COMMENTS

Comments are short papers which criticize or correct papers of other authors previously published in the **Physical Review**. *Each Comment should state clearly to which paper it refers and must be accompanied by a brief abstract. The same publication schedule as for regular articles is followed, and page proofs are sent to authors.*

Comment on "Fractal study of tertiary structure of proteins"

Yi Xiao

Department of Physics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, People's Republic of China

(Received 20 January 1993)

It is shown that the mean fractal dimension of the global structure of 90 proteins is greater than that of an unrestricted random walk. Some modifications on the method of calculating the length of the backbone of a protein chain and a possible connection of the mean fractal dimensions of the local structures with the secondary structures of proteins are discussed.

PACS number(s): 87.10.+e, 87.15.By, 05.45.+b

In a recent paper [1], Wang, Shi, and Huang calculated the fractal dimension of the tertiary structures of 90 proteins according to their structural classes. The authors found that the mean fractal dimensions (D_1 and D_2) of the local and global conformations of 90 proteins were 1.38 and 1.65, respectively. And they concluded that the mean fractal dimension D_2 of the global conformation was very close to the theoretical value $\frac{5}{3}$ of the fractal dimension of a self-avoiding random walk in three-dimensional space and was in agreement with the data calculated by Allen *et al.* [2]. In this Comment, we want to point out that the value of the fractal dimension of a self-avoiding random walk in three-dimensional space is 1.40 according to the definition of fractal dimension used by Wang, Shi, and Huang and so it is not D_2 but D_1 which is very close to this value. We shall also consider some further modifications on the method of calculating the length of the backbone of a protein chain and discuss whether the fractal dimension D_1 is determined by the secondary structures of proteins.

(i) Definition of fractal dimension of a protein chain. In the literature, there exist two kind of definitions for the fractal dimension of a protein chain. The one, used by Stapleton and co-workers [2–4], is given by

$$M \sim r(M)^d , \tag{1}$$

where $r(M)$ is the mean separation between the ith and the $(i+M)$th element of the chain. According to Eq. (1), the fractal dimensions of an unrestricted random walk and a self-avoiding random walk in three-dimensional space are 2 and $\frac{5}{3}$, respectively.

The other definition, used by Isogai and Itoh [5] and Wang, Shi, and Huang [1], is given by

$$L(M) \sim M^{1-d} , \tag{2}$$

where $L(M)$ is the length of the backbone of a protein chain consisting of N residues measured with a scale of $M(\leq N)$ residues. It has been shown by Isogai and Itoh [5] that the fractal dimension of an unrestricted random walk determined by Eq. (2) is 1.5 instead of 2. Using the same method as Isogai and Itoh, it is easy to find that the fractal dimension for a self-avoiding random walk in three-dimensional space is 1.40.

In the paper by Wang, Shi, and Huang [1], the fractal dimensions of the protein chains were calculated by using Eq. (2). Thus the mean fractal dimension $D_2(=1.65)$ obtained by them is clearly greater than the value of the fractal dimension of a self-avoiding random walk in three-dimensional space and even greater than that of an unrestricted random walk. Instead, the mean fractal dimension $D_1(=1.38)$ obtained by them is very close to the value (1.40) of the fractal dimension of a self-avoiding random walk in three-dimensional space. Since D_1 describes the folding of the local conformation of a protein chain, this implies that the local conformation is determined by the excluded volume effect. Of course, there is nothing new about this conclusion because the spatial conformation is usually modeled by a three-dimensional self-avoiding random walk. On the other hand, it is surprising that D_2 is greater than 1.5, the fractal dimension of an unrestricted random walk. This implies that the global conformation of protein chains are more compact than that of an unrestricted random-walk chain. Since there are not attractive and repulsive forces between the elements of an unrestricted random walk, this demonstrates that the global conformations of protein chains are determined by some "active" and "passive" attractive forces. For instance, the former may be the van der Walls forces between the monomers and the latter may be the hydrophobic interaction which leads to the formation of a core consisting of hydrophobic residues. Since D_1 may be determined by the secondary structures of proteins (see the following), this shows that the folding of the secondary and tertiary structures are determined by different rules. This seems to be in disagreement with the suggestion made by Chan and Dill [6] that the secondary structures (helices and sheets) may be determined

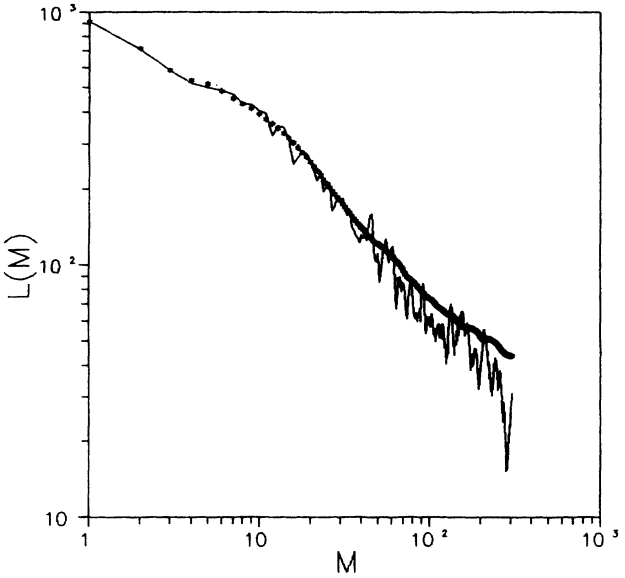

FIG. 1. The fractal diagram of carboxypeptidase *A*. The solid line is calculated by the method used by Wang, Shi, and Huang and the starred by our modified method.

mainly by the hydrophobic interaction. Thus the fractal dimension can be used to describe the general characters of the forces maintaining the spatial conformations of protein molecules and give some guidelines for the protein folding problem.

