Self similarity of protein surfaces

Thomas Goetze and Jürgen Brickmann

Institut für Physikalische Chemie, Technische Hochschule Darmstadt, Petersenstr. 20, D-6100 Darmstadt, Germany

ABSTRACT   Scaling properties of the surfaces of 53 proteins are studied on the basis of experimentally determined 3-D structures of these biological macromolecules. It is found that the surfaces show self similarity within a yardstick range of $1.5\,\text{Å} < \epsilon < 15\,\text{Å}$. The self similarity is measured by the fractal dimension D of the surface. Two different algorithms for the determination of the fractal dimension are applied, both based on cubic yardstick particles. One is related to the contact surface (CS), which was first introduced by Connolly (17), while the other corresponds to the solvent accessible surface (SAS) of Richards (9). The fractal dimensions of both are different. While the CS type approach leads to relatively high values of D in the range 2.5 to 2.6, the SAS approach gives fractal dimensions of $D \approx 2$.

1. INTRODUCTION

Chemistry in heterogeneous systems is even more complex than chemistry in a homogeneous environment. There have been numerous attempts to model heterogeneous reactions taking the microscopial environment into account (1). Geometrical models for the description of surface structure, or, more generally, of the local anisotropy of a molecular environment, are generally not sufficient to describe all properties relevant to the molecule's motion in this environment. A more general approach to take local anisotropy into account can be applied when the molecular environment shows self similarity within a certain scaling range (1–14). It has been found that a variety of structures in nature are self similar, ranging from astrophysical systems down to the molecular level of proteins (2–8, 10–14). On one hand proteins are definitely molecules, i.e., one can uniquely analyze sequences of aminoacids building up the molecule, and in many cases the three-dimensional (3-D) structures are also known from x-ray analysis and 2-D NMR investigations. On the other hand, proteins have many properties in common with macroscopic microparticles. This is particularly true for the interaction of proteins with other (small) molecules. Only a limited number of atoms and chemical groups building up the complete protein are located at the surface, i.e., in the region which is accessible to the reaction partner.

The role of self similarity in absorption phenomena and heterogeneous catalysis has been investigated by a variety of authors (2–4). There have also been some attempts to transfer this concept to the study of proteins (3–7, 11–14). Pfeifer and co-workers (7) as well as Avnir and co-workers (5, 11) have analyzed the van der Waals

surface of the lysozyme molecule. They demonstrated that this molecule seems to show self similarity, and they measured this self similarity from fractal analysis of planar cuts through the surface. Lewis and Rees (4) as well as Äquist and Tapia (6) studied the fractal dimension of protein surfaces on the basis of the Connolly algorithm (17). The latter generates a contact surface (CS) of the biomolecule with a spherical particle. The authors defined a surface dimension D (the fractal dimension [15,16]) as a local surface property and tried to assign areas of high fractal dimension to places of high receptor selectivity. The use of a local surface dimension may be useful for the discussion of molecular recognition problems but self similarity alone does not explain the high selectivity of protein substrate interactions. However, the principle of surface self similarity may be one of the generic ones in biomolecular evolution. Pfeifer and co-workers argued (3, 7, 10, 13) that a fractal dimension of protein surfaces slightly larger than $D = 2$ seems to represent a compromise between two different transport processes for a molecule to the receptor site on the surface. While the transport from the bulk phase to the surface is enhanced by a surface dimension $D > 2$, the transport along the surface from the place of first contact to the place of the receptor is hindered by a high surface dimension.

It is the aim of this paper to find out whether self similarity of the surfaces is a generic property of all proteins and, if this is true, whether there are systematic trends in the fractal dimension when the size of the protein is changed. For this reason we analyze as many proteins as possible with respect to their surface proper-

ties. In section 2 some facts from fractal analysis of surfaces are reviewed. A systematic study of proteins with well known 3-D structure is only possible within acceptable numerical effort if there are effective numerical strategies available. Section 3 thus deals with a new model approach and the algorithm for the determination of the surface dimension used in this analysis. The results of the analysis of 53 proteins (the 3-D structures are from the Protein Data Bank [Brookhaven] and the coordinates of the H-Atoms are added according to standard distances and angles) are presented in section 4 while the final section contains some conclusions.

2. THE FRACTAL DIMENSION OF A SELF SIMILAR SURFACE

Pfeifer (3) called the fractal dimension the hidden symmetry of irregular (self similar) surfaces. The self similarity of a line, a surface or any other object can be determined by measuring its fractal dimension (15). This can be done, at least in principle, by covering the object with spheres of radius ϵ and counting the number $N(\epsilon)$ of spheres. The object has the fractal dimension D if $N(\epsilon)$ scales according to

$$N(\epsilon) \sim \epsilon^{-D} \tag{1}$$

for all $\epsilon > 0$.

