Multifractality, Levinthal paradox, and energy hypersurface

M. A. Moret,[1,2] P. G. Pascutti,[2] K. C. Mundim,[3] P. M. Bisch,[2] and E. Nogueira, Jr.[3]

[1]*Departamento de Física, Universidade Estadual de Feira de Santana, Campus Universitário,*
44031-460 Feira de Santana, Bahia, Brazil
[2]*Instituto de Biofísica, Universidade Federal do Rio de Janeiro, 21949-900 Rio de Janeiro, RJ, Brazil*
[3]*Instituto de Física, Universidade Federal da Bahia, Campus de Ondina, 40210-340 Salvador, Bahia, Brazil*

Multifractal properties in the potential energy hypersurface of polypeptides and proteins are investigated. Characteristic multifractal behavior for different molecular systems is obtained from the $f(\alpha)$ spectra. The analysis shows that the dimension of the phase space of the problem influences the accessibility to different parts of the potential energy hypersurface. Also, we show that it is necessary to take into account the H-bond formation between amino acids in the conformational-folding search. The present findings indicate that the $f(\alpha)$ function describes some structural properties of a protein. The behavior of the $f(\alpha)$ spectra gives an alternative explanation about the Levinthal paradox. Furthermore, the anomalous temperature dependence of the Raman spin-lattice relaxation rates can be related to the perturbations in the secondary structures.

It is well known that a molecular system has a great number of conformations (minima in the potential energy hypersurface) [1–3], that increases with the number of degrees of freedom in the molecular system. In order to solve this kind of problem, a new strategy to obtain information about conformational energies was proposed [4], which encompasses both the evaluation of the classical force field and a generalization of the stochastic procedure called simulated annealing [5]. Thus, the generalized simulated annealing (GSA) [6], that is based on Tsallis statistics [7], was proposed, to replaced the Boltzmann weights used in the former method. The GSA procedure has been applied to a variety of problems such as genetic algorithm [8], molecule optimization using classical methods [4], or semiempirical methods [9], the geophysical problem [10], the traveling salesman problem [11], and numerical data fitting [12]. GSA proves to be the most effective simulated annealing method [4,11]. In this sense, GSA is a powerful algorithm to analyze potential energy hypersurface [4].

In this Rapid Communication we are mainly interested to investigate the acceptability of the potential energy hypersurface of the polypeptides and proteins as a function of the system size. Our strategy amounts to couple a GSA routine [4] with the THOR package, which has been developed to be a comprehensive and flexible tool to investigate macromolecular structure of biological interest [14]. This computational code is based on the GROMOS classical force field [13] with two available methods: molecular dynamics [14,15] and stochastic molecular optimization (SMO) [4]. This approach enables to scan the conformational space in the regions where the energies are close to local minima. On the other hand, regions where the energy value is greater than the local minima may be discarded in the stochastic procedure, so that the regions of the Ramachandran map [16] that are not allowed must be rejected in this procedure. These not-allowed regions modify the hyper-surface acceptability.

We apply this methodology to study proteins and polypeptides, using the force field presents in the THOR package. Furthermore, the conformational energy of the molecule is made up of a sum of bond and nonbond terms. In this approach, only the covalent hydrogen bonded to oxygen or nitrogen are considered explicitly, while the CH_1, CH_2, and CH_3 groups are assumed to be an atomic unit. The energy hypersurface is rather complex, and approximations in the force field are required to make such study tractable within our current numerical capabilities. Here, we use the following approximation for the force field:

$$E = E_S + E_\varphi + E_{vdW} + E_{el}, \qquad (1)$$

or equivalently

$$E = \frac{1}{2}\sum_m K_{S_m}(r_m + r_0)^2 + \frac{1}{2}\sum_n K_{\varphi_n}(1 + \cos(m\varphi_n + \varphi_0)^2)$$

$$+ \sum_{i<j}\left[\frac{C_{12}(i,j)}{r_{i,j}^{12}} - \frac{C_6(i,j)}{r_{i,j}^6}\right] + \frac{1}{4\pi\varepsilon_0\varepsilon_r}\sum_{i<j}\frac{q_i q_j}{r_{i,j}},$$

where E_S is the sulfur bond potential, E_φ is the dihedral angle potential, E_{vdW} is the van der Waals potential, and E_{el} is the Coulomb potential term. The E_S term is considered because, in various cases, this kind of bond stabilizes the potential energy close to the native structure. The backbone bond potential and the angular bond potential have their ideal values calculated by Corey and Pauling [17]. The bond originated by these potentials were maintained close to these values. Although these approximations in Eq. (1) smooth the hypersurface [18], these potentials do not contribute substantially to the amino acid sequence (primary structure) obtaining the tertiary structure (native structure). Therefore, these bonds do not change the protein folding substantially.

