Normal mode theory and harmonic potential approximations

Konrad Hinsen

Normal Mode Theory and Harmonic Potential Approximations

Konrad Hinsen

1.1 Introduction

Normal mode analysis (NMA) has become one of the standard techniques in the study of the dynamics of biological macromolecules. It is primarily used for identifying and characterizing the slowest motions in a macromolecular system, which is inaccessible by other methods. This chapter explains what normal mode analysis is and what one can do with it without going beyond its limit of validity. The focus of this chapter is on proteins, although normal mode analysis can equally well be applied to other macromolecules (e.g., DNA) and to macromolecular assemblies ranging in size from protein–ligand complexes to a whole ribosome.

By definition, normal mode analysis is the study of harmonic potential wells by analytic means. Section 1.2 of this chapter will therefore deal with potential wells and harmonic approximations. Section 1.3 is about normal mode approaches to different physical situations, and Section 1.4 discusses how useful information can be extracted from normal modes.

1.2 Potential Wells

The fundamental restriction of normal mode analysis is its limitation to the study of dynamics in a single potential well. More specifically, normal mode analysis studies motions of small amplitude in a potential well, where "small" means "small enough that the approximations hold." What exactly that means in practice will be discussed later in this section. An immediate consequence is that normal mode analysis is not well suited to the study of conformational transitions, although it can play a complementary role to other techniques in such applications.

The starting point for normal mode analysis is one particular stable conformation of the system that represents a minimum of the potential energy surface. One then constructs a *harmonic approximation* of the potential well around this conformation. This step involves the central approximation of the method, which therefore deserves a more detailed discussion.

A harmonic potential well has the form[1]

$$U(\mathbf{r}) = \tfrac{1}{2}(\mathbf{r} - \mathbf{R}) \cdot \mathbf{K}(\mathbf{R}) \cdot (\mathbf{r} - \mathbf{R}) \tag{1.1}$$

where $\mathbf{R}$ is a $3N$-dimensional vector (N is the number of atoms) describing the stable conformation at the center of the well and $\mathbf{r}$ is an equally $3N$-dimensional vector representing the current conformation. The symmetric and positive semidefinite matrix $\mathbf{K}$ describes the shape of the potential well. A harmonic model for a potential well thus consists of $\mathbf{R}$ and $\mathbf{K}$.

Before we can describe the options for constructing a harmonic approximation, we have to review the properties of potential energy landscapes of proteins. First and foremost, the potential energy landscape of a protein has a multiscale structure (see Figure 1.1). On the length scale on which one typically considers conformations from a structural point of view (0.1 to 10 nm), a stable conformation corresponds to a local minimum of a smooth, slowly varying potential. If several local minima exist, they describe different stable conformations, and are separated by local maxima and saddle points. Looking closer (0.001 to 0.1 nm), one sees that the potential well is not smooth, but has many local minima and energy barriers of smaller height. These are referred to as conformational substates [1–3]. The differences between neighboring conformational substates are, for example, different arrangements of sidechains, whereas a different conformation would imply more important geometrical changes involving the backbone.

By far the most frequently applied method to construct a harmonic potential model consists of starting from an all- or united-atom potential $V(\mathbf{r})$ and an

[1] We limit ourselves to harmonic potentials in Cartesian coordinates. Other coordinates can be used as well, but are less convenient for numerical applications. Note that a potential that is harmonic in one coordinate set is in general *not* harmonic in other coordinates.

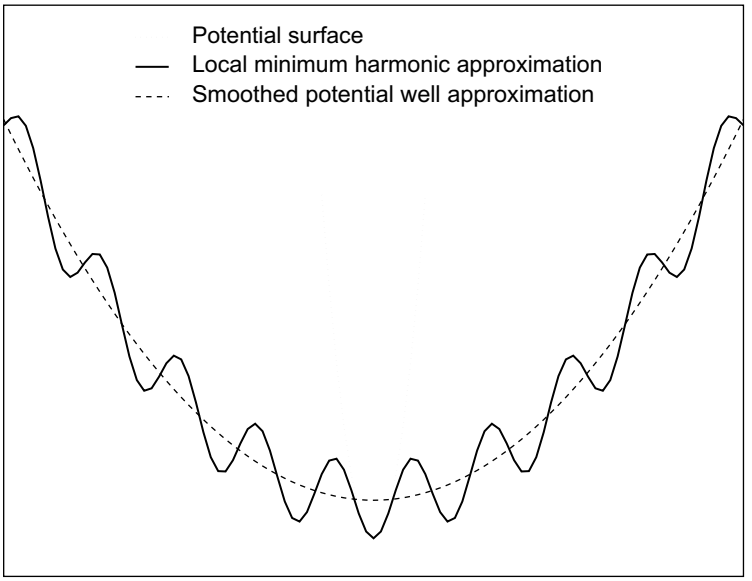

FIGURE 1.1
A schematic one-dimensional view of the potential energy surface of a protein showing two kinds of harmonic approximations: an approximation to a local minimum, and an approximation to the smoothed-out potential well.

