penalizes steric collisions between residues whose C_α atoms are within a distance of R_{col}.

After adding the collision energy, the SPs are solved by setting $\nabla F(E_{\text{ENM1}} + E_{\text{col}}, E_{\text{ENM2}} + E_{\text{col}}) = 0$ which is equivalent to solving the following SP equation (the SP is represented by X_{SP}):

$$\lambda\nabla E_{\text{ENM1}}(X_{\text{SP}}) + (1-\lambda)\nabla E_{\text{ENM2}}(X_{\text{SP}}) + \nabla E_{\text{col}}(X_{\text{SP}}) = 0. \quad (10)$$

As λ varies from 1 to 0, X_{SP} traces a pathway that connects the beginning and end conformation of a transition. Because this pathway passes all possible SPs, it gives a *universal* minimal-energy path regardless of the detailed form of $F(E_1, E_2)$. iENM outputs the above pathway as the predicted pathway for the given transition.

We solve Eq. 10 by using the following iterative procedure to find the minima of the linearly interpolated potential function $\lambda E_{\text{ENM1}} + (1-\lambda)E_{\text{ENM2}} + E_{\text{col}}$ with the Newton–Raphson method:

1. Initialization: set $n = 0$, $X_{\text{SP},0} = X_1$, which is the C_α coordinates of the beginning conformation.
2. Given $X_{\text{SP},n}$, calculate $\lambda = \lambda_n = -\frac{[\nabla E_{\text{ENM1}} - \nabla E_{\text{ENM2}}]\cdot[\nabla E_{\text{ENM2}} + \nabla E_{\text{col}}]}{|\nabla E_{\text{ENM1}} - \nabla E_{\text{ENM2}}|^2}$ to minimize $|\lambda\nabla E_{\text{ENM1}} + (1-\lambda)\nabla E_{\text{ENM2}} + \nabla E_{\text{col}}|$.
3. Calculate $R_n = \lambda_n\nabla E_{\text{ENM1}} + (1-\lambda_n)\nabla E_{\text{ENM2}} + \nabla E_{\text{col}}$.
4. If $|R_n| < 0.00001$, go to **step 7**.
5. Displace $X_{\text{SP},n}$ by

$$\delta X_{\text{SP}} = -[\lambda_n H_1 + (1-\lambda_n)H_2 + H_{\text{col}} + \varepsilon I]^{-1} R_n, \qquad (11)$$

where H_1, H_2 and H_{col} are the Hessian matrices calculated for E_{ENM1}, E_{ENM2} and E_{col}, I is identity matrix, ε is a small positive number to render the sum of matrices invertible.

6. Go to **step 3**.
7. Calculate $X_{\text{SP},n+1} = X_{\text{SP},n} + \delta X_{\text{SP}}$ and

$$\delta X_{\text{SP}} \sim -\delta\lambda[\lambda_n H_1 + (1-\lambda_n)H_2 + H_{\text{col}} + \varepsilon I]^{-1}[\nabla E_{\text{ENM1}} - \nabla E_{\text{ENM2}}], \qquad (12)$$

where H_1, H_2 and H_{col} are the Hessian matrices calculated for E_{ENM1}, E_{ENM2} and E_{col}, I is identity matrix, ε is a small positive

number to render the sum of matrices invertible, and $\delta\lambda$ is chosen so that the magnitude of δX_{SP} is small (i.e., $|\delta X_{SP}|/\sqrt{N} < 0.1$ Å).

8. Stop if $X_{SP,n+1}$ has reached X_2 which is the C_α coordinates of the end conformation, otherwise set $n \leftarrow n + 1$, then go to **step 2**.

The linear equations in Eqs. 11 and 12 are solved using a highly efficient sparse linear equation solver CHOLMOD (http://www.cise.ufl.edu/research/sparse/cholmod/) [22]. In Eq. 12 we compute an incremental structural displacement δX_{SP} based on the force-induced linear responses—δX_{SP} is calculated as a weighted sum of all normal modes of the Hessian matrix $\lambda_n H_1 + (1 - \lambda_n)H_2 + H_{col}$. Because the weight of each mode is inversely proportional to its eigenvalue, the collective motions described by the lowest modes are favorably sampled along the transition pathway.

For an example of iENM application, *see* **Note 3**.