For mathematical objects which are defined via a unique generating algorithm, the determination of D may be complicated and numerically time consuming although it can be done to any desired degree of accuracy. For real microscopic objects like proteins and other microscopic surfaces, additional problems occur that are not related to numerical efforts. Molecular surfaces are not existent per se. A molecular surface is a way to rationalize and visualize the interaction of test particles with the system and there is no way to cover the surface with a set of spheres or other yardstick objects unambiguously. Pfeifer and co-workers (7) analyzed two-dimensional cuts through the surface of the lysozyme molecule, as defined by the interactions with a water molecule in its energetically most favorable orientation, by measuring the lengths with different yardsticks, and found a line dimension of $D_L = 1.2$. Using the fact that a cut through a fractal surface with dimension D has a dimension $D_L = D - 1$ with a probability approaching one (18), the authors established an average surface dimension of $D = 2.2$ for lysozyme within a lower cutoff for the yardstick of $\epsilon = 1$ Å and a higher cutoff of $\epsilon = 20$ Å. The surface (i.e., the planar cuts) was generated using a water probe and hence the results of Pfeifer et al. are evident only for this test molecule. Considering molecules with a size different to that of water definitely leads to different surfaces and different surface dimensions (20). The situation here is similar to all cases where microscopic (i.e., molecular real) yardsticks are used: the yardstick itself defines the surface. An analysis which takes this into account was presented by Lewis and Rees (4) as well as Åquist and Tapia (6). These authors used Connolly surfaces (17), which are defined as the surface that contains all contact points of the molecule (represented as a set of spheres with well defined radii around the positions of the atoms) and a test particle (represented by a sphere of radius R [see Fig. 1]). Lewis and Rees (4) found an average of $D = 2.4$ for the fractal dimension from the relation

$$2 - D = \frac{d \log (A_S)}{d \log (R)}, \tag{2}$$

where A_S and R are the molecular surface area and probe radius, respectively, for lysozyme, ribonuclease A, and superoxide dismutase within a yardstick range 1 Å $< \epsilon < 3.5$ Å. Åquist and Tapia (6) found $D = 2.15$–2.21 for holo-RBP, the PAB monomer, the satellite tobacco necrosis virus monomer, and lysozyme within 1.5 Å $< \epsilon < 3.0$ Å from a similar type of analysis. Avnir and Farin (5) used a similar method and found a relatively high value of $D = 2.62 \pm 0.01$ for trypsine and an even higher one for the trypsine active site of $D = 2.80 \pm 0.04$. The determination of D on the basis of the Connolly surfaces takes into account that there is no uniquely defined molecular surface but that the yardstick molecule defines the surface via its interaction with the biomolecule. Unfortunately, this algorithm leads to drastically increasing computation time when the radius of the probe is increased. In the next section we describe two algorithms, which are more crude than that of Connolly, but still take the size of the test particle into account while defining the surface. The advantage of these new algorithms is that they are extremely fast and so enable us to study the surfaces of a large number of proteins for which the 3-D structures are known. It is not the aim of this paper to establish a certain value for the fractal dimension of a protein surface because this value definitely depends on the way the surface is generated. The questions we raise here are: (a) is there generally self similarity in the surface of proteins?, and if the answer is yes, (b) do all proteins have the same fractal dimension, or is there any relation between D and other quantities characterizing the biomolecule?

3. MODEL SURFACES

As was mentioned before, contact surfaces (CS) of the Connolly type (17) can only be determined with finite numerical effort if the radius of the probe particle does not exceed a value of $\epsilon = 3.5$ Å. Because, on the other hand, the lower bound of the radius is determined on physical grounds as $\epsilon = 1.5$ Å (roughly the probe of a water molecule) one has only a very limited scaling range for the analysis of self similarity, i.e., the determination of D is obtained from the slope of a small range in a log/log plot. In this work we use models that are based on the hard sphere model of proteins (see Fig. 1). The numerical data for the atomic radii used in this work are listed in Table 1. Our approach is similar to that of the contact surface generation, but instead of rolling a test particle over this hardsphere surface, we simply use cubic grids with different cube sizes to define the bulk and the surface of the biomolecule. The procedures are schematically described in Figs. 2 and 3.

Model A. The molecule is embedded in a 3-D grid of cubes with edge lengths ϵ. A cube belongs to the molecule if its center is an inner point of at least one sphere defining the hard sphere model of the protein (see Fig. 2). This model has some similarity with the contact (CS) surface generated by spherical particle. For the Connolly surface the curvature of concave areas is exactly that of the test sphere, and canyons and cavities in the surface have an extension which is larger than or equal to the diameter of the test sphere. The latter statement is also true for the present model, i.e., the model should give the right scaling behavior when the surface topology is studied. The curvature of the surface cannot be extracted from our treatment. However, the surface area has a meaning similar to the one in the Connolly approach: it represents the common surface of a cubic test particle with the protein.

Model B. Contact surfaces obtained with the Connolly scheme or the algorithms of Model A are not well suited for the determination of the distance a particle of given size has to travel in order to get from one surface point to another. For this reason one needs the solvent accessible surface (SAS), first introduced by Richards (9), i.e., a surface that is generated by the centers of the test particles rolling along the protein. With increasing radius of the test sphere the volume
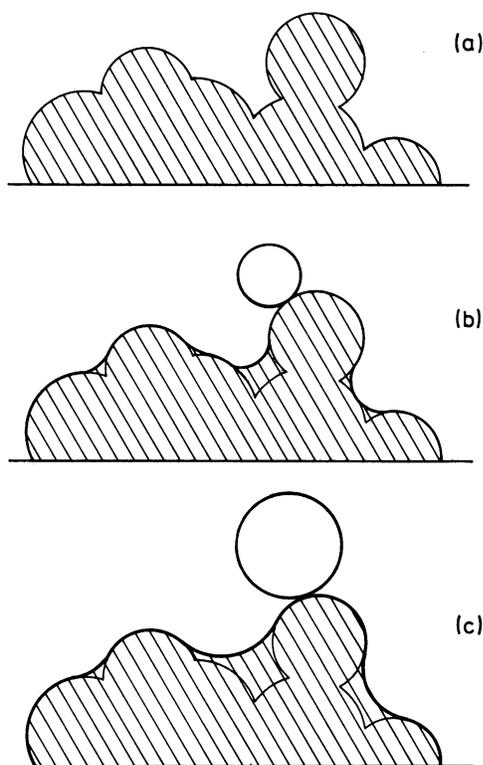
FIGURE 1 Hard sphere models. (a) For a molecule. (b) Connolly type surfaces. (c) Connolly type surfaces with test sphere of different radius.

