www.shd.org.rs

Journal of the Serbian Chemical Society

JSCS@tmf.bg.ac.yu • www.shd.org.rs/JSCS

J. Serb. Chem. Soc. 73 (8–9) 805–813 (2008)
JSCS–3763

UDC 577.112:54.02:519.2
Original scientific paper

Structural features of proteins as reflected by statistical scaling laws

LAURA PITULICE, ADRIANA ISVORAN* and ADRIAN CHIRIAC

Department of Chemistry, West University of Timisoara, 16 Pestallozi, 300115, Timisoara, Romania

(Received 31 December 2007, revised 28 March 2008)

Abstract: Within this paper, statistical scaling laws for the radius of gyration with the residues number, the surface area with the probe radii and the backbone length with the interval of residues for a set of 60 proteins are revealed. The proteins belong to three different structural classes: *alpha*, *beta* and *alpha* plus *beta* class (20 proteins for each) according to the SCOP database classification, which takes into account the composition in the elements of their secondary structure. The shape and the surface roughness of proteins seem to be independent of the protein content in the secondary structure elements. On the contrary, the protein packing density shows a strong correlation with this composition.

Keywords: protein; backbone; fractal dimension; radius of gyration; scaling law.

INTRODUCTION

Even though the general nature of the interatomic forces involved in the building the spatial structures of proteins is already known, there is still a necessity for further research to be performed in the field in order to reveal the details of the folding processes of proteins. In this respect, the application of the concepts of fractal geometry would be useful. These concepts have been widely applied for the study of the structural and dynamical properties of proteins.[1–17]

With regards to the structure of proteins, the fractal aspects refer either to the protein surface or to its backbone length. Protein surfaces have been shown to be fractal with different fractal dimensions on the micro and macro scale.[3–11] The fractal properties of 5526 tertiary structures of different proteins were investigated and a mean fractal dimension of their surface of about 2.47 was obtained.[15]

Protein backbones show two fractal dimensions, one corresponds to local folding and the other to global folding, suggesting that different molecular features are responsible for the two different spatial organizations.[2,6,12–14,16,17]

The surface of a protein is generated from the van der Waals radii of the component atoms. The surface area of proteins has been shown to be fractal,

* Corresponding author. E-mail: aisvoran@cbg.uvt.ro
doi: 10.2298/JSC0809805P

having a scaling behavior with the probe radius rolled onto it.[3–11,16,17] To determine the accessible surface (*AS*) of a protein, the ball-rolling model[18] may be applied. It uses a sphere of radius *R*, which is rolled onto the surface of the protein maintaining contact with the van der Waals surface. As a result, the following scaling law is obtained:

$$AS \approx R^{2-D_S} \tag{1}$$

where D_S is the surface fractal dimension that can be determined from the slope of the line of the plot log *AS versus* log *R*.

To reveal the fractal aspects of protein backbone, the concept of backbone length was employed. The length (*L*) of the backbone is determined by connecting step by step the carbon-alpha (C^{α}) atoms of the protein for different sequence intervals and summing these distances:

$$L_m = \sum_{i \neq m}^{j} \sqrt{(x_{i+m} - x_i)^2 + (y_{i+m} - y_i)^2 + (z_{i+m} - z_i)^2} \tag{2}$$

where x_i, y_i and z_i refer to the spatial coordinates of the i^{th} carbon-alpha atom, *j* is the integer part of the ratio *N*/*m* with *N* the total number of amino acids in sequence and *m* is the interval of the residues (Fig. 1). The scaling law is given by Eq. (3):

$$L \approx m^{\frac{1}{D}-1} \tag{3}$$

where *D* is also a non-integer value, named the backbone fractal dimension. Similarly, *D* may be calculated from the plot log *L versus* log *m*.

