Normal Mode Analysis Techniques in Structural Biology

José Ramón López-Blanco, *Department of Biological Physical Chemistry, Instituto de Química Física Rocasolano, Madrid, Spain*

Osamu Miyashita, *Department of Biochemistry & Molecular Biophysics, The University of Arizona, Tucson, Arizona, USA*

Florence Tama, *Department of Biochemistry & Molecular Biophysics, The University of Arizona, Tucson, Arizona, USA*

Pablo Chacón, *Department of Biological Physical Chemistry, Instituto de Química Física Rocasolano, Madrid, Spain*

Based in part on the previous version of this eLS article 'Normal Mode Analysis Techniques in Structural Biology' (2007) by Osamu Miyashita and Florence Tama.

Advanced article

Article Contents

Online posting date: 15th October 2014

The dynamic simulation of macromolecular systems with biologically relevant sizes and time scales is critical for understanding macromolecular function. In this context, normal mode analysis (NMA) approximates the complex dynamical behaviour of a macromolecule as a simple set of harmonic oscillators vibrating around a given equilibrium conformation. This technique, originated from classical mechanics, was first applied to investigate the dynamical properties of small biological systems more than 30 years ago. During this time, a wealth of evidence has accumulated to support NMA as a successful tool for simulating macromolecular motions at extended length scales. Today, NMA combined with coarse-grained representations has become an efficient alternative to molecular dynamics simulations for studying the slow and large-amplitude motions of macromolecular machines. Interesting insights into these systems can be obtained very quickly with NMA to characterise their flexibility, to predict the directions of their collective conformational changes, or to help in the interpretation of experimental structural data. The recently developed methods and applications of NMA together with an introduction of the underlying theory will be briefly reviewed here.

eLS subject area: Structural Biology

How to cite:
López-Blanco, José Ramón; Miyashita, Osamu; Tama, Florence; and Chacón, Pablo (October 2014) Normal Mode Analysis Techniques in Structural Biology. In: eLS. John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0020204.pub2

Introduction

A detailed knowledge of the structure and dynamics of macromolecular systems permits a deep understanding of biological processes and leads to advances in the rational discovery of disease treatments. The main functions of living cells (replication, transcription, translation, folding and protein turnover) are usually governed by large macromolecular complexes (polymerases, ribosomes, chaperonins and proteasomes). For example, the ribosomal machinery produces new proteins according to the genetic code; the chaperonin proteins assist with the folding of these newly formed proteins; and tubulin and actin filaments support cellular shape. These macromolecules are in perpetual motion, following thermal fluctuations and undergoing energy-dependent conformational rearrangements to accomplish biological functions. The magnitude of such motions ranges from a few Angstroms ($\text{Å} = 10^{-10}$ m) to hundreds of Å, and their associated time scale ranges from picoseconds to seconds (see **Figure 1**). Thermal fluctuations of bond lengths and angles occur on a relatively small scale (< 1 Å) but occur very fast (picoseconds). On the other end of the spectrum, large-scale rearrangements occur on a much longer time scale, from hundreds of nanoseconds to even seconds. Such rearrangements include folding of the protein from the nascent polypeptide chain, or changes in the 3D structure due to interaction with a ligand or another macromolecule. For example, in **Figure 1**, two crystallographic structures of the adenylate kinase in different conformations evidence the large conformational rearrangements undergone upon ligand binding. Unfortunately, the direct experimental observation of the functional motions is often not possible using current high-resolution techniques; therefore, computational methods are the only way to study these

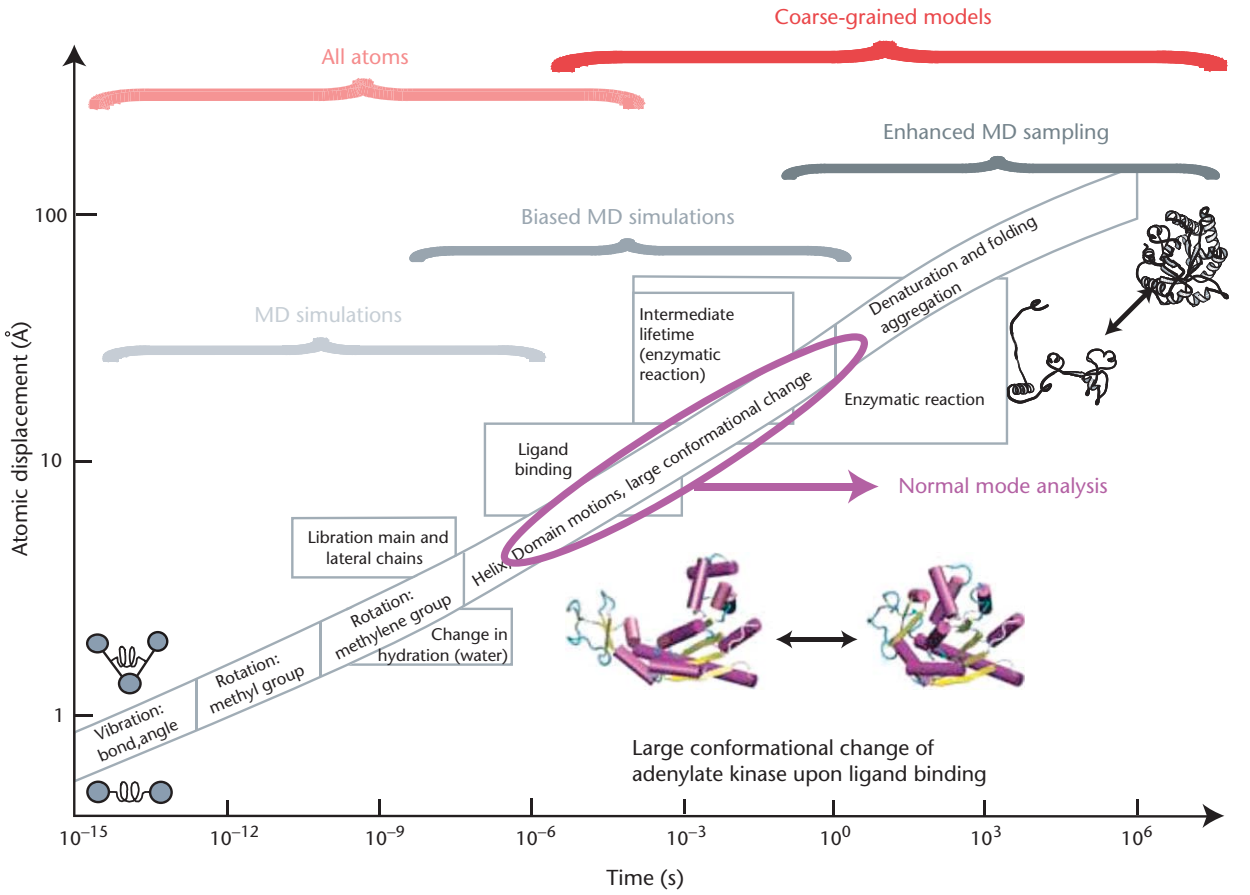

Figure 1 Dynamics of biological molecules and computational approaches that can be used to investigate dynamical properties.