experimentally or otherwise obtained initial conformation. An energy minimization algorithm is then applied to find a local minimum $\mathbf{R}_{\min}$ near the initial structure. Finally, the matrix $\mathbf{K}$ is obtained as the second derivative of the potential:

$$\mathbf{K}_{ij} = \left[\frac{\partial^2 U}{\partial \mathbf{r}_i \partial \mathbf{r}_j} \right]_{\mathbf{r}=\mathbf{R}_{\min}} \tag{1.2}$$

The resulting harmonic model is thus an approximation to a conformational substate, valid for very small motions around the local minimum. However, such models have been routinely used in the study of larger amplitude motions, for example, the opening/closing motions that control the access of ligands to the active site in enzymes. Most of the criticism aimed at normal mode analysis concerns this use of a model for a conformational substate beyond its theoretical limit of applicability. However, other kinds of harmonic models exist, as will be shown below, and even the use of conformational substate models can be justified empirically because the outcome of the subsequent normal mode analysis usually yields results that are in agreement with experimental data. The low-energy motions in the local minima and in the global potential must therefore be very similar in shape. This is in fact plausible, because the motions that separate conformational substates and those that characterize large-amplitude motions are very different.

A low-energy motion on a large scale should also be a low-energy motion on a smaller scale.

Alternatively, one can directly construct a harmonic model around a given $\mathbf{R}_{\min}$ (e.g., an experimental conformation) by fitting the remaining parameters to experimental or simulation data. This approach has been used in particular for simplified protein models in which only the C_α atoms are represented explicitly. A reasonable and simplifying assumption is

$$U(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_{\text{all pairs } \alpha, \beta} U_{\alpha\beta}(\mathbf{r}_\alpha - \mathbf{r}_\beta) \tag{1.3}$$

with

$$U_{\alpha\beta}(\mathbf{r}) = \tfrac{1}{2} k(|\mathbf{R}_\alpha - \mathbf{R}_\beta|)(|\mathbf{r}| - |\mathbf{R}_\alpha - \mathbf{R}_\beta|)^2 \tag{1.4}$$

that is, the harmonic potential consists of a sum of pair terms that represent springs whose force constants $k(r)$ decrease with an increasing distance between the two atoms in the configuration that represent the minimum.

Such a potential, with a step function for $k(r)$, was first used with an all-atom model by Tirion [4], who showed that it reproduces the low-frequency end of the density of states rather well. Hinsen [5, 6] then used another variant (with $k(r)$ exponentially decreasing and a reduced description of the backbone by the C_α atoms) for characterizing slow protein motions by dynamical domains. The Anisotropic Network Model [7], although derived in a different way, is also equivalent to a potential of the form (1.3) for the C_α atoms, again with a step function for $k(r)$. As long as only an identification of the low-frequency modes is required, the form of $k(r)$ is indeed not critical.

On the other hand, a quantitative description of a potential well requires a more careful approximation. By fitting to a local minimum (substate) of the Amber 94 force field [8], Hinsen et al. [9] obtained the form

$$k(r) = \begin{cases} 8.6 \times 10^5 \dfrac{\text{kJ}}{\text{mol nm}^3} \cdot r - 2.39 \times 10^5 \dfrac{\text{kJ}}{\text{mol nm}^2}, & \text{for } r < 0.4 \text{ nm} \\ \dfrac{128 \text{ kJ nm}^4/\text{mol}}{r^6}, & \text{for } r \geq 0.4 \text{ nm} \end{cases} \tag{1.5}$$

and found that the global potential well can be described by scaling the local potential well down by a factor that must be evaluated for each protein individually. The special case for $r < 0.4$ nm takes care of nearest neighbors along the backbone, which are strongly bound through the very rigid peptide group. For other pairs, the interaction is mediated mostly by a large number of sidechain atoms. This model has been shown to reproduce the long-time dynamics of proteins remarkably well, as will be shown in Section 1.3.2.

1.3 Normal Modes

The basic idea of normal modes is illustrated in Figure 1.2 for a system with two coordinates, labeled r_1 and r_2. The harmonic potential well shown has two special directions, labeled e_1 and e_2, which correspond to the normal modes. Imagine the potential well as a real bowl in which a small ball moves around. The normal mode directions are special because the ball can move along any one of them back and forth. If it starts along any other direction (say, r_1), then it will be deflected by the potential along the perpendicular direction (r_2) as well, and thus move along both directions. Only the normal mode directions are independent. This independence greatly simplifies the analysis of the motions. In particular, oscillations of the ball along any one of the normal mode directions have a well-defined frequency, which is related to the curvature of the potential along the direction of motion. Any compound motion contains both frequencies. Knowing the normal modes thus permits the explicit evaluation of all possible vibrational frequencies in a system, assuming of course that the system has vibrational dynamics, that is, that friction can be neglected.