The search for minima (global and local), or the scan of the potential energy hypersurface consists in comparing the geometry obtained by GSA routine with the energy measured by THOR. GSA builds up random geometry changing the dihedral angles $\varphi(t)$ [4], as follows:

(i) Fix the parameters (q_A, q_v), related to acceptance and visiting index. Start, at $t=1$, with a high enough value for $T(1)$ (temperature) and cool as in Ref. [6].

(ii) Generate φ_{t+1} from φ_t ($\varphi_{t+1} = \varphi_t + \Delta\varphi_t$) using the visiting distribution probability g_{q_v} as defined in Ref. [4].

(iii) Evaluate the conformational energy values $E(\varphi_{t+1})$ using the force field, Eq. (1) and accepting it according to the generalization of the Metropolis criterion [6].

(iv) Calculate the new temperature $T(t)$ and go back to (ii) until $E(\varphi_t)$ converges within the desired precision.

In Ref. [4] the procedure to build up a GSA routine is reviewed. Recently, the SMO strategy has been applied to study helix folding where the Ramachandran map was used as constraints [19]. The SMO results are in close accordance with experimental [20] and theoretical [21] observations. In this paper, SMO was applied on various molecular systems and, in the minimization process, if $E(\varphi_{t+1}) < E(\varphi_t)$ these energy values were analyzed.

We performed an exhaustive potential hypersurface scanning with different initial GSA parameters, i.e., $T_0 = [1,60\,000]$, $q_v = [1.1,2.7]$ and $q_A = 1.1$. We have searched the minima (locals and globals ones) in the energy hypersurface for the following molecules: (i) polyalanines (with 5 to 20 residues of alanine), (ii) a Barnase fragment, (iii) fragments of the Insect Defensins, (iv) the complete Insect Defensin A, and (v) cysteine-protease inhibitor Leupetine.

The obtained results for the energy profiles associated with the potential energy hypersurface shows that the local energy values has a nonhomogeneous structure, which are distributed over the hypersurface. Further on, different regions in the energy hypersurface present similar profiles, when they are mapped using different scales. Different systems present this behavior that, at first sight, seems to be a statistically self-similarity one. This characteristic is typical of fractal objects. In this context we were motivated to characterize the energy profiles using a multifractal formalism [22]. This formalism is appropriated to describe self-similar local quantities that can be interpreted as measures. The analysis was mostly developed to describe a broad class of objects generated by physical processes and characterized by normalized stationary distributions (measures) assigned upon fractals sets [23]. In Ref. [24] it was shown that the fractal behavior in the proteins arises from the analysis of the stereochemical properties of tetrapeptides segments of the proteins. From molecular dynamics it was observed a universal fractal behavior due the temporal analysis [25].

Multifractal properties associated with the energy profiles can be obtained through the singularities of the $f(\alpha)$ spectrum as a function of the number of dihedral angles. One of the most efficient algorithms to calculated the $f(\alpha)$ spectrum was proposed in Ref. [26]. The implementation of this algorithm requires one to define the normalized measures ($\xi_i = E_i / \Sigma_i E_i$) and construct a parameterized family of normalized measures ($\mu_i = \xi_i^q / \Sigma_i \xi_i^q$). The $f(\alpha)$ spectrum is obtained by varying the parameter q and evaluating

$$f(\alpha(q)) = \lim_{\varepsilon \to 0} \left\{ -\frac{\sum_i \mu_i(q)\ln[\mu_i(q)]}{\ln \varepsilon} \right\}, \qquad (2)$$

$$\alpha(q) = \lim_{\varepsilon \to 0} \left\{ -\frac{\sum_i \mu_i(q)\ln \xi_i}{\ln \varepsilon} \right\}. \qquad (3)$$

