3

The Gaussian Network Model: Theory and Applications

A.J. Rader, Chakra Chennubhotla, Lee-Wei Yang, and Ivet Bahar

CONTENTS

3.1 Introduction

A view that emerges from many studies is that proteins possess a tendency, encoded in their three-dimensional (3D) structures, to reconfigure into

functional forms, that is, each native structure tends to undergo conformational changes that facilitate its biological function. An efficient method for identifying such functional motions is normal mode analysis (NMA), a method that has found widespread use in physical sciences for characterizing molecular fluctuations near a given equilibrium state. The utility of NMA as a physically plausible and mathematically tractable tool for exploring protein dynamics has been recognized for the last 20 years [1, 2]. With recent increase in computational power and speed the application of NMA to proteins has gained renewed interest and popularity.

Contributing to this renewed interest in utilizing NMA has been the introduction of simpler models based on polymer network mechanics. The Gaussian network model (GNM) is probably the simplest among these. This is an elastic network (EN) model introduced at the residue level [3, 4], inspired by the full atomic NMA of Tirion with a uniform harmonic potential [5]. Despite its simplicity, the GNM and its extension, the anisotropic network model (ANM) [6], or similar coarse-grained EN models combined with NMA [7–9], have found widespread use since then for elucidating the dynamics of proteins and their complexes. Significantly, these simplified NMAs with EN models have recently been applied to deduce both the machinery and conformational dynamics of large structures and assemblies including HIV reverse transcriptase [10, 11], hemagglutinin A [12], aspartate transcarbamylase [13], F1 ATPase [14], RNA polymerase [15], an actin segment [16], GroEL-GroES [17], the ribosome [18, 19], and viral capsids [20–22].

Studying proteins with the GNM provides more than the dynamics of individual biomolecules, such as identifying the common traits among the equilibrium dynamics of proteins [23], the influence of native state topology on stability [24], the localization properties of protein fluctuations [25], or the definition of protein domains [26, 27]. Additionally, GNM has been used to identify residues most protected during hydrogen–deuterium exchange [28, 29], critical for folding [30–34], conserved among members of a given family [35], or involved in ligand binding [36].

The theoretical foundations of the GNM will be presented in this chapter, along with a few applications that illustrate its utility. The following questions will be addressed. What is the GNM? What are the underlying assumptions? How is it implemented? Why and how does it work? How does the GNM analysis differ from NMA applied to EN models? What are its advantages and limitations compared to coarse-grained NMA? What are the most significant applications and prospective utilities of the GNM, or the EN models in general?

To this end, the chapter begins with a brief overview of conformational dynamics and the relevance of such mechanical motions to biological function. Section 3.2.1 is devoted to explaining the theory and assumptions of the GNM as a simple, purely topological model for protein dynamics. The casual reader may elect to skip over Sections 3.2.2 to 3.2.4 where the derivation of the GNM is presented using fundamental principles from statistical mechanics. In Section 3.3, attention is given to how the GNM is implemented

(Section 3.3.1), and to what extent it can be or cannot be applied to proteins in general. An interpretation of the physical meaning of both fast and slow modes is presented (Section 3.3.2) with examples for a small enzyme, ribonuclease T1. Section 3.3.3 describes how the GNM differs from the ANM (i.e., from the NMA of simplified EN models) and discusses when the use of one model is preferable to the other. Finally, results are presented in Sections 3.3.4 and 3.3.5 for two widely different applications: specific motions of supramolecular structures and classification of motions in general through the iGNM online database of GNM motions. The chapter concludes with a discussion of potential future uses.

3.1.1 Conformational Dynamics: A Bridge Between Structure and Function

With recent advances in sequencing genomes, it has become clear that the canonical sequence-to-function paradigm is far from being sufficient. Structure has emerged as an important source of additional information required for understanding the molecular basis of observed biological activities. Yet, advances in structural genomics have now demonstrated that structural knowledge is not sufficient for understanding the molecular mechanisms of biological function either. The connection between structure and function presumably lies in *dynamics*, suggesting an encoding paradigm of sequence to structure to dynamics to function.

Not surprisingly, a major endeavor in recent years has been to develop models and methods for simulating the dynamics of proteins, and relating the observed behavior to experimental data. These efforts have been largely impeded, however, by the memory and time cost of molecular dynamics (MD) simulations. These limitations are particularly prohibitive when simulating the dynamics of large structures or supramolecular assemblies.

3.1.2 Functional Motions of Proteins Are Cooperative Fluctuations Near the Native State

While accurate sampling of conformational space is a challenge for macromolecular systems, the study of protein dynamics benefits from a great simplification: proteins have uniquely defined native structures under physiological conditions, and they are functional only when folded into their native conformation. Therefore, while the motions of macromolecules in solution are quite complex and involve transitions between an astronomical number of conformations, those of proteins near native state conditions are much simpler, as they are confined to a subset of conformations, or microstates, near the folded state. These microstates usually share the same overall fold, secondary structural elements, and even tertiary contacts within individual domains. Typical examples are the open and closed forms of enzymes, usually adopted in the unliganded and liganded states, respectively. Exploring the *fluctuation*

dynamics of proteins near native state conditions is a first step toward gaining insights about the molecular basis and mechanisms of their function; and fluctuation dynamics can be treated to a good approximation by linear models — such as NMA.

Another distinguishable property of protein dynamics — in addition to confinement to a small subspace of conformations — is the *collective* nature of residue fluctuations. The fluctuations are indeed far from random, involving the correlated motions of large groups of atoms, residues, or even entire domains or molecules whose concerted movements underlie biological function. An analytical approach that takes account of the collective coupling between all residues is needed, and again NMA emerges as a reasonable first approximation.

3.2 The Gaussian Network Model

3.2.1 A Minimalist Model for Fluctuation Dynamics

Most analytical treatments of complex systems dynamics entail a compromise between physical realism and mathematical tractability. A challenge is to identify the simplest, yet physically plausible, model that retains the physical and chemical characteristics, which are needed for the time and length scales of interest. Clearly, as the size and length scales of the processes of interest increase, it becomes unnecessary to account for many of the microscopic details in the model. The inclusion of these microscopic details could, on the contrary, tend to obscure the dominant patterns characterizing the biological function of interest.

The GNM was proposed by Bahar et al. [3] within such a minimalist mindset to explore the role and contribution of *purely topological constraints*, defined by the 3D structure, on the collective dynamics of proteins. Inspired by the seminal work of Flory and collaborators applied to polymer gels [37], each protein is modeled by an EN (Figure 3.1), the dynamics of which is entirely defined by network topology. The position of the nodes of the EN are defined by the C^α-atom coordinates, and the springs connecting the nodes are representative of the bonded and nonbonded interactions between the pairs of residues located within an interaction range, or cutoff distance, of r_c. The cutoff distance is usually taken as 7.0 Å, based on the radius of the first coordination shell around residues observed in PDB structures [38, 39].

3.2.2 GNM Assumes Fluctuations Are Isotropic and Gaussian

If we define equilibrium position vectors of a node, i, by $\mathbf{R}_i^0$, and the instantaneous position by $\mathbf{R}_i$, the fluctuations, or deformations, from this mean position can then be defined by the vector $\Delta \mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$. The fluctuations
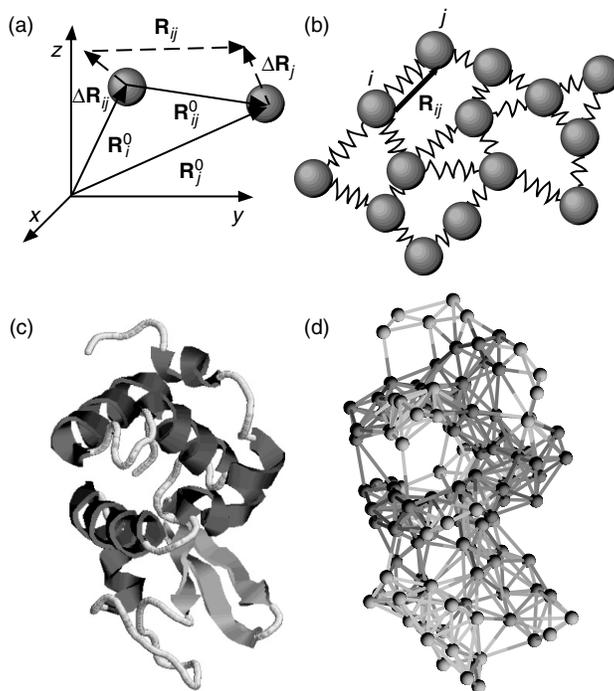
FIGURE 3.1

