Journal of Advances in Applied Sciences and Technology

ISSN : 2393-8188 (print)
2393-8196 (online)
www.milliyasrcollege.org.journal.php

SELFSIMILARITY AND SCALE INVARIANCE IN BIOMOLECULES

Yusuf H Shaikh[1] , Kapsae S.K [2]., Wasim Ahmed Hyderi [3], Gulam Rabbani[4]*

[1]Shivaji Arts, Commerce and Science College Kannad.
[2]S.B. Science College Aurangabad

[3]Maharashtra College, Mumbai

[4]Dr. Rafiq Zakaria Campus, Dr. Rafiq Zakaria marg, Rauza Bagh, Aurangabad.

*Corresponding Author:dr_grabbani@gmail.com

ABSTRACT:

Quantification of complex random shapes poses difficulties as they can not be modeled using Euclidian geometry. We used the concept of fractals and fractal geometry for the purpose of quantification of complexity of shape and structure associated with complex shape of large biomolecules. For this purpose two dimensional projection of wire-frame of biomolecules is used and box counting technique is implemented for quantifying the complexity of shape and structure. It is demonstrated that the concept of fractals and fractal dimension can be used for quantification of complex structures like biomolecules of porous materials. Details are presented and findings discussed.

KEYWORDS: Biomolecule, Self-similarity, Scale invariance, Fractal, Fractal Dimension

1. INTRODUCTION:

With the advancement in technology and availability of sophisticated instrumentation and new computational techniques, the most important area of life sciences i.e. the Biotechnology has seen new horizons. Never before techniques have been evolved and deeper insight into phenomena could be realized that revolutionized the field of healthcare and medicine. Newer and potential biological techniques are made available in a variety of forms, this was a result of rapid research in the field of microbiology, bioinformatics etc. These days almost all new techniques rely one way or the other on computers, microcontrollers or advanced electronics. The task of investigation of complex Biomolecules is challenging as these molecules have more than tens of thousands of atoms and thus that many bonds and related complex structure. Large molecules like proteins and enzymes have an involved structure with associated complexity of shape and structure.

Fractal geometry [1 to 3] is a branch of science and technology that proved to be very useful in characterizing complex shapes and structures where other techniques are either insufficient or fail to yield reliable results. We tried to use the concept of fractals and fractal dimension [4, 5] to characterize the complex shapes of Biomolecules using a simple approach employing the two dimensional projection of the shape of Biomolecules.

2.METHODOLOGY:

For the purpose of characterization of complex shapes of Biomolecules we used a

technique of extracting a two dimensional projection of biomolecule. Standard biomolecules and their details of chains and sequences are available in the form of open source data bases at number of places[6 to 8]. These data base of desired molecules can be downloaded and visualized in three dimension using rendering software like arguslab, rasmol, raswin etc [9, 10]. These are open source freeware applications made available by institutions and agencies for public use. Online tools for rendering and some standard data processing are also available [11].

The two dimensional projection of the biomolecule is obtained by rendering the molecule of interest in a suitable software and capturing the view of the molecule from certain angle. These images are stored in an image or picture file, preferably in bitmap file with .BMP extension. These images are then converted into two colour bitmap images that can be used by other software for further analysis. We have developed programs for reading in the two colour bitmap images and converting them into a matrix form that is representative of that image. Box counting algorithm is implemented on this matrix representing the biomolecule. The program used different boxes of different sizes (r) and scans the whole image to find out the number of boxes required (N) to completely cover the image. This process is repeated for different selected sizes of boxes and the results are recorded into a text file in the form of table of r, N, log(r) and log(N). Most of the naturally occurring shapes and patterns exhibit Selfsimilarity and scale invariance[12, 13], almost all the molecules exhibit self similarity and scale invariance over certain range of scale.

To establish scale invariance a graph is plotted using log (N) at the y-axis and log(r) on the axis of x. The presence of Selfsimilarity is ascertained from the resulting graph. If the power law holds and scale invariance is present, the resulting graph is a straight line. Slope of this straight line gives the power law exponent from which the fractal dimension can be found. The power law exponent or the fractal dimension is related to the degree of complexity of structure and texture associated with the pattern under study[14, 15]. A higher fractal dimension implies a higher level of complexity associated with the shape and a lower fractal dimension corresponds to a lower degree of complexity. Thus the fractal analysis of the pattern i.e. two dimensional projection of the image of a complex biomolecule[16, 18] reveals the associated degree of complexity of shape quantitatively and from the fractal dimension or the power law exponent information related to the complexity of shape can be derived.

