J. Chem. Inf. Comput. Sci. 1983, 23, 135-149.
(6) Warr, W. A. Diverse Uses and Future Prospects for Wiswesser Line-Formula Notation. J. Chem. Inf. Comput. Sci. 1982, 22, 98-101.
(7) IUPAC Nomenclature of Organic Chemistry, Sections A-F and H; Pergamon: Oxford, 1979.
(8) Rush, J. E. Status of Notation and Topological Systems and Potential Future Trends. J. Chem. Inf. Comput. Sci. 1976, 16, 202-210.
(9) Wiswesser, W. J. A Linear-Formula Chemical Notation; Crowell: New York, 1954.
(10) Wiswesser, W. J. Historic Development of Chemical Notations. J. Chem. Inf. Comput. Sci. 1985, 25, 258-263.
(11) Communication, Storage and Retrieval of Chemical Information; Ash, J. E., Chubb, P. A., Ward, S. E., Welford, S. M., Willett, P., Eds.; Ellis Horwood: Chichester, 1985.
(12) Huixiao, H.; Xinquan, X. ESSESA: An Expert System for Elucidation of Structures from Spectra. 1. Knowledge Base of Infrared Spectra and Analysis and Interpretation Programs. J. Chem. Inf. Comput. Sci. 1990, 30, 203-210.
(13) Huixiao, H.; Xinquan, X. ESSESA: An Expert System for Structure Elucidation from Spectra. II. A Novel Algorithm of Perception of the Linear Independent Smallest Set of Smallest Rings. Anal. Chim. Acta, in press.

Hausdorff Dimension as a Quantification of Local Roughness of Protein Surfaces

CARL-DIETER ZACHMANN and JÜRGEN BRICKMANN*

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

Received August 27, 1991

Based on structural information from the Brookhaven Protein Data Bank, the contact surfaces of the proteins lysozyme, trypsin, and BPTI (bovine pancreatic trypsin inhibitor) with a spherical test particle of the size of a water molecule have been calculated and systematically analyzed. It is our purpose to establish (i) self-similarity as a statistical concept for the characterization of surface roughness and (ii) the Hausdorff dimension as a measure of the local surface complexity. It is found that the proteins statistically show self-similarity within a yardstick range 1.2 Å < R < 20 Å, and that this concept also holds reasonably for parts of the surface which are not too small.

INTRODUCTION

Up to now, there was no way to predict the three-dimensional structure of a protein with a given sequence of amino acids based only on the knowledge of a molecular force field, which—at least in principle—can be derived from first principles. Additional information from structural investigations (X-ray, NMR) for fragments are necessary for any successful attempt. Consequently, the building principles of protein structures presently can not be completely understood on the basis of first principles (in contrast to small- and medium-sized molecules). A systematic analysis of 3D protein data may help to understand the building principles which have been working during the evolution of these biopolymers. There are several arguments for the statement saying that one of these principles is related to the surface complexity of these molecules.[1-5]

The "surface roughness" (or geometrical complexity) of biological macromolecules influences the transport of substrate molecules from the aqueous bulk phase to the surface as well as the migration on the surface. While by an increase of surface roughness the first process is speeded up, the second is slowed down.[1,2] The roughness of molecular surfaces also plays a role in intermolecular recognition. Recognition is related to geometrical selectivity and to the specificity of point interactions. One molecule can only be positively identified by a second one when two conditions are fulfilled, namely

(i) that the shapes of the molecular repulsion surfaces fit in the contact region

(ii) that specific point links via directed hydrogen bonds can be formed as a consequence of reasonable proton donor-acceptor arrangements

This paper deals with the quantification of "surface roughness". It has been argued[1-5] that the concept of fractals may be helpful for the characterization of the surface complexity of proteins. These surfaces seem to be self-similar within a certain yardstick range. Moreover, it has been suspected[1,2] that the selectivity of a receptor site may be related to local fractality of the protein surface in the receptor-site region.

The purpose of this paper is to re-examine whether protein surfaces really exhibit self-similarity, i.e., whether the concept of fractals holds at least in a statistical sense. If the answer is yes, we shall attempt to use it to quantify the roughness (i.e., geometrical complexity) of such a surface as a local property (a quantity which can be assigned to a local surface area of the protein). The fractal concept can only be used for correlation studies describing the selectivity of an active-site region if it can be shown to be meaningful to describe a local property of the surfaces.

