Temperature Dependence of Normal Mode Reconstructions of Protein Dynamics

Francesco Piazza and Paolo De Los Rios
Laboratoire de Biophysique Statistique, SB ITP, Ecole Polytechnique Fédérale de Lausanne - EPFL, CH-1015, Lausanne, Switzerland

Fabio Cecconi
SMC-INFM Center for Statistical Mechanics and Complexity (CNR) and Istituto dei Sistemi Complessi CNR, Via dei Taurini 19, 00185 Rome, Italy.

(Received 26 October 2008; published 29 May 2009)

Normal mode (NM) analysis is a widely used technique for reconstructing conformational changes of proteins from the knowledge of native structures. In this Letter, we investigate to what extent NMs capture the salient features of the dynamics over a range of temperatures from close to $T = 0$ to above unfolding. We show that the use of normal modes at room temperature is justified provided proteins are cooperative, that is, globular and highly structured. On the other hand, it is imperative to consider several modes in order to eliminate the unpredictable temperature dependence of single-mode contributions to the protein fluctuations.

DOI: 10.1103/PhysRevLett.102.218104 PACS numbers: 87.15.−v, 87.10.Tf, 87.14.E−

The functional effectiveness of proteins strongly depends on their ability to explore different conformations. In many cases, these correspond to global deformations of the molecule, although not necessarily of large amplitude. In such circumstances, the small-amplitude dynamics can be adequately described through normal modes (NM) analysis in order to decipher the structure-dynamics-function relation [1,2].

NM analysis of macromolecules has a venerable history, dating back to the early 1980's [3,4], and has enjoyed several renaissances until its latest variant, the anisotropic network model and its derivatives [5,6]. The cornerstone of the NM approach is the assumption that functional dynamics of proteins can be captured by the lowest-frequency normal modes as long as the native-state deformations are of small amplitude. Therefore, the total potential energy $V(\{\vec{r}\})$ of a protein can be expanded, as a function of the atomic coordinates $\vec{r}_i$, $i = 1, \ldots, N$, to second order around the native state

$$V(\{\vec{r}\}) \simeq \frac{1}{2} \sum_{i\alpha, j\beta} \partial^2_{i\alpha, j\beta} V|_{\vec{R}_i} \Delta r_{i\alpha} \Delta r_{j\beta}, \qquad (1)$$

where $\Delta r_{i\alpha} = r_{i\alpha} - R_{i\alpha}$, $R_{i\alpha}$ being the native-state coordinates. The dynamical matrix of the system is defined as $D = M^{-1/2} H M^{-1/2}$, with M the diagonal mass matrix and H the Hessian of the potential. The eigenvectors $\hat{\psi}_k$, $k = 1, \ldots, 3N$, of D are the normal modes of the protein.

However, proteins live in a free-energy minimum, that does not necessarily coincides with the global minimum of $V(\{\vec{r}\})$. Therefore, although NM analysis has been applied with success also to predict large-scale deformations, its appropriate context remains the small-amplitude regime and extrapolation to larger fluctuations should be rigorously checked to avoid arbitrary interpretations. In fact, at $T = 300$ K, a mere 10% below typical protein unfolding temperatures (T_f), displacements well beyond the harmonic approximation must be expected. A possible way to find suitable collective coordinates to describe large fluctuations is to use the "essential modes", the eigenvectors of the covariance matrix of the displacements from the native state, computed in a molecular dynamics simulation [7]. Unfortunately, the convergence of the covariance matrix toward a stationary solution is very slow [8].

In this Letter, we establish on a firm basis the requirements for the validity of the NM description of protein dynamics at room temperature and unveil some remarkable

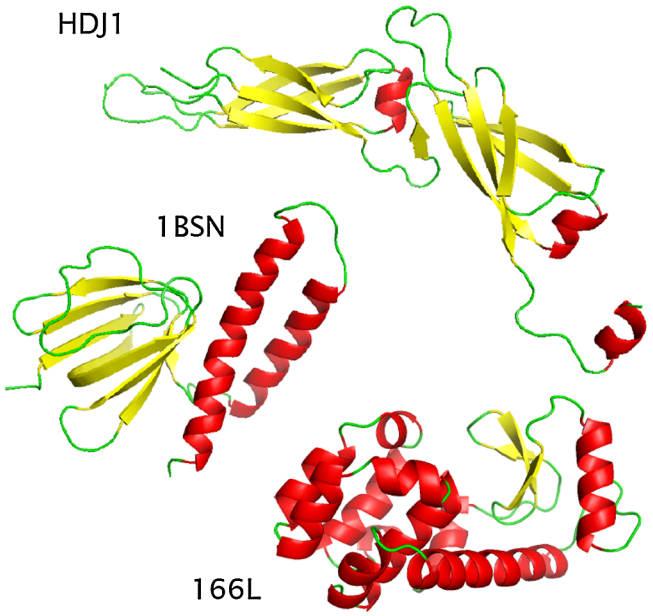

FIG. 1 (color online). Cartoon representations of three of the analyzed structures, along with the corresponding PDB codes.

features of this approach as temperature is increased from very low to beyond T_f.

