

Surface characterization of proteins using multi-fractal property of heat-denatured aggregates

Tapobrata Lahiri¹*, Hrishikesh Mishra¹, Subrata Sarkar¹, Krishna Misra²

¹Indian Institute of Information Technology, Allahabad, India; ²Centre of Biotechnology, University of Allahabad, Allahabad, India; Tapobrata Lahiri* - E-mail: tlahiri@iiita.ac.in; * Corresponding author

received March 07, 2008; revised May 05, 2008; accepted May 31, 2008; published June 23, 2008

Abstract:

Multi-fractal property of heat-denatured protein aggregates (HDPAs) is characteristic of its individual form. The visual similarity between digitally generated microscopic images of HDPAs with that of surface-images of its individual X-ray structures in protein databank (PDB) displayed using Visual Molecular Dynamics (VMD) viewer is the basis of the study. We designed experiments to view the fractal nature of proteins at different aggregate scales. Intensity based multi-fractal dimensions (ILMFD) extracted from various planes of digital microscopic images of protein aggregates were used to characterize HDPAs into different classes. Moreover, the ILMFD parameters extracted from aggregates show similar classification pattern to digital images of protein surface displayed by VMD viewer using PDB entry. We discuss the use of irregular patterns of heat-denatured aggregate proteins to understand various surface properties in native proteins.

Keywords: protein structure; multi-fractal dimension; protein aggregate; digital image processing; light microscopy; degree of differentiation; graphical pair-wise class comparison

Background:

The study of protein surface is emerging as a novel approach in the field of protein-ligand interactions [1] where part of the surface plays the dominant role of binding site in such interactions. A ligand may be a protein, a small molecule, DNA, RNA, or any other drug like molecule. These surface sites are either rough or deep or cleft like zones. Protein-protein interactions occur via flat surface on proteins. But for large molecules like DNA, interaction occurs through convex surface or elevations on the protein surface. Drug binding sites are also found as rough areas on protein surface [1]. Therefore, quantitative measure of surface is needed to differentiate and identify various types of protein-ligand binding sites.

Surface roughness means presence of several elevations and depressions in a small surface-area. Roughness of binding sites indicates the complex local shapes needed for efficient use of specific interactions in binding sites [1]. Roughness of the surface allows more Van der Waals' contacts to occur between the binding site and the interacting molecule. So, rougher the binding site, stronger is the interaction between protein and binding molecule. Moreover, it is well-known that interaction between surface pockets and ligands is not just one-to-one type because, one ligand may have affinity towards more than one protein surfaces and different ligands may bind to the same protein surface [2].

The importance of studies on protein surface roughness mainly lies in defining functional sites of proteins from

their surface. The description of this surface as indexed surface parameter is possible only through the availability of 3D structures of protein derived by X-ray crystallography or nuclear magnetic resonance (NMR) or molecular modeling. These studies are helpful for further extraction of active or functional site of proteins. They are also sometimes used for structural classification [3]. The bottleneck is the prerequisite of already evaluated protein structures. This is not possible for 90% of proteins available now.

We observed that the direct listing of surface parameters in proteins without X-ray structures is the key issue for deriving protein functional sites. Our study shows that there is a possibility to detect some protein surface properties without X-ray structures. The alternative approach is to use easily available heat denatured protein aggregates. Several studies show their specificity for corresponding aggregates. Bohr and colleagues (1997) [4] showed the resultant aggregate structures of proteins are strongly influenced by the shape of their individual molecule using electronic, atomic force and ordinary microscope. Jakubowski [5] reviews that these aggregates are not as non-specific as believed. Patro and Przybycien [6], simulated protein aggregation and showed that the extent and orientation of monomer hydrophobic surface area, the magnitude of protein-protein interaction energies and the configurational entropy loss are the factors that considerably impact the structure of kinetically irreversible protein aggregates. These were manifested in terms of change in free energy of

the system. The aggregate parameters like jamming limit, protein-protein contacts distribution, solvent accessible hydrophobic surface area, porosity and short or long range order are considered in these studies. These parameters of protein aggregation can influence specific structure of aggregates. Here, we describe the use of irregular patterns of heat-denatured aggregate proteins to understand various surface properties in native proteins.