2.4 Assessment of Motional Order Using Reaction Coordinates

The predicted transition pathway allows us to determine the motional order of different parts/domains of a protein. For this purpose, the following reaction coordinate (RC) is defined for an intermediate conformation of a given part S [13]:

$$RC_S = (\delta X_S \bullet \delta X_{S,obs})/|\delta X_{S,obs}|^2, \qquad (13)$$

where δX_S is the displacement vector of part S from the beginning conformation of a transition to a given intermediate conformation, and $\delta X_{S,obs}$ is the observed displacement of part S from the *beginning* conformation to the end conformation of a transition. RC_S measures the motional progress of part S in the direction of a transition. $RC_S = 0$ at the beginning of a transition, and $RC_S = 1$ at the end of a transition. For two different parts (named S_1 and S_2) in an intermediate conformation, if $RC_{S_1} > RC_{S_2}$, then S_1's motion precedes S_2's motion.

For an example of RC calculations, *see* **Note 3**.

3 Conclusion

We have demonstrated the use of a coarse-grained ENM for analyzing conformational transitions in protein or protein-DNA complex. The predicted order of structural events has been validated using structural data. This method is highly efficient—it takes only 3 min to run the entire ATP cycle of NS3hel using a dual-core workstation. This method will be useful for future simulations of a variety of molecular motors including many monomeric and ring-shaped helicases. Both ENM-based NMA and iENM are available via a Web server at http://enm.lobos.nih.gov.

4 Notes

1. *Introduction to NS3hel*

We will illustrate the usage of ENM-based NMA and iENM using NS3hel as an example. To unwind double-stranded DNA/RNA, NS3hel assembles on a 3′ end of ssDNA/RNA tail, and actively translocates along ssDNA/RNA in the 3′–5′ direction [23]. Several crystal structures of NS3hel bound with ssDNA and various ATP analogs have been solved [24–29], which correspond to three biochemical states of its work cycle (apo, ATP, and ADP-Pi, where Pi represents inorganic phosphate). The structure of NS3hel consists of three domains (*see* Fig. 1). The ATP binds at the cleft between domains 1 and 2, while the ssDNA binds in a groove between domains 1, 2 and domain 3 (*see* Fig. 1). From the structural data, an inchworm model has emerged for the translocation of NS3hel along ssDNA (complemented by a ratchet action) [24, 28–31] (*see* Fig. 1): first, ATP binding induces a closure