where $\varepsilon = (1/2\pi)^D$ and D is the number of dihedrals angles optimized. In the next, we discuss the results from various systems analyzed using this formalism.
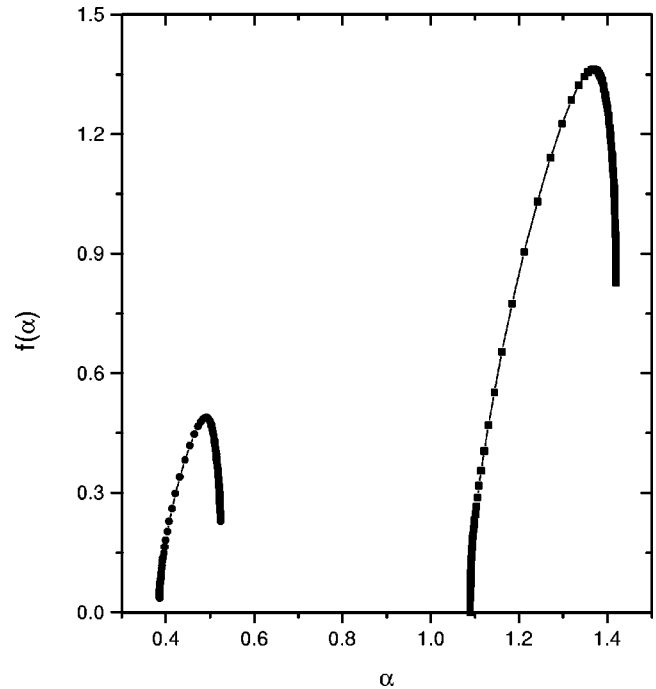

FIG. 1. The $f(\alpha)$ spectrum for 5 (squares) and 13-alanine (circles).

Polyalanine is a theoretical polypeptide probe to study helix folding [27–31]. Figure 1 depicts the 5 and 13-alanine $f(\alpha)$ spectrum, respectively, which shows a characteristic multifractal behavior of the polyalanines. These results point out that the maximum of the $f(\alpha)$ spectrum decreases as the dimension of the optimization problem increases.

The second system studied, Barnase, is a member of the microbial ribonuclease family. In Ref. [32] the 85-102 Barnase fragment (a β-hairpin) is studied by molecular dynamics. The authors obtained three different unfolded conformations ($C1$, $C2$, and $C3$). We observed that all $f(\alpha)$ calculated for these conformations are very similar.

The third system studied, Insect Defensin A, is a basic 4 kDa protein that, *in vivo*, is excreted in the hemolymph of the flesh fly *Phormia terramovae* larvae [33,34]. This small protein presents one α-helix and two β-strands, stabilized with three disulfide bridges [35]. This protein is an interesting system because the existence of the three disulfide bridges. From the simulations we observed that this kind of bonds (E_S) does not change the multifractal behavior, which means some independence of the $f(\alpha)$ spectrum in relation to the bond potential.

Note that all systems discussed above, present secondary structures (α-helix or β-sheet) as a minimum. To observe the influence of secondary structures in the $f(\alpha)$ spectrum we analyzed a class of larger polypeptides. In this analysis, we simulated polypeptides with 13 to 20 residues of alanine and 13-23 fragment of the Defensin, which lead to α-helix structures. On the other hand, the β-sheet structures obtained from the 85-102 fragment of the Barnase and 26-39 fragment (Defensin) present the same $f(\alpha)$ shape. In case of α-helix structure polypeptides, it was observed that a stable unfolded conformation (a random coil) increases from 5 to 15 Kcal/
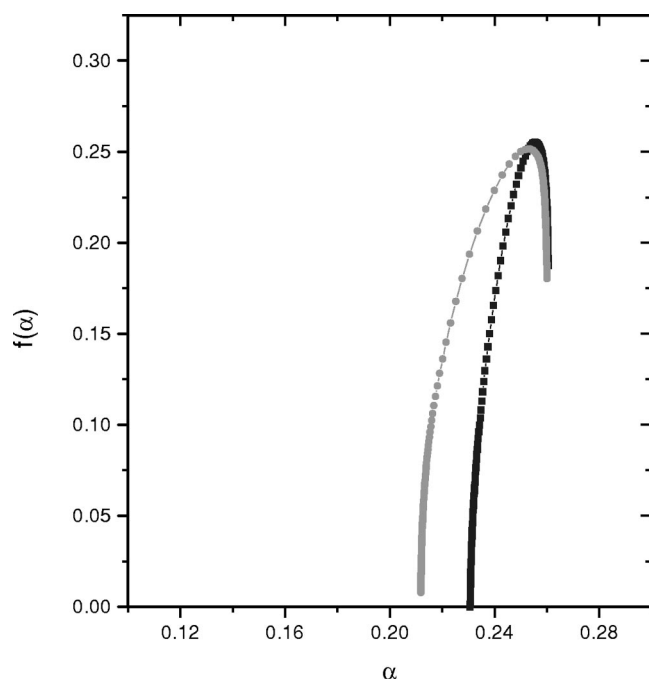
FIG. 2. The $f(\alpha)$ spectrum observed from the β-sheet of the Defensin. The correct β-sheet $f(\alpha)$ spectrum (squares) turns more sharply than the random coil one (circles).

