Topological Thermal Instability and Length of Proteins

Raffaella Burioni,[1] Davide Cassi,[1] Fabio Cecconi,[2] and Angelo Vulpiani[2]*
[1]*Dipartimento di Fisica and INFM, Università di Parma, Parco Area delle Scienze 7A, 43100 Parma, Italy*
[2]*Dipartimento di Fisica and INFM (UdR and SMC), Università "*La Sapienza*," P.le A. Moro 2, 00185 Roma, Italy*

ABSTRACT **We present an analysis of the effects of global topology on the structural stability of folded proteins in thermal equilibrium with a heat bath. For a large class of single domain proteins, we computed the harmonic spectrum within the Gaussian Network Model (GNM) and determined their spectral dimension, a parameter describing the low frequency behavior of the density of modes. We found a surprisingly strong correlation between the spectral dimension and the number of amino acids in the protein. Considering that larger spectral dimension values relate to more topologically compact folded states, our results indicate that, for a given temperature and length of protein, the folded structure corresponds to a less compact folding, one compatible with thermodynamic stability. Proteins 2004;55:529–535.** © 2004 Wiley-Liss, Inc.

Key words: thermal fluctuations; spectral dimension; topological instability; Gaussian Network Model (GNM)

INTRODUCTION

The role of geometry has recently been considered as a factor of primary importance for the study of several physical properties of proteins and other biological macromolecules. In particular, since the topology of the folded state is known to influence the folding properties of the protein,[1–8] a great deal of work has been devoted to the study of those theoretical aspects that describe the networks of links among amino acids in folded proteins.[9–11] Furthermore, relevant features of protein conformations seem to follow the geometrical principles of the optimal packing problem[12,13] and mathematical concepts from graph theory have been interestingly applied to identify flexible and rigid regions of folded states.[14]

Starting from the primary, linear structure (the sequence of amino acids), a protein evolves during the folding process until it reaches a final state (native state) whose geometrical shape is crucial to the function of the protein itself. However, the problem of the geometrical arrangement of proteins in their native states cannot be regarded as a purely static issue. Indeed, a massive accumulation of experimental data collected from X-ray, NMR and neutron spectroscopy has revealed that protein native states are dynamic structures wherein amino acids constantly move around their equilibrium positions. This motion crucially involved in protein function,[15,16] is usually examined and investigated through normal modes analysis[17] (NMA) or essential dynamics.[18] However, the study of collective motion of large scale proteins is generally difficult due to limited access to realistic all-atom NMA,[19] and simplified or approximate approaches are usually welcome. Tirion[20] first proposed the possibility of replacing, in protein normal mode computations, complicated empirical potentials with Hookian pairwise interactions depending on a single parameter. This approach stems from the observation that low-frequency dynamics, which are mainly associated with protein-domain motion, are generally insensitive to the finer details of atomic interactions. Much of the subsequent literature[12–26] has confirmed the success of simple harmonic models in the study of the slow vibrational dynamics of large biological macromolecules, and they have become a viable alternative to heavy and time-consuming all-atom NMA. This success results from the striking agreement of predictions with experiments, the presence of few adjustable parameters and the fast and easy numerical implementation on computers. For these reasons, harmonic models are also utilized for the systematic analysis of large data sets of proteins.

The topological stability of macromolecules is far from being a purely mechanical problem as it closely involves thermodynamics. Indeed, the relevant thermodynamic potential that must be minimized in order to find a stable configuration is not energy but free energy. This is due to the interaction of molecules with the environment (schematized as a thermal bath), which is generally not negligible, especially for biological macromolecules that have a stable phase in the solvent. In particular, water is a very efficient medium for the transfer of thermal energy on the microscopic scale (i.e. oscillations and molecular rotations).

With these considerations in mind, in this work we apply NMA as an approach to investigate the influence of the global native state topology on the thermal stability of proteins.

Vibrational thermal instability is a well-known topic of study in solid state physics. Since the initial classical analysis by Peierls,[27] it has been recognized that equilibrium with a thermal bath can dramatically influence the possible topological arrangements of large geometrical structures. Up to now, the most striking consequence of

*Correspondence to: Angelo Vulpiani, Dipartimento di Fisica and INFM (UdR and SMC), Università "*La Sapienza*," P.le A. Moro 2, 00185 Roma, Italy. E-mail: vulpiani@roma1.infn.it.

Received 12 June 2003; 10 October 2003, 22 November 2003; Accepted 24 November 2003

Published online 14 April 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20072

Peierls' instability has concerned low-dimensional crystals: for one- and two-dimensional lattices, the mean square displacement of a single atom at a finite temperature diverges at the thermodynamic limit, i.e. with an increasing number of atoms. When the displacement exceeds the order of magnitude of the lattice spacing, the topological arrangement of the lattice is unstable and the crystal becomes a liquid. For real structures, formed by a finite number of units and far from the thermodynamic limit, the divergence sets a maximal stability size, which is negligible for one-dimensional lattices and typically mesoscopic for two-dimensional lattices.