important conformational changes in detail. The most common computational technique used to study the dynamical properties of biological molecules is molecular dynamics (MD). In MD, the system evolves as a function of time by iteratively integrating Newton's equations of motion. However, even though computational methodology and processing power have been improving significantly, the application of MD to large-scale macromolecular assemblies is limited to relatively short time scales, due to the computational complexity of all-atom MD simulations (see **Figure 1**). Time scales of functional motions in large macromolecular assemblies are still computationally intensive and highly impractical. For example, the 100 ns simulation for a relatively small protein filament in explicit solvent (approximately 2000 amino acids and 300 000 atoms) takes 3 days using 512 processors in state-of-the-art computational resources. Thus, it would take too long to reach more relevant time scales (milliseconds) for large-scale rearrangements. An alternative approach to extend simulation times is the use of coarse-grained representations (Ingólfsson *et al.*, 2014), this reduces the number of atoms necessary for simulation. This type of simplification enables microsecond time scales to be reached, at least for small proteins. However, such calculations are still computationally expensive for large macromolecular assemblies and slow/large-amplitude motions. **See also**: Molecular Dynamics

A less time-consuming alternative to simulate large or slow conformational rearrangements for large biological molecules is normal mode analysis (NMA) (Bastolla, 2014; Cui and Bahar, 2010). This approach, commonly used in physics, was introduced in structural biology around 30 years ago to study the dynamics of the biological macromolecules (Brooks and Karplus, 1983; Go *et al.*, 1983; Levitt *et al.*, 1985). Although MD approximately solves the equations of motion using a realistic force field, NMA obtains the exact solutions, but for a simplified force field. The basic assumption (and limitation) of the vibrational analysis is that the potential energy of the system varies quadratically about a given minimum energy conformation. Based on this harmonic approximation, NMA analytically solves the equations of motion (Lagrangian or Hamiltonian) in a matter of minutes. The resulting solutions are a set of orthogonal displacement vectors (or normal modes) with their corresponding frequencies, which encode all possible motions around the initial conformation. The modes are sorted according to the energy required for their movement; while the high-frequency modes represent high-energy localised displacements, low-frequency modes correspond to low-energy collective

conformational changes. These collective movements are closely related to functional motions detected experimentally by crystallography and nuclear magnetic resonance (NMR) or observed in MD simulations. Therefore, NMA can be used as a frequency filter for reducing the dimensionality of the system and for separating the essential (collective low-frequency modes) from the nonessential (local high-frequency modes) movements. Such dimensionality reduction facilitates the interpretation of low-resolution experimental data. In summary, the characterisation of the essential motions permits us to obtain useful predictions about the dynamics, long-range coupling, allosteric regulation, and elastic properties of biological molecules. Furthermore, the necessary flexibility required during catalysis or when two or more biomolecules interact can be often approximated by a few normal modes.

In this article, we will introduce the computational technique of NMA, which has widely proven useful for analysing the functional motions of large biological molecules. The focus is limited to introducing NMA theory with a brief description of recently developed methods and their applications. Comprehensive overviews, applications, and discussions about the NMA methodology can be found in Further reading materials.

Normal Mode Theory

Using the classical mechanics formulation of NMA (Goldstein *et al*., 2002), the complex dynamical behaviour of a macromolecule can be approximated as a simple set of harmonic oscillators vibrating around a given equilibrium conformation (Bastolla, 2014; Cui and Bahar, 2010). This mechanical system consists of N atoms under a given force field and located at positions $\mathbf{r} = (\mathbf{r}_1,\ldots, \mathbf{r}_n,\ldots, \mathbf{r}_N)$, where $\mathbf{r}_n$ represents the Cartesian coordinates (x_n,y_n,z_n) of the atom n. The time evolution of the system is uniquely defined by the Hamiltonian, which is simply the sum of the kinetic $K(\mathbf{r})$ and the potential $U(\mathbf{r})$ energy terms:

$$\mathscr{H}(\mathbf{r}) \cong K(\mathbf{r}) + U(\mathbf{r}) \tag{1}$$

The Taylor expansion of the potential energy function around an equilibrium conformation ($\mathbf{r}^0$) gives:

$$U(\mathbf{r}) \cong U(\mathbf{r}^0) + \sum_i^{3N} \left.\frac{\partial U}{\partial r_i}\right|_{\mathbf{r}=\mathbf{r}^0} (r_i - r_i^0) + \frac{1}{2!}\sum_i^{3N}\sum_j^{3N} \left.\frac{\partial^2 U}{\partial r_i \partial r_j}\right|_{\mathbf{r}=\mathbf{r}^0} \times (r_i - r_i^0)(r_j - r_j^0) + \cdots \tag{2}$$

where r_i and r_j are the $3N$ Cartesian coordinates of $\mathbf{r}$. Because $\mathbf{r}^0$ is by definition at a minimum of the energy function, $\partial U/\partial r_i(\mathbf{r}^0)$ vanishes. Moreover, the terms beyond the second order can be neglected for small displacements, that is, assuming that the potential energy of the system varies quadratically about $\mathbf{r}^0$. This basic assumption (and a latent limitation) is founded on the observation that biomolecules behave, more than expected, as if the energy surface were parabolic, even though the potential contains many local minima (see **Figure 2**). Then, after defining the potential energy of the reference structure as $U(\mathbf{r}^0) = 0$, the potential energy function can be approximated as:

$$U(\mathbf{r}) \cong \frac{1}{2}\sum_i^{3N}\sum_j^{3N} \left.\frac{\partial^2 U}{\partial r_i \partial r_j}\right|_{\mathbf{r}=\mathbf{r}^0} (r_i - r_i^0)(r_j - r_j^0) \tag{3}$$

The kinetic energy function is defined by:

$$K(r) = \frac{1}{2}\sum_i^{3N} m_i \left(\frac{dr_i}{dt}\right)^2 \tag{4}$$

where m_i corresponds to the mass of the atom with coordinate i. For convenience, the Hamiltonian (eqn (1)) is rewritten in mass-weighted coordinates, $X_i = m_i^{1/2}(r_i - r_i^0)$, from eqns (3) and (4):

$$\mathscr{H}(X) \cong \frac{1}{2}\sum_i^{3N} \dot{X}_i^2 + \frac{1}{2}\sum_i^{3N}\sum_j^{3N} \left.\frac{\partial^2 U}{\partial X_i \partial X_j}\right|_{\mathbf{X}=\mathbf{X}^0} X_i X_j \tag{5}$$

where the dot over the X indicates the time derivative. The oscillatory motions corresponding to this Hamiltonian are coupled. In other words, the displacement of a given coordinate depends on the displacements of the others (see **Figure 2**). Fortunately, the motions can be reformulated as a superposition of independent (uncoupled) harmonic oscillators by choosing the appropriate normal mode coordinates ($\mathbf{q}$). The mass-weighted Cartesian and normal mode coordinates are linearly related by:

$$\mathbf{X} = \mathbf{A}\mathbf{q} \tag{6}$$

where $\mathbf{A}$ is an orthonormal transformation matrix; thus satisfying:

$$\mathbf{A}^{\mathbf{T}}\mathbf{A} = \mathbf{I} \tag{7}$$

The main consequence of using these new coordinates is that the kinetic and potential energy terms from eqn (5) can be further simplified. Using eqns (6) and (7) to simplify the kinetic energy, and converting the summation into matrix form, we get:

$$K(q) = \frac{1}{2}\sum_i^{3N} \dot{X}_i^2 = \frac{1}{2}\dot{\mathbf{X}}^T\dot{\mathbf{X}} = \frac{1}{2}\dot{\mathbf{q}}^T\mathbf{A}^T\mathbf{A}\dot{\mathbf{q}} = \frac{1}{2}\dot{\mathbf{q}}^T\dot{\mathbf{q}} \tag{8}$$