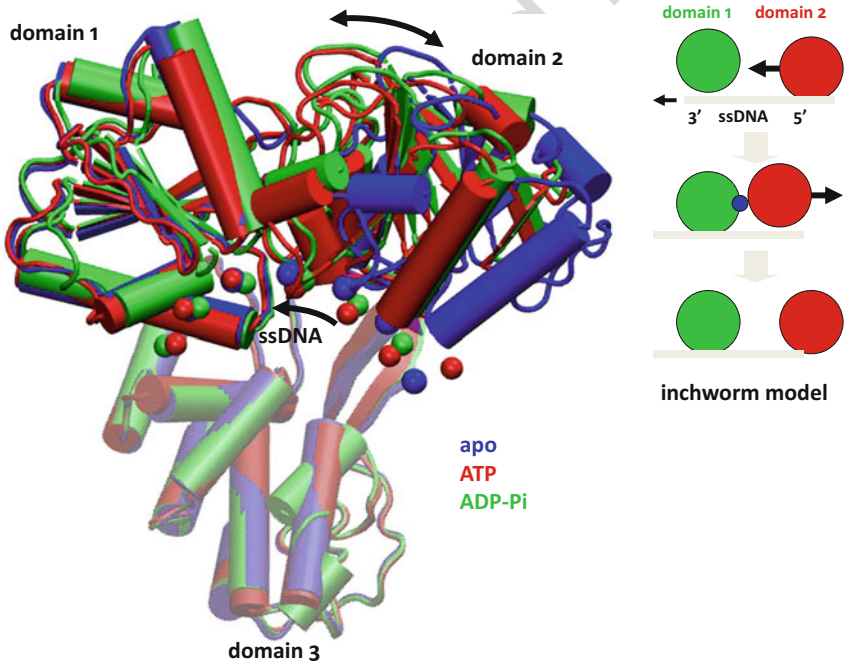

Fig. 1 The structures of NS3hel in three biochemical states (apo, ATP, and ADP-Pi, colored in *blue*, *red*, and *green*, respectively). NS3hel is shown in *cartoon* representation, and ssDNA is shown as a *chain of beads* located at C4′ atoms. The three domains of NS3hel and ssDNA are labeled. The opening/closing motions of domain 2 and the 5′–3′ sliding of ssDNA are marked by *arrows*. The three structures are aligned along domain 3 which is shown as *transparent*. *Inset*: a schematic cartoon illustrates the inchworm model (domain 1, domain 2, ATP, and ssDNA are colored *green*, *red*, *blue*, and *gray*, respectively; the opening/closing motions of domain 2 and the sliding of ssDNA are marked by *arrows*; domain 3 is not shown)

t.1 **Table 1**
Comparison between the lowest ten normal modes and the crystallographically observed conformational changes in HCV NS3hel

t.2 PDB_chain id of two NS3hel structures	RMSD (Å)	Mode#	Overlap
t.3 3kqk_AD → 3kqu_B	3.26	#3	0.49
t.4		#4	0.57
t.5		#1–10	0.92
t.6 3kqu_BM → 3kql_A	1.03	#1	0.61
t.7		#1–10	0.72
t.8 3kql_AE → 3kqk_A	3.37	#5	0.45
t.9		#6	0.50
t.10		#1–10	0.82

motion of domain 2 toward domain 1, with domain 1 releasing its grip on ssDNA and sliding along it, while domain 2 maintains its grip on ssDNA; second, following ATP hydrolysis and the release of ADP and Pi, domain 2 opens again as it releases its grip on ssDNA and slides along it, while domain 1 maintains its grip on ssDNA. The net effect is the translocation of NS3hel along ssDNA by one base in the 3′–5′ direction, consuming one ATP per step. The details of the conformational transitions between apo, ATP, and ADP-Pi state remain largely unknown.

2. *Normal mode analysis of ENM*

To validate the use of ENM for NS3hel, we compare the domain motions predicted by NMA of ENM with the observed conformational changes between NS3hel structures in different states (for results, *see* Table 1). Here we will focus on the conformational transition from apo to ATP state.

We have performed NMA for an ENM constructed from an NS3hel-ssDNA structure in apo state (PDB id: 3kqk), and then calculated the overlaps between each mode and the observed conformational changes from 3kqk to an NS3hel-ssDNA-ADP · BeF_3 structure in ATP state [28] (PDB id: 3kqu). Encouragingly, 84 % of the observed conformational changes are captured by the lowest ten modes (with cumulative overlap $C_{10} = 0.92$), among which mode #3 and #4 contribute most (with overlap $I_3 = 0.49$ and $I_4 = 0.57$, respectively). To visualize the domain motions predicted by these two modes, we have deformed the 3kqk structure along the directions given by the eigenvectors of these two modes, and then compared the deformed structures with 3kqk using the VMD program (http://www.ks.uiuc.edu/Research/vmd/).

Mode #3 describes coupled rotations of domains 1 and 2 relative to domain 3, which result in the opening of domains 1–3 interface, the closing of domains 1–2 interface, and the sliding of ssDNA toward its 3′ end (*see* Fig. 2a).

