Chapter 7

Elastic Network Models For Biomolecular Dynamics: Theory and Application to Membrane Proteins and Viruses

Timothy R. Lezon, Indira H. Shrivastava, Zheng Yang and Ivet Bahar*

Department of Computational Biology, School of Medicine, University of Pittsburgh, Suite 3064, Biomedical Science Tower 3, 3051 Fifth Ave., Pittsburgh, PA 15213

7.1. Introduction

Elastic network models (ENMs) have over the last decade enjoyed considerable success in the study of macromolecular dynamics. These models have been used to predict the global dynamics of a variety of proteins and protein complexes, ranging in size from single enzymes to macromolecular machines (Keskin *et al.* 2002), ribosomes (Tama *et al.* 2003, Wang *et al.* 2004) and viral capsids (Tama & Brooks, 2002, Tama & Brooks, 2005, Rader *et al*. 2005). They have provided insights into a wide range of protein behaviors, such as mechanisms of allosteric regulation (Ming & Wall, 2005, Bahar *et al*. 2007, Chennubhotla *et al*. 2008), protein-protein binding (Tobi and Bahar 2005), anisotropic response to uniaxial tension and unfolding (Eyal and Bahar, 2008, Sulkowska *et al*. 2008), co-localization of catalytic sites and key mechanical sites (e.g., hinges) (Yang & Bahar 2005), interactions at the binding sites (Ming & Wall, 2006), and energetics (Miller *et al.* 2008), to name a few. ENMs allow the global motions of a molecule to be quickly calculated, making them an ideal complement to conventional molecular dynamics (MD) simulations. Increasingly, variants of ENMs are being applied to non-equilibrium situations, such as the prediction of transition pathways between functional states separated by low energy barriers (Zheng *et al.* 2007) or driving MD simulations (Isin *et al.* 2008).

At its core, the ENM provides a simplified representation of the potential energy function of a system, in this case a macromolecule or macromolecular

*Corresponding Author: *bahar@ccbb.pitt.edu http://www.ccbb.pitt.edu*

assembly, near equilibrium. The nodes of the network are the building blocks, such as atoms, nucleotides or amino acids, from which the system is composed (Fig. 7.1). Each node is typically represented as a point particle in three dimensions (3D), and the edges of the network, or the springs joining the nodes, represent harmonic restraints on displacements from the equilibrium structure. Thus, the ENM provides an intuitive and quantitative description of behavior near equilibrium: The starting conformation resides at the bottom of a harmonic well, and any deviations from equilibrium will increase the energy and result in a linear net force directed toward restoring the system to its lowest energy state.

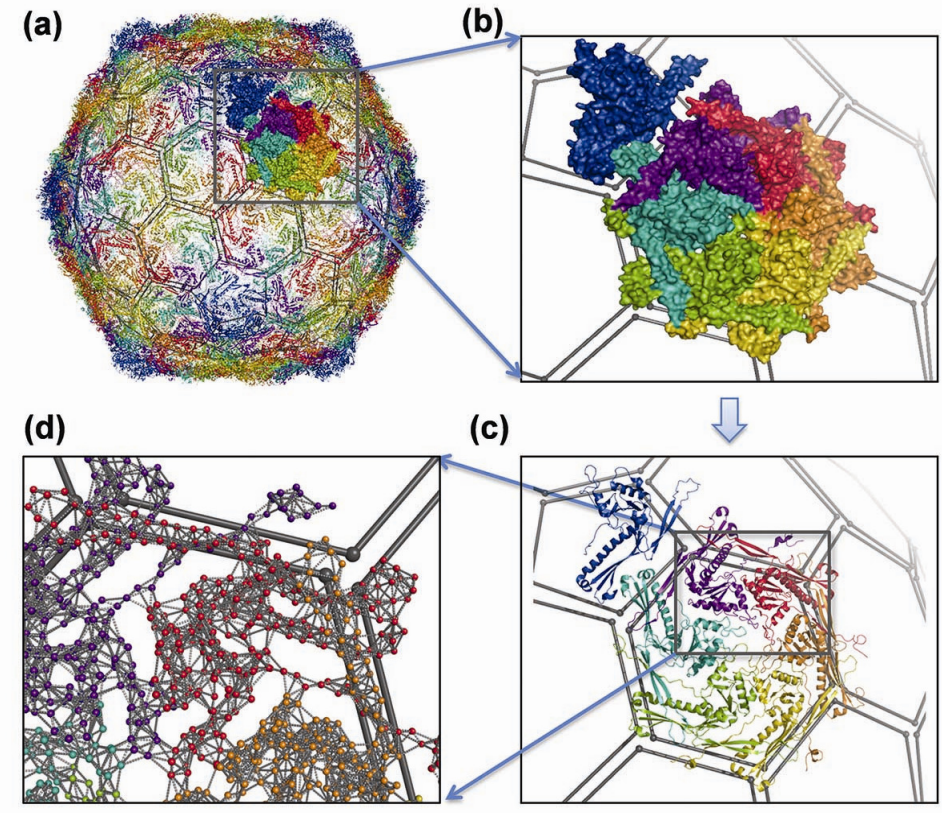

Fig. 7.1. ***From protein assemblies to network models.*** (a) External view of the intact viral capsid HK97 colored by chain, generated using the PDB file 2FT1 deposited by Johnson and coworkers (Gan *et al.* 2006). The capsid consists of 420 identical proteins arranged into 12 pentamers and 60 hexamers. (b) One asymmetric unit from panel (a) is enlarged. Each chain is in a distinctive color, indicating a possible scheme for rigid building blocks. (c) One asymmetric unit shown as secondary structures, in the same viewpoint and color scheme of panel (b). (d) A cartoon of the ENM in which the nodes are C^{α} atoms (spherical dots) and the edges represent the springs (or elastic couplings) connecting pairs of nodes located within a distance of r_c.

The harmonic approximation is ubiquitous in physics, and similar models have been used for calculating elastic properties of bulk polymers and classical lattice vibrations in crystals (see, for example, Marder 2000). Proteins, on the other hand, are small and somewhat flexible, and it was not until Tirion (1996) demonstrated that a harmonic potential faithfully captures the global dynamics of proteins that ENMs saw wide use in theoretical investigations of protein dynamics. Subsequent studies reinforced this observation, bestowing the ENM with particular relevance to proteins (Cui & Bahar 2006). It is well established that a relationship exists between protein structure and protein function. Here we have a simple network model that enables the calculation of global dynamics from structure alone, suggesting that protein dynamics is intermediate to structure and function, and ENMs provide an easily employable conduit between structure and dynamics.

An attractive feature of ENMs that keeps them in continuous use is that they provide a wealth of information at low computational cost. Construction of the EN is a matter of straightforwardly defining and linking nodes provided that information on structure, or simply on inter-residue contact topology, is available. The standard technique for determining dynamics or statistical distributions from an ENM is to conduct a mode decomposition using spectral graph theory and methods (e.g., Gaussian Network Model (GNM) (Bahar *et al.* 1997, Bahar *et al.* 1998) inspired by polymer network theory (Flory, 1976)) or a normal mode analysis (NMA) with uniform harmonic potentials, both of which provide analytical solutions to the equations of motion, bypassing the need to sample conformation space. Although NMA can also be applied to potentials derived from more detailed force fields, such calculations either require an initial energy minimization that inevitably distorts the input conformation, or else they risk producing energetically unstable solutions. ENMs, on the other hand, can take any conformation as input and guarantee that it resides at a minimum of the potential and therefore has physical motions. Furthermore, the few parameters used in ENMs can be easily adjusted as seen fit, giving ENMs uncommon adaptability. There are, however, some limitations to ENMs as predictive tools. Although ENMs robustly predict collective global motions, they do not fare as well in providing reliable descriptions of local motions. Also, the harmonic approximation requires a potential minimum, limiting the utility of ENMs for modeling non-equilibrium dynamics. A corollary of this second point is that only motions in the neighborhood of the global energy minimum can be accurately predicted by the ENMs, and one must be careful when interpreting the results from the model not to exceed the limits of the model.

Here we will outline the theory behind the ENM and some of its extensions, and then we will present some recent applications. We will focus on two groups of proteins, membrane proteins and viral capsids. Membrane proteins form one of the most ubiquitous classes of proteins, accounting for more than 25% of the proteins in most genomes (Wallin and von Heijne, 1998). Their functions cover a wide range of spectrum from transport of metabolites in prokaryotes to regulating and maintaining intra-cellular communications in eukaryotes by transporting ions. In mammalians, these proteins are responsible for maintaining the electrochemical gradients across cell-membrane, which is vital for efficient functioning of the central nervous system. Malfunctioning of membrane proteins leads to potentially fatal diseases, such as Alzheimer's, multiple sclereosis and arrhythmia (Ashcroft, 2000). Membrane proteins indeed constitute a large fraction of proteins currently targeted by approved drugs. Understanding the general principles of their structural dynamics and thereby mechanisms of function is thus essential in the rational design of therapeutics that target membrane proteins. By way of applications to a number of membrane proteins, we will illustrate in the present chapter how ENM approaches can provide insights into gating and/or signal transduction mechanisms.

Viruses constitute another group of proteins that are difficult to examine by all-atom simulations (due to their sizes of the order of Megadaltons), but are amenable to ENM analyses. The viral capsids in particular possess solid-like behavior, and can be well represented by elastic network models and their material properties. We will show how ENMs can greatly enhance our understanding of the complex dynamics of viral capsids, and open the way to simple descriptions in terms of measurable material properties. In summary, both groups of applications illustrate the utility of ENM approaches in providing simple descriptions of highly complex structures' dynamics, and gaining insights into potential mechanisms of biomolecular functions.

7.2. Theory and Assumptions

7.2.1. *Statistical mechanical foundations*

Potential energy. The elastic network model theory follows the same formalism that is commonly presented for studying small oscillations (Goldstein 1953). Here the physical system is a molecule or molecular assembly consisting of N constituent particles, where each particle may be an atom, a residue, or some other structural element acting as a node in the network. The changes in

generalized coordinates are defined by the vector $\mathbf{q} = (q_1, \ldots, q_n)^T$ of displacements from equilibrium. Typically the three Cartesian coordinates of each node are considered separately, giving $\mathbf{q}$ a total of $3N$ components (notable exceptions include the GNM, in which $\mathbf{q}$ has N components, and highly symmetric systems, such as viral capsids, for which symmetry can be exploited to reduce the dimensionality of $\mathbf{q}$). Near the equilibrium structure, the potential energy can be expanded as a power series in $\mathbf{q}$ as

$$V(\mathbf{q}) = V(0) + \sum_i \left(\frac{\partial V}{\partial q_i}\right)_0 q_i + \frac{1}{2}\sum_{ij}\left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_0 q_i q_j + \cdots . \tag{7.1}$$

The first term of the above expression is a constant that may be set to zero, and the second term is identically zero at a potential minimum. To second order, the potential is a sum of pairwise potentials

$$V(\mathbf{q}) = \frac{1}{2}\sum_{ij}\left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_0 q_i q_j \tag{7.2}$$

$$= \frac{1}{2}\sum_{ij} q_i U_{ij} q_j \tag{7.3}$$

$$= \frac{1}{2}\mathbf{q}^T \mathbf{U}\mathbf{q}\,, \tag{7.4}$$

where $\mathbf{U}$ is the matrix of second derivatives of the potential with respect to the generalized coordinates. It should be noted that $\mathbf{U}$ is symmetric and nonnegative definite.