For simplifying the potential energy, the Hessian matrix $\mathbf{H}(X)_{ij} = \partial^2 U/\partial X_i \partial X_j$ can be transformed into the normal mode coordinates using $\mathbf{A}$:

$$\mathbf{H}(q) = \mathbf{A}^T\mathbf{H}(X)\mathbf{A} = \mathbf{\Lambda} \tag{9}$$

where the matrix to be determined $\mathbf{A}$ diagonalizes $\mathbf{H}$, and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_{3N})$ is the resulting diagonal

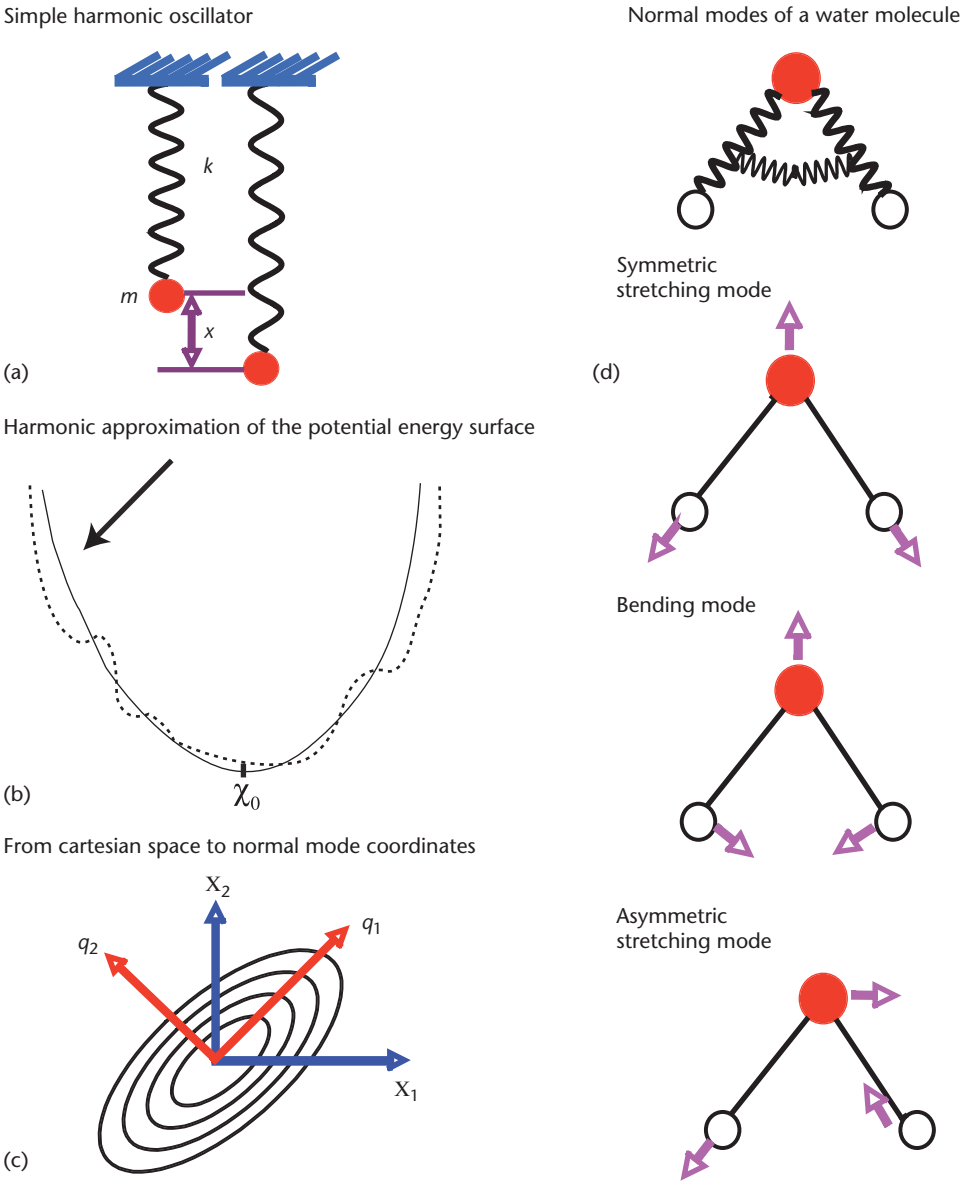

Figure 2 Normal mode theory. (a) Description of a simple harmonic oscillator: A particle *m* is attached to a spring with a force constant *k* and its displacement is *x*. (b) Normal mode analysis: harmonic approximation of the potential energy surface. For any biological system, the real energy surface is rugged (dotted line) but for the normal mode analysis, the surface is approximated as a harmonic surface (plain line). (c) Normal mode coordinates are independent (uncoupled) but not Cartesian coordinates (coupled). Although the contour lines represent the equipotential points of a parabolic force field in a two-dimensional space, the blue and red axis correspond to Cartesian (X) and normal mode (*q*) systems of coordinates, respectively. When some particle is released at any of the normal mode axes, its trajectory stays on this axis. In contrast, when the particle is released at some other point, its motion needs to be described by both Cartesian axes. (d) Simple normal mode vectors for the water molecule. Each arrow represents the direction of motion that each atom will undergo as obtained from normal mode theory. The three distinct motions predicted by NMA for the water molecule, i.e. symmetric and asymmetric stretching modes plus a bending mode, are in agreement with experimental observations.

matrix. Thus, the potential energy in normal mode basis becomes:

$$U(q)=\frac{1}{2}\sum_{k}^{3N}\sum_{l}^{3N} q_k \mathbf{A}^T \mathbf{H}(X)\mathbf{A}q_l=\frac{1}{2}\mathbf{q}^T\mathbf{\Lambda q} \qquad (10)$$

where indices k and l correspond to the $3N$ normal mode coordinates. Finally, after reverting the matrix notation of eqns (8) and (10) to summations, the transformed Hamiltonian uncouples as a set of independent harmonic oscillators:

$$\mathscr{H}(q)\cong\frac{1}{2}\sum_{k}^{3N}\dot{q}_k^2+\frac{1}{2}\sum_{k}^{3N}\lambda_k q_k^2 \qquad (11)$$

Note that the cross-terms found in the Hamiltonian of eqn (5) have disappeared. In practice, the transformation

matrix $\mathbf{A}$ and the diagonal matrix $\mathbf{\Lambda}$ are determined by solving the standard eigenvalue problem, that is, diagonalizing the Hessian matrix $\mathbf{H}(X)$:

$$\mathbf{HA}=\mathbf{\Lambda A} \tag{12}$$

The transformation $\mathbf{A}=(\mathbf{a}_1,\ldots,\mathbf{a}_k,\ldots,\mathbf{a}_{3N})$ contains the eigenvectors ($\mathbf{a}_k$), and the diagonal matrix $\mathbf{\Lambda}$ contains the corresponding eigenvalues (λ_k). Each pair of eigenvector and the associated eigenvalue ($\mathbf{a}_k$, λ_k) is known as a normal mode and represents one independent oscillator. Although the eigenvector provides the relative amplitudes of the collective atomic oscillations in mass-weighted Cartesian coordinates, the eigenvalue λ_k determines the oscillation frequency ($v_k=\lambda_k^{1/2}/2\pi$) which is the same for all atoms. Six of the modes have zero frequency (null modes) and correspond to the rigid body motions (three translations and three rotations) of the macromolecule. Since they represent trivial motions they are usually removed from summations. Thus, the dynamics of the system can be concisely described as a linear combination of $3N-6$ independent normal mode oscillators:

$$q_k=b_k\cos(2\pi v_k t+\varphi_k) \tag{13}$$

The b_k and φ_k are the maximum-amplitude and phase variables, respectively, and are determined by the initial conditions. To convert the motions from normal coordinates to Cartesian coordinates, eqn (6) must be employed first to obtain the mass-weighted Cartesian displacements, and then the simple formula $(r_i-r_i^0)=m_i^{-1/2}X_i$ to revert such weighting.

