PROTEINS: Structure, Function, and Bioinformatics 55:635–645 (2004)

Accurate and Efficient Description of Protein Vibrational Dynamics: Comparing Molecular Dynamics and Gaussian Models

Cristian Micheletti,[1,2] Paolo Carloni,[1] and Amos Maritan[1,2,3]
[1]*International School for Advanced Studies (SISSA) and INFM, Trieste, Italy*
[2]*Abdus Salam International Center for Theoretical Physics, Trieste, Italy*
[3]*Dip. di Fisica "Gi Galilei", Padua, Italy*

ABSTRACT **Current all-atom potential based molecular dynamics (MD) allows the identification of a protein's functional motions on a wide-range of timescales, up to few tens of nanoseconds. However, functional, large-scale motions of proteins may occur on a timescale currently not accessible by all-atom potential based MD. To avoid the massive computational effort required by this approach, several simplified schemes have been introduced. One of the most satisfactory is the Gaussian network approach based on the energy expansion in terms of the deviation of the protein backbone from its native configuration. Here, we consider an extension of this model that captures in a more realistic way the distribution of native interactions due to the introduction of effective side-chain centroids. Since their location is entirely determined by the protein backbone, the model is amenable to the same exact and computationally efficient treatment as previous simpler models. The ability of the model to describe the correlated motion of protein residues in thermodynamic equilibrium is established through a series of successful comparisons with an extensive (14 ns) MD simulation based on the AMBER potential of HIV-1 protease in complex with a peptide substrate. Thus, the model presented here emerges as a powerful tool to provide preliminary, fast yet accurate characterizations of protein near-native motion. Proteins 2004;55:635–645.**

Key words: HIV-1 protease; molecular dynamics; Gaussian models for vibrational dynamics; drug-resistance; protein flexibility

INTRODUCTION

Considerable insight into the biological activity of a protein can be gained by identifying its large-scale functional movements. Ideal tools for a detailed characterization of such dynamical properties are constituted by computational techniques such as molecular dynamics (MD) simulations based on effective all-atom potentials.[1] By these means, it is possible, at present, to follow numerically the dynamical evolution of a protein of a few hundred residues in its surrounding solvent over time intervals of tens of nanoseconds.

Such timescales allow us to gain considerable insight into important aspects of protein dynamics and to make quantitative connections with experimental quantities such as NMR order parameters[2,3] and Trp fluorescence spectra.[4] Other complex conformational changes, however, are difficult or impossible to observe. Examples include protein–protein molecular recognition, rearrangements occurring upon ligand binding, and so forth, which all involve timescales of the order of 1 μs or longer.[5] In addition, the simulated trajectory might not be sufficiently long that thermodynamic averages can be legitimately replaced with dynamical ones.[6]

Several studies have attempted to bridge the gap between the timescales of feasible MD simulations and the ones of biologically relevant protein motion by recoursing to a mesoscopic rather than a microscopic approach.[7] In fact, the large-scale dynamical features encountered in MD trajectories can be conveniently interpreted, at a first approximation, as a superposition of independent harmonic modes.[5] This observation was complemented by Tirion,[7] who pointed out that, in a normal mode analysis of protein vibrations, the detailed classical force field could be replaced by suitable harmonic couplings with the same spring constants. These results stimulated a variety of studies where the elastic properties of proteins were described through coarse-grained models in which amino acids are replaced by effective centroids corresponding to the C_α atoms and the energy function is reduced to harmonic couplings between pairs of spatially close centroids. These approaches, in particular the Gaussian and anisotropic network models (GNM and ANM), have been found to be in accord with both experimental and MD results.[8–10]

Here, we introduce an extended network Gaussian model that, at variance with previous approaches, incorporates effective C_β centroids "tethered" to the C_α atoms. The presence of the effective C_β's allows good control of the

Grant sponsor: INFM; Grant sponsor: INFM-Democraticos; Grant sponsor: MIUR Cofin 2003.

*Correspondence to: Cristian Micheletti, International School for Advanced Studies (SISSA/ISAS), Via Beirut 2-4, I-34014 Trieste, Italy. E-mail: michelet@sissa.it.

Received 26 June 2003; Accepted 3 October 2003

Published online 5 March 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20049

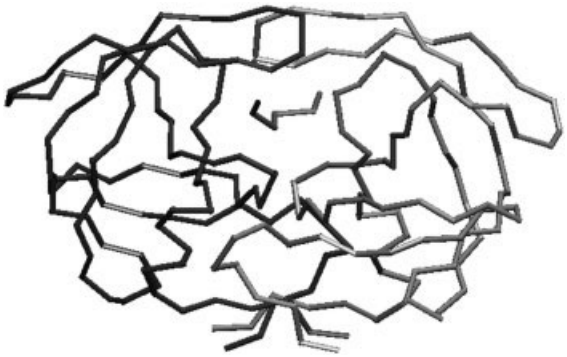

Fig. 1. Backbone trace of the complex formed by the HIV-1 protease homodimer and the bound substrate. Each monomer is composed of 99 residues; the model substrate consists of six amino acids.[11,12]

directionality of pairwise interactions in the protein and thus leads to an improved vibrational description. Furthermore, the fact that the side-chain degrees of freedom are entirely controlled by the C_α's has the crucial implication that the computational effort required to characterize the system is exactly the same as for models based only on C_α's.

The model equilibrium dynamics is compared against MD simulations based on all-atom effective potentials. We provide several quantitative estimates for how, and in what sense, the simplified quadratic approaches can provide a complement of the more accurate but also much more computationally demanding all-atom potential-based MD calculations. From a general point of view, the ideal term of reference would be constituted by a direct experimental determination of the quantities of interest here, such as the correlation of residues' motion. However, since this is not presently feasible, such detailed information can only be obtained from MD simulations. Although such simulations do not provide an absolute term of comparison, they are certainly an adequate reference for the much simplified protein models and energy functionals considered here.

The reference system considered here is the complex formed by the HIV-1 protease dimer with a bound model substrate (Fig. 1). This protein appears to be suitable to serve as reference for our model calculations in many respects. First, its dynamics in aqueous solution has been extensively investigated over more than 10 ns by all-atom effective potentials MD simulation.[11,12] Second, a large-scale motion analysis based on this MD simulation has been already performed. Third, this protein–substrate complex does not contain metal ions, prosthetic groups, cofactors, or nonstandard amino acids; therefore, it is amenable as a first test system for our model calculations, which consider only standard amino acids. Finally, this protein is of outstanding pharmaceutical relevance (it is one of the two targets currently used in anti-AIDS therapy) and its dynamics has been revealed as a key ingredient for the enzymatic function and for rationalizing resistance data.[11,13–15]

The data obtained from the MD simulation by Piana et al.[11,12] have been used here to evaluate all the MD-related quantities used for comparison against the results for quadratic models. In support of the general applicability of Gaussian approaches to capture the details of protein essential motion, we also report a comparison of the model predictions against data from a recent MD study of the NGF-trkA complex formed by the nerve growth factor and the tyrosin Kinase A receptor.[16]

The computer program implementing the Gaussian network model discussed here may be obtained from us.

