Surface fractality as a guide for studying protein–protein interactions

J Åqvist and O Tapia*

Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala Biomedical Centre, Box 590, S-75124 Uppsala, Sweden

The concept of fractal dimension is applied to protein surfaces. Satellite tobacco necrosis virus, prealbumin, retinol binding protein and lysozyme have been studied. A residue fractal index has been defined, which provides a suitable colour code when using computer graphics for visualizing surfaces. Some provisions are made that render the MS algorithm useful to calculate protein surface fractal dimensions. It has been found that a correlation exists between regions of high fractal dimension and those involved in protein–protein interactions. The usefulness of surface fractality in this context is demonstrated by a molecular docking experiment.

Keywords: proteins, protein surfaces, fractal dimension, satellite tobacco necrosis virus, prealbumin, retinol binding protein, lysozyme

Received 21 October 1986
Accepted 28 October 1986

The surfaces of proteins are involved in important biological processes. Specific associations of multimeric proteins, polypeptide hormone–receptor interactions and antigenic regions on protein hormone–receptor interactions and antigenic regions on protein surfaces are illustrations of the important role that they play. The origin of specificity and recognition in biomacromolecular interactions is an important problem, the solution of which may have important consequences for rational protein design and engineering. In this respect, theoretical and experimental methods for analysing surface properties may be extremely useful.

A number of quantitative methods for representing protein (molecular) surfaces have been devised[1–4]. Recently, computational methods attempting to quantitate the topographic structure of a protein surface have been reported[5,6]; Connolly's algorithm is an interesting attempt to predict protein complexes from the coordinates of the individual proteins[5]; Fanning *et al.*[6] have proposed a characterization of the variegated surface topography with a molecular cartographic method. Lewis and Rees[7] and Pfeiffer *et al.*[8] have extended the concept of surface fractality (self-similarity with respect to the measured length scale) to protein surfaces. The molecular surface (MS) program[4] was used by Lewis and Rees to calculate surfaces, which were divided into subregions by spherical projections. Fractal dimensions were then calculated for each of these subregions. It has been reported[7] that surface regions of high dimension in superoxide dismutase were found to correlate to those regions involved in the dimer interface, and that the surface of lysozyme cross-reacting with monoclonal antibodies consists, in part, of one of the regions with high fractal dimension. However, the global fractal dimensions reported are dependent on the radius of the probe sphere in Connolly's algorithm[4]. This fact is in contradiction with the definition of fractal (Hausdorff) dimension. In this paper the elimination of this inconsistency will be discussed and it will be shown that a number of proteins have a unique (surface) dimension.

A step forward is taken here by defining residue fractal indices. By colour coding the protein molecule according to these, the fractal properties of the surface can be viewed on a graphics display. The hypothesis of Lewis and Rees, namely that zones of high fractality are more likley to interact and bind specifically can thus be tested more easily. This procedure may be of use for predicting macromolecular complexes, as has been shown in a docking experiment of retinol binding protein (RBP) with the prealbumin (PAB) tetramer, recently carried out by the authors[9].

FRACTAL DIMENSION

A fractal curve can be characterized by the special relationship between its length L and the incremental length ε.

$$L(\varepsilon) \propto \varepsilon^{-\gamma} \tag{1}$$

where γ is a positive constant, at least for some range of ε. The exponent γ can be defined with respect to the topological dimension D_T as $D - D_T$, where D is the fractal dimension.

More generally, the fractal dimension of a compact set may be defined through the relation:

$$N(\varepsilon) \propto \varepsilon^{-D} \tag{2}$$

where $N(\varepsilon)$ denotes the number of 'balls' of radius ε needed to cover the set. For a surface, such as that

*To whom all correspondence should be addressed

 0263-7855/87/010030-05 $03.00

of a protein ($D_T = 2$), the area would be related to the fractal dimension D as:

$$A(\varepsilon) \propto \varepsilon^{2-D} \quad (3)$$

The dimension D of a set is roughly the amount of information needed to specify a point of the set accurately[10]. In order for the fractal dimension to be a meaningful quantity, the set should possess self-similarity, i.e. invariance to scale transformations.