In order to measure the spectral weight of protein fluctuations as a function of temperature, we introduce the thermal involvement coefficients (TICs) $\Theta_k(T)$ [9]. They are defined as $\Theta_k(T) = \langle Y_k(T)\rangle_{\mathcal{N}}$, with

$$Y_k(T) = \frac{\langle(\vec{\Delta}\cdot\hat{\psi}_k)^2\rangle_t}{\sum_{k'}\langle(\vec{\Delta}\cdot\hat{\psi}_{k'})^2\rangle_t}, \quad (2)$$

where $\vec{\Delta} = \vec{r} - \vec{R}$ is the $3N$-dimensional deviation of the protein structure from the native fold configuration, which is projected onto the NMs, and averaged both over time $\langle\ldots\rangle_t$ and over $\mathcal{N}$ different initial conditions drawn from the same equilibrium distribution at temperature T, $\langle\ldots\rangle_{\mathcal{N}}$. We have explicitly verified that the two averages commute, so as to strengthen the statistical significance of the sampling and averaging procedures. Statistical uncertainties were computed as the standard errors on the realization averages $\Delta\Theta_k(T) = \sqrt{[\langle Y_k^2(T)\rangle_{\mathcal{N}} - \langle Y_k(T)\rangle_{\mathcal{N}}^2]/\mathcal{N}}$.

Protein trajectories are simulated within the isokinetic scheme [10], which provides a correct sampling of the configuration space [11]. Forces are computed using the coarse-grained $G\bar{o}$ model introduced in Ref. [12], where each amino-acid is replaced by a bead with the average mass of 110 a.m.u., in equilibrium at its C_α site. Successive beads along the backbone are connected by stiff harmonic springs, while nonbonded interactions between non consecutive particles are modeled with Lennard-Jones 12-10 potentials if the inter-residue distances in the native state are smaller than a given cutoff R_c and with purely repulsive interactions otherwise. The cutoff is fixed as $R_c = 7.5$ Å, but for proteins with $N < 100$, where we take $R_c = 6.5$ Å. This prevents smaller structures from being *artificially* rigidified. The force field is completed by standard harmonic angle-bending interactions, plus the dihedral potential energy

$$V_{dh} = \sum_{i=3}^{N-2}[k_\phi^{(1)}(1 - \cos\Delta\phi_i) + k_\phi^{(3)}(1 - \cos 3\Delta\phi_i)], \quad (3)$$

where $\Delta\phi_i$ are the dihedral angle deviations from the native values [13]. NMs are computed from the dynamical matrix of the $G\bar{o}$ model, after a conjugate-gradient relaxation of the PDB structures. In order to determine T_f, we have performed thermal unfolding simulations coupled to multiple-histogram reconstructions [14] of the specific heat profiles.

We have analyzed a number of protein structures [15], ranging in length from $N = 57$ to $N = 177$ and heterogeneous for α and β content (see Fig. 1). As $T \to 0$, TICs are ordered following the equipartition theorem, namely $\Theta_k(0) = \omega_k^{-2}/\sum_m \omega_m^{-2}$. Upon raising the temperature, one expects nonlinear effects to come into play and cause energy flow among NMs. The first notable finding is that a restricted number of low-frequency normal modes is still able to capture a sizable fraction of the equilibrium fluctuations over a broad temperature range, from $T = 0$ to about 10%–20% below T_f. This is clearly shown for two representative proteins in Fig. 2. However, while the TICs of Lysozyme (166L) show the signs of little energy redistribution, those of the LUSH protein (1OOI) undergo a significant reshuffling, with the fifth mode raising alone to a maximum at room temperature (about 10% below T_f). Remarkably, TICs of other collective modes either collapse as the structure unfolds (first mode) or remain strikingly close to their equilibrium value (second mode). It is interesting to note that, at a closer look, also the fifth TIC of lysozyme increases with temperature, even though not as dramatically as for the LUSH protein. These results manifestly warn on reconstructions of conformational changes through zero-temperature NMs, as these are not guaranteed to provide meaningful descriptions of fluctuations at physiological temperatures.