Equation of motion. The kinetic energy can similarly be written in compact form as

$$T = \frac{1}{2}\dot{\mathbf{q}}^T \mathbf{M}\dot{\mathbf{q}}\,, \tag{7.5}$$

where the elements of the diagonal matrix $\mathbf{M}$ are the masses of the nodes and $\dot{\mathbf{q}}$ is the time derivative of $\mathbf{q}$. The equations of motion of the system are

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{U}\mathbf{q} = 0\,. \tag{7.6}$$

Here the double dot denotes the second derivative with respect to time.

Analytical solution. We solve Eq. (7.6) by transforming to mass-weighted coordinates, $\mathbf{r} = \mathbf{M}^{1/2}\mathbf{q}$, $\mathbf{K} = \mathbf{M}^{-1/2}\mathbf{U}\mathbf{M}^{-1/2}$, which yield

$$\ddot{\mathbf{r}} = -\mathbf{K}\mathbf{r}\,, \tag{7.7}$$

Note that the potential energy (Eq. (7.4)) can be expressed in terms of the mass-weighted coordinates and the mass-weighted stiffness matrix **K** as $V(\mathbf{r}) = \frac{1}{2}\mathbf{r}^T\mathbf{K}\mathbf{r}$. The solution to Eq. (7.7) is

$$\mathbf{r}(t) = \mathbf{a}e^{-i\omega t}. \tag{7.8}$$

From Eq. (7.7) and Eq. (7.8), we find that the coefficients **a** solve the eigenvalue equation $\mathbf{Ka} = \lambda\mathbf{a}$, where **a** is a vector of displacements along a normal mode of vibration, and the eigenvalue, λ, is the square of the normal mode frequency ω.

In most cases, **K** is not invertible, but has a well-defined number of eigenvalues that are identically zero. This occurs because the potential energy only depends on internal degrees of freedom and places no energetic restrictions on rigid-body rotations and translations. With this in mind, the inverse of **K**, when required, is replaced by the pseudo-inverse, defined as

$$\mathbf{K}^{-1} = \sum\nolimits_{\lambda_k \neq 0} \frac{\mathbf{v}_k \mathbf{v}_k^T}{\lambda_k}, \tag{7.9}$$

where λ_k are the nonzero eigenvalues of **K**, and $\mathbf{v}_k$ are their associated eigenvectors.

ENM partition function. The system's partition function can be calculated by integrating the potential over all possible changes in structure:

$$Z = \int d^n r \exp\left\{-\frac{1}{2k_B T}\mathbf{r}^T\mathbf{K}\mathbf{r}\right\} \tag{7.10}$$

$$= (2\pi k_B T)^{n/2}\left[\det(\mathbf{K}^{-1})\right]^{1/2}, \tag{7.11}$$

where k_B is the Boltzmann constant and T is the absolute temperature. As $\det(\mathbf{K}^{-1})$ is simply the product of the reciprocal nonzero eigenvalues of **K**, the lowest frequency modes contribute most to the partition function. These modes are also of highest interest when seeking to determine the most probable *global* fluctuations of a molecule. Indeed, the low-frequency, or 'slow', modes of an ENM are robust to variations in network topology, the level of resolution adopted in describing the network (see for example Doruker *et al.* 2002) and the force field adopted in NMA, and they reflect the intrinsically accessible motions that are endowed upon the molecule by its structure.

Mean-square fluctuations and cross-correlations. Expectation values for dynamical variables predicted by the ENM can be directly compared to experimental measurements. X-ray temperature factors (B-factors) provide a measure of the mean-square fluctuations of individual atoms. Similarly, an

ensemble of NMR structures can be used to calculate the correlations between the displacements of atoms. The correlations between node fluctuations in the ENM are given by

$$\langle r_i r_j \rangle = \frac{1}{Z} \int d^n r \exp\left\{ -\frac{1}{2k_B T} \mathbf{r}^T \mathbf{K} \mathbf{r} \right\} r_i r_j \tag{7.12}$$

$$= k_B T (\mathbf{K}^{-1})_{ij} , \tag{7.13}$$

and from Eq. (7.9) it is clear that the most significant contributions to the inter-residue correlations are also from the slowest modes.

7.2.2. *Anisotropic network models*

The most common ENMs are anisotropic network models (ANM) (Doruker, Atilgan and Bahar 2000; Atilgan *et al.* 2001; Tama and Sanejouand 2001) that use the $3N$ mass-weighted coordinates of the nodes as generalized coordinates: $\mathbf{r} = (\Delta x_1, \Delta y_1, \Delta z_1, \ldots, \Delta x_N, \Delta y_N, \Delta z_N)^T$, where $\Delta x_i = x_i - x_i^0$ is the x-component of the displacement of node i from its equilibrium position, $\mathbf{r}_i^0$. In this case the interaction matrix $\mathbf{K}$ is the $3N \times 3N$ Hessian matrix, $\boldsymbol{\mathcal{H}}$, of mixed second derivatives of the potential with respect to the coordinates of the residues. The Hessian might be thought of as an $N \times N$ matrix of 3×3 submatrices, each of which describes the energetic contribution from the interaction of two nodes. The elements of $\boldsymbol{\mathcal{H}}$ can be calculated from the potential energy,

$$V = \frac{1}{2} \sum\nolimits_{ij} \gamma_{ij} (R_{ij} - R_{ij}^0)^2 , \tag{7.14}$$

where γ_{ij} is the spring constant between nodes i and j, R_{ij} is their distance, and R_{ij}^0 is their equilibrium distance. The second derivatives of the potential function have the general form

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = -\frac{\gamma_{ij} (x_j - x_i)(y_j - y_i)}{R_{ij}^2} , \tag{7.15}$$

where x_i and y_j are the x- and y-coordinates of nodes i and j, respectively. Using the notation $x_{ij} = (x_j - x_i)$, and similarly for y_{ij} and z_{ij}, the off-diagonal super-elements of $\boldsymbol{\mathcal{H}}$ are

$$\mathbf{H}_{ij} = -\frac{\gamma_{ij}}{R_{ij}^2} \begin{bmatrix} x_{ij}^2 & x_{ij} y_{ij} & x_{ij} z_{ij} \\ x_{ij} y_{ij} & y_{ij}^2 & y_{ij} z_{ij} \\ x_{ij} z_{ij} & y_{ij} z_{ij} & z_{ij}^2 \end{bmatrix} , \tag{7.16}$$

and the diagonal super-elements satisfy

$$\mathbf{H}_{ii} = \sum_{j;\, j \neq i} \mathbf{H}_{ij} \,. \tag{7.17}$$

Diagonalization of $\mathcal{H}$ yields $3N - 6$ normal modes, each of which has a 3-vector component for every node. The remaining 6 modes have zero eigenvalue and correspond to rigid-body rotations and translations of the system.

The spring constants γ_{ij} are the only adjustable parameters in this model, and a variety of methods are used to select their values. Pairwise interactions are predominantly local, and a common practice is to assign a uniform spring constant, $\gamma_{ij} = \gamma$, to all pairs of nodes separated by less than some cutoff distance, and $\gamma_{ij} = 0$ for all others. It has been found empirically (Eyal *et al.* 2006) that when the nodes are taken to be the α-carbons of a protein, a cutoff distance of about 15Å results in residue mean-square fluctuations that correlate well with experimental B-factors. An alternative approach (Hinsen, 1998) that agrees comparably with experiments is to assign spring constants that decay with distance. Recent studies (Kondrashov, Cui and Phillips 2006) show that the adoption of stiffer force constants for the springs that connect first neighbors along the sequence further enhances the correlation with B-factors.

7.2.3. *Gaussian network model*

A simplification of the above-described model is the Gaussian network model (GNM) (Bahar *et al.*, 1997). This model uses the assumptions that node fluctuations are isotropic and Gaussian to reduce the interaction matrix from a $3N \times 3N$ Hessian to an $N \times N$ Kirchhoff matrix. Interestingly, this model often agrees better with experimental data than does its anisotropic counterpart, because the underlying potential penalizes the vectorial changes $\Delta\mathbf{R}_{ij} = \mathbf{R}_{ij} - \mathbf{R}_{ij}^0$ in internode distances (as opposed to penalizing the changes in the magnitudes $|\mathbf{R}_{ij}| - |\mathbf{R}_{ij}^0|$ only, as in the ANM; see Eq. (7.14)).

The inherent assumption of vibrational isotropy allows GNM to predict the size of motions and their cross-correlations, but not their directions. It also predicts the displacements along normal coordinates (e.g., slow modes) and permits us to define the domains engaged in concerted motions in the global modes, but not their mechanism/direction of concerted rearrangements. Note that the GNM uses only a single parameter, the elastic constant γ that defines the interactions between nodes that are separated by a distance less than a cutoff distance, r_c. When applied to proteins at the residue level, a value of r_c between

6.5Å and 7.5Å, corresponding to the radius of the first coordination shell, is typically selected for use in GNM.