There is another important feature of normal modes that can be seen in Figure 1.2. The thick line describes a particular constant energy value. A ball that is dropped from a position on that line will bounce back to the same energy level again (assuming the absence of friction). If the ball moves along the lower normal mode (e_1, the one with the lower curvature and lower oscillation frequency), it can move further away from the minimum at a given energy than if it moved along the higher normal mode. This illustrates that the low normal modes describe large-amplitude motions. In a molecular system, the level of available energy is defined by the temperature.

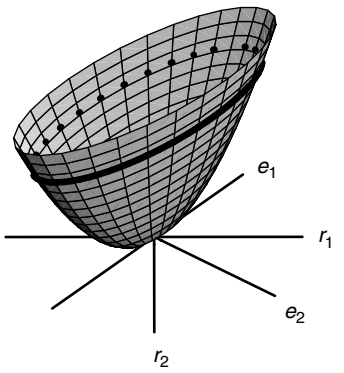

FIGURE 1.2
A two-dimensional harmonic potential well. The two Cartesian coordinate axes of the system are r_1 and r_2, the two normal mode directions are e_1 and e_2.

In the case of a protein with N atoms, there are $3N$ Cartesian coordinates and thus also $3N$ normal mode directions. It is useful to consider the $3N$-dimensional space defined by the $3N$ Cartesian coordinates, which is called configuration space. A $3N$-dimensional vector in this space can either represent a point, that is, a configuration of the protein, or a direction, that is, the change of a configuration. Normal mode vectors represent directions, as do velocity vectors and force vectors. A normal mode vector thus describes in which direction each atom moves, and how far it moves relative to the other atoms. However, a normal mode vector does *not* describe an absolute amount of displacement for any atom. Additional information (e.g., the temperature) is required for fixing the global amplitude of the atomic displacements.

Mathematically, the normal mode vectors are obtained as the eigenvectors $\mathbf{e}_i$ of the matrix $\mathbf{K}$, which are defined by

$$\mathbf{K} \cdot \mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad i = 1, \ldots, 3N \tag{1.6}$$

The $3N$ numbers λ_i are the associated eigenvalues that describe the curvature of the potential along the normal mode directions.

The independence of the normal modes makes it possible to rewrite the harmonic potential in the simpler form

$$U(\mathbf{c}) = \tfrac{1}{2}\mathbf{c} \cdot \mathbf{\Lambda} \cdot \mathbf{c} \tag{1.7}$$

The new interaction matrix $\mathbf{\Lambda}$ is diagonal and has the eigenvalues λ_i as its elements. The new coordinates $\mathbf{c}$ are given by

$$c_i = (\mathbf{r} - \mathbf{R}) \cdot \mathbf{e}_i \tag{1.8}$$

and the original coordinates $\mathbf{r}$ can be recovered through

$$\mathbf{r} = \mathbf{R} + \sum_{i=1}^{3N} c_i \mathbf{e}_i \tag{1.9}$$

Each of the coordinates c_i measures the distance from the minimum along one of the normal mode directions.

More important to us is, however, the physical interpretation of the normal modes. The eigenvalue λ_i describes the energetic cost of displacing the system by one length unit along the eigenvector $\mathbf{e}_i$. Normal mode analysis therefore classifies the possible deformations of a protein by their energetic cost. For realistic potentials, low-energy deformations correspond to *collective* or *delocalized* deformations, whereas high-energy modes are *local* deformations. This is a consequence of the nonlinearity of the interaction terms, plus the fact that short-range interactions (e.g., bond stretching) are stronger than long-range interactions (e.g., electrostatic).

This can be illustrated by a simple example: a linear chain of N equidistant particles, each of which interacts with its two neighbors through a spring of equilibrium length d, the pair potential is then given by Equation (1.4) with $k(r) = k_0$. Displacing a particle in the middle by a distance a causes two pair terms to increase by $\frac{1}{2}k_0a^2$. Displacing a group of ten particles in the middle by the same distance a (all in the same direction) also causes two pair terms to increase, by exactly the same amount. However, we should be comparing $3N$-dimensional displacement vectors of the same length, that is, the same norm in $3N$-dimensional space. Moving a group of M particles as a unit by a distance a yields a displacement vector with a norm of $\sqrt{M}a$. The incurred energy increase is thus proportional to $1/M$, that is, collective motions (large M) are energetically cheaper than local ones. This would not be the case if the potential were linear in the pair distance, local and global motions would then have equal energetic costs. A potential with a less than linear growth would even favor local moves. However, such potentials do not exist at the atomic scale. Finally, global displacements would be penalized if there were strong interactions at longer distances, beyond nearest neighbors. But such situations are not found on the atomic scale, the short-range interactions (the chemical bond structure) are the strongest ones.

