Self-similarity and protein compactness

M. A. Moret,[1,2,*] M. C. Santana,[2] G. F. Zebende,[1,2] and P. G. Pascutti[3]
[1]*Programa de Modelagem Computacional, SENAI CIMATEC, Salvador, BA, Brazil*
[2]*Departamento de Física, UEFS, 44031-460 Feira de Santana, BA, Brazil*
[3]*Instituto de Biofísica, UFRJ, 21949-900 Rio de Janeiro, RJ, Brazil*
(Received 13 March 2009; revised manuscript received 14 August 2009; published 6 October 2009)

The hydrophobic effect is the major factor that drives a protein toward collapse and folding. As a consequence of the folding process a hydrophobic core is shielded by the solvent-accessible surface area of the protein. We analyze the solvent-accessible surface area of 1825 nonhomolog protein chains deposited in the Brookhaven Protein Data Bank. This solvent-accessible surface area presents an intrinsic self-similarity behavior. The comparison between the accessible surface area as function of the number of amino acids and the accessible surface area as function of gyration radius supplies a measure of the scaling exponent close to the one observed by volume as function of radius of gyration or by mass-size exponent. The present finding indicates that the fractal analysis describes the protein compactness as an object packing between random spheres in percolation threshold and crumpled wires.

DOI: 10.1103/PhysRevE.80.041908 PACS number(s): 87.15.−v, 89.75.Da, 05.45.Df

The fractal behavior assists for the explanation of several structural and dynamics aspects of biological molecules. Thus, self-similarity was uncovered in, e.g., clusters dimension of proteins [1], anomalous temperature dependence of the Raman spin-lattice relaxation rates [2,3], relation between the fractal dimension and number of hydrogen bridges [4], multifractality in the energy hypersurface of the proteins and a possible alternative explanation of the Levinthal paradox [5], packing of small protein fragments [6], loss of the accessible surface area of amino acids and hydrophobicity scale [7], and degree of compactness of the proteins [8,9], among others. Furthermore, the fractal methods corroborate to identify different states of the same system according to its different scaling behaviors; e.g., the fractal dimension is different for structures with (without) hydrogen bonds [4,5] or different long-range correlations in a liquid-vapor-phase transition of the solvent [10]. Then, the correct interpretation of the scaling results obtained by the fractal analysis is crucial for understanding the intrinsic geometry of the systems under study [11]. In this point of view the mass-size relation is one of the most important measures, because if self-similarity appears, we have scale invariance like

$$\text{mass} \propto R^D, \tag{1}$$

where D is a fractionary exponent interpreted as the fractal mass dimension of the system.

Protein folding is driven by hydrophobic forces, apolar amino acids associate to form a hydrophobic core, packing the protein in the folded state. Thus, during folding, hydrophobic groups are expelled spontaneously from water, thereby engendering a sequestered, solvent-shielded core. In this sense, there are a myriad number of ways in which the internal residues can pack together efficiently. Packing in proteins was first studied by using a Voronoi analysis for proteins in a space-filling model [12], where each atom is taken to be a sphere with a fixed radius given by the van der Waals one. Several empirical rules of the protein folding can be deduced from information obtained from studies of protein *in vitro* [13]. To study protein folding is necessary taking into account that a molecular system has a great number of minima in the energy hypersurface (molecular conformations) [5,13–23], which increases with the number of the degrees of freedom in the molecular system.