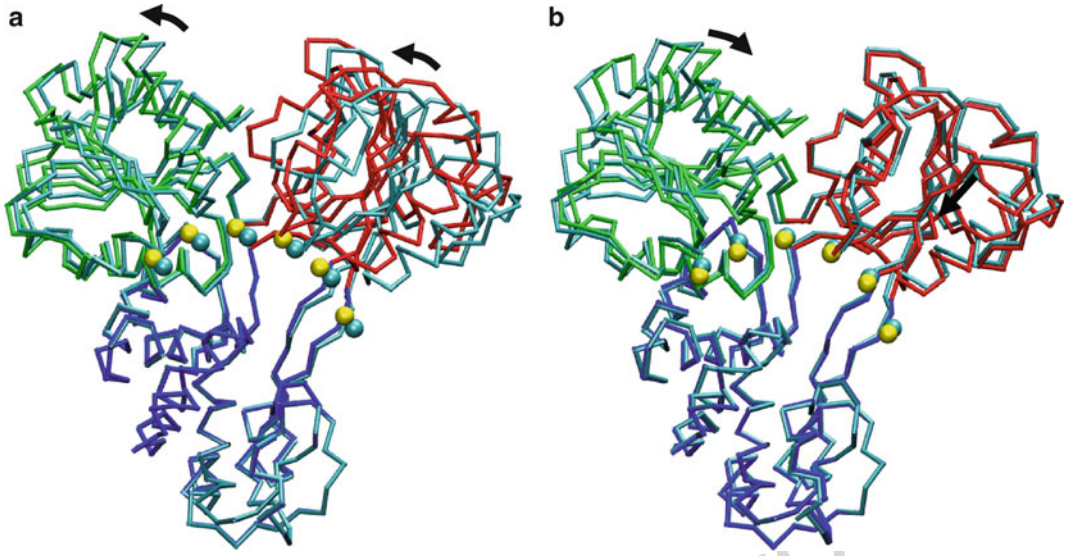

Fig. 2 Conformational changes in NS3hel as predicted by the following ENM-based normal modes: (**a**) mode #3, (**b**) mode #4, which are solved from the apo-state NS3hel-ssDNA structure (PDB id: 3kqk). The original NS3hel-ssDNA structure is colored *cyan*. For the deformed NS3hel-ssDNA structure after the conformational changes, domain 1, 2, and 3 are colored *green*, *red*, and *blue*, respectively, and ssDNA is shown as a *chain of yellow beads* located at C4′ atoms. The two structures of NS3hel are superimposed along domain 3. The domain rotations are shown by *arrows*

Mode #4 describes simultaneous rotations of domains 1 and 2 toward domain 3, which cause the closing of domains 1–2 interface and a shift of ssDNA toward domain 3 (*see* Fig. 2b). Although the lowest ten modes accurately capture the observed conformational changes in NS3hel, they do not correctly predict the translocation of NS3hel along ssDNA which requires domain 1 to slide along ssDNA while domain 2 holds ssDNA. Instead, in both modes #3 and #4, domains 1 and 2 maintain their grip on ssDNA with no sliding between domain 1 and ssDNA (*see* Fig. 2). This is not surprising because domain 1 and ssDNA are linked by elastic springs in ENM which disfavor sliding between them. Therefore, to accurately describe the conformational dynamics underlying NS3hel translocation, one has to modify ENM and account for the anharmonicity of protein–DNA interactions (*see* Subheading 2).

3. *Transition pathway modeling by iENM*

Next, we use iENM to simulate the conformational transitions in NS3hel between three biochemical states (apo, ATP, and ADP-Pi), which are captured by three crystal structures of NS3hel [28] (PDB ids: 3kqk, 3kqu, 3kql) (*see* Fig. 1). We validate the use of iENM for NS3hel by checking if it correctly predicts the order of inter-domain motions observed among crystal structures of NS3hel in different states. To this end,

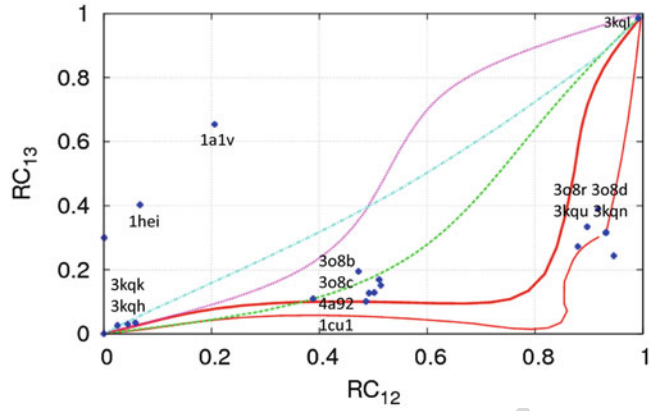

Fig. 3 Comparison between calculated transition pathways and crystal structures of NS3hel using two reaction coordinates (RC_{12} and RC_{13}). The calculated pathways are obtained using the following methods: iENM (*solid lines*), mixed-ENM server (*dashed line*), Morph server (*dot dashed line*), MinActionPath server (*dotted line*). The crystal structures are shown as *points* (PDB&chain ids: 1a1vA, 1cu1A, 1cu1B, 1heiA, 1heiB, 3o8bA, 3o8bB, 3o8rA, 3o8rB, 3o8dA, 3o8dB, 3o8cA, 3o8cB, 3kquA, 3kquB, 3kquC, 3kquD, 3kquE, 3kquF, 3kqlA, 3kqlB, 3kqkA, 3kqkB, 3kqnA, 3kqhA, 3kqhB, 4a92A, 4a92B). The following three pathways are calculated by iENM: 3kqk → 3kql (*thick solid line*), 3kqk → 3kqu (*thin solid line*), 3kqu → 3kql (*thin solid line*)