As a simple example, **Figure 2d** shows the resulting normal mode vectors obtained for a water molecule. NMA reveals three well known motions of the water molecule, that is, the symmetric and asymmetric stretching modes plus the bending mode. The frequencies obtained from these modes can be directly related to infrared experiments for which bond bending and stretching can be experimentally observed.

If the system is in thermal equilibrium, statistical thermodynamics theory states that the average energy of each mode is equal to $k_BT/2$, where T is the absolute temperature and k_B the Boltzmann constant. Thus, the average squared fluctuations of the q_k normal mode coordinate can be estimated using the potential energy of a single mode ($U_k=\lambda_k\langle q_k^2\rangle/2$):

$$\langle q_k^2\rangle=\frac{k_BT}{\lambda_k}=\frac{k_BT}{(2\pi v_k)^2} \tag{14}$$

The average of the squared atomic fluctuations in Cartesian coordinates can be obtained from eqns (6) and (14) after reverting the mass-weighting:

$$\langle(r_i-r_i^0)^2\rangle=\frac{1}{m_i}\sum_k^{3N-6}a_{ki}^2\langle q_k^2\rangle=\frac{k_BT}{m_i}\sum_k^{3N-6}\frac{a_{ki}^2}{(2\pi v_k)^2} \tag{15}$$

where a_{ki} is the i-th component of the k-th eigenvector ($a_{ki}=\partial X_i/\partial q_k$). From this equation, it is evident that the largest contribution to the atomic displacement comes from the lowest frequency/energy normal modes. These modes represent the most collective motions, that is, a large number of atoms with significant displacements (a_{ki}). Conversely, only a few atoms contribute to the motion (local) in high-frequency/energy eigenvectors. Lowest frequency modes actually correlate well with experimentally observed conformational changes in proteins and nucleic acids. Probably, such modes are relevant for biological functions because large conformational changes can be induced at a lower energetic cost by perturbations such as ligand binding or environmental changes (pH, ionic strength, temperature, etc.). Thus, studies employing NMA generally focus on these modes (Tama and Sanejouand, 2001). **Figure 3** shows several collective and local normal modes of the adenylate kinase protein. In low-frequency modes (top) almost all the atoms are experiencing a concerted motion however, in high-frequency modes (bottom), only a few are moving together.

Potential Energy Functions for Normal Mode Analysis

NMA is usually performed using a high-resolution structure of the biological molecule determined from X-ray crystallography or NMR. In the classical NMA approach, the potential energy terms for atomic interactions are defined by standard MD force fields. This NMA approach requires an initial energy minimisation step to ensure that the structure is at a minimum of the potential energy function. Otherwise, negative frequency modes may arise due to unstable equilibrium conditions. These minimisation procedures are not computer intensive compared with MD simulations but they require user time and expertise. Instead of using a detailed force field, Tirion (1996) pioneered the combination of NMA with a simplified protein representation (the so-called 'elastic network model' (ENM)) to reproduce the low-frequency normal modes calculated from detailed potential energy functions. In the ENM, the potential energy is assumed quadratic in the displacements, as in Hookean springs (see **Figure 2**), and corresponds to a three-dimensional elastic network of harmonic springs that keeps the atoms together. It is defined by:

$$U(\mathbf{r})=\sum_{n<m}^{N}\frac{1}{2}f_{nm}r_{nm}^2 \quad \text{if} \quad |\mathbf{r}_n^0-\mathbf{r}_m^0|\le R \tag{16}$$

where r_{nm} is the distance increment from the reference position of atoms n and m ($r_{nm}=|\mathbf{r}_n-\mathbf{r}_m|-|\mathbf{r}_n^0-\mathbf{r}_m^0|$), f_{nm} is the force (or stiffness) constant of the corresponding spring, and R is a distance cutoff to neglect long-range interactions. The spring constants are typically assumed to be the same for all interacting pairs ($f_{nm}=k$), but they can be tuned to improve the predictions in proteins (Orellana *et al*., 2010) and nucleic acids (Setny and Zacharias, 2013). The distance cutoff is typically set from 5 to approximately

