Self-similarity and protein chains

M. A. Moret,[1,2] J. G. V. Miranda,[3] E. Nogueira, Jr.,[3] M. C. Santana,[2] and G. F. Zebende[1,2]
[1]*CEPPEV, Fundação Visconde de Cairu, 40226-900 Salvador, BA, Brazil*
[2]*Departamento de Física, UEFS, 44031-460 Feira de Santana, Bahia, Brazil*
[3]*Instituto de Física, UFBA, 40210-340 Salvador, Bahia, Brazil*
(Received 13 January 2004; revised manuscript received 16 November 2004; published 27 January 2005)

Fractal properties of 5526 different protein chains are investigated. Characteristic fractal behavior for different molecular systems is obtained from the fractal dimension analysis, which shows that the dimension is $\delta=2.47$. This dimension gives a measure of the protein compactness. The present finding indicates that the fractal analysis describes some structural properties of proteins and corroborates the explanation about multifractality in the energy hypersurface.

DOI: 10.1103/PhysRevE.71.012901 PACS number(s): 87.14.Ee, 05.45.Df, 87.15.Aa, 02.50.−r

It is well known that a molecular system has a great number of conformation minima in the energy hypersurface [1–5] that increases with the number of degrees of freedom in the molecular system. In this context, the biological activity depends on the spatial conformation taken by the macromolecules in the physiological medium. The action of hormones and drugs is also dependent on the molecular three-dimensional structure of the target molecules. However, in analyzing the self-similarity [6] of the protein chains, we notice that the fractal dimension is independent of the number of degrees of freedom, and biological activity, among others characteristics of the proteins. Therefore, the fractality is an intrinsic and universal characteristic of protein chains.

Protein folding is driven by hydrophobic forces [7]. Lattman and Rose [8] analyzed globular proteins and concluded that the native fold determines the packing but packing does not determine the native fold. This view is corroborated by the widespread occurrence of protein families whose members assume the same fold without having a sequence similarity. Evidently, there is a large number of ways in which the internal residues can pack together efficiently.

As a consequence of steric constraints in compact polymers, helical and sheets structures appear [9]. Exhaustive simulations of the conformations indicate that the proportion of helices and sheets increases dramatically with the number of intrachain contacts [7]. In this way, the folding funnel theory [10] describes the thermodynamics and kinetic behavior of the transformation of unfolded molecule to the native state, and increases the number of native contacts. The folding funnel theory shows that any polypeptide chain explores the folding routes toward the native structure through intermediates consisting of population of partially folded species whose number decreases as the protein navigates down to the minimum of the energy landscape [10].

There is considerable discussion in the literature regarding the "old" and "new" views of protein folding [11–19]. The old view assumes that a small number of well-defined folding pathways exists and that folding is a hierarchical assembly process; e.g., the random coil first forms secondary structures, which are then organized into the native tertiary structure. In the new view, structurally less well-defined ensembles progress to the native state along multiple pathways. Zhou and Karplus [17,18] show that the specific behavior for small α-helical proteins depends on the optimization of the system. Thus, for a highly optimized model, the old view provides a satisfactory description. Folding proceeds through a small number of structurally well-defined pathways that begin with helix formation; this is followed by the formation of two-helix microdomains and docking of the third helix on two-helix microdomains to form the native structure. Less optimized models have a collapse to a disordered globulelike intermediate and do not have a well-defined folding pattern. There are many different ways for the collapsed globule to reach the native state, in accordance with the new view of protein folding. It thus appears that even for small α-helical proteins a wide range of mechanisms that encompasses both the old and new views are possible.

In this paper we are mainly interested in investigating the self-similarity present in 5526 different protein chains deposited in the Brookhaven Protein Data Bank. Our strategy is to measure the average radius of each chain as a function of the mass of each protein chain. This intrinsic characteristic of the protein structures must be responsible for the explanation of several aspects of those molecules, like the high compactness of those molecules, that has been discussed in the last decade in several different contexts [9,10,15,20–23]. This has finally been measured, as this compactness is given by the relation between the average radius and the mass of the protein chains. We describe this compactness via the mass-size exponent (fractal dimension) [24], i.e., via the relation between the average radius and the mass of protein chains, as the following:

$$M_i^j = k_i \langle R_i^j \rangle^{\delta}, \tag{1}$$

where M_i^j is the mass of the jth protein chain belonging to the ith protein and $\langle R_i^j \rangle$ is the corresponding average radius, i.e., the average distance from the geometric center for all coordinates.

