Fractal study of tertiary structure of proteins

Cun Xin Wang
Center for Fundamental Physics, University of Science and Technology of China, Hefei, Anhui, People's Republic of China

Yun Yu Shi
Department of Biology, University of Science and Technology of China, Hefei, Anhui, People's Republic of China

Fu Hua Huang
Department of Modern Chemistry, University of Science and Technology of China, Hefei, Anhui, People's Republic of China

(Received 12 July 1989; revised manuscript received 28 December 1989)

A refined method for calculating the fractal dimensions of proteins is introduced in this article. The fractal dimensions of 90 proteins covering four structural classes of proteins are established using this method. The relationship between the fractal dimension and tertiary structure of proteins is analyzed. The mean value of the fractal dimension D_2 for the global structure of proteins is 1.65, which is very close to the theoretical value $\frac{5}{3}$ associated with a self-avoiding random walk in three-dimensional Euclidean space.

Recently there has been a great deal of interest in studies of the fractal properties of macromolecular chains.[1-6] We have found that the results of Ref. 6 by Isagai and Itoh are evidently different from the experimental and simulation data of Stapleton *et al.*[2,4] Here we report some refinement in the method and the reasonable computational results.

According to the positions of the α carbons, we can draw a zigzag line using the method of Ref. 6. The backbone length $L(M)$ of a protein molecule consisting of N residues measured with a scale of M can be written by

$$L(M)=L_z(M)+nL_z(M)/(MK)\ , \tag{1}$$

where $L_z(M)$ is the length of the zigzag line, the second term is the correction term in which n and K are the number of the left unconnected residues and the folding number of the zigzag line, respectively. Since

TABLE I. Fractal dimensions of each protein for the α class.

Protein	Residue number	Fractal dimension D_1	D_2
Myohemerythrin	118	1.41±0.03	1.60±0.08
Hemerythrin (azido)	118	1.41±0.03	1.60±0.07
Hemerythrin (met)	113	1.41±0.02	1.59±0.06
Cytochrome B562	103	1.41±0.03	1.56±0.06
Hemoglobin (aquomet)	136	1.42±0.04	1.61±0.09
Hemoglobin (cyanomet)	136	1.42±0.03	1.61±0.08
Hemoglobin (erythrocruorin)	136	1.42±0.04	1.61±0.09
Hemoglobin (deoxy)	141	1.41±0.04	1.63±0.08
Hemoglobin *S* (deoxy)	141	1.41±0.04	1.63±0.09
Hemoglobin (carbonmonoxy)	141	1.42±0.03	1.62±0.07
Hemoglobin (sickle cell, deer)	141	1.43±0.04	1.61±0.08
Hemoglobin *A* (oxy)	141	1.42±0.04	1.63±0.08
Leghemoglobin (acetate, met)	153	1.41±0.03	1.62±0.09
Leghemoglobin (aquo, met)	153	1.41±0.03	1.62±0.08
Leghemoglobin (deoxy)	153	1.41±0.03	1.62±0.07
Calcium-binding parvalbumin *B*	108	1.44±0.01	1.66±0.07
Actinidin (domain 1)	147	1.36±0.01	1.59±0.12
Citrate synthase	437	1.39±0.02	1.67±0.05
Papain (domain 1)	111	1.37±0.02	1.56±0.05

$1 \leq M \leq N-1$, the $N-1$ different values of $L(M)$ could be obtained. Plotting $\log_{10}M$ and $\log_{10}L(M)$ on the abscissa and the ordinate, respectively, a fractal diagram can be drawn. Defining $H(M)$ as the slope of a fractal diagram obtained with linear regression, a local fractal dimension $D_0(M)$ is given by

$$D_0(M)=1-H(M) . \tag{2}$$

Although there are some self-similarities in the polypeptide chain of proteins, proteins lack rigorous self-similarity, hence the resulting value of fractal dimension depends somewhat on the choice of the scale M; the results are therefore averaged over the number of various choices of the scale M.[1,4] The mean fractal dimension $D(j)$ is defined as

$$D(j)=\frac{1}{j}\sum_{M=1}^{j} D_0(M) , \tag{3}$$

where j is the number of local fractal dimension.