The potential in Eq. (7.14) is a sum over pairwise potentials, each of which depends on the difference between the instantaneous distance between two nodes and their equilibrium separation. By assuming isotropic fluctuations, we can separate the spatial components of each node's motion, resulting in the GNM potential

$$V_{GNM} = \frac{1}{2}\sum_{ij}\gamma_{ij}[(R_{ij}-R_{ij}^0)\cdot(R_{ij}-R_{ij}^0)] \tag{7.18}$$

$$= \frac{1}{2}\sum_{ij}\gamma_{ij}\left[(x_{ij}-x_{ij}^0)^2+(y_{ij}-y_{ij}^0)^2+(z_{ij}-z_{ij}^0)^2\right] \tag{7.19}$$

$$= \frac{\gamma}{2}\sum_{ij}\Gamma_{ij}(\Delta x_{ij}^2+\Delta y_{ij}^2+\Delta z_{ij}^2) \tag{7.20}$$

$$= \frac{\gamma}{2}(\Delta\mathbf{x}^T\Gamma\Delta\mathbf{x}+\Delta\mathbf{y}^T\Gamma\Delta\mathbf{y}+\Delta\mathbf{z}^T\Gamma\Delta\mathbf{z}) \tag{7.21}$$

In the preceding lines we use the notation $\Delta x_{ij} = x_{ij} - x_{ij}^0$, and similarly for Δy_{ij} and Δz_{ij}. Likewise, we used the notation $\Delta\mathbf{x} = (\Delta x_1, \Delta x_2, \Delta x_3, \ldots, \Delta x_N)^T$ and similar expressions for $\Delta\mathbf{y}$ and $\Delta\mathbf{z}$, where Δx_i is the x-component of the vector Δr_i describing the fluctuation in the position of node i. The Kirchhoff adjacency matrix, $\mathbf{\Gamma}$, has elements

$$\Gamma_{ij=}\begin{cases} -1 \text{ if } R_{ij}\le r_c \text{ and } i\neq j \\ 0 \text{ if } R_{ij} > r_c \text{ and } i\neq j \\ -\sum_{k;k\neq i}\Gamma_{ik} \text{ if } i=j \end{cases}. \tag{7.22}$$

The equations of motion separate into three identical equations, one for each spatial coordinate, and the normal modes are obtained by diagonalizing $\mathbf{\Gamma}$. The assumption of isotropy essentially reduces the system to one dimension, and $\mathbf{\Gamma}$ has $N-1$ non-zero eigenvalues. Correlations between nodes can be found as before with

$$\left\langle \Delta r_i \Delta r_j \right\rangle = 3k_BT(\Gamma^{-1})_{ij}, \tag{7.23}$$

which is identical to Eq. (7.13) with $\mathbf{\Gamma}$ taking the place of $\mathbf{K}$ and a factor of three from the summation over spatial coordinates.

7.2.4. *Rigid block models*

One of the advantages of ENMs over more detailed force fields and simulations is their ability to produce analytical results with a coarse-grained model. One can quickly calculate the dynamics of a moderately large macromolecule simply by building and diagonalizing its Hessian matrix, an $O(N^3)$ computation. Occasionally, though, the memory requirements of the calculation overwhelm contemporary computers, and further coarse-graining is necessary. For example, if one wishes to consider the motion of side chains, a Hessian based on α-carbon coordinates alone is insufficient, and an ENM using all atoms must be used instead. Such refinement increases the system size approximately tenfold, possibly beyond the tolerance of available computational resources. Similarly, very large molecular assemblies such as viral capsids can contain upwards of 10^5 residues, and their Hessian matrices cannot be easily handled by conventional computing resources.

One way to overcome the problem of excessive system size is to bundle several elements of the physical system into a single node. This method faithfully reproduces the global dynamics of the system (Doruker *et al.* 2000), but does not produce detailed motions for all of the original nodes. Mixed models (Kurkcuoglu *et al.* 2003) that rely on multiple levels of coarse-graining can provide detailed results only for specific regions of interest. A good method for surmounting the problem of finding the normal modes of very large systems is the rotations and translations of blocks (RTB) (Tama *et al.* 2000), also called block normal mode (BNM) (Li and Cui, 2002) method. This method assumes that the system is constructed of n_b rigid blocks, and that the normal modes of the system can be expressed as rigid body rotations and translations of its constituent blocks. Each block has 6 degrees of freedom, and the approximation reduces the size of the system from $3N$ to $6n_b$.

Consider a system of N nodes that can be collected into $n_b < N$ rigid blocks connected by elastic springs. As before, the generalized mass-weighted coordinates form the $3N$-vector $\mathbf{r}$, and the system's Hessian, $\mathcal{H}$, is calculated from the topology of the elastic network. We define the $3N \times 6n_b$ projection matrix, $\mathbf{P}$, from the $3N$-dimensional space of all nodes into the $6n_b$-dimensional space of rotations and translations of the rigid blocks. The relationship between linear motion of a rigid body and the motion of its constituent components is captured by the conservation of linear momentum in mass-weighted coordinates:

$$\sqrt{M}\,\dot{\mathbf{r}}_{CM} = \sum_k \sqrt{m_k}\,\dot{\mathbf{r}}_k\,, \tag{7.24}$$

where the summation is over all nodes in the rigid block, m_k and $\mathbf{r}_k$ are the mass and position of node k, and $M = \Sigma_k m_k$ and $\mathbf{r}_{CM}$ are the mass of the block and position of its center of mass. Note that Eq. (7.24) holds not only for velocities, but also for positions and higher time derivatives. The matrix elements projecting from the full $3N$-dimensional space to translations in the block space are

$$\frac{\partial(\dot{\mathbf{r}}_{CM})_\mu}{\partial(\dot{\mathbf{r}}_k)_\nu} = \sqrt{\frac{m_k}{M}}\delta_{\mu\nu}, \tag{7.25}$$

where μ and ν indicate x-, y- or z-components of the vector (or tensor) enclosed in parentheses, and $\delta_{\mu\nu}$ is the Kronecker delta function. Similarly, angular momentum conservation gives

$$\mathbf{I}^{1/2}\dot{\theta} = \sum_k \sqrt{m_k}(\mathbf{r}_k \times \dot{\mathbf{r}}_k), \tag{7.26}$$

where $\mathbf{I}$ and θ are the block's moment of inertia tensor (the elements of which are given by $I_{\mu\alpha} \equiv \Sigma_k\, m_k\, (r_k{}^2\delta_{\mu\alpha} - (\mathbf{r}_k)_\mu\, (\mathbf{r}_k)_\alpha)$ and angular displacement vector. The components of $\dot{\theta}$ are

$$\dot{\theta}_\mu = \sum_k \sqrt{m_k} \sum_\alpha (\mathbf{I}^{-1/2})_{\mu\alpha}(\mathbf{r}_k \times \dot{\mathbf{r}}_k)_\alpha \tag{7.27}$$

$$= \sum_k \sqrt{m_k} \sum_\alpha (\mathbf{I}^{-1/2})_{\mu\alpha} \sum_{\beta\nu} (\mathbf{r}_k)_\beta (\dot{\mathbf{r}}_k)_\nu \varepsilon_{\alpha\beta\nu}, \tag{7.28}$$

where $\varepsilon_{\alpha\beta\nu}$ is the permutation symbol ($\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1$; $\varepsilon_{213} = \varepsilon_{321} = \varepsilon_{132} = -1$; otherwise $\varepsilon_{\alpha\beta\nu} = 0$), also known as Levi-Civita symbol. Differentiating the components of $\dot{\theta}$ with respect to the components of $\dot{\mathbf{r}}_k$, the matrix elements projecting from the full $3N$-dimensional space to rotations in the block space are

$$\frac{\partial\theta_\mu}{\partial(\dot{\mathbf{r}}_k)_\nu} = \sum_{\alpha\beta} \sqrt{m_k}(\mathbf{I}^{-1/2})_{\mu\alpha}(\mathbf{r}_k)_\beta \varepsilon_{\alpha\beta\nu}. \tag{7.29}$$

Using Eqs. (25) and (29) and the notation of Li and Cui (2002), the elements of the $3N \times 6n_b$ projection matrix $\mathbf{P}$ are

$$P^\mu_{J,j\nu} = \begin{cases} \sqrt{m_j / M_J}\,\delta_{\mu\nu} & \text{for } \mu = 1,2,3 \\ \sum_{\alpha\beta} (\mathbf{I}^{-1/2})_{(\mu-3),\alpha} \sqrt{m_j}\,(r_j - r_J{}^0)_\beta \varepsilon_{\alpha\beta\nu} & \text{for } \mu = 4,5,6 \end{cases} \tag{30}$$

where J is the rigid block index, $\mathbf{r}_J{}^0$ is the position vector of the mass center of block J, j is the node index, μ designates the rigid block translation ($1 \le \mu \le 3$) or rotation ($4 \le \mu \le 6$), and ν is the node displacement component ($1 \le \nu \le 3$). The

Hessian is projected into the space of rigid blocks with the transformation $\mathcal{H}_{BLK} = \mathbf{P}^T \mathcal{H} \mathbf{P}$, $\mathcal{H}_{BLK}$ is diagonalized with $\mathbf{V}_{BLK}^T \mathcal{H} \mathbf{V}_{BLK} = \Lambda_{BLK}$, and the resulting eigenvectors are projected back into the full *3N*-dimensional space with the inverse projection $\mathbf{V} = \mathbf{P}^{\mathrm{T}} \mathbf{V}_{BLK}$.

An example application of the RTB/BNM formalism is to virus maturation. Tama and Brooks (2005) modeled viral capsids of different sizes and symmetries using an ENM in which each capsomer was taken to be a rigid block. Their analysis showed that the conformational changes that occur during viral maturation can be largely accounted for with only a few icosahedrally symmetric slow modes. Another example application is to molecular motors. Li and Cui (2004) found that the conformational changes that occur in myosin and the calcium transporter Ca^{2+}-ATPase are dominated by a small number of slow modes, suggesting that Brownian motions have an essential role in the function of molecular motors.

7.2.5. *Treatment of perturbations*

It is often interesting to compare the global dynamics of a system in the absence and presence of some environmental perturbation applied at a given position, such as ligand binding. In such cases the perturbation takes the form of additional nodes, and the Hessian is calculated for the extended system, including these additional nodes. Comparing the normal modes of the original system to those of the perturbed system is not straightforward: Including the perturbation provides additional degrees of freedom, so the normal modes of the perturbed system are not necessarily orthogonal when projected into the space of the unperturbed system. It is useful to have an effective Hessian that will account for the influence of the perturbation without modifying the size of the system. This can be calculated as follows.

