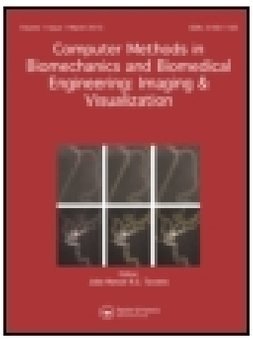


Taylor & Francis
Taylor & Francis Group

Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization

ISSN: 2168-1163 (Print) 2168-1171 (Online) Journal homepage: http://www.tandfonline.com/loi/tciv20

Fractal and mathematical morphology in intricate comparison between tertiary protein structures

Ranjeet Kumar Rout, Pabitra Pal Choudhury, Santi Prasad Maity, B. S. Daya Sagar & Sk. Sarif Hassan

To cite this article: Ranjeet Kumar Rout, Pabitra Pal Choudhury, Santi Prasad Maity, B. S. Daya Sagar & Sk. Sarif Hassan (2016): Fractal and mathematical morphology in intricate comparison between tertiary protein structures, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, DOI: 10.1080/21681163.2016.1214850

To link to this article: http://dx.doi.org/10.1080/21681163.2016.1214850

Published online: 07 Nov 2016.

COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING: IMAGING & VISUALIZATION, 2016
http://dx.doi.org/10.1080/21681163.2016.1214850

Taylor & Francis
Taylor & Francis Group

Fractal and mathematical morphology in intricate comparison between tertiary protein structures

Ranjeet Kumar Rout[a], Pabitra Pal Choudhury[b], Santi Prasad Maity[c], B. S. Daya Sagar[d] and Sk. Sarif Hassan[e]

[a]Department of Computer Science and Engineering, National Institute of Technology, Jalandhar, India; [b]Applied Statistics Unit, Indian Statistical Institute(ISI), Kolkata, India; [c]Information Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India; [d]Systems Science and Informatics Unit, Indian Statistical Institute (ISI), Bangalore, India; [e]Department of Mathematics, College of Engineering Studies, University of Petroleum and Energy Studies, Dehradun, India

ABSTRACT

Intricate comparison between two given tertiary structures of proteins is as important as the comparison of their functions. In literature, several algorithms have been devised to compute the similarity and dissimilarity among protein structures. But, these algorithms compare protein structures by structural alignment of the protein backbones which are usually unable to determine precise differences. In this paper, an attempt has been made to compute the similarities and dissimilarities among tertiary protein structures using the fundamental mathematical morphology operations and fractal geometry which can decipher the intricate differences. In doing so, two techniques have been used here in determining the global as well as local similarity in atomic level of the protein molecules. This intricate structural difference would provide an insight to biologists to understand the protein structures and their functions in more precise fashion.

ARTICLE HISTORY
Received 11 August 2015
Accepted 16 July 2016

KEYWORDS
Tertiary protein structure; geodesic dilation; skeleton and fractal dimension

1. Introduction

Proteins are made of amino acids chain with its length typically ranging from 50 to more than 3000. A carbon atom (called C) is connected to a carboxyl (–COOH) group, an amine ($–NH_2$) group, a hydrogen atom and a residue (which depends on the specific amino acid) to formulate a single amino acid (Anfinsen 1973; Alberts et al. 1983; Branden & Tooze 1991; Chiu 2002; Petsko & Ringe 2004). The amine group of an amino acid is covalently bonded by polypeptide bond with the carboxyl group of another amino acid to form a protein. The sequence of carbon atoms forms the backbone of the protein. Whenever the protein is left in its natural environment, it folds to a specific three-dimensional structure. This is due to the forces between the amino acids such that the total free energy is minimised (Lancia & Istrail 2003). This renders a stable three-dimensional protein structure. Thus, a protein can either be considered as polypeptides sequence of 20 amino acids occurring naturally which popularly known as primary structure of protein or as a tertiary structure into which a particular protein folds (Chew 2006). It is a known fact that the two proteins having more than 30% amino acid sequence identity would not necessarily mean protein functional similarity and vice versa (Chew 2006; Galgonek et al. 2011). Therefore, it is evident that the comparison of protein functionality requires the intricate comparison of their tertiary structures. The search for an effective solution for tertiary protein structural similarity is justified because such tools can be of aid to scientists for prediction of the functions of a hypothetical protein, which would enable to drug design and to understand evolutionary network among proteins (Koehl 2001; Krasnogor & Pelta 2003). In the literature, as in many cases, there is not even a single superposition that reveals all regions of similarity between compared proteins (root mean square deviation – RMSD, DALI, ProSup) (Zemla 2003). Also, there are many conceptual difficulties associated with various methods (RMSD, ad hoc scores based on local secondary structure, hydrogen bonding pattern, burial status or interaction environment), which have not been resolved (Goldman et al. 1999). Classical criteria such as the RMSD fail to identify similar shapes in a consistent way (Guyon & Tuffry 2010). To add on, various systems have been proposed for structural classification, such as structural classification of proteins (Murzin et al. 1995), class architecture topology homology (Orengo et al. 1997), Families of Structurally Similar Proteins (FSSP) (Holm & Sander 1993), and others. The similarity scores, in their cases are computed using structural alignment algorithms such as DALI, CE, VAST, SSAP (Holm et al. 1992; Holm & Sander 1993, 1996; Gibrat et al. 1996; Shindyalov & Bourne 1998) and others. Most of these methods are computationally intensive and time consuming, especially when searching large databases, due to intrinsic complexity of structural alignment (Choi et al. 2004). Also, the prevailing practice in the protein crystallographic community for computing structural differences is highly inappropriate, in particular when

CONTACT Ranjeet Kumar Rout ranjeetkumarrout@gmail.com

the medium- and the low-resolution structures are involved (Kleywegt 1999). Geometrical feature like fractal dimension of C of the backbone structure of one peptide chain proteins are considered in (Cui et al. 2005). Petros (Daras et al. 2006) applied the spherical trace transform to a protein shape to produce a rotation invariant shape descriptor. In order to address the problem of structure reconstruction, a method has been developed Contact Map Reconstruction (COMAR) (Vassura et al. 2008). In the approach given in (Cui et al. 2004), the interaction between amino acids determines the final structure of protein. So, proteins have an intrinsic self-similarity in the compactness and the packing of their structure. In (Cui et al. 2004, 2005), the fractal dimension of the protein backbone structure has been used as a single feature in the structural level, which cannot be a robust representative of its geometric features, the geodesic distance between the whole structures in different faces and fractal dimension in atomic level have been introduced as different features. There are huge numbers of algorithms, which extract some features from the alternative representatives of the polypeptide chain such as three-dimensional Spline (Can & Wang 2003; Bhattacharya et al. 2004). In this paper, these problems have been tried to resolve by introducing two different methods based on two approaches on using Fractal Geometry (fractal dimension) and Mathematical Morphology (geodesic dilation) which yield the desired output.

The organisation of the paper is as follows: In Section 2, the basic review of mathematical morphology operations; in Section 3, data specification and source of data has been discussed; in Section 4, the result and analysis based on the two methods are discussed and conclusion is drawn in Section 5.