(See color insert following page 136) Description of the GNM. (a) Schematic representation of the equilibrium positions of the ith and jth nodes, $\mathbf{R}_i^0$ and $\mathbf{R}_j^0$, with respect to a laboratory-fixed coordinate system (xyz). The instantaneous fluctuation vectors, $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$, are shown by the dashed arrows, along with the instantaneous separation vector $\mathbf{R}_{ij}$ between the positions of the two residues. $\mathbf{R}_{ij}^0$ is the equilibrium distance between nodes i and j. (b) In the EN of GNM every residue is represented by a node and connected to spatial neighbors by uniform springs. These springs determine the $N-1$ degrees of freedom in the network and the structure's modes of vibration. (c) Three dimensional image of hen egg white lysozyme (PDB file 1hel [46]) showing the C^α trace. Secondary structure features are indicated by pink for helices and yellow for β-strands. (d) Using a cutoff value of 10 Å, all connections between C^α nodes are drawn for the same lysozyme structure to indicate the nature of the EN analyzed by GNM.

in the distance vector $\mathbf{R}_{ij}$ between residues i and j, can in turn be expressed as $\Delta\mathbf{R}_{ij} = \mathbf{R}_{ij} - \mathbf{R}_{ij}^0 = \Delta\mathbf{R}_j - \Delta\mathbf{R}_i$ (Figure 3.1[a]). By assuming that these fluctuations are *isotropic* and *Gaussian* we can write the potential of the network of N nodes (residues), V_{GNM}, in terms of the components ΔX_i, ΔY_i, and ΔZ_i of $\Delta\mathbf{R}_i$, as

$$V_{\text{GNM}} = \frac{\gamma}{2}\left[\sum_{i,j}^{N}\Gamma_{ij}[(\Delta X_i - \Delta X_j)^2 + (\Delta Y_i - \Delta Y_j)^2 + (\Delta Z_i - \Delta Z_j)^2]\right]$$

(3.1)

where Γ_{ij} is the ijth element of the Kirchhoff (or connectivity) matrix of inter-residue contacts defined by

$$
\Gamma_{ij} =
\begin{cases}
-1, & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\
0, & \text{if } i \neq j \text{ and } R_{ij} > r_c \\
-\sum\limits_{j,j \neq i} \Gamma_{ij}, & \text{if } i = j
\end{cases}
\tag{3.2}
$$

and γ is the force constant taken to be uniform for all network springs. Expressing the X-, Y-, and Z-components of the fluctuation vectors $\Delta \mathbf{R}_i$ as three N-dimensional vectors $\Delta \mathbf{X}$, $\Delta \mathbf{Y}$, and $\Delta \mathbf{Z}$, Equation (3.1) simplifies to

$$
V_{\text{GNM}} = \frac{\gamma}{2} [\Delta \mathbf{X}^{\text{T}} \boldsymbol{\Gamma} \Delta \mathbf{X} + \Delta \mathbf{Y}^{\text{T}} \boldsymbol{\Gamma} \Delta \mathbf{Y} + \Delta \mathbf{Z}^{\text{T}} \boldsymbol{\Gamma} \Delta \mathbf{Z}]
\tag{3.3}
$$

where $\Delta \mathbf{X}^{\text{T}}$, $\Delta \mathbf{Y}^{\text{T}}$, and $\Delta \mathbf{Z}^{\text{T}}$ are the row vectors $[\Delta X_1, \Delta X_2, \ldots, \Delta X_N]$, $[\Delta Y_1, \Delta Y_2, \ldots, \Delta Y_N]$, and $[\Delta Z_1, \Delta Z_2, \ldots, \Delta Z_N]$, respectively. The total potential can alternatively be expressed as

$$
V_{\text{GNM}} = \frac{\gamma}{2} [\Delta \mathbf{R}^{\text{T}} (\boldsymbol{\Gamma} \otimes \mathbf{E}) \Delta \mathbf{R}]
\tag{3.4}
$$

where $\Delta \mathbf{R}$ is the $3N$-dimensional vector of fluctuations, $\Delta \mathbf{R}^{\text{T}}$ is its transpose, $\Delta \mathbf{R}^{\text{T}} = [\Delta X_1, \Delta Y_1, \ldots, \Delta Z_N]$, $\mathbf{E}$ is the identity matrix of order 3, and $(\boldsymbol{\Gamma} \otimes \mathbf{E})$ is the direct product of $\boldsymbol{\Gamma}$ and $\mathbf{E}$, found by replacing each element Γ_{ij} of $\boldsymbol{\Gamma}$ by the 3×3 diagonal matrix $\Gamma_{ij}\mathbf{E}$. One should note that by construction the eigenvalues for this $3N \times 3N$ matrix, $\Gamma \otimes \mathbf{E}$, are threefold degenerate. This degeneracy arises from the isotropic assumption, further explored in the following section.

3.2.3　Statistical Mechanics Foundations of the GNM

What we are primarily interested in is determining the mean-square (ms) fluctuations of a particular residue, i, or the correlations between the fluctuations of two different residues, i and j. These respective properties are given by

$$
\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = \langle \Delta X_i^2 \rangle + \langle \Delta Y_i^2 \rangle + \langle \Delta Z_i^2 \rangle
\tag{3.5}
$$

and

$$
\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \langle \Delta X_i \Delta X_j \rangle + \langle \Delta Y_i \Delta Y_j \rangle + \langle \Delta Z_i \Delta Z_j \rangle
\tag{3.6}
$$

Thus, if we know how to compute the component fluctuations $\langle \Delta X_i^2 \rangle$ and $\langle \Delta X_i \Delta X_j \rangle$ then we know how to compute the residue fluctuations and their cross-correlations.

In the GNM, the probability distribution of all fluctuations, $P(\mathbf{\Delta R})$ is *isotropic* (Equation [3.7]) and *Gaussian* (Equation [3.8]), that is,

$$P(\mathbf{\Delta R}) = P(\mathbf{\Delta X}, \mathbf{\Delta Y}, \mathbf{\Delta Z}) = p(\mathbf{\Delta X})p(\mathbf{\Delta Y})p(\mathbf{\Delta Z}) \tag{3.7}$$

and

$$\begin{aligned}
p(\mathbf{\Delta X}) &\propto \exp\left\{-\frac{\gamma}{2k_\mathrm{B}T}\mathbf{\Delta X}^\mathrm{T}\mathbf{\Gamma}\mathbf{\Delta X}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\mathbf{\Delta X}^\mathrm{T}\left(\frac{k_\mathrm{B}T}{\gamma}\mathbf{\Gamma}^{-1}\right)^{-1}\mathbf{\Delta X}\right)\right\}
\end{aligned} \tag{3.8}$$

where k_B is the Boltzmann constant and T is the absolute temperature. Similar forms apply to $p(\Delta \mathbf{Y})$ and $p(\Delta \mathbf{Z})$. $\mathbf{\Delta X} = [\Delta X_1, \Delta X_2, \ldots, \Delta X_i, \ldots, \Delta X_N]$ is therefore a multidimensional Gaussian random variable with zero mean and covariance $(k_\mathrm{B}T/\gamma)\mathbf{\Gamma}^{-1}$ in accord with the general definition [40]

$$W(\mathbf{x}, \boldsymbol{\mu}, \mathbf{\Xi}) = \frac{1}{(2\pi)^{N/2}|\mathbf{\Xi}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\mathrm{T}\mathbf{\Xi}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{3.9}$$

for multidimensional Gaussian (normal) probability density function associated with a given N-dimensional vector $\mathbf{x}$ having mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Xi}$. Here, the term in the denominator, $(2\pi)^{N/2}|\mathbf{\Xi}|^{1/2}$, is the partition function that ensures the normalization of $W(\mathbf{x}, \boldsymbol{\mu}, \mathbf{\Xi})$ upon integration over the complete space of accessible $\mathbf{x}$, and $|\mathbf{\Xi}|$ is the determinant of $\mathbf{\Xi}$. Similarly, the *normalized* probability distribution $p(\Delta \mathbf{X})$ is

$$p(\mathbf{\Delta X}) = \frac{1}{Z_\mathbf{X}}\exp\left\{-\frac{1}{2}\left(\mathbf{\Delta X}^\mathrm{T}\left(\frac{k_\mathrm{B}T}{\gamma}\mathbf{\Gamma}^{-1}\right)^{-1}\mathbf{\Delta X}\right)\right\} \tag{3.10}$$

where $Z_\mathbf{X}$ is the partition function given by

$$Z_\mathbf{X} = \int \exp\left\{-\frac{1}{2}\left(\mathbf{\Delta X}^\mathrm{T}\left(\frac{k_\mathrm{B}T}{\gamma}\mathbf{\Gamma}^{-1}\right)^{-1}\mathbf{\Delta X}\right)\right\}\mathrm{d}\mathbf{\Delta X} = (2\pi)^{N/2}\left|\frac{k_\mathrm{B}T}{\gamma}\mathbf{\Gamma}^{-1}\right|^{1/2} \tag{3.11}$$