The state vector for an *N* node system is $\mathbf{r} = (r_1, \ldots, r_{3N})^T$, and the state vector for the same system in the presence of a perturbation, $\mathbf{e} = (e_1, \ldots, e_{3n})^T$ by a system of *n* nodes is $\mathbf{r}' = (s_1, \ldots, s_{3N}, e_1, \ldots, e_{3n})^T = (\mathbf{r}^T \ \mathbf{e}^T)^T$, where the first $3N$ components refer to the original system, and the last $3n$ to the environment or perturbing molecule. As demonstrated by Ming and Wall (2005), and by Zheng and Brooks (2005), the Hessian of a molecule within a specific environment can be decomposed as follows:

$$\mathcal{H} = \begin{pmatrix} \mathbf{H}_{ss} & \mathbf{H}_{se} \\ \mathbf{H}_{se}^T & \mathbf{H}_{ee} \end{pmatrix}, \tag{7.31}$$

where $\mathbf{H}_{ss}$ contains contributions from interactions of the original system with itself, $\mathbf{H}_{ee}$ accounts for interactions of the environment with itself, and $\mathbf{H}_{se}$ contains interactions between the system and its environment. Note that $\mathbf{H}_{ss}$ is not simply the unperturbed Hessian, but has different diagonal super-elements due to environmental contributions. The potential energy can be written as

$$V = \frac{1}{2}(\mathbf{r}')^T \mathcal{H}\mathbf{r}' \tag{7.32}$$

$$= \frac{1}{2}\mathbf{r}^T\mathbf{H}_{ss}\mathbf{r} + \mathbf{r}^T\mathbf{H}_{se}\mathbf{e} + \frac{1}{2}\mathbf{e}^T\mathbf{H}_{ss}\mathbf{e}\,, \tag{7.33}$$

using $\mathbf{e}^T\,\mathbf{H}_{se}{}^T\,\mathbf{r} = \mathbf{r}^T\,\mathbf{H}_{se}\,\mathbf{e}$. At equilibrium, $\partial V / \partial e_i = 0$ for all environmental nodes, giving

$$0 = \mathbf{H}_{es}\mathbf{r} + \mathbf{H}_{es}\mathbf{e}\,, \tag{7.34}$$

which yields

$$\mathbf{e} = -\mathbf{H}_{ee}^{-1}\mathbf{H}_{es}\mathbf{r}\,. \tag{7.35}$$

Substitution of this expression into Eq. (7.33) permits us to write the potential energy in terms of the $3N$ components of $\mathbf{r}$ as

$$V = \frac{1}{2}\mathbf{r}^T\bar{\mathbf{H}}\mathbf{r} \tag{7.36}$$

where $\bar{\mathbf{H}}$ is a pseudo-Hessian with the same dimensionality as the unperturbed Hessian, but which includes the environmental effects:

$$\bar{\mathbf{H}} = \mathbf{H}_{ss} - \mathbf{H}_{se}\mathbf{H}_{ee}^{-1}\mathbf{H}_{se}^T\,. \tag{7.37}$$

Diagonalizing $\bar{\mathbf{H}}$ leads to the normal modes in the presence of the perturbation, and these can be directly compared to the modes of unperturbed system. This approach has been used to examine conformational changes in myosin and kinesin nucleotide-binding pockets. Zheng and Brooks (2005) employed a model in which the binding pockets of motor proteins constituted the system, and the remainder of the protein made up its environment. They showed that the dynamics relevant to the myosin binding pocket are coupled to its global modes, in agreement with hypothesized pathways between actin binding and force generation. Ming and Wall (2006) used this method to demonstrate that substrate allosteric proteins usually bind their substrates at sites that induce significant perturbation in the collective dynamics.

7.2.6. *Langevin dynamics*

The equations of motion that are most commonly adopted in ENM studies do not generally account for frictional forces (see Eq. (7.6)). Nonetheless, biomolecules exist in viscous environment, and viscous drag may alter their normal modes of motion. It is therefore useful to have a technique for calculating normal modes of motion in the presence of damping forces. Perhaps the simplest way to introduce viscous drag is through the Langevin equation:

$$\mathbf{M}\ddot{\mathbf{q}} = -\mathbf{U}\mathbf{q} - \zeta\dot{\mathbf{q}} + \boldsymbol{\xi}(t)\,. \tag{7.38}$$

Here ζ is a velocity-dependent damping term and ξ(t) is a time-dependent vector of random forces, also called white noise, which satisfies the conditions

$$\left\langle \xi_i(t) \right\rangle = 0 \tag{7.39}$$

$$\left\langle \xi_i(t)\xi_j(t') \right\rangle = 2\zeta_{ij}\delta(t-t')k_BT\,, \tag{7.40}$$

In mass-weighted coordinates, Eq. (7.38) becomes

$$\ddot{\mathbf{r}} = -\mathbf{K}\mathbf{r} - \mathbf{F}\dot{\mathbf{r}} + \mathbf{R}(t)\,, \tag{7.41}$$

with **K** as defined earlier, $\mathbf{F} = \mathbf{M}^{-1/2}\zeta\,\mathbf{M}^{-1/2}$ is the mass-weighted friction matrix, and $\mathbf{R} = \mathbf{M}^{-1/2}\boldsymbol{\xi}$.

Defining the $6N \times 6N$ matrix (Miller *et al.* 2008)

$$\mathbf{A} = \begin{pmatrix} 0 & I \\ -\mathbf{K} & -\mathbf{F} \end{pmatrix}, \tag{7.42}$$

in which **I** is the $3N \times 3N$ identity matrix, Eq. (7.41) may be re-written as

$$\begin{pmatrix} \dot{\mathbf{r}} \\ \ddot{\mathbf{r}} \end{pmatrix} = \mathbf{A}\begin{pmatrix} \mathbf{r} \\ \dot{\mathbf{r}} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{R}(t) \end{pmatrix}, \tag{7.43}$$

and the normal modes can be solved analytically by diagonalizing **A**. The first $3N$ components of the eigenvectors of **A** provide the displacements along the normal modes, and the last $3N$ components correspond to the mode velocities. The eigenvalues of **A** are complex; their imaginary parts are the oscillatory frequencies of the modes, and their real parts are the exponential decay constants of their amplitudes. This approach has been used by Miller *et al.* (2008) to estimate the fractional free energy loss in the myosin power stroke.

In the limit of strong friction, all of the modes are over-damped and the system obeys Brownian dynamics. Hinsen *et al.* (2000) demonstrated that the

normal modes in this limit are found by diagonalizing the friction-weighted force constant matrix,

$$\hat{\mathbf{U}} = \zeta^{-1/2}\mathbf{U}\zeta^{-1/2}. \tag{7.44}$$

In such a case, the system does not oscillate, but displacements from the equilibrium conformation will relax along the eigenvectors of $\hat{\mathbf{U}}$ with relaxation constants given by the corresponding eigenvalues. This technique has been used to calculate scattering functions of proteins, and to investigate the sources behind damping in global protein motions (Hinsen *et al.* 2000).

7.3. Applications

7.3.1. *Membrane proteins*

Membrane proteins are typically composed of three domains: an extracellular (EC) domain exposed to the periplasm, an intracellular/cytoplasmic (IC or CP) domain buried in the cytoplasm, and a transmembrane (TM) domain embedded in the lipid bilayer. Some membrane proteins, known as receptors, are involved in signal transmission via recognition and binding of substrate/ligand to the EC domain, which triggers conformational changes in the CP domain. The allosteric coupling between different domains or the concerted motions, permit the protein to recognize, bind or translocate substrates. Other membrane proteins serve as ion channels or substrate transport. Permeations of ions and/or substrates thus require collective relaxation mechanism or cooperative motions, which are usually amenable to ENMs. Here we focus on recent progress made in delineating the dynamics of four groups of membrane proteins, potassium channels, acetylcholine receptors, rhodopsin and mechanosensitive channels using ENM-based methods.

Potassium channels: Common gating mechanism observed in different potassium channels. The TM domain of K^+ channels is composed of a bundle of eight α–helices contributed by four identical monomers (Fig. 7.2). At the center of these helices is a narrow *selectivity filter* (towards the EC region), followed by a large cavity in the middle, and a long gating region, also called *pore*, that connects to the CP region (Fig. 7.3a). MD studies have provided us with insights in to the mechanism of function at the selectivity filter, including the preferential selectivity of potassium over sodium (Shrivastava and Sansom, 2000, Shrivastava *et al.* 2002, Bernèche and Roux, 2000) and the free energy

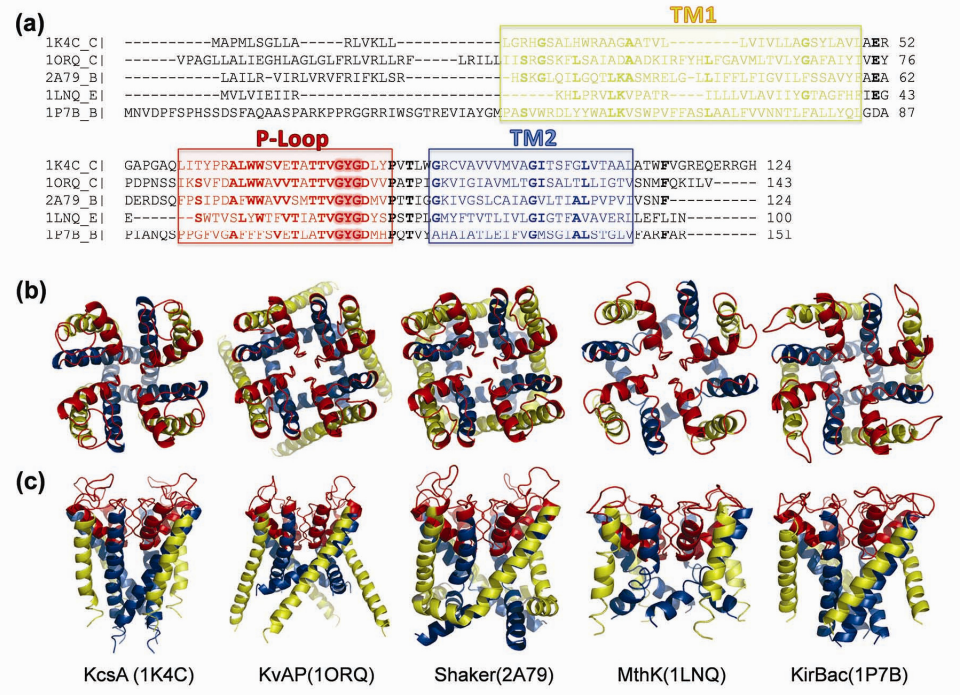

Fig. 7.2. ***Sequence and structure of the pore region of five structurally known potassium channels.*** (a) Alignment of the pore region sequences. The regions corresponding to the helices TM1 and TM2 and the selectivity filter are indicated by yellow, blue and red blocks, respectively. The alignment was performed using ClustalW (Thompson *et al.*, 1994). Fully or highly conserved regions are shown in bold. The signature motif GYG at the selectivity filter, is highlighted and shaded in red (b) Structural comparison of the pore forming regions of the K^+ channels aligned in panel a. These are all tetrameric structures, the monomers of which contain either 2 TM helices (KcsA, MthK and KirBac) colored yellow (TM1) and blue (TM2), or 6 TM helices (KvAP and Shaker) denoted as S1-S6. Only the pore forming helices S5 and S6, equivalent to TM1 and TM2, are displayed here, along with the selectivity filter, (which is colored red in all the structures).

profile along the selectivity filter (Bernèche and Roux, 2003). Yet, a fundamental question that remained unanswered until recently has been the mechanism of *gating*, i.e., the conformational events that allow for the transfer of ions from the central cavity to the CP region through the channel-pore. In the X-ray structures, the radius of the pore is too small to let the ions through. The question was: how does the narrow pore open up to permit the permeation of potassium ions?