FIGURE 2 Hard sphere models. (a) For a molecule. (b) and (c) Model surfaces generated with cubic yardsticks of different sizes according to model A (see text).

of such an object increases systematically. To generate a model which uses cubes of a given size (vertex length ϵ) as test particles, and which takes the systematic increase of the volume into account, we used the same procedure as in model A, but changed the hard sphere model by replacing the individual radii r_i for the protein atoms by $r_i + \epsilon/2$ (see Fig. 3, *dotted surfaces*).

These two models are particularly suited for the numerical analysis. Both the coordinates and the bookkeeping about the status of a cube can be treated within a simple but very effective integer algorithm. The details are published elsewhere (19). For both models three quantities are counted as functions of the cube size:

$N_B(\epsilon)$, the number of cubes belonging to the molecule;

$N_S(\epsilon)$, the number of cubes with at least one outer square (surface cube);

$N_A(\epsilon)$, the number of surface square.

For regular objects, i.e., those with volume dimension $D_B = 3$ and surface dimension $D_S = 2$, one expects for both models A and B

$$\lim_{\epsilon \to 0} (N_B(\epsilon) \cdot \epsilon^3) = V \qquad (3)$$

TABLE 1   van der Waals radii used in this work

Atom	H	C	N	O	Mg	P	S	Mn	Fe	Co
$r/\text{Å}$	1.30	1.80	1.65	1.60	1.70	2.18	2.11	1.70	1.70	1.70

and

$$\lim_{\epsilon \to 0} (N_S(\epsilon) \cdot \epsilon^2) \leq S, \qquad (4)$$

where V is the volume and S the surface area of the object, respectively. The validity of Eq. 4 is schematically demonstrated in Fig. 4 in a two-dimensional representation. The number of squares counted as surface squares in the algorithm described above (Fig. 4 a) is less than or equal to those necessary to cover the curve (Fig. 4 b). As a consequence of Eq. 4, $N_S(\epsilon)$ is not well suited to study scaling properties of the surface. Instead, we define a quantity $N_A(\epsilon)$ which is more appropriate for this investigation. The limit

$$\lim_{\epsilon \to 0} (N_A(\epsilon) \cdot \epsilon^2) = A \qquad (5)$$

is generally not identical to the surface area S for regular objects. A is the sum of areas $A_{x1}, A_{x2}, \ldots A_{y1}, \ldots A_{z1}, A_{z2} \ldots$ which occur as single valued projections of surface regions A_i onto the planes $x = 0, y = 0$ and $z = 0$, respectively, where (x, y, z) are the axes of the orthogonal coordinate system of the cubic grid. Note that a sphere with radius d has a surface area of $S = 4 \cdot \pi \cdot d^2$ which is definitely smaller than the sum of projections:

$$A = 6 \cdot \pi \cdot d^2 = \frac{3}{2} \cdot S. \qquad (6)$$

Following the argumentation of Kappraff (21) such a proportionality relation can be established for all objects that can be defined by a generating algorithm (dependent on a scaling parameter $\Delta = \epsilon^2$,
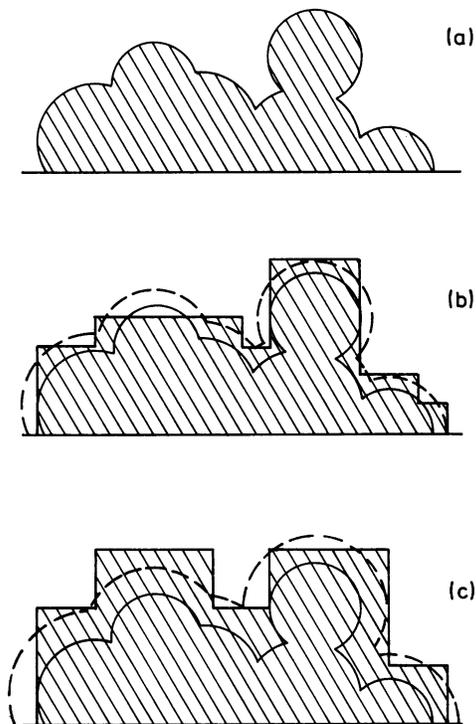
FIGURE 3 Hard sphere models. (*a*) For a molecule. (*b*) and (*c*) Model surfaces generated with cubic yardsticks of different sizes according to model B (see text).