In the GNM, the determinant of the Kirchhoff matrix is zero, and the inverse, $\mathbf{\Gamma}^{-1}$, which scales with the covariance, cannot therefore be directly computed. $\mathbf{\Gamma}^{-1}$ is found instead by eigenvalue decomposition of $\mathbf{\Gamma}$ and reconstruction of the inverse using the $N - 1$ nonzero eigenvalues and associated eigenvectors. The same expression in Equation (3.11) is valid for $Z_\mathbf{Y}$ and $Z_\mathbf{Z}$ such

that the overall GNM partition function (or configurational integral) becomes

$$Z_{\text{GNM}} = Z_X Z_Y Z_Z = (2\pi)^{3N/2} \left| \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right|^{3/2} \tag{3.12}$$

Now we have the statistical mechanical foundations to write the expectation values of the residue fluctuations, $\langle \Delta X_i^2 \rangle$ and correlations, $\langle \Delta X_i \Delta X_j \rangle$. It can be verified that the $N \times N$ covariance matrix $\langle \Delta \mathbf{X} \Delta \mathbf{X}^T \rangle$ is equal to $(k_B T/\gamma)\mathbf{\Gamma}^{-1}$, using the statistical mechanical average[1]

$$\langle \Delta \mathbf{X} \Delta \mathbf{X}^T \rangle = \int \Delta \mathbf{X} \Delta \mathbf{X}^T p(\Delta \mathbf{X}) \mathrm{d}\Delta \mathbf{X} = \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \tag{3.13}$$

Because

$$\langle \Delta \mathbf{X} \Delta \mathbf{X}^T \rangle = \langle \Delta \mathbf{Y} \Delta \mathbf{Y}^T \rangle = \langle \Delta \mathbf{Z} \Delta \mathbf{Z}^T \rangle = \tfrac{1}{3} \langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle \tag{3.14}$$

we obtain

$$\langle \Delta \mathbf{R}_i^2 \rangle = \frac{3 k_B T}{\gamma} (\mathbf{\Gamma}^{-1})_{ii}$$

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{3 k_B T}{\gamma} (\mathbf{\Gamma}^{-1})_{ij} \tag{3.15}$$

as the ms fluctuations of residues and correlations between residue fluctuations. It should be noted that the assumption of *isotropic fluctuations* (Equation [3.8]) is intrinsic to the GNM. Thus the $3N$-dimensional problem (Equation [3.4]) can be reduced to an N-dimensional one described by Equation (3.15).

3.2.4 Influence of Local Packing Density

The diagonal elements of the Kirchhoff matrix, $\mathbf{\Gamma}_{ii}$, are equal to the residue coordination numbers, z_i ($1 \leq i \leq N$), which represent the degree of the EN nodes in graph theory. Thus z_i is a direct measure of *local packing density* around the ith residue. To better understand this, it is possible to express $\mathbf{\Gamma}$ as a sum of two matrices $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$, consisting exclusively of the diagonal and off-diagonal elements of $\mathbf{\Gamma}$, respectively. Using these two matrices, $\mathbf{\Gamma}^{-1}$ may

---

[1]Note that solving Equation (3.13) involves the ratio of the multidimensional Gaussian counterparts for the two integrals $\int \exp\{-ax^2\}\mathrm{d}x = \frac{1}{2}\sqrt{(\pi/a)}$ and $\int x^2 \exp\{-ax^2\}\mathrm{d}x = (\sqrt{\pi}/4)a^{3/2}$ in the range $(0, \infty)$ such that $\langle x^2 \rangle = (\sqrt{\pi}/4)a^{-3/2}/\frac{1}{2}\sqrt{(\pi/a)} = 1/2a$. For the simplest case of a single spring, subject to harmonic potential $\frac{1}{2}\gamma x^2$, $a = \gamma/2k_B T$, and $\langle x^2 \rangle = k_B T/\gamma$.

be written as

$$\Gamma^{-1} = [\Gamma_1 + \Gamma_2]^{-1} = [\Gamma_1(E + \Gamma_1^{-1}\Gamma_2)]^{-1} = (E + \Gamma_1^{-1}\Gamma_2)^{-1}\Gamma_1^{-1}$$

$$= (E - \Gamma_1^{-1}\Gamma_2 + \cdots)\Gamma_1^{-1} = \Gamma_1^{-1} - \Gamma_1^{-1}\Gamma_2\Gamma_1^{-1} + \cdots \qquad (3.16)$$

if one assumes that the invariants of the product $(\Gamma_1^{-1}\Gamma_2)$ are small compared to those of the identity matrix, E, which is a valid approximation for protein Kirchhoff matrices. Thus, the information concerning local packing density and distribution of contacts is dominated by the diagonal matrix, Γ_1^{-1}, which is the leading term in a series expansion for Γ^{-1} in Equation (3.16). Consequently, application of Equation (3.15) indicates that $\langle(\Delta R_i)^2\rangle$ scales with $[\Gamma_1^{-1}]_{ii} = 1/z_i$, to a first approximation. Thus the *local packing density* as described by the coordination numbers is an important structural property contributing to the ms fluctuations of residues [41]. However, these coordination numbers represent only the leading order and not the entire set of dynamics described by Equation (3.15).


3.3 Method and Applications

3.3.1 Equilibrium Fluctuations

The ms fluctuations of residues are experimentally measurable (e.g., x-ray crystallographic *B*-factors, or root mean-square [rms] differences between different models from NMR), and as such, have often been used as an initial test for verifying and improving computational models and methods. Beginning with the original GNM paper [3], several applications have demonstrated that the fluctuations predicted by the GNM are in good agreement with experimental *B*-factors [6, 23, 39, 42–44]. The *B*-factors are related to the expected residue fluctuations and calculated according to

$$B_i = \frac{8\pi^2}{3}\langle(\Delta R_i)^2\rangle = \frac{8\pi^2 k_B T}{\gamma}[\Gamma^{-1}]_{ii} \qquad (3.17)$$

Figure 3.2(a) illustrates the agreement between the *B*-factors predicted by the GNM (solid curve) and those calculated from experimental data (open circles) for an example protein, ribonuclease T1 (RNase T1), where Γ has been constructed from the C^α coordinates for RNase T1 deposited in the Protein Data Base (PDB) [45]. Panel B compares the rms fluctuations predicted by the GNM and those observed across the 20 NMR models deposited in the PDB for reduced disulphide-bond formation facilitator (DsbA) [46]. The correlation coefficient between GNM results and experimental data for these two example proteins are 0.769 and 0.823 in the respective panels A and B. An extensive comparison of experimental and theoretical GNM *B*-factors for a series of PDB
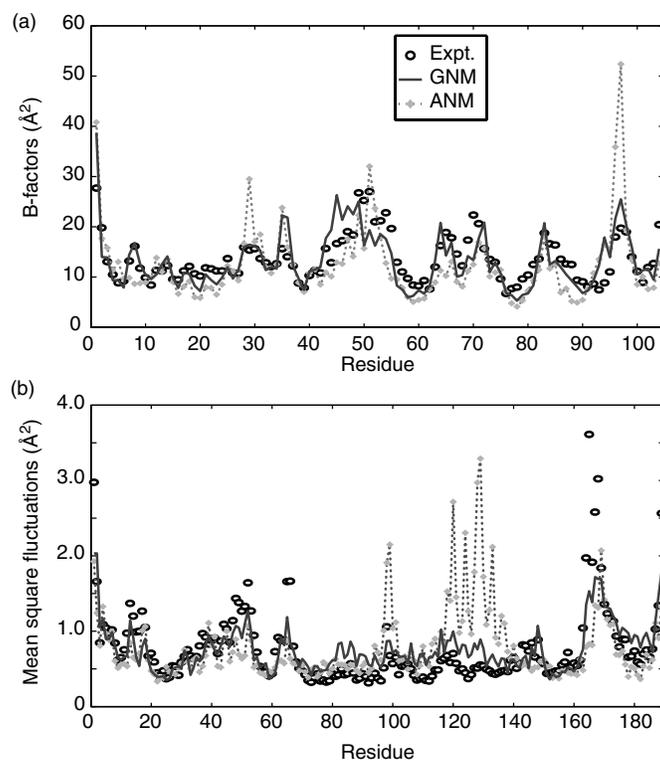
FIGURE 3.2
Comparison of ms fluctuations predicted by GNM and ANM with experimental observations. (a) Experimental x-ray crystallographic *B*-factors (open circles) reported for ribonuclease T1 (PDB file 1bu4 [53]) plotted with calculated values from GNM (solid line) and ANM (dotted line) against residue number. (b) Root mean square deviation between C^α coordinates of NMR model structures (open circles) deposited for the reduced disulphide-bond formation facilitator (DsbA) in the PDB file 1a24 [46].

structures by Phillips and coworkers has shown that GNM calculations yield an average correlation coefficient of about 0.65 with experimental *B*-factors provided that the contacts between neighboring molecules in the crystal form are taken into account. The agreement with NMR data is also remarkable, pointing to the consistency between the fluctuations undergone in solution and those inferred from x-ray structures.