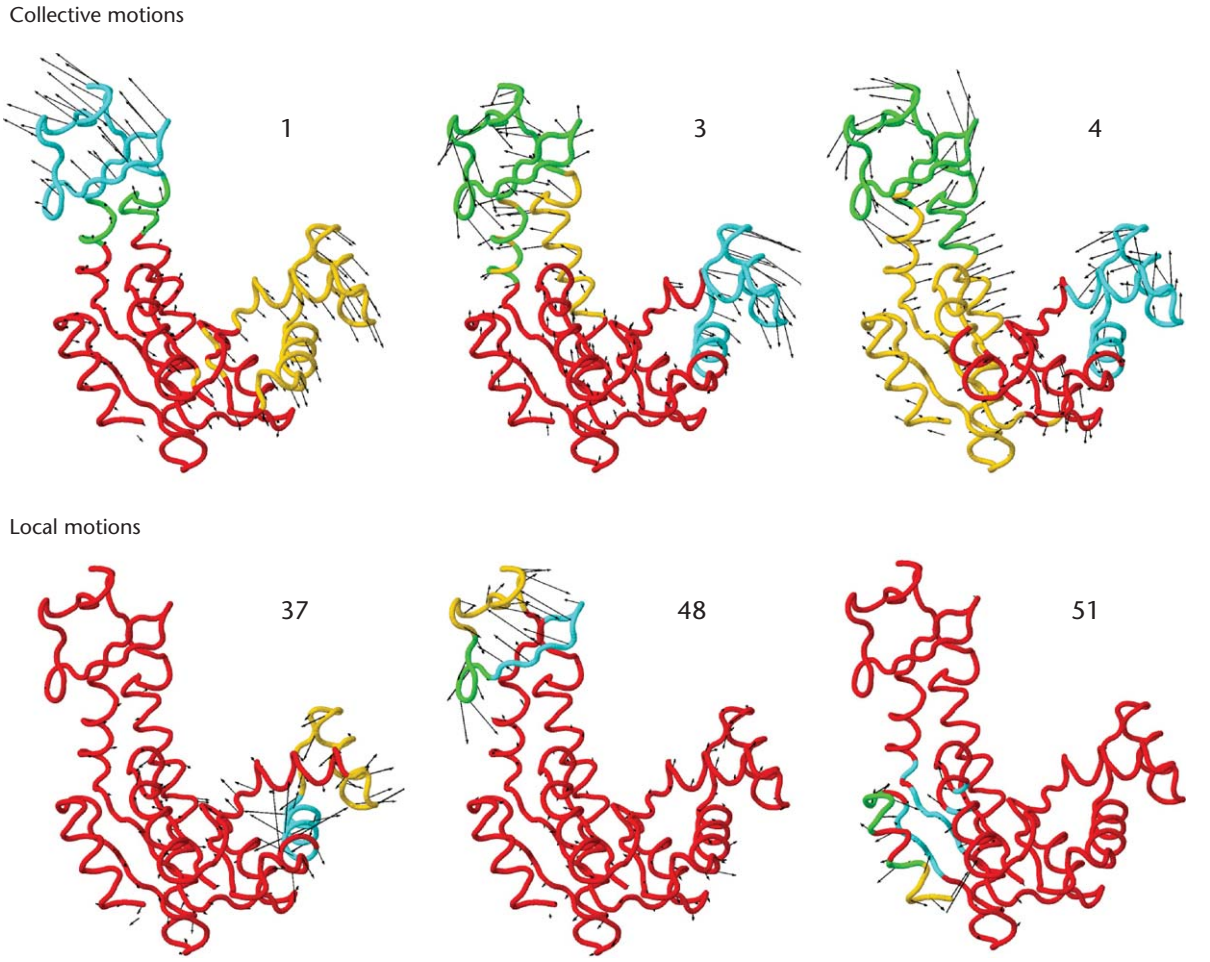

Figure 3 Normal mode analysis applied to a biological macromolecule. Representative atomic displacements corresponding to collective and local motions are shown for the adenylate kinase protein (chain A from PDB ID 4ake), the enzyme that catalyses the reaction: 2ADP ↔ ATP+AMP. Numbers are the mode indices sorted from low to high frequencies. Whereas the arrows represent the direction and relative amplitude of the motions, the different colours indicate the regions that are moving together. In the lowest frequency normal modes (collective motions), large groups of atoms experience a concerted motion, whereas in higher-frequency modes (local modes), only some small groups are moving together (local motions). It is worth noting that the arrow lengths have been normalised for visualisation purposes, otherwise local motions should be smaller. These images were generated with the iMODS server (López-Blanco *et al.*, 2014) (http://imods.chaconlab.org).

12 Å depending on the case. Notably, Tirion's potential energy function is, by definition, at the energy minimum for any chosen input conformation. Thus, in practice, NMA can be performed directly on crystallographic or NMR structures without any prior minimisation step.

The idea of combining NMA and coarse-grained (CG) representations was further refined, validated, and extended by several research groups (Bahar *et al.*, 2010b; Bahar and Rader, 2005; Hinsen *et al.*, 2000; Yang *et al.*, 2009b). One of the simplest and best performing CG models reduces each amino acid to a single pseudo-atom (or bead) located at the Cα position. These coarse-grained representations are critical for the study of large biological molecules because they effectively reduce the size of the $3N \times 3N$ Hessian matrix (where N is the number of pseudo-atoms), leading to a dramatic improvement in computational efficiency. Excellent agreement with experimental data has been obtained with this Cα model (Atilgan *et al.*, 2001; Tama and Sanejouand, 2001) or with even coarser models (Bahar *et al.*, 2010b).

Inspired by the ENM, but based on the thermodynamic theory of random networks of polymers, Bahar developed the Gaussian network model (GNM) where the atoms experience isotropic fluctuations according to a Gaussian distribution (Bahar *et al.*, 1997). In this case, the potential energy is defined as:

$$U(\mathbf{r}) = \sum_{n<m}^{N} \frac{1}{2} f_{nm} (\Delta\mathbf{r}_{nm} - \Delta\mathbf{r}^0_{nm}) \cdot (\Delta\mathbf{r}_{nm} - \Delta\mathbf{r}^0_{nm}) \quad \text{if} \quad |\mathbf{r}^0_n - \mathbf{r}^0_m| \leq R \tag{17}$$

where $\Delta\mathbf{r}_{nm}$ is the difference vector between the position of atoms n and m ($\Delta\mathbf{r}_{nm} = \mathbf{r}_n - \mathbf{r}_m$). Consequently, the GNM

potential penalises not only the changes in the interatomic distance (r_{nm}), such as in Tirion's model, but also any change in the direction of the interatomic vector ($\Delta\mathbf{r}_{nm}$). In contrast with NMA modes that also retain the directional information, the GNM modes only contain information about the magnitude of the fluctuations. However, the isotropic assumption of GNM effectively reduces the Hessian matrix size to $N \times N$, thus improving the efficiency of studies where the directional information is not required. As in the ENM case, GNM has been used to successfully predict the global dynamics of a variety of macromolecular complexes.

Normal Mode Computations

The solution of the standard eigenvalue problem (eqn (12)), that is, the diagonalisation of the Hessian, constitutes the main computational bottleneck in NMA methods. Furthermore, it has been prohibitive for large systems given that the computational cost scales as N^3. This has limited studies to proteins of approximately 300 amino acids until the early 1990s. Improvements in the NMA formulations to effectively reduce the number of variables, such as using the dihedral angles as internal coordinates (Levitt *et al*., 1985; Noguti and Go, 1983), the rotation translation of blocks (RTB) (Tama *et al*., 2000), or the explicit consideration of symmetry based on group theory (Van Vlijmen and Karplus, 2005) enable the NMA of very large biological molecules in a very short amount of time. In addition, new more efficient computational diagonalisation techniques also contribute to extending the applicability of NMA (López-Blanco *et al*., 2013).

Normal Modes are Properties of the Shape of Biomolecules

There exist many studies evidencing that the collective motions encoded in the low-frequency modes from different CG-ENMs effectively characterise biologically relevant conformational changes. The high accordance between these results strongly suggests that low-frequency normal modes are predominantly a property of the shape of the molecular system (Tama and Brooks, 2006). The idea that the characteristics of these low-frequency normal modes are mainly caused by the properties of the shape of the biological molecule is very intriguing. If this argument is true, it would mean that an atomic representation of the biological molecule is not needed to obtain its collective dynamical properties, rather only its shape would be necessary. For example, it has been demonstrated that the lowest frequency normal modes obtained from a low-resolution density map, where atoms cannot be distinguished, agree very well with those computed from the corresponding structure at atomic resolution (Chacón *et al*., 2003). Furthermore, the low-frequency modes are particularly robust to changes in the potential energy function (Lu and Ma, 2005), the CG model (López-Blanco *et al*., 2011), and sequence variations (Zheng *et al*., 2006). Moreover, flexibility profiles of homologous proteins are conserved at family and superfamily levels, even for pairs of proteins with nonsignificant sequence similarity (Maguid *et al*., 2006). All these findings suggest that macromolecular machines have evolved to adopt a specific shape that favours of the biological function.

Applications of Normal Mode Analysis to Structural Biology

Interesting insights into the mechanical properties of the molecules at extended time scales can be obtained very quickly using NMA in contrast to the more demanding MD simulations. Many studies are indicative of the impact of NMA in structural biology, especially for large biological systems. In the next sections, we comment on examples of illustrative applications of NMA for the characterisation of macromolecular flexibility, the prediction of collective conformational changes, and the interpretation of structural experimental data. In **Table 1**, several freely available NMA tools for solving these important problems have been summarised.