However, thermal instability is present not only in crystals but also in structurally inhomogeneous systems, such as glasses, fractals, polymers and non-crystalline structures. Here, the problem is much more complex. Generalizing the Peierls approach to mesoscopic disordered structures, we are able to apply this argument to the thermal stability of macromolecules. In this article, we describe how this can be done in the case of proteins; we predict the existence of a critical stability size depending on a global topological parameter (the spectral dimension) and compare our predictions with experimental data.

Theory

In a recent paper[28] generalizing the Peierls result, we showed that thermodynamic instability also appears in inhomogeneous structures and is determined by the spectral dimension $\bar{d}$. The parameter $\bar{d}$[29] is defined according to the asymptotic behavior of the density of harmonic oscillations at low frequencies. More precisely, using $g(\omega)$ to denote the density of modes with frequency ω, then

$$g(\omega) \sim \omega^{\bar{d}-1} \tag{1}$$

as $\omega \to 0$. The spectral dimension is the most natural extension of the usual Euclidean dimension d to disordered structures as far as dynamic processes are concerned. It coincides with d in the case of lattices, but in general, it can assume non-integer values between 1 and 3. The spectral dimension represents a useful measure of the effective connectedness of geometrical structures on a large scale, because large values of $\bar{d}$ correspond to high topological connectedness. Moreover, it characterizes not only harmonic oscillations, but it also relates to diffusion, phase transitions and electrical conductivity, allowing a variety of both experimental and numerical methods for its determination.[30,31] The relevance of $\bar{d}$ in connection with the anomalous density of vibrational modes in proteins has also been considered in refs. [32] and [33].

In the case of thermal instability, we demonstrated that, for $\bar{d} \geq 2$, the mean square displacement $\langle r^2 \rangle$ of a structural unit (an atom, molecule or supra-molecular structure, according to the studied case) of a system composed of N elements, diverges in the limit $N \to \infty$. Using T to represent the temperature of the heat bath, k_B the Boltzmann constant, and γ the interaction energy scale, the divergence is given by the asymptotic law

$$\langle r^2 \rangle \propto \frac{k_B T}{\gamma} N^{2/\bar{d}-1} \tag{2}$$

when $\bar{d} < 2$. When $\bar{d} = 2$, the mean square displacement diverges logarithmically, $\langle r^2 \rangle \propto k_B T/\gamma \ln(N)$, as in the case of the Peierls result for a two-dimensional crystal. Notice that the divergence in $\langle r^2 \rangle$ is only determined by $\bar{d}$. Now, at any given temperature T, there will exist a threshold value $N(T)$ beyond which $\langle r^2 \rangle^{1/2}$ exceeds the typical spacing between the nearest neighbors, making the solid structure unstable. Therefore, at a large enough value of N, the solid will experience a structural reorganization that can lead to either a homogeneous liquid phase at sufficiently high temperatures or a disordered three-dimensional solid, which is homogeneous on a large scale and inhomogeneous on a small scale. In general, the threshold values of N are very small with respect to the typical order of magnitude of macroscopic systems comparable to the size of large complex macromolecules such as biopolymers.

This poses an intriguing question concerning proteins. Indeed, to exploit their biological function proteins must keep a specified geometric and topological arrangement and cannot afford even partial large-scale geometric fluctuations such as happen to swollen polymeric chains in a good solvent.[34] This makes thermodynamic stability crucial and suggests a possible correlation between spectral dimension and length in protein chains.

Vibrational stability in proteins has been analyzed with the Gaussian network model (GNM) proposed by Bahar et al.[35] and widely applied because it yields results in agreement with principal X-ray spectroscopy experiments. This approach generally considers proteins to be elastic networks whose nodes correspond to the positions of the alpha-carbons (C_α) in the native structure, and the interactions among nodes are modelled as harmonic springs. The only information required to implement the method is knowledge of the native structure. Two parameters are introduced, the spring constant and the interaction cut-off, which turn out to be related whenever the model is applied to fit experimental data. The GNM can be defined by the quadratic Hamiltonian equation

$$\sum_i \frac{\mathbf{p}_i^2}{2M} + \frac{\gamma}{2} \sum_{ij} \Delta_{ij} (\delta \mathbf{r}_i - \delta \mathbf{r}_j)^2 \tag{3}$$

where the first term represents the kinetic energy of the system, γ represents the strength of the springs that are assumed to be homogeneous, $\mathbf{R}_i$ and $\delta \mathbf{r}_i$ represent the equilibrium position and the displacement with respect to $\mathbf{R}_i$ of the i-th C_α atom, respectively. The model is eventually defined by the contact matrix Δ with the following entries: $\Delta_{ij} = 1$ if the distance $|\mathbf{R}_i - \mathbf{R}_j|$ between two C_α atoms, in the native conformation, is below the cutoff R_0, while $\Delta_{ij} = 0$ otherwise.