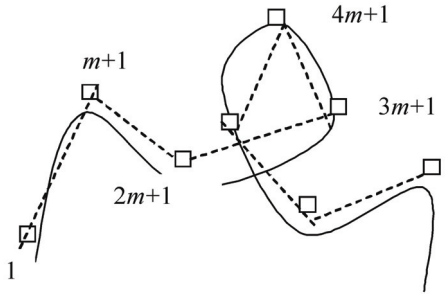

Fig. 1. Schematic representation of the algorithm to determine the backbone length (*L*) of a protein for different intervals of amino acids (*m*) (adapted from Dewey[6]).

Another fractal feature of the structure of a protein may be revealed through the dependence of its radius of gyration on the number of amino acids in the protein sequence. It has been demonstrated that the radius of gyration scales with the residue number for different types of proteins.[1,6–11,17] The radius of gyration, R_g, is defined as the root mean squared distance of the mass constituents of a polymer from the center of its mass. It scales with the residue number (*N*) according to the Eq. (4):

$$R_g \approx N^{\frac{1}{D_f}} \quad (4)$$

where D_f is a non-integer value that may be obtained as the slope of the line of the plot log R_g *versus* log N. The scaling properties of the radius of gyration with the residue number have been demonstrated for globular proteins,[6] proteases and non-proteases,[10] and for *O*-glycosidases.[11]

The above-mentioned scaling laws (Eqs. (1), (3) and (4)) have usually been used to study and describe the shape, surface roughness and packing compactness of different proteins. In the present study, an attempt was made to show if these dependences correlate with the composition in the secondary structural elements of proteins. Thus, the scaling laws for the above-mentioned properties were examined considering 60 proteins belonging to three different structural classes, as given by the SCOP database,[19] *i.e.*, all *alpha* (essentially formed by α-helices), all *beta* (dominated by β-sheets) and *alpha* plus *beta* class (in which α-helices and β-strands are largely segregated).

EXPERIMENTAL

For this study, three unbiased sets of 20 proteins for each considered structural class were randomly chosen. The structural data necessary for the calculations were retrieved from the Protein Data Bank of the RCSB (http://www.rcsb.org/pdb/home/home.do)[20]and the following files were considered:

− 1MOH, 19GS, 1A56, 1A59, 1A7W, 1AEW, 1AH7, 1AIN, 1A6G, 1A7D, 1DAV, 1TZV, 1BKR, 1NOL, 1AD6, 1ALU, 1B1B, 1C20, 1CMZ, 2UTG for all *alpha* class;

− 2AVG, 1A42, 1A18, 1A3K, 1A1X, 1A3Z, 1AG6, 1AMX, 9ILB, 1A21, 1A25, 1A3Y, 1A9V, 1AAC, 1AG4, 1AQB, 1B8E, 4PEP, 43C9, 2SOD for all *beta* class;

− 1A0K, 1A4V, 1A6F, 1A70, 1APS, 1RLO, 1ABA, 4YAS, 16PK, 1A8Q, 1AJZ, 1ARL, 1B0J, 1B7E, 1BAM, 1BCO, 1CEX, 1CJC, 1CP7, 5ULL for *alpha* plus *beta* class.

In order to avoid a reduction of the statistical significance of the results due to the analysis of similar sequences, sequences alignment of these proteins was performed using the Clustalw software accessible on-line.[21] The similarity in the sequences was predicted to be lower than 35 %, which is considered as satisfactory.

The radius of gyration was calculated using the on-line facilities provided by the Protein Dipole Moments Server under the Weizman Institute web page (http://bioportal.weizmann.ac.il/dipol/dipol2.html)[18] and the molecular weights were calculated using the ProtParam tool under the Swiss Prot web page.[22]

The proteins backbone lengths were determined using an in-house Pascal 7.0 program. It uses the spatial coordinates of carbon-alpha atoms as they are given in Protein Data Bank in order to calculate the distances between them for different intervals of amino acids. On summing these distances, the length of the backbone is obtained for every interval, according to Eq. (2).

Finally, the molecular surface was calculated using the Getarea on-line free software[18] (http://www.pauli.utmb.edu/cgi-bin/get_a_form.tcl), for which probe radii of 1.0, 1.2, 1.4, 1.6, 1.8 and 2.0 Å were used.

