BIOINFORMATICS ORIGINAL PAPER Vol. 22 no. 21 2006, pages 2619–2627
doi:10.1093/bioinformatics/btl448

Structural bioinformatics

Anisotropic network model: systematic evaluation and a new web interface

Eran Eyal, Lee-Wei Yang and Ivet Bahar*

Department of Computational Biology, School of Medicine, University of Pittsburgh, PA 15213, USA

Received on June 2, 2006; revised on August 1, 2006; accepted on August 16, 2006
Advance Access publication August 23, 2006
Associate Editor: Anna Tramontano

ABSTRACT

Motivation: The Anisotropic Network Model (ANM) is a simple yet powerful model for normal mode analysis of proteins. Despite its broad use for exploring biomolecular collective motions, ANM has not been systematically evaluated to date. A lack of a convenient interface has been an additional obstacle for easy usage.

Results: ANM has been evaluated on a large set of proteins to establish the optimal model parameters that achieve the highest correlation with experimental data and its limits of accuracy and applicability. Residue fluctuations in globular proteins are shown to be more accurately predicted than those in nonglobular proteins, and core residues are more accurately described than solvent-exposed ones. Significant improvement in agreement with experiments is observed with increase in the resolution of the examined structure. A new server for ANM calculations is presented, which offers flexible options for controlling model parameters and output formats, interactive animation of collective modes and advanced graphical features.

Availability: ANM server (http://www.ccbb.pitt.edu/anm)

Contact: bahar@ccbb.pitt.edu

1 INTRODUCTION

1.1 The anisotropic network model

Recent years have seen a renewed interest in normal mode analysis (NMA) of equilibrium structures, with the realization of the potential utility of elastic network (EN) models for deriving functional motions from structure (Cui and Bahar, 2006; Nicholay and Sanejouand, 2006). Anisotropic network model (ANM) is an EN introduced in 2000 (Atilgan *et al*., 2001; Doruker *et al*., 2000), inspired by the pioneering work of Tirion (1996), succeeded by the development of the Gaussian network model (GNM) (Bahar *et al*., 1997; Haliloglu *et al*., 1997), and by the work of Hinsen (1998) who first demonstrated the validity of performing EN NMA at the residue level. The main advantage of NMA and, particularly, EN models over traditional methods such as molecular dynamics is the fast computational time and the ability to predict large-scale fluctuations or the so-called global modes resulting from the collective participation of entire domains or substructures. In the ANM, a uniform force constant γ is adopted for all springs; nodes are identified by the positions of C^{α} atoms, and the overall potential of the system is a sum of harmonic potentials:

$$V = \frac{\gamma}{2\sum_{j|j\neq i}(\mathbf{\Gamma}_{ij})(R_{ij}-R_{ij}^0)^2}, \tag{1}$$

where γ is uniform spring constant, R_{ij} and R_{ij}^0 are the respective instantaneous and equilibrium distances between nodes i and j, $\mathbf{\Gamma}_{ij}$ is the ij-th element of the Kirchhoff matrix $\mathbf{\Gamma}$ of inter-residue contacts (Bahar *et al*., 1997), equal to 1 if nodes i and j are within an interaction cutoff distance r_c, zero otherwise. $\mathbf{\Gamma}$ provides a complete description of inter-residue contact topology in the folded state. The Hessian matrix $\boldsymbol{H}$ for a network of N nodes is a $3N \times 3N$ matrix composed of $N \times N$ super elements, $\boldsymbol{H}_{ij}$ $(i \neq j)$ of the form

$$\boldsymbol{H}_{ij} = \frac{\gamma\mathbf{\Gamma}_{ij}}{(R_{ij}^0)^2}\begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix} \tag{2}$$

found from the second derivatives of V with respect to residue positions. Here, X_{ij}, Y_{ij} and Z_{ij} are the components of the distance vector $\boldsymbol{R}_{ij}^0$. The diagonal super elements of $\boldsymbol{H}$ are 3×3 submatrices

$$\boldsymbol{H}_{ii} = -\sum_{j|j\neq i}\boldsymbol{H}_{ij}. \tag{3}$$

Information regarding the mean-square fluctuations of individual residues and correlations between their fluctuations are conveyed by the $3N \times 3N$ covariance matrix $\boldsymbol{C}$ of the multivariate Gaussian distribution. The covariance matrix is proportional to the inverse of $\boldsymbol{H}$. $\boldsymbol{H}$ is not, however, invertible, by definition [Equation (3)]. A pseudo-inverse is constructed instead, based on its $3N-6$ nonzero eigenvalues λ_i and corresponding eigenvectors $\boldsymbol{u}_i$ as

$$\boldsymbol{H}^{-1} = \sum_{i=1}^{3N-6}\frac{1}{\lambda_i}\boldsymbol{u}_i\boldsymbol{u}_i^{\mathrm{T}}. \tag{4}$$

The inverse $\boldsymbol{H}^{-1}$ is also organized in $N \times N$ submatrices of size 3×3, each. The ij-th submatrix $\boldsymbol{H}_{ij}^{-1}$ defines the covariance between the fluctuations of residues i and j. The cross-correlation $\langle \Delta\boldsymbol{R}_i \bullet \Delta\boldsymbol{R}_j \rangle$ between the equilibrium fluctuations of residues i and j is expressed in terms of the trace (tr) of these submatrices as

$$\boldsymbol{C}_{ij} \equiv \langle \Delta\boldsymbol{R}_i \bullet \Delta\boldsymbol{R}_j \rangle = \frac{\mathrm{tr}(\boldsymbol{H}_{ij}^{-1})}{\sqrt{\mathrm{tr}(\boldsymbol{H}_{ii}^{-1})\ \mathrm{tr}(\boldsymbol{H}_{jj}^{-1})}}. \tag{5}$$

*To whom correspondence should be addressed.