THEORY

The Model

The starting point of the present analysis is the expansion of the Hamiltonian in terms of the deviations of the amino acids from their reference native positions. The underlying assumption is that a protein immersed in aqueous solution vibrates around its native state with amplitudes sufficiently small to justify the quadratic expansion around the minimum of the potential energy function. A posteriori, this does not seem to be a drastic restriction,[5,9,17] despite both the large-amplitude of motion observed in dynamical trajectory and the existence of several native substates, instead of a unique energy minimum.[18]

To reduce the spatial degrees of freedom of a protein, we adopt a coarse-grained model where a two-particle representation is used for each amino acid: Besides the C_α atom, an effective C_β centroid is employed to capture, in the simplest possible way, the side-chain orientation in a given amino acid (except for GLY, for which only the C_α atom is retained). To distinguish the proposed model from those based on the C_α-only representation, we shall refer to it as the β Gaussian model (βGM for brevity). Since our focus is to study the concerted vibration of various amino acids around the native state of the system, the Hamiltonian that is adopted incorporates, accordingly, pairwise interactions between all pairs of particles that are sufficiently spatially close in the native state. Formally, the system energy function evaluated on a trial structure, Γ, takes on the form

$$\mathcal{H}(\Gamma) = \mathcal{H}_{BB}(\Gamma) + \mathcal{H}_{\alpha\alpha}(\Gamma) + \mathcal{H}_{\alpha\beta}(\Gamma) + \mathcal{H}_{\beta\beta}(\Gamma), \qquad (1)$$

where

$$\begin{aligned}
\mathcal{H}_{BB}(\Gamma) &= k \sum_i V^{CA-CA}(d_{i,i+1}^{CA-CA}) \\
\mathcal{H}_{\alpha\alpha}(\Gamma) &= \sum_{i<j} \Delta_{ij}^{CA-CA} V^{CA-CA}(d_{i,j}^{CA-CA}) \\
\mathcal{H}_{\alpha\beta}(\Gamma) &= \sum_{i,j} \Delta_{ij}^{CA-CB} V^{CA-CB}(d_{i,j}^{CA-CB}) \\
\mathcal{H}_{\beta\beta}(\Gamma) &= \sum_{i<j} \Delta_{ij}^{CB-CB} V^{CB-CB}(d_{i,j}^{CB-CB}). \qquad (2)
\end{aligned}$$

In expressions [Eq. (2)], Δ_{ij}^{XY} is the native contact matrix that takes on the values of 1 [0] if the native separation of the effective particles of type X and Y, belonging respectively to residues i and j, is below [above] a certain cutoff value, R. On the other hand, Δ_{ij}^{XY} denotes the actual

separation of the particles in the trial structure, Γ. The indices i and j run over all integer values ranging from 1 up to the protein length, N. In particular, the interaction between particles in consecutive amino acids, $|i - j| = 1$, leads to a simple treatment of the protein chain connectivity. However, to account for the much higher strength of the peptide bond with respect to noncovalent contact interactions between amino acids, we have added in [Eq. (1)] an explicit chain term, $\mathcal{H}_{BB}$, where the interaction of consecutive C_α's is controlled by $k > 0$.

By construction, the minimum of the various interaction terms is attained for the native separation of each pair of particles. This ensures that the native state is at the global energy minimum. For small fluctuations around the native structure, the potential interaction energy of two particles, i and j, can be expanded in terms of the deviations from the native distance-vector, $\vec{r}_{ij}$. If we indicate the deviation vector as $\vec{x}_{ij}$, so that the total distance vector is $\vec{d}_{ij} = \vec{r}_{ij} + \vec{x}_{ij}$, we can approximate the pairwise interaction as

$$V(d_{ij}) \approx V(r_{ij}) + \frac{V''(r_{ij})}{2} \sum_{\mu,\nu} \frac{r_{ij}^{\mu} r_{ij}^{\nu}}{r_{ij}^2} x_{ij}^{\mu} x_{ij}^{\nu}, \qquad (3)$$

where μ and ν denote the Cartesian components, x, y, and z, and V'' is the second derivative of V. Several models have been introduced previously, where the quadratic expansion [Eq. (3)] was used in a context where only interactions among C_α centroids where considered, such as in the ANM, recently introduced to study the vibrational spectrum of proteins.[10]

Based on this quadratic expansion the Hamiltonian of Eq. (1) can be approximated as,

$$\tilde{\mathcal{H}} = \frac{1}{2} \sum_{ij,\mu\nu} x_{i,\mu}^{CA} M_{ij,\mu\nu}^{CA-CA} x_{j,\nu}^{CA}, \qquad (4)$$

where M is a $3N \times 3N$ symmetric matrix. The elastic response of the system is uniquely dictated by the eigenvalues and eigenvectors of M.

What differentiates the βGM from several previous studies is the presence of the interactions between C_α and C_β and $C_\beta - C_\beta$ (besides the extra strength of the chain term). The introduction of the C_β centroids in the protein description leads, in principle, to a more complicated Hamiltonian, with the additional C_β's degrees of freedom:

$$\mathcal{H} = \frac{1}{2} \sum_{ij,\mu\nu} x_{i,\mu}^{CA} M_{ij,\mu\nu}^{CA-CA} x_{j,\nu}^{CA} + \sum_{ij,\mu\nu} x_{i,\mu}^{CA} M_{ij,\mu\nu}^{CA-CB} x_{j,\nu}^{CB} + \frac{1}{2} \sum_{ij,\mu\nu} x_{i,\mu}^{CB} M_{ij,\mu\nu}^{CB-CB} x_{j,\nu}^{CB}. \qquad (5)$$

However, the location of the C_β atoms in a protein structure is almost uniquely specified by the geometry of the peptide chain. An accurate method that predicts the location of the C_β atoms from the CA trace of a protein is the geometric construction of Park and Levitt,[19] which assigns the ith C_β location given the positions of the C_α's of residues $i - 1$, i and $i + 1$, allows us to place the fictitious C_β at a distance of 0.3 Å from the crystallographic location. Such excellent agreement clarifies that the degrees of freedom of the C_β centroids should not be considered independent from the C_α ones. On the contrary, the C_β's can be viewed as rigidly "tethered" to the C_α and, hence, the fluctuations of the former are dictated by those of the latter.

Although in principle one could use the original rule of Park and Levitt,[19] we have adopted a simpler construction scheme, which places the C_β exactly in the plane specified by the local C_α trace. This simplifies the construction of the M matrices, which remain "diagonal" in the Cartesian components (e.g., the x component of the reconstructed C_β depends only on the x components of the neighbouring C_α's. More precisely, the location of the ith C_β is given by

$$\vec{r}_{CB}(i) = \vec{r}_{CA}(i) + l \frac{2\vec{r}_{CA}(i) - \vec{r}_{CA}(i+1) - \vec{r}_{CA}(i-1)}{|2\vec{r}_{CA}(i) - \vec{r}_{CA}(i+1) - \vec{r}_{CA}(i-1)|}. \qquad (6)$$

For reasons of self-consistency of the model, this construction rule is used to determine the contact matrices involving the effective C_β centroids that are used in place of those depending on the crystallographic C_β locations in Eq. (2). To leading order in the deviations of the C_α atoms, the deviations of $r_{CB}(i)$ thus becomes

$$\vec{x}_{CB}(i) \approx l \frac{2\vec{x}_{CA}(i) - \vec{x}_{CA}(i+1) - \vec{x}_{CA}(i-1)}{|2\vec{r}_{CA}(i) - \vec{r}_{CA}(i+1) - \vec{r}_{CA}(i-1)|}, \qquad (7)$$

where $l = 3$ Å. By using this rule, one parametrizes, in terms of the C_α positions, the effective C_β location of all residues except for GLY, and for the initial and final residues that lack one of the flanking C_α's. When the resulting expressions [Eq. (7)] are substituted in Eq. (5), one obtains an effective quadratic Hamiltonian which, as in Eq. (4), involves only the C_α deviations, but coupled through a new effective matrix, $\tilde{M}$, which is distinguished from previous matrices by a tilde. The bookkeeping operations necessary to calculate the elements of such a matrix are conveniently implemented with the aid of a computer. Thus, the computational cost and difficulty characterizing the elastic response of the protein is reduced to exactly the same as models with C_α atoms only. In spite of the same computational cost, the βGM appears to have several advantages in terms of the ability to capture the low-frequency motion and other vibrational properties of proteins, as will be seen below.

