Computational Biology and Chemistry 47 (2013) 16–23

Contents lists available at SciVerse ScienceDirect

Computational Biology and Chemistry

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

Research article

Computational structure analysis of biomacromolecule complexes by interface geometry

Sedigheh Mahdavi[a,b,*,1], Ali Salehzadeh-Yazdi[a,1], Ali Mohades[b], Ali Masoudi-Nejad[a,**]

[a] Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran
[b] Laboratory of Algorithms and Computational Geometry, Department of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 18 November 2012
Received in revised form 11 June 2013
Accepted 12 June 2013

Keywords:
Alpha complex
Computational geometry
Protein–nucleic acid complexes
Protein–protein complexes
Voronoi diagram

ABSTRACT

The ability to analyze and compare protein–nucleic acid and protein–protein interaction interface has critical importance in understanding the biological function and essential processes occurring in the cells. Since high-resolution three-dimensional (3D) structures of biomacromolecule complexes are available, computational characterizing of the interface geometry become an important research topic in the field of molecular biology. In this study, the interfaces of a set of 180 protein–nucleic acid and protein–protein complexes are computed to understand the principles of their interactions. The weighted Voronoi diagram of the atoms and the Alpha complex has provided an accurate description of the interface atoms. Our method is implemented in the presence and absence of water molecules. A comparison among the three types of interaction interfaces show that RNA–protein complexes have the largest size of an interface. The results show a high correlation coefficient between our method and the PISA server in the presence and absence of water molecules in the Voronoi model and the traditional model based on solvent accessibility and the high validation parameters in comparison to the classical model.

1. Introduction

Protein–nucleic acid and protein–protein interactions are a critical phenomenon in many biological mechanisms. DNA–protein interactions play an essential role in many cellular mechanisms such as DNA replication, transcription and nucleosome remodeling (Cartharius et al., 2005; Johnson and McKnight, 1989; Kamei et al., 1996; Sancar et al., 2004). RNA–protein interactions are involved in several regulatory processes such as RNA splicing, transport, replication, translation and post-transcriptional control (Glisovic et al., 2008). Protein–protein interactions are vitally important in a range of cellular processes, including DNA replication and transcription, RNA splicing, signal transduction and metabolic networks (Wu et al., 2007). Identification of macromolecular interaction sites is significant for understanding macromolecular functions and drug design studies (Bahadur et al., 2008). Various studies have been carried out on the structural analysis of biomacromolecule interaction interfaces. Varshney et al. (1995) extracted the interface surface from the power diagram of atoms of the molecule in protein–protein complexes (Varshney et al., 1995). Lo Conte et al. (1999) analyzed the atomic structure of the recognition sites of protein–protein interactions which contain the size and chemical properties of a protein-buried surface at the interfaces, packing density of atoms and polar interactions through hydrogen bonds and interface water molecules (Lo Conte et al., 1999). They calculated the Voronoi volume of buried atoms at the interface and compared it with buried atom inside proteins. Samudrala and Moult (1998) suggested all-atom distance-dependent discriminatory function for the prediction of nucleic acid binding proteins (Samudrala and Moult, 1998). Moont et al. (1999) used pair potentials for the screening of predicted docked protein–protein complex (Moont et al., 1999). Allers and Shamoo (2001) developed and used the program ENTANGLE to study RNA–protein interactions (Allers and Shamoo, 2001). Glaser et al. (2001) studied the residue frequencies and pairing preferences at protein–protein Interfaces (Glaser et al., 2001). Nadassy et al. (2001) evaluated the atomic volumes of DNA portion in DNA–protein interfaces (Nadassy et al., 2001). Chakrabarti and Janin (2002) showed that the protein–protein interaction recognition sites have an amino acid composition similar to the overall protein surface (Chakrabarti and Janin, 2002). Jeong et al. (2003) have analyzed the direct and water-mediated hydrogen-bonding properties of RNA–protein complexes (Jeong

* Corresponding author at: Laboratory of Algorithms and Computational Geometry. Department of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran.
** Corresponding author. Tel.: +98 21 6695 9256; fax: +98 21 6640 4680.
E-mail addresses: s_mahdavi@aut.ac.ir (S. Mahdavi), amasoudin@ibb.ut.ac.ir (A. Masoudi-Nejad).
URL: http://LBB.ut.ac.ir (A. Masoudi-Nejad).
[1] These authors contributed equally to the manuscript.

1476-9271/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiolchem.2013.06.003