In practice, one must first calculate the surfaces, using some suitable algorithm. The variation of the surface area with the probe radius should then obey a relationship such as equation (3) if the calculated surface possesses a true fractal structure. If this is not the case, such as in Reference 7, it may be that the surface is not a fractal one, or that the algorithm used does not produce the 'correct' surface. A different approach to calculate the dimension of protein surfaces has been proposed by Pfeiffer *et al.*[8]. They have determined the fractal surface dimension of lysozyme, using a method based on sectioning of the molecule into a number of parallel 2D layers. For each section, the fractal dimension of the curve corresponding to the transition between attractive and repulsive interaction, experienced by a water molecule, is determined by counting the number of steps $N(\varepsilon)$ of length ε. Using the theorem stating that the dimension of the intersection of a set with a plane is $D - 1$, where D is the (fractal) dimension of the set, the protein surface dimension can be obtained. With this method, the limit $\varepsilon \rightarrow 0$ can be taken since the potential energy curves are continuous, thereby providing a route for a rigorous examination of the fractal set.

Connolly's molecular surface algorithm can be used to obtain surfaces that satisfy, within a limited range, the idea of self-similarity. The MS program[4] is used here with different probe sizes to obtain values of D for holo-RBP, the PAB monomer, the STNV (satellite tobacco necrosis virus) monomer and lysozyme (Lym1, Brookhaven data bank coordinates).

Figure 1(a) shows the dependence of the surface area on the probe radius for 1.0Å $< r <$ 3.0Å for RBP, if the molecular surface program is used. As is evident from the Figure, two length-scales (with different D) are clearly separated. This behaviour corresponds indeed to a breakdown of the self-similarity hypothesis for $r <$ 1.4–1.5Å, and can be understood after visually inspecting the generated surfaces on a graphics display. The transition is basically due to the algorithm used to calculate the molecular surface. The MS probe is placed adjacent (the distance determined by the atomic radii) to each protein atom at a number of positions. If the probe can be placed adjacent to one, two or three atoms (corresponding to convex, saddle and concave faces, respectively) without penetrating any other, surface points are generated. This is done irrespectively of whether the probe is on the outside or inside of the surface. So, even if there are no channels allowing the probe to move from the outside to the inside, surface points may be generated in the interior of the protein, provided there is room for the probe (see also Reference 5). When such isolated interior surface blobs appear, the topology of the surface is changed, and equation (3) is no longer valid. This transition occurs at slightly different probe radii for different proteins. The authors