To explore the relationship between the fractal dimension and tertiary structure of proteins, we have selected 90 proteins from the Protein Data Bank[7] and obtained the fractal diagrams of each protein by the above mentioned procedure. The 90 proteins cover four structural classes:[8] (1) antiparallel α domains (α), (2) antiparallel β domains (β), (3) parallel α-β domains (α-β), and (4) small disulfide-rich or small metal-rich domains (SD). Figures 1(a)–1(d) show the fractal diagrams of four proteins covering four structural classes. It is interesting to note from Fig. 1 that the part of the diagram within the range of $M \leq 15$ is smooth and the slope of the fractal curve does not change much, but the part of the curve within the range of $M > 15$ is jagged. These features are also observed in the other proteins included in this work. Since the fractal diagram within the range of smaller M reflects the local folding of the protein backbone and that within the range of larger M reflects the global folding of the

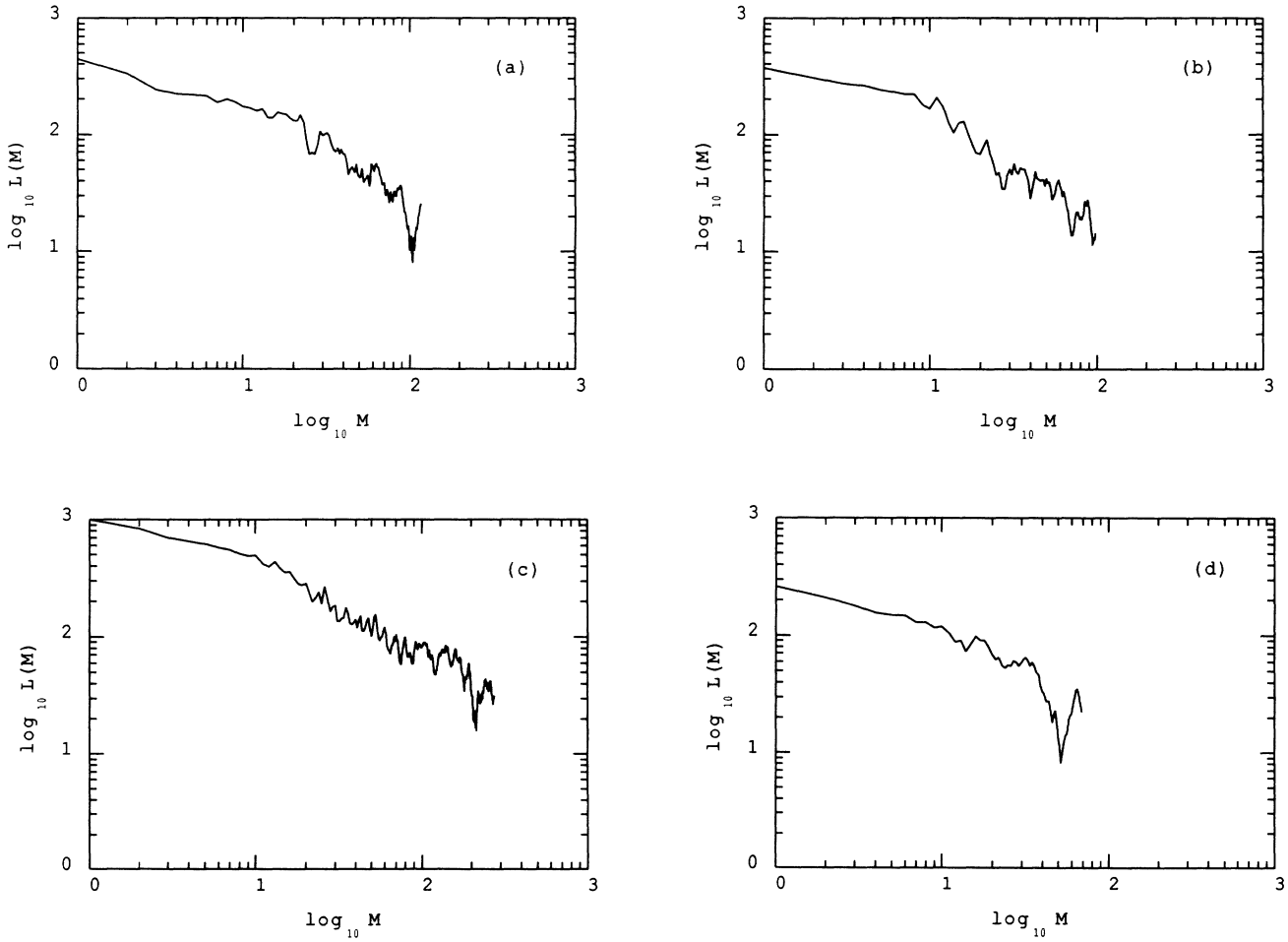

FIG. 1. The fractal diagrams of four proteins selected from 90 proteins: (a) myohemerithrin (α class), (b) plastocyanin (β class), (c) subtilisin BPN (α-β class), and (d) insulin (SD class).