2. Basics of mathematical morphology and fractal dimension

Mathematical morphology is a widely used paradigm in the field of image processing (Serra 1982; Matheron & Serra 2004). Morphological tools are popular for image segmentation, image decomposition. Morphological operations include erosion, dilation, opening, closing are used for processing images. Let A be a slice of protein structure decomposed from tertiary protein structure and S be the structuring element with certain characteristic information. In this technical note, each slice is subject to use some morphological transformation. The dots "." in each slice are in black colour pixels (A) and the background is with white pixels (A^c). The definitions of these basic morphological operators are as follows (Teo & Sagar 2006).

Erosion: Erosion transformation (1) of the slice A by S is defined as:

$$A \ominus S = \{a - s : a \in A, s \in S\} = \bigcap_{s \in S} A_s \tag{1}$$

Dilation: Dilation transformation (2) of the atoms in the slice A by S increases the size of atoms.

$$A \oplus S = \{a + s : a \in A, s \in S\} = \bigcup_{s \in S} A_s \tag{2}$$

Opening: Opening transformation (3) of the slice A by S is shown.

$$A \circ S = (A \ominus S) \oplus S \tag{3}$$

Multiscale opening: Multiscale morphological opening (4) can be performed by increasing the size of S. Where the nth size of S is defined as:

$$\underbrace{S \oplus S \oplus \ldots \oplus S}_{n\text{- times}} = nS$$

$$A \circ nS = (A \ominus nS) \oplus nS \tag{4}$$

$A \ominus nS$, $A \oplus nS$ and $A \circ nS$ denotes morphological erosion, dilation and opening by S of size n.

2.1. Morphological skeleton

Morphological skeleton of every geometrical structure is a subset of the original structure which has the same connectivity as the original structure from which inference can be drawn. From each point of the skeleton, the distance to the boundary of the original set is the radius of a maximal circle (whose centre is at a point of the skeleton) which touches the boundary at least two different points. The skeleton of an object gives a clear idea about the shape of the object. For a geometric shape A, and the structuring element S, the skeleton (Maragos & Schafer Ronald 1986; Avnir et al. 1998; Klette 2002) can be constructed through the operation as shown in (5).

$$SK_n(A) = (A \ominus nS) \backslash (A \ominus nS) \circ S, \text{ for } n = 1, 2, \ldots, N \tag{5}$$

2.2. Geodesic dilation

Geodesic dilation is a morphological transformation to operate only some part of the image (as marker) to grow until the boundary of the image on which it is applied (Sagar et al. 2000; Sagar & Tien 2004; Radhakrishnan et al. 2004). This transformation is used to find out similarity between two 2D images. The advantage of this transformation is that the structuring element can vary at each pixel, according to the image. The geodesic dilation δ_x of an image Y inside X is defined as the intersection of the dilation of Y (with respect to a structuring element S) with the image X (6).

$$\delta_x^n = (Y \oplus nS) \cap X \text{ where } n = 1, 2, \ldots, N \tag{6}$$

So geodesic dilation terminates when all the connected components of X are constructed i.e. convergence is reached and defined in (7):

$$\forall n > n_0, \delta_x^{(n)}(Y) = \delta_x^{(n_0)}(Y) \tag{7}$$

2.3. Fractal dimension

A fractal dimension is an index for characterising fractal patterns or sets. The patterns illustrate self-similarity and the fractal dimension indicates the extent to which the fractal objects fill a particular Euclidean space in which it is embedded. The most commonly used technique in determining the fractal dimension is Box Counting Method which is briefly stated as follows.

2.3.1. Box-counting method

This method computes the number of cells required to entirely cover an object, with grids of cells of varying size. Practically, this is performed by superimposing regular grids over an object and

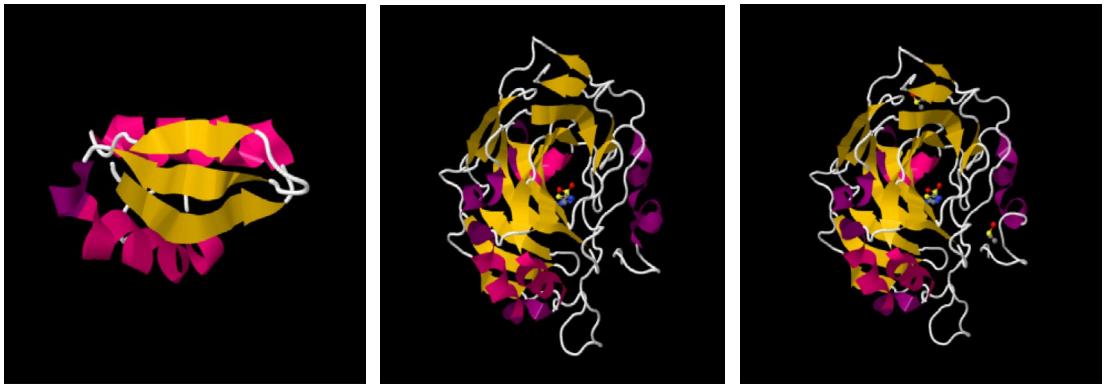


Figure 1. Tertiary structure of three proteins (a) for 2LEP, (b) for 3V2J and (c) for 3V2M.

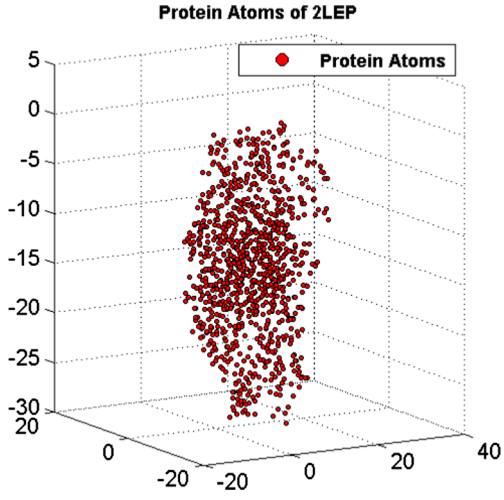


Figure 2. Atoms of protein 2LEP with respect to *z*-axis.

by counting the number of occupied cells. The logarithm of $N(r)$, the number of occupied cells, versus the logarithm of $\frac{1}{r}$, where r is the size of one cell, gives a line whose gradient corresponds to the box dimension (Avnir et al. 1998; Develi & Babadagli 1998). To calculate the dimension for a fractal Sk_n, the Box-Counting dimension is defined as in (8).

$$\mathrm{Dim}_{\mathrm{box}}(Sk_n) = \lim_{r\to 0} \frac{\log N(r)}{\log \frac{1}{r}} \tag{8}$$

3. Data used and specification

The Protein Data Bank PDB (Berman et al. 2000) (http://www.rcsb.org/pdb/home/home.do) is the largest and most commonly used repository; from which information regarding proteins is utilised. The most popular techniques in obtaining protein tertiary structure include the X-ray crystallography and nuclear magnetic resonance (Robillard et al. 1976; Billeter 1992; Drenth 1999). From the PDB (Berman et al. 2000) database, three proteins viz. 2LEP, 3V2J and 3V2M in the standard .pdb format are collected and are shown in Figure 1. The tertiary structure is represented in a standard Cartesian coordinates of the atoms presented in the protein 2LEP is shown in Figure 2.

In the present work, the tertiary structure of protein is decomposed, with reference to *z*-axis, into non-overlapping slices. The methodology of the slice decomposition is explained in Section 4. Some of the slices of protein 2LEP are depicted in Figure 3.