RESULTS

In order to assess the manner in which the shapes of the investigated proteins from the three distinct structural classes compare with each other, the radius of

gyration of each protein was computed and represented *versus* the residue number in a double logarithmical plot, as shown in Fig. 2.

A linear dependence of log R_g *versus* log N was obtained for each structural class of proteins. The global fractal dimensions obtained by fitting linearly the points according to Eq. (3) are: 1.272±0.087 for the *alpha* class, 1.161±0.074 for the *beta* class and 1.172±0.023 for the *alpha* plus *beta* class. The exponent values obtained in this study are in the same range as those previously reported[6,9–11,16] and indicate a structure of the proteins closer to that of an extended ideal polymer.[6]

The fractal diagrams for the backbone of the studied proteins were obtained according to the algorithm presented by Dewey.[6] As an example, the fractal diagram for the bacteriophage T4 glutaredoxin (code entry 1ABA) is presented in Fig. 3.

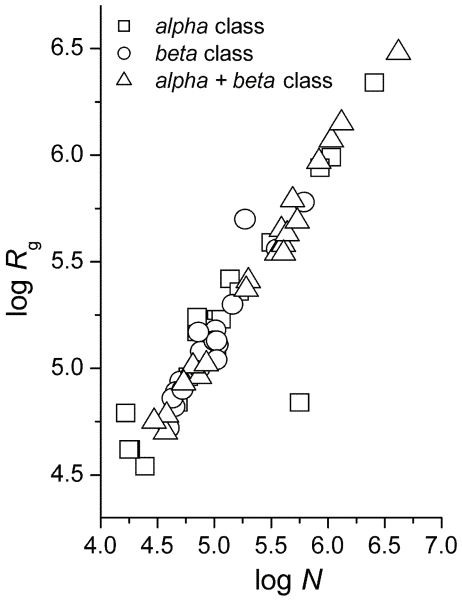

Fig.2. The double logarithmical plot of the radius of gyration *versus* the residue number (□ – *alpha* class; ○ – *beta* class; △ – *alpha* + *beta* class).

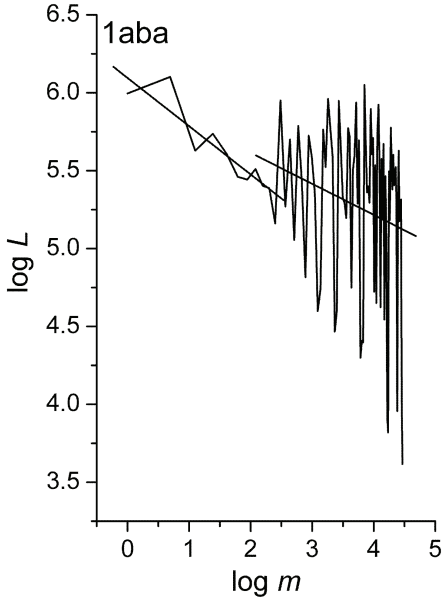

Fig.3. The fractal diagram for bacteriophage T4 glutaredoxin showing the linear fits for the two regions.

Within this plot, two regions can be seen, as reported in the literature for many others proteins.[2,6,12–14,16,17] For each region, an automatic linear fitting was performed. The first region is clearly linear (linear correlation coefficient, R^2, is 0.9516) and is associated with local folding of the protein.[2,6,10–14,16,17] The slope of this line enabled the local fractal dimension (D_l) for the protein backbone to be calculated according to Eq. (2). The values obtained for the studied proteins are presented in Table I. The mean values are: 1.687±0.102 for the

alpha class, 1.367±0.130 for the *beta* class and 1.506±0.119 for the *alpha* plus *beta* class. The Student t-test, implemented under the Origin 7.0 package, showed that these values are significantly different, with only a 0.05 probability that the differences may be caused by chance. The second region of the fractal diagram contains a lot of noise and the errors are high; the R^2 value for the linear fit was 0.4218 in this case. Thus, it was decided not to use this part of the graph in the analysis for accuracy reasons.