Toward gaining an understanding of the potential mechanism of pore opening, we recently examined five K^+ channels, KcsA, KirBac, MthK, KvAP and Shaker (Shrivastava & Bahar, 2006). Figure 7.2a shows a comparison of the

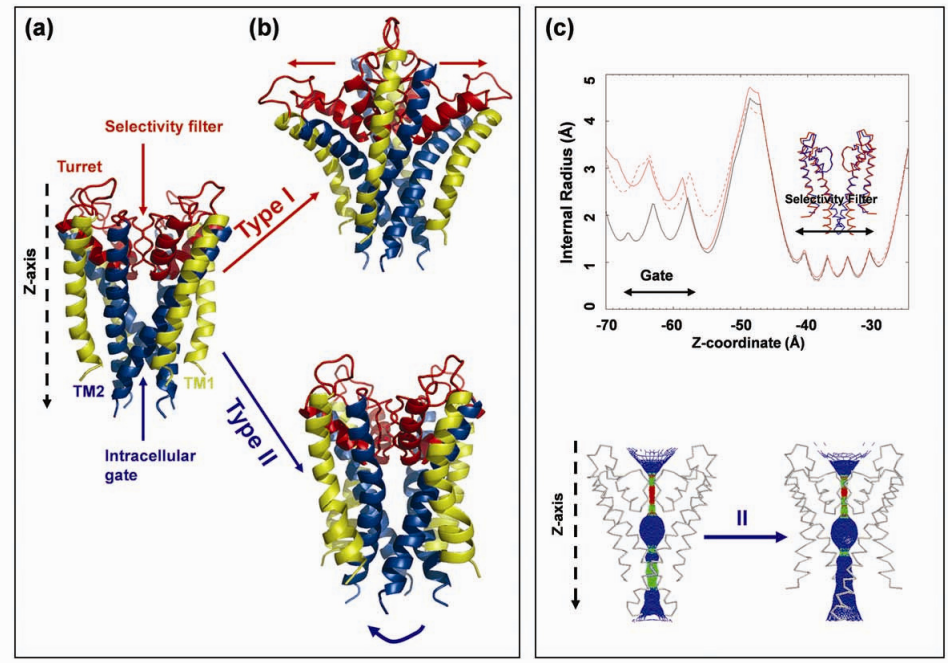

Fig. 7.3. ***Mobilities in global modes shown by color-coded diagrams for KcsA.*** (a) The different regions of the protein are indicated, namely, the inner and outer helix (TM1 and TM2 respectively), the putative intracellular gate, the selectivity filter and the turret region. (b) The slowest global mode (mode type I), which is two fold degenerate (top), is an opening/stretching motion, with the turrets opening and closing above the selectivity filter. The second lowest frequency mode (mode type II) (bottom), corresponds to a twisting/torsion motion, leading to a widening of the pore, shown in the next panel. (c) Pore-radius profiles (top panel) as a function of the position along the cylindrical (Z-) axis. Bottom panel shows a solid-sphere representation of the inner surface of the channel at the pore region for the crystal structure (left), for the model of the open form (right). The color code for the solid-sphere representation of the pore region is: red, pore radius < 1.15 Å; green, 1.15 Å < radius < 2.30 Å; and blue, radius > 2.30 Å. The pore radius profiles were generated using HOLE (Smart *et al.*, 1993). In the inset of panel (c) is the backbone of the crystal structure (blue) superimposed onto the model of the open form (red). Two monomers have been deleted for clarity.

sequences and pore region structures of these five channels. The observed structural similarities in the pore-forming region suggested a common gating mechanism, which was indeed verified by GNM/ANM calculations. The equilibrium dynamics of these five channels were found to obey similar patterns on a global scale. Mainly, two types of highly cooperative motions were identified at the low frequency end of the mode spectrum, shared by all five structures: The first (referred to as type I) is an alternating expansion/contraction of the EC and/or CP via anti-correlated fluctuations of oppositely located pairs of

monomers (Fig. 7.3a). The second (type II) is global torsion of the helical bundle similar to a cork-screw mechanism, with the net result of inducing an enlargement of the pore region (Figs. 7.3b and 7.3c). The change in the relative spacing of the TM2 helices (Shrivastava & Bahar, 2006) was observed to be in accord with the models based on site-directed spin labeling and EPR spectroscopy data (Perozo *et al.* 1999) and experimental structures of the open form.

This study reinforces the observation that proteins have an inherent ability to undergo conformational changes required for their biological function (Ma, 2005, Bahar *et al.* 2007, Tama and Sanejouand, 2001, Xu *et al.* 2003). Computational studies performed by Sansom and coworkers for investigating the dynamics of inward rectifying potassium channels (Kirs) (Haider *et al.* 2005) further indicated a good agreement (correlation coefficient of 0.87) between the mean-square fluctuations of Kir3.1 residues obtained from molecular dynamics (MD) simulations and those predicted by ANM. The lowest frequency mode from ANM indicated an asymmetric dimer-of-dimers motion, which is also in agreement with that inferred from MD simulations, suggesting that this mechanism of motion is a robust property of the structure. (Sansom *et al.* 2005). A good correlation was also reported more recently between ENM results and site-directed mutagenesis experiments for KcsA and Mthk (Haliloglu and Ben-Tal, 2008).

***nAcetyl Choline Receptor* (*nAChR*): *Gating via global twist of the quaternary structur*e**. As a member of the receptor family of membrane proteins, this homo- or hetero-pentamer switches between ion-permeable and –impermeable conformations upon binding or releasing its neurotransmitter substrate, acetylcholine (ACh). The ACh binding site is located at the boundary between the subunits in the EC region. Binding of ACh promotes a transient opening of the channel. Several models have been proposed for the structural transition mediating the signal transmission (Changeux and Edelstein, 1998; Taly *et al.* 2005; Liu *et al.* 2008, Szarecka *et al.* 2007). Normal mode analysis on the complete structure revealed a concerted symmetric quaternary twist motion (Taly *et al.* 2005), with the EC and IC domains rotating in opposite directions resulting in a wide opening of the pore, compatible with experimental observations. GNM/ANM analyses (Szarecka *et al.* 2007) of structural models based on cryo-electron microscopy data (Unwin, 2005) also revealed two types of quasi-symmetric twisting motions: Type I inducing a twisting of ligand binding domain (LBD) in opposite direction to that of TM and IC domains; and Type II where the

central TM domain undergoes counter-rotation with respect to the EC and CP domains (Figs. 7.4a and 7.4c). Both motions induce an increase in the pore diameter (Figs. 7.4b and 7.4d).

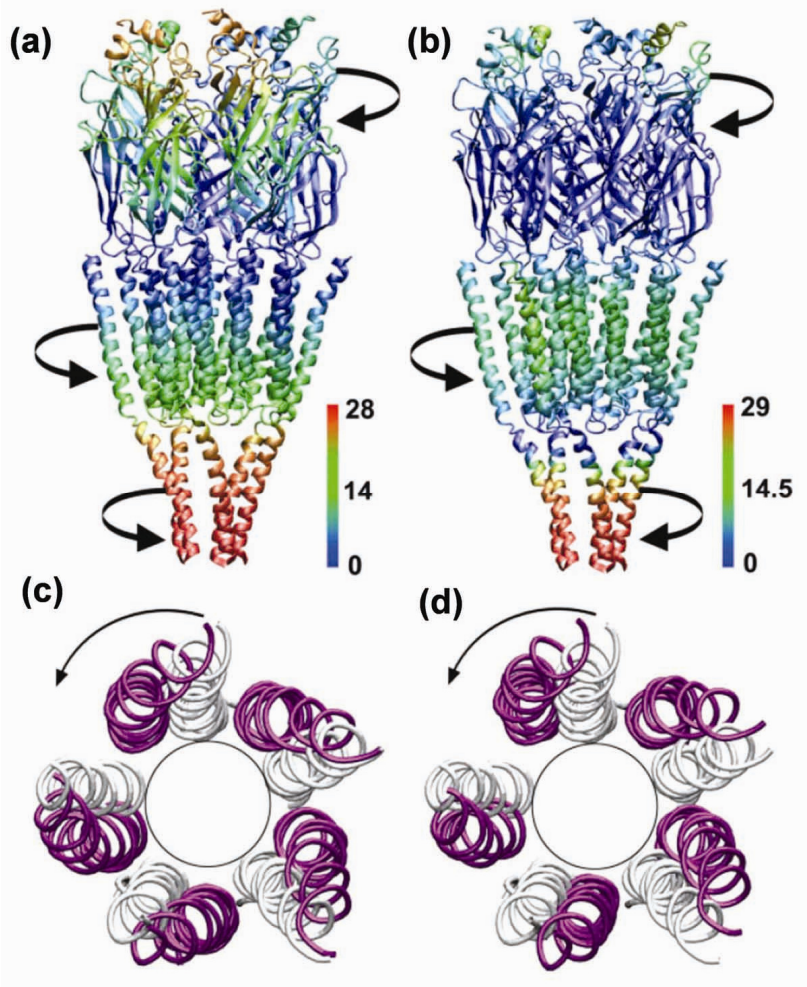

Fig. 7.4. ***Quasisymmetric twisting motion of heteropentameric nAChR.*** GNM mean-square (MS) fluctuations of the global mode types I and II are shown, mapped onto a ribbon diagram of the protein. The fluctuation values are color-coded on the nAChr structure (red:high, green:moderate, blue:low). The black arrows indicate the phases of twisting directions from the correlated ANM modes. The TM2 domains from the models of open-pore structures (magenta) calculated from ANM eigenvectors and eigenvalues are compared with the starting close-pore structure (gray) for type I (c) and type II (d) twists. Note the widening of the pore as a result of the twists (*Figure adapted from* ***Szarecka et al. 2007***).

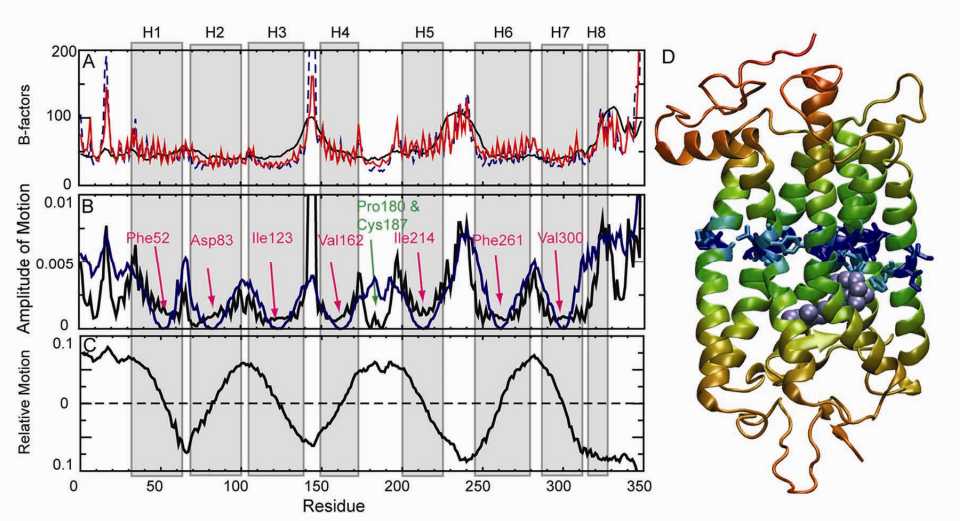

Fig. 7.5. ***Dynamics of rhodopsin predicted by the GNM and ANM.*** (a) Experimental(black), GNM (red) and ANM (dashed blue) predicted thermal B-factors. (b) Distribution of square displacements of residues predicted by GNM (blue) and ANM (black). The non-TM regions exhibit higher mobilities, in general. ANM yields two additional minima: Pro180 and Cys187 near the EC entrance to the chromophore binding pocket. (c) The global mode eigenvector calculated with the GNM, indicating relative motions of different regions of the proteins along the principal mode coordinate. Positive and negative regions delineate structural blocks subject to concerted motions. The locations of the helices (1-8) are indicated on the upper abscissa and distinguished by gray bands. (d) Ribbon diagram of rhodopsin color-coded according to the relative motions in (b) in order of increasing mobilities: blue (lowest mobility), cyan, green, yellow, orange, red (highest mobility). Side chains are shown for the seven GNM hinges labeled in (b) and 11-*cis*-retinal is shown in light blue space-filling representation.