Normal modes as predictive tools

The exploration of the normal modes from a single atomic structure can yield insights, at an atomic level, into the fluctuations of macromolecular complexes and the mechanisms of the large-scale rearrangements that occur upon binding to ligands or to other macromolecules (Bahar *et al*., 2010a; Bahar *et al*., 2010b). The mobile or static regions can be directly estimated using the relative amplitudes of the thermal fluctuations predicted by NMA theory (Cui and Bahar, 2010). Furthermore, rigid, flexible, and hinge regions can be inferred, taking into account the directionality (Kovacs *et al*., 2004) or the covariance (Flores *et al*., 2008) of the atomic motions. These predictions are highly correlated with the experimental thermal fluctuations provided by MD simulations (Rueda *et al*., 2007) and by crystallography or NMR (Yang *et al*., 2009a).

The motions between two distinct experimental structures observed in the Protein Data Bank lie mostly in the direction of the two lowest frequency modes (Krebs *et al*., 2002). Thus, conformational transition trajectories, that is, the feasible pathways connecting two distinct atomic structures, can be generated using a linear combination of some of the lowest frequency normal modes (see **Figure 4**). Although the trajectory structures are only representatives of possible intermediates, they provide tentative models that are useful for a better understanding of the functional transitions and can be used as initial models for further modelling approaches.

Table 1 Several NMA-based free programs and servers

Software	URLs	Comments
Bahar's	http://www.csb.pitt.edu/Faculty/bahar/index.php	*ProDy*: free library for NMA (ANM, GNM) and PCA of proteins *NMWiz*: NMA and PCA plugin for VMD *oGNM*: GNM server for proteins and nucleic acids *ANM*: interactive ANM sever for proteins *coMD*: hybrid MD and NMA method to generate transition paths
Chacon's	http://chaconlab.org	*DFprot*: interactive NMA server for proteins *iMODS*: interactive NMA and transition path generation server in dihedral coordinates for proteins and nucleic acids *iMOD*: NMA and transition path generation in dihedral coordinates for proteins and nucleic acids *iMODFIT*: flexible fitting of protein and nucleic acid structures into EM maps in dihedral coordinates
ElNémo	http://www.igs.cnrs-mrs.fr/elnemo	Noninteractive NMA server for proteins and nucleic acids
FlexServ	http://mmb.irbbarcelona.org/FlexServ	Interactive NMA server for proteins
Gerstein's	http://molmovdb.org	*MolMovDB*: database of macromolecular movements and noninteractive NMA server for proteins *StoneHinge*: hinge prediction of proteins
HingeProt	http://www.prc.boun.edu.tr/appserv/prc/hingeprot	Interactive NMA server for hinge prediction of proteins
Hinsen's	http://dirac.cnrs-orleans.fr/plone/software	*MMTK*: free library for molecular modelling, including NMA *Domain Finder*: interactive NMA-based program to characterise the dynamical properties of protein domains *DensityFit*: flexible fitting of atomic structures into EM maps
KOSMOS	http://bioengineering.skku.ac.kr/kosmos	Interactive NMA and transition path generation server for proteins and nucleic acids
NMSim	http://cpclab.uni-duesseldorf.de/nmsim	Interactive server to generate transition paths using NMA-based geometric simulations
NOMAD-Ref	http://lorentz.immstr.pasteur.fr/nma/submission.php	Noninteractive NMA server for proteins and nucleic acids
PARS	http://bioinf.uab.cat/pars	Prediction of protein allosteric and regulatory sites
ProMode	http://promode.pdbj.org/promode_elastic/index.do	Pre-computed interactive animations of normal modes in dihedral coordinates and other NMA-based results
SPACER	http://allostery.bii.a-star.edu.sg	Analyse allosteric communication between different sites
TMM@	http://services.cbu.uib.no/tools/tmma	NMA server for the analysis of trans-membrane α-helices
WebNM@	http://apps.cbu.uib.no/webnma/home	Interactive NMA server for proteins
Zheng's	http://enm.lobos.nih.gov	Several servers based on NMA or elastic networks: *AD-ENM*: interactive NMA server for proteins and nucleic acids *DC-ENM*: builds atomic models satisfying distance constraints *PATH-ENM* and *iENM*: generate transition paths *EMFF*: flexible fitting server of atomic structures into EM maps