FIGURE 4 Two dimensional representation of the algorithm. (*a*) To calculate the number of surface cubes N_S in a grid representation as was used in the model approach. (*b*) Representation of those cubes which are necessary for a complete cover of the surface (schematically).

which, for simplicity, may be a surface element of given area), i.e.,

$$A(\Delta) = G \cdot S(\Delta) \tag{7}$$

and that $S(\Delta)$ follows a scaling law

$$S(\Delta) = C \cdot \Delta^{\frac{2-D}{2}} \tag{8}$$

with an arbitrary C. Then $A(\Delta)$ also scales according to

$$A(\Delta) = C \cdot G \cdot \Delta^{\frac{2-D}{2}} \tag{9}$$

if and only if the factor G does not depend on the scaling parameters Δ. For practical applications, i.e., the determination of the fractal dimension D from $A(\Delta)$, it is sufficient that G does not systematically depend on Δ, but only fluctuates around a mean value G with varying Δ. To decide whether such a situation is found for protein surfaces or, if not, how such a situation can be generated, one has to know the geometrical meaning of the factor G.

Let B be an object and $F(\Delta)$ be a triangle representation of the surface of B with resolution Δ. $F(\Delta)$ may be a set of $N_S(\Delta)$ triangles of equal surface area Δ. The total surface area is then given by

$$S(\Delta) = N_S(\Delta) \cdot \Delta \tag{10}$$

while

$$A(\Delta) = \sum_{i=1}^{N_S(\Delta)} (A_{xi} + A_{yi} + A_{zi}) \tag{11}$$

with the projections A_{xi}, A_{yi}, A_{zi} of the triangle i onto the x-, y-, and z-plane, respectively. These projections are simply given as

$$A_{wi} = |\cos \alpha_{wi}| \cdot \Delta$$
$$w = x, y, z \tag{12}$$

with the direction cosine $\cos \alpha_{xi}$, $\cos \alpha_{yi}$, $\cos \alpha_{zi}$ of the normal vector on triangle i with respect to the coordinate axes. From Eqs. 7 and 10–12 one directly obtains

$$G(\Delta) = \frac{1}{N_S(\Delta)} \cdot \sum_{i=1}^{N_S(\Delta)} (|\cos \alpha_{xi}| + |\cos \alpha_{yi}| + |\cos \alpha_{zi}|). \tag{13}$$

For isotropically oriented triangles $G(\Delta)$ approaches the value $G_\infty(\Delta) = \frac{3}{2}$ for $\Delta \to 0$ and $N_S(\Delta) \to \infty$, as in the case of the sphere Eq. 6.

A triangle representation of a protein surface with given resolution does not, in general, show isotropically distributed normal direction of the triangles even if one could say that the distribution should not be far away from this average for globular proteins. However, for the determination of D from the scaling property of $A(\Delta)$ it is not necessary that the directions of the surface elements are isotropically distributed. It is only important that $G(\Delta)$ does not systematically depend on Δ and that fluctuations are small as Δ is varying.

In this study we determine $A(\epsilon)$ for a given protein as an average of n independent calculations

$$\bar{A}(\epsilon) = \frac{1}{n} \sum_{\alpha=1}^{n} A^\alpha(\epsilon), \tag{14}$$

wherein the grid coordinate systems of a pair of calculations $A^\alpha(\epsilon)$ and $A^\beta(\epsilon)$ differ by a random displacement of the origin and a random rotation. The procedure is schematically shown in Fig. 5. This procedure has two main advantages.

(a) The replacement of the protein bulk by a set of cubes is a relatively crude approximation when the cube size is increased up to a value of 5–15 Å. In this range, the shape and the volume is affected by the relative position of the molecule with respect to the grid. This effect can be drastically reduced if the number of averages n is chosen sufficiently large. In the numerical tests we found that there are no changes of $\overline{A}(\epsilon)$ larger than 1% if $n = 35$ is chosen.

(b) If there are small cavities and canyons in the molecular surface, individual calculations may give misleading results (see for example Fig. 2). A gap in the surface of volume ϵ^3 is indicated when the center of the cube incidently is located in this narrow channel, while a Connolly type algorithm would automatically exclude most of the channel area. Only a few of the members in the sample which have the "wrong" shape will contribute to the average, and the effect will be smoothed out.

4. RESULTS

We have analyzed the surfaces of 53 proteins within the contact surface (CS) approach of model A as well as within the solvent accessible surface (SAS) approach of model B.

In a first series of test calculations the cube edge was varied in the interval $0.1 \text{ Å} < \epsilon < 10 \text{ Å}$. The results from model A for the 1fx1 molecule are shown in Fig. 6, a–c. It is seen that there are two ranges with different slope in the $\log N_A(\epsilon)/\log \epsilon$ plots. For $\epsilon < 1.5$ Å the numerical data can be well fitted with $D = 2.089$ (Fig. 6 b). Here $\overline{N}_A(\epsilon)$ was averaged over 35 independent runs. For $\epsilon > 1.54$ Å a value of $D = 2.581$ was obtained from the

(a)

(b)
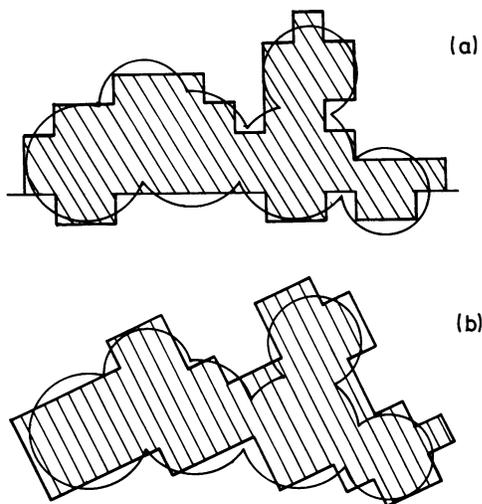
FIGURE 5 Surfaces a and b generated with cubic yardsticks of identical resolution but with relative displacement and rotation.

scaling property of the cube surfaces (Fig. 6 c). The low dimension in the range of small ϵ values is not of any importance for any protein molecule interaction because there are no real molecules, or molecular parts in this scaling range. We concentrated the analysis for the set of molecules studied here to the range of high values. The lowest value in all studies was taken as $\epsilon = 1.5$ Å while the upper limit was determined from the condition

$$\frac{\overline{N}_S(\epsilon)}{\overline{N}_B(\epsilon)} < 1, \qquad (15)$$

i.e., the cube size was increased as long as there were cubes left without any surface element.