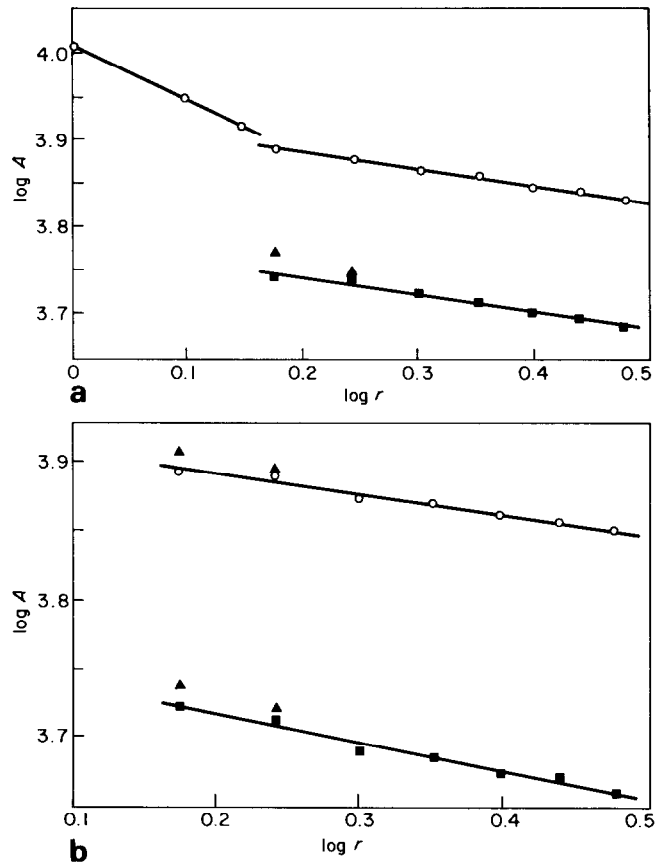

Figure 1. Plot of the dependence of the total surface area on the probe radius for (a) RBP: (○), D = 2.18; lysozyme: (■) D = 2.19; and (b) STNV: (○), D = 2.15; and PAB: (■) D = 2.21. The areas obtained for probe sizes 1.5Å < r < 3.0Å without the correction discussed in the text are denoted by (▲). The standard deviation of the regression coefficient for the fitted lines is ⩽ 0.02

chose 1.5 Å as the lower bound for the probe radius, when calculating the *D:s*. This choice still requires removal of some isolated (nonaccessible) surfaces for the other three proteins examined.

As can be seen in Figure 1 the correction gives a linear relationship between log(A) and log(r), in all cases. The overall surface dimensions are given in the Figure. For RBP, the authors have examined probe sizes up to 10 Å with the MS program[4], and the linear behaviour holds over this range. For even larger probes, the MS program is not suitable, because the requirement that the area should tend to zero in the large r limit is not fulfilled. The program will rather give some finite area for large probes, which above some limit will be independent of r. With the method of Pfeiffer *et al.*[8], the fractality can be examined over larger ranges of the measure length. They examined incremental lengths up to 20 Å, thus establishing fractal behaviour for the lysozyme surface over this range.

Lewis and Rees[7] have obtained a fractal dimension of 2.44 for lysozyme, which is at variance with the present results (D = 2.19) and those of Pfeiffer *et al.*[8] (D = 2.17). One reason for this discrepancy is likely to be the breakdown of self-similarity (see Figure 1). There is also some difference regarding the total molecular surface area of lysozyme. In the authors' calculations they obtained a total area of about 6200–6900

Å^2 for the lysozyme (coordinate sets Lyz6 and Lym1, Brookhaven data bank) using both the molecular surface program and the one by Lee and Richards[1] with the same atomic radii and a probe size of 1.4 Å. This is in reasonable agreement with the results of Chothia[11], who obtained an accessible surface area of 6480 Å^2 for lysozyme using the same probe size. At r = 1.5 Å the surface area obtained by Lewis and Rees is about 8000 Å^2 (as judged from Figure 1(b) of their paper).

On the other hand, the authors' results for lysozyme are in excellent agreement with those of Pfeiffer *et al.*[8]. The fact that two different methods for calculating the fractal dimension give essentially the same result reinforces the idea that this concept is a well defined property of the surface, and does not depend on computational procedures.

FRACTAL INHOMOGENEITY OF PROTEIN SURFACES

The procedure of Lewis and Rees[7] for dividing the surface into subregions is roughly followed here. The centre of mass of the protein molecule is placed at the origin of a spherical coordinate system. A grid is then defined by suitable steps in the polar angles (θ, φ); for each gridpoint all atoms with position vectors within an angle δ (30° was used) of the point contribute to the surface area associated to the gridpoint. A calculation of D is then performed for each (θ, φ) to obtain a fractal map of the protein surface. It would be desirable to be able to assign a value of D to each residue, rather than to spatial regions. It turns out, however, that the surface area of a single residue is too small to produce significant statistics for the fractal dimension. A larger portion of the surface must be included in order to lower the relative errors of the D:s. The grid procedure may, although it is not an optimal subdivision of the surface, be used to assign approximate D:s to each residue. It must be noted that the cones corresponding to (θ, φ) gridpoints will overlap if the entire surface is to be covered. At present the authors have used a weighted average over the gridpoint D:s to obtain an approximation for each amino-acid.

$$D_i = \frac{\sum_k w_{ik} D_k}{\sum_k w_{ik}} \qquad (4)$$

Here, D_i and D_k denote the fractal dimension of residue i and gridpoint (cone) k, respectively. The weight w_{ik} is the fraction of the total exposed surface area of residue i contributing to the area of cone k.