When normal mode analysis is applied to an isolated protein, the first six eigenvalues λ_i are zero. They describe the six rigid-body movements of the protein (translation along three independent axes plus rotation around three independent axes) that incur no energetic cost at all. They are usually of no interest and ignored in the analysis, such that "the lowest-energy modes" in practice means "the lowest-energy modes with nonzero energies."

1.3.1 Vibrational Modes

If one assumes that the atoms in a molecule are classical particles, then the equations of motion for a molecule with a harmonic interaction potential of the form (1.1) are given by

$$\mathbf{M} \cdot \ddot{\mathbf{r}} = -\mathbf{K} \cdot (\mathbf{r} - \mathbf{R}) \tag{1.10}$$

The matrix $\mathbf{M}$ is a $3N \times 3N$ diagonal matrix, which contains the masses of the atoms on its diagonal, each mass being repeated three times, once for each of the three Cartesian coordinates. A system with these equations of motion is known as a $3N$-dimensional harmonic oscillator and is discussed in all textbooks on classical mechanics (see, e.g., [10]). We will therefore only give a summary of the solution.

With the introduction of *mass-weighted coordinates*,

$$\tilde{\mathbf{r}} = \sqrt{\mathbf{M}} \cdot \mathbf{r} \tag{1.11}$$

$$\tilde{\mathbf{R}} = \sqrt{\mathbf{M}} \cdot \mathbf{R} \tag{1.12}$$

$$\tilde{\mathbf{K}} = \sqrt{\mathbf{M}}^{-1} \cdot \mathbf{K} \cdot \sqrt{\mathbf{M}}^{-1} \tag{1.13}$$

the equations of motion can be rewritten as

$$\ddot{\tilde{\mathbf{r}}} = \tilde{\mathbf{K}} \cdot (\tilde{\mathbf{r}} - \tilde{\mathbf{R}}) \tag{1.14}$$

The $3N$ independent solutions of these equations have the form

$$\tilde{\mathbf{r}}(t) = \tilde{\mathbf{R}} + \tilde{\mathbf{A}}_i \cos(\omega_i t + \delta_i), \quad i = 1, \ldots, 3N \tag{1.15}$$

where δ_i is an arbitrary phase factor and ω_i and $\tilde{\mathbf{A}}_i$ are the solutions of the eigenvalue equation

$$\tilde{\mathbf{K}} \cdot \tilde{\mathbf{A}}_i = \omega_i \tilde{\mathbf{A}}_i \tag{1.16}$$

This is identical to Equation (1.6) except for the use of the mass-weighted force constant matrix.

The combination of the $3N$-dimensional vector $\mathbf{A}_i$ and the eigenvalue ω_i is known as a vibrational normal mode. Since this was historically the first type of normal mode analysis, and remains the most frequently used one, it is common to use the term "normal mode" for this form only.

The physical interpretation of $\mathbf{A}_i$ and ω_i can be obtained from Equation (1.15): ω_i is a vibrational frequency, and $\mathbf{A}_i$ is an amplitude vector that specifies how far and in what direction each individual atom moves. Vibrational normal mode analysis thus classifies all possible motions around a stable equilibrium state by vibrational frequency. Note that since the range of atomic masses is much smaller than the range of eigenvalues, the difference between energetic (Equation [1.6]) and vibrational (Equation [1.16]) analysis is not very large. Low-frequency modes are therefore to a very good approximation also low-energy modes, and vice versa. For historical reasons (normal mode analysis in chemistry was originally developed for describing the vibrational spectra of small molecules), most published normal mode studies on proteins use vibrational modes, even though the interpretation is often in terms of energetic modes.

Figure 1.3 shows the frequency spectrum of three proteins, crambin, lysozyme, and myoglobin, obtained from vibrational normal mode analysis using a conformational substate approximation to the Amber 94 potential [8]. The main observation is that the three spectra are nearly identical. The reason for this is that most of the modes describe motions that are common to all proteins, ranging from hydrogen vibrations (the well-separated block beyond 85 ps^{-1}) at the high end through internal vibrations of single amino acids down to vibrations of secondary-structure elements (helices, β-sheets). The small differences are due to the different amino acid distributions and different percentages of secondary structure motifs. The motions that are specific to a particular protein, and thus of interest for understanding its function, are at the far lower end of the spectrum.