Each of the tertiary structure of the protein in Figure 1(b) and (c) has been converted into backbone structure by *JMol* viewer (http://www.rcsb.org/pdb/explore/jmol.do). The backbone structures of the two protein structures 3V2J and 3V2M (Figure 1(b) and (c)) are shown in Figure 4(a) and (b).

The three tertiary protein structures shown in Figure 1(a)–(c) are consequently used to demonstrate the two approaches, respectively, based on fractal dimension and geodesic dilation. First approach is demonstrated on the thirty slices decomposed from 2LEP (Figure 2) are shown in Figure 3. The 3V2J and 3V2M protein structure and their corresponding backbone structures (Figure4(a) and (b)) were used to demonstrate the approach based on geodesic dilation.

4. The intended methods

In this section, two different methods are proposed to compare between two tertiary structures in intricate level on the basis of mathematical morphology and fractal geometry. These methods, which are explained in this section, have been discussed for the sake of clarity on three tertiary protein structure acquired from PDB database (Berman et al. 2000).

4.1. Tertiary structure skeleton and its fractal dimension

The details of method are described as follows:

Step 1. The slices (non-overlapping) of the tertiary structure of protein f are obtained by slicing the tertiary structure along z-axis on its m ordinates with m planes $P_1, P_2, P_3, \ldots, P_m$ that are perpendicular to the z-axis. These slices are represented as different planner functions and are denoted as $X^1, X^2, \ldots, X^m$ such that $\bigcup_{i=1}^{m} X^i = f$, where f contains almost all atoms of the tertiary protein structures. Each slice X^i contains the protein atoms that share the same z-coordinate. The slices consisting of atoms of the tertiary structure of 2LEP (Figure 2) is represented

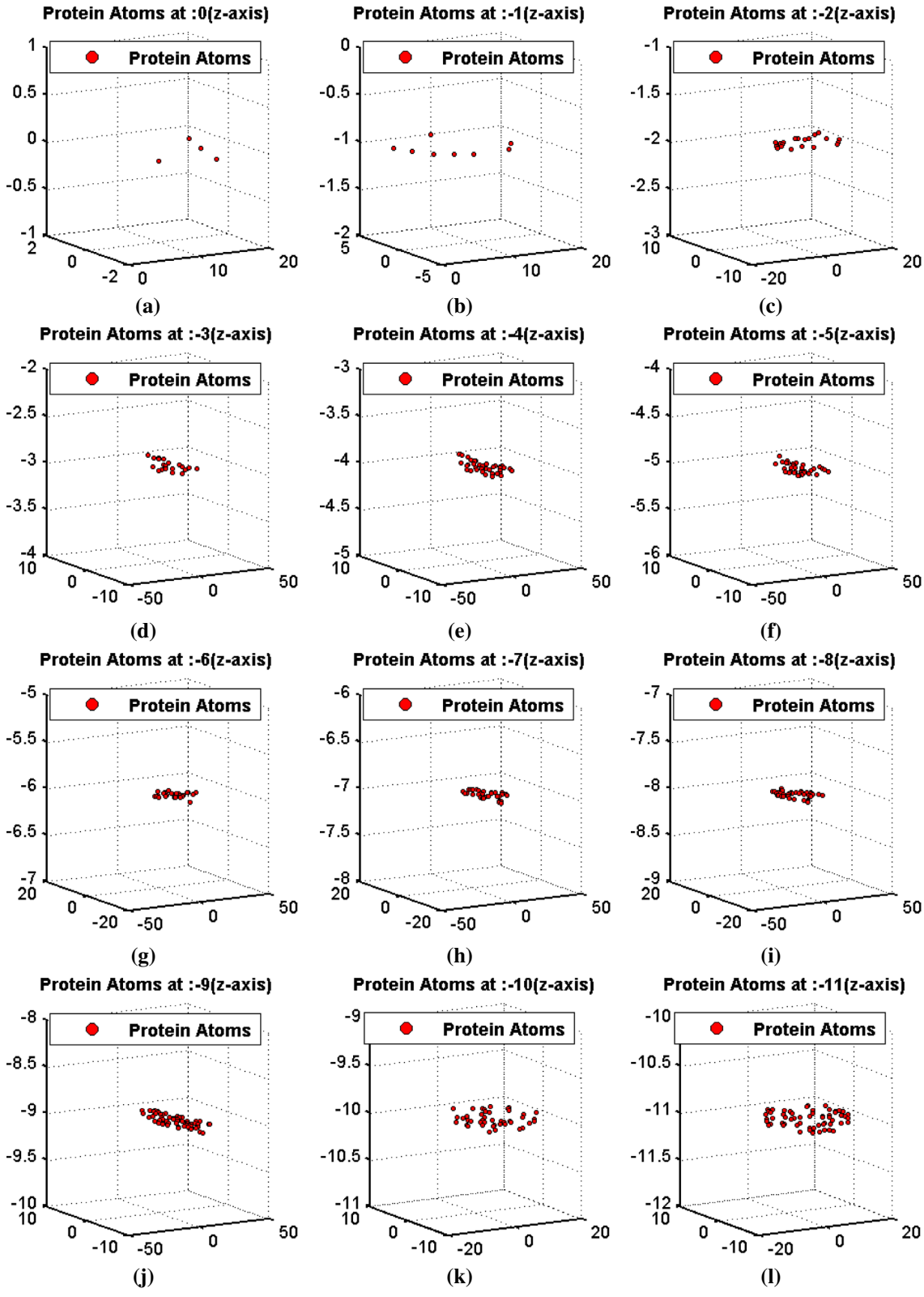
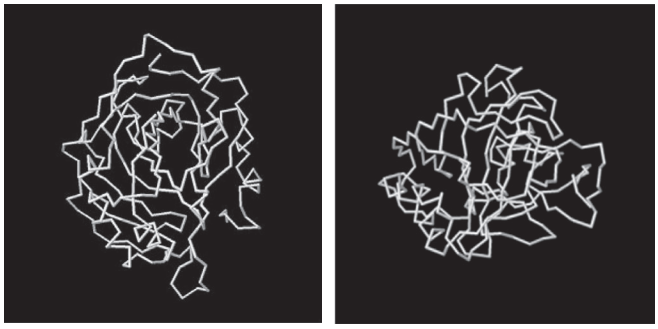


Figure 3. (a)–(l) Atoms of protein 2LEP at different *z*-axis.

by f. The function f could be decomposed with thirty slices like $X^0, X^1, X^2, \ldots, X^{29}$,and a few slices are shown in Figure 3.

Step 2. The connected component of X^i obtained by opening the X^i with respect to the structuring element S of size $n = 1, 2, \ldots, N$ is denoted for simplicity as X_c^i. The slice X^7 contains all the atoms whose z coordinate is at −0.700 is shown in (Figure 3(h)). The 2D representation of the slice X^7 is shown in Figure 5. By performing the multi-scale opening on X^7 (Figure 5) slice until the



(a) (b)

Figure 4. 3D backbone structures of proteins (a) for 3V2J and (b) for 3V2M.

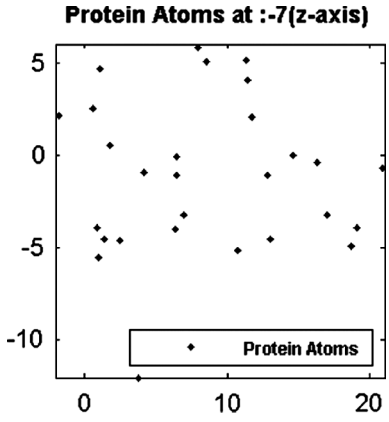


Figure 5. Slice of protein 2LEP at *z*-axis = −7.