Despite their simplicity, coarse-grained NMA with ANMs proved in many applications to be a promising tool for describing the collective dynamics of a wide range of biomolecular systems (Cui and Bahar, 2006; Bahar and Rader, 2005; Chennubhotla *et al*., 2005). Recent notable applications include the studies of hemoglobin (Chunyan *et al*., 2003), influenza virus hemagglutinin A (Isin *et al*., 2002), tubulin (Keskin *et al*., 2002), HIV-1 reverse transcriptase complexed with different inhibitors (Temiz and Bahar, 2002), HIV-1 protease (Micheletti *et al*., 2004; Vincenzo *et al*., 2006), DNA polymerase (Delarue and Sanejouand, 2002), motor proteins (Zheng and Brooks, 2005a,b; Zheng and Doniach, 2003), membrane proteins including potassium channels (Shrivastava and Bahar, 2006), rhodopsin (Rader *et al*., 2004), nicotinic acetylcholine receptor (Hung *et al*., 2005; Taly *et al*., 2005) and mechanosensitive channel MscL (Valadie *et al*., 2003), or supramolecular structures such as the ribosomal complex (Tama *et al*., 2003; Wang *et al*., 2004) and viral capsids (Rader *et al*., 2005; Tama and Brooks, 2005). Other coarse-grained EN models were also introduced and successfully used in NMA: some group atoms or residues into rigid blocks [e.g. rigidly rotating/translating blocks (RTB) method (Durand *et al*., 1994; Tama *et al*., 2000; Hinsen, 1998; Li and Cui, 2002; Tama *et al*., 2000)], others take account of the positions of β-carbons in addition to α-carbons (Micheletti *et al*., 2004). A distance-dependent form has been introduced by Hinsen (1998) for inter-residue force constants, the extensive use of which by Gerstein and co-workers (Alexandrov *et al*., 2005; Krebs *et al*., 2002) has provided evidence for the correlation between the predicted (by coarse-grained NMA) and observed (from experiments) structural transitions. In the present study, we examine on a database scale the performance of ANM as a function of its parameters and provide an assessment of the accuracy of prediction for different residue types and secondary structures.

1.2 Implementation of ANM server

Implementation of user-friendly programs and web interfaces is of crucial importance to the adequate usage and application of computational models, especially by nonexpert users. It is not trivial to take output data from NMA, usually in the form of a set of eigenvectors and eigenvalues, and to gain an understanding of the dynamics of the system. A number of web servers have indeed been implemented to permit automated NMA of biomolecules.

At the atomic level, the Vibrations Evaluation Server (MoViES, http://ang.cz3.nus.edu.sg/cgi-bin/prog/norm.pl; Cao *et al*., 2004) offers a detailed NMA. The calculations may, however, be slow and the results are sent by email. The users can obtain information on the free energy, kinetic and potential energies of individual residues, along with the frequencies and shapes of each mode. No graphical interface exists to assist in the analysis of the results.

NOMAD-REF is a new server for atomic NMA (Lindahl *et al*., 2006). Coordinate files can be saved and later viewed by other applications. Graphs of atomic fluctuations in different modes are accessible. The major strength of this site is the implementation of the NMA methodology as a tool for timely tasks in structural biology including refinement and optimization of docked complexes. The server is capable of calculating the overlap of the eigenvectors and eigenvalues describing the dynamics of two different conformations.

The ProMod database (Wako *et al*., 2004) offers results from NMA for a fairly large collection of proteins (1900 of them). In contrast to other servers that use Cartesian coordinates, the motions are derived in the space of dihedral angles, subject to fixed bond lengths and bond angles, which permits to reduce the number of degrees of freedom, while retaining backbone atomic structural data. Results are animated using Chime plug-in. Other graphical options for analyzing the fluctuations are available.

A server that performs NMA at residue level is WEBnm@ (http://www.bioinfo.no/tools/normalmodes; Hollup *et al*., 2005). WEBnm@ calculates the slowest six nonzero mode shapes and associated deformation energies using the MMTK package (Hinsen, 2000). Visualization includes gif animated movies.

AD-ENM (http://enm.lobos.nih.gov), additional simplified ENM tool, allows for calculation of up to 20 modes and permits to calculate the contribution of each mode to the conformational change between two given structures.

ElNemo (http://igs-server.cnrs-mrs.fr/elnemo/index.html; Suhre and Sanejouand, 2004) is another server that performs residue level calculations. A methodology similar to ANM is adopted therein, with the major difference being the use of the RTB model. This approximation allows for analyzing relatively large proteins, although the calculations may still take up to several hours. Outputs include mean-square fluctuations, cross-correlations, degree of collectivity and an option to derive the normal modes responsible for the conformational changes between known structures.

Finally, within the frame of the DB of macromolecular movements (MolMovDB, http://molmovdb.org/; Alexandrov *et al*., 2005) which also uses the MMTK package (Hinsen, 2000), up to five slowest mode shapes are presented as static or animated gif files and as snapshots of coordinates.

Despite the large number of online tools for NMA, the graphical options to visualize the results—the most important feature for many users–are lacking. Most of the above listed servers offer only animated gif formats, typically very limited in resolution and smoothness of motions. The user cannot freely rotate, translate or zoom the molecule and is unable to vary the size and frequency of vibrations for visualization purposes. Some of the servers such as WEBnm@ and NOMAD-REF provide the option to download files that can be interactively visualized in other applications, VMD (Humphrey *et al*., 1996) and PyMol (DeLano, 2002; http://www.pymol.org), respectively. This, however, is feasible only for one mode at a time, and requires other applications to be installed and knowledge to utilize them. ProMod offers interactive online animation based on Chime, but this is available only for a limited number of structures in the DB and without the options to control the model parameters.

We have recently developed (Yang *et al*., 2005) a database for analyzing protein dynamics, based on the GNM (Bahar *et al*., 1997), which includes GNM results for all Protein Data Bank (PDB) structures (http://ignm.ccbb.pitt.edu/; Yang *et al*., 2005), as well as an online computation server, *o*GNM (Yang *et al*., 2006) for analyzing structural models submitted in the PDB format. Both tools allow for graphical analysis of the size of fluctuations and their cross-correlations. The mean-square fluctuations of residues and their mobilities in different modes can be color-mapped onto ribbon diagrams using either Chime or Jmol (http://jmol.sourceforge.net/). No movies are provided, inasmuch as GNM does not convey information on the directions of fluctuations, but their size.