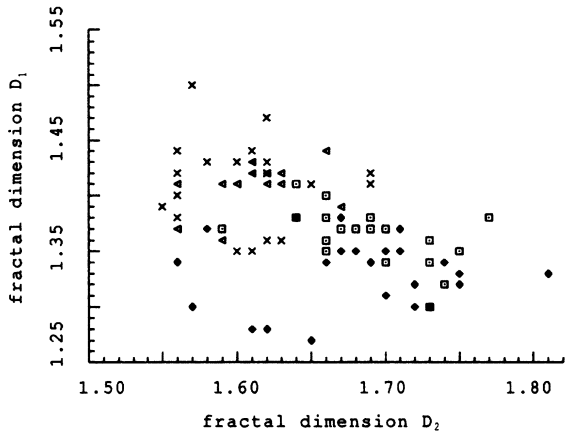

FIG. 2. The scatter diagram of the fractal dimension D_1 and D_2. Each point on the diagram represents a protein of the corresponding fractal dimension D_1 and D_2. The four structural classes are distinguished by the different marks: triangle for the α class, diamond for the β class, square for the α-β class, and cross for SD class.

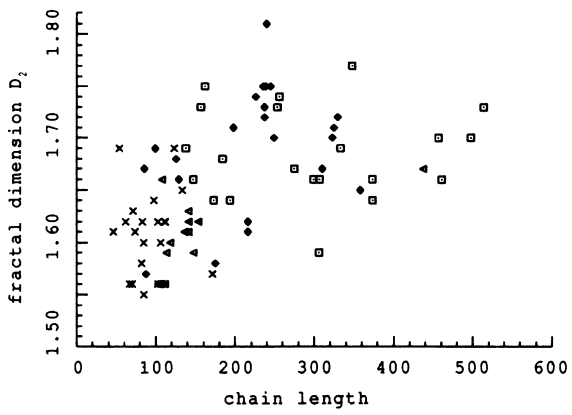

FIG. 3. Dependence of the fractal dimension D_2 on the chain length. The number of amino acid residues is taken as the chain length. The meanings of the marks are the same as those in Fig. 2.

protein backbone, the local and global foldings of proteins seem to be dictated by different rules. From Eq. (3), the mean fractal dimension within the range of $1 \leq M \leq 15$ $(j=15)$ and $1 \leq M \leq N-1$ $(j=N-1)$ are named fractal dimensions D_1 and D_2, which represent the mean fractal dimensions of the local and global structure of protein, respectively.

In Tables I–IV, the values of the mean fractal dimension and the standard deviations for 90 proteins are reported. The fractal dimensions D_2 for ferric-iron-containing heme and some nonheme proteins are in con-

TABLE II. Fractal dimensions of each protein for the β class.

Proteins	Residue number	Fractal dimension D_1	D_2
Plastocyanin	99	1.34±0.05	1.69±0.09
Azurin (oxidized)	129	1.34±0.03	1.66±0.07
Azurin	126	1.35±0.04	1.68±0.08
Chymotrypsin *A*	226	1.34±0.03	1.74±0.07
Chymotrypsin	245	1.32±0.03	1.75±0.07
Alpha chymotrypsin *A* (tosylated)	236	1.32±0.03	1.75±0.08
Concanavalin (demetallized)	237	1.30±0.04	1.72±0.08
Concanavalin *A*	237	1.30±0.04	1.73±0.08
Tosyl-elastase	240	1.33±0.03	1.81±0.10
Immunoglobulin FC (chain *D*)	249	1.31±0.02	1.70±0.07
Immunoglobulin FAB	216	1.28±0.03	1.61±0.06
Immunoglobulin G1	216	1.28±0.03	1.62±0.05
Alpha-lytic protease	198	1.37±0.05	1.71±0.06
Acid protease (penicillopepsin)	323	1.35±0.04	1.70±0.05
Acid protease (rhizopuspepsin)	325	1.35±0.04	1.71±0.06
Acid protease (endothiapepsin)	330	1.32±0.03	1.72±0.06
CTP-liganded aspartate carbamoyltransferase	310	1.38±0.02	1.67±0.04
Bacteriochlorophyll-*A* protein	358	1.27±0.01	1.65±0.06
Elastase complex with trifluoroacetyl	240	1.33±0.03	1.81±0.11
Gamma chymotrypsin *A*	236	1.33±0.02	1.75±0.08
Alpha-chymotrypsin	239	1.33±0.05	1.75±0.75
Gene 5/DNA binding protein	87	1.30±0.04	1.57±0.09
Papain (domain 2)	110	1.34±0.03	1.56±0.06
Hemagglutinin (HA2)	175	1.37±0.02	1.58±0.10
Actinidin (domain 2)	86	1.35±0.04	1.67±0.10

formity with the data obtained by Allen *et al.* using the different computational method.[4] On the other hand, the value of fractal dimension D_1 reflecting the local conformation of protein is smaller than the value of fractal dimension D_2 reflecting the global conformation of protein.

