Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 1104–1112
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Pocket extraction on proteins via the Voronoi diagram of spheres

Donguk Kim [a], Cheol-Hyung Cho [b], Youngsong Cho [a], Joonghyun Ryu [a], Jonghwa Bhak [c], Deok-Soo Kim [d,*]

[a] *Voronoi Diagram Research Center, Hanyang University, Seoul, Republic of Korea*
[b] *Mechatronics and Manufacturing Technology Center, Samsung Electronics, Suwon, Republic of Korea*
[c] *National Genome Information Center (NGIC), KRIBB, Daejeon, Republic of Korea*
[d] *Department of Industrial Engineering, Hanyang University, Seoul, Republic of Korea*

Received 2 October 2007; accepted 2 October 2007
Available online 7 October 2007

Abstract

Proteins consist of atoms. Given a protein, the automatic recognition of depressed regions, called pockets, on the surface of proteins is important for protein-ligand docking and facilitates fast development of new drugs. Recently, computational approaches have emerged for recognizing pockets from the geometrical point of view. Presented in this paper is a geometric method for the pocket recognition which is based on the Voronoi diagram for atoms. Given a Voronoi diagram, the proposed algorithm transforms the atomic structure to meshes which contain the information of the proximity among atoms, and then recognizes depressions on the surface of a protein using the meshes.

Keywords: Voronoi diagram of atoms; Protein; Pocket; Binding site; Docking; Protein surface; Blending mesh

1. Introduction

Molecules such as protein, DNA, or RNA consist of atoms. Given the atomic structures of molecules, analyzing interactions between molecules is important for understanding their biological functions. An example is the interaction between a protein and a small molecule and this interaction is the basis of designing new drugs.

The study of molecular interactions, such as docking or folding, can be approached from a physicochemical and/or a geometrical point of view [30]. While the physicochemical approach is to evaluate and minimize the free energy between two molecules using, for example, the area of molecular surfaces, the geometric approach is to determine whether two molecules have geometrically meaningful features for interaction.

Interaction between a protein, called a *receptor*, and a small molecule, called a *ligand*, is usually done via some depressed regions, called *pockets*, on the surface of the receptor. Since the docking of chemicals into pockets to find an appropriate complementary ligand is essential in various aspects of drug design, identifying the potential pockets on a receptor is an important first step to efficient drug design. Considering the fact that chemical database usually contain a relatively large amount of data, automated methods are preferred for the generation of candidate pockets on receptors [24]. While the efforts on the physicochemical approach have been done since the early days of science, the efforts to understand the geometry of biological systems have started relatively recently [7,13,25,34].

Geometric approaches to recognize pockets on a protein usually involve the definition of surfaces on a protein. Initial studies on the problem primarily created a 3D spatial lattice of the space occupied by a protein and used simple techniques to reason the relative relations among the grids in the lattice to extract the exterior boundary of the protein, and then they recognized the depressed regions on the surface of the protein [6,14,15,36]. Recently, researchers have started to use more rigorously defined mathematical and computational tools related to the geometry among the atoms in a protein. α-shapes are one of the most successful efforts in this avenue. Since α-shapes can be used to represent the surface of a protein quite efficiently for a fixed size probe, it has been often used in the extraction of pockets [8]. Peters et al. and Liang et al. have even tried to explain the geometric characteristics between

* Corresponding author. Tel.: +82 2 2220 0472; fax: +82 2 2292 0472.
E-mail address: dskim@hanyang.ac.kr (D.-S. Kim).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.10.002

extracted pockets and ligands to fit into the pockets using α-shapes [26,31]. However, the α-shape has certain limitations for various applications mainly due to the fact that it is defined from the centers of atoms, not from the surfaces of atoms and therefore it cannot take into account size difference among atoms.

In this paper, we will introduce another methodology to provide the definition of pockets on the surface of a protein from the geometric point of view and present an algorithm to recognize pockets automatically. Given a protein, the proposed algorithm first computes the *Voronoi diagram* of the atoms and constructs a *mesh structure* on the surface of the protein using the Voronoi diagram and a probe. We emphasize here that the Voronoi diagram of atoms is different from the Voronoi diagram of atom centers. The mesh structure defines the spatial proximity among the atoms on the surface of the protein with respect to the probe. Then, we define a *pocket primitive* which corresponds to each planar face of the convex hull of the protein. After pocket primitives are extracted, we evaluate the validity of boundaries between neighboring pocket primitives to test if two neighbors should be merged into a single pocket or not. Eventually, therefore, there will be a few pockets left on the surface of a receptor where each pocket corresponds to an appropriately depressed region.