It should be stressed that this analysis describes only vibrational motion in a conformational substate. There are larger amplitude motions along

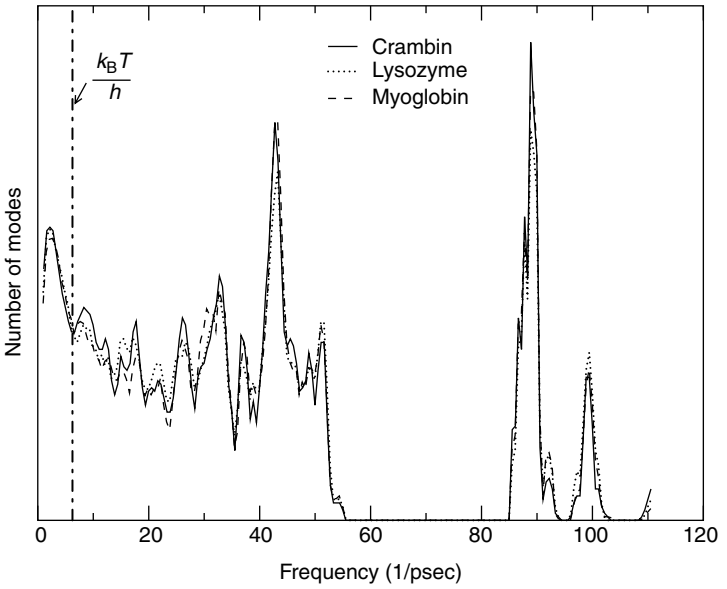

FIGURE 1.3
The vibrational frequency spectrum (number of modes per frequency interval) of three different proteins in a local minimum of the Amber 94 force field. The vertical line indicates the quantum limit for $T = 300$ K.

the lower-frequency modes as well, but they are diffusive, not vibrational. They will be discussed in Section 1.3.2.

It should also be noted that at the high frequency end, quantum effects become important. The criterion for the applicability of classical mechanics is $h\nu \ll k_B T$. At 300 K, this yields $\nu \ll 6$ psec^{-1}, which, as Figure 1.3 shows, is satisfied for only a very small part of the vibrational spectrum. However, since the transformation to normal mode coordinates remains valid in a quantum description, only the dynamic interpretation must be adapted.

1.3.2 Langevin and Brownian Modes

The real large-amplitude motions in proteins traverse many conformational substates. The transition from one conformational substate to the next requires crossing a small energy barrier. At the structural level this means, for example, that some sidechain rearrangements are necessary before the backbone motion can proceed. An explicit treatment of these barrier crossings is not desirable, and also not necessary. One can model such situations by a smoothed-out potential (see Figure 1.1) and replace the barrier crossings by the introduction of friction and random forces into the dynamics. The simplest model involving friction is known as *Langevin dynamics*. It consists

of augmenting Equation (1.10) by two terms:

$$\mathbf{M} \cdot \ddot{\mathbf{r}} = -\mathbf{K} \cdot (\mathbf{r} - \mathbf{R}) - \mathbf{\Gamma} \cdot \dot{\mathbf{r}} + \xi(t) \tag{1.17}$$

The first term, proportional to the velocities, is a friction term, defined by a $3N \times 3N$ matrix $\mathbf{\Gamma}$, called friction matrix, which will be discussed later. The second term describes a random force that satisfies the conditions

$$\langle \xi(t) \rangle = 0 \tag{1.18}$$

$$\langle \xi(t)\xi(t') \rangle = 2k_{\mathrm{B}}T\mathbf{\Gamma}\delta(t - t') \tag{1.19}$$

The second condition specifies that the random force is a white noise signal (i.e., uncorrelated in time) with an amplitude defined to add on average just as much energy to the system as is taken out by the friction term.

A method for solving this equation numerically has been given by Lamm and Szabo [11]. However, it will not be discussed here because a further useful simplification can be made for the case of large-amplitude motions in proteins. In general, Langevin modes describe damped oscillations plus random displacements along a normal mode coordinate. When the friction coefficients are very large, the oscillations become overdamped: the molecule moves slowly back toward its energetic minimum, but reaches it only asymptotically and never swings back. The random displacements become the dominant aspect of the dynamics, and one observes Brownian motion (diffusion) with preferential movements toward the minimum. This is the dynamic behavior that the large-amplitude motions of proteins display. It can be described by the formalism of Brownian Dynamics, which consists of a differential equation (known as the Smoluchowski equation) for the probability distribution of the random displacements. This equation can be solved analytically for a harmonic potential. The derivation is too lengthy to be reproduced here, the reader is therefore referred to Reference 9 and to Section 2.2 of Reference 12. The result is again an eigenvalue problem, this time for the matrix

$$\hat{\mathbf{K}} = \sqrt{\mathbf{\Gamma}}^{-1} \cdot \mathbf{K} \cdot \sqrt{\mathbf{\Gamma}}^{-1} \tag{1.20}$$

that is, a friction-weighted force constant matrix. Its eigenvalues $\hat{\lambda}_i$, $i = 1, \ldots, 3N$, are the relaxation coefficients of the Brownian modes, whose directions are again given by the eigenvectors. If the protein were deformed along Brownian mode k by an amplitude A, and if then the random forces were switched off, the protein would return toward the energetic minimum along the same direction and its position along this direction would be given by $A \exp(-\hat{\lambda}_k t)$.