In Table I, the fractal dimensions D_1 and D_2 of 14 hemoproteins are centrally distributed near 1.41 and 1.62, respectively. The same features are found in the homologous protein cytochrome *C* (see Table IV). This means that although there are some distinct differences in the primary structure of homologous proteins, the property of their three-dimensional structure tends to be conservative in the process of evolution. In other words, the fractal dimension of proteins may tend to be conservative in the process of evolution.

Table III shows the values of fractal dimensions D_1 and D_2 of the α-β class which is the largest structural grouping in the classification of protein[8] and includes three subgroupings (a), (b), and (c) (see Table III). The mean values of fractal dimensions D_1 for (a), (b), and (c) subgroupings are 1.41, 1.37, and 1.36, respectively, and that of D_2 in the same case are 1.64, 1.68, and 1.72, respectively. The magnitude of D_1 for the α-β class decreases in the order of (a), (b), and (c), but that of D_2 increases in the same order. These results would be helpful for studying the principles of folding and structure of protein.

In Table V, the mean values of fractal dimensions D_1 and D_2 for four structural classes and for 90 proteins are listed, respectively. The magnitude of the mean values of the fractal dimension D_1 for four structural classes decreases in the order of α, SD, α-β, and β, where the mean values for the α class and the SD class are almost the same because there are some α-helix structures in the SD class, such as cytochrome *C*, C2, C3, C5, C550, C551, and B5 in the type of small metal-rich domains, and insulin in the type of small disulfide-rich domains. One can put these proteins into the α class.[9] In addition, the α

TABLE III. Fractal dimensions of each protein for the α-β class.

Protein	Residue number	Fractal dimension D_1	D_2
(a) Singly wound parallel β barrels			
2-keto-3-deoxy-6-phosphogluconate aldolase	173	1.41±0.03	1.64±0.06
(b) Doubly wound parallel β sheets			
Subtilisin (BPN)	275	1.37±0.02	1.67±0.05
L-arabinose-binding protein	306	1.37±0.02	1.59±0.04
Flavodoxin	147	1.36±0.02	1.66±0.07
Glutathione peroxidase (bovine)	184	1.37±0.01	1.68±0.08
Hexokinase *A* and glucose complex	457	1.37±0.01	1.70±0.08
Adenylate kinase	194	1.38±0.02	1.64±0.07
Dihydrofolate reductase complex	162	1.35±0.04	1.75±0.09
Dihydrofolate reductase complex with methotrexate	157	1.34±0.04	1.73±0.09
Flavodoxin (oxidized)	!38	1.37±0.04	1.69±0.07
Flavodoxin (semiquinone form)	138	1.38±0.03	1.69±0.07
Apo-*d*-gyceraldehyde-3-phosphate dehydrogenase	333	1.38±0.03	1.69±0.05
Glutathione reductase	461	1.35±0.01	1.66±0.05
Apo-liver alcohol dehydrogenase	374	1.38±0.02	1.64±0.04
Holo-liver alcohol dehydrogenase	374	1.38±0.02	1.66±0.04
Catalase (beef liver)	498	1.34±0.01	1.70±0.06
(c) Miscellaneous parallel α-β			
Carbonic anhydrase *C*	254	1.30±0.02	1.73±0.09
Carboxypeptidase *B*	299	1.40±0.02	1.66±0.05
D-glucose 6-phosphate isomerase	514	1.36±0.01	1.73±0.08
D-alanyl-*d*-alanine carboxypeptidase	348	1.38±0.02	1.77±0.09
Carbonic anhydrase *B*	256	1.32±0.02	1.74±0.09
Carboxypeptidase *A*	307	1.38±0.02	1.66±0.05

TABLE IV. Fractal dimensions of each protein for the SD class.