2. Geometric models of a protein and related terminologies

A protein, consisting of atoms, can be viewed as shown in Fig. 1. The circles denote atoms constituting a protein and their radii are van der Waals radii. In the model, there are two kinds of surfaces associated with the protein: *solvent accessible surface* (SAS) and *molecular surface* (MS). SAS consists of a set of points on the space where the center of the probe is located when the probe is in contact with the protein without intersection with other atoms. The inner-most possible trajectories of points on the probe surface, then, define MS. SAS usually defines a *free-space* that a small molecule can move around without interfering with the protein and therefore plays a fundamental role for folding and/or docking [25]. On the other hand, MS, often called by another name *Connolly surface* after the name of the first researcher who developed an analytical algorithm for MS, conveniently defines the boundary between the interior and exterior volume of a protein so that the volume or the density of a protein can be calculated [5].

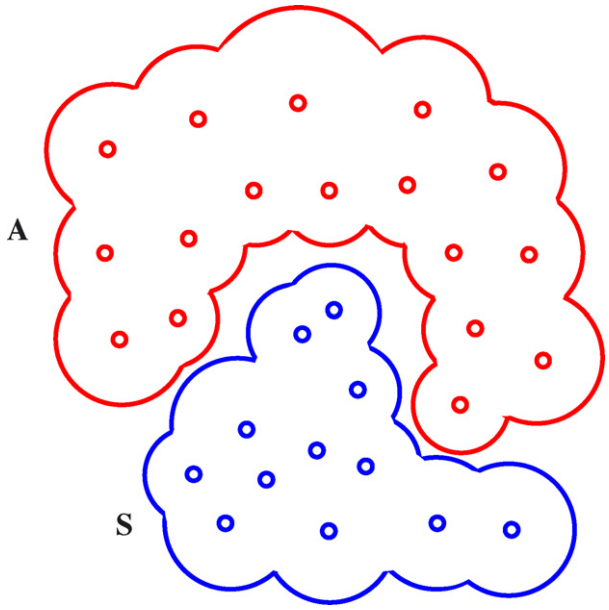

Fig. 2. A docking configuration between receptor A and ligand S.

Fig. 2 shows two molecules interacting with each other via a pocket defined on the molecular surface of molecule A. The molecule labeled A is a receptor and the small molecule labeled S is a ligand. Then, A and S interact with each other as the protruded region of S has been geometrically inserted into the depressed region, which is called a *pocket*, on the molecular surface of A.

Let $A = \{a_1, a_2, \ldots, a_n\}$ be a protein consisting of a number of atoms $a_i = (c_i, r_i)$ where $c_i = (x_i, y_i, z_i)$ and r_i define the center coordinates and the radius of an atom, respectively. In addition, suppose that $S = \{s_1, s_2, \ldots, s_m\}$ is a small molecule which also consists of a number of atoms s_j, defined similarly to a_i, and S will be docking with A. Note that $m \ll n$ in general since m is usually in the range of several dozens at most and n is between several hundreds and thousands. Usually a small molecule is approximated by the spherical probe $R = (c_R, r_R)$ and most investigations on the geometric properties for the

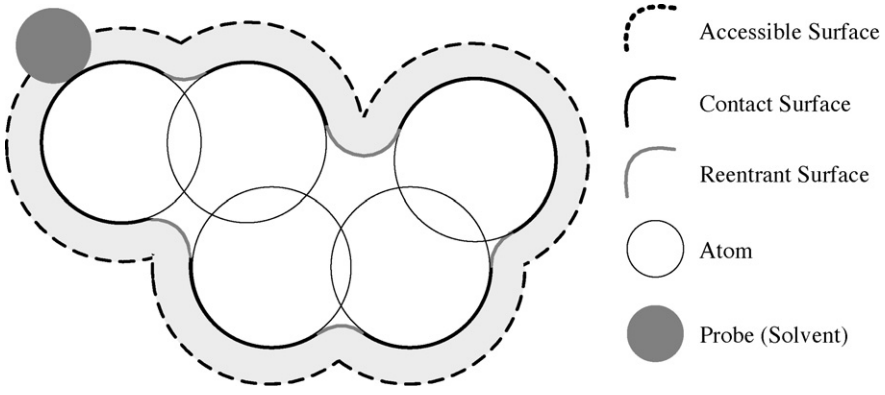

Fig. 1. The geometric model of a protein. The reentrant surface and the contact surface altogether define the molecular surface.