we have generated pathways for three transitions (apo → ATP, ATP → ADP-Pi, apo → ADP-Pi) using iENM, and then compared the predicted pathways with the crystal structures [24–26, 28, 29] using two reaction coordinates (RC) (*see* Fig. 3): RC_{12} quantifies the progress of motion between domains 1 and 2, and RC_{13} quantifies the progress of motion between domains 1 and 3. Both RCs vary from 0 to 1, where 0 corresponds to the apo state, and 1 corresponds to the ADP-Pi state. The iENM pathway for apo → ADP-Pi transition predicts that the increase of RC_{12} precedes RC_{13} during the transition (*see* Fig. 3), which implies that the domains 1–2 motion precedes the domains 1–3 motion. This order is functionally meaningful, because the domains 1–2 motion occurs upon ATP binding while the domains 1–3 motion occurs during the subsequent transition (ATP hydrolysis) (*see* Fig. 1). This prediction agrees well with the RCs of most crystal structures, which form two intermediate clusters located near the predicted pathway (*see* Fig. 3)—the first cluster includes several apo structures of full-length NS3, and the second cluster includes several crystal structures of NS3hel and full-length NS3 bound with ADP-BeF_3 (corresponding to ATP state).

Two "outlier" structures (*see* Fig. 3) may correspond to off-path intermediates trapped by crystallization conditions. It is remarkable that the iENM pathway for apo → ADP-Pi transition visits the ATP-state structures as intermediates even though these structures are not used in the modeling. For comparison, we have also analyzed the pathways for apo → ADP-Pi transition predicted by alternative methods including Yale Morph server [32], mixed-ENM server [33] and MinActionPath server [12], which do not seem to agree with the crystal structures (*see* Fig. 3).

After validating iENM for exploring conformational transitions in NS3hel, we have used it to simulate the translocation of NS3hel along ssDNA as it undergoes the following three transitions: apo → ATP → ADP-Pi → apo, which comprise the ATP cycle.

For the apo → ATP transition, as predicted by the iENM pathway, domain 2 closes toward domain 1 while it holds the ssDNA (*see* Fig. 4a). Consequently, the ssDNA moves toward its 3′ end by ~4.8 Å as it slides between domains 1 and 3 (*see* Fig. 4a). Therefore, our iENM pathway has reproduced key motions upon ATP binding as postulated by the inchworm model—a closure motion of domain 2 toward domain 1, with domain 1 releasing its grip on ssDNA and sliding along it, while domain 2 maintains its grip on ssDNA (*see* Fig. 1).

For the ATP → ADP-Pi transition, as predicted by the iENM pathway, domains 1 and 2 undergo a coupled rotation, resulting in the movement of ssDNA toward its 5′ end by ~0.8 Å relative to domain 3 (*see* Fig. 4b). Both domains 1 and 2 maintain their grip on ssDNA during this transition, so there is no sliding of ssDNA relative to domains 1 and 2 (*see* Fig. 4b).

For the ADP-Pi → apo transition, as predicted by the iENM pathway, domain 2 opens early to release its grip on ssDNA, which is followed by a small sliding (~1.7 Å) of ssDNA toward its 3′ end and a slight opening of domain 1 (*see* Fig. 4c). Our finding has largely reproduced key motions during ADP-Pi release as postulated by the inchworm model—opening of domain 2 after it releases its grip on ssDNA and slides along it, while domain 1 maintains its grip on ssDNA (*see* Fig. 1).

In sum, our iENM calculations for the above three transitions predict a net translocation of NS3hel along ssDNA in the 3′–5′ direction by ~5.7 Å, which corresponds approximately to 1-base step size.

Acknowledgment

We thank funding support from American Heart Association (grant #0835292N) and National Science Foundation (grant #0952736).