Equilibrium Properties

The derivation of the vibrational properties of a protein from the quadratic expansion of the Hamiltonian can be done, broadly speaking, in two different ways: the normal mode analysis and the Langevin analysis. What differentiates the two approaches is the view of the role of the solvent on the system dynamics. If one assumes that the motion of the protein is not significantly damped by the interaction with the solvent, then the normal modes picture can be applied to study the system dynamics around the native state by solving the Newton's dynamical equations[17,20–26]:

$$m_i \ddot{x}_{i,\mu} = \sum_{j,\nu} \tilde{M}_{ij,\mu\nu} x^{CA}_{j,\nu}. \quad (8)$$

The eigenfrequencies and eigenvectors are hence obtained by diagonalizing a matrix derived from $\tilde{M}$ by an appropriate mass weighting.[27]

Although the normal-mode analysis allows a straightforward dynamical characterization, it is of dubious applicability, since protein motion in a solvent does not resemble a superposition of pure harmonic oscillations. In fact, several theoretical, experimental, and computational studies have shown that the dynamics of a protein is severely overdamped by the interaction with the solvent.[21,23,28,29] The description of the motion in terms of overdamped dynamics appears to be particularly valid for the protein's low-frequency vibrations, which are the most interesting ones due to their expected role in proteins functional activities.[26] This observation leads to the alternative view of a heavily damped dynamics.[30]

In this case, for small deviations from the reference positions, the dynamics of the amino acids can be written as

$$\dot{x}_{i,\mu}(t) = -\sum_{j,\nu} \tilde{M}_{ij,\mu\nu} x_{j,\nu}(t) + \eta_{i,\mu}(t), \quad (9)$$

where the time unit has been implicitly chosen so that the viscosity coefficients (assumed to be equal for all particles) are set equal to 1 and the stochastic noise terms satisfy[31]:

$$\langle \eta_{i,\mu} \rangle = 0 \quad (10)$$

$$\langle \eta_{i,\mu} \eta_{j,\nu} \rangle = \delta_{i,j} \delta_{\mu,\nu} 2\kappa_B T. \quad (11)$$

These two conditions ensure, in the long run, the onset of canonical thermal equilibrium, so that the equilibrium probability of a given configuration, $\{x\}$, for the particles in the system is controlled by the Boltzmann factor:

$$e^{-\beta \mathcal{H}(\{x\})} = e^{-\beta \Sigma_{ij,\mu\nu} x_{i,\mu} \tilde{M}_{ij,\mu\nu} x_{j,\nu}}. \quad (12)$$

In this case, no periodic motion of the system can exist in the absence of external periodic excitations, since any structural deformation will be dissipated by a damped dynamics. The standard theory of stochastic processes[20,31] shows that the eigenvalues of $\tilde{M}$ are inversely proportional to the system relaxation times, and the corresponding eigenvectors indicate the actual shape of the associated distortion of the system.

Covariance Matrices and Temperature Factors

Besides identifying the elementary modes of excitation of a protein, it is important to calculate suitable thermodynamic quantities that characterize the protein dynamics once thermal equilibrium with the solvent has been established. The main observable quantity that can be calculated within the Gaussian model is the degree of correlation of the displacement from the equilibrium (native position) of pairs of C_α's. The thermodynamic average of the correlated displacements are easily obtained from the inversion of the M matrix. In fact, after setting $1/\beta = \kappa_B T = 1$, one has

$$\langle x_{i,\mu} x_{j,\nu} \rangle = \tilde{M}^{-1}_{ij,\mu\nu}, \quad (13)$$

where the brackets denote usual canonical thermodynamic averages with the weight of Eq. (12). The inverse matrix, $\tilde{M}^{-1}_{ij,\mu\nu}$, is often referred to as the covariance matrix. Since it provides directional details about the correlated motion of pairs of residues, we shall term it *full* covariance matrix to distinguish it from the *reduced* one discussed below, which incorporates only a measure of the degree of correlation (but no directional information).

The eigenvectors of the full covariance matrix represent the three-dimensional (3D) independent modes of structural distortion for the reference protein. The modes associated to the largest eigenvalues of $\tilde{M}^{-1}$ are the slowest to decay in a dissipative dynamics and, hence, make the largest contribution to the mean-square displacement of a given residues. The latter quantity is straightforwardly calculated from Eq. (13),

$$\langle |\vec{x}_i|^2 \rangle = \sum_{\mu} \tilde{M}^{-1}_{ii,\mu\mu}, \quad (14)$$

and can be directly connected to the temperature-factors (also called B-factors) measurements reported in X-ray or high-resolution NMR structural determinations.[32]

It is worth remarking that the full covariance matrix provides information about the system elasticity not only in conditions of isolation but also when an external force, $\vec{f}_i$, is applied to a given residue, i. In fact, within the Gaussian approximation, the average displacement of the jth amino acid from its reference position due to the application of $\vec{f}_i$ is given by

$$\langle x_{j,\nu} \rangle \propto \sum_{\mu} \tilde{M}^{-1}_{ji,\nu\mu} f_{i,\mu}. \quad (15)$$

As anticipated above, an important role in the analysis of molecular dynamical trajectories is also played by the reduced covariance matrix whose elements, C_{ij}, are defined as

$$C_{ij} \equiv \langle \vec{x}_i \cdot \vec{x}_j \rangle = \sum_{\mu} \tilde{M}^{-1}_{ij,\mu\mu}. \quad (16)$$

In ordinary MD simulations, the thermodynamic average in Eq. (16) is replaced with the time average taken over the simulated trajectory (ergodicity assumption). Due to the fact that the C matrix is obtained from $\tilde{M}^{-1}$ after a summation over the Cartesian components, the linear size of C is equal to the number of protein residues, N, instead of $3N$, as for $\tilde{M}^{-1}$. This 10-fold reduction of information greatly simplifies the identification of significant correlations between residues motion.