3.3.2 GNM Mode Decomposition: Physical Meaning of Slow and Fast Modes

A major utility of the GNM is the ease of obtaining the collective modes of motion accessible to structures in native state conditions. The GNM normal modes are found by transforming the Kirchhoff matrix into a product of

three matrices, the matrix $\mathbf{U}$ of the eigenvectors $\mathbf{u}_i$ of $\boldsymbol{\Gamma}$, the diagonal matrix $\boldsymbol{\Lambda}$ of eigenvalues λ_i, and the transpose $\mathbf{U}^{\mathrm{T}} = \mathbf{U}^{-1}$ of the unitary matrix $\mathbf{U}$ as in Equation (3.18).

$$\boldsymbol{\Gamma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}} \tag{3.18}$$

The eigenvalues are representative of the *frequencies* of the individual modes, while the eigenvectors define the *shapes* of the modes. The first eigenvalue, λ_1, is identically zero with the corresponding eigenvector comprised of elements all equal to a constant, $1/\sqrt{N}$, indicative of an absence of internal motions in this *zero* mode. The vanishing frequency reflects the fact that the molecule can be translated rigidly without any potential energy change.

Combining Equations (3.15) and (3.18), the cross-correlations between residue fluctuations can be written as a sum over the $N - 1$ nonzero modes ($2 \leq k \leq N$) using

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{3k_{\mathrm{B}}T}{\gamma}[\boldsymbol{\Gamma}^{-1}]_{ij} = \frac{3k_{\mathrm{B}}T}{\gamma}[\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^{\mathrm{T}}]_{ij}$$

$$= \frac{3k_{\mathrm{B}}T}{\gamma}\sum_k [\lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^{\mathrm{T}}]_{ij} \tag{3.19}$$

This permits us to identify the correlation, $[\Delta\mathbf{R} \cdot \Delta\mathbf{R}]_k$ contributed by the kth mode as

$$[\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j]_k = \frac{3k_{\mathrm{B}}T}{\gamma}\lambda_k^{-1}[\mathbf{u}_k]_i[\mathbf{u}_k]_j \tag{3.20}$$

where $[\mathbf{u}_k]_i$ is the ith element of $\mathbf{u}_k$. Because $\mathbf{u}_k$ is normalized, the plot of $[\mathbf{u}_k]_i^2$ against the residue index, i, yields the normalized distribution of ms fluctuations of residues in the kth mode, shortly referred to as the kth *mode shape* (Figure 3.3[a]). Because the residue fluctuations are related to the experimental temperature (B-factors) by Equation (3.17), these elements of $\mathbf{u}_k$ reflect the residue mobilities in the kth mode.

Note that the factor λ_k^{-1} plays the role of a statistical weight, which suitably rescales the contribution of mode k. This ensures that the slowest mode has the largest contribution. In addition to their significant contribution, the slowest motions are in general also those having the highest *degree of collectivity*. Many studies have shown that the shapes of the slowest modes indeed reveal the mechanisms of *cooperative* or *global* motions, and the most constrained residues (minima) in these modes play a critical role, such as a hinge-bending center, that govern the correlated movements of entire domains [10–13, 17, 19, 44, 47–52]. It is important to note that although these motions are slow, they involve substantial conformational changes distributed over several residues. The fastest modes, on the other hand, involve the most tightly packed and hence most severely constrained residues in the molecule. Their high frequency does not imply a definitive conformational change, because they cannot effectively relax within their severely
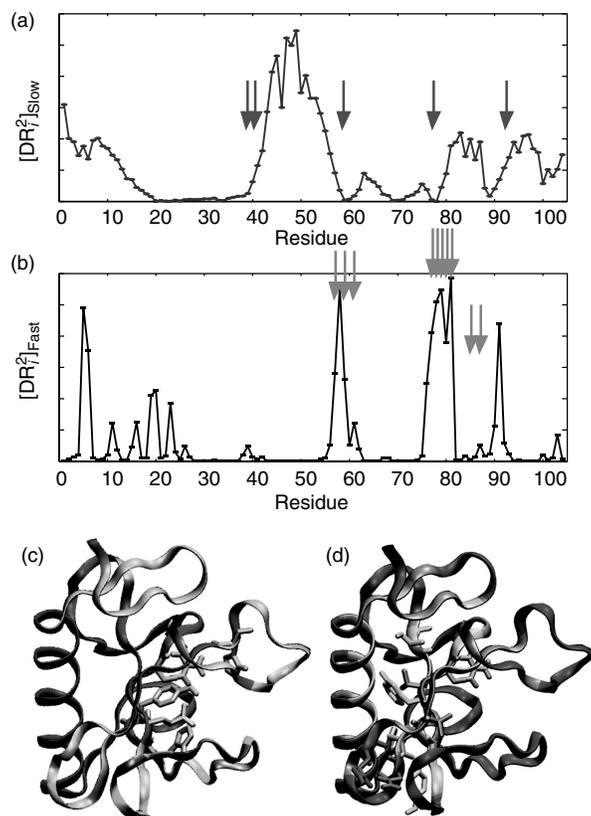
FIGURE 3.3
(See color insert following page 136) Physical meaning of slow and fast modes in GNM. (a) Distribution of squared displacements of residues in the slowest mode as a function of residue index for ribonuclease T1 (RNase T1). The red arrows identify local minima that correspond to five experimentally identified catalytic residues: Tyr38, His40, Glu58, Arg77, and His92. (b) Distribution of squared displacements averaged over the ten fastest modes for the same protein. Here the arrows indicate the residues shown by hydrogen/deuterium exchange to be the most protected and thus important for reliable folding. A majority of these critical folding residues appear as peaks in the fast modes. (c) Color-coded mapping of the slowest mode (a) onto the 3D C^{α} trace of RNase T1 (PDB file: 1bu4 [53]) where red is most mobile and blue least mobile. The side chains of the five catalytic residues are shown in pink surrounding the nucleotide binding cavity. (d) A similar color-coded mapping of the fluctuations of the ten fastest modes (b) onto the C^{α} trace. Here the side chains of the ten most protected residues from hydrogen deuterium exchange experiments are drawn explicitly showing that most of them are calculated to be mobile (red). The images in c and d were generated using VMD [74].

constrained environment. On the contrary, they enjoy extremely small conformational freedom, on a *local* scale, by undergoing fast, but small amplitude fluctuations.

Figure 3.3 illustrates the contrast between the degree of collectivity for the slowest and fastest modes for an example protein, RNase T1. As in this case, the slow modes involve almost the entire molecule as indicated by the

broad, delocalized peaks in panel A. The relative potential motion predicted by this mode is plotted onto the 3D structure in Figure 3.3(c), color-coded such that minima are blue and maxima are red. For RNaseT1, five red arrows are drawn in Figure 3.3(a) to indicate the residues identified as part of the catalytic site (Y38, H40, E58, R77, and H92) [53]. With the exception of H92, these five residues are located near minima in the slow (global) mode shape (Figure 3.3[a]) and their side chains are shown to be spatial neighbors (pink tubes) in the 3D plot of this protein (Figure 3.3[c]).

In contrast, the fastest modes are highly localized, with mode shapes that usually involve only a few peaks, as in Figure 3.3(b). These peaks refer to the residues that have a high concentration of local energy and are tightly constrained in motion. It has been noticed that these residues are often conserved across species and may form the folding nuclei [33, 34, 54]. In the application to RNase T1, the ten most protected residues (57,59,61,77–81,85, and 87), as identified by hydrogen–deuterium exchange experiments [55], are indicated by gold arrows in Figure 3.3(b) and shown with their side chains in the 3D structure, color-coded such that minima are blue and maxima are red (Figure 3.3[d]). As illustrated, many of these residues involve interactions between different strands of the central β-sheet, suggesting their potential involvement in the folding of RNase T1.