Packing in proteins is an intriguing and a current research field. The high packing density of residues in proteins ought to be manifested in some order and the packing order has not been thoroughly characterized. Thus, the packing regularity in proteins determines the internal organization of proteins. Hence, protein packing can have a dominant effect on functional dynamics and it can aid in the design, simulation, and evaluation of structures. Many aspects of protein structures relate to internal packing considerations including design, simulation, and evaluation of structures, as well as sequence conservation and even folding nucleation. Usually, in the literature there has been a strong focus on the conformations of protein backbones. However, because of competition between local and long-range interactions, it is not clear where the greatest regularity should appear. The literature about protein compactness has a lot of proposed models to explain the packing density. Richards [12] proposed that the packing inside proteins is compared to the packing of spheres in crystalline solids. Finney [24] proposed that proteins behave as liquid random spheres and Honig [25] stressed on the structural plasticity of proteins. Based on observations of high packing densities and low compressibility, protein cores are often considered to be more like solids than liquids. Jigsaw puzzle (a perfect packing), random spheres, among other approaches, are proposed models to explain the protein packing [25]. In contrast, proteins are tolerant to mutations suggesting that proteins can be regarded as plastic or liquidlike [25]. From the mass-size analysis [26], we observed that the packing density of these folded structures behaves like spheres in their percolation threshold [8,9,27]. Thus, when we analyze the different members of a protein family, we observe that deletions and/or insertions do not change the protein folding;

*mamoret@gmail.com

these mutations maintain practically the same folded structure. On the other hand, when we change an alanine to a proline in α helices, this mutation must be the ruin of the α-helix structure. Recently, protein compactness has been studied by the most varied approaches as, for instance, by a method based on Delaunay tesselation [27], a coarse-grained scale [28], mass-size exponent [8,9], and self-organized criticality [29–31], among other approaches. We notice that Ref. [28] proposed that residue clusters from PDB structures, each comprised of a central residue and all neighbors located within the first coordination shell, have been rigidly reoriented and superimposed in a self-consistent optimization. About 2/3 of residues are found to follow approximately the relative orientation preferences of face-centered-cubic (fcc) packing, when examined on a coarse-grained scale (one site per residue), while the remaining 1/3 occupy random positions. The observed regularity, which becomes more pronounced after optimal superimposition of core residues, appears to be the result of uniform sampling of the coordination space around each residue on a coarse-grained scale with hydrophobic clustering and volume exclusion to achieve packing densities close to that of the universal closest packing of identical spheres. The remaining 1/3 refer to residues that are more loosely or randomly packed being exposed in the solvent-accessible surface area [28]. However, the solvent-accessible surface area of proteins is a variable of great importance and it is not well known, what turns it a defined thermodynamic variable in an imprecise way [32]. Thus, this study can aid in the understanding of the relation between the area and the volume of proteins, which is not yet defined.

In this paper we analyze the solvent-accessible surface area of 1825 nonhomolog protein chains deposited in the Brookhaven Protein Data Bank (PDB). Thus, we measure the solvent-accessible surface area of a protein molecule by rolling a probe sphere (with a size close to water molecule) over a molecule [33]. The surface of a macromolecule can be defined to be the part of the molecule that is accessible to solvent. The probe sphere is represented by a sphere of radius 1.4 Å, i.e., close to the solvent molecule (water) radius. Figure 1 depicts the self-similarity of the solvent-accessible surface area as function of number of amino acids. We observe that the solvent-accessible surface area (A) obeys a power law when the number of amino acids (N) increases in the protein chains (scale invariance), i.e.,

$$A \propto N^{\gamma} \tag{2}$$

with exponent $\gamma=0.87$. We emphasize that the correlation coefficient of Pearson is $R=0.992$, which is very close to 1, if we take into account that we analyzed 1825 protein chains. This fractal behavior is similar to one obtained by the radius of gyration R_g as function of the number of amino acids N [34],

$$R_g \propto N^{\nu}. \tag{3}$$

As a result of a fractal analysis, Fig. 2 shows the self-similar behavior of the solvent-accessible surface area as function of the radius of gyration of protein chains R_g. Again we observe that the solvent-accessible surface area obeys a power law when the R_g increases, i.e.,