Rhodopsin: An activation mechanism coupling the EC and CP domains. As the only structurally determined member of the G-protein coupled receptors (GPCR) family, rhodopsin has been widely studied by both experimental and computational techniques. Capture of the substrate G-protein triggers a highly cooperative conformational change in rhodopsin accompanied by the isomerization of its chromophore (11-*cis*-retinal) at the TM region. The chromophore binding pocket is a highly packed region. The perturbation of the structure at this region drives the propagation of the conformational change via cooperative rearrangements of TM helices to the cytoplasmic (CP) domain, to induce the active state of rhodopsin, metarhodopsin II (Meta-II) (Isin *et al.* 2006). The application of GNM to two dark state structures of rhodopsin 1L9H (Okada *et al.* 2002) and 1U19A(Okada *et al.* 2004) yielded the B-factor profile (Fig. 7.5A) in close agreement with the experimental data (correlation coefficient

of **0.837**). The loops between helices 3 and 4 and the C-terminus are distinguished by their enhanced mobilities.

A clearer picture of the relative mobilities of the different structural components in the long-time regime is obtained upon examination of the slowest mode profile obtained by GNM in Fig. 7.5 (panels B and C). The slowest mode divides the structure into two regions subject to anticorrelated fluctuations. Mainly, the positive and negative portions of the mode 1 eigenvector plotted in Fig. 7.5C define the two anticorrelated regions. The cross-over regions between them form the minima in the square displacements profile shown in Fig. 7.5B. The corresponding residues occupy central positions in the TM helices (Fig. 7.5D). Since they also lie at the interface between the two anticorrelated regions of the molecule, these residues play an important role in transmitting conformational perturbations. Many residues lying in this critical region (e.g., D83, V162, F261) participate in the retinal binding pocket, and efficiently propagate local conformational changes between the CP and extracellular (EC) ends of the molecule (Isin *et al.* 2006). ANM analysis shows that this mode essentially drives a global twisting of the TM helices, which results in an overall expansion at the two ends. These conformational changes agree well with experimental data (Isin *et al.* 2006). In particular, the mobility of spin-labeled side chains at the buried surfaces of TM helices 1, 2, 3, 6 and 7 were found to increase upon isomerization, indicating a reduced packing consistent with the expansion of the pore, in accord with ANM results.

Mechanosensitive Channel (MscL): Channel widening upon global twisting and torsion. These proteins act as a "safety-valve" in *E. Coli*: they open up when the osmotic pressure is beyond a certain threshold (Hamill & Martinac, 2001, Anishkin & Kung, 2005), thus preventing membrane breakdown and cell lysis. The diameter of the gate region, as inferred from the X-ray structure of the closed form (Chang *et al.* 1998), is ~2 Å, whereas in the open form it is ~30-35 Å (Sukharev *et al.* 1999), suggesting a significant conformational change. ENM studies have elucidated the dynamics of this mechanosensitive protein (Valadie *et al.* 2003, Haliloglu & Ben-Tal, 2008). Two major kinds of motions were identified: Type I (Fig. 7.6), a symmetrical motion that corresponds to an overall iris-like opening, exhibited by the non-degenerate modes; and Type II, which resulted in a global bending/tilting. Notably, three non-degenerate modes (modes 11, 31 and 64) (Valadie *et al.* 2003) accessible to the closed state can alone account for 65% of the conformational change observed between the closed and open states, while the first 100 modes describe 76% of the transition. As to the opposite change, five non-degenerate modes recover 65% of the

conformational change. Interestingly, in this protein as well, the twisting and tilting motions result in widening up of the channel pore.

Although more than 25% of the human genome is made up of membrane proteins (Bond *et al.* 2007), only ~100 X-ray structures of membrane proteins are known to date (White, 2004). Moreover, most of these structures are from prokaryotic organisms. Thus insights on mechanics of biological functions of these proteins that are made from these structures are vital in understanding the functioning of the respective human homologues. In particular, the 'twist-to-open' mechanism instrumental in the gating function of most of the membrane proteins discussed here, suggests a common mechanism of pore-opening when the pore architecture exhibits a cylindrical symmetry with funnel-like organization of a bundle of helices.

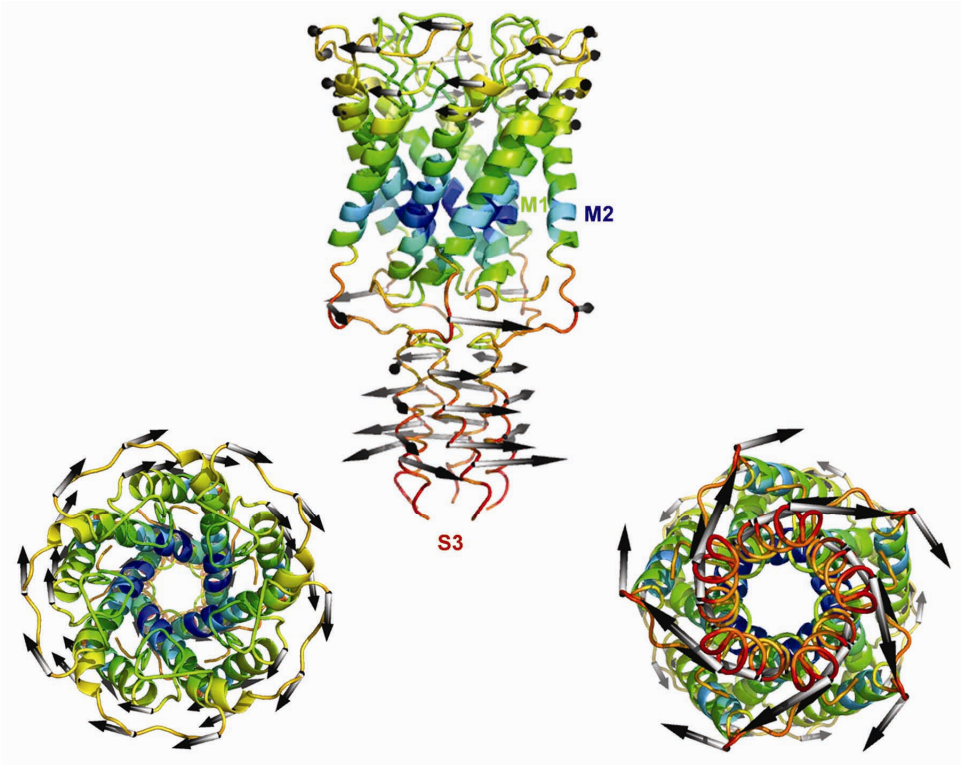

Fig. 7.6. ***Twisting and torsion MscL in the slow-frequency twisting/torsion mode.*** The top panel illustrates the side view with the protein with a vector representation of the amplitude and direction of motion predicted by the ANM (see also Valadie *et al.* 2003). The lengths of the arrows scale with the amplitude of the motion. The bottom panel shows the motion of the protein as viewed from top (left) and bottom (right).

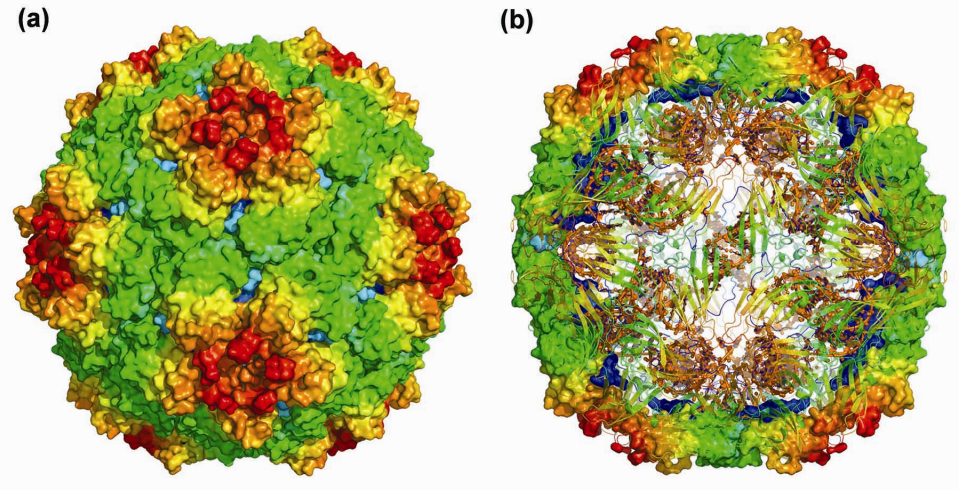

Fig. 7.7. ***The structure of STMV.*** (a) The structure of STMV represented by its capsid, colored by geometric positions to illustrate the icosahedral symmetry. (b) A cross section view of the STMV, highlighting the capsid and enclosed genetic material (RNA). The RNA, shown as cartoon representation in yellow, on which spherical dots represents the positions of atoms P, C2 and C4'. These atoms are used to build up the elastic network in ANM.

7.3.2. *Viruses*

Viral capsids are some of the largest systems to have been studied with ENMs. Their large size – on the order of a million atoms – places many viruses beyond the reach of MD. Even NMA of viral ENMs is computationally cumbersome, necessitating the use of various techniques to further simplify normal mode calculation. The RTB method was used (Tama & Brooks, 2002; Tama & Brooks 2005) to investigate swelling of viral capsids, leading to the observation that capsid maturation in several viruses can be largely accounted for with only a few non-degenerate (icoasahedrally symmetric) slow modes. The symmetry of viral capsids was exploited (Kim *et al.* 2003, Kim *et al.* 2004) to construct simplified ENMs for studies of capsid maturation, while the maturation of the HK97 bacteriophage was explored using the GNM (Rader *et al.* 2005). The dynamics of sufficiently small viruses can thus be analyzed through direct application of an ENM without further simplification, as exemplified below.