We conclude this section by discussing a technically important point. The inversion of the $\tilde{M}$ matrix used in Eqs. (13)–(16), as well as a correct interpretation of Eq. (12) are possible only within the subspace orthogonal to the eigenvectors of $\tilde{M}$ associated with zero eigenvalues. Physically this corresponds to omit the structural modifications that cost no energy (zero modes). Due to the invariance of Hamiltonian [Eq. (1)] under rotations and transla-

tions of the Cartesian reference frame, there will always be at least 6 zero modes. This number can, however, be larger if the $\tilde{M}$ matrix is sparse. The presence of additional spurious modes in Gaussian network models that incorporate only C_α coordinates is usually achieved by two means: either a reduction of the dimensionality of $\tilde{M}$ (as in GNM) or by using large interaction cutoffs in the range of 10–15 Å (as in ANM). The model discussed here allows us to use physically appealing interaction cutoffs of the order of 7 Å, as for GNM, yet retain the full 3D detail in the $\tilde{M}$ matrix. As will be shown later, these ingredients are necessary to capture the finer aspects of protein vibrations, such as the correlation of residues' motion, while, consistently with previous studies, the overall mobility of individual residues is rather insensitive to the details of the model.[33,34]

In fact, we found that GNM, ANM, and βGM have a similar performance on the prediction of experimental B-factors. This was established using high-resolution, single-chain proteins taken from the nonredundant Protein Data Base (PDB) select list.[35] We restricted to proteins length between 50 and 200 residues and excluded from the comparison the first and last 5 residues to avoid biases due to enhanced terminal mobility. Overall, we selected 36 proteins determined with X-ray and 31 with NMR. For simplicity, we summarize the level of agreement as the average of the nonparametric rank correlation, τ.[36] This analysis does not rely on the knowledge of the probability distribution from which the points (pairs of data) are taken. What matters is the agreement of the ranking of the points according to each of the two variables. In case of perfect [anti]correlation the Kendall parameter, τ takes on the value 1 [−1] and usually provides a more stringent (and robust) measure than linear correlation.[37] For the X-ray set and using GNM, τ ranged from 0.37 to 0.39 for cutoffs in the range 7.5–15.0 Å. For the same range, βGM gave $0.34 < \tau < 0.37$, while for ANM, it gave $0.30 < \tau < 0.37$ for interaction ranges 10–15 Å. The comparison against NMR temperature-factors provided higher correlations as already observed.[32] For the same cutoff ranges reported above, one has for GNM, $0.45 < \tau < 0.47$; for βGM, $0.46 < \tau < 0.48$ and for ANM, $0.43 < \tau < 0.48$.

In summary, the novel model discussed here allows us to incorporate, in an effective Hamiltonian, not only backbone–backbone interactions but also backbone–side-chain and side-chain–side-chain ones. The side-chain degrees of freedom are entirely controlled by the C_α positions. This has the crucial implication that the computational effort required to characterize the vibrational properties of the system is exactly the same as for models that incorporate only interactions between C_α pairs.

RESULTS AND DISCUSSION

In this section, we examine the extent to which suitable topology-based harmonic models can capture the details of the near-native vibrations of proteins in thermal equilibrium. Given the present impossibility of probing experimentally the various thermodynamical quantities discussed before, it is mandatory to choose as a reference the results of an all-atom MD calculation performed on the HIV-1 protease, in complex with TIMMNR peptide model substrate.[11,12] All MD dynamical averages were calculated after discarding an initial interval of a few nanoseconds over which the protease complexed to the model substrate was equilibrated.[11,12] The configuration obtained of the end of the equilibration protocol was taken as the reference structure for the Gaussian approach. The inversion of the symmetric $\tilde{M}$ matrix, necessary to characterize the system equilibrium dynamics, was done exploiting the Householder reduction[37] and took about 10 min on a personal computer.

For the comparison, it is important to mention that although MD studies can reproduce reliably a variety of experimental quantities they ultimately rely on empirical potentials which may be imperfectly parametrized. Besides this issue, it should also be noted that, usually, it is not easy to ascertain whether the simulated trajectory is sufficiently long that thermodynamic averages, as in Eq. (16), can be legitimately replaced with dynamical ones, though a recent study has indicated some valuable criteria for this purpose.[6] This potential limitations of general MD approaches should therefore be borne in mind also in the present context.

The series of tests carried out to ascertain the consistency among MD results and that of Gaussian models includes the comparison of temperature factors, covariance matrices, and essential subspaces. The findings, summarized below, provide a direct and strong indication that the βGM is apt to capture several aspects of proteins' near-native vibrational dynamics, with an accuracy that rivals with techniques based on all-atom potentials.

The generalized energy function of Eq. (1) contains several parameters; one of these, the interaction amplitude of C_α pairs, V''_{CA-CA}, can be conveniently taken as the energy unit. The other parameters are the interaction cutoff, R (which enters in the definition of the contact matrices) and the amplitudes of the C_β–C_β and C_β–C_α interactions, as well as the extra strength of the peptide term, k. For reasons of simplicity, the strength of these interactions have been chosen of the same order as $V''_{CA-CA} = k = 1$, $V''_{CA-CB} = V''_{CB-CB} = \frac{1}{2}$. This choice was made for reasons of simplicity but is not particularly restrictive due to the fact that effective interactions between side-chains are expected to be of the same order as C_α–C_α ones[38]; in addition, the precise value of k will mostly impact on the high-frequency vibrational spectrum of the system. Hence, it can be anticipated that the system elastic response should mostly depend on the value of the interaction radius, R, which has been accordingly varied in our analysis.

We first discuss the possibility to predict the temperature factors encountered in MD. This type of validation has been considered before by Doruker et al.[9] in connection with the anisotropic Gaussian model and reported a good consistency with dynamical simulations.

As a measure of the agreement between the residues mean-square fluctuations observed in MD and those predicted from the Gaussian models [see Eq. (14)], we consid-

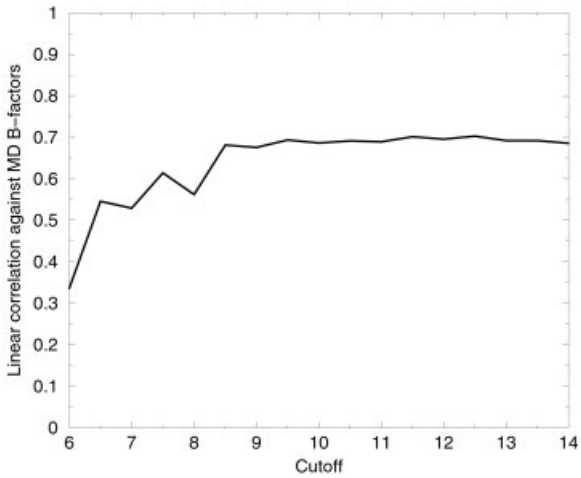

Fig. 2. Linear correlation coefficient for the temperature factors of the 204 residues in the HIV-1 PR/SUB complex.