3.3.3 What Is ANM? How Does GNM Differ from ANM?

As pointed out in Chapter 1 by Hinsen, ANM analysis is simply an NMA applied to an EN model, the potential of which is defined as [6]

$$V_{\text{ANM}} = \frac{\gamma}{2}\left[\sum_{i,j}^{N}(R_{ij} - R_{ij}^{0})^{2}H(r_{\text{c}} - R_{ij})\right] \tag{3.21}$$

where $H(r_{\text{c}} - R_{ij})$ is the heavyside step function equal to 1 if the argument is positive, and zero otherwise. $H(r_{\text{c}} - R_{ij})$ selects all residue pairs within the cutoff separation of r_{c}. In the GNM, on the other hand, the potential is given by

$$V_{\text{GNM}} = \frac{\gamma}{2}\left[\sum_{i,j}^{N}(\mathbf{R}_{ij} - \mathbf{R}_{ij}^{0})^{2}H(r_{\text{c}} - R_{ij})\right] \tag{3.22}$$

Equation (3.22) looks very similar to Equation (3.21), with the major difference that the vectors $\mathbf{R}_{ij}$ and $\mathbf{R}_{ij}^{0}$ in Equation (3.22) replace distances (scalars), R_{ij} and R_{ij}^{0}. This means that the potential, which depended upon the dot product between the fluctuation vectors in the GNM

$$(\mathbf{R}_{ij} - \mathbf{R}_{ij}^{0}) \cdot (\mathbf{R}_{ij} - \mathbf{R}_{ij}^{0}) = R_{ij}^{2} + (R_{ij}^{0})^{2} - 2R_{ij}R_{ij}^{0}\cos(\mathbf{R}_{ij}, \mathbf{R}_{ij}^{0})$$
$$= R_{ij}^{2} + (R_{ij}^{0})^{2} - 2(X_{ij}X_{ij}^{0} + Y_{ij}Y_{ij}^{0} + Z_{ij}Z_{ij}^{0}) \tag{3.23}$$

now (in the ANM) depends upon their scalar product

$$
\begin{aligned}
(R_{ij} - R_{ij}^0)(R_{ij} - R_{ij}^0) &= R_{ij}^2 + (R_{ij}^0)^2 - 2R_{ij}R_{ij}^0 \\
&= R_{ij}^2 + (R_{ij}^0)^2 - 2[X_{ij}^2 + Y_{ij}^2 + Z_{ij}^2]^{1/2} \\
&\quad \times [(X_{ij}^0)^2 + (Y_{ij}^0)^2 + (Z_{ij}^0)^2]^{1/2}
\end{aligned}
\tag{3.24}
$$

Because the scalars R_{ij} and R_{ij}^0 depend upon their components in a nonquadratic form, it is natural to end up with anisotropic fluctuations upon taking the second derivatives of the potential with respect to the displacements along the X-, Y-, and Z-axes as is done in NMA. Using Equations (3.23) and (3.24), the difference between these two potentials is

$$
V_{\mathrm{GNM}} - V_{\mathrm{ANM}} = \gamma \left[\sum_{i,j}^{N} R_{ij}R_{ij}^0 (1 - \cos(\mathbf{R}_{ij}, \mathbf{R}_{ij}^0)) H(r_{\mathrm{c}} - R_{ij}) \right]
\tag{3.25}
$$

that is, the two potentials are equal only if $\cos(\mathbf{R}_{ij}, \mathbf{R}_{ij}^0) = 1$, that is, $\mathbf{R}_{ij} = \mathbf{R}_{ij}^0$ or $\Delta\mathbf{R}_i = \Delta\mathbf{R}_j$.

Physically, this means that in addition to changes in inter-residue distances (springs), any change in the *direction* of the inter-residue vector $\mathbf{R}_{ij}^0$ is also being resisted or penalized in the GNM potential. On the contrary, the ANM potential depends exclusively on the *magnitudes* of the inter-residue distances and does not penalize any such changes in orientation. It is conceivable that within the densely packed environment of proteins, orientational deformations may be as important as translational (distance) ones, and a potential that takes account of the energy dependence associated with the internal orientational changes (i.e., V_{GNM}) is physically more meaningful than one exclusively based on distances (V_{ANM}). Not surprisingly, ANM has been observed to give rise to excessively high fluctuations compared to the GNM results or experimental data (Figure 3.2), and hence necessitated the adoption of a higher cutoff distance for interactions [6]. With a higher cutoff distance, each residue is connected to more neighbors in a more constrained and consolidated network.

Inasmuch as V_{GNM} is physically more realistic, one might prefer to adopt the GNM, rather than the ANM for a coarse-grained NMA. However, this greater realism comes at a price. Because the GNM describes the dynamics within an N-dimensional configurational space as opposed to a $3N$-dimensional one of ANM, the residue fluctuations predicted by the GNM are intrinsically isotropic. Thus GNM cannot provide information regarding the individual components: $\Delta X^{(k)}$, $\Delta Y^{(k)}$, and $\Delta Z^{(k)}$, of the deformation vectors $\Delta\mathbf{R}^{(k)}$ associated with each mode, k, but rather predicts the magnitudes, $|\Delta\mathbf{R}^{(k)}|$, induced by such deformations. The conclusion is that GNM is more accurate, and should be chosen when evaluating the deformation magnitudes, or the distribution of motions of individual residues. However, ANM is the only

possible (less realistic) model when it comes to assessing the directions or mechanisms of motions. That the fluctuations predicted by the GNM correlate better with experimental *B*-factors than those predicted by the ANM has been observed and confirmed in a recent systematic study of Phillips and coworkers [23]. The dotted curves in Figure 3.2 illustrate the ANM results, and provide a comparison of the level of agreement (with experimental data) usually achieved by the two respective models. The correlation coefficients between the GNM results and experimental data are 0.769 and 0.823 in the panels A and B, respectively, whereas their ANM counterparts are 0.639 and 0.261. We note that the two sets of computed results are themselves correlated (0.756 and 0.454, respectively), which can be expected from the similarity of the underlying models.

3.3.4 Applicability to Supramolecular Structures

A major advantage of the GNM is its applicability to large complexes and assemblies. The size of the Kirchhoff matrix is $N \times N$ for a structure of N residues, as opposed to the size $3N \times 3N$ of the equivalent Hessian matrix for a residue-level EN NMA (or ANM). The resulting computational time requirement for GNM analysis is then about 3^3 times less than for ANM, which in turn is about 8^3 times less than for NMA at atomic scale (assuming eight atoms on the average per residue). This enormous decrease in computational time permits us to use ANM, and certainly GNM, for efficiently exploring the dynamics of supramolecular structures [17, 22].

Due to limitations in computational memory and speed, efforts to analyze large structures of $\sim 10^5$ residues rely upon further coarse-graining of the structure of interest. This is now the standard approach, having been implemented in several forms by various research groups including hierarchical coarse-graining (HCG) [56], discussed below; rotations–translations of blocks (RTB) [57] or block normal mode (BNM) [9]; and substructure-synthesis method (SSM) [58], which are discussed in other chapters of this book.

For both GNM and ANM, it has been demonstrated that an HCG scheme where clusters of residues and their interactions, as opposed to individual pairs of residues, are considered as the EN nodes successfully reproduce the essential features of the full-residue GNM/ANM calculations [56]. The global dynamics of hemagglutinin A were obtained at least two orders of magnitude faster than standard GNM/ANM by coarse-graining to the level of every 40th residue ($N/40$) [56]. Notably, the minima in global mode shapes, which identify key regions coordinating the collective dynamics, were *exactly* reproduced by the $N/40$ coarse-graining.