The fractal dimension and its interpretation in terms of self similarity are basically statistical concepts (11). It has been demonstrated by Avnir et al. (11) that one should be careful in analyzing single objects. In our approach we analyzed a sample of similar objects which are generated by the same algorithm from experimental data for the structure. The number of individual calculations for the sampling was increased up to an average value of $n = 35$. A higher number was not necessary because all the scaling exponents become stable within three digits with this sampling.

For numerical control the dimension D_B of the bulk of all the molecules under consideration was calculated from the scaling of $N_B(\epsilon)$. For all molecules $D_B = 3.00 \pm 0.01$ resulted, i.e., the volumes of the objects generated according to model A scale like regular bulky objects.

Following Avnir et al. (11) we also analyzed the local slope of the log/log plots. These authors have demonstrated that this type of analysis is very helpful to identify systematic deviations from linearity in $\log N/\log \epsilon$ plots. It is seen from Fig. 6 d that there are no such systematic deviations for the surface nor for the bulk. However, the curves show increasing fluctuations with increasing value of the yardstick as a consequence of the breakdown of the statistical self similarity concept for large ϵ values.

The results for D_A and D_B, as obtained with the aid of the model A approach, are listed in Table 2.

The data obtained from model B differ from those from model A both qualitatively and quantitatively. Exponents D_A and D_B result from linear fits of the $\log N/\log \epsilon$ plots, which are substantially smaller than the corresponding values from model A. One obtains $D_A = 1.899 + 0.022$ and $D_B = 2.496 + 0.012$ for 1fx1 (see Fig. 7 a).

From Fig. 7 b it seems that both $\overline{N}_B(\epsilon)$ and $\overline{N}_A(\epsilon)$ show the same systematic trends which we assume may be adequately described by the equations

$$\overline{N}_B(\epsilon) = N_{B0} \cdot F(\epsilon) \cdot \epsilon^{-D_B}$$

$$\overline{N}_A(\epsilon) = N_{A0} \cdot F(\epsilon) \cdot \epsilon^{-D_A} \qquad (16)$$
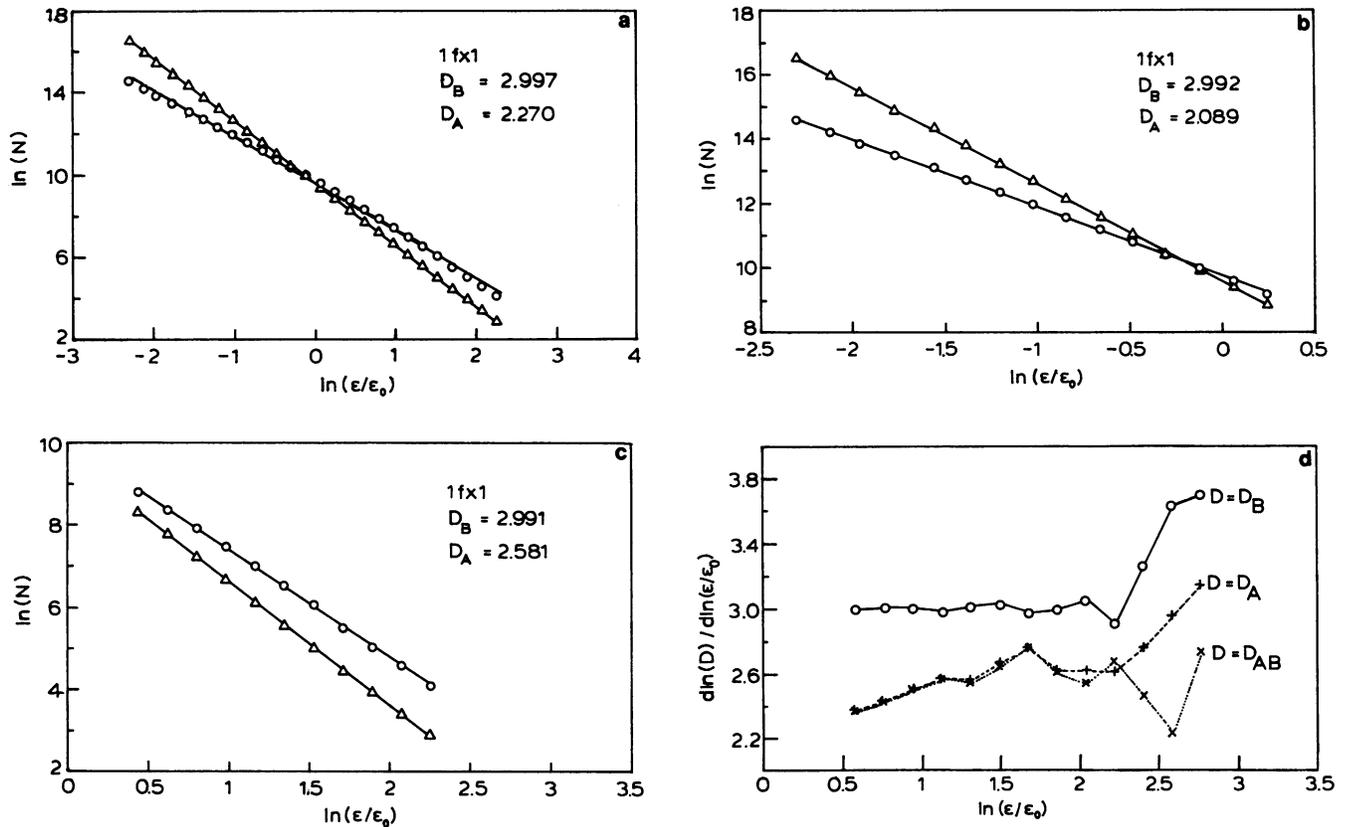
FIGURE 6 Average numbers of cubes (from 26 independent runs) $N = \overline{N_B}$ building the molecular bulk in an orthogonal grid representation of resolution as well as the average number of surface squares $N = \overline{N_A}$ as function of the yardstick ϵ for 1fx1 calculated with respect to model A. In a, the linear regression was taken over the complete yardstick range $0.2 < \epsilon < 15$ Å, while for b and c independent regressions for the ranges $0.2 < \epsilon < 1.5$ Å and $1.5 < \epsilon < 15$ Å, respectively, were performed. The bulk dimension is very close to $D_B = 3$ in the complete range as is expected, while the surface dimension converges towards $D_A = 2$ in the range of low ϵ values $\epsilon < 1.5$ Å and $D_A = 2.581$ is obtained from the slope of the linear regression in the range of high ϵ values. d Shows the local slope of the log (N)/log ϵ plots (c) and the ratio from equation 17 as obtained from a linear regression over three adjacent points.