Like other normal mode techniques, Brownian mode analysis requires a stable conformation of the protein as input and a harmonic model for the global potential well. In addition, a model for the friction matrix $\mathbf{\Gamma}$ is required. Since friction manifests itself already on short time scales, it can be measured

from molecular dynamics (MD) simulations of proteins. With the simplifying assumption that each particle of the protein has an independent friction constant (which implies that $\mathbf{\Gamma}$ is diagonal), it is sufficient to calculate the mean-square displacement of each particle from the simulation trajectory and fit the short-time behavior to a straight line in order to obtain approximate values of the elements of $\mathbf{\Gamma}$. It turns out that the friction constant can be well described by a linear function of the local density in the protein around the particle of interest, averaged over a sphere of 1.5 nm radius [9]. In a typical compact protein, the local density is uniform on that length scale, the variations are thus due to surface effects: for particles near the surface, the sphere contains water, whose density is much smaller than that of the protein itself. The correlation between friction constant and amount of protein matter in the vicinity of the particle is not surprising in view of the explanation of the origin of friction given above, that is, interactions with other atoms in the protein, in particular sidechain atoms. However, the idea that friction is a solvent effect is quite popular in the literature, although it has never been backed by any data.

Several experimentally observable quantities, in particular time correlation functions, can be calculated from Brownian modes analytically [12], which permits the study of protein dynamics at arbitrarily long time scales. Figure 1.4 shows that such a model can yield surprisingly good results. It shows the incoherent intermediate scattering function for a C-phycocyanin dimer from a two-level normal mode calculation (Brownian modes for the long-time dynamics plus vibrational modes for short-time effects) and from a standard MD trajectory. It should be noted that the MD results should tend to the same asymptotic values as the normal modes curves; the fact that they do not indicates that the trajectory of 1.6 nsec is not long enough for sampling all the motions. A look at the relaxation times obtained from the Brownian modes confirms this: the largest relaxation time is 4.5 nsec. The absence of sampling problems is in fact an important advantage of normal mode techniques in the study of slow protein dynamics.

In summary, Brownian mode calculations demonstrate that a very simple harmonic potential with few parameters can reproduce the backbone dynamics of a protein very well if an appropriate dynamical model is chosen. The major limitation is the restriction to motions around a stable energetic minimum.

1.4 Interpretation and Analysis of Normal Modes

In the study of molecular systems, normal modes are used to answer particular scientific questions. In order to draw valid conclusions, it is important to understand the methods and, in particular, their limitations.

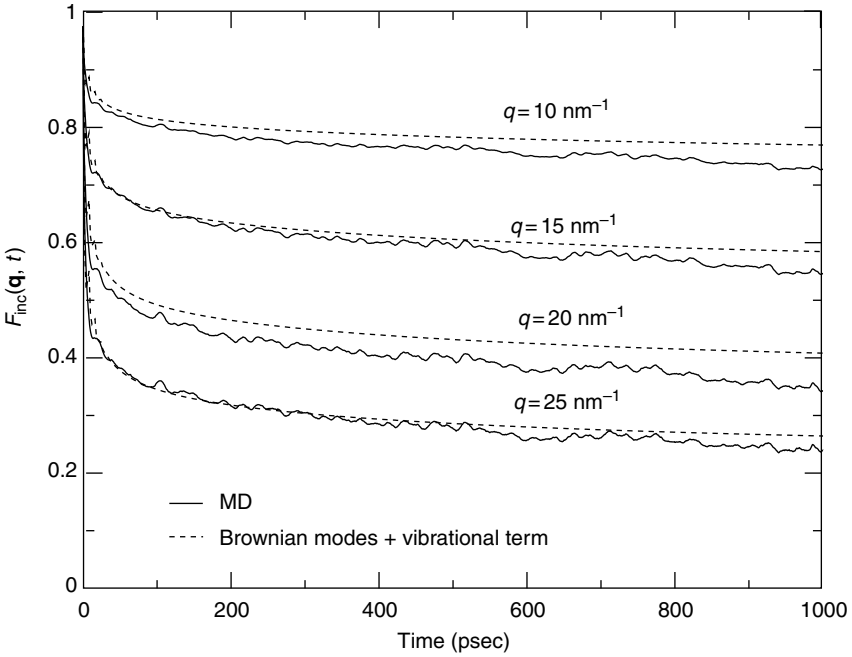

FIGURE 1.4
The incoherent intermediate scattering function $F_{\text{inc}}(\mathbf{q}, t)$, a quantity observable in neutron scattering experiments, calculated from a mixed Brownian/vibrational modes model and from an MD trajectory for a C-phycocyanin dimer. Both calculations are for a coarse-grained model in which a single point mass located at the C_α position represents a whole residue. The normal modes were calculated directly for this model, the MD trajectory was generated from an all-atom simulation.

The applications of normal modes can be broadly classified into two groups. Those in the first group use all modes or a large subset (usually the lowest energy modes) as a convenient analytical representation of the potential well. In that case the only limitations are due to the necessarily approximate nature of the harmonic model, and due to the choice of a subset. The other group contains all analyses that look at the properties of individual modes. In this case, care must be taken to avoid an overinterpretation of the data.