(ii) The length of the backbone of a protein chain. In the fractal diagram shown by both Isogai and Itoh [5] and Wang, Shi, and Huang [1], the part of the diagram with M greater than 10 or 15 was jagged. This can be removed by (a) taking $L(M)$ as the mean value of the lengths of the backbone, measured by starting from every C^{α} atoms of the chain instead of only the C^{α} atom of the N-terminal residue as Isogai and Itoh and Wang, Shi, and Huang have done. This is because the measured length of the backbone is somewhat different if taking a different C^{α} atom as the starting point. (b) Using the actual end-to-end distance as the length of the segment consisting of the remaining residues ($<M$) which are not enough to span at the final step of drawing the zigzag line. It is clearly a crude approximation to estimate the length of the segment by using the mean end-to-end distance of the segments consisting of M residues, especially for larger M. By taking these two points into account, we can get better results. Figure 1 is the fractal diagram of carboxypeptidase *A*. It can be seen that the part of the fractal diagram with M greater than 10 calculated by us is nearly a straight line.

(iii) Fractal dimension of the secondary structure of protein. In order to determine whether D_1 is related to the secondary structure of protein, it is interesting to calculate the mean fractal dimension D_1 of long ideal secondary structures of proteins (e.g., Pauling and Corey's α helix and β sheet, and so on). The mean fractal dimension D_2 for the global structure approaches one, so it is not of interest to us. Of course, the fractal dimension here is only a quantity describing the general feature of the secondary structures and does not necessarily imply

TABLE I. The mean fractal dimensions for the secondary structures of protein.

Secondary structure	Fractal dimension D_1
α helix	1.44±0.13
3_{10} helix	1.30±0.13
π helix	1.57±0.13
Collagen helix	1.14±0.06
Parallel β sheet	1.09±0.06
Antiparallel β sheet	1.06±0.04
Twisted β sheet	1.07±0.05

that the local conformations are fractal since the values taken by M are only from 1 to 15. Fractal theory provides us an alternative way to examine irregular objects. So it is more reasonable to call D_1 and D_2 "fractal exponents."

Wang, Shi, and Huang have calculated D_1 for 90 proteins [1]. For comparison, we also use Eq. (2) to calculate D_1 for the secondary structure. In Table I, we present D_1's for seven secondary structures of protein. They are α helix, 3_{10} helix, π helix, collagen helix, parallel β sheet, antiparallel β sheet, and twisted β sheet. For ideal secondary structure, the distance between nth and $(n+m)$th residues in the chain is given by

$$r(m)=\{[2R\sin(\pi m/p)]^2+(am/p)^2\}^{1/2}, \quad (3)$$

where p is the number of residues each turn, a is the pitch of helix, R is the radius of helix. Their values for seven ideal secondary structures can be found in Ref. [7].

The mean value of D_1 for α, β, α-β, and small disulfide-rich or small metal-rich domains classes are 1.41, 1.33, 1.37, and 1.41, respectively [1]. In globular protein, the most occurring secondary structures are α helix, twisted β sheet, and reverse turn. For α helix and twisted β sheet, D_1 is 1.44±0.13 and 1.07±0.05, respectively. For protein of α class, D_1 is very close to that of α helix. For β class, however, D_1 is much smaller than that of pure twisted β sheet. This may be due to the fact that the secondary structure of protein of β class also contains a certain amount of reverse turn. If we assume that half of the secondary structure is reverse turn (of course this is overestimated), and fit $D_1(=1.33)$ of β class by the average of D_1 of twisted β sheet and reverse turn, we find that the approximate D_1 of reverse turn is 1.59. Using this value, we obtain the mean value of D_1 of α helix, twisted β sheet, and reverse turn which is 1.37. This is the same as that of the α-β class given by Wang, Shi, and Huang. This is reasonable since the protein of the α-β class contains all of these three secondary structures. For the *SD* class, it can be put into α class and so they have the same mean fractal dimension D_1 [7].

Finally, it is worth noting that the largest and smallest values of D_1 for α helix and β sheet, respectively, may be the origin of the apparent dimensionality of the specific heat measured for solid polypeptide chains at low-temperature [one-dimensional (1D) for α helical polyalanine, 2D for β sheet polyalanine] [8].

In conclusion, we have shown that the definition of

fractal dimension used by Isogai and Itoh or Wang, Shi, and Huang is different from that used by Stapleton and co-workers. We have also shown that it is not the mean value of D_2 but D_1 for 90 proteins which is very close to the fractal dimension of a self-avoiding random walk in three-dimensional space. In fact, D_2 is larger than the fractal dimension of an unrestricted random walk. Finally, from D_1's of the secondary structures, we find that the values of D_1's for four structural classes are probably determined by their secondary structures.

[1] C. X. Wang,Y. Y. Shi, and F. H. Huang, Phys. Rev. A **41**, 7043 (1990).
[2] J. P. Allen, J. T. Colvin, D. G. Stinso, C. P. Flynn, and H. J. Stapleton, J. Biophys. **38**, 299 (1982).
[3] J. T. Colvin and H. J. Stapleton, J. Chem. Phys. **82**, 4699 (1985).
[4] G. C. Wagner, J. T. Colvin, J. P. Allen, and H. J. Stapleton, J. Am. Chem. Soc. **107**, 5589 (1985).
[5] Y. Isogai and T. Itoh, J. Phys. Soc. Jpn. **53**, 2162 (1984).
[6] H. S. Chan and K. A. Dill, Phys. Today **2**, 24 (1993), and references therein.
[7] G. E. Schulz and R. H. Schirmer, *Principles of Protein Structure* (Springer-Verlag, New York, 1979).
[8] L. Finegold and J. L. Cude, Nature **238**, 38 (1972).