A couple of typical biomolecules used are shown in the form of ribbon rendering in figure 1 and 2.

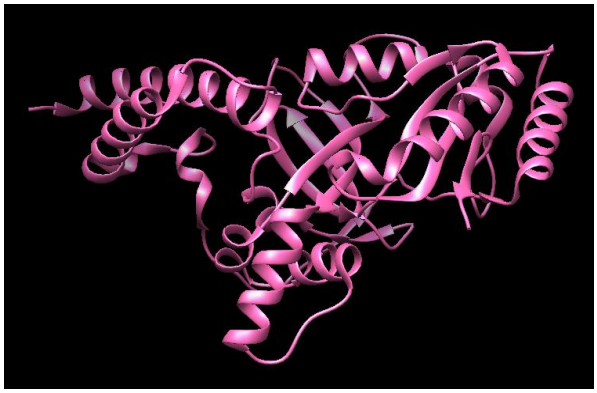

Fig. 1. Enzyme 3BA8 that is a relatively small molecule and is displayed as ribbon.

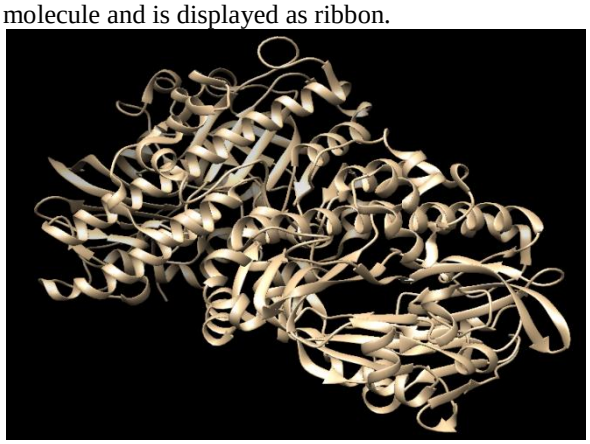

Fig. 2. Enzyme 4HK4 that is a relatively large molecule and is displayed as ribbon.

Comparison of the two molecules in Fig. 1 and 2 clearly indicate that the structure of the molecules is sufficiently complicated and the molecule in Fig. 2 is bigger in size in terms of physical size and number of atoms contained in. The ribbons themselves have lot of atoms in them and organized in a characteristic fashion to impart complexity to the shape of the structure. To have a feel of the complexity of structure of the molecules the two molecules rendered in the form of wire frame are shown in Fig. 3 and 4.

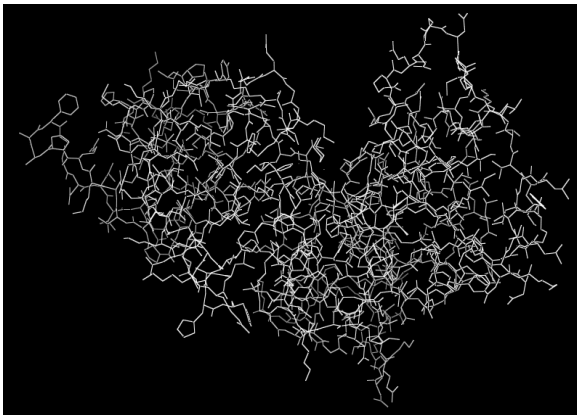

Fig. 3. Enzyme 3BA8 showing wire frame for comparison with fig. 4.

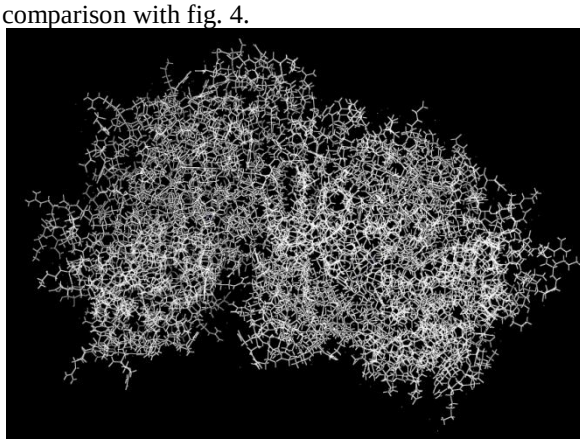

Fig. 4. Enzyme 4HK4 showing wire frame for comparison with fig. 3.