One potential pitfall of single mode analysis is discussing the differences of modes that are nearly equal in energy. In the extreme case of exactly equal energies (the modes are then called degenerate), the modes that come out of a numerical calculation represent arbitrary choices of the algorithm. Any combination of such modes would be an equally valid mode. Interpreting the characteristics of any one such mode or the differences between the degenerate modes is no more meaningful than discussing the differences between motion along the x and the y coordinates in an arbitrarily chosen Cartesian coordinate system. Although this is strictly true only for equal energies, it is also approximately true for approximately equal energies. A small difference

in energy between two modes should be considered a probably unreliable detail of the numerical model, rather than something fundamental about the system being studied. In practice, only a few of the lowest modes in a protein are sufficiently well separated to merit an individual discussion, and even that is not always the case. In all other cases, it is preferable to analyze the coordinate subspace spanned by all modes in a certain range of timescales.

A second pitfall is placing too much importance on the frequency of a mode obtained from a vibrational normal mode calculation. As discussed above, the slow modes that are characteristic of a particular protein and often related to its function show diffusional behavior on long timescales. Vibrational dynamics occurs only inside a conformational substate for a short duration and is rarely of interest. Vibrational normal mode analysis is thus useful mostly for higher frequencies, for example, when comparing to spectroscopic measurements. For assessing the time scales of slow motions, Brownian modes are the appropriate approach.

A very useful approach in the analysis of normal modes is to turn attention away from individual modes and toward the types of motion in the protein that one would like to analyze. For example, one can ask the question: "Which modes (and thus which energies and which time scales) are involved in the rotation of this domain?" Or, turning to higher modes, "Which frequencies are involved in helix bending motions?"

Such questions can be answered using projection methods [13], which are based on an important mathematical property of normal modes: the normal mode vectors $\mathbf{e}_i$ (see Equation [1.6]), being the eigenvectors of a matrix, form a basis of the $3N$-dimensional configuration space of the protein. This means that any vector d in configuration space, and thus any type of motion, can be written as a superposition of normal mode vectors with suitable prefactors p_i, which are the *projections* of $\mathbf{d}$ onto mode i. Mathematically, the projections are defined by

$$p_i = \mathbf{d} \cdot \mathbf{e}_i \tag{1.21}$$

and satisfy the relation

$$\sum_{i=1} 3N p_i^2 = 1 \tag{1.22}$$

because the normal mode vectors form a basis of configuration space. It is therefore possible to interpret p_i^2 as the contribution of mode i (and its associated energy and time scales) to the motion described by $\mathbf{d}$.

Many interesting types of motion are described by more than one degree of freedom. For example, the rigid-body translation of a helix has three degrees

of freedom, one for each independent direction in 3D-space:

$$\mathbf{d}_x = \begin{pmatrix} (0,0,0) \\ \dots \\ (0,0,0) \\ (1,0,0) \\ \dots \\ (1,0,0) \\ (0,0,0) \\ \dots \\ (0,0,0) \end{pmatrix}, \quad \mathbf{d}_y = \begin{pmatrix} (0,0,0) \\ \dots \\ (0,0,0) \\ (0,1,0) \\ \dots \\ (0,1,0) \\ (0,0,0) \\ \dots \\ (0,0,0) \end{pmatrix}, \quad \mathbf{d}_z = \begin{pmatrix} (0,0,0) \\ \dots \\ (0,0,0) \\ (0,0,1) \\ \dots \\ (0,0,1) \\ (0,0,0) \\ \dots \\ (0,0,0) \end{pmatrix} \tag{1.23}$$

The nonzero entries in these vectors correspond to the atoms that make up the helix. For the case of M vectors (in this example we have $M = 3$), the projections are defined as

$$p_i = \frac{1}{\sqrt{M}} \sum_{k=1}^{M} \mathbf{d}_k \cdot \mathbf{e}_i \tag{1.24}$$

such that the sum of p_i^2 is again 1, and p_i can again be interpreted as the quantitative contribution of mode i to the motion under consideration. A convenient graphical representation is a plot of

$$C_k = \sum_{i=6}^{k} p_i^2, \quad k = 1, \dots, 3N \tag{1.25}$$

against k, ω_k (for vibrational modes), or $\hat{\lambda}_k$ (for Brownian modes). This yields a curve that increases from 0 to 1, with the steepest increase in the time scales that contribute most to the type of motion being studied.

An example for such an analysis is shown in Figure 1.5. It is taken from a normal mode study of the dynamics and conformational changes of Ca-ATPase [14] and shows how helix translations and rotations are distributed over the normal modes. In particular, it shows that different helices move on different timescales, and also that some helices have a wider time scale spectrum than others. In the case of Ca-ATPase, the helices near the A domain are characterized by longer timescales and larger amplitudes than the other helices. No explicit time scales were obtained in this calculation, but this would have been possible by performing a Brownian mode analysis (see Section 1.3.2).