Methodology:

Proteins used in the study

The proteins chosen for the study were albumin, haemoglobin, insulin, ferritin, cytochrome-C and lysozyme. All chemicals were of analytical grade and were used without further purification. All the proteins used in this work were obtained from Sigma Aldrich Inc. (USA) in purified form. We used purified water by Millipore Water System (Model: Millipore, USA) throughout the experiment.

Aggregate formation and HDPA image dataset

In our experiment we suspended the proteins in Millipore water (50 mg/cc) and boiled it at 100°C for 15 minutes to get HDPA. HDPA drops were poured in Neubauer Chamber using micropipette and viewed in 1000× magnification of transmitted-light phase-contrast

microscope (Leica-DM-LB2). Seven images of HDPA taken at different fields of views for each protein were collected by a camera fitted to the microscope (Canon-Power Shot S50) at its optical zoom 2×. This forms HDPA image dataset for proteins.

Image dataset of native proteins

Images of 3D molecular structures for each of the proteins were visualized and collected by VMD surface visualization software. VMD is designed for modeling, visualization, and analysis of biological systems such as proteins, nucleic acids and lipid bi-layer assemblies. VMD can read PDB files and display structures. VMD provides a wide variety of options for drawing and coloring a molecule. The option used by us is VDW (Van der Waals) as a drawing method, "Name" as coloring method, "Opaque" as material with sphere scale 1.0 and sphere resolution 8 [7]. VMD images of proteins were compared with images of aggregate counterpart using the following steps (illustration is given in figure 1): (1) the orientation of the VMD-viewer-image of a protein was carefully changed to find the best match with one of its corresponding aggregate image. (2) Step 1 was repeated for different aggregates of the same protein. 15 VMD images taken at various orientations for each protein were thus collected and this forms the PDB image dataset of proteins.

