Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Minireview

Protein contact maps: A binary depiction of protein 3D structures

Arnold Emerson Isaac *, Amala Arumugam

Department of Biotechnology, School of Bio Sciences and Technology, VIT University, Vellore-632014, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 4 January 2016
Received in revised form 25 June 2016
Available online xxxx

Keywords:
Protein structure prediction
Contact map
Complex network

ABSTRACT

In recent years, there has been a considerable interest in examining the structure and dynamics of complex networks. Proteins in 3D space may also be considered as complex systems emerged through the interactions of their constituent amino acids. This representation provides a powerful framework to uncover the general organized principle of protein contact network. Here we reviewed protein contact map in terms of protein structure prediction and analyses. In addition, we had also discussed the various computational techniques for the prediction of protein contact maps and the tools to visualize contact maps.

Contents

1. Introduction

Proteins are the most abundant organic molecules in living systems. These molecules are much more diverse in structure and function than other classes of macromolecules. At any given point of time, the living system contains thousands of proteins within a single cell and each with a unique function. Proteins play a wide array of roles in a cell or organism from enzymes to hormones. The shape of a protein is typically described using four levels of structural complexity: the primary, secondary, tertiary, and quaternary levels. For some proteins, a single polypeptide chain folded in its proper three-dimensional structure creates the final protein. Protein structures are complex systems with several tens, hundreds

* Corresponding author.
E-mail address: i_arnoldemerson@yahoo.com (A.E. Isaac).

http://dx.doi.org/10.1016/j.physa.2016.08.033
0378-4371/

Please cite this article in press as: A.E. Isaac, A. Arumugam, Protein contact maps: A binary depiction of protein 3D structures, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.033

or even thousands of residues, interacting with each other to help stabilize the tertiary structures so that specific functions can be realized *in vivo*. In this sense, the network modelling approach is suitable for characterizing and analysing protein structures, in which residues correspond to vertices of the networks, and interaction (or any other type of relationship) between residues represented as an edge linking the corresponding nodes.

Understanding complex systems often requires a bottom-up approach, breaking the system into small and elementary constituents and mapping out the interactions between these components. In the biological systems, networks emerge in many disguises, from food webs in ecology to various biochemical nets in molecular biology. In particular, the wide range of interactions between genes, proteins and metabolites in a cell are best represented by various complex networks. An alternative way in which to conceptualize and model protein structures is to consider the contacts between atoms in amino acid residues as a network of interactions irrespective of secondary structure and fold type.

So far, many studies have been carried out to investigate the protein structures as complex networks of interacting residues [1–7]. Recently, we had analysed the membrane protein structures in terms of complex networks [8–10] and also investigated the structural and functional critical residues in protein structures using the k-core decomposition algorithm [11]. There have been many more detailed reviews on protein as networks, protein dynamics and the linking of topological characteristics to the protein folding [12–16]. In this review, we represent how the protein contact maps enable us to study the protein structure prediction, various tools for analysis and prediction of contact map. Protein 3D structure coordinates can be represented as a more reduced form called protein contact maps. It is possible to reconstruct the 3D coordinates of a protein using its contact map by several computational techniques [17,18].

2. Construction of protein contact map

Fig. 1 represents the steps in constructing the protein contact maps. The Cα atom of each amino acid has been considered as vertices of the corresponding protein contact network as shown in Fig. 1(a). The distances between every pair of residue were determined using Euclidean distance and the part of the distance matrix is shown in Fig. 1(b). The diagonal line in the distance matrix is always zero since the distance between the same residues is zero. To determine whether any two residues are connected, the distance between the residues should be less than or equal to the cut-off value 7 Å distance. The choice of cut-off distance was based on the range at which non-covalent interactions, which are responsible for the polypeptide chain to fold into its native-state. Various cut-offs ranging from 5 Å [7], to 7 Å [8], to 8.5 Å [6] have been used in earlier studies. The protein contact map was derived using the said cut-off value represented in 2-dimensional binary matrix (Fig. 1(c)). If any two residues are connected, then the matrix cell values are set to 1 (black colour) or else 0 (white colour) if they are not connected (Fig. 1(d)). In recent years, there had been several tools developed to analyse protein contact maps as shown in Table 1.

3. Protein structures as complex networks

Several researches have focused on modelling biological systems to be of a complex nature and more specifically to possess complex networks [34–43]. Among all biological systems such as the cells, tissues, or even the human body, 'proteins' are considered to be one of the most important macromolecules. Proteins perform a vast array of functions within living organisms including catalyzing metabolic reactions, DNA replication, responding to stimuli, and transporting molecules from one location to another. All these make proteins a very interesting system to study as a 'complex dynamical system'.

Traditionally, the three-dimensional folds of proteins have been perceived as a construct, based on elements of secondary structure and fold arrangement [44–47]. An alternative way in conceptualizing and modelling protein structures is to consider the contacts between atoms in amino acids as a network of interactions, irrespective of secondary structures and fold type. There is a natural distinction of contacts into two types: long-range and short-range interactions. Long-range interactions occur between residues that are distant from each other in the primary structure but situated at a much closer distance in the tertiary structure. These interactions are important for defining the overall topology. Short-range interactions occur between residues that are local to each other in both the primary, secondary and tertiary structures.

For most networks what is termed as a node and a link is fairly straightforward. When looking at protein transition states, several studies have considered the Cα atoms to be the nodes, and established a link between two nodes if the atoms were within 8.5 Å of each other [3,48]. In chemical terms, however, this is a simplification of the interactions within a protein. Side-chains play the pivotal role in forming and fixing a protein structure, and any information on their orientation is lost in an analysis based solely on Cα atoms. Another study conducted has used native structures considering each amino acid to be a node, and a link established between the two nodes when any two atoms from two amino acids are within a distance of 5 Å from each other [4]. These efforts deflect the studies on protein structure as complex networks representing amino acids as nodes and interactions between them as edges. Analyses of protein 3D structures as a complex network approach help us to understand in many different aspects, including its structural flexibility, key residues stabilizing its 3D structures, folding nucleus, important functional residues, mixing behaviour of the amino acids and hierarchy of the structure, etc. [3,7,48–51]. Several mathematical principles in network systems have been implicated to analyse protein contact networks. This principle includes bell-shaped Poisson distribution, small-world, scale-free, betweenness centrality and degrees of separation. Results revealed that protein structures have small-world properties and thus protein contact networks have a high clustering coefficient C with a relatively short characteristic path length L [3–6]. This concept was

Please cite this article in press as: A.E. Isaac, A. Arumugam, Protein contact maps: A binary depiction of protein 3D structures, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.033