The harmonic spectrum for each structure is given by the set of eigenvalues $\{\omega_0, \omega_1, \ldots, \omega_{N-1}\}$ of the Kirchhoff matrix (or valency-adjacency matrix), $\Gamma_{ij} = -\Delta_{ij} + \delta_{ij} \sum_{l \neq i} \Delta_{il}$. Notice that the first eigenvalue, ω_0, vanishes and corresponds to the constant eigenvector related to the trivial uniform translation.

The comparison between experimental data and predicted GNM results is obtained via X-ray crystallographic

B-factors, measuring the mean square fluctuation of C_α atoms around their native positions

$$B_i(T) = \frac{8\pi^2}{3}\langle \delta\mathbf{r}_i \cdot \delta\mathbf{r}_i\rangle,$$

with $\langle \cdot \rangle$ indicating the thermal average. In the GNM approximation, this average is easily carried out, because amounts to a Gaussian integration, and B-factors can be expressed in terms of the diagonal part of the inverse of the matrix Γ:[35]

$$\langle \delta\mathbf{r}_i \cdot \delta\mathbf{r}_j\rangle = \frac{3k_BT}{\gamma}[\Gamma^{-1}]_{ij}$$

Knowledge of the eigenvectors and eigenmodes of matrix Γ allows computation of the GNM B-factors using the following formula:

$$B_i(T) = \frac{8\pi^2 k_B T}{\gamma}\sum_k \frac{|u_i(k)|^2}{\omega_k^2},$$

where i is the residue index, the sum runs over all non-zero frequencies ω_k and $u_i(k)$ indicates the i-th component of the k-th eigenmode.

Comparison with crystallographic data is crucial for setting the correct values of the parameters R_0 and γ (see Methods and Results).

METHODS

We present a GNM harmonic analysis performed over the data set of protein native structures with different sizes downloaded from the Brookheaven Protein Data Bank. The purpose of the analysis is basically to investigate whether there exists a correlation between the spectral dimensions of native structures and the lengths of naturally occurring proteins and, if so, to verify whether the correlation can be explained in terms of the abovementioned stability criterion determined by eq. (2).

Our representative statistical sample, listed in Tables I and II, was selected according to the following criteria. First, we only considered proteins with stable large-scale geometries. This excluded multiple-domain proteins, wherein domains can undergo relative motion, giving rise to larger geometric fluctuations. Moreover, we considered only proteins not bonded to fragments of DNA, RNA or other substrates because such structures cannot be described with sufficient accuracy in terms of a simple harmonic model with only two effective parameters. Finally, we chose proteins that uniformly represent a wide length interval ranging from 100 to 3600 residues to test our prediction.

The diagonalization of the Kirchhoff matrix Γ_{ij} to obtain its eigenvalues $\{\omega_1^2, \ldots, \omega_n^2\}$ and eigenvectors was performed with the standard numerical packages.[36]

The value of the interaction cut-off for generating the contact matrix Δ was set to $R_0 = 7$ Å, as is customary in such studies. The cut-off choice, which affects the overall GNM performance, is generally tested through the correlation coefficient ρ[37]

$$\rho = \frac{\sum_{ij}(B_i - \langle B_i\rangle)(X_j - \langle X_j\rangle)}{\sqrt{\sum_{ij}(B_i - \langle B_i\rangle)^2(X_j - \langle X_j\rangle)^2}},$$

between experimental (X_i) and theoretical (B_i) B-factors. The sum runs over a number of protein residues, and $\langle B\rangle$ and $\langle X\rangle$ indicate the average values. Our data set contains only those protein structures with a coefficient ρ greater than 0.5 (see last columns of Tables I and II). This should, in principle, ensure that GNM correctly reproduces C_α fluctuations for each selected protein. However, since we studied two different cut-offs, we decided to include even those few structures, such as 9RNT, 1A47, and 1CDG, that have $\rho > 0.5$ for one cut-off, and those with $\rho < 0.5$ for the other one. The few instances of agreement between B-factors from GNM and crystallography are shown in Figure 1, where we display the best and the worst cases with respect to the coefficient ρ.

For each protein, the optimal value of the spring constant γ was obtained through a least-square fitting to the experimental B-factors, which yielded the formula

$$\frac{k_BT}{\gamma} = \frac{1}{8\pi^2}\frac{\sum_i B_iX_i}{\sum_i B_i^2}$$

The values of k_BT/γ, besides being an essential ingredient for the application of the GNM method, are also an indication of the protein global flexibility and allow for a direct comparison among the structures considered.

The spectral dimension $\bar{d}$ was estimated via a power-law fitting to the low-frequency behavior of the cumulated density of modes $G(\omega)$, namely the integral of $g(\omega)$. Indeed, due to relation (1), $G(\omega) \sim \omega^{\bar{d}-1}$ at small arguments (see Fig. 2). The harmonic spectra, obtained within the GNM, for three proteins with sizes, small, medium and large, respectively are plotted in Figure 2, where low-frequency regions clearly exhibit the expected power-law behavior whose exponent is the spectral dimension $\bar{d}$.