*N*th (Figure 6(b)–(k)) cycle yields a single connected component, such that all the atoms are contained by the connected component. Now,

$$X_c^7 = X^7 \, o \, nS$$

is the connected component of X^7 as shown in Figure 6(k).

Step 3. The morphological skeleton $Sk_n(X_c^i)$ of the X_c^i connected component is

$$Sk_n(X_c^i) = (X_c^i \ominus nS) \setminus (X_c^i \ominus nS) \, o \, S,$$

For $n = 0, 1, 2, \ldots, N$ and $i = 1, 2, \ldots, m$ is obtained. Now, the morphological skeleton $Sk_n(X_c^7)$ of the X_c^7 (Figure 6(k)) connected component is shown in (Figure 6(l)) and defined as:

$$Sk_n(X_c^7) = (X_c^7 \ominus nS) \setminus (X_c^7 \ominus nS) oS,$$

For $n = 0, 1, 2, \ldots, N$ and $i = 1, 2, \ldots, m$ is obtained.

Step 4. Similar approach has been followed for generating connected components for rest of the slices decomposed from 2LEP protein structure. The skeleton networks of these connected components of all the slices are generated and have been stacked as shown in Figure 7.

$$Sk = \bigcup_{i=1}^{m} X_c^i, \text{where } i = 1, 2, 3, \ldots, m.$$

Step 5. The fractal dimension D_p of the Skeleton Sk is determined by the Box counting method.

Step 6. In comparison between two different protein structures i and j, the difference operator $\rho = |D_p(i) - D_p(j)|$ could be used. This difference would describe the intricate structural difference.

Similar approach has been followed to compute the fractal dimension D_p of the skeleton of the corresponding tertiary protein structure and use it as the feature of tertiary protein structure. Fractal Dimension D_p of a group of protein molecules are given in Table 1 irrespective of their residue.

4.2. Geodesic dilation and its quantification

Unlike the existing algorithms, (Koehl 2001; Lancia & Istrail 2003; Choi et al. 2004; Cui et al. 2004, 2005; Galgonek et al. 2011) this method consider tertiary structure of proteins. Mathematical morphology is applied on the tertiary structure of proteins; they are converted into a collection of 2D objects. From the PDB database, the protein structures 3V2J and 3V2M are viewed using *JMol* as shown in Figure 1(b) and (c), and the protein structures are rotated depending on the 3-axis, from which the 6-faces or views (front, left, right, top, bottom and back) of each tertiary protein structures are collected respectively as shown in Figure 8.

Let the six faces of each of the backbone structure respectively denoted by X^1(front), X^2(left), X^3(right), X^4(top), X^5(bottom) and X^6(back). The six faces for each of the protein structure (3V2J and 3V2M) are shown respectively in Figure 8(a)–(l). The aim of this section is to compute the similarity index between the backbone structure faces from source structure X_s^i and the target structure X_t^j.

Similarity index for X_s^i and X_t^j computed by geodesic dilation operation is as follows:

Step 1. If $X_s^i \cap X_t^j = X_s^i \cup X_t^j$, then there exists exact similarity. In such cases, the dilation distance between X_s^i and X_t^j, and between X_t^j and X_s^i would be zero.

Step 2. By performing dilation operation with structuring element S, the minimum number of dilation required from intersection parts to the source protein structure faces and to the target protein structure faces are represented by Δ_s and Δ_t respectively and defined as $\Delta_s = \min\{n : X_s^i \subseteq (X_s^i \cap X_t^j) \oplus nS\}$.

Step 3. Similarly, by performing dilation operation

$$\Delta_t = \min\left\{n : X_j^t \subseteq \left(X_s^i \cap X_t^j\right) \oplus nS\right\}.$$

Step 4. Difference between the numbers of dilation with respect to all faces from source protein to target protein structure is defined as:

$$\Delta = \sum_{l=1}^{6} |\Delta_s - \Delta_t|$$

Here, the front views of two different proteins are considered to compute the global similarity between them. For this purpose, specifically the front views of 3V2J and 3V2M are considered. Figure 8(a) and (g), the common structural part of both the protein molecules is given in Figure 9(a) and by

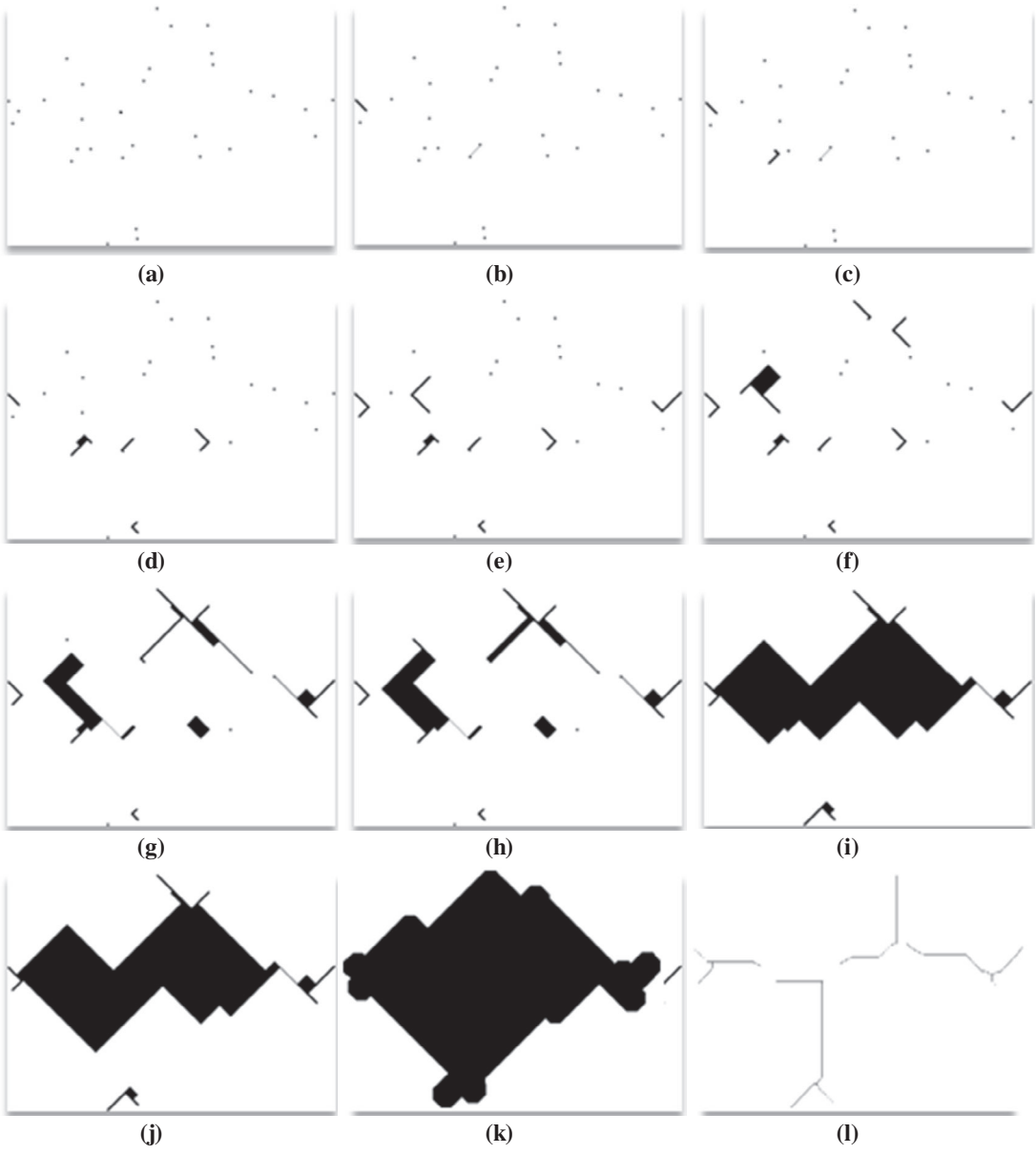


Figure 6. (a) A slice of protein 2LEP. (b–k) Different iterations of multi-scale opening of (a), and (l) the skeleton of the connected component of (k).