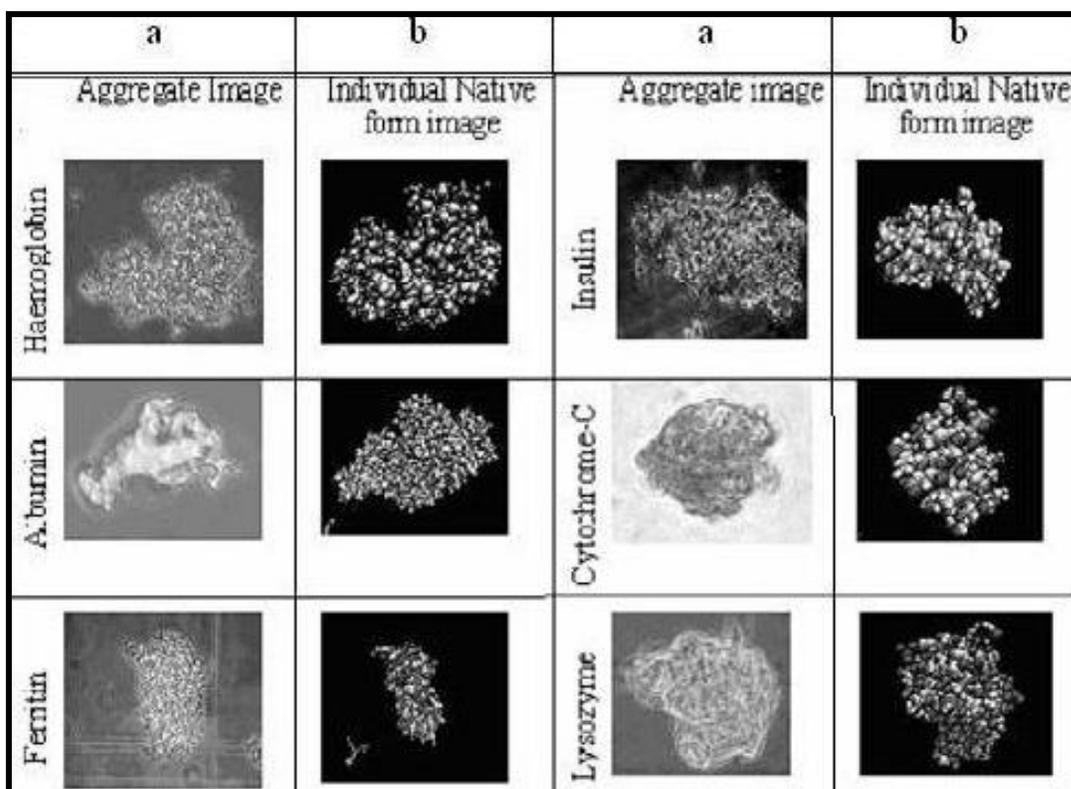


Figure 1: Representative matches of protein-aggregates with their corresponding native form are shown for six proteins. Column (a) shows the aggregate images of proteins and column (b) shows the surface map of corresponding individual native form of proteins shown by VMD viewer. The visual similarity match is obtained by manual rotation.

Computational protocol

Computation and graphics were performed using the mathematical simulation software MATLAB.

Preprocessing of images

Each image was converted to grey scale and resized to 1/3rd of the original size. Aggregates were segmented by using Adobe Photoshop (version 7.0).

Intensity plane slicing of images

Each resized grey image **I** was split into 10 binary images (B_1, B_2, \dots, B_{10}) on the basis of fixed intensity ranges S_1, S_2, \dots, S_{10} [8]. We obtained the binary image B_i by the simple rule that, pixel value of $B_i = 1$, if the corresponding pixel value of **I** remains within the range S_i and S_{i+1} . Otherwise, the pixel value of B_i is assigned to zero. The value of $S_i = 0$ for $i = 1$, otherwise $S_i = (m/10) * (i-1) + 1$, for $i > 1$ where m is the maximum intensity of the image **I**.

Multi-fractal formalism

Fractal dimensions for each intensity plane were measured using box counting method [9] and were used to calculate

multi-fractal dimension (ILMFD) using $D = \left\{ D_i \right\}_{i=0}^{10}$.

In the above expression, D is set to 10 fractal dimensions calculated from each of 10 binary images that are obtained by intensity-slicing method (vide previous section).

ILMFD for PDB structures

The preprocessing steps described above were followed for these images except for segmentation. Thus, ILMFDs were evaluated for all PDB images of proteins following the same procedure as described above.

Graphical pair-wise class comparison (GPCC)

Table 1 (see supplementary material) enables visualization of separation between all possible ILMFD-cluster-pairs within a multi-dimensional feature space for both aggregate and individual proteins. Each protein-pair circles were drawn by taking the first and second member of the pair respectively from Table 1 (see supplementary material) with their corresponding protein-ILMFD-cluster-radii R_{LHS} and R_{RHS} respectively (depends on the scale of the figure). CG to CG distances, d between these protein-ILMFD-clusters were measured. The degree of differentiation (DD) for a pair of proteins within their 10 dimensional ILMFD feature space was measured using $D_A = d / (R_{LHS} + R_{RHS})$. Naturally, if $R_{LHS} + R_{RHS} = d$, the members of the variant-pair are separated. If $R_{LHS} + R_{RHS} < d$, the members of the variant-pair are well-separated and if $R_{LHS} + R_{RHS} > d$, the variant clusters of the variant-pair are overlapped.

Discussion:

Functional significance of selected proteins

All the proteins used in the experiment are functionally different. Albumin acts as binding protein for several drug-like molecules in blood. Haemoglobin functions in the transport of oxygen from lungs to different body parts and carbon dioxide from body tissues to lungs. Ferritin is a

storage protein used for storing iron in liver cells. Insulin is a peptide hormone which regulates blood glucose level. Cytochrome-C is an enzymatic protein in various metabolic reactions. Lysozyme is a family of enzymes (EC 3.2.1.17) which damages bacterial cell walls by catalyzing hydrolysis of 1, 4-beta-linkages between N-acetyl-muramic acid and N-acetyl-D-glucosamine residues in a peptide-glycan and between N-acetyl-D-glucosamine residues in chitodextrins.

Universal aggregation and its accomplishment

As discussed earlier, we were in search of aggregate based protein markers and therefore interested to find a single method of aggregation that can be universally applicable for producing protein aggregates. Producing an aggregate of native protein is the most suitable starting point. However, it is nearly impossible to develop aggregates for most proteins. Heat denaturation is the easiest method of aggregation for proteins. Protein denaturation is sensitive to denaturing methods [9]. This encouraged us to optimize denaturation protocols universally applicable to get aggregation of all types of proteins.