Table I also contains the values of the surface fractal dimensions for the investigated proteins, calculated according to Eq. (1). The mean values are: 2.221± ±0.009 for the *alpha* class, 2.184±0.005 for the *beta* class and 2.234±0.004 for the *alpha* plus *beta* class; these values are not significantly different.

As it is to be expected that the protein surface (*AS*) should be dependent on the molecular weight, *M*, of the protein, this relation is presented in Fig. 4 for the three classes of studied proteins.

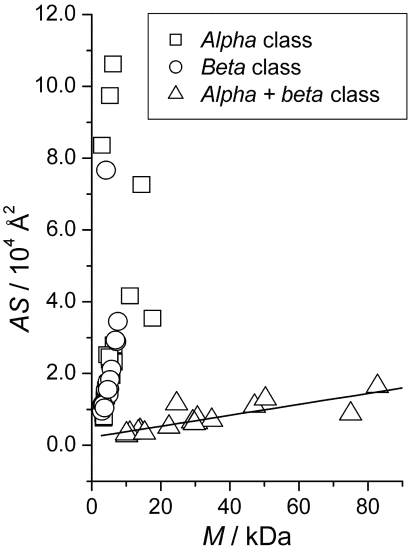

Fig. 4. The dependence of the surface area of proteins on their molecular weight.

From Fig. 4, it can be can observed that for the proteins belonging to *alpha* plus *beta* class, the dependence between the two parameters is linear (R^2 = = 0.8437) and the slope of the line is 0.152±0.024. With regard to the *alpha* and *beta* classes of proteins, linear fitting does not offer an accurate description as some data are scattered around the line. Even though the errors might be high, a slope 0.208±0.039 for the *alpha* class dependence and of 0.186±0.013 for the *beta* class was retrieved for comparison purposes.

TABLE I. The values of the local fractal dimensions and those of the surface fractal dimensions for the investigated proteins

Alpha class			*Beta* class			*Alpha* + *beta* class		
PDB code	D_1	D_S	PDB code	D_1	D_S	PDB code	D_1	D_S
1moh	1.774±0.036	2.253±0.013	2avg	1.430±0.129	2.163±0.019	1a0k	1.392±0.178	2.165±0.013
19gs	1.669±0.028	2.198±0.018	1a42	1.422±0.016	2.180±0.015	1a4v	1.402±0.245	2.169±0.018
1a56	2.107±0.162	2.372±0.024	1a18	1.307±0.014	2.238±0.022	1a6f	1.478±0.207	2.200±0.024
1a59	1.381±0.162	2.233±0.021	1a3k	1.070±0.249	2.157±0.015	1a70	1.580±0.166	2.201±0.016
1a7w	1.674±0.059	2.065±0.017	1a1x	1.092±0.168	2.167±0.011	1aps	1.428±0.139	2.180±0.021
1aew	1.614±0.091	2.207±0.016	1a3z	1.472±0.035	2.101±0.006	1rlo	1.609±0.189	2.291±0.017
1ah7	1.954±0.172	2.218±0.025	1ag6	1.293±0.186	2.071±0.013	1aba	1.448±0.049	2.127±0.025
1ain	1.283±0.139	2.190±0.025	1amx	1.325±0.016	2.118±0.026	4yas	1.489±0.022	2.194±0.025
1a6g	1.889±0.103	2.340±0.005	9ilb	1.597±0.183	2.149±0.023	16pk	1.369±0.168	2.306±0.005
1a7d	1.629±0.044	2.163±0.021	1a21	1.185±0.194	2.167±0.007	1a8q	1.603±0.022	2.284±0.021
1dav	1.380±0.259	2.166±0.013	1a25	1.310±0.019	2.156±0.015	1ajz	1.532±0.020	2.304±0.013
1tzv	1.934±0.075	2.298±0.021	1a3y	1.472±0.155	2.311±0.011	1arl	1.445±0.157	2.301±0.021
1bkr	1.686±0.029	2.124±0.023	1a9v	1.390±0.185	2.352±0.013	1b0j	1.740±0.146	2.350±0.023
1nol	1.588±0.152	2.397±0.014	1aac	1.338±0.028	2.147±0.015	1b7e	1.514±0.031	2.267±0.014
1ad6	1.489±0.173	2.179±0.004	1ag4	1.394±0.145	2.243±0.032	1bam	1.567±0.107	2.250±0.004
1alu	1.966±0.102	2.252±0.012	1aqb	1.393±0.175	2.215±0.010	1bco	1.371±0.163	2.189±0.012
1b1b	1.791±0.053	2.113±0.017	1b8e	1.463±0.130	2.237±0.021	1cex	1.388±0.163	2.154±0.017
1c20	1.462±0.068	2.300±0.007	4pep	1.526±0.138	2.244±0.013	1cjc	1.710±0.161	2.285±0.007
1cmz	1.899±0.101	2.319±0.024	43c9	1.333±0.288	2.086±0.014	1cp7	1.558±0.018	2.236±0.024
2utg	1.572±0.031	2.043±0.012	2sod	1.534±0.147	2.178±0.017	5ull	1.400±0.034	2.226±0.012