RESULTS AND DISCUSSION

Our statistical analysis for the whole data set of proteins and cut-offs with $R_0 = 7$ Å is summarized in Table I, where we report the spectral dimension and its corresponding error, our estimate for k_BT/γ, and the correlation coefficient. To test the accuracy of our results, we have repeated the same analysis at a slightly different cut-off, namely at $R_0 = 6$ Å, which yields a smaller but still significant correlation between experimental and theoretical B-factors (Table II). Error for $\bar{d}$-values, in both tables, was estimated to account for uncertainty due to the choice of fitting region for the power-law, because the slope of the linear fit (see Fig. 2) can change with variation of this region. Furthermore, error bars also take into account the correlation value ρ which indicates how GNM can faithfully reproduce the low-energy deformation of a given protein structure.

Relation (2) establishes a rather strong constraint between the spectral dimension and the maximum size $N_{\max}$ a protein can achieve. Since the stability is assumed to fail when the fluctuation $\langle r^2\rangle^{1/2}$ attains the same order of

TABLE I. List of Processed Native Protein Structures and Characteristics (I)

PDB code	Length	d^A	Error^B	K_BT/γ	Correl. (ρ)
9RNT	104	1.62	0.05	1.657	0.474
1BVC	153	1.56	0.01	0.392	0.698
1G12	167	1.89	0.01	0.793	0.584
1AMM	174	1.71	0.06	0.003	0.720
4GCR	185	1.73	0.04	0.001	0.711
1KNB	186	1.88	0.01	1.104	0.699
1CUS	197	1.86	0.01	0.914	0.731
1IQQ	200	1.84	0.01	0.480	0.626
2AYH	214	1.86	0.02	0.539	0.773
1AE5	223	1.93	0.02	0.952	0.531
1LST	239	1.77	0.01	0.982	0.647
1A06	279	1.78	0.03	2.184	0.623
1NAR	289	1.81	0.01	0.602	0.696
1A48	298	1.72	0.01	0.664	0.549
1A3H	300	1.90	0.02	0.719	0.553
1SBP	309	1.74	0.02	0.641	0.757
1A5Z	312	1.74	0.01	2.111	0.574
1A1S	313	1.89	0.01	1.068	0.600
1ADS	315	1.79	0.03	0.500	0.687
1A40	321	1.90	0.04	0.524	0.546
1A54	321	1.86	0.03	0.601	0.516
1A0I	332	1.71	0.03	1.109	0.826
3PTE	347	1.79	0.01	0.366	0.840
1A26	351	1.82	0.01	1.369	0.635
1BVW	360	1.87	0.02	0.652	0.639
8JDW	360	1.94	0.01	1.293	0.607
7ODC	387	1.92	0.01	0.859	0.620
1OYC	399	1.93	0.01	1.056	0.697
1A39	410	1.97	0.01	1.113	0.656
16PK	415	1.82	0.03	0.630	0.590
1DY4	441	1.88	0.02	0.785	0.614
1BU8	446	1.95	0.01	0.859	0.632
1AC5	483	1.87	0.01	1.091	0.709
1LAM	484	1.97	0.01	0.488	0.583
1CPU	495	1.92	0.02	0.620	0.729
3COX	500	1.92	0.02	0.491	0.670
1A65	504	2.09	0.01	1.042	0.606
1SOM	528	2.00	0.02	1.585	0.653
1E3Q	532	1.97	0.01	1.577	0.623
1CRL	534	2.00	0.01	0.969	0.652
1AKN	547	1.87	0.01	1.737	0.667
1CF3	581	2.01	0.03	1.154	0.639
1EX1	602	2.01	0.03	1.193	0.598
1A14	612	2.10	0.09	0.865	0.524
1MZ5	622	2.02	0.04	0.750	0.705
1CB8	674	1.92	0.02	1.164	0.630
1HMU	674	1.92	0.02	0.907	0.684
1A47	683	2.02	0.04	0.646	0.529
1CDG	686	1.98	0.02	1.074	0.593
1DMT	696	1.96	0.02	1.204	0.536
1A4G	780	1.98	0.03	0.591	0.567
1HTY	1014	2.07	0.05	0.646	0.766
1KCW	1017	2.05	0.03	2.130	0.638
APP1	1021	1.93	0.02	0.805	0.576
1KEK	2462	2.07	0.05	1.263	0.730
1B0P	2462	2.08	0.09	0.319	0.810
1K83	3494	2.01	0.01	2.030	0.659
1I3Q	3542	1.97	0.01	2.435	0.758
1I50	3558	1.98	0.02	2.236	0.701

[A]d represents the structures' corresponding spectral dimensions estimated by GNM with cut-off R_0 = 7 Å.
[B]Error in determining spectral dimension.