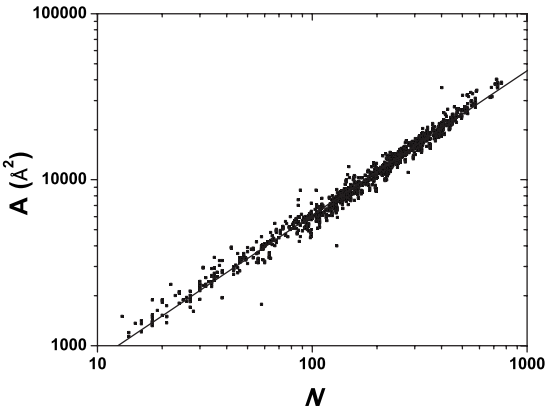

FIG. 1. The solvent-accessible surface area A as function of number of amino acids N. The continuous line represents the best fit, with $\gamma=0.87$; standard deviation is 0.04 and standard error is 0.003.

$$A \propto R_g^{\lambda} \tag{4}$$

with $\lambda=2.26$ (fractal area dimension). From the fractal behavior we note that this surface area has reentrances because the fractal dimension is greater than 2 carrying it as a crumpling object [35,36].

These Eqs. (2)–(4) or volume × radius of gyration [27] can be used to compare the relations of the surface area showed in Figs. 1 and 2 with the fractal behavior of the protein volume. Then, after some simple steps we have

$$1/\nu = \frac{\lambda}{\gamma} = 2.60. \tag{5}$$

The value of the exponent $(1/\nu)$ can be compared with the relation between the average radius $\langle R_i^j \rangle$ and the mass M_i^j of the jth protein chain belonging to the ith protein as proposed by [8,9]. In this case $M_i^j \propto \langle R_i^j \rangle^{\delta}$ with $\delta=2.47$. The difference

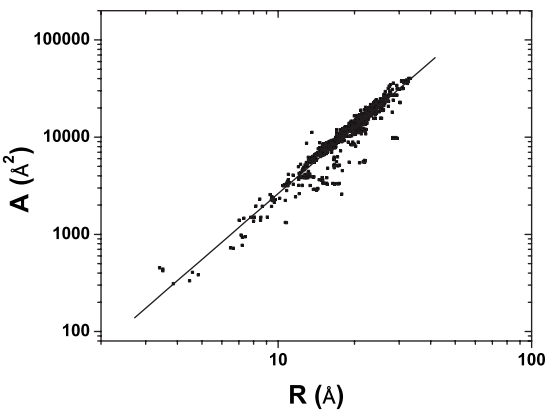

FIG. 2. The solvent-accessible solvent area A as function of radius of gyration R_g. The continuous line represents the best fit, with $\lambda=2.26$; standard deviation is 0.09, standard error is 0.007, and correlation coefficient $R=0.949$.

between the volume or mass-size dimension δ and ν^{-1} arises from a simple mean-field model [35,36]. In this model the total energy E associated with the folded structure of a protein is divided in the most coarse-grained way possible in two contributions:

(1) A term coming from an elastic energy, $E_{el}=\frac{1}{2}kR^n$, where $k>0$ is an effective elastic constant, R characterizes the radius of gyration of the protein configuration, and n is a scaling exponent. We notice that $n=2$ recovers the usual parabolic elastic energy.

(2) The second term is a self-avoidance energy E_{sa}, which is assumed to be a two-body interaction and consequently is proportional to the square of the average density of mass. This self-avoidance energy term thus reads after integration in the volume, $E_{sa}\propto\rho^2\times V\propto M^2\times R^{-3}$ ($\rho\equiv\frac{M}{V}$ and $V\propto R^3$). The protein mass M scales obviously with the number of amino acids ($M\sim N$). The self-avoidance energy term corresponds in turn to an outward nonlinear force:

$$F_{sa}=-\frac{\partial E_{sa}}{\partial R}\propto\frac{M^2}{R^4}. \tag{6}$$

In this model elastic term is associated with entropic and segregation forces, as well as with hydrophobic interactions. Hydrophobic force drives a protein toward collapse and folding; then this force tends to privilege a collapsed configuration of the protein, while the self-avoidance term, which has its microscopic origin in Pauli's exclusion principle, tends to privilege extended configurations of the protein. In our model, F_{sa} simulates the steric repulsion forces, which is one of the dominant forces in protein folding [14,15].