FIG. 3. The shape of the $f(\alpha)$ spectrum for some systems: 5-alanine (black squares), 8-alaninie (black circles), leupetine (light gray triangles), 18-alanine (black triangles), α-helix of the Defensin (black diamonds), and Insect Defensin A (gray squares).

mol the conformational energy in relation to the global minimum (an α-helix). In case of β-sheet fragment of the Defensin (26-39 fragment), it results that the global minimum is an unfolded conformation, i.e., a random coil. However, the β-sheet conformation increases the conformational energy ~ 0.8 Kcal/mol only. Despite the α-helix and β-sheet different energetic behavior the shape of the $f(\alpha)$ spectrum maintaining as a function of the secondary structure existence. The last system studied, Leupeptine, is a potent cysteine-protease inhibitor. This short polypeptide presents no secondary structures and has a great mobility. For these reasons we investigated its multifractal behavior. We found that the $f(\alpha)$ spectrum from Leupeptine is larger than those for poly-alanines. The $f(\alpha)$ spectrum maximum obtained from Leupeptine is wider than that one obtained for 9-alanine that has the same optimization dimension, because its great mobility.

As it was observed, a common (and more fascinating) feature of the analyzed systems is a multifractal spectrum of energy which can be taken as their basic signatures. Figure 2 shows the typical behavior for all systems that we analyzed. It can be observed that the shape of the $f(\alpha)$ spectrum depends on the secondary structure formation (it depends on the number of H-bonds). Moreover, if a secondary structure was obtained in the simulation, the shape of the $f(\alpha)$ spectrum is sharper than in the opposite situation. In the simulations of the Defensin, it was observed that the disulfide bridges do not change the shape of $f(\alpha)$ spectra, which means that the multifractal behavior is independent of bond potential.