Figure 3.4 illustrates the application of GNM to the wild type 70S ribosome from *Escherichia coli* [59]. The calculations were performed by considering a single node for each amino acid (on the C^α atom) and each nucleotide (on the *P* atom), yielding a total of 10,453 nodes (residues and nucleotides). Because the diameter of the *A*-form RNA double helix is 20 Å, a larger cutoff distance is
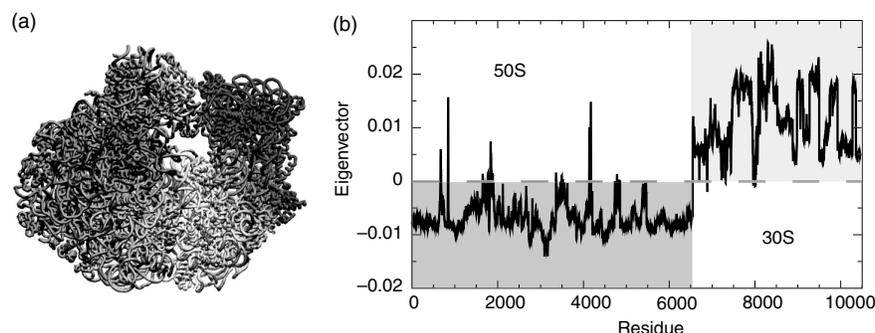
(a) (b)

FIGURE 3.4
(See color insert following page 136) Application of GNM to the 70S ribosome structure. The calculations were performed on the wild type 70S ribosome from *E. coli* (PDB files 1pnx and 1pny [59]). (a) The slowest nonzero mode for the 70S ribosome colored from −1 (red) to +1 (blue) is mapped onto the 3D structure indicating a dramatic break at the interface between the two subunits (50S and 30S). This image was generated using VMD [74]. (b) The slowest nonzero mode plotted vs. the residue number. Residues in the 50S subunit (blue) exhibit one direction of motion that is opposed to the motion in the 30S subunit (red).

required to correctly identify base-paired nucleotides solely by their *P*-atom positions [42]. To ensure adequate connectivity, two cutoff distances were adopted, 9.0 Å if both atoms were C^{α} and 21.0 Å if one or both were phosphorous, analogous to our ANM analysis of ribosome [19]. Panels a and b illustrate the slowest (nonzero) mode shape as a color-coded 3D structure and against the residue index. The coloring emphasizes the distinct difference between the motions of the 50S (red) and 30S (blue) subunits in this mode and indicates an anticorrelated motion of one subunit with respect to the other. This type of anticorrelated motion (i.e., ratcheting of one subunit with respect to the other) has been observed by cryo-EM [60].

Recently the dynamics of the HK97 bacteriophage viral capsid has been analyzed using the GNM. Two different forms of the capsid known as the pro-capsid (Prohead II) [2] and mature (Head II) [61] were considered. These structures are comprised of 420 copies of a single protein chain arranged into 12 pentamers and 60 hexamers, which expand from a spherical form (prohead) to icosahedral form (head) during maturation [62]. The GNM results obtained with a coarse-graining of $N/35$ for the first (slowest nonzero) mode of the two forms are shown in Figure 3.5(a) and Figure 3.5(b). These HCG structures have 3072 and 3360 nodes respectively. GNM computations were also performed with the complete sets of 107,520 and 117,600 residues for the respective pro-capsid and mature capsid to examine the conformational changes accompanying maturation [22]. Figure 3.5(c) and Figure 3.5(d) indicate that the slowest nonzero mode for the full capsid matches the $N/35$ results. This mode is asymmetric, yet identifies a region at each pole, pentamer-centered, as the most mobile (red) in each of these calculations.
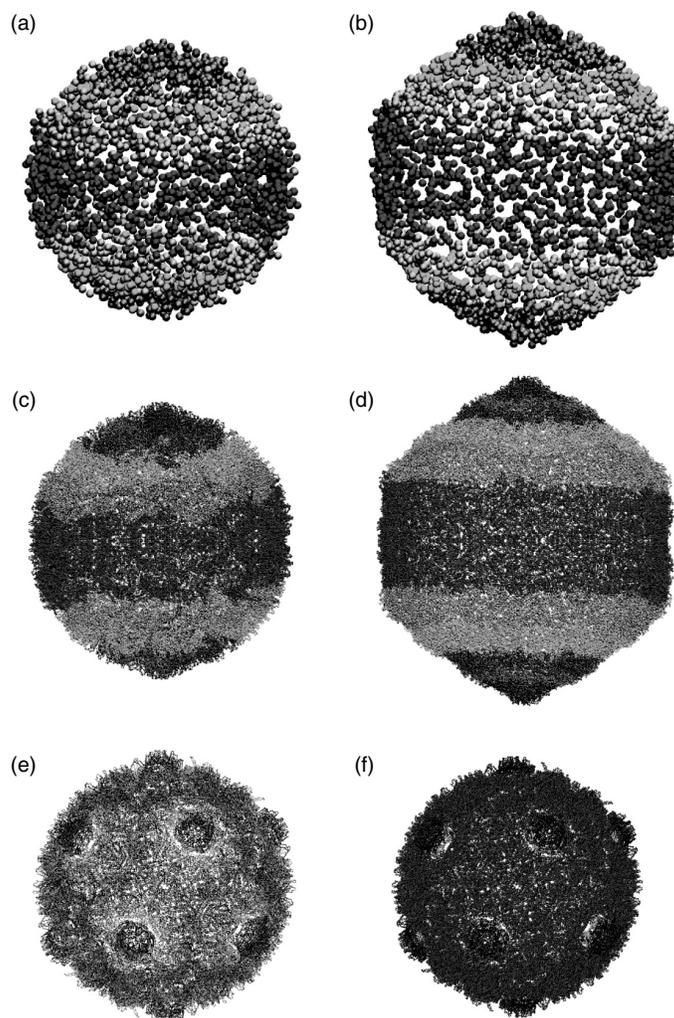
FIGURE 3.5
(See color insert following page 136) Application of GNM to the HK97 bacteriophage viral capsid. (a) The ms fluctuations from the slowest (threefold degenerate) mode for the prohead viral capsid coarse-grained by retaining only every 35th residue are colored from most mobile (red) to least mobile (blue). (b) The results for the slowest (threefold degenerate) mode of the head viral capsid calculated using a similar coarse-grained procedure of retaining every 35th residue. Both identify pentamer-centered regions at opposite poles as the most mobile regions suggesting an expansion or puckering of these residues. (c) The ms fluctuations for the slowest mode calculated over the entire (107,520 residue) prohead capsid structure (PDB file 1if0 [62]) and (d) entire (117,600 residue) head capsid structure (PDB file 1fh6 [61]) also demonstrate this high degree of mobility at the poles. (e) The weighted summation of the 11 slowest modes identifies the 12 pentamers as the most mobile regions responsible for expansion from the prohead to head form. (f) The slowest nondegenerate, symmetric mode, mode 31, also identifies these pentamers as highly mobile. These images were generated using VMD [74].

Due to the large degree of symmetry of the capsid, many of the calculated modes are degenerate, that is, have the same frequencies (eigenvalues). For example, the mode indicated in Figures 3.5(a) to Figure 3.5(d) is threefold degenerate, oriented along a different axis of the capsid. Hence a superposition of these three modes would more accurately describe the global motion of expansion that accompanies maturation. In fact, the superposition of the 11 slowest modes was found to yield icosahedrally symmetric fluctuations and identify the 12 pentamers as the most mobile (red) regions in the capsid (Figure 3.5[e]). Cryo-EM maps of intermediates between the prohead and head conformations indicate a large degree of motion for these pentamers during expansion [63]. It should be noted that the slowest nondegenerate mode, mode 31, is icosahedrally symmetric and also identifies these 12 pentamers as the most mobile regions (Figure 3.5[f]). However, because the frequency of mode 31 is at least three times that of modes 1 through 11, its contribution to the observed structural changes is small relative to that of these slower, asymmetric modes which cooperatively induce a similar set of fluctuation dynamics [22].

3.3.5 iGNM: A Database of GNM Results

With advances in computational methods for characterizing proteins, and with the recognition of the importance of modeling and understanding structural dynamics, a number of groups have recently undertaken the task of generating and making available on the Internet servers or databases for modeling or examining protein motions. One of the earliest attempts in this direction is the Database (DB) of Macromolecular Movements (MolMovDB; http://molmovdb.org/) [64] originally known as the DB of Protein Motions, constructed by Gerstein and collaborators [65]. Currently, about 4400 movies ("morphs") are available in the MolMovDB, generated by interpolating between pairs of known conformations of proteins or RNA molecules. These "morphs" are then used to classify molecules into roughly 178 motion types. Another online calculation tool based on a simplified NMA combined with the rotations–translations of blocks (RTB) algorithm [57] has been developed by Sanejouand's group with a sophisticated webserver interface (elNémo; http://igs-server.cnrs-mrs.fr/elnemo/), which includes up to 100 slowest modes for each studied structure [66]. A similar, yet more extensive work has been conducted in the lab of Wako [67] where the normal modes in the space of dihedral angles have been generated and collected in the ProMode DB (http://cube.socs.waseda.ac.jp/pages/jsp/index.jsp), for nearly 1442 single chain proteins extracted from the PDB. ProMode has been restricted to relatively small proteins (<300 residues) due to the time cost of energy minimization performed prior to NMA. In the same spirit, we have generated a DB of GNM results, iGNM (http://ignm.ccbb.pitt.edu/) [68]. Two major advantages of this DB are (i) motions are predicted, rather than interpolated between two structures, implying that knowledge of a single structure is sufficient to

generate the equilibrium fluctuations and (ii) the completeness of structure-dynamics data generated, since almost all of the existing PDB structures (22,549 PDB files as of September 15, 2003) have been analyzed. One should note that we are still in the process of cleaning and refining the resultant output files as well as augmenting the iGNM DB with newly deposited structures.