TABLE II. List of Processed Native Protein Structures and Characteristics (II)

PDB code	Length	$\bar{d}^A$	Error^B	K_BT/γ	Correl. (ρ)
9RNT	104	1.43	0.08	0.209	0.549
1BVC	153	1.36	0.02	0.186	0.626
1G12	167	1.61	0.04	0.412	0.599
1AMM	174	1.55	0.09	0.113	0.802
4GCR	185	1.48	0.09	0.002	0.689
1KNB	186	1.70	0.01	1.104	0.699
1CUS	197	1.71	0.03	0.453	0.693
1IQQ	200	1.57	0.02	0.199	0.625
2AYH	214	1.68	0.01	0.222	0.756
1AE5	223	1.66	0.02	0.396	0.537
1LST	239	1.59	0.01	0.441	0.707
1A06	279	1.52	0.01	0.907	0.621
1NAR	289	1.54	0.01	0.257	0.731
1A48	298	1.46	0.01	0.235	0.546
1A3H	300	1.71	0.01	0.342	0.414
1SBP	309	1.62	0.03	0.301	0.718
1A5Z	312	1.57	0.02	0.914	0.539
1A1S	313	1.70	0.01	0.523	0.643
1ADS	315	1.56	0.02	0.204	0.611
1A40	321	1.57	0.01	0.199	0.604
1A54	321	1.57	0.03	0.232	0.543
1A0I	332	1.60	0.01	0.492	0.799
3PTE	347	1.66	0.01	0.180	0.840
1A26	351	1.60	0.03	0.602	0.613
1BVW	360	1.73	0.02	0.297	0.527
8JDW	360	1.71	0.01	0.550	0.537
7ODC	387	1.54	0.01	0.301	0.586
1OYC	399	1.74	0.01	0.472	0.659
1A39	401	1.78	0.02	0.473	0.643
16PK	415	1.67	0.04	0.277	0.591
1DY4	441	1.84	0.02	0.357	0.535
1BU8	446	1.76	0.01	0.331	0.538
1AC5	483	1.60	0.02	0.482	0.646
1LAM	484	1.75	0.01	0.204	0.623
1CPU	495	1.67	0.01	0.235	0.546
3COX	500	1.72	0.01	0.202	0.571
1A65	504	1.86	0.03	0.421	0.701
1SOM	528	1.63	0.02	0.560	0.610
1E3Q	532	1.67	0.02	0.570	0.533
1CRL	534	1.81	0.01	0.448	0.648
1AKN	547	1.71	0.01	0.800	0.641
1CF3	581	1.73	0.05	0.473	0.560
1EX1	602	1.73	0.01	0.401	0.544
1A14	612	1.86	0.02	0.373	0.538
1MZ5	622	1.68	0.02	0.274	0.740
1CB8	674	1.66	0.02	0.425	0.627
1HMU	674	1.67	0.01	0.334	0.652
1A47	683	1.75	0.01	0.236	0.376
1CDG	686	1.76	0.01	0.402	0.454
1DMT	696	1.73	0.09	0.484	0.549
1A4G	780	1.99	0.09	0.244	0.554
1HTY	1014	1.70	0.07	0.267	0.739
1KCW	1017	1.82	0.01	0.918	0.581
APP1	1021	1.83	0.05	0.299	0.572
1KEK	2462	1.90	0.04	0.439	0.664
1B0P	2462	1.94	0.02	0.118	0.695
1K83	3494	1.90	0.04	0.781	0.631
1I3Q	3542	1.94	0.02	0.883	0.691
1I50	3558	1.96	0.05	0.816	0.653

[A]d represents the structures' corresponding spectral dimensions estimated by GNM with cut-off R_0 = 6 Å.
[B]Error in determining spectral dimension.