In this study, we present a large-scale evaluation of the performance of ANM in different systems and within different environments (solvent exposure, secondary structures, amino acids). We examine different interaction cutoffs and investigate the dependence of the spring constant on the interaction distance. Having established the optimal model parameters and the limits of accuracy for different systems, we present a new server developed for ANM calculations, the main strengths of which are the rapid computing ability and the user-friendly graphical capabilities for analyzing and interpreting the outputs.

2 METHODS

2.1 Dataset for statistical analysis

The test set of proteins was extracted from the January 2006 version of the PDB (Berman *et al*., 2000) using PDB-REPRDB (Noguchi and Akiyama, 2003) to include all crystal structures with resolution better than 2.5 Å, R-factor <0.3 Å and no missing coordinates. Redundancy filters were applied at both the sequence level (each protein in the final set shares <15% identity with all others) and the structural level (structures differ by at least 15 Å root-mean-square deviation in their backbone coordinates). The list was further filtered to ensure that proteins crystallized as monomer are indeed the functional unit based on the biological unit data reported in the PDB. The final set includes 176 proteins, the PDB codes of which are listed in http://ignmtest.ccbb.pitt.edu/anmdocs/anm_test_set.txt.

2.2 Solvent accessibility, globularity and assignment of secondary structures

Solvent accessibilities were calculated using the Voronoi polyhedra procedure (McConkey *et al*., 2002). The accessibility of residue X is defined as the ratio of its total solvent-accessible surface in the native structure to that in the peptide GGXGG with the same backbone conformation.

As a quantitative measure of the protein shape, we define a globularity index as the fraction of residues with solvent accessibility <0.16. This threshold is the median value for accessibility in globular proteins. Finally, secondary structures are assigned using DSSP (Kabsch and Sander, 1983).

2.3 Evaluation of the model

The ANM is evaluated by comparing the experimental fluctuation (B-factors) reported in the PDB file to ANM predictions

$$B_i^{\mathrm{ANM}} = \frac{8\pi^2 k_{\mathrm{B}} T}{3\gamma} \mathrm{tr}(H_{ii}^{-1}), \qquad (6)$$

where k_{B} is the Boltzmann constant and T is the Kelvin temperature. This is a standard way to evaluate ANM predictions. It should be noted that we resort to this comparison because available X-ray structures in the PDB usually contain B-factors data that may be readily compared with ANM predictions. The major utility of ANM calculations or NMAs with EN models in general, is not, however, to compute the mean-square fluctuations of amino acids, which can be obtained from experiments (or computed by simpler methods such as the GNM). Instead, the most useful feature of present calculations is to predict the most cooperative and functional directions of motions (usually driven by the low frequency modes). These are not necessarily observed by direct measurements but require comparisons of two or more PDB structures if available, as performed in previous studies on coarse-grained NMAs (Cui and Bahar, 2006; Krebs *et al*., 2002). We anticipate the ANM server to facilitate such analyses and provide insights in the absence of experimental data.

We applied two measures to compare theoretical and experimental fluctuations data. For the evaluation of entire structures' dynamics we used the correlation coefficient (CC) between the two sets of data. In this case there is no need to know the value of the spring constant γ. For integrating data from different proteins, we resorted to mean (absolute) differences between experimental and scaled theoretical values. In this case, the value of the per-protein spring constant γ is required and deduced from the ratio between the observed and the theoretical values of all residues in the protein. The goal while using this measure was to compare predictions under different conditions and not to perform an objective evaluation.

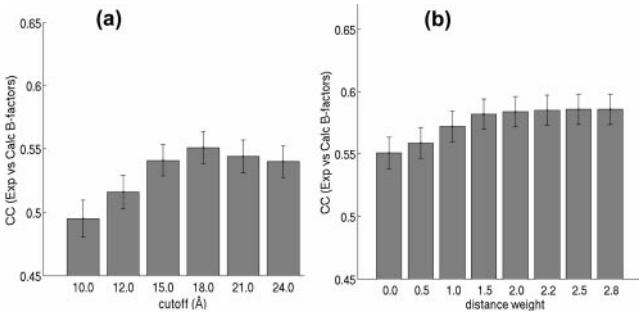

Fig. 1. (**a**) Average correlation between ANM-predicted mean-square fluctuations and the experimental values (B-factors) for different cutoff distances. The ordinate is the average correlation coefficient (CC) between the computed and measured data for the examined 176 PDB structures, and the bars indicate the standard errors for each set. (**b**) Dependence of the same CC on the power p used in Equation (8) for the distance dependence on the spring constant.

3 RESULTS

3.1 Model evaluation

The ANM was evaluated by comparing the fluctuations predicted from theory and those experimentally observed (B-factors deposited in the PDB). All proteins tested are monomers according to their crystal arrangements and have all coordinates resolved by using X-ray crystallography. Details regarding the protein set can be found in Section 2.

The first basic step is to examine the dependence of the performance of the model on its single parameter. Figure 1a shows a model that is quite robust to the cutoff distance in the range of 15–24 Å, with a CC between experiments and theory of ~0.54. The insensitivity of the GNM to the choice of cutoff distances within a certain range has been shown in previous works (Kundu *et al*., 2002; Yang, *et al*., 2006). The present results suggest that the same insensitivity is valid for ANM results too. Note that a larger cutoff distance is required in the ANM, compared with GNM, as also discussed in more detail in our previous work (Atilgan *et al*., 2001) where 7.3 Å cutoff for GNM was shown to be equivalent to 18.0 Å for ANM after calibration with experimental fluctuations. The large cutoff in ANM does not mask the topological differences between residues because the Hessian matrix essentially depends on the orientational cosines of the inter-residue vectors rather than their absolute distances [Equation (2)]. For HIV protease (e.g. PDB 1hhp), the correlation between the fluctuations predicted using $r_{\mathrm{c}} = 18.0$ Å show a CC of 0.89 and 0.78 with those obtained with 10.0 Å and 30.0 Å cutoffs, respectively. Nevertheless, the ANM server (Section 3.6) allows users to interactively select the cutoff over a broad range, so that the robustness of the results may be tested when detailed case studies are conducted.