A better comparison can be had viewing the rendering of the same molecules in the form of ball and stick as shown in Figure 5 and 6. Fig. 5 being a small molecule is displayed as more of open structure in magnified form to fill the screen whereas that in Fig. 6 is shown on a reduced scale. For actual comparison either fig. 6 is to be magnified or Fig 5 is to be compressed.

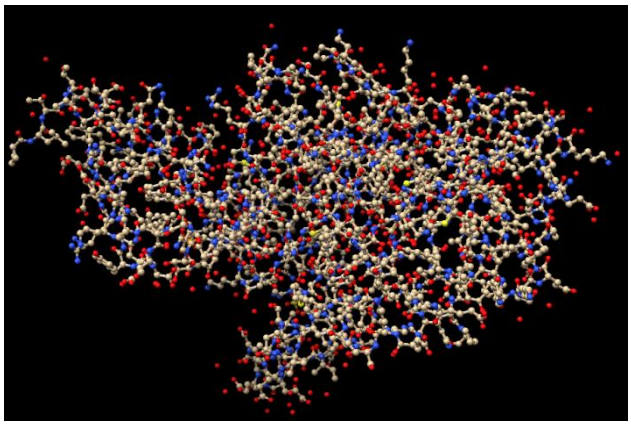

Fig. 5. Enzyme 3BA8 showing Ball and Stick for comparison with fig. 4.

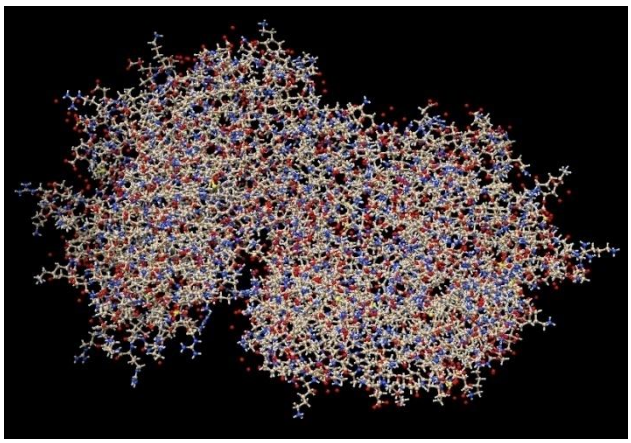

Fig. 6. Enzyme 4HK4 showing ball and stick for comparison with fig. 5.

For the purpose of fractal analysis we implemented box-counting technique using the wire frames of the selected molecules that served as a two dimensional projection of the image of the skeleton of the molecule. This technique makes use of different sizes (r) of square boxes and scan the entire image to determine the total number of boxes (N) required covering the image. Computer programs are developed for the implementation of the box-counting technique, the program performs the entire work using different box sizes (r) and finally saves the results in a file in tabular form for further processing. The results of box-counting for the two images shown in Fig. 3 and 4 for Enzyme molecules 3BA8 and 4HK4 are presented in Table – 1. A comparison of wire frame structure shown in Fig. 3 and 4 suggests that the two figures are on different scale, for visual comparison scaling of the

two patterns to same size would be more realistic. However in the present case, these complex patterns happen to be fractal in nature and thus possess scale invariance and self similarity. The presence of scale invariance provides the flexibility and freedom to use the images at a suitable (available) size thus it is not necessary to scale up or down the patterns to come to the same scale.

Table – 1 : Results of box-counting for images from 3BA8 and 4KH4.