The iGNM DB provides information on protein dynamics beyond those experimentally provided by *B*-factors (for x-ray structures) or rms fluctuations (NMR structures). We have developed an Internet-based query system to retrieve information on *B*-factors, GNM mode shapes (eigenvectors corresponding to both the slowest and fastest ends of the vibrational spectrum), ribbon diagrams color-coded according to the residue mobilities, and cross-correlation maps describing the strength and types of correlations between residue fluctuations. The retrieved data are presented using a Chime plug-in (for 3D visualization of molecular structures) and Java applets (for graphical two-dimensional [2D] plots of fluctuation distributions). In addition to retrieving the data stored in the DB, the user has the ability to compute GNM dynamics for newly deposited structures as well as user-generated structures (e.g., from comparative modeling) through an automated online calculation server. The raw data for modes, frequencies, and residue correlations can be downloaded from the iGNM DB. Figure 3.6 illustrates the different types of outputs accessible from the iGNM DB, for phospholipase A2, as an example. For visualization, the residue mobility in the 20 slowest and fastest modes for each protein is mapped onto a ribbon diagram of the 3D structure (Figure 3.6[a]) color-coded from blue (least mobile) to red (most mobile). For slow modes, the blue portion of the molecule indicates the least mobile regions (minima), which are usually relevant to biological function (see Section 3.3.2). Similarly, by selecting the fast modes from the menu, the users are able to view the residues predicted to be critical for folding (maxima) colored red. Figure 3.6(b) illustrates a 2D plot of this same information against residue number for a given mode. Using the Java-applet windows (Figure 3.6[b] insets) the user is able to identify residues of interest, and scale the curves in order to compare multiple modes or different proteins. Figure 3.6(c) demonstrates the case where theoretical (red) and experimental (yellow) *B*-factors are plotted against the residue index simultaneously. The correlation map between the motions of residue pairs is illustrated in Figure 3.6(d). The colors on the map reflect the correlation cosines between the fluctuations of residue pairs — dark blue for anticorrelated (coupled but opposite direction), red for correlated (coupled, same direction), and green for uncorrelated motions.

3.4 Future Prospects

As discussed earlier, EN models in general have proven effective in characterizing large-scale motions in biomolecules. The simplicity of these models
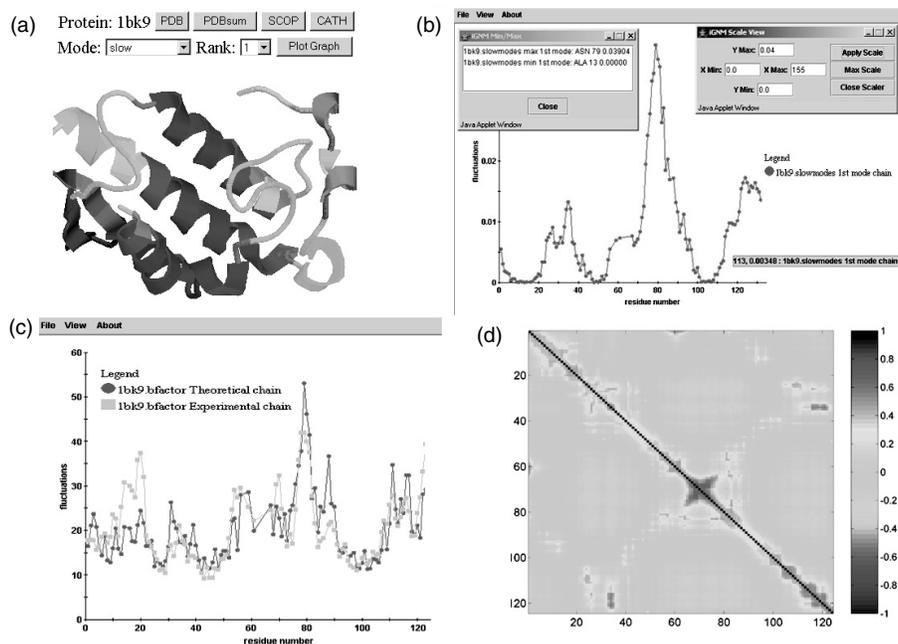
FIGURE 3.6

(See color insert following page 136) Visualization of dynamics data stored in the iGNM DB. The results shown are of the example phospholipase A2 (PDB:1bk9 [75]). (a) The slowest motions (slow1) are color-coded and mapped on structure of 1bk9 in dark blue, green, orange to red in the increasing order of mobility. (b) The window shows the mobility of the slowest mode with scalable range of view, max/min value info window and pop-up tag that shows the residue number and coordinates. (c) The experimental and predicted B factors are compared. (d) The cross-correlation of residue pairs of all modes. The perfect concerted motion (+1) is colored dark red while the perfect anticorrelated motion (−1) is colored dark blue.

sacrifice atomic-level predictions, in favor of performing efficient calculations of large structures, which capture the essential dynamics and functional motions encoded in the native structure. Such models are necessary to infer dynamics and function in an age where structural information about larger and larger complexes outpaces the computational ability to apply traditional atomic-level simulations.

There are at least five major areas in which GNM, or EN models in general, are anticipated to be useful. These include (i) the applications to low resolution structural data (e.g., cryo-EM), both for structure refinement and dynamics assessment [69–71]; (ii) exploring the machinery of supramolecular structures or multi-molecular assemblies including protein–DNA, protein–protein, and membrane protein–lipid complexes [19]; (iii) high throughput examination of the collective dynamics of families of proteins toward extracting common dynamic patterns and design principles selected and conserved across genomes or within subfamilies for functional requirements; (iv) investigation

of unfolding pathways and kinetics by including a temperature dependence for contacts [31, 36, 72]; and (v) integration with more detailed models and simulations [32, 73] (e.g., MD) for efficiently exploring the physical and chemical transitions between different functional forms further discussed in the chapter by Liu and coworkers.

References

1. Go, N., Noguti, T., and Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes, *Proc. Natl Acad. Sci. USA,* 80, 3696, 1983.

2. Brooks, B. and Karplus, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor, *Proc. Natl Acad. Sci. USA,* 80, 6571, 1983.

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

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

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

6. Atilgan, A.R. et al. Anisotropy of fluctuation dynamics of proteins with an elastic network model, *Biophys. J.,* 80, 505, 2001.

7. Hinsen, K. and Kneller, G.R. A simplified force field for describing vibrational protein dynamics over the whole frequency range, *J. Chem. Phys.,* 111, 10766, 1999.

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

9. Li, G.H. and Cui, Q. A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca2+-ATPase, *Bipohys. J.,* 83, 2457, 2002.

10. Bahar, I. et al. Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function, *J. Mol. Biol.,* 285, 1023, 1999.

11. Temiz, N.A. and Bahar, I. Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase, *Proteins,* 49, 61, 2002.

12. Isin, B., Doruker, P., and Bahar, I. Functional motions of influenza virus hemagglutinin: a structure-based analytical approach, *Biophys. J.,* 82, 569, 2002.

13. Thomas, A. et al. Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study, *Proteins,* 34, 96, 1999.

14. Cui, Q. et al. A normal mode analysis of structural plasticity in the biomolecular motor F-1-ATPase, *J. Mol. Biol.,* 340, 345, 2004.

15. Van Wynsberghe, A., Li, G., and Cui, Q. Normal-mode analysis suggests protein flexibility modulation throughout RNA polymerase's functional cycle, *Biochemistry,* 43, 13083, 2004.

16. Ming, D.M. et al. Simulation of F-actin filaments of several microns, *Biophys. J.,* 85, 27, 2003.

17. Keskin, O. et al. Molecular mechanisms of chaperonin groel–groes function, *Biochemistry,* 41, 491, 2002.

18. Tama, F. et al. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy, *Proc. Natl Acad. Sci. USA,* 100, 9319, 2003.

19. Wang, Y. et al. Global ribosome motions revealed with elastic network model, *J. Struct. Biol.,* 147, 302, 2004.

20. Tama, F. and Brooks III, C.L. The mechanism and pathway of ph induced swelling in Cowpea Chlorotic mottle virus, *J. Mol. Biol.,* 318, 733, 2002.