Protein	Residue number	Fractal dimension D_1	D_2
Small disulfide-rich domains			
Insulinlike growth factor I	70	1.38±0.10	1.56±0.08
Insulinlike growth factor II	67	1.40±0.02	1.56±0.08
Phospholipase A2	123	1.41±0.01	1.69±0.10
Crambin	46	1.44±0.03	1.61±0.09
Alpha-bungarotoxin	74	1.35±0.05	1.61±0.09
Erabutoxin *B*	62	1.36±0.07	1.62±0.10
Alpha cobratoxin	71	1.36±0.07	1.63±0.09
Ferredoxin (spirulina platensis)	98	1.38±0.05	1.64±0.07
Wheat germ agglutinin (isolectin 2)	171	1.50±0.05	1.57±0.03
Small metal-rich domains			
Cytochrome *C*	111	1.42±0.02	1.62±0.07
Cytochrome *C* (oxidized)	103	1.43±0.02	1.62±0.06
Cytochrome *C* (reduced)	103	1.43±0.02	1.62±0.07
Ferrocytochrome *C*	103	1.44±0.01	1.56±0.05
Cytochrome C2 (oxidized)	112	1.42±0.02	1.62±0.06
Cytochrome C2 (reduced)	112	1.42±0.01	1.62±0.06
Cytochrome C3	107	1.42±0.02	1.56±0.04
Cytochrome C5 (oxidized)	83	1.47±0.03	1.62±0.07
Cytochrome C550	134	1.41±0.02	1.65±0.08
Cytochrome C551 (oxidized)	82	1.43±0.02	1.58±0.06
Cytochrome C551 (reduced)	82	1.43±0.02	1.58±0.06
Cytochrome B5 (oxidized)	85	1.39±0.02	1.55±0.05
Ferredoxin (peptococcus aerogenes)	54	1.42±0.06	1.69±0.12
Feeredoxin (nonheme Fe protein)	106	1.35±0.04	1.60±0.07
Oxidized high potential iron protein	85	1.43±0.03	1.60±0.07

and β classes show the largest and the smallest values of fractal dimension D_1, respectively, since the local structure of the β class is extended more than that of α class and that of the α-β class is in between. In the last column of Table V, the mean values of fractal dimensions D_1 and D_2 for 90 proteins are 1.38 and 1.65, respectively, where $D_2 = 1.65$ is very close to the theoretical value $\frac{5}{3}$ associated with a self-avoiding random walk in three-dimensional Euclidean space.[3,10]

Finally, we rule out the possibility of the correlations between the fractal dimensions D_1 and D_2 and between the fractal dimension and the chain length by plotting the scatter diagrams of the 90 proteins. Figure 2 is the scatter diagram of the fractal dimensions D_1 and D_2 for 90 proteins. As shown in Fig. 2, no correlation is found between the fractal dimensions D_1 and D_2. This consequence is the same as Ref. 6. Since the local structure of protein is mainly determined by the local interactions, no correlation between fractal dimension D_1 and the chain length is a natural consequence. Therefore the scatter diagram of the fractal dimension D_1 and the chain length is not reported in the paper. Figure 3 shows the dependence of the fractal dimensions D_2 on the chain length. In Fig. 3, the fractal dimension D_2 seems to be independent of the chain length. But distinct negative correlation between the fractal dimension D_2 and the chain

TABLE V. The mean values of fractal dimensions D_1 and D_2 for four structural classes and 90 proteins.

	α	β	α-β	SD	90 proteins
Mean values of D_1	1.41	1.33	1.37	1.41	1.38
Mean values of D_2	1.61	1.69	1.69	1.61	1.65

length has been obtained by Isogai and Itoh[6], if four points are omitted from figure when the chain length is greater than 400 residues.[6] Note that Isogai and Itoh[6] did not evaluate the value of mean fractal dimension D_2 and only examined 43 proteins in their work.

Computer simulation has been performed on the Micro VAX 3500 and VAX 8700 computer of the University of Science and Technology of China. This work has been supported by Grant No. 9388006-5 from the Chinese National Science Foundation.

[1]S. Havlin and D. Ben-Avraham, J. Phys. A **15**, L311 (1982).

[2]H. J. Stapleton, J. P. Allen, C. P. Flynn, D. G. Stinso, and S. R. Kurtz, Phys. Rev. Lett. **45**, 1456 (1980).

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

[4]J. P. Allen, J. T. Colvin, D. G. Stinson, C. P. Flynn, and H. J. Stapleton, J. Biophys. **38**, 299 (1982).

[5]R. Elber and M. Karplus, Phys. Rev. Lett. **56**, 394 (1986).

[6]Y. Isogai and T. Itoh, J. Phys. Soc. Jpn. **53**, 2162 (1984).

[7]Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, New York.

[8]*Advances in Protein Chemistry,* edited by C. B. Anfinese, J. T. Edsall, and F. M. Richards (Academic, New York, 1981), Vol. 34, pp. 168–339.

[9]*Principles of Protein Structure,* edited by G. E. Schulz and R. H. Schirmer (Springer-Verlag, New York, 1979), p. 88.

[10]J. C. LeGuillou and J. Zinn-Justin, Phys. Rev. Lett. **39**, 95 (1977).