Figure 3 depicts the main characteristic of all systems discussed here, using the geometric (dihedral angles) analysis. This figure shows that the maximum of the $f(\alpha)$ function decreases as the dimension (real) of the optimization prob-
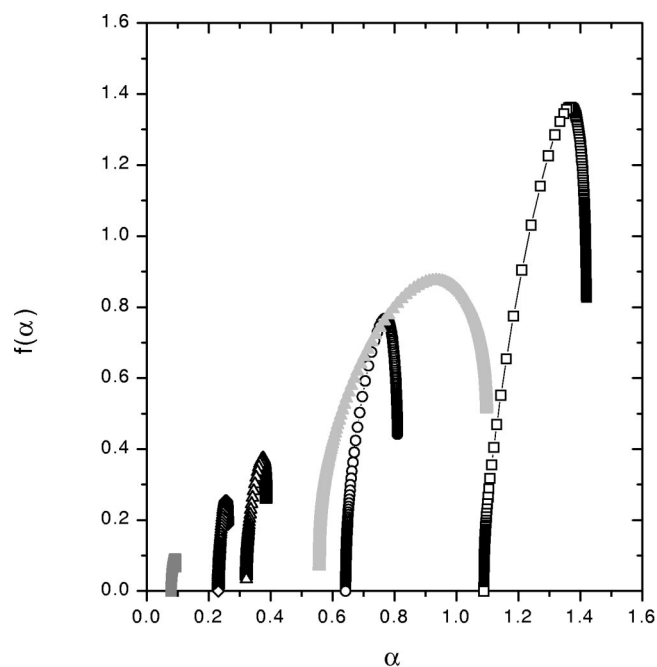
lem increases. This behavior helps the protein in order to search the native state. The evaluation of the potential energy hypersurface, using the procedure proposed, shows that the shape of the $f(\alpha)$ spectrum depends on the degrees of freedom of the problem. This result points out that the ability to populate the potential energy hypersurface depends on the dimension of the problem. Therefore, as the hypersurface increases with the number of degrees of freedom in the system, these polypeptides present a poor capability to populate this hypersurface. Initially this result seems to be contradictory but, if we take into account the Levinthal paradox [36] (namely, that a polypeptide chain can find its unique native state in spite of the astronomical number of configurations in the denatured state), it helps with understanding the protein folding, because the hypersurface presents a great number of regions that are not allowed . Thus, according to our simulations, the proteins adopt conformations only in the allowed regions by the $f(\alpha)$ spectrum. We recall that this result is in accordance to protein folding pathways because these pathways make inaccessible a great number of regions in the potential energy hypersurface. On the other hand, we observe that if we analyze the multifractality as a function of time (Monte Carlo step) the $f(\alpha)$ spectrum always presents the same shape and its maximum is close to 1. Thus, we find a universal multifractal behavior, but this behavior does not allow us to analyze the intrinsic properties of the protein hypersurfaces.

In summary, we investigate the multifractal properties of the potential energy hypersurface of the proteins. The simulations show that the $f(\alpha)$ spectra depend on the number of H-bonds. This result is more precise than one obtained by Ref. [37] using a lattice model, since in the present paper we performed simulations using a more accurate model for pro-

tein. The results for the lattice model indicate that the system is fractal and that dimension depends on the number of H-bonds [37]. Using a more precise model we have observed that, when the system analyzed has H-bonds, the $f(\alpha)$ spectrum shape is different in comparison to those systems that do not produce H-bonds. Therefore, the H-bond formation induces differences to the $f(\alpha)$ spectra shape. We want to stress that the shape of the $f(\alpha)$ spectra does not depend on the GSA parameters. These parameters are used only to map the energy hypersurface. Their choice does not alter the main conclusions of this paper significantly.

On the other hand, as the secondary structures (α-helices and β-sheets) depend on the environment, a protein tertiary structure (native structure) depends on this too. Then, a solution on the aqueous solvent may perturb the protein tertiary structure and, therefore, change the multifractal response. As it was shown, the multifractal response-behavior changes by perturbing the secondary structures. This feature seems to be related to the anomalous temperature dependence of the Raman spin-lattice relaxation rates [26,38], because an ionic strength solution changes the temperature dependence.

Therefore, the behavior of the anomalous temperature dependence of the relaxation rates can be related to the behavior of the secondary structures that changes the shape of $f(\alpha)$ spectra.

In relation to the $f(\alpha)$ spectra as a function of the polypeptide size, it was shown that the proteins adopt conformations in the potential energy hypersurface only in allowed regions from the $f(\alpha)$ spectrum. Thus, the potential energy hypersurface presents allowed (and not-allowed) regions that depend on the protein size, i.e., the number of degrees of freedom in the molecular system. This behavior shows an alternative explanation for the Levinthal paradox, since a great number of possible configurations must be in the not-allowed regions, making the number of allowed configurations decrease and restricting the protein to navigate over the folding routes only.

[1] J. M. Yon, Cell. Mol. lif. Sci. **53**, 557 (1997).

[2] P. G. Wolynes *et al.*, Science **267**, 1619 (1995).

[3] V. S. Pande, A. Y. Grosberg, T. Tanaka, and D. S. Rokhsar, Curr. Opin. Struct. Biol. **8**, 68 (1998).

[4] M. A. Moret, P. G. Pascutti, P. M. Bisch, and K. C. Mundim, J. Comput. Chem. **19**, 647 (1998).

[5] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).

[6] C. Tsallis and D. A. Stariolo, Physica A **233**, 395 (1996).