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

22. Rader, A.J., Vlad, D.H., and Bahar, I. Maturation dynamics of bacteriophage HK97 capsid, *Structure,* 13, 413, 2005.

23. Kundu, S. et al. Dynamics of proteins in crystals: comparison of experiment with simple models, *Biophys. J.,* 83, 723, 2002.

24. Burioni, R. et al. Topological thermal instability and length of proteins, *Proteins,* 55, 529, 2004.

25. Wu, Y. et al. Universal behavior of localization of residue fluctuations in globular proteins, *Phys. Rev. E.,* 67, 041909, 2003.

26. Kundu, S., Sorensen, D.C., and Phillips Jr., G.N. Automatic domain decomposition of proteins by a Gaussian network model, *Proteins: Struct. Funct. Bioinf.* 57, 725, 2004.

27. Kundu, S. and Jernigan, R.L. Molecular mechanism of domain swapping in proteins: an analysis of slower motions, *Bipohys. J.,* 86, 3846, 2004.

28. Bahar, I. et al. Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model, *Biochemistry,* 37, 1067, 1998.

29. Jaravine, V.A., Rathgeb-Szabo, K., and Alexandrescu, A.T. Microscopic stability of cold shock protein examined by NMR native state hydrogen exchange as a function of urea and trimethylamine n-oxide, *Protein Sci.,* 9, 290, 2000.

30. Ortiz, A.R. and Skolnick, J. Sequence evolution and the mechanism of protein folding, *Bipohys. J.,* 79, 1787, 2000.

31. Micheletti, C., Lattanzi, G., and Maritan, A. Elastic properties of proteins: insight on the folding process and evolutionary selection of native structures, *J. Mol. Biol.,* 321, 909, 2002.

32. Micheletti, C., Carloni, P., and Maritan, A. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models, *Proteins,* 55, 635, 2004.

33. Rader, A.J. and Bahar, I. Folding core predictions from network models of proteins, *Polymer,* 45, 659, 2004.

34. Rader, A.J. et al. Identification of core amino acids stabilizing rhodopsin, *Proc. Natl Acad. Sci. USA,* 101, 7246, 2004.

35. Chen, S.C. and Bahar, I. Mining frequent patterns in protein structures: a study of protease families, *Bioinformatics,* 20, i77–i85, 2004.

36. Micheletti, C. et al. Crucial stages of protein folding through a solvable model: predicting target sites for enzyme-inhibiting drugs, *Protein Sci.,* 11, 1878, 2002.

37. Flory, P.J. Statistical thermodynamics of random networks, *Proc. R. Soc. Lond. A,* 351, 351, 1976.

38. Miyazawa, S. and Jernigan, R.L. Estimation of effective inter-residue contact energies from protein crystal structures: quasi-chemical approximation, *Macromolecules,* 18, 534, 1985.

39. Bahar, I. and Jernigan, R.L. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation, *J. Mol. Biol.,* 266, 195, 1997.
40. Papoulis, A. Sequences of random variables, in *Probability, Random Variables, and Stochastic Processes,* Kogakusha Company, Tokyo, 1965, p. 233
41. Halle, B. Flexibility and packing in proteins, *Proc. Natl Acad. Sci. USA,* 99, 1274, 2002.
42. Bahar, I. and Jernigan, R.L. Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms, *J. Mol. Biol.,* 281, 871, 1998.
43. Bahar, I. Dynamics of proteins and biomolecular complexes: inferring functional motions from structure, *Rev. Chem. Eng.,* 15, 319, 1999.
44. Keskin, O., Jernigan, R.L., and Bahar, I. Proteins with similar architecture exhibit similar large-scale dynamic behavior, *Biophys. J.,* 78, 2093, 2000.
45. Berman, H.M. et al. The protein data bank, *Nucleic Acids Res.,* 28, 235, 2000.
46. Wilson, K.P., Malcolm, B.A., and Matthews, B.W. Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme, *J. Biol. Chem.,* 267, 10842, 1992.
47. Bahar, I. et al. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability, *Phys. Rev. Lett.,* 80, 2733, 1998.
48. Jernigan, R.L., Demirel, M.C., and Bahar, I. Relating structure to function through the dominant slow modes of motion of DNA topoisomerase II, *Int. J. Quant. Chem.,* 75, 301, 1999.
49. Jernigan, R.L. et al. Relating the structure of HIV-1 reverse transcriptase to its processing step, *J. Biomol. Struct. Dyn.,* 11, 49, 2000.
50. Keskin, O. et al. Relating molecular flexibility to function: a case study of tubulin, *Biophys. J.,* 83, 663, 2002.
51. Xu, C., Tobi, D., and Bahar, I. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T<-> R2 transition, *J. Mol. Biol.,* 333, 153, 2003.
52. Sluis-Cremer, N., Temiz, N.A., and Bahar, I. Conformational changes in HIV-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding, *Curr. HIV Res.,* 2, 323, 2004.
53. Loris, R. et al. Conserved water molecules in a large family of microbial ribonucleases, *Proteins,* 36, 117, 1999.
54. Demirel, M.C. et al. Identification of kinetically hot residues in proteins, *Protein Sci.,* 7, 2522, 1998.
55. Mullins, L.S., Pace, C.N., and Raushel, F.M. Conformational stability of ribonuclease T1 determined by hydrogen–deuterium exchange, *Protein Sci.,* 6, 1387, 1997.
56. Doruker, P., Jernigan, R.L., and Bahar, I. Dynamics of large proteins through hierarchical levels of coarse-grained structures, *J. Comput. Chem.,* 23, 119, 2002.
57. Tama, F. et al. Building-block approach for determining low-frequency normal modes of macromolecules, *Proteins,* 41, 1, 2000.
58. Ming, D. et al. Substructure synthesis method for simulating large molecular complexes, *Proc. Natl Acad. Sci. USA,* 100, 104, 2003.
59. Vila-Sanjurjo, A. et al. X-ray crystal structures of the wt and a hyper-accurate ribosome from *Escherichia coli*, *Proc. Natl Acad. Sci. USA,* 100, 8682, 2003.
60. Frank, J. and Agrawal, R.K. A ratchet-like inter-subunit reorganization of the ribosome during translocation, *Nature,* 406, 318, 2000.

61. Wikoff, W.R. et al. Topologically linked protein rings in the bacteriophage HK97 capsid, *Science,* 289, 2129, 2000.
62. Conway, J.F. et al. Virus maturation involving large subunit rotations and local refolding, *Science,* 292, 744, 2001.
63. Lata, R. et al. Maturation dynamics of a viral capsid: visualization of transitional intermediate states, *Cell,* 100, 253, 2000.
64. Echols, N., Milburn, D., and Gerstein, M. MolMovDB: analysis and visualization of conformational change and structural flexibility, *Nucleic Acids Res.,* 31, 478, 2003.
65. Gerstein, M. and Krebs, W. A database of macromolecular motions, *Nucleic Acids Res.,* 26, 4280, 1998.
66. Suhre, K. and Sanejouand, Y.H. elNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement, *Nucleic Acids Res.,* 32, W610–W614, 2004.
67. Wako, H., Kato, M., and Endo, S. ProMode: a database of normal mode analyses on protein molecules with a full-atom model, *Bioinformatics,* 20, 2035, 2004.
68. Yang L.W., Lin, X., Jursa, C.J., Holliman, M., Rader, A.J., Karimi, H., and Bahar, I. iGNM: A Database of Protein Functional Motions Based on Gaussian Network Model, *Bioinformatics,* 21, 2978–2987, 2005.
69. Ming, D. et al. How to describe protein motion without amino acid sequence and atomic coordinates, *Proc. Natl Acad. Sci. USA,* 99, 8620, 2002.
70. Tama, F., Wriggers, W., and Brooks III, C.L. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory, *J. Mol. Biol.,* 321, 297, 2002.
71. Delarue, M. and Dumas, P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models, *Proc. Natl Acad. Sci. USA,* 101, 6957, 2004.
72. Portman, J.J., Takada, S., and Wolynes, P.G. Microscopic theory of protein folding rates. I. Fine structure of the free energy profile and folding routes from a variational approach, *J. Chem. Phys.,* 114, 5069, 2001.
73. Zhang, Z.Y., Shi, Y.Y., and Liu, H.Y. Molecular dynamics simulations of peptides and proteins with amplified collective motions, *Biophys. J.,* 84, 3583, 2003.
74. Humphrey, W., Dalke, A., and Schulten, K. VMD: visual molecular dynamics, *J. Mol. Graph.,* 14, 33, 1996.
75. Zhao, H. et al. Structure of a snake venom phospholipase A2 modified by p-bromo-phenacyl-bromide, *Toxicon,* 36, 875, 1998.