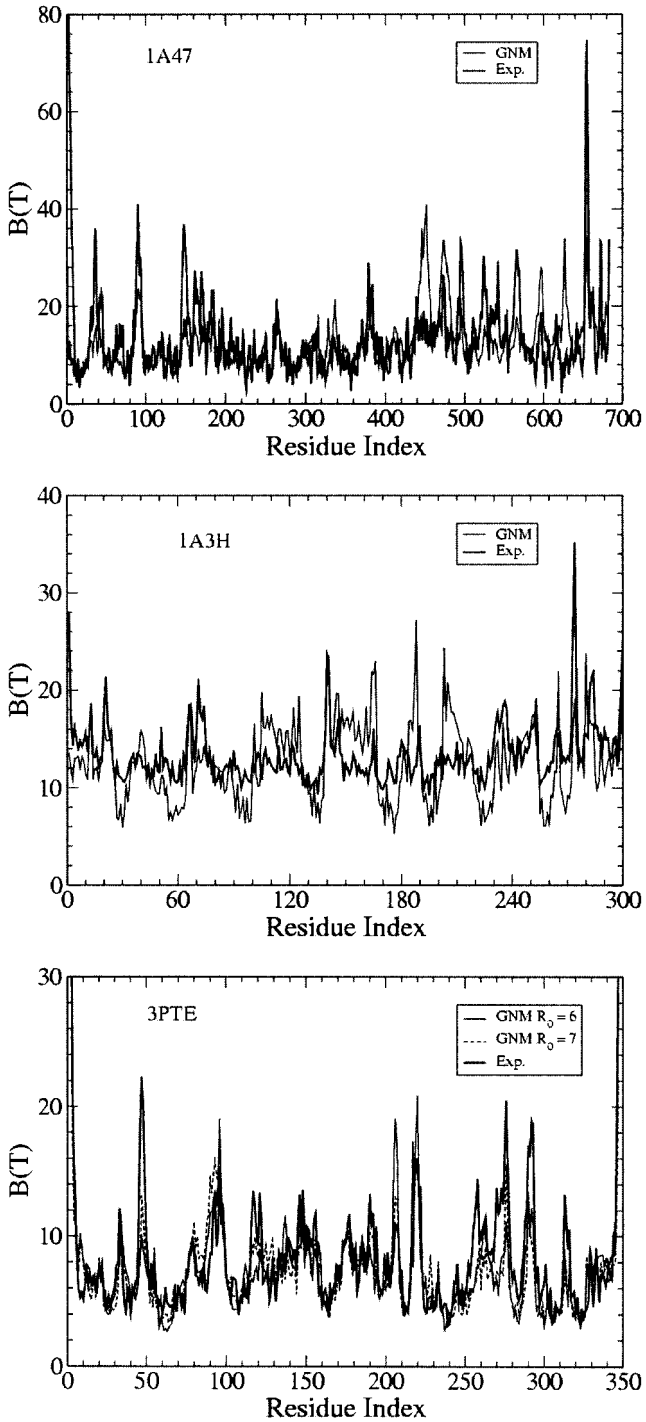

Fig. 1. Comparison between experimental B-factors and mean square fluctuations of C_α by GNM for structures 1A47 (lowest correlation) and 3PTE (highest correlation) at cut-off R_0 = 7Å, and structures 9RNT (lowest correlation) and 3PTE (highest correlation) at cutoff R_0 = 6Å. Heavy solid line refers to crystallographic data, while thin and dashed lines refer to GNM approximation.

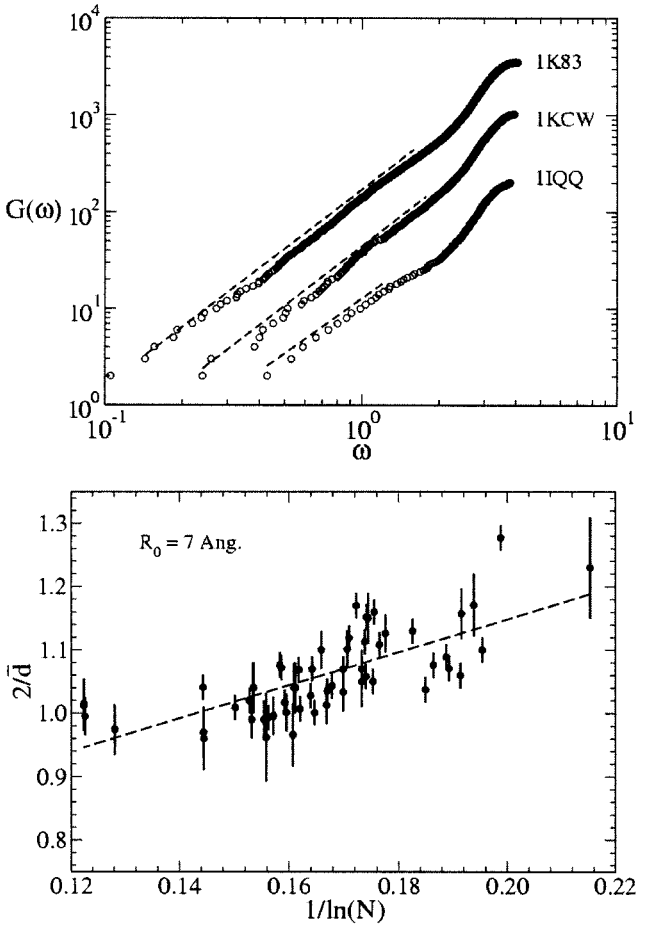

Fig. 2. Log-log plot of GNM-harmonic spectrum of three proteins with different sizes, 1IQQ (N = 200), 1KCW (N = 1017) and 1K83 (N = 3494). On the vertical axis, we report the cumulated distribution $G(\omega)$ of vibrational modes. Low frequency regions clearly exhibit a power-law behavior, and dashed lines indicate the best-fit of the power-law, whose exponent is the spectral dimension.

magnitude as the mean distance between non-consecutive amino acids (about 7 Å), one can assume that

$$\frac{2}{\bar{d}} = 1 + \frac{b}{\ln(N_{\max})}.$$

The proportionality constant b depends on the mean amino acid spacing, the spring elastic constant γ and the temperature T. However, this dependence is expected to be very weak (i.e. only logarithmic), and this allows for comparison among different proteins without computation of the specific parameters. It should be stressed that eq. (9), based solely on thermodynamic stability, can actually be regarded as an upper bound prediction only.