It should be noted that GNM fluctuation predictions agree better with experiments than those computed with ANM. The CC between

GNM results and experimental data averaged over a large representative set is at least 0.6 under the same evaluation conditions and reach up to 0.74 upon distinguishing between bonded- and nonbonded-pairs and inclusion of crystal contacts (Kondrashov *et al*., 2006). The higher performance of GNM has been attributed to its underlying potential (Chennubhotla *et al*., 2005):

$$V = \frac{\gamma}{2} \sum_{j|j\neq i} (-\mathbf{\Gamma}_{ij})(\boldsymbol{R}_{ij} - \boldsymbol{R}^0_{ij}) \bullet (\boldsymbol{R}_{ij} - \boldsymbol{R}^0_{ij}), \qquad (7)$$

which takes into account the orientational deformations, in addition to distance changes.

3.2 Distance weight for the force constants

From Equation (2) it is clear that the distance between residues comes into play in the Hessian only through the ratios S_{ij}/R^0_{ij} where $S = X$, Y or Z, i.e $\boldsymbol{H}$ depends on the spatial orientations of inter-residue distance vectors, rather than their sizes. On the other hand, it is reasonable to assume that closer residues would be subject to stronger interactions and, thereby, stiffer springs. Hinsen (1998) indeed suggested a force constant that decays with the square distance. We have considered a slightly different distance weighting scheme as described by the new Hessian off-diagonal super elements

$$H'_{ij} = H_{ij}/(R^0_{ij})^p, \qquad (8)$$

where p is an empirical parameter. Figure 1b shows the average CC between theoretical and experimental B-factors obtained using Equation (8) with different P-values. Weighting the interactions improves the correlation gradually up to 0.58 at $P = 2.5$, approximately. We also tried different forms of exponential decays but could not obtain any improvement in the correlation. The results presented in the following refer to a distant weight with $P = 2.5$, unless stated otherwise.

3.3 Globular proteins are easier targets

ANM depends on the topology of the EN, which in turn describes the overall shape of the protein. It is therefore natural to check the influence of the protein shape on the performance of the model. As described above, we estimated the globularity of the proteins in our test set using a simple index. Figure 3 shows the mean correlation for proteins in four bins of globularity. The prediction is clearly better for more globular proteins with a difference of almost 15% between the extreme bins.

More globular proteins have, by definition, larger fraction of buried residues. The fluctuations predicted for buried residues, would, in turn, be expected to be in better agreement with experiments than solvent-exposed residues. This is because the fluctuation spectrum is more constrained, or well-defined, in a high coordination number environment (protein core), as opposed to surface regions where large amplitude fluctuations can be sampled. In addition, experimental B-factors of surface residues are biased due to crystal packing, making it difficult to achieve a high correlation with theoretical calculations in vacuum. The results shown in Figure 2 can be understood on these grounds.

We also examined directly how the accuracy of the predicted fluctuations is related to solvent accessibilities. Figure 3 demonstrates that the fluctuations of buried residues are slightly underestimated on average, while the theoretical fluctuations of exposed

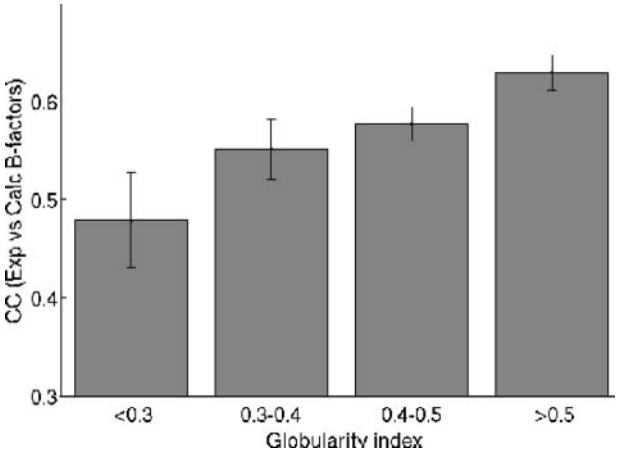

Fig. 2. Dependence of the correlation between ANM predictions and experimental data (B-factors) on the globularity of the examined proteins. Highest correlation is achieved with most globular proteins.

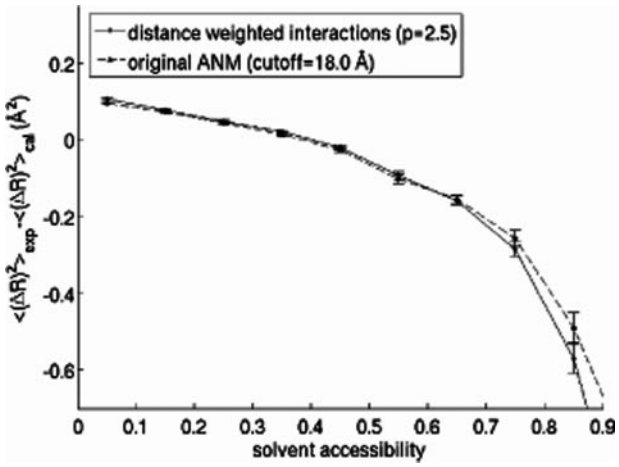

Fig. 3. Influence of solvent accessibility on the difference between experimental and theoretical mean-square fluctuations of residues, computed with two different methods (as labeled). Fluctuations are slightly underestimated in the protein core and overestimated on the surface.

residues are larger than their experimental counterparts at least partially because the fluctuations in the crystal do not reflect real motion in solution/vacuum. In absolute terms, the predictions for buried residues are in significantly better agreement with the experimental data.