For the proteins examined here, the fractal dimensions were in all cases found to vary considerably over the surface, with D:s ranging from 2.0 to about 2.4. The observed correlation between high D regions and parts of the surface involved in protein–protein interactions, such as subunit interfaces, will now be examined. It is not within the scope of this paper to describe the various protein structures in detail, and reference should therefore be made to the relevant structural papers (see References 12–15).

STNV

The STNV surface provides a test case for studying the relationships between fractal indices and regions involved in protein–protein interactions, since each monomer has several contact areas with other subunits. The entire virus particle has icosahedral symmetry, resulting in subunit contacts of two-fold, three-fold and five-fold symmetry[12]. In Colour Plate 1 the STNV subunit is shown coded according to the residue fractal indices (see Colour Plate caption).

The region involved in the three-fold interactions (i.e. part of the α-helix and strands B and I together with the C–D loop)[12] is particularly well mapped by having a high fractal dimension ($D > 2.25$). Also most of the two-fold interactions (the end of strand E and the E–F and G–H loops) are found to be of relatively high dimension. An exception are residues 121–122, between strands E and F, where D is approximately 2.2. The five-fold contacts are not as well mapped by the fractal dimension as the two- and three-fold ones. In particular the F–G loop and residues 113–115 are in regions of relatively low D. The five-fold contacts involving the C-terminal part of the α-helix and parts of the D and I strands correlate, however, fairly well to high D values. Solvent molecules identified in the electron density map have, however, not been included in the calculations. This may have influence on certain parts of the surface and requires further study.

PAB

The 3D crystal structure of prealbumin[13] is tetrameric (see Colour Plate 2), which is also the form in which the protein occurs in the plasma. The two thyroxine binding sites are located in a channel running through the dimer–dimer interface[13]. This channel lies horizontally as depicted in Colour Plate 2. The association of two monomers is described as two molecules forming a dimer, interacting via a monomer–monomer interface. The tetramer is then made up of two such dimers, interacting through a dimer–dimer interface[13].

The entire dimer–dimer interface, including residues 19–23 and 110–112, has a rather high value of D (2.23–2.29). For the monomer–monomer interface, however, the interaction region consisting of residues 87–95 in strand F has an average D of 2.15, which is below the overall average. The other part of the monomer–monomer interface (strands G and H), which is the same backbone region partly involved in the dimer–dimer contacts, has a fairly high fractal dimension.

Interestingly, the largest part of the tetrameric surface with low fractal dimension coincides with the region that was proposed as a possible binding site for DNA[14]. The two F strands of a dimer (discussed above), belonging to the monomer–monomer interface, lie at the bottom of the proposed DNA binding groove.

Apart from the backbone region concerned in subunit interactions, there are also other parts of the chain with high D. These form four large symmetry related stretches, each involving three subunits. Two such adjacent stretches can be seen in Colour Plate 2 (one at the upper right and the other at the lower left). Each of these regions consists of the E–F loop (around Gln 63), the F–G loop (around Pro 102) of one subunit and residues 80–85 (C-terminal of α-helix) and part of

the G–H loop (around Tyr 114) from the other subunit in the same dimer. The region around Arg 21 of the A–B loop from one subunit of the second dimer also belongs to the same patch of high D. In the docking experiment carried out in Reference 9, this region is the candidate for binding RBP.