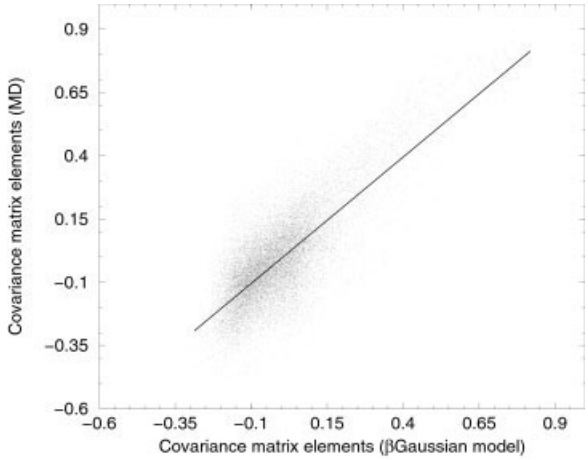

Fig. 3. Scatter plot of corresponding entries of the covariance matrices obtained within the βGM ($R = 7.5$ Å) and from the 14-nsMD simulation on the HIV-1 PR/SUB complex. The linear correlation coefficient over the 2×10^4 data points is 0.80.

ered the linear correlation coefficient. The degree of correlation as a function of the interaction cutoff radius for βGM is shown in Figure 2.

The performance of βGM is particularly stable starting from an interaction cutoff of about 8.0 Å. Given the large number of residues in the system (198 for the protein and 6 for the substrate) over which the B-factors are calculated, it is certainly possible to conclude that the correlation coefficient approaching 0.7 visible in Figure 2 is statistically significant. A more quantitative assessment of the statistical significance could be done using the Student's t-test or related methods,[37] although such analyses usually rely on the assumption that the joint distribution of the correlated variables is binormal (which is not necessarily satisfied for B-factors). An alternative way is to recourse to the nonparametric Kendall test mentioned before,[37] as recently proposed by Halle.[36] The Kendall correlation coefficient among the B-factors of the simulation and those of the β Gaussian model for ($R \approx 7.5$) Å is $\tau \approx 0.61$, and amply satisfies all ordinary criteria for statistical significance.

The successful comparison of the B-factors confirms the general agreement between the overall residues' motion in MD and the equilibrium dynamics predicted by Gaussian models[9]; in fact, for the same cutoff used above and against the same MD data of the HIV-1PR/SUB complex, GNM provides a linear correlation coefficient of 0.61, while the Kendall parameter τ is equal to 0.59. However, the finer details of such accord have not, to the best of our knowledge, been explored yet and, hence, become the focus of our subsequent analysis, which is based on a comparison of the MD and βGM covariance matrices. This test is particularly important due to the wealth of biological and chemical information that can been extracted from the covariance (essential dynamics) analysis.[39–41]

In the context of HIV-1 Pr,[13,14,42–49] these motions have a direct mechanical bearing on the structural modulation of the active site, even though they are located remotely from it.[11,12,15]

As a first case, we consider the reduced covariance matrices. To allow a straightforward comparison of the theoretical and numerical results, rather than working directly in terms of the reduced covariance matrix, it is useful to focus on the normalized version, which is dimensionless:

$$\tilde{C}_{ij} = \frac{\langle \vec{x}_i \cdot \vec{x}_j \rangle}{\sqrt{\langle |\vec{x}_i|^2 \rangle \langle |\vec{x}_j|^2 \rangle}}. \tag{17}$$

The scatter plot of Figure 3 summarizes the degree of accord between the normalized covariance matrix of the MD simulation and that of the βGM, for a cutoff of 7.5 Å. The number of entries in the plot is about 2×10^4, equal to the number of distinct entries in the 204×204 $\tilde{C}_{ij}$ matrix. To avoid introducing artificial biases in the correlation, the diagonal elements of the normalized matrices (which are all equal to 1) have been omitted from the plot.

The linear correlation coefficient among the two sets of data is 0.80 (and is stable in the neighborhood of $R = 7.5$ Å). We do not attempt to provide a quantitative measure for the statistical significance of the linear correlation in Figure 3. In fact, on one hand, as is visible in Figure 3, the joint distribution for covariance matrix elements is only approximately binormal; hence, the traditional tests of linear regression significance are of dubious applicability. On the other, the Kendall correlation measure requires us to consider all possible pairs of points in the scatter plot: This makes the analysis impractical and disproportionate to the main goal, which is to ascertain the existence of the accord between topological Gaussian models and MD results. In fact, rather than measuring the correlation in absolute terms, we compare the accord of Gaussian models and MD simulations against the degree of "internal" consistency of the simulated dynamical trajectory itself.

We conclude the discussion of the scatter plot of Figure 3 by mentioning that in case of perfect correlation of two $\tilde{C}_{ij}$ sets, due to the normalization condition of Eq. 17, the data would align along the diagonal of the graph in Figure 3.

Interestingly, despite the scatter visible in the same figure, the interpolating line lies very close to the diagonal, having a slope of $s = 0.97$ (taking the βGM covariance as the independent variable). This fact is useful in illustrating the effects of the cutoff, R, on the accord with MD covariance elements. While for $R = 10$ Å, the slope is still good, $s = 1.05$, it deteriorates for $R = 15$ Å, where $s = 1.64$. This effect is even more pronounced if the C_β's are not included in the model. For example, for $R = 15$ Å, the observed slope was $s = 3.41$.

The quantification of the self-consistency of MD trajectories is an extremely important issue, since it can provide an a posteriori indication of whether the dynamical sampling of the phase-space was sufficiently to obtain reliable thermodynamic averages. The analysis usually starts from the calculation of two covariance matrices pertaining to the first and second halves of the MD trajectory.

The two matrices could then be compared entry by entry, as done above. However, the most appropriate procedure is not to compare the corresponding matrix elements, but rather the physically important (essential) eigenspaces, that is, the linear spaces spanned by the eigenvectors of M^{-1} associated to the largest eigenvalues. A number of studies have suggested ways of measuring this consistency.

The first method of comparison that we will be taken into account is the one introduced by Amadei et al.,[50] which focuses on the top n eigenvectors of the covariance matrices under comparison. These eigenvectors describe the most significant modes of vibration of the molecule in the 3D space. We stress here that the covariance matrix considered here is not the reduced one of Eq. (16), whose size is $N \times N$, but is the full one (the M^{-1} matrix) of size $3N \times 3N$ that contains the 3D information about correlated motion of pairs of residues; see, for example, Eq. (13).

By denoting the top n eigenvectors of the two matrices under comparison of $\{\vec{\eta}\}$ and $\{\vec{\nu}\}$, the degree of overlap of the essential subspaces is defined as the root-mean-square inner product (RMSIP) of all pairs of eigenvectors in the two sets[50]:

$$RMSIP = \sqrt{\frac{1}{n} \sum_{i,j} |\vec{\eta}_i \cdot \vec{\nu}_j|^2}. \qquad (18)$$

Customarily, the analysis is restricted to the top $n = 10$ eigenvectors. We have measured the RMSIP in Eq. (18), when $\{\eta\}$ and $\{\nu\}$ come from the essential subspaces of the first and second halves of the 14-ns MD trajectory of the HIV-1 PR/SUB complex. The calculated RMSIP value, Eq. 18, was 0.71. Amadei et al.[50] have also proposed a series of approximate tests to ascertain the statistical relevance of the observed overlap. Based on their analysis, we can conclude that the value obtained here for the given system size of 204 amino acids has a probability to have arisen by chance that is, by far, inferior to the conventional threshold of 1%. This supports the fact that the MD trajectory was sufficiently long to contain significant physical information about the system equilibrium dynamics.