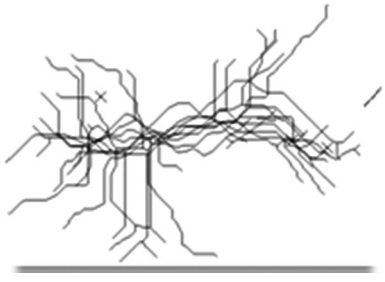


Figure 7. Skeleton network constructed for all the 30 slices of 2LEP protein.

Table 1. Fractal dimension of protein molecules.

Protein ID	Residue number	D_p
3V2J	260	1.6612
3SMK	236	1.6201
3T0O	238	1.6465
4ECS	435	1.6495
3V2M	260	1.6561
3SV1	190	1.6054
4AG2	226	1.6959
1CAH	259	1.6611
1CAI	259	1.6605
4BIJ	476	1.6802
2LEP	69	1.5496
1CGI	245	1.6359
4EYM	371	1.6495
2CBC	260	1.6614

performing dilation operation with structuring element S, the minimum number of dilation required from intersection parts to the source protein structure 3V2J front faces and to the target protein structure 3V2M front faces are given in Figure 9(b)–(e) and (f)–(i), respectively. The Geodesic dilation as a device determines the structural similarity between those protein molecules. The number of dilations required from the intersection part $\left(X_s^i \cap X_t^i\right)$ to both the protein molecules towards constructing the similar structure, i.e. the

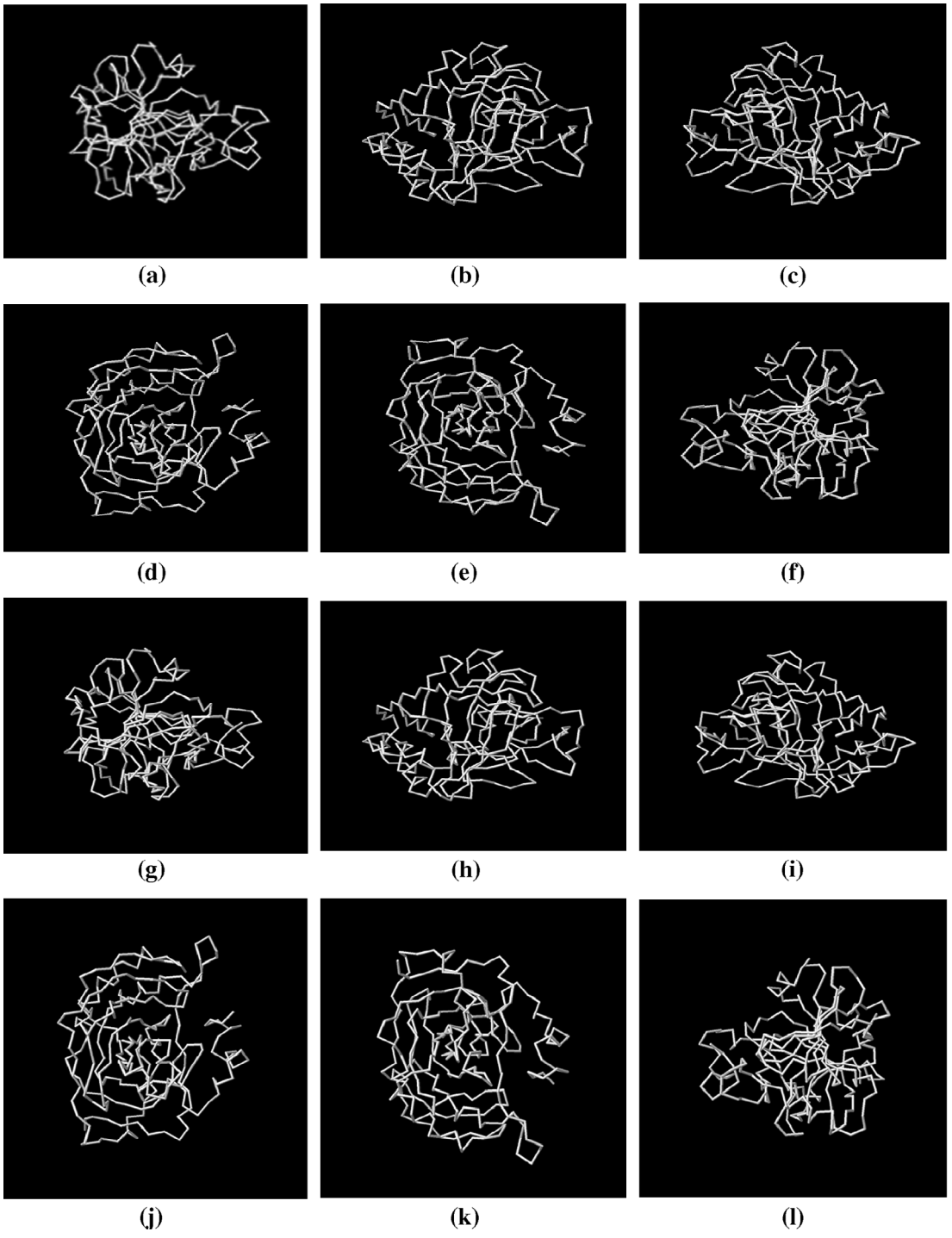


Figure 8. (a)–(f) Six different 2D faces of 3V2J protein and (g)–(l) six different 2D-faces of 3V2M protein.

number of geodesic dilation from $X_s^i \cap X_t^j$ towards 3V2J and 3V2M is four for each. Similarly, such geodesic dilation distance for other faces are given in Table 2.

Table 2 provides the geodesic dilation distance between the markers ($X_s^i \cap X_t^j$), intersection of source and target faces of corresponding backbone structure) and the source backbone structure as well as the geodesic dilation distance between marker and target backbone structure. The Δ is the difference between the source and target backbone structure.

Further, we extend this experiment over a set of 19 proteins of Homo sapiens having different residue which are {4AE8, 4AFQ, 4AFS, 4AFU, 4AFZ, 4AG1, 4AG2, 4AUO, 4AYT, 4AYX, 4ECQ, 4ECR, 4ECS, 4ECT, 4ECX, 4ECY, 4ED1, 4ED2, 4ED3}. For each pair of proteins from the set $Q \times Q$, we did find the difference operators ρ and Δ. The tables of these two operators are given in the Appendix 1. Also the PDB result of structural comparisons is also given at the beginning of the Appendix 1. The experimental result shows that if $\rho \leq 0.008$ and $\Delta \leq 12$, two protein molecules are very similar in tertiary structures.