3.4 Performance for different amino acids and secondary structures

ANM does not consider the residue types by any means during the calculation. Yet, it is important to evaluate the theoretical fluctuations for each residue. This may provide hints with respect to the limitations or biases of the simple potential in the coarse-grained approach. Interestingly, the model is capable of producing a near native fluctuation pattern for the different amino acids, as shown in Figure 4a. This proves that, in general, the fluctuations are a direct consequence of the architecture of the protein and not an intrinsic

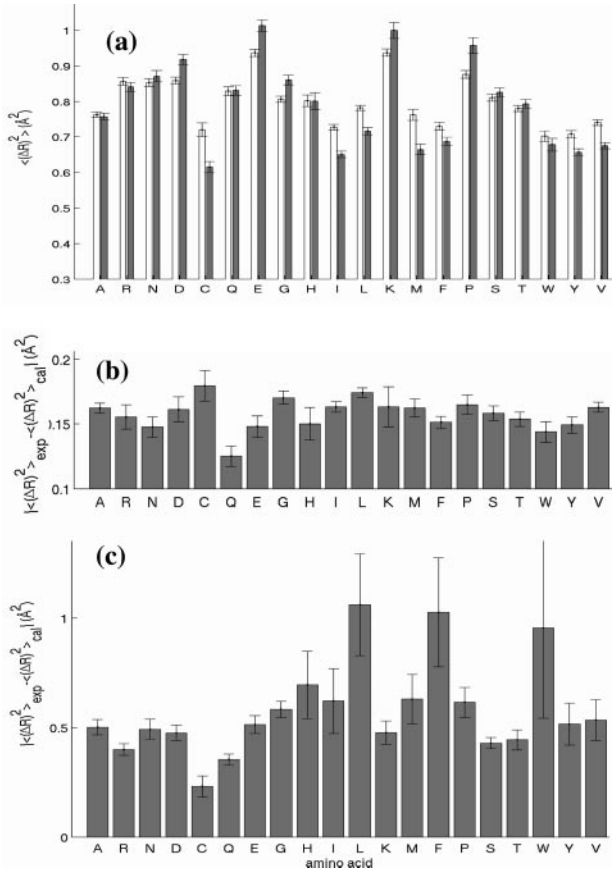

Fig. 4. Fluctuations of different amino acids. (**a**) Experimental (open bars) and theoretical (filled bars) fluctuations for different amino acids. (**b**) Level of agreement between the theoretical mean-squared fluctuations and the experimental ones in the protein core. (**c**) Level of agreement on the protein surface (accessibility >0.6).

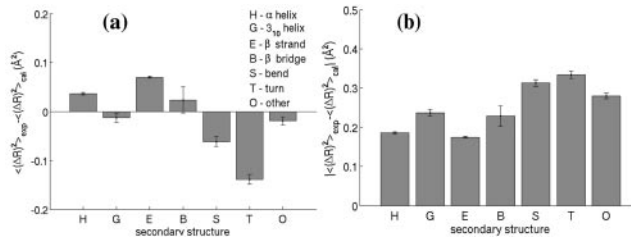

Fig. 5. (**a**) Difference between theoretical and experimental mean-square fluctuations for different secondary structures. (**b**) Mean absolute differences.

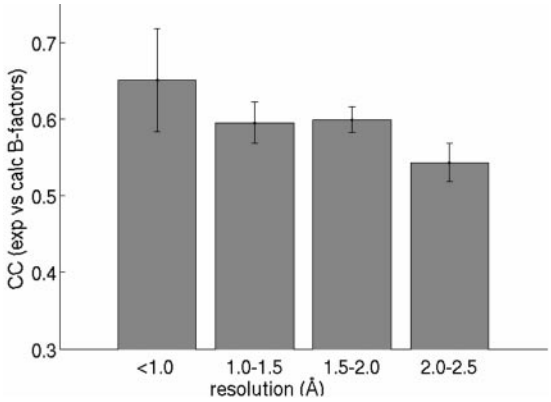

Fig. 6. Decrease in the correlation between theoretical (ANM) predictions and experimental data (B-factors) with decrease in the resolution of the X-ray crystallographic structure. Theoretical mean-squared fluctuations and the experimental ones.

property of amino acids. The overall amino acid sequence determines the architecture of the protein, but given that architecture, the local fluctuations at each site are not affected much by the identity of the residue. Figure 4b shows that the differences in the accuracy between residue types in the protein interior are small. In general, the predictions for larger amino acids (notably glutamine) are slightly more accurate than those of small amino acids. More pronounced differences are observed for exposed residues (Fig. 4c). These differences appear to be related to the physicochemical properties of the residues. Polar/charged residues are more accurately predicted than hydrophobic ones, a possible consequence of the involvement of surface hydrophobic residues in crystal contacts. Interestingly, the theoretical fluctuations of cysteines are by far more accurate than all other residues. In the protein interior, in contrast, fluctuations of cysteines are the least accurate. This might have to do with properties of the disulfide bonds not considered by the model: 38% of the cysteines in the database are involved in disulfide bonds.

Analysis of ANM results for different secondary structure elements reveals that the fluctuations are slightly underestimated for α-helices and β-strands (Fig. 5a); although, these departures are minimal in terms of absolute sizes (Fig. 5b). For other less organized regions (turns/loops), the theoretical fluctuations are relatively less accurate and tend to be larger than those observed in X-ray structures. Note that these are the regions that would be most strongly affected by crystal contacts.

We also checked if there are differences in the prediction of dynamics according to structural classes based on SCOP definitions (Murzin *et al*., 1995). No significant differences are found among the four major structural classes (α, β, $\alpha+\beta$, α/β).

3.5 Factors affecting the evaluation

B-factors are composed of several components which also reflect static disorders and inaccuracies in the crystals that mask the real dynamics component. It is therefore expected that the agreement between B-factors and theoretical predictions is limited, irrespectively of the quality of the model. Moreover, the temperature factors are also biased by non-natural crystal contacts as was shown previously (Eyal *et al*., 2005). This explains the lower fluctuations experimentally observed in exposed regions, compared with those predicted by ANM (Fig. 3a).

The level of agreement between theory and experiments is also affected by the resolution of the examined structure. The B-factors, similar to the coordinates, are refined based on diffraction patterns. More data (better resolution) should result in more accurate B-factors. The better agreement found between calculated fluctuations and B-factors for high-resolution structures (Fig. 6) confirms this and probably indicates about the accuracy of the B-factors more than that of the theoretical fluctuations.

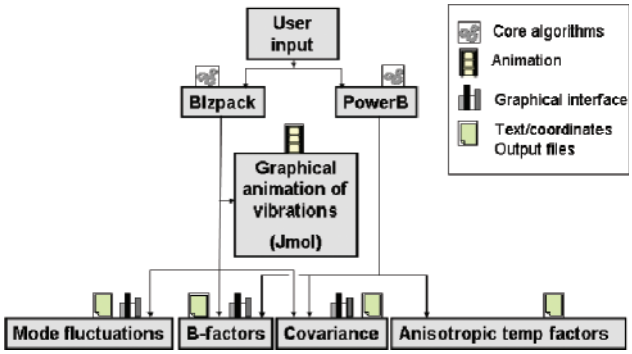

Fig. 7. ANM web server architecture.