Figure 3 verifies the prediction drawn form the thermodynamic stability argument and shows the final result of our analysis. We plot the quantity $2/\bar{d}$ versus $1/\ln(N)$ as suggested by relation (9). Indeed, if eq. (6) holds, we should obtain a straight line crossing the y-axis at 1 for a zero abscissa.

As matter of fact, our data are well fitted by a straight line, but there is an offset with respect to eq. (6):

$$\frac{2}{\bar{d}} = a + \frac{b}{\ln(N)}.$$

For a cut-off R_0 = 7 Å, the best-fit values of the parameters are $a = 0.63$, $b = 2.61$, with a correlation coefficient of 0.73.

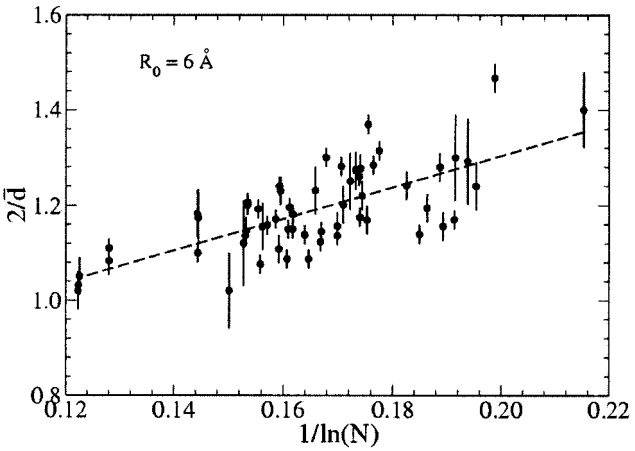

Fig. 3. Linear plot showing the dependence of spectral dimension on protein size. The dashed line, indicating the behavior (7), is the best fit, with a correlation coefficient 0.73 and 0.72 for cut-offs 7Å and 6Å respectively.

For the cut-off $R_0 = 6$ Å, we obtained the values $a = 0.63$, $b = 3.40$, with a correlation of 0.72. Interestingly, the linear behavior predicted by eq. (9) was confirmed for two different cut-offs with a correlation larger than 0.7, providing evidence of the repeatability of the result.

CONCLUSIONS

We applied GNM to investigate the influence of native state topology on thermodynamic stability for a set of folded proteins with sizes ranging from 100 to 3600. Employing GNM is appropriate in this type of study because such a model correctly accounts for the topological features of the native protein conformations. Our results show that the spectral dimension $\bar{d}$, which is sensitive to the large-scale topology of a geometric structure, is one parameter governing the low-energy fluctuations of a given protein structure. As a consequence, one can derive an instability criterion for proteins, based only on topological considerations, which is analogous to Peierls' criterion developed for ordered crystalline structures. The criterion easily predicts the non-trivial logarithmic dependence of the spectral dimension on the length of a protein. This further confirms the lack of universality for the spectral dimensions of proteins,[32] an issue addressed in previous studies.[24] We verified that such a logarithmic dependence is really observed, within statistical and systematic error, for the whole set of selected proteins. Furthermore, the dependence is robust because it applies even with alteration of the interaction cut-off, which is the most critical parameter to GNM applicability. We can conclude that the relation between spectral dimension and length of proteins is not a consequence of a particular cut-off choice, providing that a significant correlation is maintained between experimental and theoretical *B*-factors. We verified that, at a larger cut-off value, the scaling behavior (10) is preserved, although the spectral dimension grows due to the increase of the average connectivity of the elastic network.

The result expressed by eq. (10) requires some comments.

Eq. (10) is in agreement with the upper bound represented by eq. (9), supporting the relevance of topological thermal instability as a constraint to protein geometry. More importantly, not only is the upper bound satisfied, but the experimental points lie on a straight line parallel to the upper bound line of eq. (9). This suggests a more fundamental role of topological stability: the protein tends to arrange topologically in such a way to reach the minimum value compatible with stability constraints. In other words, for any fixed length, it tends to the most swollen state, which remains stable with respect to thermal fluctuations.

An interesting point is the meaning of the offset $a - 1$, which is predicted to be 0 according to eq. (9). Its positive value could have different explanations, but its universal nature (it is a "protein-independent" because is a global shift) must be due to a very general mechanism. A rather obvious reason is the contribution of anharmonic interactions at finite temperatures; a more intriguing one could be an effective longer range interaction due to the presence of bound water molecules around the external amino acids, which could change the effective form of the interaction matrix Δ. This hypothesis is also suggested by the physical interpretation of b as an anomalous dimension exponent, typically related to renormalized interactions.[38] However, the most intriguing evidence relies on the regression coefficient. Independent of the physical origin of b, its high value strongly supports the existence of a thermodynamic stability threshold, dependent on the topology of the folded state for the size of proteins.