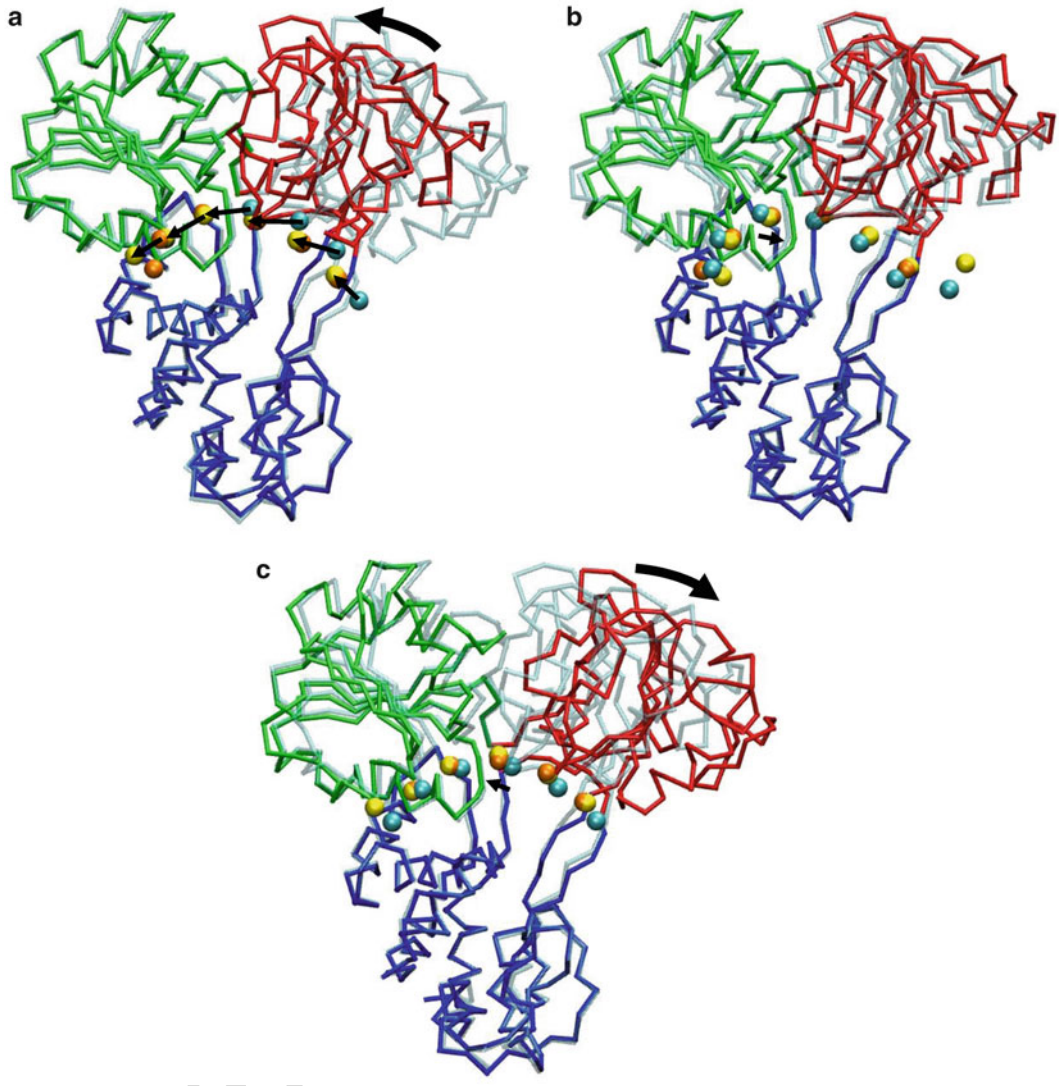

Fig. 4 Snapshots at the beginning and end of iENM pathways for the following transitions in NS3hel: (**a**). apo → ATP transition; (**b**). ATP → ADP-Pi transition; (**c**). ADP-Pi → apo transition. In the end conformation of NS3hel, domain 1, 2, and 3 are colored *green*, *red*, and *blue*, respectively; the beginning conformation of NS3hel is shown as transparent; ssDNA is shown as a chain of beads located at C4′ atoms (for ssDNA, the beginning conformation, end conformation, and target conformation are colored *cyan*, *yellow*, and *orange*, respectively); the movements of domain 2 and ssDNA are marked by *arrows*