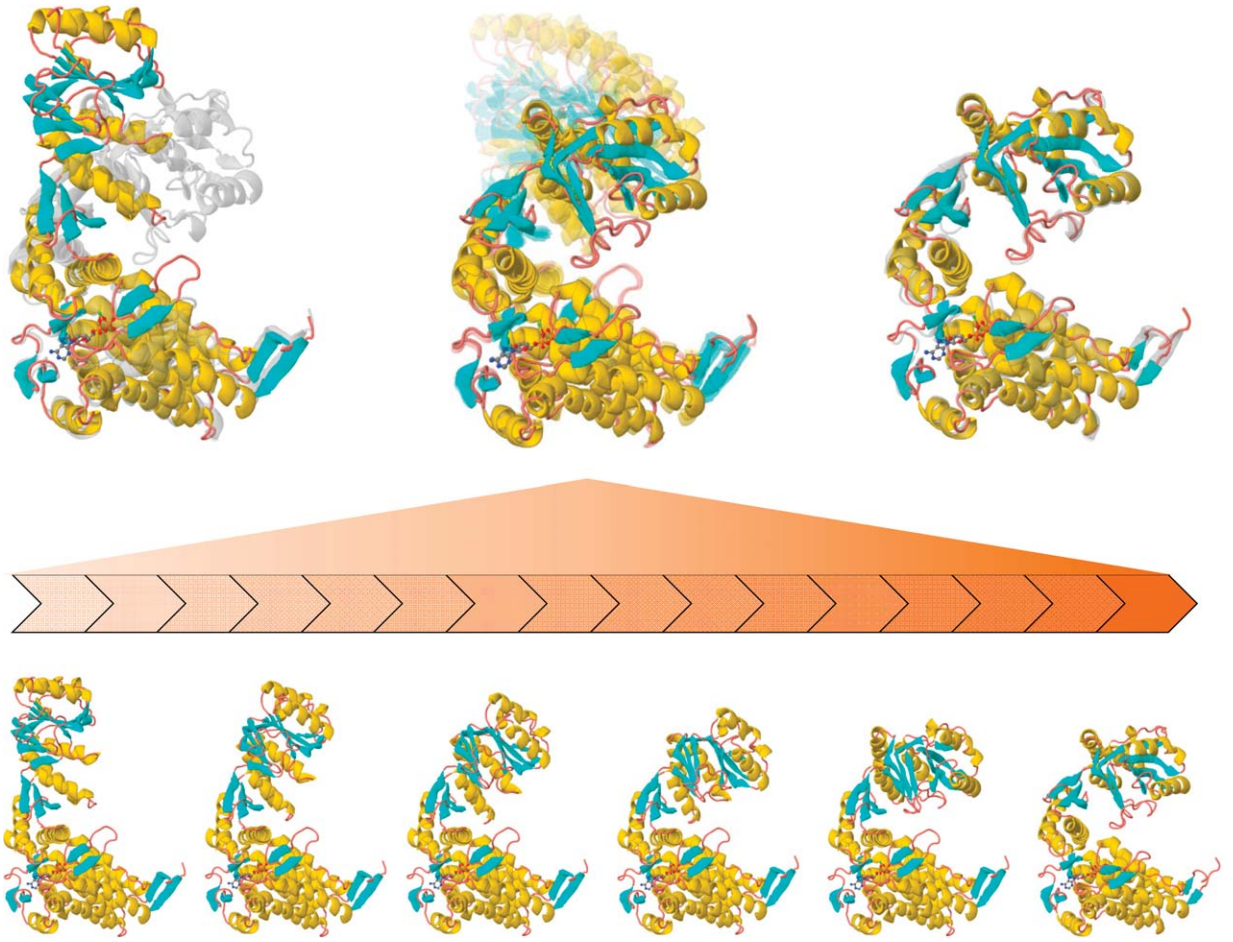

Figure 4 Conformational open-to-closed transition pathway of the GroEL protein based on NMA. Starting from the open monomer (chain A from PDB ID 1sx4) (coloured structure at top left corner), a combination of low-frequency modes were used iteratively to generate the intermediate structures (middle and bottom rows). Only those modes that reduced the differences with the closed conformation (chain H from PDB ID 1sx4) (grey) were employed. This transition was generated with the morphing tool of iMODS server (http://imods.chaconlab.org).

NMA also represents a promising alternative for modelling flexibility in macromolecular docking (Zacharias, 2010). Macromolecular docking is a computational technique for predicting how two or more interacting biomolecules can form a stable complex from the unbound structures. In any of the docking variants, that is, protein–protein, protein–ligand, or protein–nucleic acid, the accurate modelling of the complex is a difficult problem, where the major challenge is in dealing with molecular flexibility. Fortunately, only a few low-frequency normal modes are required to describe about one third of the conformational changes experienced upon association (Stein *et al*., 2011). These modes have been successfully applied either to generate conformational ensembles or to directly include flexibility in docking simulations (Meireles *et al*., 2011).

In allosteric regulation, the union of an effector molecule to an enzyme usually leads to conformational changes in the active site that modulate its activity. The combination of the collective fluctuations predicted by the GNM with graph and information theories permits the identification of such active sites (e.g. catalytic or metal-binding residues) (Eyal *et al*., 2011). These NMA predicted key sites seem to be highly prone to efficient communication with the rest of the structure as evidenced by the small number of steps needed to transmit information to any other residue.

Normal modes are key to interpreting experimental data

New computational techniques for NMA have also opened ways to complement structural data from different experimental sources, from which atomic models cannot be directly constructed or refined. Mainly, normal mode vectors can be used as search directions to mimic protein dynamics and achieve a better fit to the experimental data. Early applications of NMA used the normal modes to refine the crystallographic B-factors obtained from X-ray crystallography. NMA was also employed to improve the molecular replacement technique used in X-ray

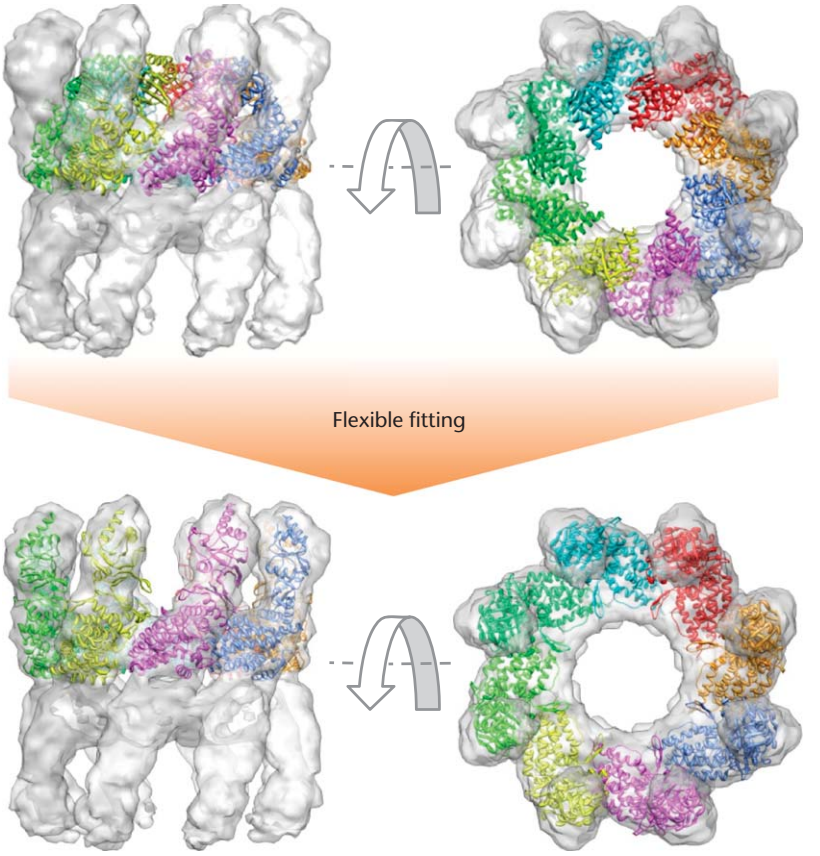

Figure 5 Flexible fitting of an atomic structure into a low-resolution density map based on NMA. The high-resolution structure of the thermosome (PDB ID 1a6d) (rainbow colours) has been flexibly fitted into a low-resolution cryo-Electron Microscopy density map (EMDB 1396) (grey transparency) with iMODFIT (Lopéz-Blanco and Chacón, 2013) by using the low-frequency modes to maximise the overlap between the map and the structure.

crystallography. In this technique, structures of unknown macromolecules are modelled using data from known structural templates. Given the conformation found in the crystal does not exactly match the conformation of the templates, several candidate structures can be first generated from NMA and then evaluated against crystallographic data to obtain possible solutions for molecular replacement. More recently, new structural refinement methodologies have also benefited from the use of NMA in conjunction with lower resolution structural information such as cryo-electron microscopy (Lopéz-Blanco and Chacón, 2013; Tama *et al*., 2004), small angle X-ray scattering (Miyashita *et al*., 2011), fibre diffraction data, and distance constraints. The approach is similar to the computation of conformational transition trajectories described previously, but rather than targeting another atomic structure, the target is defined by the low-resolution structural data. For example, NMA enables the interpretation of new functional states captured only by electron microscopy (EM) from available X-ray structures. In **Figure 5**, the closed atomic structure of the thermosome is flexibly fitted into a cryo-EM map in an open conformation using only the lowest frequency modes. This and other similar NMA-based approaches open up new ways for the atomic-level interpretation of large conformational changes and their functional implications.

NMA also provides a reasonable description of the macromolecular mechanical responses that contribute to the interpretation of single-molecule experiments and elucidates the relationship between mechanical stability and biological function. For example, it has been shown that the effective stiffness calculated from NMA correlates well with the force required to unfold the protein using single-molecule manipulation techniques (Eyal *et al*., 2011). Finally, normal modes can be used to predict conformational changes by matching experimental distance constraints from fluorescence or NMR (Zheng and Brooks, 2006).