Table 2. Geodesic dilation of different faces.

Protein ID	$(X_t^j) \cap (X_t^j)$	Front	Left	Right	Top	Bottom	Back	Δ_s / Δ_t	Δ
3V2J(X_s^i)	3V2J ∩ 3V2M	4	4	5	6	5	3	27	2
3V2M(X_t^j)		4	4	4	6	4	3	25	
2LE8(X_t^j)	2LE8 ∩ 2LLS	10	6	6	7	8	10	47	64
2LLS(X_t^j)		19	19	21	18	15	19	111	
1CAH(X_t^j)	1CAH ∩ 2CBC	1	1	1	1	1	1	6	1
2CBC(X_t^j)		1	1	1	2	1	1	7	
1CAH(X_t^j)	1CAH ∩ 4EYM	10	10	9	18	18	11	76	15
4EYM(X_t^j)		20	6	6	9	11	9	61	
3V2J(X_t^j)	3V2J ∩ 3T0O	12	8	9	12	12	11	64	29
3T0O(X_t^j)		4	5	5	5	11	5	35	

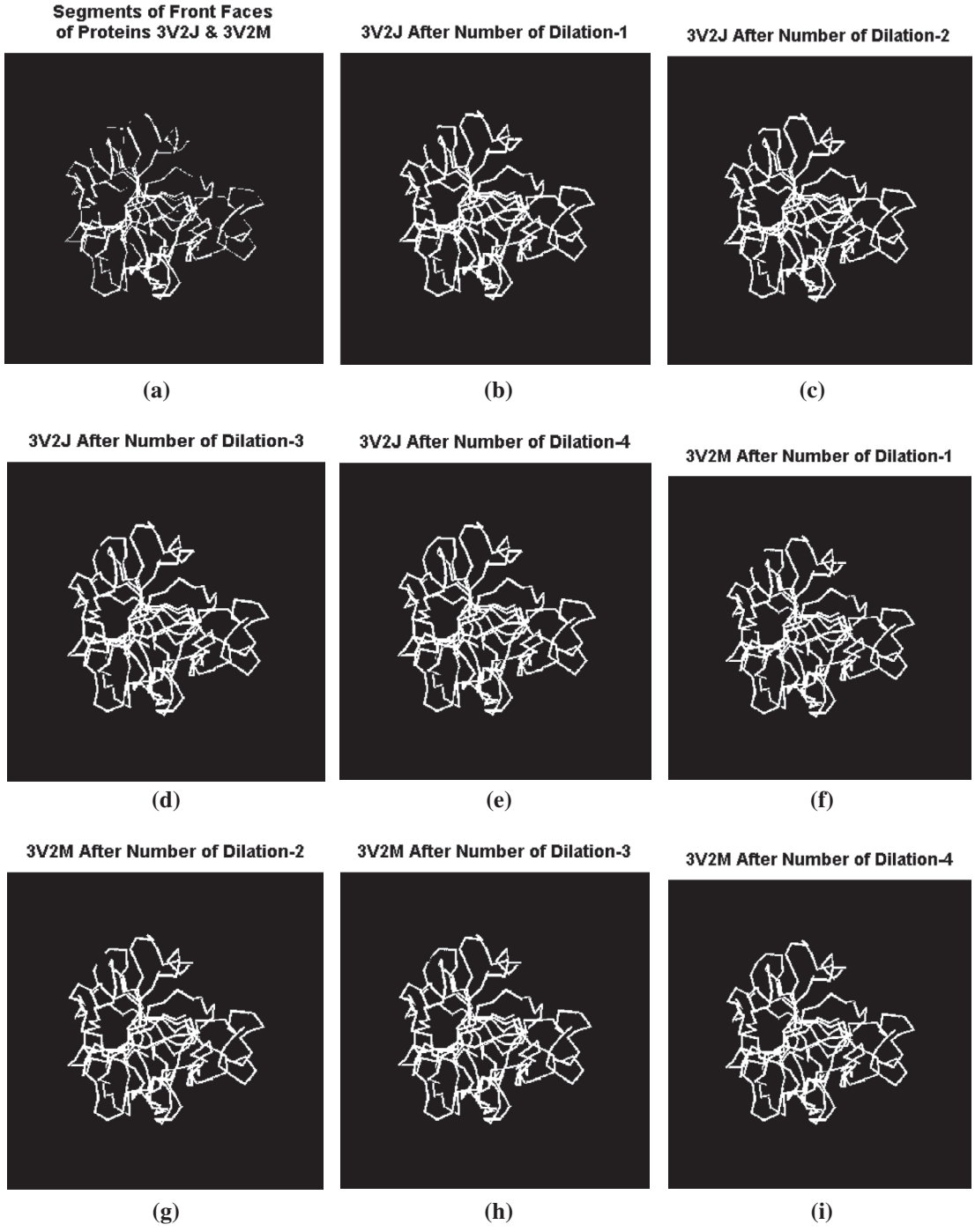


Figure 9. (a) Intersection segments between front face of 3V2J (Figure 8(a)) and front face of 3V2M (Figure 8(g)) protein structure, (b)–(e) reconstructed image after different dilation of intersection segments (Figure 9(a)) by taking front face of 3V2J (Figure 8(a)) as target structure and (f)–(i) reconstructed image after different dilation of intersection segments (Figure 9(a)) by taking front face of 3V2M (Figure 8(g)) as target structure.

Table 3. Difference between fractal dimensions of compared proteins pairs.

Protein ID-1	$D_i(p)$	Protein ID-2	$D_j(p)$	ρ	Δ	PDB result (%)
3V2J	1.6612	3SMK	1.6201	0.0411	24	12
		3T0O	1.6465	0.0147	29	18
		4ECS	1.6494	0.0118	31	2
		3V2M	**1.6561**	**0.0051**	**2**	**100**
		3SV1	1.6053	0.0559	28	28
		4AG2	1.6959	0.0347	39	31
1CAH	1.6611	**1CAI**	**1.6605**	**0.0006**	**5**	**100**
		4BIJ	1.6802	0.0191	21	41
		2LEP	1.5496	0.1114	43	57
		1CGI	1.6359	0.0252	35	50
		4EYM	1.6495	0.0117	15	39
		2CBC	**1.6613**	**0.0002**	**1**	**100**

Note: The bold value signifies 100% similarity, between two compared protein structures.

4.3. Result and discussion

The intricate comparison between two protein structures *i* and *j* has been captured using two difference operators ρ and Δ. The difference between the fractal dimensions essentially measures the difference between the structural complexities. As ρ approaches to zero, the structures are closed to be similar. It found to be true that if $\rho \leq 0.008$ and $\Delta \leq 12$, two protein molecules are very similar in structures. This has been clearly demonstrated with examples in the Table 3 including the Tables A1 and A2 of the Appendix 1. In the Table 3, it is shown that the protein *1CAH* is compared with a list of proteins viz. *1CAI*, *2LEP*, *2CBC* and so on. The difference ρ and Δ in the case of two proteins *1CAI* and *1CAH*are: 0.0006 and 5, respectively. Similarly, the difference ρ and Δ for the proteins *2CBC* and *1CAH* are 0.0002 and 1, respectively. As it is remarked earlier, these two pair of proteins is very structurally similar. The web server like FATCAT server and CE server says that the two pair of proteins is 100% similar. It is to be remarked that these pair of proteins are very close structurally but the degree of similarity is different as mentioned in Table 3 in each of the pair of proteins which is not clear from PDB result.