References

1. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9(9):646–652
2. Tozzini V (2005) Coarse-grained models for proteins. Curr Opin Struct Biol 15(2): 144–150
3. Hinsen K (1998) Analysis of domain motions by approximate normal mode calculations. Proteins 33(3):417–429
4. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2001) Anisotropy of

fluctuation dynamics of proteins with an elastic network model. Biophys J 80(1):505–515

5. Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. Protein Eng 14(1):1–6
6. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 77(9): 1905–1908
7. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu HY, Gerstein M (2002) Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. Proteins 48 (4):682–695
8. Yang L, Song G, Jernigan RL (2007) How well can we understand large-scale protein motions using normal modes of elastic network models? Biophys J 93(3):920–929
9. Miyashita O, Onuchic JN, Wolynes PG (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci U S A 100 (22):12570–12575
10. Maragakis P, Karplus M (2005) Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. J Mol Biol 352(4):807–822
11. Zheng WJ, Brooks BR, Hummer G (2007) Protein conformational transitions explored by mixed elastic network models. Proteins 69 (1):43–57
12. Franklin J, Koehl P, Doniach S, Delarue M (2007) MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape. Nucleic Acids Res 35(Web Server issue):W477–W482
13. Tekpinar M, Zheng W (2010) Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. Proteins 78(11): 2469–2481
14. Zheng W (2010) Multiscale modeling of structural dynamics underlying force generation and product release in actomyosin complex. Proteins 78(3):638–660
15. Zheng W (2011) Coarse-grained modeling of conformational transitions underlying the processive stepping of myosin V dimer along filamentous actin. Proteins 79(7):2291–2305
16. Zheng W (2012) Coarse-grained modeling of the structural states and transition underlying the powerstroke of dynein motor domain. J Chem Phys 136(15):155103
17. Zheng W, Auerbach A (2011) Decrypting the sequence of structural events during the gating transition of pentameric ligand-gated ion channels based on an interpolated elastic network model. PLoS Comput Biol 7(1):e1001046
18. Zheng W, Liao JC, Brooks BR, Doniach S (2007) Toward the mechanism of dynamical couplings and translocation in hepatitis C virus NS3 helicase using elastic network model. Proteins 67(4):886–896
19. Zheng W (2010) Computer modeling of helicases using elastic network model. Methods Mol Biol 587:235–243
20. Zheng W (2008) A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations. Biophys J 94(10):3853–3857
21. Pyle AM (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. Annu Rev Biophys 37:317–336
22. Chen YQ, Davis TA, Hager WW, Rajamanickam S (2008) Algorithm 887: CHOLMOD, supernodal sparse cholesky factorization and update/downdate. ACM Trans Math Software 35(3):1–14
23. Frick DN (2007) The hepatitis C virus NS3 protein: a model RNA helicase and potential drug target. Curr Issues Mol Biol 9(1):1–20
24. Kim JL, Morgenstern KA, Griffith JP, Dwyer MD, Thomson JA, Murcko MA, Lin C, Caron PR (1998) Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure 6(1):89–100
25. Yao N, Reichert P, Taremi SS, Prosise WW, Weber PC (1999) Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. Structure 7(11):1353–1363
26. Yao N, Hesson T, Cable M, Hong Z, Kwong AD, Le HV, Weber PC (1997) Structure of the hepatitis C virus RNA helicase domain. Nat Struct Biol 4(6):463–467
27. Cho HS, Ha NC, Kang LW, Chung KM, Back SH, Jang SK, Oh BH (1998) Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. J Biol Chem 273 (24):15045–15052
28. Gu M, Rice CM (2010) Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. Proc Natl Acad Sci U S A 107(2):521–528
29. Appleby TC, Anderson R, Fedorova O, Pyle AM, Wang R, Liu X, Brendza KM, Somoza JR (2011) Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV. J Mol Biol 405(5):1139–1153

30. Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell 97(1):75–84

31. Lee JY, Yang W (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell 127(7): 1349–1360

32. Krebs WG, Gerstein M (2000) The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. Nucleic Acids Res 28 (8):1665–1675

33. Zheng W, Brooks BR, Hummer G (2007) Protein conformational transitions explored by mixed elastic network models. Proteins 69 (1):43–57