Some work has been done to establish the fractal-surface dimension of proteins. Lewis and Rees[3] calculated the contact surfaces $A(R)$ of some proteins with spherical test molecules (hard spheres) following the algorithm of Connolly[8] and using sphere radii between R = 1.0 Å and R = 3.5 Å. They determined the fractal dimension D from the average gradient in a log (A)/log (R) plot and found D = 2.44, 2.44, and 2.43 for lysozyme, ribonuclease A, and superoxide dismutase, respectively. The curves in their diagrams are, however, not straight lines, as expected for self-similar objects.[5-7] Consequently, from the work of these authors, it seems questionable whether the concept of fractals is applicable in this case, and, in particular, whether the "local" fractal dimensions calculated by the authors are really meaningful. Their conclusion that molecular regions involved in the formation of tight complexes (such as antibody-combining regions) appear to be more irregular (with high-D values) than regions involved in the formation of transient complexes (such as active sites) is open to discussion. Åqvist and Tapia[4] re-investigated the surface fractality as a guide for studying protein-protein interactions. Starting from the same type of surface generation and surface analysis as done by Lewis and Rees,[3] they indeed found linear plots in the log (A)/log (R) diagrams (with standard deviations of the regression coefficient < 0.02) for a range 1.5 Å < R < 3.0 Å. The authors obtained a fractal dimension of D = 2.19 for lysozyme, which is at variance with the results of Lewis and Rees (D = 2.44), but in remarkable agreement with

* Author to whom correspondence should be addressed.

0095-2338/92/1632-0120$03.00/0

the data of Pfeiffer et al. (D = 2.2).[2] The latter authors analyzed two-dimensional cuts through the surface of the lysozyme molecule, as defined by the interaction with a water molecule in its energetically most-favorable orientation, by measuring the length with different yardsticks. They found a line dimension of D_L = 1.2 by using the fact that a section through a fractal surface with a dimension D has a dimension $D_L = D - 1$ with a probability approaching one.[9] The authors established a surface dimension of D = 2.2 for lysozyme within a lower cutoff of the yardstick of 1 Å and a higher cutoff of 20 Å.

Molecular surfaces are not uniquely defined. Consequently, the approaches of Lewis and Rees[3] and Åqvist and Tapia[4] may lead to different surface dimensions if compared to that of Pfeiffer and co-workers.[2] In the first case, the surface is generated by the yardstick particle itself while in the second a surface is generated by a test particle of given size and then analyzed with different yardsticks.

CALCULATIONS

In this article, we give a short outline of our methodology and present results obtained for the proteins lysozyme, trypsin, and BPTI. More detailed information on the algorithms used for the calculations and results for other proteins will be published elsewhere.[10]

In accordance with Pfeiffer et al.[2] we base our considerations here on the contact surface of a protein with a spherical test molecule (hard sphere), which has the van der Waals radius R = 1.4 Å of a water probe. Our surface analysis relies on the fact that the mass $M(R)$ included in a sphere of radius R around an inner point of a self-similar object with fractal dimension D (the Hausdorff mass[6]) scales like

$$M(R) = CR^D \quad (1)$$

with a constant C. The Hausdorff mass approach has the advantage that one can test whether the self-similarity concept is valid within a distance range around a given surface point. It allows a local fractal dimension to be established if a linear relationship between log (M) and log (R) can be found.

Calculation of Hausdorff Mass for Individual Surface Points. First, we triangulated the Connolly surface with approximately 3 points per Å^2 in order to obtain an analytically defined surface.[11] We determined $M(R)$ for 1.2 Å < R < 50 Å for all surface points as reference centers. The R values were chosen such that equidistant values occur in the log (R) scale. In addition to this analysis, we analyzed the area increase as a function of R. Only for a fractal object and for fixed ratio $R^{(k)}/R^{(k+1)}$ of two adjacent values in the R scale, one obtains[5]

$$\Delta M^{(k)} = M[R^{(k+1)}] - M[R^{(k)}] = C'[R^{(k)}]^D \quad (2)$$

with a constant C'. It has been demonstrated by Avnir and co-workers[5] that the latter relation is the more sensitive one for the identification of fractal objects. The results for two selected surface points for a BPTI molecule (bovine pancreatic trypsin inhibitor; coordinates are from Brookhaven Data Bank, file 2PTC.pdb; hydrogen positions are generated according to standard distances and standard angles) are shown in Figure 1. It is seen that there are surface points for which the self-similarity concept holds within large distance ranges from this point (Figure 1a, an area close to residue Leu-6), but there are many others where a linear relationship in the log (M_i)/log (R) diagrams is not found (Figure 1b, close to residue Arg-1). We did not find a systematic criterion for "fractal" and "non-fractal" points. In other words, those points for which the concept is well fulfilled, and those for which this is not the case, are more or less randomly distributed on the surface.