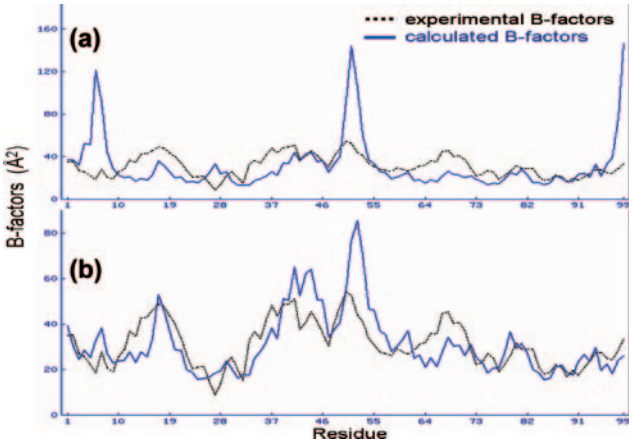

Fig. 8. Calculated and experimental B-factors for HIV protease (PDB code 1hhp). Graphs were taken from the ANM website. Theoretical results were obtained using (**a**) the coordinates of the single subunit reported in the PDB (CC 0.26) or (**b**) the coordinates of the homodimer (CC 0.74).

3.6 A web interface for the analysis of protein dynamics

To facilitate the use of the ANM model, especially by nonexpert users, we developed a new web interface for both calculations and analysis of the results. Its architecture is shown in Figure 7.

The core calculations are performed therein using a Block Lagrange algorithm implemented in Blzpack (http://crd.lbl.gov/%7Eosni/) to obtain the slowest 20 modes. The PowerB method (Yang *et al*., 2006) is applied for calculating the entire covariance matrix [Equation (5)] based on perturbation and inversion of the Hessian matrix followed by subtraction of the contribution of the six dominant eigenmodes of the perturbed matrix. These algorithms are very efficient. Calculations for proteins with hundreds of residues are completed within seconds and for proteins of thousands of residues within a few minutes on a single processor, 1 GHz, 1 GB machine (Yang *et al*., 2006). Advanced graphical analysis options including interactive animations of the vibrations, graphs of mean-square fluctuations, 2D plots for the correlations $\langle \Delta \boldsymbol{R}_i \bullet \Delta \boldsymbol{R}_j \rangle$ between the fluctuations of residue pairs as well as the fluctuations $\langle (\Delta \boldsymbol{R}_{ij})^2 \rangle$ in inter-residue distance vectors and the changes in inter-residue distances, $\langle (R_{ij} - R_{ij}^0)^2 \rangle$.

3.6.1 Input data The server offers considerable flexibility in controlling input data. Users can submit their own structures (in the PDB format) or use PDB structures. Notably, the structure of the putative multimeric states of proteins (biological units) recently deposited by the PDB can also be submitted. Understanding the functional role of the fluctuations requires consideration of the dynamics of the entire complex. This is now an easy task using our server. It is also important to consider the intact complex when comparing predicted fluctuations with experimentally observed ones (through the B-factors). The coordinates reported in the PDB files do not necessarily include all the molecules of the complex. The experimental temperature factors, however, reflect the effect of all subunits in the entire complex. This point is demonstrated in Figure 8, which shows the experimental and theoretical residue fluctuations taken from the ANM website for HIV protease as a monomer (the coordinates explicitly deposited in the PDB file 1hhp) and as a homodimer (the native state or biological unit). Significantly better agreement with experimental data is achieved when the homodimer structure is taken into consideration in the ANM computations. Inasmuch as the explicit coordinates deposited in the PDB file include only one subunit, the first step in computations should usually be to determine the coordinates of the symmetrically related subunit/monomer(s). The user can also choose to perform computations for all polypeptide chains or a single chain. In case of a multi-model PDB file, any of the models or all of them can be chosen. The parameters of the model, the interaction cutoff and the distance power, previously discussed, can also be interactively defined.

3.6.2 Graphical interactive animations using Jmol An important feature of the website is the ability to interactively explore the movements associated with individual modes using Jmol—an open source, rapidly developed molecular graphics software, specifically oriented for web applications. This is a significant improvement over other servers that release gif movies, only. Apart from their low resolution, these figures are noninteractive and display the molecule in a single fixed orientation.

Moreover, using Jmol, it is feasible to view in single session information on many vibration modes. This makes it possible to easily compare the different modes with only a button click to switch between them. Apart from the animation, the motion can be visualized from a set of arrows, the size and direction of which are proportional to the mobilities and reorientation of the individual residues, as given by the eigenvectors of $\boldsymbol{H}$. The mobility of residues can also be visualized by color-coded diagrams.

Some of these features are illustrated in Figure 9. The cartoon color-coded by the relative amplitude of motions induced in the lowest frequency mode (nonzero mode 1) is displayed in Figure 9a for the apo form of HIV protease, along with arrows indicating the directions of motions of all residues. The well-known enhanced fluctuations of the flap regions can be clearly distinguished. Figure 9b shows the mode profiles (normalized square displacements of residues) induced by the slowest four modes. It is worth noting that catalytic sites coincide with the minima in the first mode, which are also termed the global hinge sites, consistent with the recent observation of a general coupling between chemistry and