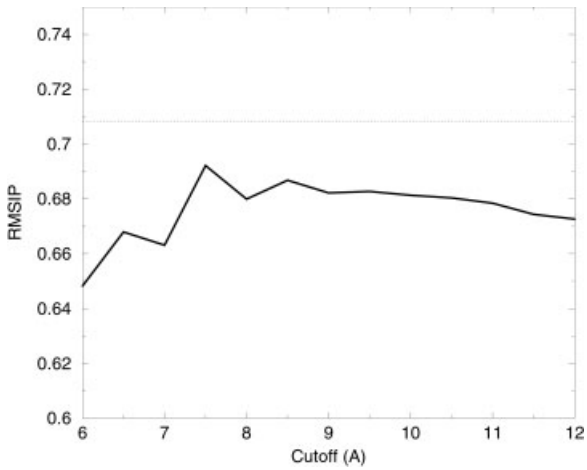

Fig. 4. Degree of correlation (root-mean-square-inner product), see Eq. (18), of the essential subspaces of the MD simulation and of the Gaussian models as a function of the interaction cutoff, R. The thick curve denotes the performance of the βGM, while the horizontal dotted line indicates the overlap of the essential subspaces of the first and second halves of the MD trajectory.

Having in mind the level of internal consistency of the present reference MD trajectory, we have turned to measuring the RMSIP between the essential spaces of the whole dynamical trajectory and those of the βGM.

The resulting trend for the subspaces overlap is shown in Figure 4 as a function of the interaction cutoff, R. The best performance of the model is obtained for a cutoff of $R \approx 7.5$ Å. The corresponding overlap value of 0.68 is very close to the internal overlap of the MD trajectory. We wish to remark that such values of RMSIP are highly nontrivial due to the large size of the full covariance matrix (612×612). This implies that the probability to observe a given overlap, q, between two random unit vectors decreases extremely rapidly as q approaches 1.[50] As a consequence, the MD simulation time required to reach a given target value for the internal RMSIP consistency, $\bar{q}$, grows very rapidly with $\bar{q}$.

This observation clarifies the utility of the Gaussian approach. With a modest computational investment, required by the diagonalization of a $3N \times 3N$ matrix, one obtains a description of the protein essential dynamics that, within a MD framework where all atom effective potentials are used, requires a considerably heavier computational investment. Further improvements over the βGM performance are obviously possible within MD, but at the price of a rapidly growing computing time.

The usefulness of the βGM is further supported by yet another type of analysis, which concludes the series of tests carried out here. This last approach follows a more precise measure for the agreement of the essential subspaces recently introduced by Hess.[6] The new measure is an improvement over the definition of Eq. (18), since it removes both the subjectivity of the choice of n, assigns more importance to the physically relevant eigenspaces, and deals correctly with the presence of spectral degeneracies.

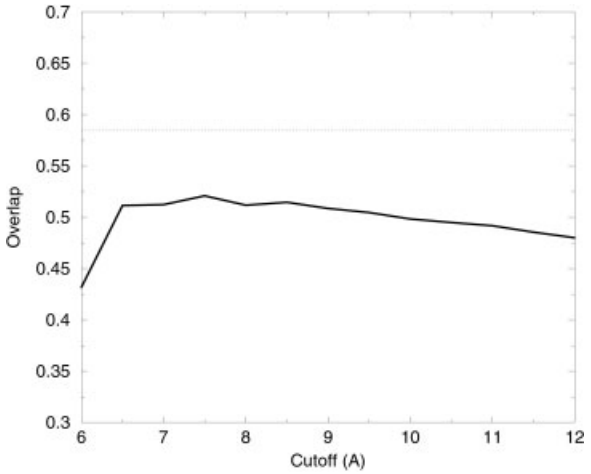

Fig. 5. Degree of overlap according to the measure introduced by Hess[6] of the essential subspaces of the MD simulation and of the Gaussian models as a function of the interaction cutoff, R. The thick curve denotes the performance of the βGM, while the horizontal dotted line indicates the overlap of the essential subspaces of the first and second halves of the MD trajectory.

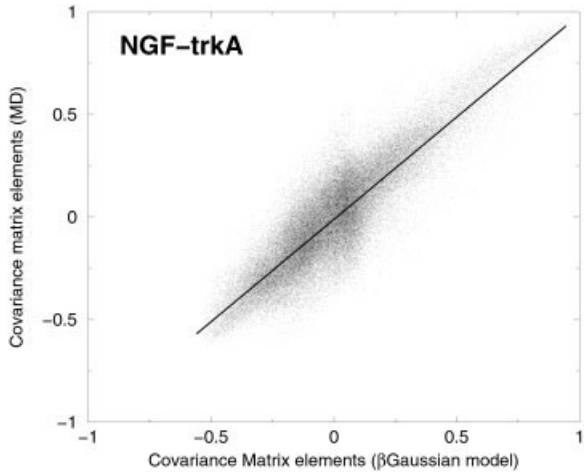

Fig. 6. Scatter plot of corresponding entries of the covariance matrices obtained within the βGaussianM (R = 7.5 Å) and from the 2.6-nsMD simulation on the NGF-trkA complex.[16] The linear correlation coefficient over the nearly 10^5 data points is 0.86.

Unlike the measures of Amadei et al.,[50] that of Hess is sensitive to the actual values of the eigenvalues of M^{-1}. Since the energy units of the Gaussian models considered here have been chosen arbitrarily, a proper normalization of the spectrum of the covariance matrices has to be carried out in order to use the measure of Hess for the comparison against MD. For this reason, the degree of overlap of the matrices was carried out after having uniformly rescaled the eigenvalues of M^{-1}, so that the trace of M^{-1} was equal to 1 for both systems. Physically, this corresponds to a normalization of the average residues mean-square displacement.

The results for this analysis are shown in Figure 5. The enhanced stringency of this test, which is extended to the whole vibrational spectrum and not just to the top 10 modes, is reflected in an overall decrease of the overlap with respect to Figure 4. This finer measure also reflects better than RMSIP the higher degree of inner consistency of the MD trajectory, as opposed to the MD-βGM accord. Although the best reference for our model would be provided by a simulation run sufficiently long that the inner MD overlap approaches 1, this is not presently feasible due to the slow (approximately logarithmic) increase of the inner overlap with the length of the MD run. Despite this fact, the results fully confirm the previous conclusions, namely, that the βGM can predict, with good statistical confidence, equilibrium dynamical properties of proteins and, in particular, identify the relevant modes of vibration of the system.

The reference system used here for the comparison between the model and MD simulations was chosen for both its biological significance and for the availability of MD data collected over the rather long simulation time. The βGM is, however, of general applicability, and to confirm the robustness of the strategy, we have considered another biologically important reference system, the NGF-trkA complex. This is constituted by a protein dimer, the nerve growth factor (NGF), complexed with the tyrosine kinase A receptor (trkA); altogether, the system comprises 431 residues. The dynamics of the complex in aqueous solution was recently simulated for a time span of 2.6 ns using all-atom potentials.[16] The resulting normalized covariance matrix was compared with the one obtained from the βGM. The linear correlation coefficient over the nearly 10^5 corresponding distinct entries of the matrices is 0.86, as is visible in Figure 6.

This result confirms the viability of the Gaussian approach to capture the details of the large-scale protein movements. However, due to the simplicity of the model interaction potentials (all pairs of C_α's and C_β's interact with the same strength), one may foresee that the model is not suitable for modeling the vibrational dynamics of proteins where electrostatic effects or disulfide bonds play an important role for native stability or functionality.