Calculation of Hausdorff Mass for Entire Surfaces. In a second series of calculations, we studied whether the fractal dimension can be well established as a statistical measure for

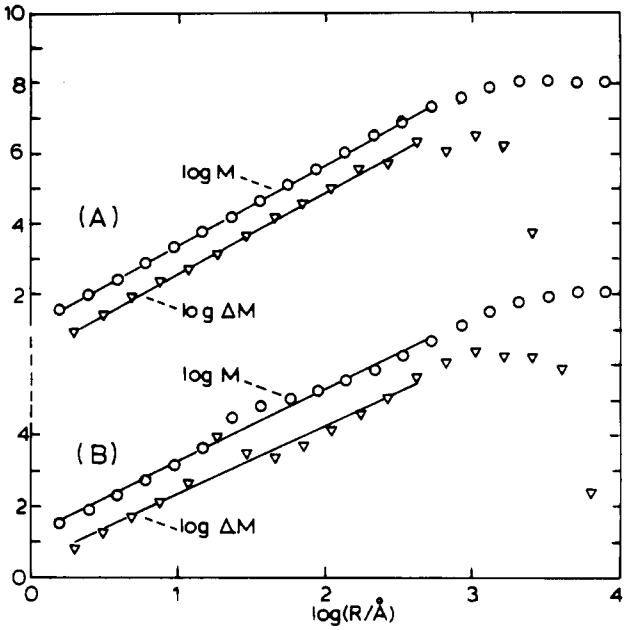

Figure 1. Hausdorff mass $M(R)$ (in arbitrary mass units, see eq 1) as well as mass increase ΔM (see eq 2) as functions of the distance R from selected surface points of the Leu-6 residue (a) and the Arg-1 residue (b) of the bovine pancreatic trypsin inhibitor (BPTI).

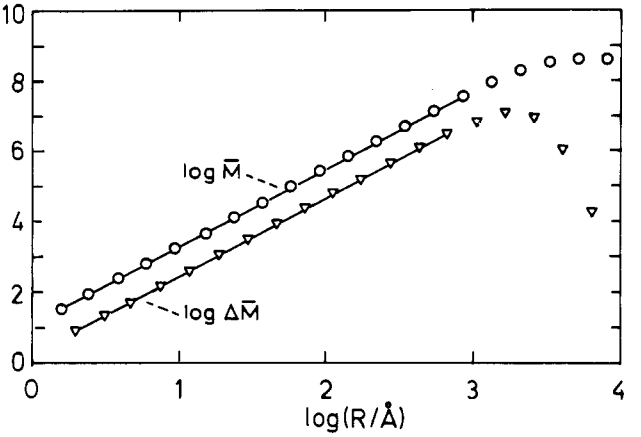

Figure 2. Average Hausdorff mass $\bar{M}$ (see eq 3) and mass increase $\Delta\bar{M}$ (as defined in eq 2) as functions of the distance R from the individual surface points of lysozyme.

the complexity of the *entire* protein surface. For this reason the scaling property of the average Hausdorff mass as a function of the distance from the reference point

$$\bar{M}(R) = \sum_i M_i(R)/N \quad (3)$$

was analyzed. N is the total number of surface points (N is in the order of 10^4). Nearly perfect linear relations resulted now in the log (M)/log (R) plots and also in the more sensitive area increase diagrams (see Figure 2) within a range of 1.2 Å < R < 20 Å. For lysozyme (Brookhaven Protein Data Bank file 7LYZ.pdb) D = 2.204 ± 0.005 (mean error of the linear coefficient) with a correlation coefficient C_c = 0.999971 results from eq 1, while D = 2.218 ± 0.006 (C_c = 0.999953) was obtained from eq 2. Our data are in excellent agreement with the results of Åqvist and Tapia[4] and Pfeiffer et al.[1,2] We also analyzed the trypsin molecule (D = 2.190 ± 0.007 from eq 1 and D = 2.218 ± 0.005 from eq 2) and the BPTI molecule (D = 2.192 ± 0.004 and D = 2.203 ± 0.007, respectively, coordinates for both proteins from Brookhaven Data Bank file 2PTC.pdb). The correlation coefficients for all linear regressions are of the same magnitude as those for lysozyme. It is remarkable that the overall fractal dimension of trypsin

and its natural inhibitor are identical within the statistical errors.

Calculation of Hausdorff Mass for Parts of a Surface. In a third series of calculations the contact area of trypsin with BPTI was analyzed along the lines described above. The average Hausdorff mass $\bar{M}$ (see eq 3) was now determined for all surface points of the contact area (as calculated from the structure of the trypsin–BPTI complex). It turned out that the concept of self-similarity holds reasonably well for the surface area of trypsin centered around residue Ser-190 [$D = 2.226 \pm 0.011$ (eq 1)], $D = 2.259 \pm 0.018$ (eq 2) and for the one of BPTI around residue Lys-15 ($D = 2.116 \pm 0.021$ and $D = 2.045 \pm 0.041$, respectively) but that the fractal dimensions of both areas are different. While the fractality of the binding region of trypsin is higher than the average dimension of this molecule, the corresponding value in BPTI is smaller. From this result it seems to be questionable whether the local fractality of a surface area can be used as a guideline for the identification of receptor sites or preferential locations for protein–protein interactions. This latter statement is supported by the fact that there are surface areas of the BPTI (close to residue Ala-40) which have nearly the same D-value [$D = 2.237 \pm 0.007$ (eq 1)] as the trypsin inhibitor site but are far away from the contact area. The same is true for the trypsin molecule. We found an area (close to residue Ser-110) with the low value $D = 2.144 \pm 0.003$ (from eq 1), which is very close to the fractal dimension of the BPTI-binding region.