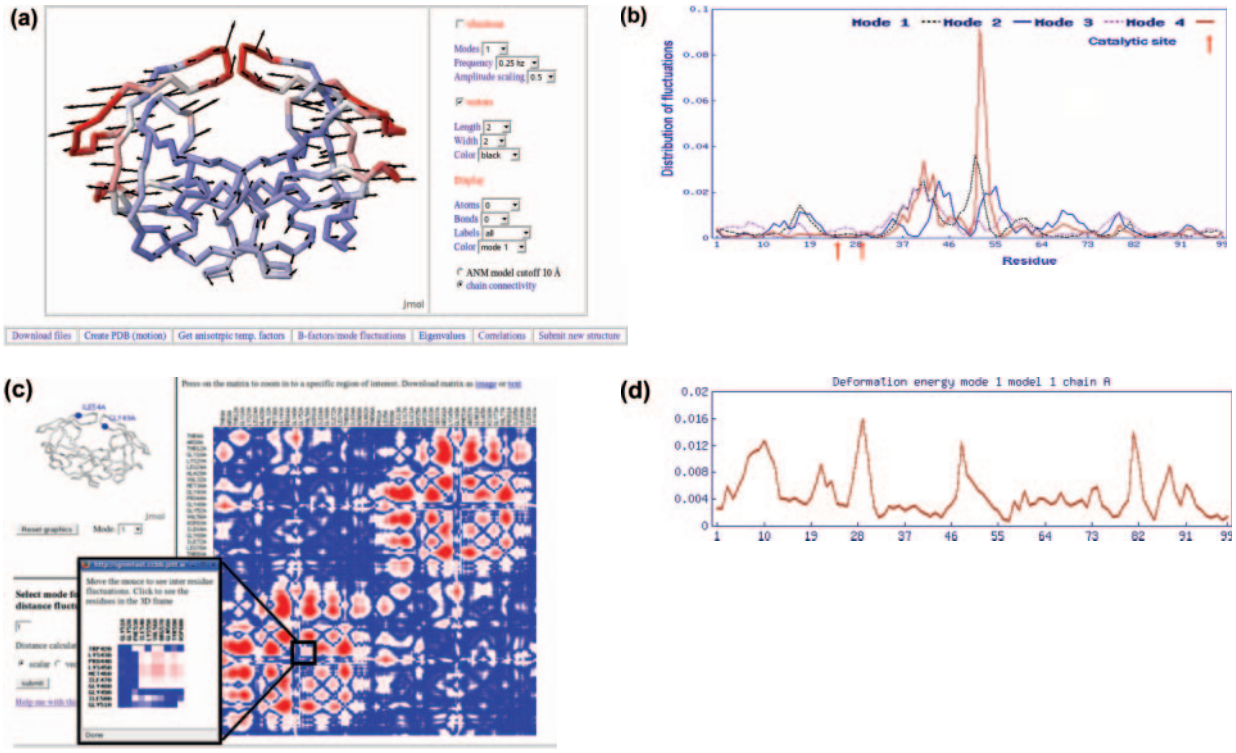

Fig. 9. ANM server graphical representation of the equilibrium dynamics of HIV protease (1hhp). (**a**) Ribbon diagram color-coded by mobilities in the slowest ANM mode. This is an interactively animated Jmol diagram, with arrows indicating the directions of motions. (**b**) Fluctuation profiles as a function of residue number corresponding to four lowest frequency modes. Results are computed for the dimer, but displayed for one of the monomers, both monomers exhibiting identical dynamics. Large fluctuations are found in the 'flap' region (near residues 48–53) in agreement with many experimental studies (Perryman *et al*., 2004; Piana *et al*., 2002), and the catalytic sites indicated by the arrows lie at the minima. (**c**) Inter-residue distance fluctuations in the slowest mode. The figure demonstrates how the user can zoom into a region of interest and how associated distances can be mapped on to the structure. (**d**) Deformation energies of residues associated with the motions driven by the first mode.

mechanics for controlling enzymatic activity (Yang and Bahar, 2005).

3.6.3 Cross-correlations and deformation energies The analysis of the cross-correlations between individual residues (reflecting the tendency of the residue to move in the same direction) is another feature of the website. Correlations can be computer based on individual modes, subset of modes or all modes. In addition, the fluctuations in inter-residue distances can also be visualized. These provide valuable information on the type of collective domain motions induced in different modes and the associated deformation energies. The user can graphically explore the distance fluctuations between any two residues and map the residues, colored according to the value on top of the vibrating structure, as illustrated in Figure 9c. The deformation energy for each residue is directly proportional to the sum of differences in squared distance with interacting residues comparing to the equilibrium state. The deformation energy profile of the first mode of HIV protease is shown in Figure 9d.

3.6.4 Output text and coordinate files The ANM website offers various types of output files, including coordinate files in the PDB format for alternative conformations generated by deforming the equilibrium state along the directions of the dominant modes. These files can be downloaded and read by other applications capable of showing/animating PDB structures. The user has a full control over the amplitude of the fluctuations and the number of frames in the animation (related to the smoothness of the motion). A single file with the input coordinates and the 20 most significant eigenvectors is also available, formatted according to Gamess (Schmidt *et al*., 1993) and readable by Jmol.

The model predicts the autocorrelations [diagonal terms in Equation (2); anisotropic temperature factors] and the covariance between fluctuations in different directions (off-diagonal terms). PDB files with C^{α} coordinates and their anisotropic temperature factors are available for download. Similarly, text files including the raw data such as eigenvalues and eigenvectors, as well as the fluctuations in individual modes, theoretical B-factors, calculated spring constant values are also readily accessible.

4 SUMMARY

The first part of this study focuses on a critical evaluation of the ANM and assessment of the limits of applicability and associated errors applied to different structures, structural regions (core versus solvent-exposed), secondary structural elements, amino acid types, resolution, and the effect of the choice of different functional forms and parameters for defining the network topology and spring constants. ANM is shown to better predict residue fluctuations in globular proteins, in the protein core and in regular secondary structural

elements. Having established the accuracy and limitations of the model, we have constructed a new web server, for a user-friendly assessment of equilibrium dynamics for the PDB structures and models submitted by users. By employing efficient core algorithms and advanced interactive graphical environments our server is very practical and powerful for simple NMA calculations and analysis, and for improving our understanding on the basic mechanisms of motions that underlie the functional dynamics of individual proteins as well as families of proteins, or biomolecular complexes and assemblies.

ACKNOWLEDGEMENTS

We gratefully acknowledge the assistance of Dr A. J. Rader, Indiana University–Purdue University, Indiana, in the development of the software and for insightful advice, and Dr Chakra Chennubhotla for critical reading of the manuscript. I.B. acknowledges the support from NIH grant no. 1 R33 GM068400-01A2. Funding to pay the Open Access publication charges was provided by the University of Pittsburgh.

Conflict of Interest: none declared.