[7] C. Tsallis, J. Stat. Phys. **52**, 479 (1988); E. M. F. Curado and C. Tsallis, J. Phys. A **24**, L69 (1991); **24**, 3187(E) (1991); **25**, 1019 (1992); C. Tsallis, Phys. Lett. A **206**, 277 (1995); Braz. J. Phys. **29**, 1 (1999).

[8] M. A. Moret, P. M. Bisch, and F. M. C. Vieira, Phys. Rev. E **57**, R2535 (1998).

[9] K. C. Mundim and C. Tsallis, Int. J. Quantum Chem. **58**, 373 (1996).

[10] K. C. Mundim, T. Lemaire, and A. Bassrei, J. Phys. A **252**, 405 (1998).

[11] T. J. P. Penna, Phys. Rev. E **51**, R1 (1995).

[12] T. J. P. Penna, Comput. Phys. **9**, 341 (1995).

[13] W. V. van Gusteren and H. J. C. Berendsen, GROMOS (Groningen Molecular Simulation) software package.

[14] P. G. Pascutti, K. C. Mundim, A. S. Ito, and P. M. Bisch, J. Comput. Chem. **20**, 971 (1999).

[15] E. P. G. Arêas, P. G. Pascutti, S. Schreier, K. C. Mundim, and P. M. Bisch, J. Phys. Chem. **99**, 14 882 (1995).

[16] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, J. Mol. Biol. **24**, 95 (1963).

[17] R. B. Corey and L. Pauling, Proc. R. Soc. London, Ser. A **141**, 10 (1953).

[18] D. J. Wales and H. A. Scheraga, Science **285**, 1368 (1999).

[19] M. A. Moret, P. M. Bisch, K. C. Mundim, and P. G. Pascutti, Biophys. J. (unpublished).

[20] K. R. Shoemaker, P. S. Kim, E. J. York, J. M. Stewart, and R. L. Baldwin, Nature (London) **326**, 563 (1987).

[21] N. C. Rogers, in *Prediction of Protein Structure and Principles of Proteins Confirmations*, edited by Gerald D. Fasman (Plenum Press, New York, 1989).

[22] T. Vicsek, *Fractal Growth Phenomena*, 2nd ed. (World Scientific, Singapore, 1992).

[23] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, Phys. Rev. A **33**, 1141 (1986).

[24] G. C. Wagner, J. T. Covin, J. P. Allen, and H. J. Stapleton, J. Am. Chem. Soc. **107**, 5589 (1985).

[25] D. A. Lidar, D. Thirumalai, R. Elber, and R. B. Gerber, Phys. Rev. E **59**, 2231 (1999).

[26] A. B. Chhabra and R. V. Jensen, Phys. Rev. Lett. **62**, 1327 (1989).

[27] V. Daggett, P. A. Kollman, and I. D. Kuntz, Biopolymers **31**, 1115 (1991).

[28] D. J. Tobias and C. L. Brooks, Biochem. J. **30**, 60 590 (1991).

[29] S.-S. Sung, Biophys. J. **66**, 1796 (1994).

[30] S.-S. Sung, Biophys. J. **68**, 826 (1995).

[31] C. L. Brooks, J. Phys. Chem. **100**, 2546 (1996).

[32] M. Prévost and I. Ortmans, Proteins **29**, 212 (1997).

[33] J. Lambert, E. Keppi, J.-L. Dimarq, C. Wicker, J.-M. Reichhart, B. Dunbar, P. Lepage, A. van Dorsselaer, J. Hoffmann, J. Fothrgill, and D. Hoffmann, Proc. Natl. Acad. Sci. U.S.A. **86**, 262 (1989).

[34] S. Cociancich, A. Ghazi, C. Hetru, J. A. Hoffmann, and L. Letellier, J. Biol. Chem. **268**, 19 239 (1993).

[35] B. Cornet, J.-M. Bonmatin, C. Hetru, J. A. Hoffmann, M. Ptat, and F. Vovelle, Structure (London) **3**, 435 (1995).

[36] C. Levinthal, J. Chem. Phys. **65**, 44 (1968).

[37] J. S. Helman, A. Coniglio, and C. Tsallis, Phys. Rev. Lett. **53**, 1195 (1984).

[38] J. T. Colvin and H. J. Stapleton, J. Chem. Phys. **82**, 4699 (1985).