3BA8				4HK4			
r	N	Log(r)	log(N)	r	N	Log(r	log(N)
1	16447	0.000	4.216	1	18460	0.000	4.266
2	8421	0.301	3.925	2	9954	0.301	3.998
3	5062	0.477	3.704	3	6316	0.477	3.800
4	3394	0.602	3.531	4	4431	0.602	3.647
5	2405	0.699	3.381	5	3299	0.699	3.518
6	1813	0.778	3.258	6	2556	0.778	3.408
7	1406	0.845	3.148	7	2042	0.845	3.310
8	1129	0.903	3.053	8	1679	0.903	3.225
9	925	0.954	2.966	9	1387	0.954	3.142
11	645	1.041	2.810	11	1008	1.041	3.004
13	487	1.114	2.688	13	763	1.114	2.883
15	373	1.176	2.572	15	592	1.176	2.772
17	299	1.230	2.476	17	479	1.230	2.680
20	220	1.301	2.342	20	353	1.301	2.548
23	174	1.362	2.241	23	281	1.362	2.449
26	142	1.415	2.152	26	226	1.415	2.354
30	109	1.477	2.037	30	173	1.477	2.238
34	88	1.532	1.945	34	141	1.532	2.149
39	69	1.591	1.839	39	110	1.591	2.041
44	56	1.644	1.748	44	88	1.644	1.945
50	45	1.699	1.653	50	74	1.699	1.869
57	40	1.756	1.602	57	58	1.756	1.763
64	31	1.806	1.491	64	49	1.806	1.690
72	26	1.857	1.415	72	42	1.857	1.623
81	23	1.909	1.362	81	33	1.909	1.519
91	19	1.959	1.279	91	27	1.959	1.431
103	16	2.013	1.204	103	23	2.013	1.362
116	14	2.065	1.146	116	18	2.065	1.255
131	10	2.117	1.000	131	17	2.117	1.230
147	9	2.167	0.954	147	16	2.167	1.204
165	8	2.218	0.903	165	12	2.218	1.079
186	6	2.270	0.778	186	9	2.270	0.954
209	7	2.320	0.845	209	9	2.320	0.954

The table shows the size of box (r) in pixels in the first column and the number of such boxes required to cover the entire image (N) in the second column. Log(r) and log(N) are tabulated in the next two columns i.e. column 3 and 4, similarly the information for the next molecule 4HK4 is tabulated in the next consecutive columns. A graph is plotted using log(N) on the y-axis and log(r) on the axis of x. This plot shows if there is scale invariance and self similarity in the pattern. The presence of self similarity and scale invariance indicated by the line joining these point. If the points lie along a straight line, power law is obeyed and Selfsimilarity and scale invariance is present. Log(N) versus log(r) plots for Fig. 3 and 4 are shown in Fig. 7 and 8.

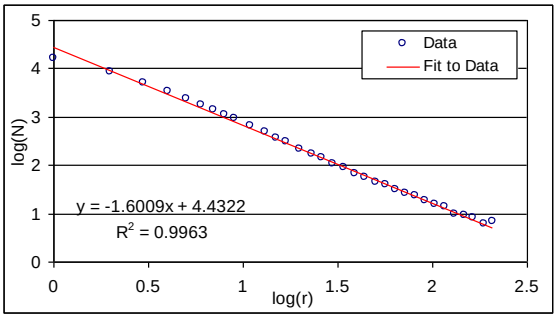

Fig. 7 Log(N) versus log(r) plot for pattern shown in Fig. 3.

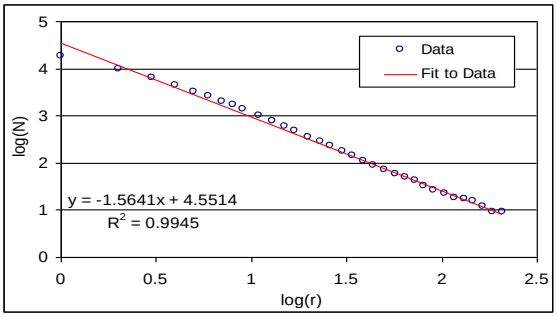

Fig. 8 Log(N) versus log(r) plot for pattern shown in Fig. 4.

Fig. 7 and 8 are the log(N) versus log(r) plots for Fig. 3 and 4, the data plotted is shown in Table – 1, the points plotted are the actual data from Table – 1 and the straight line joining these point is the least square fit straight lines joining these points. It is clearly seen from both the plots that all the data points lie along a straight line which is also seen from the value of R^2 close to unity. The equation of

the straight line best fitting the data points and the resulting value of R2 are also shown on the graph in the inset. From the equation of the straight line, the slope can be found which is the power law exponent showing scale invariance.