The compact folded structure presents an elegant fractal behavior that determines how compact the folded protein is, as shown in Fig. 1. This figure depicts the behavior of the average radius as a function of the mass of 5526 analyzed protein chains. From this figure, the fractal dimension of the protein chains is $\delta=2.47\pm0.03$. This fractal dimension sup-

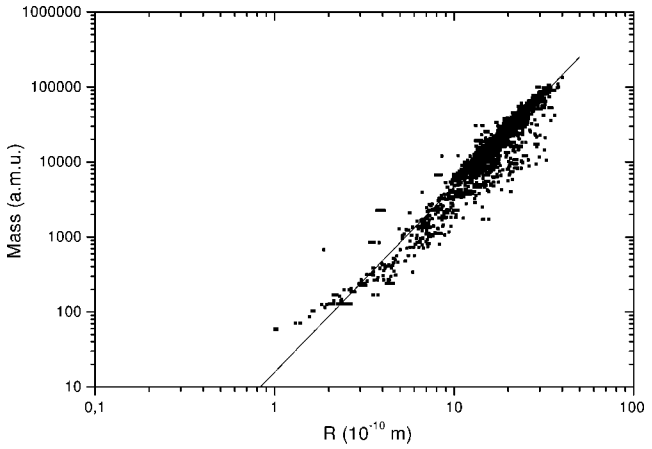

FIG. 1. The behavior of the average radius as a function of the mass of 5526 proteins chains. Each chain obeys the fractal dimension $\delta=2.47\pm0.03$, with the correlation coefficient of Pearson $R=0.91$.

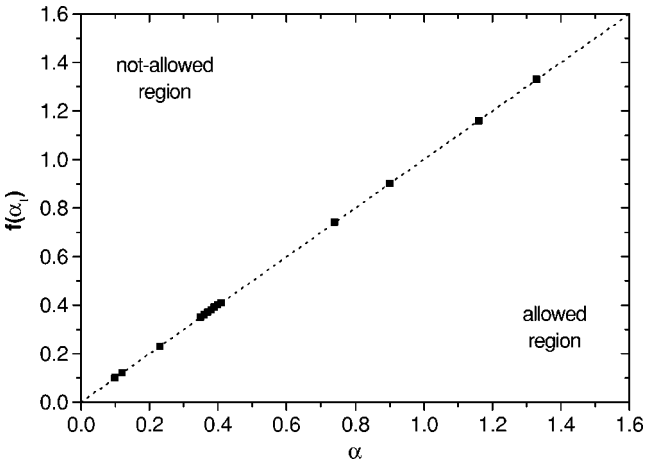

FIG. 2. The shape of the information dimension for some systems [4]. Above the tangent line defined by $f(\alpha_1)/\alpha_1$ no values of $f(\alpha)$ are possible.

plies us with the measure of how compact the proteins are. Using Voronoi tesselation [25], it is possible to show that the volume of the molecular surface in the function of the radius obeys a power law with $\delta=2.47$, which is exactly the same as we found. It is interesting to note that this value of the fractal dimension is smaller than the value obtained by normal objects in the normal ordinary Euclidean space.

Therefore, the atoms that compose the protein are distributed in this fractal object that represents the volume of the macromolecule. In this context, the interactions between these atoms carry over into a multifractality in the energy hypersurface. We recall that the main variables in the multifractal formalism are obtained from the relations

$$N_q(\varepsilon) \sim \varepsilon^{-f(\alpha_q)} \tag{2}$$

and

$$P_q \sim \varepsilon^{\alpha_q}, \tag{3}$$

where $f(\alpha_q)$ is the fractal dimension of the subset and α_q measures the intensity.

Recently, the multifractality of the energy hypersurface of peptides and of the proteins was measured [4]. It was shown that the proteins adopt conformations in the energy hypersurface only in allowed regions from the $f(\alpha)$ spectrum. Therefore, the energy hypersurface determines allowed (and not-allowed) regions that depend on the protein size, i.e., the number of degrees of freedom in the molecular system, as shown in Fig. 2. From this figure we observe that the interactions between atoms allow the proteins to map the space according to their size, i.e., as the hypersurface increases with the number of degrees of freedom in the system, these polypeptides present a poor capability to populate this hypersurface [4].

We recall that the energy hypersurface is a multifractal object not only due to the different normal modes of vibration existent in the protein structure, but also due to the support being a fractal object. In several scales the interactions among the atoms that compose the protein transform the energy hypersurface into a multifractal object.