CONCLUSION

We have demonstrated that the Hausdorff mass approach is well suited for the systematic study of the surface complexity of proteins. Surfaces of proteins do show self-similarity, which is a global statistical information. This fact is relevant for the study of transport phenomena of substrates from the bulk phase to the surface as well as parallel to the surface.[1,2] The concept of fractals can also be applied to the study of selected surface areas (like the trypsin–BPTI interaction site), but it is only of minor importance for the identification of such sites. The biological key and the lock may have different fractality even when there is a highly specific interaction.

ACKNOWLEDGMENT

We thank J. Schnitter, M. Waldherr-Teschner, and W. Heiden for computational assistance as well as J. Schnitter and Ph. Bopp for carefully reading the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft, Bonn, and the Fonds der Chemischen Industrie, Frankfurt.

Registry No. BPTI, 9087-70-1; lysozyme, 9001-63-2.

REFERENCES AND NOTES

(1) Pfeiffer, P. *Appl. Surf. Science* **1984**, *18*, 146–164.
(2) Pfeiffer, P.; Welz, U.; Wippermann, H. *Chem. Phys. Lett.* **1985**, *113*, 535–540.
(3) Lewis, M.; Rees, D. C. *Science* **1985**, *230*, 1163–1165.
(4) Aqvist, J.; Tapia, O. *J. Mol. Graphics* **1987**, *5*, 30–34.
(5) Farin, D.; Peleg, S.; Yavin, D.; Avnir, D. *Langmuir* **1985**, *1*, 399–407.
(6) Mandelbrot, B. B. *The Fractal Geometry of Nature*; Freeman: San Francisco, 1982.
(7) Feder, J. *Fractals*; Plenum Press: New York, 1988.
(8) Connolly, M. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
(9) Mattila, P. *Ann. Acad. Sci. Fenn.* **1975**, *1*, 227–244.
(10) Zachmann, C.-D.; Brickmann, J., manuscript in preparation.
(11) Heiden, W.; Schlenkrich, M.; Brickmann, J. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 255–269.

COMPUTER SOFTWARE REVIEWS

Spectrochimica Acta Electronica[1,2]

STEPHEN R. HELLER

U.S. Department of Agriculture, ARS, Building 005, Beltsville, Maryland 20705-2350

Received November 10, 1991

Spectrochimica Acta Electronica is an unusual product, a computer-readable supplement to a printed journal, *Spectrochimica Acta*. Pergamon Press is to be commended for this interesting experiment in electronic publishing. Spectrochimica Acta Electronica, otherwise known as Part B of *Spectrochimica Acta*, contains data or programs that are intimately associated with papers published in the printed section of the Journal.

The first issue of this new section, Volume 46B of *Spectrochimica Acta*, appears as 3 disks which are related to two papers in Part A of the Journal. The first paper, entitled Simulation of Atomic Spectra. I. Profiles and Signal-to-Background Ratios of 350 Prominent Lines Emitted from an Inductively-Coupled Plasma or "Whatever" Customized Atomic Emission Lines appears in hard copy with an appendix and also a $3^1/_2$ in. diskette containing a program, data files, and a manual. The printed text details the purpose of the work, with emphasis on its spectroscopic aspects, and the appendix provides the information needed to access the diskette and make use of the program and data. Additional tutorial guidance is in the manual which is on the diskette.

The program concerns primarily simulation of the 350 prominent lines derived from 65 elements in an inductively-coupled plasma as a function of the spectral bandwidth of the spectrometer and the Doppler temperature of the source. A spectrum is displayed as a function of wavelength in terms of (SBR + 1), where SBR is the signal-to-background ratio. A spectral window of 80 ppm is covered. In addition, the spectral data that are displayed encompass, inter alia, the SBR and background equivalent concentration (BEC) in the maxima of the spectral structure, the full width at half-maximum (FWHM) of the peaks, the standard values of SBR and BEC, and the detection limit. The main purpose of the program that is provided is to facilitate the conversion of data on SBRs, BECs, and detection limits from instruments of different spectral bandwidth and thus to allow unbiased comparisons