Chapter 26
Visualization of Protein 3D Structures in 'Double-Centroid' Reduced Representation:
Application to Ligand Binding Site Modeling and Screening

Vicente M. Reyes
School of Life Sciences, College of Science, Rochester Institute of Technology, Rochester, USA

Vrunda Sheth
School of Life Sciences, College of Science, Rochester Institute of Technology, Rochester, USA

ABSTRACT

This article is of two parts: (a) the development of a protein reduced representation and its implementation in a Web server; and (b) the use of the reduced protein representation in the modeling of the binding site of a given ligand and the screening for the model in other protein 3D structures. Current methods of reduced protein 3D structure representation such as the Cα trace method not only lack essential molecular detail, but also ignore the chemical properties of the component amino acid side chains. This chapter describes a reduced protein 3D structure representation called "double-centroid reduced representation" and presents a visualization tool called the "DCRR Web Server" that graphically displays a protein 3D structure in DCRR along with non-covalent intra- and intermolecular hydrogen bonding and van der Waals interactions. In the DCRR model, each amino acid residue is represented as two points: the centroid of the backbone atoms and that of the side chain atoms; in the visualization Web server, they and the non-bonded interactions are color-coded for easy identification. The visualization tool in this chapter is implemented in MATLAB and is the first for a reduced protein representation as well as one that simultaneously displays non-covalent interactions in the molecule. The DCRR model reduces the atomicity of the protein structure by ~75% while capturing the essential chemical properties

DOI: 10.4018/978-1-60960-491-2.ch026

of the component amino acids. The second half of this chapter describes the application of this reduced representation to the modeling and screening of ligand binding sites using a data model termed the "tetrahedral motif." This type of ligand binding site modeling and screening presents a novel type of pharmacophore modeling and screening, one that depends on a reduced protein representation.

INTRODUCTION

There exist different methods of protein 3D structure representation and visualization methods, the most popular of which being the all-atom representation (AAR), ribbon or 'spaghetti' representations, and space-filling models (please see refs. Sayle & Milner, 2000; DeLano, 2002; Guex, et al., 1999; Schwede, et al., 2003; Richardson & Richardson, 1992). The AAR model such as the wireframe and ball-and-stick models display every atom of the protein. But, even though all chemical information of the component amino acid residues are accounted for, the display is too crowded and overwhelming. On the other hand, van der Waals (VDW) surface representations such as the space-filling model are a good way to view the surface properties of the protein and locate shape complementarity involved in protein interactions but they fail to clearly show secondary structures, loops, functional sites and non-covalent interactions. Finally, ribbon and spaghetti models and the like provide a good view of the secondary structures and loops but do not show any side chain structural elements. In this paper we describe "double-centroid reduced representation" (DCRR), a reduced protein representation wherein amino acid residues in the protein are represented by two point coordinates: the centroid of the backbone atoms (N, Cα C' and O), and the centroid of the side chain atoms (CB and beyond). This method is similar to, but not identical to and independently conceived from, that proposed by Kolinski (2004) and Liwo, et al. (1997). In these two models, the Cα position is used instead of the centroid of the backbone atoms, and additionally in the Liwo, at al. (1997) method, a 'united peptide group' is inserted between two consecutive Cα atoms, to which the corresponding 'united sidechain group' is attached by a virtual bond.

We further develop a graphical visualization tool implemented in MATLAB that displays the reduced representation of the input protein PDB file, while simultaneously showing the intramolecular H-bonds and VDW interactions, as well as intermolecular ones with any bound ligands and water molecules. Our other aim was to develop a way of modeling ligand binding sites and to screen for these models in other proteins. We thought that this might find applications in pharmacophore modeling and screening that is quite different or even improved relative to current methods (Guner et al., 2004; Guner, 2005; Hopfinger, 2000; Khedkar et al., 2007; Mason, et al., 2001; Sun, 2008).

We thus proceeded to apply the DCRR method to the modeling of ligand binding sites (LBS) in proteins. Our LBS model is composed of the four most dominant amino acid centroids of the protein (in DCRR) which interact with the ligand atoms. These interactions may be in the form of hydrogen bonds or van der Waals interactions. The four centroids form a tetrahedron in 3D space, hence we term the model 'tetrahedral motif' model.

Finally we developed a screening method for the tetrahedral motif in any given protein, in order to predict whether the given protein would bind the ligand whose binding site tetrahedral motif is being sought. The screening procedure is composed of a series of Fortran programs that takes in two inputs, namely, a protein PDB structure file in DCRR, and the dimensions and centroid identities of the tetrahedral motif under query. The programs then either outputs the coordinates of four centroids in the protein that closely matches the tetrahedral motif if it finds one, or outputs null if it does not.