Feder and co-workers [26,27] studied the aggregation of immunoglobulin proteins of IgG type using light scattering. They concluded that the cluster fractal dimension is $\delta=2.53\pm0.3$ for the IgG aggregates. The main difference from our analysis is that they did not calculate the interface volume; therefore, the fractal dimension was overestimated, in relation to the mass-size fractal dimension of the proteins. From the Flory exponent, Creamer *et al.* [28] showed the compactness in protein segments (for $26\leqslant N\leqslant 40$) without measuring the compactness of the proteins. On the other hand, fractal behavior is related to the anomalous temperature dependence of the Raman spin-lattice relaxation rates [4,29–31], because an ionic strength solution changes the temperature dependence.

Hence the compactness of the protein chains turns the geometry of these molecules fractal. It is interesting to point out that independent of the origin of the molecule, its family, or the organism where it was expressed, each chain of a protein obeys a power law with the characteristic exponent $\delta=2.47\pm0.03$. In this fractal geometry the atoms that compose the protein interact, making the energy hypersurface a multifractal object.

Finally, we would like to comment that both the "old" and "new" views of protein folding are contemplated in this multifractal structure of the energy hypersurface. Experiments are crucial for determining the best description of the folding mechanism for specific proteins.

[1] J. M. Yon, Cell. Mol. Life Sci. **53**, 557 (1997).
[2] M. A. Moret, P. G. Pascutti, P. M. Bisch, and K. C. Mundim, J. Comput. Chem. **19**, 647 (1998).
[3] D. J. Wales and H. A. Scheraga, Science **285**, 1368 (1999).
[4] M. A. Moret, P. G. Pascutti, K. C. Mundim, P. M. Bisch, and E. Nogueira, Jr., Phys. Rev. E **63**, 020901(R) (2001).
[5] M. A. Moret, P. M. Bisch, K. C. Mundim, and P. G. Pascutti, Biophys. J. **82**, 1123 (2002).
[6] T. Vicsek, *Fractal Growth Phenomena*, 2nd ed. (World Scientific, Singapore, 1992).
[7] D. Voet and J. Voet, *Biochemistry*, 2nd ed. (John Wiley & Sons, Inc., New York, 1995).
[8] E. E. Lattman and G. D. Rose, Proc. Natl. Acad. Sci. U.S.A. **90**, 439 (1993).
[9] K. A. Dill, Biochemistry **29**, 7133 (1990).
[10] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[11] C. Anfinsen, Science **181**, 23 (1973).
[12] R. L. Baldwin, J. Biomol. NMR **5**, 103 (1995).
[13] P. G. Wolynes, Z. Luthey-Schulten, and J. N. Onuchic, Chem. Biol. **3**, 425 (1996).
[14] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[15] A. Y. Pande, V. S. Grosberg, T. Tanaka, and D. S. Rokhsar, Curr. Opin. Struct. Biol. **8**, 68 (1998).
[16] M. Karplus, Folding Des. **2**, S68 (1997).
[17] Y. Zhou and M. Karplus, J. Mol. Biol. **293**, 917 (1999).
[18] Y. Zhou and M. Karplus, Nature (London) **401**, 400 (1999).
[19] G. A. Artega, C. T. Reimann, and O. Tapia, Chem. Phys. Lett. **350**, 277 (2001).
[20] K. A. Dill, Curr. Opin. Struct. Biol. **3**, 99 (1993).
[21] A. Sali, E. I. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).
[22] E. I. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1994).
[23] J. E. Shea and C. L. Brooks, Annu. Rev. Phys. Chem. **52**, 499 (2001).
[24] M. A. F. Gomes, J. Phys. A **20**, L283 (1987).
[25] J. Liang and K. Dill, Biophys. J. **81**, 751 (2001).
[26] J. Feder, T. Jossang, and E. Resenqvist, Phys. Rev. Lett. **53**, 1403 (1984).
[27] T. Jossang, J. Feder, and E. Resenqvist, J. Chem. Phys. **120**, 1 (1984).
[28] T. P. Creamer, R. Srinivasan, and G. D. Rose, Biochemistry **34**, 16245 (1995).
[29] J. S. Helman, A. Coniglio, and C. Tsallis, Phys. Rev. Lett. **53**, 1195 (1984).
[30] J. T. Colvin and H. J. Stapleton, J. Chem. Phys. **82**, 4699 (1985).
[31] G. C. Wagner, J. T. Colvin, J. P. Allen, and H. J. Stapleton, J. Am. Chem. Soc. **107**, 5589 (1985).