BIOLOGICAL IMPLICATIONS

In the MD calculations by Piana et al.,[11,12,15] the essential dynamics analysis has allowed us to identify the sites that, despite being spatially distant from the active site, have a strong mechanical influence on the structural modulation of the regions binding the substrate. We now compare the findings obtained with MD with those obtained with our model. The degree of mechanical coupling observed in the MD trajectory between the substrate motion and the HIV-1 protease subunits is visible in the top panel of Figure 7. The two curves in the plot represent the profile of the reduced covariance matrix, C_{ij}, between the two central atoms of the peptide and the 198 protease residues. Interestingly, the regions that correlate significantly with the substrate motion are those that have been indicated as rather underconstrained by studies where the theory of rigidity has been applied to characterize the enzyme elasticity.[51] The two facts provide a consistent picture for the HIV-1 PR mechanics, since any functionally relevant mechanical coupling must intuitively involve

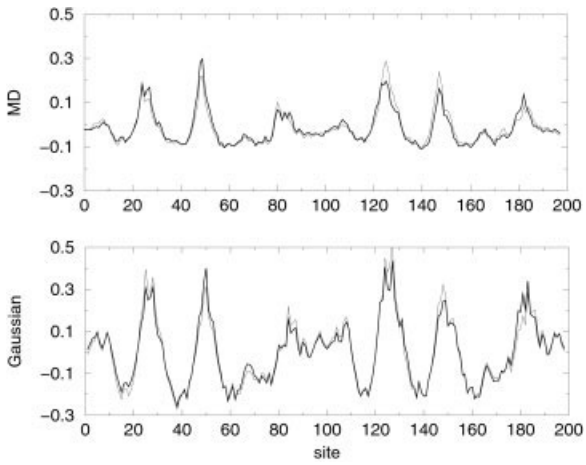

Fig. 7. The two curves in each panel indicate the degree of correlation of the motion between the two central atoms of the substrate and the 198 residues in the subunits of the HIV-1 protease. The top panel reports the MD findings, while the bottom one pertains to the βGM.

mobile (and hence underconstrained) regions. The identification of the mobile regions of HIV-1 protease has also been previously addressed with Gaussian network models in a series of studies that also allowed identification of the residues important for protein stability,[52,53] or otherwise taking part in crucially important networks of key native contacts.[54]

The bottom panel of Figure 7 represents the corresponding correlation profiles calculated within the βGM for a cutoff $R = 7.5$ Å (which we take as an optimal value from previous analysis). The degree of agreement of the profiles across the two panels is remarkable, and the main difference appears to be due to an overemphasis of negative correlations in the Gaussian profiles.

In both sets of data, one observes a strong positive correlation between the substrate motion and the regions 24–30 and 45–55. This direct mechanical coupling is immediately interpreted due to the fact that the first region comprises the cleavage site, while the second involves the tips of the protease flaps.

The essential-space analysis of the MD trajectory revealed that the two regions embrace the substrate and involve it in a rotational "nutcracker-like" motion. As a consequence of this rotation, the regions near the flaps elbows, 37–41 and 61–73, undergo a countermovement that results in a negative correlation with the substrate motion. This effect is clearly visible in the Gaussian profiles of Figure 7, while it is less pronounced, but still significant, in the MD results.

As noted in the Theory section, the intimate connection between the linear response theory and the covariance matrix allows us to conclude that a force applied in correspondence of sites around residues 40 and 63 should affect the protease–substrate coupling. This effect was indeed observed in the MD simulation, where the motion of the substrate toward the cleavage site was strongly affected by the constraints on these regions (see Fig. 6 in Piana et al.[11]).

It is by virtue of this mechanically important coupling that it is possible to rationalize the emergence of mutations causing drug resistance in correspondence of sites far from the cleavage region (e.g., M63I, M46I-L, and L47V). In fact, as the explicit MD calculation has shown, the high degree of such coupling between such sites and the cleavage is such that the detailed chemical identity of the former strongly influences the substrate binding affinity of the latter. In particular, the mutations observed in clinics[45–49] are arguably the result of a chemical fine-tuning that retains the native functionality of the enzyme, while decreasing its affinity for inhibiting drugs.

The details of how the enzymatic reaction kinetics changes upon amino acid mutations is beyond the reach of the topological models presented here. The Gaussian scheme, in fact, is entirely adequate to identify which sites influence mechanically the active site motion, but due to the fact that all amino acids (except for GLY, which lacks the C_β centroid) are treated equally, we cannot explore the ramifications of changing the amino acid identity into the cleavage-region and substrate coupling. The effect of one of such mutations (M46I) on substrate motion has instead been fully taken into account.[12] Further improvements of the βGM may be possible by optimizing the distance of the C_β centroids from their respective C_α's (to better capture the displacement of the side-chain centers of mass). The model could also be extended to include, at the simplest possible level, the effects of thermal denaturation through a self-consistent, temperature-dependent weakening of the strength of the harmonic couplings, as done earlier.[32,55]

CONCLUSIONS

We have examined the extent to which the dynamical properties of a protein in thermodynamic equilibrium can be accounted for through solvable models. The starting point of our analysis is the quadratic approximation of the free energy landscape in terms of the deviations of amino acids from their reference positions in the known native state. We have adopted a novel description of the amino acids, which allows us to consider the presence of effective C_α centroids whose degrees of freedom are entirely controlled by the C_α atoms. As a result, the model that is considered is able to account for the directionality of amino acid side-chains, while retaining the same degree of complexity as models based on C_α representation only. Various equilibrium quantities that are apt to characterize the most relevant modes of vibrations of proteins are considered. In particular, we focused on (in increasing order of complexity and detail) the B-factors, the covariance matrices, and the essential dynamical subspace. Our results have been compared against the analogous quantities obtained through a 14-ns MD simulation carried out on the HIV-1 PR enzyme in complex with a TIMMNR peptide substrate.

As far as overall equilibrium dynamical properties are concerned, the βGM provides a picture that is in remarkable agreement with the MD results. In fact, the essential subspace predicted theoretically appears to have a degree

of consistency with MD results that is close to the "inner consistence" of a 14-ns MD simulation with itself.

This provides a strong indication that suitable quadratic models can provide a powerful and accurate tool for characterizing the vibrational motions of proteins near their native state, while requiring only a modest investment of computational resources. Other important properties of protein dynamics and functionality that strongly depend on the sequence composition or on out-of-equilibrium conditions are, at the moment, beyond the reach of such simplified approaches considered here. For this reason, we believe that the Gaussian approach would be ideally used in conjunction with MD by providing, prior to investing significant computational resources in all-atom simulations, a fast but accurate characterization of a protein's near-native motion.

ACKNOWLEDGMENTS

We are indebted to Stefano Piana, Michele Cascella, Giorgio Colombo, Paolo De Los Rios, Gianluca Lattanzi, Luca Marsella, and Gianni Settanni for useful suggestions and advice.