The reshuffling of TICs at intermediate temperatures strongly suggests that a global measure gathering spectral weight from several low-frequency NMs would be a more stable and reliable indicator over the whole thermal span [see Figs. 2(b) and 3]. Hence, subspace-based, rather than single-mode-based reconstructions carry the relevant spectral weight. We remark nonetheless that the dimension of the relevant subspace is likely to be a protein-specific feature. Incidentally, this sheds further light on the observed correlation between low-frequency subspaces of NMs and essential modes pointed out in Ref. [16].

As T_f is approached, a loss of spectral weight is invariably observed for all NMs. Above T_f, TICs seem no longer sensitive to structural rearrangements. However, the nature of such transitions appears rooted in the special structural features of proteins. More precisely, TICs turn out to relax more or less abruptly. In this regard, crucial issues seem to be (i) the degree of structural arrangement at the secondary level and (ii) deviations from globularity of the scaffold. (i) To address the first point, we generated for a few structures self-avoiding polymers of the same length as the corresponding proteins and confined in the same ellipsoid as calculated through the corresponding inertial tensor. The necessary excluded volume constraints and local persistence length properties were imposed by requiring that the three-body radii be never lower than 2.7 Å, as described in Ref. [17]. By doing so, the ensuing structures show the same internal static two- and three-body correlations as the original proteins, while almost lacking the distinctive occurrence of α and β motifs. As shown in Fig. 3(a), TICs from the *mother* structure seem to set an upper bound for the temperature-dependent spectral content within the ensemble of surrogates, as if the true protein maximized the ability of a few collective coordinates to capture a sizable fraction of its fluctuations up to room temperature. This has been verified for different proteins in

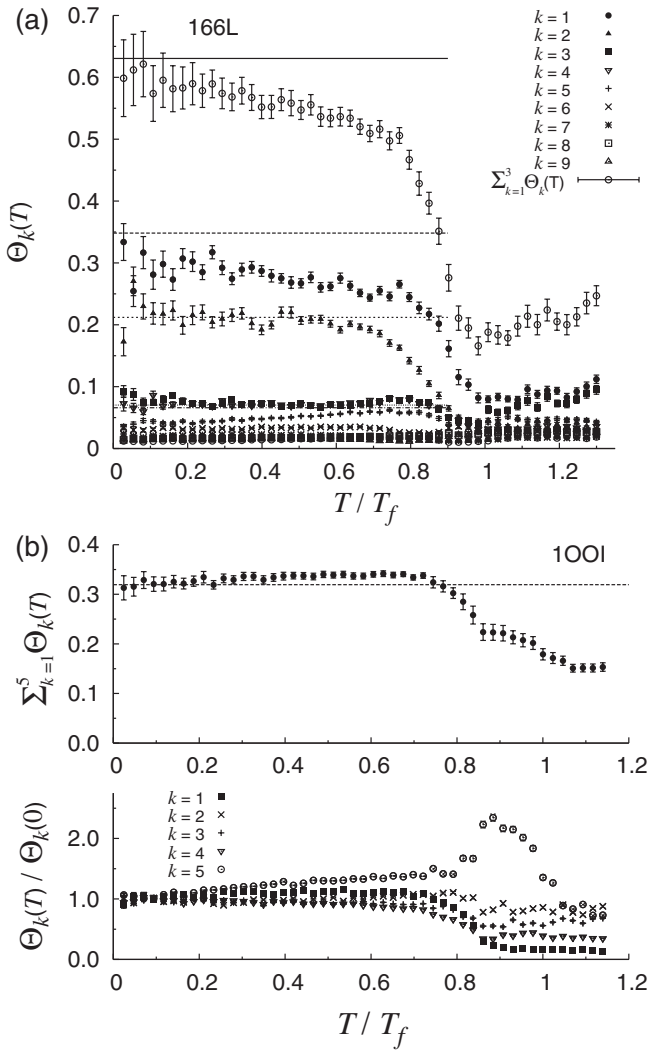

FIG. 2. (a) First nine thermal involvement coefficients for lysozyme (PDB code 166L), $N = 162$, $T_f = 0.75$, as functions of the reduced temperature. (b) Relative TICs for the odorant binding protein LUSH (PDB code 1OOI), $N = 124$, $T_f = 0.86$. In the lower panel TICs are normalized to the $T = 0$ values (see text). In all graphs horizontal lines mark single $[\Theta_k(0)]$ and aggregated $[\sum_{m=1}^{3,5} \Theta_m(0)]$ TICs.