with constant values of N_{A0} and N_{B0} and a function $F(\epsilon)$ which smoothly depends on the yardstick ϵ. This assumption is indeed justified.

Considering the scaling of the number of surface elements per volume element, one obtains

$$N_{AB}(\epsilon) = \frac{\overline{N_A}(\epsilon)}{\overline{N_B}(\epsilon) \cdot \epsilon^3}$$

$$= \frac{N_{A0}}{N_{B0}} \cdot \epsilon^{-(3-D_A+D_B)}, \quad (17)$$

i.e., for 1fx1 $N_{AB}(\epsilon)$ scales with an exponent $D_{AB} = 2.401 + 0.002$. The same value (within three digits) is obtained from a log $N_{AB}(\epsilon)$/log ϵ plot when the surface fraction $N_{AB}(\epsilon)$ is determined independently for each individual run and then averaged over 35 runs (Fig. 7 c). The numerical results for D_B, D_A, and D_{AB} as obtained from model B are also listed in Table 2.

There are large differences in the numerical values obtained for the surface dimension from model A and model B, but this fact is not surprising. Model A is related to the contact surface, the fractality of which may be relevant for adsorption phenomena, e.g., for the calculation of the number of molecules of given size which can be adsorbed in contact. Obviously this surface shows a larger amount of complexity than that calculated with Model B. The latter is more relevant for transport processes along the surface, for which the distance of the molecule's center between two positions, measured along the surface, is important. It is remarkable that the surface/volume ratio scales with a very similar exponent for both models.

We do not think that the actual value of the exponent D (the fractal dimension), which is obtained with one particular algorithm for the generation of the surface of a protein, has any universal importance. This value

