ESTIMATION OF PROTEIN PACKING DENSITY FROM PROTEIN CONTACT MAP

R. Mageswari¹, K. Srinivasa Rao², K. Sivakumar³
¹Assistant Professor, Department of Mathematics
E-mail: mageswari78@gmail.com
²Associate Professor, Department of Mathematics
³Associate Professor, Department of Chemistry
Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya (SCSVMV University),
Kanchipuram - 631 561, Tamilnadu, INDIA.

Abstract

Packing of protein atoms is an indicative factor for understanding its stability and functionality. In this paper, a new approach is proposed for estimating the packing density of a protein from its contact map. Maximum number of contacts for each residue within the particular threshold is approximated as the irregular packing of identical spheres within the sphere of newly calculated radii. Contact map is constructed from the tertiary structure of a protein using the distances between the alpha carbon atoms. Alpha carbon atoms of all residues are taken as vertices of the graph obtained from its contact map, the edges of the graph depends on the distance between the carbon alpha atoms.

AMS Subject Classification: 97K30

Key Words and Phrases: alpha carbon, contact map, packing density, protein structure.
1 Introduction

Volume occupied by the proteins and voids existing in the proteins are the main source of information required for understanding the structure and stability of protein [2, 3, 4]. Occupied volume of a protein [5] can be calculated using the different geometrical methods reported in the literature. Among that, Voronoi diagram is also a geometric construction to find the occupied volume and the packing density of a protein [6, 7].

The ratio of the volume enclosed by the Van der walls envelope [4, 6, 8] of a given molecule or atom to the actual volume of space that it occupies gives the packing density of the molecule. Volumes occupied by all the atoms or groups of atoms can be estimated using voronoi polyhedra using the atomic co-ordinates [9]. But, voronoi method is useful only for buried residues since it requires an atom to be surrounded by neighboring atoms. The overall packing of a protein is estimated from the mean length of the extended normal [10] for all residues in a protein structure. Protein packing is analyzed through the newly calculated radii and volumes [11] of atomic groups and it is also defined as the ratio between the number of edges in the protein graph and the number of edges in the maximum packing polymer [14]. It is found that packing of proteins is dependent on the size of the protein, the content of secondary structure elements and the amino acid composition, using occluded surface algorithm [12].

Among the different mathematical methods for protein structure identification, graph theoretical methods are easy to estimate. Contact map plays an important role in understanding proteins structure. Using contact map of the protein, specific folds [13] in the protein are identified.

In this paper, we propose a new method for estimating the packing density of protein using the [15] contact map and the polymer having maximum number of contacts within the particular threshold.
2 Preliminaries

2.1 Alpha Carbon

The alpha carbon (α carbon) in organic molecules refers to the first carbon atom that attaches to a functional group.

2.2 Contact maps

In general, protein 3D structures are transfigured into a graph [14] by assuming Cα atoms as nodes and the distance between two Cα atoms as edges. The distance between two residues may be defined by the distance between two carbon alpha (Cα) atoms or between two carbon beta (Cα) atoms or it may be the minimum distance between any pair of atoms belonging to the side chain or to the backbone of two residues.

A contact between two given atoms (or residues) exists when a certain distance is below a given threshold.

Let $P$ be a protein with $n$ atoms which are labeled $1, 2, 3 \ldots n$. The contact map of the protein is defined as a matrix,

$$T = (t_{ij})_{1 \leq i, j \leq n}$$

where

$$t_{ij} = \begin{cases} 
1 & \text{if } i \neq j \text{ and } d_{ij} \leq 6\text{Å} \\
0 & \text{otherwise.} 
\end{cases}$$

A protein can be considered as a graph $G = (V, E)$ for which each vertex $v_i \in V$ represents a residue of the protein and each $v_i - v_j \in E$ represents a contact between two residues $v_i$ and $v_j$. On the other hand, there is an edge $v_i - v_j \in E$ if $t_{ij} = 1$.

3 Methodology for Converting a Protein into a Graph

3.1 Protein extraction from protein database

Protein with PDB ID 2WPY is taken for constructing a contact map from which we can find the graph of the protein. Protein sequence and tertiary structure of the protein with PDB ID 2WPY is extracted from the database PDB.
3.2 Contact map of the Protein

Contact map of the protein 2WPY is obtained as a matrix of order $32 \times 32$ in which the entries are 0’s and 1’s where the entry 1 represents the distances between the two alpha carbons are less than or equal to 6Å and the entry 0 represents the distances between the two alpha carbons are greater than 6Å.

3.3 Graph of the protein

The graph of the protein 2WPY having 32 residues is obtained from its contact map as per the definition in section 2.2 and it is shown below for different $D_{cutoff}$ values.

![Figure 1: Graph of the Protein 2WPY (6Å cutoff)](image)

![Figure 2: Graph of the Protein 2WPY (5Å cutoff)](image)

3.4 Calculating Packing density of a protein

A mathematical conjecture by Kepler says that the maximum possible density amongst both regular and irregular arrangements is 74%. Here, protein packing is compared with a random sphere packing problem. Also, a random packing of equal spheres
Figure 3: Graph of the Protein 2WPY (7 Å cutoff)

(residues) has a density around 64%. Now, protein packing problem is to find the maximum number of contact residues for the protein within the particular threshold value. Contact between two carbon alpha atoms exists if the distance between two atoms is less than or equal to 6 Å. Maximum number of contacts for each residue (sphere) within the particular threshold value i.e., within the sphere of radius \( R = D_{\text{cutoff}} + r \), where \( r \) is the radius of the identical sphere (alpha carbon), is found from the approximate packing density of 64% for the random packing of equal spheres. Here, a residue is approximated as a ball of radius 1.5 Å to avoid overlapping.

Therefore, the maximum number of possible contacts for each residue is obtained as

\[
NC_{\text{mp}} = \frac{64\% \times 4/3\pi R^3}{4/3\pi r^3}
\]

Hence, the protein having \( N \) residues has maximum of \((N \times NC_{\text{mp}})/2\) number of contacts.

Thus, packing density of a protein is defined as \( P_d = \frac{|E(G)|}{N \times NC_{\text{mp}}/2} \), where \(|E(G)|\) denotes the number of edges in the graph derived from the protein. Hence, the packing density of the protein 2WPY is obtained

(i) With 6 Å cutoff as 0.4792.
(ii) With 5 Å cutoff as 0.2159.
(iii) With 7 Å cutoff as 0.3688.

4 Conclusion

In this research, a new method is introduced to find the packing density of a protein from its contact map. Maximum number
of contacts for each residue within the particular threshold is approximated as the irregular packing of identical spheres within the sphere of newly calculated radii. Packing density is calculated from the graph of the protein using graph theoretical concepts and it is observed that the packing density of a protein $P_d$ decreases when the number of residues increases. Hence, the protein with minimum length has high packing density.

References