GPCC of protein-aggregates with native forms

Figure 1 shows the representative matches of protein-aggregates with native form as described in methodology. Visual similarity is quite intriguing to start exploring the possibility of having an analytical basis of such relationship.

Visual comparison of aggregates with native form

As shown in figure 1, we were interested in visually comparing individual protein molecules with their aggregates. The shapes of different aggregates of a particular protein are presumed as statistically self-similar. Thus, for a particular protein, distribution of its different-aggregates within microscopic field of view should closely resemble different-orientations of statistically same aggregate of a particular protein. If the shapes of protein aggregates are presumed to be similar to its corresponding native form then for each orientation of the aggregate its shape should match with the native protein shape. This is viewed through the VMD viewer after proper tuning of its orientation by VMD options.

Degree of differentiation (DD)

Table 1 (see supplementary material) contains DD-values, D_A for the protein-aggregates and D_I for the individual native protein-pairs. Cells on both sides of the diagonal show the same combinations or pairs of proteins. The cells with un-bold values on left side of diagonal are used to represent DD values for PDB images (D_I) and cells on the right side of the diagonal with bold values represent DD values for aggregate images (D_A). Both of these values show overall efficiency of the ILMFD feature in differentiating one protein from another. As we have already discussed in methodology, the average values of ($D_A^{AV} = 0.4015$) and D_I ($D_I^{AV} = 0.2356$) are indicative of differentiating capability of the feature of our concern

namely ILMFD. The mean DD values greater than or equal to 1 indicate full differentiation (methodology section) which was not fully satisfied in this case. The significance in fulfilling the goal was followed from the comparative analysis. In our bid to find the correspondence of protein-aggregates with their native form, we compared their DD values, D_A and D_I for different combinations or pairs as marked by their different indices, omitting the data for two pairs (shown in figure 2).

The indices of protein pairs used in Figure 2 are: (1) albumin and lysozyme, (2) albumin and insulin, (3) albumin and haemoglobin, (4) albumin and ferritin, (5)

lysozyme and cytochrome-C, (6) lysozyme and insulin, (7) lysozyme and hemoglobin, (8) lysozyme and ferritin, (9) cytochrome-C and insulin, (10) cytochrome-C and haemoglobin, (11) cytochrome-C and ferritin, (12) insulin and haemoglobin, and (13) haemoglobin and ferritin. Protein-pairs omitted in this figure are (a) albumin and cytochrome-C and (b) insulin and ferritin. The striking similarity of pattern generated by D_A with that generated by D_I further strengthened our contention that there is a possibility of having an aggregate-surface based signature of individual protein and the aggregate parameter mapped to parameters of individual native form of the concerned protein.

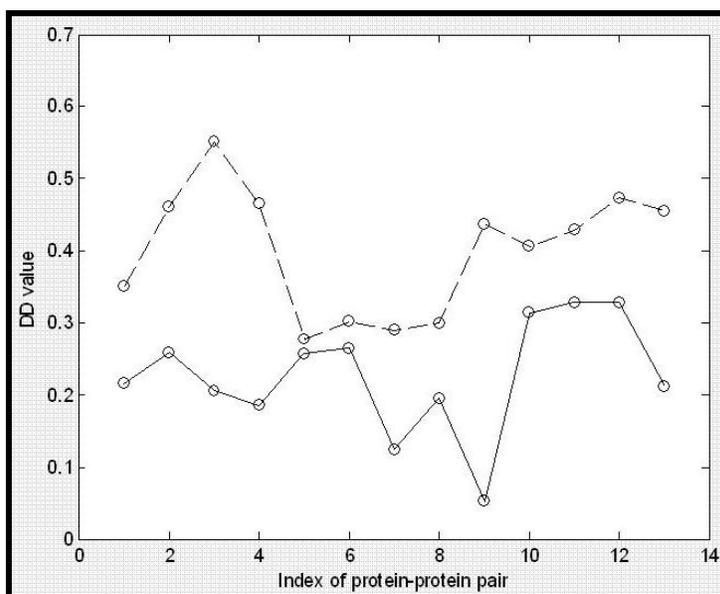


Figure 2: Pattern generated by D_A (fragmented line and circle) and D_I (solid line and circle) for all possible pairs of proteins (as shown in Table 1 under supplementary material) omitting data for 2 protein-pairs.