REFERENCES

Alexandrov,V. *et al*. (2005) Normal modes for predicting protein motions: a comprehensive database assessment and associated web tool. *Protein Sci*., **14**, 633–643.

Atilgan,A.R. *et al*. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J*., **80**, 505–515.

Bahar,I. *et al*. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des*., **2**, 173–181.

Berman,H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res*., **28**, 235–242.

Cao,Z.W. *et al*. (2004) MoViES: molecular vibrations evaluation server for analysis of fluctuational dynamics of proteins and nucleic acids. *Nucleic Acids Res*., **32**, W679–W685.

Chennubhotla,C. *et al*. (2005) Elastic network models for understanding biomolecular machinary: from enzymes to supramolecular assemblies. *Phys. Biol*., **2**, S173–S180.

Chunyan,X. *et al*. (2003) Computational prediction of allosteric structural changes by a simple mechanical model: application to hemoglobin T to R transition. *J. Mol. Biol*., **333**, 153–168.

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

DeLano,W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.

Delarue,M. and Sanejouand,Y.H. (2002) Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J. Mol. Biol*., **320**, 1011–1024.

Doruker,P. *et al*. (2000) Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins*, **40**, 512–524.

Durand,P. *et al*. (1994) A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers*, **34**, 759–771.

Eyal,E. *et al*. (2005) The limit of accuracy of protein modeling: influence of crystal packing on protein structure. *J. Mol. Biol*., **351**, 431–442.

Haliloglu,T. *et al*. (1997) Gaussian dynamics of folded proteins. *Phys. Rev. Lett*., **79**, 3090–3093.

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

Hinsen,K. (2000) The molecular modeling toolkit: a new approach to molecular simulations. *J. Comput. Chem*., **21**, 79–85.

Hollup,S.M. *et al*. (2005) WEBnm@: a web application for normal mode analysis of proteins. *BMC Bioinformatics*, **11**, 52.

Humphrey,W. *et al*. (1996) VMD: visual molecular dynamics. *J. Mol. Graph*., **14**, 27–38.

Hung,A. *et al*. (2005) Molecular dynamics simulation of the M2 helices within the nicotinic acetylcholine receptor transmembrane domain: structure and collective motions. *Biophys. J*., **88**, 3321–3333.

Isin,B. *et al*. (2002) Functional motions of influenza virus hemagglutinin: a structure-based analytical approach. *Biophys. J*., **82**, 569–581.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Keskin,O. *et al*. (2002) Relating molecular flexibility to function: a case study of tubulin. *Biophys. J*., **83**, 663–680.

Kondrashov,D.A. *et al*. (2006) Optimization and evaluation of a coarse-grained model of protein motion using X-ray crystal data. *Biophys. J*., doi:10.1529/biophysj.106.85894.

Krebs,W.G. *et al*. (2002) Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins*, **48**, 682–695.

Kundu,S *et al*. (2002) Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophys. J*., **83**, 723–732.

Li,G. and Cui,Q. (2002) A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca^{2+}-ATPase. *Biophys. J*., **83**, 2457–2474.

Lindahl,E. *et al*. (2006) NOMAD-Ref: visualization, deformation, and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res*., **34**, W52–W56.

McConkey,B.J. *et al*. (2002) Quantification of protein surfaces, volumes and atom–atom contacts using a constrained Voronoi procedure. *Bioinformatics*, **18**, 1365–1373.

Micheletti,C. *et al*. (2004) Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. *Proteins*, **55**, 635–645.

Murzin,A.G. *et al*. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol*., **247**, 536–540.

Nicholay,S. and Sanejouand,Y.-H. (2006) Functional modes of proteins are among the most robust. *Phys. Rev. Lett*., **96**, PMID: 16606146.

Noguchi,T. and Akiyama,Y. (2003) PDB-REPRDB: a database of representative protein chains from the Protein Data Bank (PDB) in 2003. *Nucleic Acids Res*., **31**, 492–493.

Perryman,A.L. *et al*. (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci*., **13**, 1108–1123.

Piana,S. *et al*. (2002) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J. Mol. Biol*., **319**, 567–583.

Rader,A.J. *et al*. (2004) Identification of core amino acids stabilizing rhodopsin. *Proc. Natl Acad. Sci. USA*, **101**, 7246–7251.

Rader,A.J. *et al*. (2005) Maturation dynamics of bacteriophage HK97 Capsid. *Structure*, **13**, 413–421.

Schmidt,M.W. *et al*. (1993) General atomic and molecular electronic structure system. *J. Comput. Chem*., **14**, 1347–1363.

Shrivastava,I.H. and Bahar,I. (2006) Common mechanism of pore opening shared by five different potassium channels. *Biophys. J*., **90**, 3929–3240.

Suhre,K. and Sanejouand,Y.-H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res*., **32**, W610–W614.

Taly,A. *et al*. (2005) Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism. *Biophys. J*., **88**, 3954–3965.

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

Tama,F. *et al*. (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins*, **41**, 1–7.

Tama,F. *et al*. (2003) Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl Acad. Sci. USA*, **100**, 9319–9323.

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

Valadie,H. *et al*. (2003) Dynamical properties of the MscL of *Escherichia coli*: a normal mode analysis. *J. Mol. Biol*., **332**, 657–674.

Vincenzo,C. *et al*. (2006) Convergent dynamics in the protease enzymatic superfamily. *J. Am. Chem. Soc*., **128**, 9766–9772.

Wako,H. *et al*. (2004) ProMode: a database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics*, **20**, 2035–2043.

Wang,Y. *et al*. (2004) Global ribosome motions revealed with elastic network model. *J. Struct. Biol*., **147**, 302–314.

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

Yang,L.-W. *et al*. (2005) iGNM: a database of protein functional motions based on Gaussian Network Model. *Bioinformatics*, **21**, 2978–2987.

Yang,L.-W. *et al*. (2006) oGNM: online computation of structural dynamics using the Gaussian network model. *Nucleic Acids Res*., **34**, W24–W31.

Zheng,W. and Brooks,B.R. (2005a) Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *J. Mol. Biol*., **346**, 745–759.

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

Zheng,W. and Doniach,S. (2003) A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl Acad. Sci. USA*, **100**, 13253–13258.