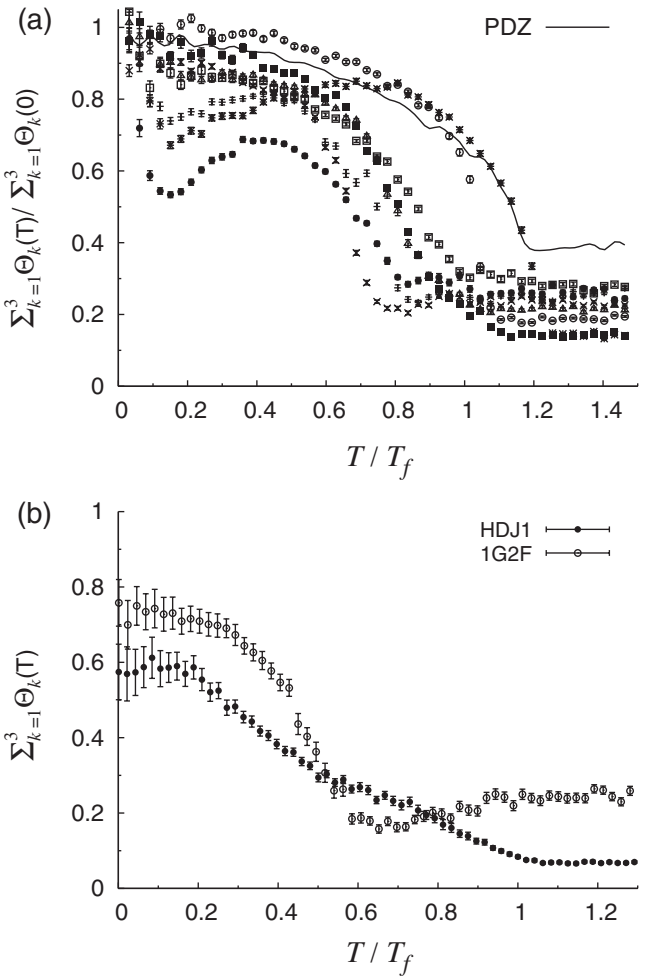

FIG. 3. Cumulative involvement coefficient over the first three modes. (a) PDZ III domain, $N = 85$, $T_f = 0.67$, PDB code 1BFE (solid line) and different independent surrogates (symbols). (b) Peptide-binding fragment from Hsp40, $N = 171$, $T_f = 0.96$ (PDB code HDJ1) and TATA-box Zinc-finger binding protein, $N = 177$, $T_f = 0.89$ (PDB code 1G2F).

our ensemble. Hence, we reaffirm the principle that functional motions that can be captured through NM analysis are rooted in the special arrangements of protein folds [9,18–20]. This is the second important point that we stress here. (ii) Not all proteins possess globular folds, some DNA- and peptide- binding fragments, for example, being characterized by less regular shapes, necessary to adapt to their targets upon binding. This, in turn, implies a varying degree of structure-related cooperativity in the dynamics. Therefore, it is instructive to assess the impact of deviations from globularity on the temperature trend of TICs in such cases. In Fig. 3(b), we show that the TICs describing fluctuations of two highly nonglobular proteins (see cartoon in Fig. 1), are indeed characterized by a weak degree of cooperativity. In particular, it is clear that they no longer represent a faithful measure of the overall dynamics already well below T_f.

Besides shape, we expect that the energy landscape features have a crucial impact on the temperature dependence of TICs. To prove this, we have performed the same simulations by switching off the dihedral terms (3), which confer substantial cooperativity to the folding process within a $G\bar{o}$ scheme [21]. We expect that, due to the lack of the dihedral term, TICs will show a behavior similar to that of nonglobular proteins, also little cooperative. Figure 4 documents the validity of our inference in the case of two highly structured proteins. Note that the strong cooperativity and thermal resilience of TICs are both profoundly eroded starting from relatively small temperatures. This completes the series of our findings.