Conclusion:

Aggregation of proteins is predominantly a protein surface mediated phenomenon. So the idea of protein aggregates as a useful tool to extract information of individual proteins is feasible. The simple microscopic experiment based study described here demonstrated systematic analysis of microscopic images of irregular protein aggregates. This may help to explore individual protein property if a suitable feature is available and is mapped from aggregate to individual native form. We used multi-fractal property for 2D images of both proteins and their aggregates. This is because 2D intensity roughness map for object of interest (i.e., proteins and their aggregates) is actually representation of their surface and surface roughness in two dimensional projections. Moreover, the fractal property is conventionally a representative of roughness [10]. Therefore, the property or the feature ILMFD reflects the surface roughness of both proteins and their aggregates.

ILMFD is a rotation independent feature [11]. Therefore, it can be applied without difficulties for studying orientation of the object. These facts describe the basis to use ILMFD as a surface feature. This is a protocol-design for application in the area of structural proteomics. Protein structure determination by X-Ray crystallography is laborious and time consuming. Here, we show the possibility of utilizing irregular assembly of proteins namely HDPA. The method hypothesized here can be applied for any proteins without the need for X-ray data.

Acknowledgement:

We are grateful to Dr. Pinakpani Chakrabarti, Head, Bioinformatics Centre, Bose Institute, Kolkata, India for helpful discussion. We acknowledge the significant contribution made by Anjan Kumar Dasgupta, Head, Department of Biochemistry, University of Calcutta, for improving the manuscript to its current form. We also thankfully acknowledge the grant received from MHRD,

Government of India (No.F.26-14/2003.T.S.V.Dated 14.01.2004) for the work.

References:

- [01] F. K. Pettit and J. U. Bowie, *J. Mol. Biol.*, 285: 1377 (1999) [PMID: 9917382]
 [02] A. Via *et al.*, *Cell. Mol. Life Sci.*, 57: 1970 (2000) [PMID: 11215522]
 [03] S. Singha *et al.*, *Online Journal of Bioinformatics*, 7: 74 (2006)
 [04] H. Bohr *et al.*, *Z. Phys. D.*, 40: 513 (1997)
 [05] <http://employees.csbsju.edu/hjakubowski/classes/ch331/protstructure/olprotaggreg.html>
 [06] S. Y. Patro and T. M. Przybycien, *Biophysical Journal*, 66: 1274 (1994) [PMID: 8061184]
 [07] http://www.ks.uiuc.edu/Research/vmd/allversions/w hat_is_vmd.html
 [08] A. W. Vermeer and W. Norde, *Biophysical Journal*, 78: 394 (2000) [PMID: 10620303]
 [09] E. Kawaguchi and R. Taniguchi, *IEEE Transactions on Systems, Man and Cybernetics*, 19: 1321 (1989)
 [10] B. B. Mandelbrot *et al.*, *Nature*, 308: 721 (1984)
 [11] R. Singh *et al.*, *Lecture Notes in Computer Science*, 3832: 713 (2005)

Edited by P. Kanguane

Citation: Lahiri *et al.*, *Bioinformatics* 2(9): 379-383 (2008)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material

	Albumin	Cytochrome C	Ferritin	Hemoglobin	Insulin	Lysozyme
Albumin		0.2564	0.4651	0.5522	0.4608	0.3519
Cytochrome C	0.2650		0.4291	0.4058	0.4371	0.2765
Ferritin	0.1846	0.3277		0.4546	0.1567	0.2999
Hemoglobin	0.2067	0.3133	0.2108		0.4733	0.7013
Insulin	0.2598	0.0527	0.3272	0.3291		0.3016
Lysozyme	0.2164	0.2571	0.1947	0.1241	0.2648	

Table 1: D_A and D_I for all possible pairs of protein are shown. Cells with un-bold numbers show DD values for PDB images while cells with bold numbers show DD values for aggregate images in corresponding protein-protein pair.