From the Tables A1 and A2, it is observed that there are protein pairs for which the difference operator ρ is less than 0.008 ($\Delta > 12$) and Δ is less than 12 ($\rho > 0.008$) hence such pair of proteins are failed to be very similar which is assured from the PDB result. Therefore, ρ is less than 0.008 and Δ is less than 12 are two necessary conditions for making a pair of proteins very similar. It is our strong conviction that the converse result is true as well, that is if two proteins are structurally (tertiary) highly similar, then corresponding ρ is less than 0.008 and Δ is less than 12. It is noted that the same experiment is done over a set of 500 pair of proteins (data are not shown) which were collected from PDB and it is found that the highly structural pair of proteins are following the conditions ρ is less than 0.008 and Δ is less than 12.

5. Conclusions

In this present work, the intricate details in the form of two-dimensional slices of the three-dimensional protein structures are being fetched out and which further have been used to understand the whole structural comparison of the protein tertiary structures. In past works, only the outer morphological structures were taken to compare two given proteins, whereas in the present article, two important techniques using the tools (fractal dimension and geodesic dilation) have been devised and employed to compute the structural similarity of protein structures. In the experiment, the atoms of all the protein structures are divided into slices by fixing the *z* co-ordinate values (along *z*-axis) and therefore no scaling effect does matter as such in doing the comparisons of the protein structures. It is needless to mention that the present work is based on area-based morphological analysis and so one is free to make slices along the *x* and *y* axes too and in that case, the result would not be affected as reported here. In near future, the volumetric analysis of the tertiary protein structures is planned to study.

List of symbols

Slice of tertiary protein structure	A
Structuring element	S
Dilation operator	$\oplus$
Erosion operator	$\ominus$
Opening	O
Slices at *z* coordinates	X^i
Connected components	X_c^i
Skeleton of connected components	Sk
Fractal dimension	D_p
Source tertiary protein structure faces	X_s^i
Target tertiary protein structure faces	X_t^j
Dilation of X_s^i	Δ_s
Dilation of X_t^j	Δ_t
Fractal dimension (difference)	Δ
Geodesic dilation distance (difference)	ρ

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. 1983. Molecular biology of the cell. New York (NY): Garland Publishing.

Anfinsen CB. 1973. Principles that govern the folding of protein chains. Science 181:223–230.

Avnir D, Biham O, Lidar O, Malcai O. 1998. Applied mathematics: is the geometry of nature fractal? Science 279:39–40.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. Nucleic Acids Res. 28:235–242.

Bhattacharya A, Can T, Kahveci T, Singh AK, Wang YF. 2004. ProGreSS: simultaneous searching of protein databases by sequence and structure. Proc PacificSymp Biocomputting; 2004 Jan 6–10; Hawaii, USA; p. 264–275.

Billeter M. 1992. Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals, Q Rev Biophys. 25: 325–377.

Branden CI, Tooze J. 1991. Introduction to protein structure. 2nd ed. New York (NY): Garland Publishing.

Can T, Wang YF. 2003. CTSS: a robust and efficient method for protein structure alignment based on local geometrical and biological features. Proc IEEE CS Conf Bioinformatics (CSB); Stanford (CA);p. 169–179.

Chew LP. 2006. Exact computation of protein structure similarity. Proceeding SCG '06 Proceedings of the Twenty-second Annual Symposium on Computational Geometry; Sedona (AZ); p. 468–474.

Chiu W. 2002. Deriving folds of macromolecular complexes through electron cryomicroscopy and bioinformatics approaches. Curr Opin Struct Biol. 12:263–269.

Choi IG, Kwon J, Kim SH. 2004. Local feature frequency profile: a method to measure structural similarity in proteins. PNAS. 101:3797–3802.

Cui C, Wang D, Shi J. 2004. Comparing 3-D protein structures similarity by using fractal features. Proc IEEE Computational Systems Bioinformatics Conf (CSB 04); Stanford; p. 698–699.

Cui C, Wang D, Yuan X. 2005. 3D protein structures similarity matching based on fractal features. SPIE. 5637:567–572.
Daras P, Zarpalas D, Axenopoulos A, Tzovaras D, Strintzis MG. 2006. Three-dimensional shape-structure comparison method for protein classification. IEEE/ACM Trans Comput Biol Bioinf. 3:193–207.
Develi K, Babadagli T. 1998. Quantification of natural fracture surfaces using fractal geometry. Math Geol. 30:971–998.
Drenth J. 1999. Principles of protein x-ray crystallography. New York, (NY): Springer-Verlag.
Galgonek J, Hoksza D, Skopal T. 2011. SProt: sphere-based protein structure similarity algorithm. Proteome Sci. 9:1–12.
Gibrat JF, Madej T, Bryant H. 1996. Surprising similarities in structure comparison. Curr Opin Struct Biol. 6:377–385.
Goldman D, Papadimitriouz CH, Istraily S. 1999. Algorithmic aspects of protein structure similarity. FOCS 99 Proceedings of the 40th Annual Symposium on Foundations of Computer Science; New York; 512 pp.
Guyon F, Tuffry P. 2010. Assessing 3D scores for protein structure fragment mining. Open Access Bioinf. 2:67–77.
Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. J Mol Biol. 233:123–138.
Holm L, Sander C. 1996. Mapping the protein universe. Science 273.
Holm L, Ouzounis C, Sander C, Tuparev G, Vriend G. 1992. A database of protein structure families with common folding motifs. 1:1691–1698.
Klette G. 2002. Skeletons in digital image processing. Glen Innes: Computer Science Department, University of Auckland; p. 21. Available from CITR. Glen Innes, Auckland, NewZealand;CITR-TR-112.
Kleywegt GJ. 1999. Experimental assessment of differences between related protein crystal structures. Acta Crystallogr D Biol Crystallogr. 55:1878–1884.
Koehl P. 2001. Protein structure similarities. Curr Opin Struct Biol. 11:348–353.
Krasnogor N, Pelta DA. 2003. Measuring the similarity of protein structures by means of the universal similarity metric. Bioinformatics 20:1015–1021.
Lancia G, Istrail S. 2003. Protein structure comparison: algorithms and applications. Lect Notes Comput Sci. 2666:1–33.
Maragos PA, Schafer Ronald W. 1986. Morphological skeleton representation and coding of binary images. IEEE Trans Acoust Speech Signal Process. 34:1228–1244.
Matheron G, Serra J. 2004. History of mathematical morphology. Available from: http://cmm.ensmp.fr/serra/pdf/birth-of-mm.pdf
Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 247:536–540.
Orengo CA, Michie AD, Tones JD, Swindells MS, Thornton JM. 1997. CATHA hierarchic classification of protein domain structures. Structure 5:1093–1108.
Petsko GA, Ringe D. 2004. Protein structure and function. Sunderland (MA): Sinauer Associates.
Radhakrishnan P, Teo LL, Sagar BSD. 2004. Estimation of fractal dimension through morphological decomposition. Chaos Solitons Fractals. 21:563–572.
Robillard GT, Tarr CE, Vosman F, Berendsen HJC. 1976. Similarity of the crystal and solution structure of yeast tRNAPhe. Nature 262: 363–369.
Sagar BSD, Tien TL. 2004. Allometric power-law relationships of hortonian fractal digital elevation model. Geophys Res Lett. 31:L06501.
Sagar BSD, Venu M, Srinivas D. 2000. Morphological operators to extract channel networks from digital elevation models. Int J Remote Sens. 21:21–30.
Serra J. 1982. Image analysis and mathematical morphology. London: Academic.
Shindyalov IN, Bourne PE. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng Des Sel. 11:739–747.
Teo LL, Sagar BSD. 2006. Modeling, description, and characterization of fractal pore via mathematical morphology. Discrete Dyn Nat Soc. 2006, Article ID 89280; 1–24.
Vassura M, Margara L, Lena PDI, Medri F, Fariselli P, Casadio R. 2008. Reconstruction of 3D structures from protein contact maps. IEEE/ACM Trans Comput Biol Bioinf. 5:357–367.
Zemla A. 2003. LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res. 31:3370–3374.