By minimizing the total energy $E=E_{el}+E_{sa}$ with respect to R, i.e., ($\frac{\partial E}{\partial R}=0$), we obtain the mass-size dependence at the equilibrium as the power law ($M\propto R^D$), with $D=(3+n)/2$. The case of mass-size exponent uses $n=2$ which corresponds to a linear elastic (Hooke) force, while all other cases are associated with nonlinear forces. Thus a continuous spectrum of fractal dimension, say in the experimental interval, will be allowed for n varying in the satisfying $0<n<3$.

From our analysis we observe that the two protein fractal dimensions, one obtained by mass as function of radius ($\delta=2.47$) and the other obtained by number of amino acids as function of radius ($1/\nu=2.60$), are completely correlated. The main difference between these two exponents is the elastic-force contribution to the inside of proteins $n\cong 2$ and to the solvent-accessible surface area $n\cong 2.2$. Then, here we show that the folded structures of proteins behave on average as spheres in percolation threshold inside the protein and as irreversibly crushed wires in the solvent-accessible surface area. Furthermore, a protein not necessarily behaves as a traditional liquid or a solid crystalline object.

In summary, we observe that proteins have two different self-similar behaviors, the self-similarity obtained by the mass-size exponent [8], by the volume as function of radius [27] and by the solvent-accessible surface area as a function of the number of amino acids. In relation to the solvent-accessible surface area we obtain two different behaviors, i.e., the solvent-accessible surface area as a function of the number of amino acids (Fig. 1) and the fractal dimension of the solvent-accessible surface area (Fig. 2). These power laws supply us a measure of the roughness of the solvent-accessible surface area and corroborate the explanation about the fractal behavior of the protein volume. Therefore, the scaling exponent $\delta=2.47$ can be viewed as amino acids filling out a protein in the same way as random spheres in percolation threshold [9,37–39]. In this sense, solvent-accessible surface area is hydrophilic region that is in contact with physiologic solvent medium [40]. Comparing the fractal behavior of the volume with the accessible surface area we note that accessible surface area has a larger roughness characterized by the power law ($\lambda=2.26$) and the protein volume presents voids ($\delta=2.47$).

Therefore, amino acids over the solvent-accessible surface area behave on average as irreversibly crushed wires. From the elastic scaling exponent we observe that the solvent-accessible surface area restoring force is greater than the core restoring one. Then, proteins present an elegant and simple behavior. Thus, the inside protein packing behaves like random spheres in the percolation threshold [8,9]. On the other hand, the solvent-accessible surface area packing behavior is equivalent to the irreversibly crushed wires.

Finally, we recall that in general proteins are globular structures. These macromolecules can be composed for one protein chain or more chains. In this paper we analyzed proteins with one monomer and protein chains of the other cases (dimer, trimer, etc.) only. Nevertheless, this result can be generalized to proteins independently of number of monomers.

This work received financial support from CNPq (Brazilian federal grant agency) and from FAPESB (Bahia state grant agency).

[1] J. Feder, T. Jossang, and E. Resenqvist, Phys. Rev. Lett. **53**, 1403 (1984).
[2] J. T. Colvin and H. J. Stapleton, J. Chem. Phys. **82**, 4699 (1985).
[3] G. C. Wagner, J. T. Colvin, J. P. Allen, and H. J. Stapleton, J. Am. Chem. Soc. **107**, 5589 (1985).
[4] J. S. Helman, A. Coniglio, and C. Tsallis, Phys. Rev. Lett. **53**, 1195 (1984).
[5] M. A. Moret, P. G. Pascutti, K. C. Mundim, P. M. Bisch, and E. Nogueira, Phys. Rev. E **63**, 020901(R) (2001).
[6] T. P. Creamer, R. Srinivasan, and G. D. Rose, Biochemistry **34**, 16245 (1995).
[7] M. A. Moret and G. F. Zebende, Phys. Rev. E **75**, 011920 (2007).
[8] M. A. Moret, J. G. V. Miranda, E. Nogueira, M. C. Santana, and G. F. Zebende, Phys. Rev. E **71**, 012901 (2005).
[9] M. A. Moret, M. C. Santana, E. Nogueira, and G. F. Zebende, Physica A **361**, 250 (2006).

