Fast and Accurate Calculation of Protein Depth by Euclidean Distance Transform

Dong Xu[1,2], Hua Li[3], and Yang Zhang[2]

[1] Bioinformatics & Systems Biology Program, Sanford-Burnham Medical Research Institute
dxu@sanfordburnham.org
[2] Department of Computational Medicine and Bioinformatics, University of Michigan
zhng@umich.edu
[3] Integration Application Center, Institute of Computing Technology,
Chinese Academy of Sciences
lihua@ict.ac.cn

Abstract. The depth of each atom/residue in a protein structure is a key attribution that has been widely used in protein structure modeling and function annotation. However, the accurate calculation of depth is time consuming. Here, we propose to use the Euclidean distance transform (EDT) to calculate the depth, which conveniently converts the protein structure to a 3D gray-scale image with each pixel labeling the minimum distance of the pixel to the surface of the molecule (i.e. the depth). We tested the proposed EDT method on a set of 261 non-redundant protein structures. The data show that the EDT method is 2.6 times faster than the widely used method by Chakravarty and Varadarajan. The depth value by EDT method is also highly accurate, which is almost identical to the depth calculated by exhaustive search (Pearson's correlation coefficient≈1). We believe the EDT-based depth calculation program can be used as an efficient tool to assist the studies of protein fold recognition and structure-based function annotation.

Keywords: Euclidean distance transform, fold recognition, molecular visualization, protein depth, protein tertiary structure, solvent accessibility.

1 Introduction

For a given protein tertiary structure, many residue level attributions can be extracted, such as the secondary structure type, dihedral angle and solvent accessibility. Those structural features help establish the properties of different amino acid types and categorize protein structure folds. For example, Ramachandran plot [1] revealed that the distribution of backbone dihedral angles (or the secondary structure) was highly regulated. Solvent accessibility (SA) evaluates the hydrophobicity of amino acids in different protein structures, which can be calculated accurately by EDTSurf [2] or approximately by DSSP [3].

However, SA usually specifies the residues in a binary form. For the residues that are completely buried in protein, it does not describe where the residues locate inside

M. Deng et al. (Eds.): RECOMB 2013, LNBI 7821, pp. 304–316, 2013.

the molecule. Depth, which measures the distance of each atom/residue to the solvent accessible surface in a continuous form, greatly complements the missing information by SA. In fact, the depths of residues in a protein are highly related to their effects of mutations on protein stability and on protein-protein interactions [4]. The residue depth has also been widely used to specify protein folds in protein structure prediction [5-7] and assist structure-based protein function annotation [8].

Despite the importance, by far there are very few methods which can calculate the depth for protein structures efficiently at either an atom level or a residue level. In Ref. [4], Chakravarty and Varadarajan proposed to calculate the residue depth by rotating the protein in a box where the closest water molecule is identified for each atom in the protein. The accuracy of the method is compromised since the calculated depth value depends on the positions of the water molecules. One can also calculate the depth by first generating the explicit solvent accessibility surface (e.g. by EDTSurf or MSMS [9]) and then identifying the vertex on the triangulated surface which is the closest one to the atom [10-11]. However, the computation of this kind of method is quite time-consuming since all the atoms in the protein need to be searched against the huge number of vertices on the surface.

In a recent study, we have established the relationships between the three kinds of macromolecular surfaces and Euclidean distance transform (EDT) theoretically and developed a fast algorithm for generating their triangulated surfaces precisely [2]. In this work, we apply the EDT technique to the calculation of protein atom depth and residue depth. The algorithm is fast since the explicit triangulated surface is not required. To investigate the efficiency and accuracy of this method, we compare the computational time and depth value with that by Chakravarty and Varadarajan (CV). We also analyze the relations of the depth with the commonly-used radius of gyration and solvent accessibility. The source code and executable program are freely available at http://zhanglab.ccmb.med.umich.edu/EDTSurf/.

2 Material and Method

2.1 Depth Definition

Atom depth is the shortest distance between the center of the atom and the outer solvent accessible surface (SAS) of the molecule, as illustrated in Fig. 1. SAS is the area traced out by the center of a probe sphere when it is rolled over the whole molecule [12]. When one atom is exposed (e.g. atom i in the figure), its depth will equal to the sum of the van de Waals radius and the radius of the probe sphere r_p which is often set to 1.4 Å. For atoms which are completely buried inside (e.g. atoms j and k in the figure), their solvent accessibilities are all equal to zero, but their depths may be different. Residue depth is the average value of the atom depths of all the atoms in a residue.

The definition of depth by Chakravarty and Varadarajan is a little different, which is the shortest distance to the explicit bulk water rather than the solvent accessible surface. Since water molecules don't have spherical shapes and may have different poses around the molecule, this difference will result in the slightly different depth values.