Appendix 1.

Table A1. PDB result of structural similarity percentage of 19 proteins.

Protein-ID	4AE8 (%)	4AFQ (%)	4AFS (%)	4AFU (%)	4AFZ (%)	4AG1 (%)	4AG2 (%)	4AUO (%)	4AYT (%)	4AYX (%)	4ECQ (%)	4ECR (%)	4ECS (%)	4ECT (%)	4ECX (%)	4ECY (%)	4ED1 (%)	4ED2 (%)	4ED3 (%)
4AE8	100	50	43	31	48	37	34	47	59	51	42	37	37	38	42	37	42	42	41
4AFQ		100	100	98	99	98	99	42	35	37	27	27	27	28	27	27	27	27	27
4AFS			100	98	99	98	99	43	36	36	21	22	23	23	29	21	22	21	23
4AFU				100	100	99	100	42	18	28	22	22	29	27	30	22	30	21	23
4AFZ					100	98	99	42	35	35	29	28	30	25	27	28	22	29	29
4AG1						100	100	40	29	29	20	23	31	23	23	22	20	20	22
4AG2							100	43	31	34	30	31	28	31	31	28	31	31	31
4AUO								100	39	43	37	37	35	31	37	37	33	33	37
4AYT									100	100	41	41	41	41	41	41	41	41	41
4AYX										100	41	41	41	41	41	41	41	41	41
4ECQ											100	100	100	100	100	100	100	100	100
4ECR												100	100	100	100	100	100	100	100
4ECS													100	100	100	100	100	100	100
4ECT														100	100	100	100	100	100
4ECX															100	100	100	100	100
4ECY																100	100	100	100
4ED1																	100	100	100
4ED2																		100	100
4ED3																			100

Table A2. ρ is calculated for 19 pair of proteins and reddish coloured box indicates that ρ is less than 0.008.

Protein -ID	4AE8	4AFQ	4AFS	4AFU	4AFZ	4AG1	4AG2	4AUO	4AYT	4AYX	4EC Q	4EC R	4EC S	4EC T	4EC X	4EC Y	4ED1	4ED2	4ED3
4AE8	0	0.042841	0.057092	0.04005	0.02122	0.02758	0.01847	0.03634	0.06732	0.06717	0.02884	0.02719	0.0279	0.02473	0.03257	0.02445	0.02504	0.02765	0.02853
4AFQ		0	0.000851	0.00021	0.00594	0.00669	0.00131	0.04918	0.02448	0.02433	0.01401	0.01565	0.01494	0.01811	0.01028	0.0184	0.0178	0.0152	0.01431
4AFS			0	0.00064	0.00509	0.00754	0.00216	0.06343	0.01023	0.01008	0.02826	0.0299	0.02919	0.03236	0.02453	0.03265	0.03205	0.02945	0.02857
4AFU				0	0.00573	0.0069	0.00152	0.04639	0.02728	0.02712	0.01121	0.01286	0.01215	0.01532	0.03748	0.0156	0.01501	0.0124	0.01152
4AFZ					0	0.0126	0.00725	0.00512	0.06854	0.06839	0.03006	0.02841	0.02912	0.02595	0.03379	0.02567	0.02626	0.02887	0.02975
4AG1						0	0.00537	0.03392	0.03974	0.03959	0.02125	0.03039	0.02032	0.01285	0.04498	0.01314	0.04254	0.03206	0.01595
4AG2							0	0.01213	0.0858	0.08565	0.04731	0.04566	0.04637	0.0432	0.05104	0.04292	0.04351	0.04612	0.047
4AUO								0	0.07366	0.07351	0.03518	0.03353	0.03424	0.03107	0.03891	0.03079	0.03138	0.03399	0.03487
4AYT									0	0.00015	0.03849	0.04013	0.03943	0.0426	0.03476	0.04288	0.04228	0.03968	0.0388
4AYX										0	0.03834	0.03998	0.03928	0.04245	0.03461	0.04273	0.04213	0.03953	0.03865
4EC Q											0	0.00165	0.00094	0.00411	0.00373	0.00439	0.00379	0.00119	0.00031
4EC R												0	0.00071	0.00246	0.00538	0.00275	0.00215	0.00046	0.00134
4EC S													0	0.00317	0.00467	0.00345	0.00286	0.00025	0.00063
4EC T														0	0.00784	0.00028	0.00031	0.00292	0.0038
4EC X															0	0.0081	0.00752	0.00492	0.00404
4EC Y																0	0.0006	0.0032	0.00408
4ED1																	0	0.0026	0.00349
4ED2																		0	0.00088
4ED3																			0

Table A3. Δ is calculated for 19 pair of proteins and green coloured box indicates that Δ is less than 12.

Prote i n -ID	4AE8	4AFQ	4AFS	4AFU	4AFZ	4AG1	4AG2	4AUO	4AYT	4AYX	4EC Q	4EC R	4EC S	4EC T	4EC X	4EC Y	4ED1	4ED2	4ED3
4AE8	0	16	12	10	20	18	12	40	16	15	13	19	15	10	10	19	15	16	16
4AFQ		0	10	10	10	5	10	30	21	23	18	20	17	20	21	21	33	40	38
4AFS			0	9	11	4	5	23	28	27	18	17	18	19	18	17	34	35	33
4AFU				0	10	5	10	42	12	11	10	14	11	12	13	13	27	31	30
4AFZ					0	11	11	44	30	29	30	28	28	28	29	28	30	29	31
4AG1						0	6	51	18	16	24	26	23	25	26	27	35	37	41
4AG2							0	19	23	22	25	27	23	28	28	27	37	35	37
4AUO								0	42	51	47	52	45	49	49	49	72	66	68
4AYT									0	7	25	26	23	27	26	26	22	23	23
4AYX										0	28	25	20	26	26	26	18	21	20
4EC Q											0	3	9	0	0	2	11	7	10
4EC R												0	8	1	2	4	10	5	9
4EC S													0	8	7	8	11	9	11
4EC T														0	3	0	10	7	10
4EC X															0	1	10	5	9
4EC Y																0	10	7	10
4ED1																	0	1	1
4ED2																		0	1
4ED3																			0