DISCUSSION

By analyzing unrelated proteins belonging to three distinct structural classes, consistent structural similarities regarding their shape and surface roughness were noticed. On the contrary, their packing density seems to be a unique feature of each protein class.

Shape and compactness are critical properties characterizing the folding process. On the one hand, the shape of a protein may be quantitatively interpreted by its radius of gyration. In this study, very similar values of the exponent of the scaling law log R_g *versus* log N were obtained for each studied class of proteins (see Fig. 2). On the other hand, the calculated surface fractal dimension (D_S), which measures the surface roughness of a protein, does not differ significantly from one class to another (see Table I). These results suggest a weak dependence of the global folding of these proteins on their composition in the elements of the secondary structure. The rationale of this observation might be that even hard proteins from the same structural class share a similar secondary structure packing, the arrangement of these secondary structural units along the polypeptide chain is different. Moreover, the proteins shape and surface roughness are not dependent on their secondary structure composition. For EF-hand calcium binding proteins, which have the same secondary structure arrangements, the shape and surface roughness are significantly different for those adopting dissimilar spatial conformations: extended and compact, respectively.[17] This indicates that there are other factors responsible for the global folding of a protein, in addition to the composition of the secondary assembly, such as electrostatic and/or hydrophobic interactions. This may bring arguments for the fact that some mutations, which alter the secondary arrangement, do not have important consequences on the function of the protein.

With reference to the protein compactness, its analysis was performed using the local fractal dimension of the backbone of the proteins and the surface area per residue. These parameters should be regarded on a local and global level. The obtained data for the average local fractal dimension of the proteins backbone indicate the highest local packing density for the *alpha* class and the lowest for *beta* class of proteins. This result is not unexpected because a *beta*-strand is the local conformation in which the backbone is extended as much as possible. Also, lattice model studies indicated that compactness could induce polymer chains to develop protein-like secondary structures and lead to a considerable stabilization of them.[22] On the global level, the values of the average surface area per residue show that the highest packing density corresponds to the *alpha* plus *beta* class and the lowest to the *alpha* class. This information is corroborated by the slope of the line $AS = f(M)$ for *alpha* plus *beta* class, the value of which is lower than those corresponding to the other two classes of proteins (see Fig. 4). Thus, the packing compactness seems to be in strong correlation with the composition of the se-

condary structure of the proteins. A similar result was obtained for proteases and non-proteases, which showed the same surface fractal dimensions (2.17) but with clearly different packing densities due to their distinct composition in the elements of their secondary structures.[10]

CONCLUSIONS

In summary, this study supports the fact that statistical scaling laws may be employed in describing structural features of proteins on both the local and global levels. Moreover, it suggests that the protein shape and surface roughness do not depend on the composition of the secondary structure elements but the packing density strongly depends on it. This is thus an appropriate tool to distinguish proteins possessing distinct secondary structures.

