doi:10.1016/j.jmb.2006.07.022 J. Mol. Biol. (2006) 362, 441–458

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Travel Depth, a New Shape Descriptor for Macromolecules: Application to Ligand Binding

Ryan G. Coleman and Kim A. Sharp*

The Johnson Research Foundation and Department of Biochemistry and Biophysics University of Pennsylvania Philadelphia, PA 19104, USA

Genomics and Computational Biology Graduate Group University of Pennsylvania Philadelphia, PA 19104, USA

*Corresponding author

Depth is a term frequently applied to the shape and surface of macromolecules, describing for example the grooves in DNA, the shape of an enzyme active site, or the binding site for a small molecule in a protein. Yet depth is a difficult property to define rigorously in a macromolecule, and few computational tools exist to quantify this notion, to visualize it, or analyze the results. We present our notion of travel depth, simply put the physical distance a solvent molecule would have to travel from a surface point to a suitably defined reference surface. To define the reference surface, we use the limiting form of the molecular surface with increasing probe size: the convex hull. We then present a fast, robust approximation algorithm to compute travel depth to every surface point. The travel depth is useful because it works for pockets of any size and complexity. It also works for two interesting special cases. First, it works on the grooves in DNA, which are unbounded in one direction. Second, it works on the case of tunnels, that is pockets that have no "bottom", but go through the entire macromolecule. Our algorithm makes it straightforward to quantify discussions of depth when analyzing structures. High-throughput analysis of macromolecule depth is also enabled by our algorithm. This is demonstrated by analyzing a database of protein-small molecule binding pockets, and the distribution of bound magnesium ions in RNA structures. These analyses show significant, but subtle effects of depth on ligand binding localization and strength.

Keywords: depth; molecular surface; ligand binding; structural genomics; computational geometry

Introduction

Depth is a term frequently applied to the shape and surface of macromolecules. For example, enzyme active sites are routinely described as shallow or deep. Small ligand binding sites on proteins are also frequently described in term of depth. Depth is just one facet of the property "binding pocket shape" one would like to define quantitatively, to aid for example, in screening a large library of potential ligands, or in docking of a candidate ligand. Groove depth is one of the fundamental terms used to describe the differences in structure of the *A*, *B* and *Z* forms of DNA.[1–3] In spite of the common use of the term depth, it is a surprisingly difficult property to define rigorously in a macromolecule. Discussions of depth in the literature, although intuitively reasonable, are usually qualitative. The concept of depth is thus difficult to subject to rigorous analysis or to extract the most information from. A large part of the difficulty in analyzing depth is due to the complexity and range of shapes adopted by macromolecules. Protein surfaces are fractal in nature,[4] adding to the difficulty. To illustrate some of the difficulties, consider first the issue of a reference point or level. In geodesy, mountain peaks and ocean depths are referenced to the mean sea level, providing a standard reference level (although not without regional difficulties: mean sea level either side of the Panamanian isthmus differs considerably, for example). There is no equivalent to mean sea level in a molecule. Second, consider the case of deep pockets involving overhangs or that re-approach the molecule surface at some point away from their origin. Euclidean distance of the bottom of the pocket to the nearest surface, while easy to define and compute, will be a very misleading and grossly underestimating measure of depth. These difficulties are reflected

Abbreviation used: PDB, Protein Data Bank.

E-mail address of the corresponding author: sharpk@mail.med.upenn.edu

in the fact that there are few computational tools to quantify the concept of depth, to visualize it, or analyze the results. To address this problem, we present here our notion of travel depth, simply put the physical distance a solvent molecule would have to travel from a surface point to a suitably defined reference surface. The concept of travel depth was designed to avoid the "short circuiting" error described above, and also to solve the problem of a reference level. We first define the concept of travel depth, and the reference level used by it, then present a fast, robust approximation algorithm to compute travel depth to every surface point. Selected examples using very different molecular shapes are used to demonstrate that our definition of depth works for special cases, and that it conforms to our intuition, so confirming that we have introduced a "good" definition for depth and that our approximate numerical implementation of it is reasonable. We then describe some applications of our algorithm, including a high throughput application to a small molecule binding database.

Theory and Algorithm

Definition of travel depth

Any measure of depth must start with the questions: Depth of what, and from what? Here, we are concerned with the depth of any point on the molecule's surface. Two definitions of surface predominate for macromolecules, the solvent accessible surface,[5] and the molecular surface.[6] In both cases a crucial parameter is the probe radius, which is almost universally taken to be that of water (usually values between 1.4 Å and 1.8 Å are used). Many algorithms exist for computing these idealized surfaces. Most, but not all, produce a triangulated form of the surface, primarily for display using standard computer graphic routines.[7–10] Our algorithm assumes a simple closed triangulated surface. The surface must be orientable and connected, though these are not strong requirements; the latter disallows only cavities. For the broadest applicability of our method, we make no other assumption about how the surface was produced, or what it should look like. In practice we use the molecular surface as generated by the algorithm in the GRASP macromolecular graphics program[10] implemented as a stand-alone program[11] using a probe radius of 1.8 Å and standard atomic radii.[12] Though we test only this surface generation scheme and the resulting triangulated surfaces, our definition and algorithm generalize to any triangulated surface generation scheme.

Our definition of travel depth is that for each point on the molecular surface, the travel depth is the minimum distance a solvent probe would have to travel through the solvent from that surface point to get to the reference level. A natural and parameter-independent reference level is provided by the convex hull of the molecular surface. The convex hull is a standard construct in computational geometry. In three dimensions, the convex hull is the smallest volume convex polyhedron that contains all the surface points.[13–15] In terms of molecular surfaces, the convex hull is equivalent to the molecular surface produced by an infinite solvent probe radius. Algorithms and code for convex hull computation have been well studied and are fast and reliable.[13–15]

The next step is to compute the minimal distance from every surface point to the convex hull while respecting the boundary of the molecular surface. In other words, the travel path along which the distance is computed must lie outside the molecular surface in the solvent. We note that computing such a minimal distance between two points while avoiding obstacles is exactly the shortest path planning problem commonly encountered in robotics, and that an exact solution to the problem is NP-hard. Our solution, described below, is to approximate this minimal distance in such a way that it was easy to code and run in a short time so that we could establish what the depth measure would look like on real examples, and whether it would be useful in structural analysis.

Calculation of travel depth: preprocessing

The first step is to remove cavities, defined as completely enclosed solvent pockets in the molecular surface. The triangles that represent these cavities are removed from the surface and are not used in later calculations. Since there is no way for the solvent probe to travel from a closed cavity surface to the convex hull without passing through the macromolecule itself, travel depth does not apply to these surfaces. We note, though, that simple Euclidean distance to the nearest part of the external molecular surface would provide a satisfactory definition of the minimum depth of a closed cavity.

Two important pre-processing steps are done at this stage. First, the longest edge of any triangle in the surface is found and the length saved for later. Also, all the points on the surface are put into a two-dimensional orthogonal range search tree structure oriented along one grid axis.[14] This helps improve the running time, as described later, but it is non-essential to the algorithm.

Calculation of travel depth: mapping onto grid

The macromolecule and a region of the surrounding solvent are embedded in a cubic grid of dimensions $K \times L \times M$. For convenience, the grid extends to one grid cube beyond the minimum and maximum coordinate of the molecular surface in each orthogonal direction, so that the border is completely outside the surface. The default grid spacing used in our algorithm is 1 Å, however the algorithm and code generalize to any spacing. The only consideration is for the spacing to be small enough to approximate well the topology of the