RBP

The RBP molecule forms an eight-stranded up-and-down antiparallel β-barrel[15], that encapsulates the retinol molecule. There is also an α-helix and an N- and a C-terminal coil. Holo-RBP forms a complex in the plasma with the thyroxine binding prealbumin tetramer[16]. The molecular details of this protein–protein interaction have not been investigated by X-ray crystallography yet. However, a proposition for the PAB binding site on RBP has been made, based on molecular dynamics calculations of both the holo and apo forms of RBP[17,18]. This region involves the loops at the β-barrel entrance.

The RBP molecule, colour coded according to the residue fractal indices, is shown in Colour Plate 3. Interestingly, the highest fractal dimension of the molecule corresponds to two of these loops (the ones connecting strands C–D and E–F). Together with part of the E-strand they make up a patch of high dimension (D = 2.3 for δ = 30°). There is also a small region on the opposite side of the molecule, consisting of a few residues of the N-terminal loop and C-terminal of the α-helix.

SURFACE FRACTALITY AND DOCKING

It was assumed that the regions of RBP and PAB involved in the formation of the plasma transport complex correspond to molecular surfaces of high fractal dimension, and a docking experiment was therefore carried out. The docking was done manually, visualizing relevant parts of the surfaces on a graphics display, using the program Hydra[19]. A large zone, fully including one and part of another symmetry related high-D patch of PAB, was selected together with the high-D region consisting of two of the entrance-loops of RBP.

As can be seen in the Colour Plates 4–6 the regions of high fractal dimension on RBP and the PAB tetramer possess also topographical complementarity. This preliminary model of the complex was derived only from rigid body docking, and no modification of the sidechain positions has been made. The molecular details of the docking model will be given and discussed in relation to available experimental data in a forthcoming publication[9]. This example, whether a lucky strike or not, shows that fractal dimensionality as a guide line for interactive molecular docking may be a useful tool.

DISCUSSION

The concept of surface fractal dimension has been extended to define residue fractal indices. By eliminating nonaccessible internal surfaces, generated by the molecular surface algorithm[4], the authors have shown that this procedure is suitable for calculating fractal surface dimensions. However, the spherical projection method used here (see also Reference 7) for division of the surface is not optimal, and can certainly be improved further. The fractal surface dimensions were calculated for lysozyme, STNV, RBP and prealbumin, using probe radii in the range 1.5–3.0 Å. It has been found that a correlation exists between regions of high fractal dimension and those involved in specific protein–protein interactions. The residue fractal indices, when used as colour code on a graphics display, were found to be very helpful in the docking RBP and PAB[9].

Initial clues of the value of fractals in connection with protein–protein and protein–ligand interactions were provided by Lewis and Rees[7]. However, local fractality may be sensibly altered by the presence of isolated interior surfaces around given residues. This is the case for lysozyme where, for instance, the region around residues 21–22 is affected by the correction to the molecular surface. This region is found to be one of low fractal dimension. Actually, this surface, as viewed on a graphics display, is quite smooth. No clear correlation was found between high fractality and antigenic regions for this protein.

Pfeiffer *et al.*[8] have discussed the relation between diffusion controlled ligand adsorption on a surface and the dimensionality of the surface. While the capture rate is enhanced by larger values of D, the diffusion along the surface, e.g. to an active site, is slowed by higher fractal dimension. The observation[7] that active sites are smoother than average would thus suggest the importance of the surface diffusion processes. The correlation of high values of D with protein–protein interfaces, on the other hand, may reflect the fact that the total interaction energy between the elements of a complex is maximized. For a given volume in space, the total interaction surface area may be increased by a high surface dimension. In a pictorial sense, the degree of roughness or irregularity corresponds to the local fractal dimension. The interaction specificity may be enhanced. This is because interfaces of high fractality would reduce the relative orientational freedom between the interacting surfaces.