The satellite tobacco mosaic virus (STMV) is one of the smallest viruses known, consisting of 60 identical protein subunits arranged in an icosahedral shell about a single-stranded RNA with 1058 nucleotide bases (Figs. 7.7a and 7.7b) (Dodds, 1998, Day *et al.* 2001). Recent studies of STMV include the structural analysis

of its RNA, MD simulations of the intact capsid (Freddolino *et al.* 2006) and direct measurements of its elastic properties. We analyzed the normal modes of STMV using the ANM (Doruker *et al.* 2000; Atilgan *et al.* 2001). Starting from the PDB (Berman *et al.* 2000) structure 1A34 (Larson *et al.* 1998), we built an elastic network using the atoms C^{α} on the capsid proteins, and the atoms P, C2 and C4' on the RNA nucleotides as the network nodes. A cutoff distance of 15 Å was used to define connections between nodes, and the force constant was set to 1 N/m (note the conversion factor 1 N/m = 1.44 kcal/mol/ $Å^2$).

Results from ANM analysis of STMV. The group of rotations that preserve icosahedral symmetry has finitely many irreducible representations: 1, 3, 4 and 5 dimensional representations (Tinkam, 1964, Widom *et al.* 2007). Since the dimensionality of the irreducible representations determines the degeneracy of each normal mode, the allowed degeneracies of vibrational frequency are simply 1 (nondegenerate), 3, 4 and 5. No other degeneracies may occur. Figure 7.8 shows the eigenvalues calculated using the ANM in the cases of STMV with RNA and the STMV protein coat alone.

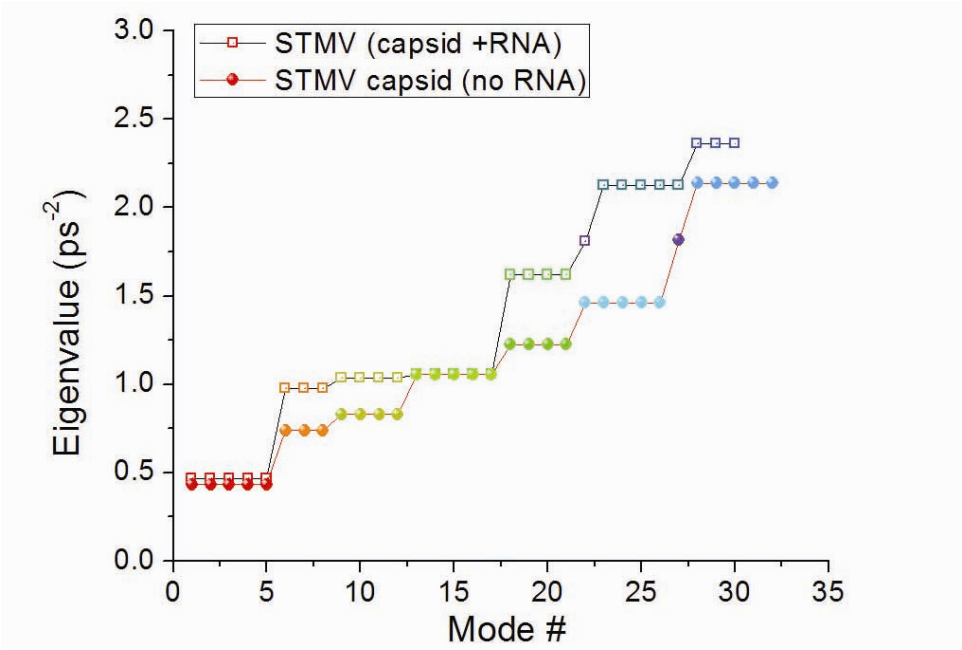

Fig. 7.8. ***Dispersion of mode frequencies.*** The curves display the eigenvalues calculated using the ANM for STMV with RNA (empty dots) and STMV protein coat alone (solid spheres). They are colored in groups. Each distinctive color represents a type of motion, with certain degeneracy (1, 3, 4 or 5). Higher resolution figure can be provided upon request.

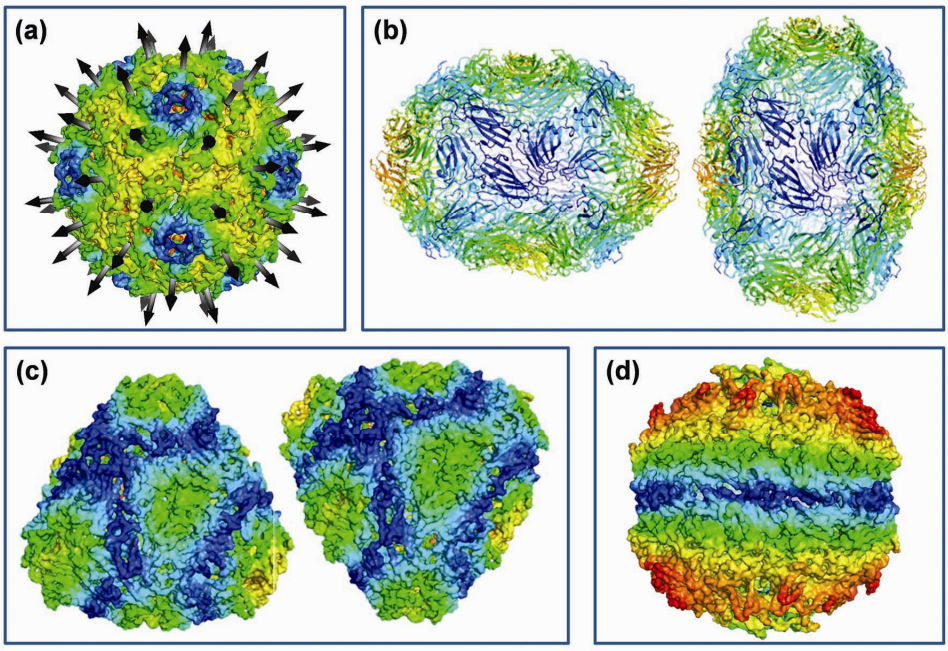

Fig. 7.9. ***Global mode shapes predicted for the intact STMV.*** *Results are based on ANM* (a) The $l = 0$ breathing mode, mode 22 in Fig. 7.8. (b) The $l = 2$ squeezing mode, 5-fold degenerate modes 1-5. (c) The $l = 3$ squeezing mode, modes 6-9. (d) The torsional mode (13–17). All panels are colored according to their mobilities from small (blue) to large (red).

The collective eigenmodes can be divided into two distinctive types (Coccia *et al.* 1998). The first type is torsional modes, in which the deformations have no radial component. An example of this type of motion is modes13-17, shown in Fig. 7.9d. The other type is spheroidal modes (e.g. modes 1-5, Fig. 7.9b), in which the eigenvectors contain both tangential and radial components. Each mode can be described by a wave number, l, that corresponds to the degree of the spherical harmonic that best aligns with the mode. Qualitatively, the wave number may be thought of as the degree of symmetry of a mode: $l = 0$ indicates spherical symmetry, $l = 2$ indicates deformation along a single axis, and so on. Figures 7.9a–c illustrates the spheroidal modes with wave numbers 0, 2 and 3, respectively.

Figure 7.9a displays a typical $l = 0$ spheroidal mode, corresponding to mode 22 of STMV (capsid + RNA in the mode frequency distribution shown in Fig. 7.8. It is non-degenerate because it preserves icosahedral symmetry. Such modes correspond to purely radial motions — shrinking or swelling (*breathing*) of the entire structure. This type of deformation occurs in response to strong

internal pressure, such as that exerted by the genome encapsulation, or external pressure, such as osmotic pressure. It is notable that the eigenvalue of this mode is higher than other low frequency modes (Fig. 7.8), implying that the capsid exhibits a relatively stronger resistance against deformations of this type.

Figure 7.9b shows the slowest mode with five-fold degeneracy (l = 2), i.e., modes 1-5 in Fig. 7.8. The motion induced in these modes can be visualized physically as the result of *squeezing* a sphere radially inwards at the poles, allowing it to bulge outwards at the equator, and vice versa. Such a deformation occurs when a molecule is probed with an atomic force microscope. Our recent calculations (Yang *et al.* 2008) showed that hollow spheres are quite soft in response to this mechanism of deformation (Michel *et al.* 2006; Kol *et al.* 2006). This is the top ranking (lowest frequency) mode in both STMV and STMV+RNA, which indicates that this kind of deformation is highly favorable from an energetic point of view.

The l = 3 mode (Fig. 7.9c) is similar to the l = 2 mode above, but the deformation direction is split into three. This mode involves two groups of degenerate modes, 6-8 and 9-12**.** The torsional mode illustrated in Fig. 7.9d is a fivefold degenerate mode with l = 2. This mode can be visualized as the result of twisting the upper and lower hemispheres in opposite directions.

Effect of RNA. There is no difference between the first 21 modes of the STMV capsid (alone) and those of the capsid with RNA in so far as the mechanism of motion is concerned, but their orders (or relative frequencies) exhibit slight changes. The eigenvalues of the capsid with RNA are slightly higher than their counterparts for the protein coat only (Fig. 7.8), but this is essentially due to the larger number of nodes and higher mass of the capsid with RNA. The eigenvalue of the l = 2 spheroidal squeezing modes (modes 1-5) increases by 8% in the presence of the genome, whereas that of the l = 2 torsional twisting modes (modes 13-17) increases by only 0.4%. Differences appear after the 21st mode; for example, the breathing mode appears earlier in the case of the capsid with RNA.

7.4. Conclusion

Elastic network models lead to a unique analytical solution and provide a thorough sampling of the energy landscape near the energy minimum. Advances in accurate representation of systems and in validation procedures have highlighted coarse-grained approaches as valuable tools for analysis, allowing a direct comparison with experimental results on macromolecular dynamics.

Based on a simplified harmonic potential, ENMs have been enhanced to account for a variety of effects that influence the dynamics of biological molecules. An advantage of ENMs is that they usually provide an accurate description of the mechanisms of motions, although no absolute time scale and size of motions can be predicted. Another advantage is to simplify our understanding of the complex and diverse interactions in biological systems with the help of a simple model, and a small number of parameters. It should be noted, however, that in a strict sense they are applicable to the close neighborhood of the native (or equilibrium) state. They essentially inform us on the intrinsic dynamic preferences of biomolecular systems, which are verified in numerous applications to be functionally relevant (Bahar *et al.* 2007). ENMs also provide us with physical insights: The overall topology of the protein plays a major role in the mechanical behavior of the protein, implying that proteins related by evolution are expected to show similar quantitative behavior, as seen in the case of potassium channels, discussed above. Basic research in biology and biochemistry along with statistical mechanical and analytical methods will thus lead to improved transferability and predictability of such approaches.

References

Ashcroft, F.M. (2000) Ion Channels and Disease. Academic Press, San Diego pp 472.

Atilgan, A.R., Durell, 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. *Folding & Design* 2: 173-181.