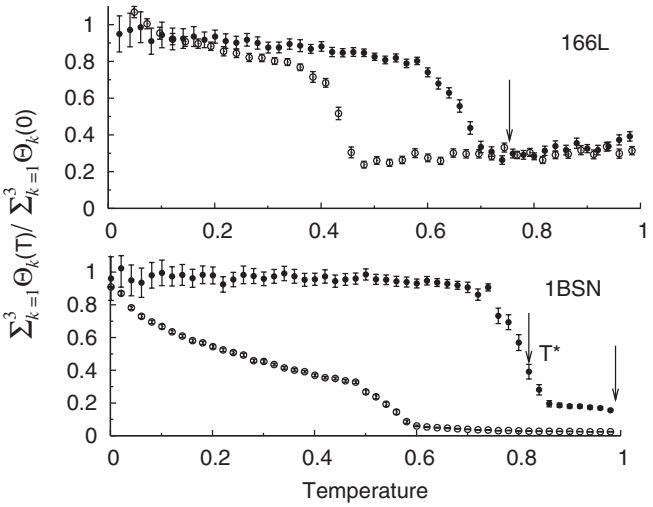

FIG. 4. Cumulative involvement coefficient with the full potential (filled circles) and without dihedral forces (empty circles) normalized to the zero-temperature values. Arrows mark the unfolding temperature calculated with the full force field. Upper panel: Lysozyme, (64%α, 9%β), $N = 162$, $T_f = 0.754$ (PDB code 166L). Lower panel: ϵ-subunit of the F1-ATPase, (26%α, 37%β), $N = 138$, $T_f = 0.99$ (PDB code 1BSN). Also marked is the temperature T^*, at which a first structural rearrangement occurs with the loss of about 25% of native contacts.

In this Letter we have determined upon what constraints and for what systems NMs can faithfully reproduce the dynamics of proteins from $T = 0$, where they are rigorously defined, up to temperatures close to the protein unfolding temperature, T_f. We have inquired into the temperature sensitivity of a NM description using the thermal involvement coefficients, that gauge how much of the protein thermal fluctuations around the native state are captured by NMs. We have shown that sizable values and thermal persistence of TICs, and thus the reliability of the NMs description, are deeply rooted in the presence of secondary motifs and in the degree of their packing within protein scaffolds. The fluctuations of more irregular, less compact and less cooperative proteins are manifestly harder to capture through normal modes. Although our results confirm the validity of NM analysis of protein functional motions, we have shown that subspaces, typically spanned by the first few low-frequency modes, rather than single normal modes, ought to be considered as identifying the relevant functional motions. In fact, the *spectral capacity* of individual modes, as gauged by TICs, often differs substantially at physiological temperatures with respect to the same measure at zero temperature, as a consequence of nonlinearity-induced intermode energy flow. Moreover, we have brought to the fore a crucial *liaison* between the meaningfulness of a NM-based description of functional fluctuations at room temperature and the degree of cooperativity of proteins.

The authors wish to thank Bingdong Sha for making the HDJ1 structure available prior to publication.

[1] I. Bahar and A. J. Rader, Curr. Opin. Struct. Biol. **15**, 586 (2005).
[2] D. Case, Curr. Opin. Struct. Biol. **4**, 285 (1994).
[3] N. Go, T. Noguti, and T. Nishikawa, Proc. Natl. Acad. Sci. U.S.A. **80**, 3696 (1983).
[4] B. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **80**, 6571 (1983).
[5] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[6] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).
[7] A. Amadei, A. Linssen, and H. Berendsen, Proteins: Struct. Funct. Genet. **17**, 412 (1993).
[8] M. Balsera, W. Wriggers, Y. Oono, and K. Schulten, J. Phys. Chem. **100**, 2567 (1996).
[9] P. De Los Rios, F. Cecconi, A. Pretre, G. Dietler, O. Michielin, F. Piazza, and B. Juanico, Biophys. J. **89**, 14 (2005).
[10] G. P. Morriss and C. P. Dettmann, Chaos **8**, 321 (1998).
[11] Since the isokinetic scheme conserves both linear and angular momenta, no alignment is required before computing conformational changes, if both total momenta are set to zero at $t = 0$.
[12] C. Clementi, H. Nymeyer, and J. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[13] The dihedral angle at site i is defined by the two adjacent planes formed by the four consecutive C_α's at $i - 2$, $i - 1$, i, $i + 1$.
[14] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[15] PDB codes: 166L, 1FAS, 1OOI, HDJ1, 1BFE, 1BSN, 1CHN, 1FVQ, 1OPC, 1AWO, 1TIT, 1UBI, 1G2F.
[16] C. Micheletti, P. Carloni, and A. Maritan, Proteins: Struct. Funct. Bioinf. **55**, 635 (2004).
[17] J. R. Banavar, A. Flammini, D. Marenduzzo, A. Maritan, and A. Trovato, Complexus **1**, 4 (2003).
[18] F. Tama and Y. H. Sanejouand, Protein Engineering Design and Selection **14**, 1 (2001).
[19] C. Chennubhotla, A. J. Rader, L. W. Yang, and I. Bahar, Phys. Biol. **2**, S173 (2005).
[20] F. Tama and C. Brooks, Annu. Rev. Biophys. Biomol. Struct. **35**, 115 (2006).
[21] M. Knott, H. Kaya, and H. S. Chan, Polymer **45**, 623 (2004).