TABLE 2   Fractal surface dimensions from different models

Mol name	No. of atoms	Model A			Model B		
		D_B	D_A	D_{AB}	D_B	D_A	D_{AB}
1ABP	4678	3.011	2.616	2.598	2.512	1.958	2.445
1BP2	1844	2.998	2.524	2.519	2.470	1.914	2.443
1CAC	4023	3.004	2.603	2.594	2.515	1.996	2.450
1CC5	1129	3.020	2.522	2.495	2.481	1.907	2.426
1CCR	1710	3.000	2.541	2.534	2.491	1.921	2.428
1CHO	4256	3.006	2.606	2.594	2.514	1.937	2.423
1CN1	7133	3.005	2.606	2.598	2.521	1.965	2.442
1CPV	1606	3.015	2.564	2.543	2.495	1.923	2.426
1CRN	642	3.015	2.470	2.447	2.419	1.825	2.404
1CSE	4833	3.010	2.602	2.587	2.536	1.922	2.386
1CTS	6893	3.000	2.610	2.607	2.492	1.971	2.478
1CTX	1066	3.031	2.522	2.482	2.399	1.876	2.476
1CY3	1754	3.005	2.462	2.454	2.391	1.948	2.555
1ETU	2711	3.000	2.571	2.565	2.479	1.905	2.425
1FBJ	6510	2.997	2.602	2.600	2.504	1.984	2.480
1FDH	4402	3.000	2.592	2.584	2.498	1.984	2.485
1FDX	724	3.036	2.515	2.472	2.477	1.862	2.384
1FX1	2144	2.991	2.582	2.578	2.496	1.899	2.401
1HCO	4384	2.023	2.619	2.589	2.508	1.980	2.471
1HDS	8841	3.005	2.624	2.614	2.550	2.045	2.495
1LH1	2388	2.989	2.549	2.556	2.495	1.918	2.422
1MCP	6674	3.000	2.605	2.601	2.500	1.968	2.466
1MLT	868	3.017	2.440	2.421	2.378	1.862	2.483
1TIM	5353	3.003	2.612	2.603	2.511	1.967	2.455
2ABX	2171	2.991	2.543	2.548	2.426	1.952	2.525
2APP	4552	2.992	2.591	2.595	2.529	1.939	2.409
2APR	4714	3.008	2.625	2.613	2.525	1.946	2.419
2AZA	3864	2.993	2.574	2.588	2.492	1.908	2.416
2CAB	3957	2.991	2.591	2.593	2.531	1.936	2.404
2CGA	7154	3.007	2.628	2.616	2.494	1.947	2.453
2CI2	1076	3.003	2.534	2.523	2.436	1.864	2.426
2CNA	3554	3.011	2.624	2.607	2.504	1.923	2.418
2CPV	1621	3.000	2.559	2.551	2.493	1.925	2.430
2CTS	6888	3.004	2.606	2.599	2.491	1.968	2.476
2TAA	7175	3.000	2.629	2.625	2.553	1.981	2.426
3ADK	3105	2.991	2.566	2.569	2.481	1.966	2.483
3CLN	2184	3.000	2.517	2.511	2.395	1.894	2.498
3CNA	3558	3.001	2.603	2.594	2.497	1.931	2.432
3CPA	4825	3.000	2.598	2.594	2.537	1.907	2.368
3CPV	1621	3.026	2.571	2.540	2.496	1.926	2.482
3ICB	1203	3.008	2.552	2.535	2.480	1.898	2.417
3PGK	6381	3.000	2.582	2.579	2.478	1.993	2.541
3WGA	4220	3.007	2.498	2.490	2.402	1.970	2.569
4ATC	14548	2.998	2.616	2.616	2.458	1.981	2.523
4CPA	4793	2.995	2.618	2.616	2.537	1.905	2.368
4CTS	13787	3.004	2.653	2.644	2.565	2.008	2.441
5API	5969	2.996	2.626	2.625	2.523	1.942	2.417
6ADH	11277	ˋ3.005	2.632	2.622	2.527	1.971	2.442
6PAD	3271	3.006	2.597	2.585	2.522	1.925	2.402
6PCY	1445	3.005	2.547	2.536	2.487	1.870	2.381
7ADH	5136	3.007	2.607	2.593	2.518	1.976	2.458
7CAT	7859	3.000	2.607	2.602	2.452	1.973	2.521
7LYZ	1958	3.017	2.556	2.542	2.503	1.915	2.410

Mol name: Name of dataset from Brookhaven database. No. of atoms: Number of atoms in dataset. D_B: D_{bulk}. D_A: D_{area}. D_{AB}: $D_{area/bulk}$.

creased as the radii are increased. A similar result was found by Zachmann et al. (20), who analyzed the Hausdorff dimensions of contact surfaces of one protein, but generated with different test particle sizes.

The important result of this investigation is the fact that all protein surfaces studied here showed a clear exponential scaling law with respect to the size of the yardstick particles. This may be interpreted as a manifestation of self similarity of the protein surfaces within a yardstick range of 1.5 Å < ϵ < 15 Å. The surface dimensions as obtained from model A are depicted in Figs. 9 and 10 as functions of the number of atoms of the protein. A correlation is found, namely that the average fractal dimension increases with increasing size of the biomolecule. Large molecules seem to appear less smooth than smaller ones to a potential substrate molecule of given size.

FIGURE 7 (a) Average number of cubes (from 34 independent runs) $N = \bar{N}_B$, as well as average number of surface elements $N = \bar{N}_A$ as functions of the yardstick for 1fx1 calculated with respect to model B. (b) Local slope calculated from a linear regression over three points from the log $\bar{N}$/log ϵ diagram (a) or the ratio from Eq. 17. (c) Averaged value for N_{AB} as a function of the yardstick (see Eq. 17).

changes when the algorithm is changed, as can be easily demonstrated by applying the procedures described above to a protein model for which the atomic radii are slightly modified. This is shown in Fig. 8. The surface dimension of the 1fx1 molecule is systematically de-

FIGURE 9 Surface dimensions D_A of proteins as functions of the molecular size (measured by the number of atoms in the molecule) for data from model A.
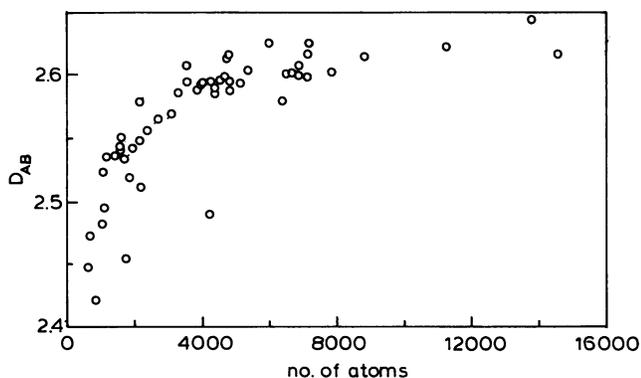
FIGURE 10   Scaling exponents D_{AB} (see Eq. 17) as functions of the molecular size as in Fig. 9 for data from model A.