REFERENCES

1. Gø N, Scheraga HA. On the use of classical statistical mechanics in the treatment of polymer chain conformation. Macromolecules 1976;9:535–549.
2. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1999;277:985–994.
3. Makarov DE, Keller CA, Plaxco KW, Metiu H. How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. Proc Natl Acad Sci USA 2002;99:3535–3539.
4. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and theory highlight role of native state topology in SH3 folding. Nature Struct Biol 1999;6:1016–1024.
5. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
6. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 1998;298: 937–953.
7. Klimov DK, Thirumalai D. Native topology determines force-induced unfolding pathways in globular proteins. Proc Natl Acad Sci USA 2000;97:7254–7259.
8. Cecconi F, Micheletti C, Carloni P, Maritan A. Molecular dynamics studies on HIV-1 protease drug resistance and folding pathways. Proteins: Struc Func Gen 2001;43:365–372.
9. Kabakcioglu A, Kanter I, Vendruscolo M, Domany E. Statistical properties of contact vectors. Phys Rev E 2002;65:041904.
10. Park K, Vendruscolo M, Domany E. Toward an energy function for the contact map representation of proteins. Proteins: Struct Func Gen 2000;40:237–248.
11. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world

view of the amino acids that play a key role in protein folding. Phys Rev E 2002;65:061910.
12. Maritan A, Micheletti C, Trovato A, Banavar JR. Optimal shapes of compact strings. Nature 2000;406:287–290.
13. Banavar JR, Maritan A. Geometrical approach to protein folding: a tube picture. Rev Mod Phys 2003;75:23-34.
14. Jacobs DJ, Rader AJ, Kunh LA, Thorpe MF. Protein flexibility prediction using graph theory. Proteins: Struct Funct Genet 2001;44:150–165.
15. Frauenfelder H, Petsko GA, Tsernoglou D. Temperature dependent X-ray diffraction as a probe as of protein structural dynamics. Nature 1979;280:558–563.
16. Frauenfelder H, McMahon B. Dynamics and functions of proteins: the search of general concepts. Proc Natl Acad Sci USA 1998;95: 4795–4797.
17. Karplus M, McCammon J. Dynamics of Proteins: Elements and Functions. Ann Rev Biochem 1983;53:263–300.
18. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins: Struct Func Gen 1993;17:412–425.
19. Levitt M, Sanders C, Stern PS. Protein normal-mode dynamics; trypsin inhibitor, crambin, ribonuclease, and lysozyme. J Mol Biol 1985;181:423–447.
20. Tirion MM. Low-amplitude elastic motions in proteins from a single-parameter atomic analysis. Phys Rev Lett 1996;77:1905–1908.
21. Hinsen K. Analysis of domain motion by approximate normal mode calculations. Proteins: Struct Funct Genetics 1999;33:417–429.
22. 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.
23. 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.
24. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett 1997;79:3090–3093.
25. Micheletti 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.
26. Micheletti C, Lattanzi G, Maritan A. Elastic properties of proteins: insight on the folding processes and evolutionary selection of native structures. J Mol Biol 2002;321:909–921.
27. Peierls RE. Bemerkung über Umwandlungstemperaturen. Helv Phys Acta 1934;7:S81–S83.
28. Burioni R, Cassi D, Fontana MP, Vulpiani A. Vibrational thermodynamic instability of recursive networks. Europhysics Lett 2002; 58:06–810.
29. Alexander S, Orbach RL. Density of states on fractals: fractons. J Phys Lett 1982;43:625–L631.
30. Burioni R, Cassi D. Universal properties of spectral dimension. Phys Rev Lett 996;76:1091–1093.
31. Saviot L, Duval E, Surotsev N, Jal JF, Dianoux AJ. Propagating to nonpropagating vibrational modes in amorphous polycarbonate. Phys Rev B 1999;60:18–21.
32. Ben-Avraham D. Vibrational normal-mode spectrum of globular proteins. Phys Rev B 1993;47:14559.
33. Elber R, Karplus M. Low frequency modes in proteins: use of effective-medium approximation to interpret fractal dimension observed in electron-spin relaxation measurements. Phys Rev Lett 1986;56:394–397.
34. De Gennes PG. Scaling concepts in polymer physics. Cornell University Press: Ithaca, 1979.
35. 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.
36. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes. Cambridge University Press: Cambridge, 1993.
37. Kundu S, Melton JS, Sorensen DC, Phillips GN Jr. Dynamics of proteins in crystal: comparison of experiment with simple models. Biophys J 2002;34:723–732.
38. Goldenfeld N. Lectures on phase transitions and the renormalization group. Frontiers in physics 85. Addison-Wesley Publishing Company: 1992.