Fractal dimension [ ]obtained from the slope of the straight line for molecule 3BA8 is 1.601 and that for 4HK4 is 1.564. A higher fractal dimension indicated higher degree of complexity with the pattern, this implies that the complexity of shape associated with the patter for 3BA8 is more than that of 4HK4. Apparently Fig. 4 and 6 corresponding to 4HK4 looks more crowded with complex structure as compared to that of Fig. 3 and 5 but in reality pattern in Fig 3 for 3BA8 is little more complex as compared to that in Fig. 4 for 4HK4.

3. CONCLUSION

Characterization of complex structures like those of porous materials and complex molecules in terms of morphology poses difficulties. We used a simple technique base on the concept of fractals and fractal dimensions for characterization of complex shapes in a quantitative manner. Box counting technique is implemented on two dimensional projection of two biomolecules with lot of complexity of structure and successfully demonstrated that the technique is capable of quantifying complexity of structure in such cases. Molecule shown in Fig. 5 appears to be a loose structure as compared to that shown in Fig. 6 because the first one is relatively small and shown on a different scale. The complexity of structure associated with the molecule of Fig 5 (3BA8) is high as compared to that in Fig. 6 (4HK4) and this is seen from their log(N) versus log(r) plots shown in Fig. 7 and 8. The fractal dimsnsions estimated from the power law exponent for the two molecules are 1.6009 and 1.5641. A higher value of fractal dimension represents a higher degree of complexity of shape and structure.

REFERENCES:

[1] Benoit B. Mandelbrot, The Fractal Geometry of Nature, W. H. Freeman and Company New York (1983)

[2] B.B.Mandelbrot, Fractal geometry:what does it do ? Proc. R.Soc. Lond. A 423, 3 -16 (1989)

[3] Paul S. Addison, Fractals and Chaos: An Illustrated Course Published by The Institute of Physics (IOP), London Ltd 1997

[4] Walter Beneson, John W.Harris, Horst Stocker, Holger Lutz, Hand book of Physics, published by Springer Verlag,Germany (2000)

[5] Kenneth Falconer, 'Fractal Geometry: Mathematical Foundations and Application, John Wiley & Sons, Ltd United Kingdom(2003)

[6]http://www.academia.edu/1114817/Biomolecules_Introduction_Structure_and_Functions_-_Carbohydrates

[7] Biomolecules 2014, 4, 56-75; doi:10.3390/biom4010056

[8] Brian Maher, Andreas A. Albrecht, Martin Loomes, Xin-She Yang and Kathleen Steinh, biomolecules, 'A Firefly-Inspired Method for Protein Structure Prediction in Lattice Models':www.mdpi.com/journal/biomolecules

[9] Software Arguslab : http://arguslab.en.softonic.com

[10]Software Rasmol: http://www.chemistryteaching.com/rasmol.htm

[11] Standard data processing tool: http://www.ehow.com/facts_5030382_methods-data-processing.html

[12] Benoit B. Mandelbrot, *Science*, 637 (**1967**).

[13] Witten T. & Sander L. M., *Phys Rev lett*. 47, 1400 (**1981**)

[14] Sander LM, *Scientific American*, 256, 94(**1987**)

[15] Morency C, Chapleau R. *Harmonic and Fractal Image Analysis* **2003**;30.

[16] Krishna Gopalakrishnan; Najarian, K., "Prediction of protein function using signal processing
of biochemical properties", Bioinformatics Conference, **2003**. CSB **2003**. Proceedings of the
2003 IEEE Volume, Issue, 11-14 Aug. **2003** Page(s): 536 – 538

[17] Armin B, Shlomo H. 'Fractals in science' _. Berlin Heidelberg: Springer-Verlag; **1994**.

[18] Nazneen Akhter, Yusuf Talib, Shaikh Yusuf H, **'**Fractal application for characterisation of structure of macromolecules like carbohydrate'J. Microbiol. Biotech. Res., 2012, 2 (1):108-114