5. SUMMARY AND CONCLUSIONS

Two algorithms are presented for the definition of a protein surface. Both use a rectangular grid of varying grid size. The first one (model A) defines the contact surface as it would be seen from a cubic yardstick particle, while the second one (model B) counts the surface elements which are defined as the set of points that can be reached by the center of the yardstick cube. The latter weakly corresponds to the solvent accessible surface (SAS). Both algorithms are applyed to 53 proteins for which the structures are taken from the Brookhaven Protein Data Bank. Hydrogen atoms are added according to standard binding data. For any protein up to 35 independent runs have been performed in order to minimize errors related to the position and orientation of the orthogonal grid. The results can be summarized as follows.

(a) Both the surface areas and the volume, as calculated by model A, show an exponential scaling law of the form

$$X = X_0 \cdot \epsilon^{-D} \qquad (18)$$

within a range of $1.5\ \text{Å} < \epsilon < 15\ \text{Å}$, where X is one of the properties mentioned above. D is interpreted as the fractal dimension. The volume dimension of model A is very close to $D = 3$ (within at least three digits for all molecules), while D becomes significantly less than 3 when model B is applied. A similar trend is observed for the surface dimension. The data resulting from model B are significantly smaller than those from model A. Moreover, the data from model B show systematic (but identical for $\overline{N}_A$ and $\overline{N}_B$) deviations from linearity in $\log N/\log \epsilon$ plots. For both models, however, the surface/volume ra-

tios scale with an exponent which is approximately equal.

(b) The fractal dimension of a protein surface generically increases when the size of the biomolecules is increased. This seems to show that small proteins appear less rough to molecular partners than large ones.

(c) In contrast to the fractal dimension of the backbone (9) the actual value D for the surface dimension of a protein is not a generic property of the biomolecule. The latter depends on the model approach. However, the fact that $2 < D < 3$ always results from different investigations manifests that the concept of fractal surfaces can be well applied for the investigation of surface related properties. In this work we did not extend the concept of fractal surfaces to the investigation of local properties like the surface complexity of receptor sites because, for this type of study additional information would be necessary. Because different model approaches result in different fractal dimensions (as global statistical properties) of the surface, one definitely has to prove which model is the relevant one for a particular question before one tries to use the statistical concept for the study of local properties.

REFERENCES

1. Somorjai, G. A. 1981. Chemistry in Two Dimensions: Surfaces. Garnell University Press. Ithaca, New York.

2. Avnir, D., editor. 1989. The Fractal Approach to Heterogeneous Chemistry—Surfaces, Colloids, Polymers. John Wiley and Sons, New York.

3. Pfeifer, P., and D. Avnir. 1983. Chemistry in noninteger dimensions between two and three. I. Fractal theory of heterogeneous surfaces. *J. Chem. Phys.* 79:3558–3565.
   Avnir, D., D. Farin, and P. Pfeifer. 1983. Chemistry in noninteger dimensions between two and three. II. Fractal surfaces of adsorbents. *J. Chem. Phys.* 79:3566–3571.

4. Lewis, M., and D. C. Rees. 1985. Fractal Surfaces of Proteins. *Science (Wash. DC).* 230:1163–1165.

5. Avnir, D., and D. Farin. 1990. Fractal scaling laws in heterogeneous chemistry. Part I: adsorptions, chemisorptions and interactions between adsorbates. *New J. Chem.* 14:197–206.

6. Åqvist, J., and O. Tapia. 1987. Surface fractality as a guide for studying protein–protein interactions. *J. Mol. Graphics.* 5:30–34.

7. Pfeifer, P., U. Welz, and H. Wippermann. 1985. Fractal surface dimensions of proteins: lysozyme. *Chem. Phys. Lett.* 113:535–540.

8. MacDonald, M., and N. Jan. 1986. Fractons and the fractal dimension of proteins. *Can. J. Phys.* 64:1353–1355.

9. Richards, F. M. 1977. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.

10. Pfeifer, P. 1984. Fractal dimension as working tool for surface-roughness problems. *Applic. Surf. Science.* 18:146–164.

11. Farin, D., S. Peleg, D. Yavin, and David Avnir. 1985. Applications and limitations of boundary-line fractal analysis of irregular surfaces: proteins, aggregates, and porous materials. *Langmuir.* 1:399–407.

12. Fushman, D. 1990. Surface fractality of proteins from theory and NMR data. *J. Biomol. Struct. Dynamics.* 7:1333–1344.

13. Elber, R. 1989. Fractal analysis of proteins. *In* The Fractal Approach to Heterogeneous Chemistry. D. Avnir, editor. John Wiley and Sons, New York. 407–424.

14. Zumhofen, G., A. Blumen, and J. Klafter. 1990. The role of fractals in chemistry. *New J. Chem.* 14:189–196.

15. Mandelbrot, B. B. 1982. The Fractal Geometry of Nature. Freeman, New York.

16. Feder, J. 1988. Fractals. Plenum Press, New York. 236–243.

17. Connolly, M. 1983. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 16:548–558.

18. Mattila, P. 1975. Hausdorff dimension, orthogonal projections and intersections with planes. *Annu. Acad. Sci. Fennicae.* 1:227–244.

19. Goetze, T. 1991. Fraktale Dimension von Proteinen. Dissertation. TH Darmstadt.

20. Zachmann, C. D., and J. Brickmann. 1992. The Hausdorff dimension as a quantification of local roughness of protein surfaces. *J. Chem. Inform. Comp. Sci.* In press.

21. J. Kappras. 1991. Connections. The Geometric Bridge between Arts and Science. McGraw-Hill, New York. p. 37.