Fractal models of protein structure have been proposed in quite a different context by Stapleton and coworkers[20–22]. Electron spin relaxation data from iron-containing proteins have been analysed in terms of a fractal model for the protein backbone. This model results in a density of vibrational states, different from that of Debye, that can account for the experimental data. It is important to realize that the fractal dimension in this case refers to the backbone chain in 3D space. Allen *et al.*[21] used X-ray data for 17 proteins to calculate a fractal dimension of the polypeptide backbone. The dimension obtained for lysozyme is 1.76 ± 0.05. It should be noted that there is no obvious relationship between this value of d and the dimension of the protein's surface. It is, however, interesting that both the curves traced by the backbone chain and the protein surface seem to possess fractal properties (statistical self-similarity). Besides, Avnir *et al.*[23] have shown experimentally that, in the molecular size range, the surfaces of most materials are fractals, i.e. surface geometric irregularities and defects are self-similar with respect to the variation of the resolution.

The fractal nature of protein surfaces is becoming an established fact. The capability of residue indices, based on this concept, to predict candidate zones for protein–protein or protein–ligand interactions has to be

documented with further work. In the present paper the authors have referred to one case where, combined with theoretical, experimental and molecular computer graphics techniques, the concept has been a useful guide in obtaining a reasonable model of a protein–protein complex.

ACKNOWLEDGEMENTS

The authors wish to thank Dr R E Hubbard for making the Hydra program available. Support from the Swedish Natural Science Research Council is gratefully acknowledged. They also acknowledge computing facilities provided by a study contract with IBM Sweden.

REFERENCES

1 **Lee, B and Richards, F M** *J. Mol. Biol.* Vol 55 (1971) pp 379–400
2 **Schrake, A and Rupley, J A** *J. Mol. Biol.* Vol 79 (1973) pp 351–372
3 **Finney, J L** *J. Mol. Biol.* Vol 96 (1975) pp 721–732
4 **Connolly, M L** *J. Appl. Cryst.* Vol 16 (1983) pp 548–558
5 **Connolly, M L** *Biopolymers* Vol 25 (1986) pp 1229–1247
6 **Fanning, D W et al.** *Biopolymers* Vol 25 (1986) pp 863–883
7 **Lewis, M and Rees, D C** *Science* Vol 230 (1985) pp 1163–1165
8 **Pfeiffer, P et al.** *Chem. Phys. Lett.* Vol 113 (1985) pp 535–540
9 **Åqvist, J et al.** (1987) to be published
10 **Eckmann, J-P and Ruelle, D** *Rev. Mod. Phys.* Vol 57 (1985) pp 617–656
11 **Chothia, C** *Nature* Vol 254 (1975) pp 304–308
12 **Jones, T A and Liljas, L** *J. Mol. Biol.* Vol 177 (1984) pp 735–767
13 **Blake, C C F et al.** *J. Mol. Biol.* Vol 121 (1978) pp 339–356
14 **Blake, C C F and Oatley, S J** *Nature* Vol 268 (1977) pp 115–120
15 **Newcomer, M E et al.** *EMBO J.* Vol 3 (1984) pp 1451–1454
16 **Peterson, P A** *Eur. J. Clin. Invest.* Vol 1 (1971) pp 437–444
17 **Sandblom, P et al.** *Biochem. Biophys. Res. Comm.* Vol 139 (1986) pp 564–570
18 **Åqvist, J et al.** *J. Mol. Biol.* Vol 192 (1986) pp 593–604
19 **Hubbard, R E** in 'The representation of protein structure' *Proc. Computer-Aided Molecular Design Conf.* Oyez, UK (18–19 October 1984) pp 99–106
20 **Stapleton, H J et al.** *Phys. Rev. Lett.* Vol 45 (1980) pp 1456–1459
21 **Allen, J P et al.** *Biophys. J.* Vol 38 (1982) pp 299–310
22 **Wagner, G C et al.** *J. Am. Chem. Soc.* Vol 107 (1985) pp 5589–5594
23 **Avnir, D et al.** *Nature* Vol 308 (1984) pp 261–263