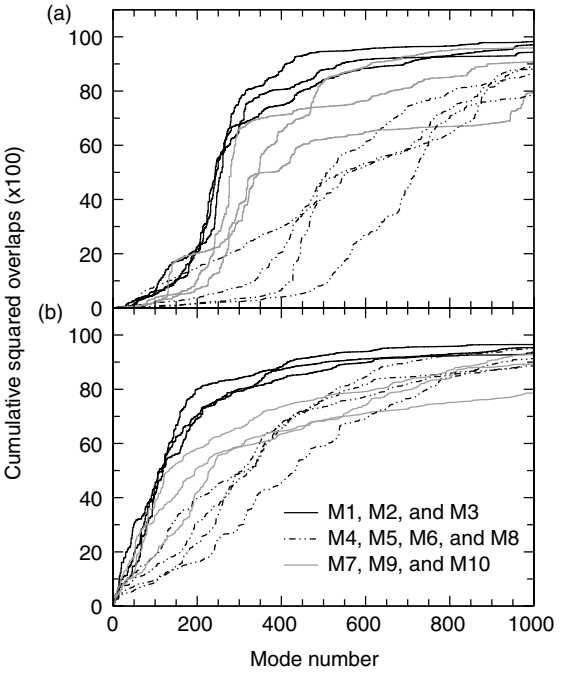

FIGURE 1.5
The cumulative projections C_k (see Equation [1.25]) of rigid-body translations (a) and rotations (b) of the transmembrane helices in Ca-ATPase onto the normal modes. Only translations along and rotations around the helix axes were taken into account. The plot shows the different timescales and amplitudes that characterize the motions of the different helices.

1.5 Conclusion

The goal of this chapter is to give an overview of the harmonic models and normal mode techniques that are used in studying the behavior of proteins. Any such overview is necessarily incomplete, and this chapter is no exception. Quasi-harmonic analysis, which derives a force constant matrix from thermodynamic calculations obtained from an MD trajectory, was left out because it is a technique for analyzing trajectories rather than an independent method. Normal mode calculations on continuous deformable media models were left out as well, because they are of interest mainly to the community of electron microscopists. Other rather specialized techniques have not been mentioned either. Finally, the actual numerical algorithms that are useful for identifying normal modes were not covered because they are either straightforward textbook algorithms (for sufficiently small systems) or specialized techniques discussed in Chapters 17 and 18. As for applications, the possibilities are numerous and the reader can find ample inspiration in the other chapters of this book.

References

1. Frauenfelder, H., Parak, F., and Young, R.D. Conformational substates in proteins, *Ann. Rev. Biophys. Biophys. Chem.*, 17: 451–479, 1988.
2. Elber, R., and Karplus, M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin, *Science*, 235: 318–321, 1987.
3. Kitao, A., Hayward, S., and Go, N. Energy landscape of a native protein: jumping-among-minima model, *Proteins*, 33: 496–517, 1998.
4. Tirion, M.M. Low-amplitude elastic motions in proteins from a single-parameter atomic analysis, *Phys. Rev. Lett.*, 77: 1905–1908, 1996.
5. Hinsen, K. Analysis of domain motions by approximate normal mode calculations, *Proteins*, 33: 417–429, 1998.
6. Hinsen, K. Thomas, A., and Field, M.J. Analysis of domain motions in large proteins, *Proteins*, 34: 369–382, 1999.
7. Atilgan, A.R. Durell, S.R. Jernigan, R.L., Demirel, M.C. Keskin, O., and Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model, *Biophys. J.*, 80: 505–515, 2001.
8. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz Jr, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. A second generation force field for the simulation of proteins and nucleic acids, *J. Am. Chem. Soc.*, 117: 5179–5197, 1995.
9. Hinsen, K., Petrescu, A.-J., Dellerue, S., Bellissent-Funel, M.C., and Kneller, G.R. Harmonicity in slow protein dynamics, *Chem. Phys.*, 261: 25–38, 2000.
10. Goldstein, H. "*Classical Mechanics*," Addison-Wesley Pub. Co., Reading, MA, 1980.
11. Lamm, G. and Szabo, A. Langevin modes of macromolecules, *J. Chem. Phys.*, 85: 7334–7348, 1986.
12. Kneller, G.R. Inelastic neutron scattering from damped collective vibrations of macromolecules, *Chem. Phys.*, 261: 1–24, 2000.
13. Hinsen, K. and Kneller, G.R. Projection methods for the analysis of complex motions in macromolecules, *Mol. Sim.*, 23: 275–292, 2000.
14. Reuter, N., Hinsen, K., and Lacapre, J.-J. Transconformations of the SERCA1 Ca-ATPase: a normal mode study, *Biophys. J.*, 85: 2186–2197, 2003.