ИЗВОД

СТРУКТУРНЕ КАРАКТЕРИСТИКЕ ПРОТЕИНА ДОБИЈЕНЕ НА ОСНОВУ СТАТИСТИЧКИХ РАЗМЕРНИХ ЗАКОНИТОСТИ

LAURA PITULICE, ADRIANA ISVORAN и ADRIAN CHIRIAC

Department of Chemistry, West University of Timisoara, 16 Pestallozi, 300115, Timisoara, Romania

У овом раду су приказане статистичке размерне законитости за везу средњег полупречника са бројем остатака, површине и пробних полупречника и дужине протеинског ланца са интервалом остатака за 60 протеина. Ови протеини припадају трима различитим структурним класама: *алфа*, *бета*, и *алфа* плус *бета* класи (по 20 протеина од сваке класе) према SCOP класификационој бази података, која узима у обзир састав елемената њихове секундарне структуре. Облик и површинска храпавост протеина изгледа да не зависи од састава протеинских секундарних структурних елемената. С друге стране, густина паковања протеина показује изразиту зависност од овог састава.

(Примљено 31. децембра 2007, ревидирано 28. марта 2008)

REFERENCES

1. M. Lewis, D. C. Rees, *Science* **230** (1985) 1163
2. C. X. Wang, F. H. Huang, *Phys. Rev. A* **41** (1990) 7043
3. D. Fushman, *J. Biomol. Struct. Dynam.* **7** (1990) 1333
4. T. Goetze, J. Brickman, *Biophys. J.* **61** (1992) 109
5. A. M. Tamburro, A. De Stradis, L. D'Alessio, *J. Biomol. Struct. Dynam.* **12** (1995) 1161
6. G. T. Dewey, *Fractals in molecular biophysics,* Oxford University Press, New York, 1997
7. A. A. Timchenko, O. V. Galzitskaya, I. N.Serdyuk, *Proteins Struct. Funct. Genet.* **28** (1997) 194
8. R. Fraczkiewicz, W. Braun, *J. Comput. Chem.* **19** (1998) 319
9. M. Daniel, S. Baskar, M. M. Latha, *Phys. Scripta* **60** (1999) 270
10. E. W. Stawiski, A. E. Baucom, S. C. Lohr, L. M. Gregoret, *Proc. Natl. Acad. Sci. U.S.A.* **97** (2000) 3954
11. E. W. Stawisky, Y. Mandel-Gutfreund, A. C. Lowenthal, L. M.Gregoret, *Pacific Symp. Biocomput.* **7** (2002) 637
12. A. Isvoran, A. Licz, L. Unipan, V. V. Morariu, *Chaos, Solitons & Fractals* **12** (2001) 757

13. A. Isvoran, A. Licz, V. V. Morariu, *Rev. Roum. Chim.* **46** (2001) 857
14. A. Isvoran, *Chaos Solitons Fractals* **19** (2004) 141
15. M. A. Moret, J. G. V. Miranda, Jr., E. Nogueira, M. C. Santana, G. F. Zebende, *Phys. Rev. E* **71** (2005) 012901
16. A. Isvoran, L. Pitulice, C. T. Craescu, A. Chiriac, *Chaos Solitons Fractals* **35** (2008) 960
17. L. Pitulice, A. Isvoran, G. T. Craescu, A. Chiriac, *Chaos Solitons Fractals* (2007), doi:10.1016/j.chaos.2007.08.016
18. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **28** (2000) 235
19. A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia, *J. Mol. Biol.* **247** (1995) 536
20. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, *The proteomics protocols handbook*, Humana Press, Hatfield, 2005, p. 571
21. R. Chenna, H. Sugawara, T. Koilee, R. Lopez, T. J. Gibson, D. G. Higgins, J. D. Thompson, *Nucleic Acids Res.* **31** (2003) 3497
22. L. M. Gregoret, *J. Mol. Biol.* **219** (1991) 109.