Limitations of Normal Mode Analysis

Despite of the usefulness of NMA for modelling macromolecular flexibility, the underlying harmonic approximation leads to important limitations (Ma, 2005). NMA fails to effectively predict the absolute time scale and amplitude of the motions, mainly as a consequence of the anharmonicities imposed by the solvent and the multiple energy barriers and minima of the energy landscape. The

refolding events and other local rearrangements are poorly predicted by NMA because they require large displacements that are too far from equilibrium. On the contrary, large collective motions, such as hinge or shear movements of domains, correspond to minor rearrangements in the atomic neighbourhood that can be well captured by the harmonic approximation. In any case, significant distortions in the covalent structure and steric clashes may appear when the normal modes are animated with too large amplitudes. This is mainly a consequence of the straight line trajectories described by the Cartesian coordinates modes. To minimise such distortions, the covalent structure can be either explicitly regularised by adjusting the covalent geometry or implicitly preserved by using the dihedral angles as internal coordinates in the NMA formulation (Lopéz-Blanco *et al*., 2011).

Conclusion

NMA is a very powerful method that has shown its proficiency in analysing and studying the dynamics of large biological molecules. Although NMA has some limitations for studying specific biological problems, due to its simplistic harmonic approximation, it represents a popular and very efficient alternative to other costly techniques for modelling collective and large-amplitude motions. We encourage the reader to take a look at the free resources provided in **Table 1** to experience the usefulness of NMA in structural biology.

References

Atilgan AR, Durell SR, Jernigan RL *et al*. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal* **80**: 505–515.

Bahar I, Atilgan AR and Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design* **2**: 173–181.

Bahar I, Lezon TR, Bakan A and Shrivastava IH (2010a) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chemical Reviews* **110**: 1463–1497.

Bahar I, Lezon TR, Yang LW and Eyal E (2010b) Global dynamics of proteins: bridging between structure and function. *Annual Review of Biophysics* **39**: 23–42.

Bahar I and Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. *Current Opinion in Structural Biology* **15**: 586–592.

Bastolla U (2014) Computing protein dynamics from protein structure with elastic network models. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**(5): 488–503.

Brooks B and Karplus M (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences of the USA* **80**: 6571–6575.

Chacón P, Tama F and Wriggers W (2003) Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. *Journal of Molecular Biology* **326**: 485–492.

Cui Q and Bahar I (2010) *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*. Boca Raton, FL: Chapman & Hall/CRC.

Eyal E, Dutta A and Bahar I (2011) Cooperative dynamics of proteins unraveled by network models. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**: 426–439.

Flores SC, Keating KS, Painter J *et al*. (2008) HingeMaster: normal mode hinge prediction approach and integration of complementary predictors. *Proteins: Structure, Function and Genetics* **73**: 299–319.

Go N, Noguti T and Nishikawa T (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proceedings of the National Academy of Sciences of the USA* **80**: 3696–3700.

Goldstein H, Poole CP and Safko JL (2002) *Classical Mechanics*, 3rd edn. San Francisco, CA: Addison-Wesley.

Hinsen K, Petrescu AJ, Dellerue S, Bellissent-Funel MC and Kneller GR (2000) Harmonicity in slow protein dynamics. *Chemical Physics* **261**: 25–37.

Ingólfsson HI, Lopez CA, Uusitalo JJ *et al*. (2014) The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**: 225–248.

Kovacs JA, Chacón P and Abagyan R (2004) Predictions of protein flexibility: first-order measures. *Proteins: Structure, Function and Genetics* **56**: 661–668.

Krebs WG, Alexandrov V, Wilson CA *et al*. (2002) Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins: Structure, Function and Genetics* **48**: 682–695.

Levitt M, Sander C and Stern PS (1985) Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of Molecular Biology* **181**: 423–447.

López-Blanco JR, Aliaga JI, Quintana-Ortí ES and Chacón P (2014) iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Research* **42**: W271–W276.

Lopéz-Blanco JR and Chacón P (2013) iMODFIT: Efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *Journal of Structural Biology* **184**: 261–270.

Lopéz-Blanco JR, Garzón JI and Chacón P (2011) iMod: Multipurpose normal mode analysis in internal coordinates. *Bioinformatics* **27**: 2843–2850.

López-Blanco JR, Reyes R, Aliaga JI *et al*. (2013) Exploring large macromolecular functional motions on clusters of multicore processors. *Journal of Computational Physics* **246**: 275–288.

Lu M and Ma J (2005) The role of shape in determining molecular motions. *Biophysical Journal* **89**: 2395–2401.

Ma J (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **13**: 373–380.

Maguid S, Fernández-Alberti S, Parisi G and Echave J (2006) Evolutionary conservation of protein backbone flexibility. *Journal of Molecular Evolution* **63**: 448–457.

Meireles L, Gur M, Bakan A and Bahar I (2011) Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Science* **20**: 1645–1658.

Miyashita O, Gorba C and Tama F (2011) Structure modeling from small angle X-ray scattering data with elastic network normal mode analysis. *Journal of Structural Biology* **173**: 451–460.

Noguti T and Go N (1983) Dynamics of native globular proteins in terms of dihedral angles. *Journal of the Physical Society of Japan* **52**: 3283–3288.

Orellana L, Rueda M, Ferrer-Costa C *et al*. (2010) Approaching elastic network models to molecular dynamics flexibility. *Journal of Chemical Theory and Computation* **6**: 2910–2923.

Rueda M, Chacón P and Orozco M (2007) Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure* **15**: 565–575.

Setny P and Zacharias M (2013) Elastic network models of nucleic acids flexibility. *Journal of Chemical Theory and Computation* **9**: 5460–5470.

Stein A, Rueda M, Panjkovich A, Orozco M and Aloy P (2011) A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks. *Structure* **19**: 881–889.

Tama F and Brooks CL III (2006) Symmetry, form, and shape: Guiding principles for robustness in macromolecular machines. *Annual Review of Biophysics and Biomolecular Structure* **35**: 115–133.

Tama F, Gadea FX, Marques O and Sanejouand YH (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Structure, Function and Genetics* **41**: 1–7.

Tama F, Miyashita O and Brooks CL III (2004) Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *Journal of Structural Biology* **147**: 315–326.

Tama F and Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. *Protein Engineering* **14**: 1–6.

Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical Review Letters* **77**: 1905–1908.

Van Vlijmen HWT and Karplus M (2005) Normal mode calculations of icosahedral viruses with full dihedral flexibility by use of molecular symmetry. *Journal of Molecular Biology* **350**: 528–542.

Yang L, Song G and Jernigan RL (2009a) Comparisons of experimental and computed protein anisotropic temperature factors. *Proteins: Structure, Function and Bioinformatics* **76**: 164–175.

Yang L, Song G and Jernigan RL (2009b) Protein elastic network models and the ranges of cooperativity. *Proceedings of the National Academy of Sciences of the USA* **106**: 12347–12352.

Zacharias M (2010) Accounting for conformational changes during protein-protein docking. *Current Opinion in Structural Biology* **20**: 180–186.

Zheng W and Brooks BR (2006) Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes. *Biophysical Journal* **90**: 4327–4336.

Zheng W, Brooks BR and Thirumalai D (2006) Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proceedings of the National Academy of Sciences of the USA* **103**: 7664–7669.

Further Reading

Dill KA and Bromberg S (2003) *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*. New York and London: Garland Science.