[10] G. F. Zebende, M. V. S. da Silva, A. C. P. Rosa, A. S. Alves, J. C. O. de Jesus, and M. A. Moret, Physica A **342**, 322 (2004).
[11] J. Feder, *Fractals* (Plenum, New York, 1988).
[12] F. Richards, J. Mol. Biol. **82**, 1 (1974).
[13] J. M. Yon, Cell. Mol. Life Sci. **53**, 557 (1997).
[14] K. A. Dill, Curr. Opin. Struct. Biol. **3**, 99 (1993).
[15] N. Byrne and C. A. Angell, J. Mol. Biol. **378**, 707 (2008).
[16] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[17] M. A. Moret, P. G. Pascutti, P. M. Bisch, and K. C. Mundim, J. Comput. Chem. **19**, 647 (1998).
[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. **82**, 1123 (2002).
[20] M. A. Moret, P. M. Bisch, E. Nogueira, and P. G. Pascutti, Physica A **353**, 353 (2005).
[21] F. P. Agostine, D. D. O. Soares-Pinto, M. A. Moret, C. Osthoff, and P. G. Pascutti, J. Comput. Chem. **27**, 1142 (2006).
[22] M. A. Moret, P. M. Bisch, and F. M. C. Vieira, Phys. Rev. E **57**, R2535 (1998).
[23] M. A. Moret, P. G. Pascutti, P. M. Bisch, M. S. P. Mundim, and K. C. Mundim, Physica A **363**, 260 (2006).
[24] J. Finney, J. Mol. Biol. **96**, 721 (1975).
[25] B. Honig, J. Mol. Biol. **293**, 283 (1999).
[26] M. A. F. Gomes, J. Phys. A **20**, L283 (1987).
[27] J. Liang and K. Dill, Biophys. J. **81**, 751 (2001).
[28] Z. Bagci, R. L. Jernigan, and I. Bahar, J. Chem. Phys. **116**, 2269 (2002).
[29] A. E. Kister and J. C. Phillips, Proc. Natl. Acad. Sci. U.S.A. **105**, 9233 (2008).
[30] J. C. Phillips, Proc. Natl. Acad. Sci. U.S.A. **106**, 3107 (2009).
[31] J. C. Phillips, Proc. Natl. Acad. Sci. U.S.A. **106**, 3113 (2009).
[32] D. Voet and J. G. Voet, *Biochemistry* (John Wiley and Sons, New York, 1994).
[33] B. K. Lee and F. M. Richards, J. Mol. Biol. **55**, 379 (1971).
[34] P. H. Figuerêdo, M. A. Moret, E. Nogueira, and S. Coutinho, Physica A **387**, 2019 (2008).
[35] M. A. F. Gomes, V. P. Brito, and M. S. Araújo, J. Braz. Chem. Soc. **19**, 293 (2008).
[36] M. A. F. Gomes, V. P. Brito, A. S. O. Coelho, and C. C. Donato, J. Phys. D **41**, 235408 (2008).
[37] D. Stauffer, Phys. Rep. **54**, 1 (1979).
[38] D. Stauffer, *Introduction to Percolation Theory* (Taylor & Francis, London, 1985).
[39] T. Vicsek, *Fractal Growth Phenomena*, 2nd ed. (World Scientific, Singapore, 1992).
[40] J. Zhang, R. Chen, C. Tang, and J. Liang, J. Chem. Phys. **118**, 6102 (2003).