REFERENCES

1. Karplus M. Molecular dynamics simulations of biomolecules. Acc Chem Res 2002;35:321–323.
2. Lipari G, Szabo A. Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: 1. Theory and range of validity. J Am Chem Soc 1982;104:4546–4559.
3. Lipari G, Szabo A. Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: 2. Analysis of experimental results. J Am Chem Soc 1982;104:4559–4570.
4. Brooks CL, Karplus M, Pettitt BM. Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. New York: Wiley; 1988.
5. Noguti T, Go N. Collective variable description of small-amplitude conformational fluctuations in a globular protein. Nature 1982;296:776–778.
6. Hess B. Convergence of sampling in protein simulations. Phys Rev E 2002;65:031910.
7. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
8. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. Fold Des 1997;2:173–181.
9. Doruker P, Atilgan A, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. Proteins 2000;40:512–524.
10. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
11. Piana S, Carloni P, Parrinello M. Role of conformational fluctuations in the enzymatic reaction of hiv-1 protease. J Mol Biol 2002;319:567–583.
12. Piana S, Carloni P, Rothlisberger U. Drug resistance in hiv-1 protease: flexibility-assisted mechanism of compensatory mutations. Protein Sci 2002;11:2393–2402.
13. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero JC, Rhodes A, Robbins HL, Roth E, Shivaprakash M, Titus D, Yang T, Teppler H, Squires KE, Deutsch PJ. In-vivo emergence of hiv-1 variants resistant to multiple protease inhibitors. Nature 1995;374:569–571.
14. Patrick AK, Mo H, Markowitz M, Appelt K, Wu B, Musick L, Kalish V, Kaldor S, Reich S, Ho D, Webber S. Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. Antimicrob Agents Chemother 1996;40:292–297.
15. Piana S, Carloni P, Rothlisberger U. Reaction mechanism of hiv-1 protease by hybrid car-Parrinell MD/classical MD simulations, 2003. In press.
16. Settanni G, Cattanea A, Carloni P. Molecular dynamics simulations of the ngf-trka domain 5 comples and comparison with biological data. Biophys J 2003;84:2282–2292.
17. Horiuchi T, Go N. Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: human lysozyme. Proteins 1991;10:106–116.
18. Frauenfelder H, Sligar SG, and Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–1603.
19. Park B, Levitt M. Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. Proteins 1996;258:367–392.
20. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 1985;181:423–447.
21. Brooks B, Karplus M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. Proc Natl Acad Sci USA 1985;82:4995–4999.
22. Case DA. Normal mode analysis of protein dynamics. Curr Opin Struct Biol 1994;4:285–290.
23. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.
24. Tirion MM, ben Avraham D. Normal mode analysis of g-actin. J Mol Biol 1993;230:186–195.
25. Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. Collective motions in hiv-1 reverse transcriptase: Examination of flexibility and enzyme function. J Mol Biol 1999;285:1023–1037.
26. Haliloglu T, Bahar I. Structure based analysis of protein dynamics: comparison of theoretical results for hen lysozyme with X-ray diffraction and NMR relaxation data. Proteins 1999;37:654–667.
27. Goldstein H. Classical mechanics. 2nd ed. Reading, MA: Addison-Wesley, 1980.
28. McCammon JA, Gelin BR, Karplus M, Wolynes PG. The hinge-bending mode in lysozyme. Nature 1976;262:325–326.
29. Swaminathan S, Ichiye T, van Gusteren W, Karplus M. Time dependence of atomic fluctuations in proteins: analysis of local and collective motions in bovine pancreatic trypsin inhibitor. Biochemistry 1982;21:5230–5241.
30. Howard J. Mechanics of motor proteins and the cytoskeleton. Sunderland, MA: Sinauer Associates; 2001.
31. Chandrasekhar S. Stochastic problems in physics and astronomy. Rev Mod Phys 1943;15:1–89.
32. Micheletti C, Lattanzi G, Maritan A. Elastic properties of proteins: insight on the folding process and evolutionary selection of native structures. J Mol Biol 2002;321:909–921.
33. Doruker P, Jernigan R, Bahar I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. J Comput Chem 2002;23:119–127.
34. Doruker P, Jernigan R, Navizer I, Hernandez R. Important fluctuation dynamics of large protein structures are preserved upon renormalization. Int J Quantum Chem 2002;90:822–837.
35. Hobohm U, Sander C. Enlarged representative set of protein structures. Protein Sci 1994;3:522–524.
36. Halle B. Flexibility and packing in proteins. Proc Natl Acad Sci USA 2002;99:1274–1279.
37. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes. Cambridge, UK: Cambridge University Press; 1999.
38. Micheletti C, Seno F, Banavar JR, Maritan A. Learning effective amino acid interactions through iterative stochastic techniques. Proteins 2001;42:422–431.
39. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.
40. Garcia A. Large-amplitude nonlinear motions in proteins. Phys Rev Lett 1992;68:2696–2699.
41. Rod TH, Radkiewicz JL, Brooks CL. Correlated motion and the effect of distal mutations in dihydrofolate reductase. Proc Natl Acad Sci USA 2003;100:3954–3959.
42. Ala PJ, Huston EE, Klabe RM, Jadhav PK, Lam PYS, Chang CH. Counteracting HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with XV638 and SD146, cyclic urea amides with broad specificities. Biochemistry 1998;37:15042–15049.
43. Gulnik S, Erickson JW, Xie D. Vitamins and hormones—advances in research and applications. Vitam Horm 2000;58:213–256.
44. Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1

protease. Annu Rev Biochem 1993;62:543–585, and references therein.
45. Reddy P, Ross J. Amprenavir—a protease inhibitor for the treatment of patients with HIV-1 infection. Formulary 1999;34:567–675.
46. Brown AJL, Korber BT, Condra JH. Associations between amino acids in the evolution of HIV type 1 protease sequences under indinavir therapy. AIDS Res Hum Retroviruses 1999;15:247–253.
47. Boucher C. Rational approaches to resistance: using saquinavir. AIDS 1996;10:S15–S19.
48. Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo HM, Markowitz M, Chernyavskiy T, Niu P, Lyons N, Hsu A, Granneman GR, Ho DD, Boucher CAB, Leonard JM, Norbeck DW, Kempf DJ. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. Nat Med 1996;2:760–766.
49. Markowitz M, Mo HM, Kempf DJ, Norbeck DW, Bhat TN, Erickson JW, Ho DD. Selection and analysis of human-immunodeficiency-virus type-i variants with increased resistance to abt-538, a novel protease inhibitor. J Virol 1995;69:701–706.
50. Amadei A, Ceruso MA, Nola AD. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. Proteins 1999;36:419–424.
51. Jacobs D, Rader A, Kuhn L, Thorpe M. Protein flexibility predictions using graph theory. Proteins 2001;44:150–165.
52. Bahar I, Atilgan AR, Demirel MC, Erman B. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. Phys Rev Lett 1998;80:2733–2736.
53. Kurt N, Scott W, Schiffer C, Haliloglu T. Cooperative fluctuations of unliganded and substrate-bound hiv-1 protease: a structure-based analysis on a variety of conformations from crystallography and molecular dynamics simulations. Proteins 2003;51:409–422.
54. tk;4Micheletti C, Cecconi F, Flammini A, Maritan A. Crucial stages of protein folding through a solvable model: predicting target sites for enzyme-inhibiting drugs. Protein Sci 2002;11:1878–1887.
55. Micheletti C, Banavar J, Maritan A. Conformations of proteins in equilibrium. Phys Rev Lett 2001;87:DOI:088102-1.