et al., 2003). Ma et al. (2003) studies the distinguished properties of conserved residues between the binding sites and exposed protein surfaces at protein–protein interfaces (Ma et al., 2003). Ahmad et al. (2004) suggested the usage of moment information in the prediction of DNA-binding proteins (Ahmad and Sarai, 2004). Ban et al. (2004) computed the interface surface based on geometric topology (Ban et al., 2004). Rodier et al. (2005) analyzed the water molecules, immobilized at the protein–protein interfaces (Rodier et al., 2005). Siggers et al. (2005) introduced a method that describes the interface geometry in terms of the spatial relationships between the individual amino acid-nucleotide pairs in DNA–protein interfaces (Siggers et al., 2005). Terribilini et al. (2006) predicted the amino acids of an RNA binding protein, participating in RNA–protein interactions (Terribilini et al., 2006). Cazals and Proust (2006) defined the protein–protein interface surface based on the alpha complex and presented geometric and topological descriptions (Cazals et al., 2006). Kim et al. (2006), instead of using a power diagram, proposed another method based on the Voronoi diagram of atoms (Kim et al., 2006). Robertson and Varani (2007) developed all-atom statistical potential function for the prediction of DNA–protein interactions from their structures (Robertson and Varani, 2007). Headd et al. (2007) generated the protein–protein interface using computational geometry and topological tools and used the statistical analysis for its residue composition (Headd et al., 2007). Darnell et al. (2007) identified the binding of hot spots in protein–protein interaction automated decision-tree approach (Darnell et al., 2007). Bahadur et al. studied the size of RNA–protein interaction interface and its composition (Bahadur et al., 2008). Gao and Skolnick (2008) extended a knowledge-based method, which could perform the DNA-free (apo) version of DNA-binding protein prediction (Gao and Skolnick, 2008). Maetschke and Yuan (2008) examined graph theoretic properties of the residue contact maps, derived from the protein structures, to improve the prediction RNA–protein binding sites (Maetschke and Yuan, 2008). Bouvier et al. (2009) introduced shelling order of Voronoi facets as a measure for an atom's depth inside the protein–protein interaction interface (Bouvier et al., 2009). Zhou and Yan (2010) used the alpha shape to represent the surface structure of the DNA–protein complex and introduced an interface-atom curvature-dependent conditional probability discriminatory function for the prediction of DNA–protein interaction (Zhou and Yan, 2010). Lise et al. (2011) predicted the hot spot residues at protein–protein interfaces, using the method of support vector machines (Lise et al., 2011). Mahdavi et al. (2012) computed the RNA–protein interaction interface surface via the weighted Voronoi diagram of atoms and two filter operations (Mahdavi et al., 2012).

In this study, we propose an algorithm to compute the biomacromolecule interaction interface surface by using weighted Voronoi at the atomic level. The interaction interface surface is defined by removing the facets based on two lose accessibility surface and alpha-neighbor Voronoi filters. We aim to improve our previous method (Mahdavi et al., 2012) by applying lose an accessibility surface filter; we can define the accessibility surface filter that is present on the solvent accessible surface of each biomacromolecule in a complex. While the Voronoi atoms (that are not in the classical interface) are decreased by applying lose an accessibility surface filter. The lose accessibility surface filter uses the alpha complex which is a better approximation of the concave shape of the biomacromolecule and compute solvent accessible surface of each biomacromolecule in the interaction complex via 3D alpha complex that is a valuable geometric method for protein-surface analysis (Albou et al., 2008). Earlier studies of molecular structure have used alpha complex for many applications such as void measurement, automatic search of ligand binding sites, defining and characterizing the protein surface, molecular shape characterization and the representation of protein contact interactions (Albou et al., 2008; Kasson et al., 2007; Peters et al., 1996; Wilson et al., 2009; Li et al., 2003). Furthermore, we have compared the data sets of three different biological complexes (RNA–protein, DNA–protein and protein–protein complexes) interface, considering the role of water molecules in the interaction complexes, computing the volume of protein atoms at the biomacromolecule interaction interface, their packing density index and biophysical property of amino acids that involved in complexes interface.

2. Materials and methods

2.1. Dataset of biomacromolecule complexes

One hundred and eighty structures of known protein–nucleic acid and protein–protein complexes are extracted from the Protein Data Bank (PDB), listed in Table 1. Table 1 is divided into three classes: class A has sixty known RNA–protein complexes; class B has sixty known DNA–protein complexes, and class C contains sixty known protein–protein complexes.

2.2. Background knowledge

2.2.1. The Voronoi and Delaunay triangulation

Suppose that we have a finite set P of points in R^3, the ordinary Voronoi diagram can be represented $VD = (V,E,F,C)$ where V, E, F, and C are the sets of Voronoi vertices,Voronoi edges Voronoi faces, and Voronoi cells respectively. Each ordinary Voronoi cell of $p \in P$ can be formulated by $V_p = \{x \in R^3 | \forall q \in P - \{p\}, ||x - p|| \leq ||x - q||\}$. It consists of the part of space closest to its corresponding point. The ordinary Voronoi diagram of P is constructed by generating all cells of points $p \in P$. In R^3, an ordinary Voronoi facet lies in the bisector plane perpendicular to the line segment joining two points; a ordinary Voronoi edge is equidistant from three points and its vertices are located at the centers of empty spheres passing through four points in the P. The space filling model is a geometric model of the biomacromolecule. Atoms of biomacromolecule are represented by the amount of space that they occupy in this model. Each atom is modeled by a sphere in three-dimensional space where the radius of the sphere is the van der Waals radius of the corresponding atom. Geometric model of the biomacromolecule is constructed by the union of spheres (balls) of all atoms in biomacromolecule (Connolly, 1983; Edelsbrunner, 1995; Lee and Richards, 1971). Since the atoms have different van der Waals radii, the ordinary Voronoi diagram can't be used in a biomacromolecule geometric model and treats all atoms as spheres with same radii. The Weighted Voronoi diagram $V(P)$ is referred as the power diagram was usually used in this model which can provide method to accommodate different radii (Aurenhammer, 1987). In the power diagram, the Euclidean distance is changed by the power distance with respect to a sphere where the power distance is length of the tangent line segment to the surface of a sphere. The power diagram decomposes the union of spheres into convex regions. The dual structure of the Voronoi diagram is the Delaunay triangulation. The regular triangulation is the dual structure of power diagram. The regular triangulation decomposes union of spheres into nonoverlapping tetrahedral vertices are sphere centers. Whenever two power diagram cells share a common facet then the edge linked by two nearest corresponding spheres belongs to the regular triangulation. Similarly whenever three or four power diagram cells share a common edge or vertex respectively then the triangle or tetrahedron spanned by three or four nearest corresponding spheres belongs to the regular triangulation respectively. This efficient geometrical approaches, Voronoi diagram and Delaunay triangulation and various extended versions of them, are used in various biology