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

Bahar, I., Chennubhotla, C., and Tobi, D. (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Cur. Op. Struct. Biol.* 17: 633-640.

Bahar, I., and Rader, A.J. (2005) Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol.* 15:1-7.

Bernèche, S. and Roux, B.(2003) A microscopic view of ion conduction through the K^+ channel.

Proc. *Natl. Acad. Sci. USA*. 100:8644-8648.

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.

Coccia, E., Fafone, V., Frossati, G., Lobo, J.A., Ortega, J.A. (1998) Hollow sphere as a detector of gravitational radiation. *Phys. Rev. D* 57: 2051-2060.

Changeux, J.P. and Edelstein, S.J. (1998) Allosteric receptors after 30 years. *Neuron*. 21:959-980.

Chennubhotla, C., Z Yang,Z., and Bahar, I. (2008) Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol. Biosystems* 4: 287-292.

Cui, Q., and Bahar, I. (2006) Normal Mode Analysis: Theory and Applications to Biological and *Chemical Systems.* Boca Raton: Chapman & Hall/CRC.

Day, J., Kuznetsov, Y.G., Larson, S.B., Greenwood, A., McPherson, A. (2001) Biophysical Studies of the RNA Cores of Satellite Tobacco Mosaic Virus. *Biophys J* 80: 2364-2371.

De Pablo, P.J., Schaap, I.A.T., MacKintosh, F.C., Schmidt, C.F. (2003) Deformation and Collapse of Microtubules on the Nanometer Scale. *Phys. Rev. Lett.* 91: 098101.

Dodds, J.A .(1998) Satellite Tobacco Mosaic Virus. *Ann. Rev. Pathol.* 36: 295-310.

Doruker, P., Atilgan,A.R., and Bahar, I. (2000) Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins* 40: 512-524.

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

Eyal, E., and Bahar, I. (2008) Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys. J.* 94: 3424-3435.

Eyal, E., Yang, L.-W., and Bahar, I. (2006) Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* 22: 2619-2627.

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

Freddolino, P.L., Arkhipov, A.S., Larson, S.B., McPherson, A., Schulten, K. (2006) Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Structure* 14: 437-449

Gan, L., Speir, J.A., Conway, J.F., Lander, G., Cheng, N., Firek, B.A., Hendrix, R.W., Duda, R.L., Liljas, L., Johnson, J.E. (2006) Capsid Conformational Sampling in HK97 Maturation Visualized by X-Ray Crystallography and Cryo-EM. *Structure* 14: 1655-1665

Goldstein, H. (1953) *Classical Mechanics.* Cambridge: Addison-Wesley.

Haider, S., Grottesi, A., Hall, B.A., Ashrcoft, F.M. and Sansom, M.S.P. (2005) Conformational dynamics of the ligand-binding domain of inward rectifier K channels are revealed by molecular dynamics simulations: toward an understanding of Kir-channel gating. *Biophys. J.* 88:3310-3320.

Haliloglu, T. and Ben-Tal, N. (2008) Cooperative transition between open and closed conformations in potassium channels. *PLoS Comp. Biol.* 4: Pcbi.1000164.

Hinsen, K. (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins* 33 : 417-429.

Hinsen, K, Petrescu, A.-J., Dellerue, S., Bellissent-Funel, M.-C., and Kneller, G.R.. (2000) Harmonicity in slow protein dynamics. *Chem. Phys.* 261: 25-37.

Isin, B., Rader, A.J., Dhiman, H.K., Klein-Seetharaman, J. and Bahar, I. (2006) Predisposition of the dark state of rhodopsin to functional changes in structure. *Proteins: Struct. Func. Bioinf.* 65:970-983.

Isin, B., Schulten, K., Tajkhorshid, E. and Bahar, I. (2008) Mechanism of signal propagation upon retinal isomerization: insights from molecular dynamics simulations of rhodopsin restrained by normal modes. *Biophys. J.* 95: 789-803.

Keskin, O., Bahar, I., Flatow, D., Covell, D.G., and Jernigan, R.L. (2002) Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry* 4: 491-501.

Kim, M.K., and Jernigan, R.L. (2005) Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. *Biophys. J.* 89: 43-55.

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.

Kol, N., Gladnikoff, M, Barlam, D., Shneck, R.Z., Rein, A., Rousso, I. (2006) Mechanical Properties of Murine Leukemia Virus Particles: Effect of Maturation. *Biophys J* 91: 767-774.

Kondrashov, D.A., Cui,Q., and Phillips, G.N. (2006) Optimization and evaluation of a coarse-grained model of protein motion using x-ray crystal data. *Biophys. J.* 91: 2760-2767.

Kurkcuoglu, O, Jernigan RL, and P Doruker. (2003) Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions. *Polymer* 45: 649-657.

Kurkcuoglu, O., Jernigan, and Doruker, P. (2005) Collective dynamics of large proteins from mixed coarse-grained elastic network model. *QSAR Comb. Sci.* 24: 443-448.

Larson, S.B., Day, J., Greenwood, A., McPherson, A. (1998) Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *J. Mol. Biol.* 277: 37-59.

Li, G., and Cui, Q. (2002) A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to CA2+-ATPase. *Biophys. J.* 83: 2457-2474.

Li, G., and Cui, Q. (2004) Analysis of functional motions in brownian molecular machines with an efficient block normal mode approach: myosin-II and Ca2+-ATPase. *Biophys. J.* 86: 743-763.

Liu, X., Xu ,Y., Li ,H., Wang HYPERLINK "http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.0040019" \l "aff1" - ,X.,

Jiang,H. and Barrantes, F.J. (2008) Mechanics of Channel Gating of the Nicotinic Acetylcholine Receptor. *PLoS. Comp. Biol.* 4:pcbi0040019

Ma, J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* 13:373-380. Ashcroft, F.M. (2000) Ion Channels and Disease. Academic Press, San Diego pp 472.

Marder, M.P. (2000) *Condensed Matter Physics.* New York : John Wiley & Sons.

Michel, J.P., Ivanovska, I.L., Gibbons, M.M., Klug, W.S., Knobler, C.M., Wuite, G.J.L.,

Schmidt, C.F. (2006) Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. *Proc. Natl. Acad. Sci.* U S A 103: 6184-6189.

Miller, B.T., Zheng, W., Venable,R.M., Pastor, R.W., and Brooks, B.R. (2008) Langevin network model of myosin. *J. Phys. Chem. B* 112: 62746281.

Ming, D., and Wall, M.E. (2005) Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* 95: 198103.

Ming, D., and Wall, M.E. (2006) Interactions in native binding sites cause a large change in protein dynamics. *J. Mol. Biol.* 358: 213-223.

Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J. Landau, E. M. and Shichida, Y. (2002) Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. *Proc. Nat. Acad. Sci. USA*. 99: 5982-5987.

Okada, T., Sugihara, M., Bondar, A.N., Elstner, M., Entel, P. and Byss, V. (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. J. Mol. Biol. 342:571-583.

Perozo, E., Cortes, D.M. and Cuello, L.G. (1999) Structural rearrangements underlying K^+ channel activation. *Science* 285:73-78.

Rader, A.J., Vlad, D.H., and Bahar, I. (2005) Maturation dynamics of bacteriophage HK97 capsid. *Structure* 13: 413-421.

Roux, B., Bernèche, S. and Im,W. (2000) Ion Channels, Permeation, and Electrostatics: Insight into the Function of KcsA. *Biochemistry,* 39 (44), 13295 -13306.

Shrivastava, I.H. and Sansom, M.S.P. (2000) Simulations of Ion Permeation through a Potassium Channel: Molecular Dynamics of KcsA in a Phospholipid Bilayer. *Biophys J.* 78:557-570

Shrivastava, I.H., Tieleman, D.P., Biggin, P.C. and Sansom**,** M.S.P.(2002) K^+ versus Na^+ Ions in a K Channel Selectivity Filter: A Simulation Study**.** *Biophys J,* 83:633-645.

Shrivastava, I.H. and Bahar, I. (2006) Common mechanism of pore opening shared by five different potassium channels. *Biophys J* 90: 3929-3940.

Sulkowska, J.I., Kloczkowski, A., Sen, T.Z., Cieplak, M., and Jernigan, R.L. (2008) Predicting the order in which contacts are broken during single molecule protein stretching experiments. *Proteins* 71: 45-60.

Szarecka, A., Xu, Y., and Tang, P. (2007) Dynamics of heteropentameric nictotinic acetylecholine receptor:Implications of the gating mechanism. *Proteins* 68:948-960.

Taly, A., Delarue, M., Grutter, T., Nilges, M., Le Novère, N., Corringer, P.J., and Changeux, J.P. (2005) Normal mode analysis suggests a quartenary twist model for the Nicotine Receptor gating mechanism. *Biophys J.* 88:3954-3965.

Tama, F., and Brooks III, C.L. (2002) The mechanism and pathway of pH induced swelling in Cowpea Chlorotic mottle virus. *J. Mol. Biol.* 318: 733-737.

Tama, F., and Brooks, III, C.L. (2005) Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J. Mol. Biol.* 345: 299-314.

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., Marques,O., and Sanejouand, Y.-H. (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* 41: 1-7.

Tama, F., Vall,e M., Frank, J., and Brooks, C.L. (2003) Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA* 100: 9319-9323.

Tinkham, M. (1964) Group Theory and Quantum Mechanics: McGraw Hill.

Tirion, M. M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77: 1905-1908.

Tobi, D., and Bahar, I. (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA* 102: 18908-18913.

Unwin, N. (2005) Refined structure of nicotinic acetylcholine receptor at 4Å resolution. *J. Mol. Biol*. 346:967-989.

Valadie, H., Lacapere, J.J., Sanejouand-, Y.-H., Etchebest, C. (2003) Dynamics properties of MScL of Escherirchia coli: A normal mode analysis *J. Mol Biol*. **332**, 657-674.

Wallin, E. and von Heijne, G.(1998) Genome-wide analysis of int*egral membrane proteins from eubacterial, archeana nd eukaryotic organisms.* Protein Sci. 7: 1029-1038.

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

Widom, M., Lidmar, J., David, R.N. (2007) Soft modes near the buckling transition of icosahedral shells. *Phys. Rev. E.* 76: 031911.

White, S.H. (2004) The progress of membrane protein structure determination. *Protein Sci.* 13: 1948-1949.

Xu, C., Tobi, D. and Bahar, I. (2003) Allosteric Changes in Protein Structure Computed by a Simple Mechanical Model: Hemoglobin T --> R2 Transition. *J. Mol. Biol.* 333:153-168

Yang, L.-W., and Bahar, I. (2005) Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure* 13: 893-904.

Zheng, W., and Brooks, B.R. (2005) Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin. *Biophys. J.* 89: 167-178.

Zheng, W., Brooks, B.R., and Hummer. G. (2007) Protein conformational transitions explored by mixed elastic network models. *Proteins* 69: 43-57.