SHORT COMMUNICATION

Identifification of Substrate Channels and Protein Cavities

Thomas Exner [1], Matthias Keil [1], Gerd Moeckel [2], and Jürgen Brickmann[1,3]

[1]Department of Physical Chemistry, Darmstadt University of Technology, Petersenstraße 20, D-64287 Darmstadt, Germany. E-mail: brick@pc.chemie.tu-darmstadt.de

[2]LION Bioscience AG, Im Neuenheimer Feld 515-517, D-69120 Heidelberg, Germany.

[3]Darmstadt Center for Scientific Computing, Darmstadt University of Technology, D-64287 Darmstadt, Germany.

Abstract A grid-based algorithm is presented for identifying and displaying channels and cavities in a given protein structure. A point on a regular Cartesian grid is defined to be located in a cavity or a channel if the point has neighbouring grid points in the direction of at least two Cartesian axes, which are located inside the molecule. The cavities or channels can be displayed as clusters of these grid points or as the part of the surface and the amino acid residues lying next to these grid points. The algorithm was tested successfully using the gramicidin A dimer, different cytochrome P450's, and the proteolytic enzyme trypsin.

Keywords Protein, Substrate channels, Cavities, Active sites

Introduction

Many signalling, recognition, and catalytic functions within a cell are performed through interaction of the active site of an enzyme with other molecules. These active sites tend to be concave or lie inside the enzyme, only accessible through a small channel. The size and shape of these channels and cavities play a crucial role in the binding specificity. Thus it is important to be able to locate and characterize them. Levitt and Banaszak introduced an interactive program for identifying, displaying, and manipulating cavities in a known protein structure [1]. Regions are defined as cavities where a test sphere can be placed without touching the center of any protein atoms and is surrounded by regions with protein contact. Similar approaches use test spheres which are adjusted in size until they touch the nearest atoms [2,3]. The cavities are defined as clusters of overlapping spheres. The method of Delaney uses a pattern-recognition technique based on cellular logic operations to distinguish between convex and concave regions of the protein [4].

In this work a new grid based method is presented. All grid points are defined as cavity points if they are surrounded by grid points lying inside the molecule in the direction of at least two Cartesian axes. These cavity points are combined to clusters building the cavities or channels of the protein structure.

Computational Methods

In this paper we use a picture where the molecules are represented by the solvent accessible surface introduced by Richards [5] and Connolly [6]. According to this definition a set of points representing the molecular surface is generated by rolling a test sphere (water molecule, radius 1.4 Å) along a CPK model. The proposed algorithm can be used to identify all surface points, building the channels and cavities and residues lying next to these points.

In this procedure, the molecule is placed in a regular Cartesian grid with a space width of 0.5 Å. Each grid point lying within the molecular surface is labelled as "in", all others are "out". The "out" labelled grid points are the selected ones. The neighbouring points of the selected point (up to a distance of 12 Å) in the positive and negative direction of each Cartesian axis are labelled. If there are at least two axes with neighbouring points labelled as "in", in the positive and negative direction, the center point is labelled as a "cavity point" (see Figure 1).

Cavity points are combined to clusters with all their neighbours, also labelled as cavity points. Two operations of the cellular logic, called contraction and expansion, are then performed on these clusters [4]. The contraction removes a monolayer of points from the border of the clusters. Any cavity point with one or more neighbours, labelled as noncavity points, is removed. Expansion is essentially the reverse of contraction. Any grid point with at least one neighbouring cavity point is added to the corresponding cluster. With these two operations, small clusters are removed, while big clus-ters, linked only by a small amount of grid points, are subdivided into smaller ones. The clusters are sorted by the number of grid points. Surface points and amino acid residues lying next to these clusters are identified and can be displayed for further investigations.

Results

The algorithm was used to identify and display the substrate channel, leading to the active site of different cytochrome P450's. The surfaces building the channel were subtracted from the molecular surface. The shape of the channel surface and physicochemical properties mapped on this surface, e.g. the electrostatic potential and the molecular lipophilicity, were compared to characterize the differences in the substrate specificity of the different P450's [7]. A comparative analysis of the molecular properties of different substrates placed into the active site of the access-model allowed to define common characteristics of nitroamines, which can be activated by P450 enzymes. N-Methyl-N'-nitrosopiperazine, for which only equivocal data on carcinogenicity are available, shows good agreement with the carcinogenic nitrosamine activated by P450 enzymes. With these examples the suitability of the active-site-access-model of cytochrome $P450_{cam}$ and the used concept of molecular surfaces as maps displaying calculated molecular properties for the investigation of enzyme-substrate interaction is demonstrated [7].

The active site of the proteolytic enzyme trypsin could also be identified within the three largest cavities of the pro-


Figure 1 *Identification of cavity points (schematically). The neighbor points (red and green) of the selected point (blue) in the direction of each Cartesian axis are identified. The selected point is labeled as a cavity point, if there are at least two axes with neighboring points labeled as "in" (green), in the positive and negative direction*
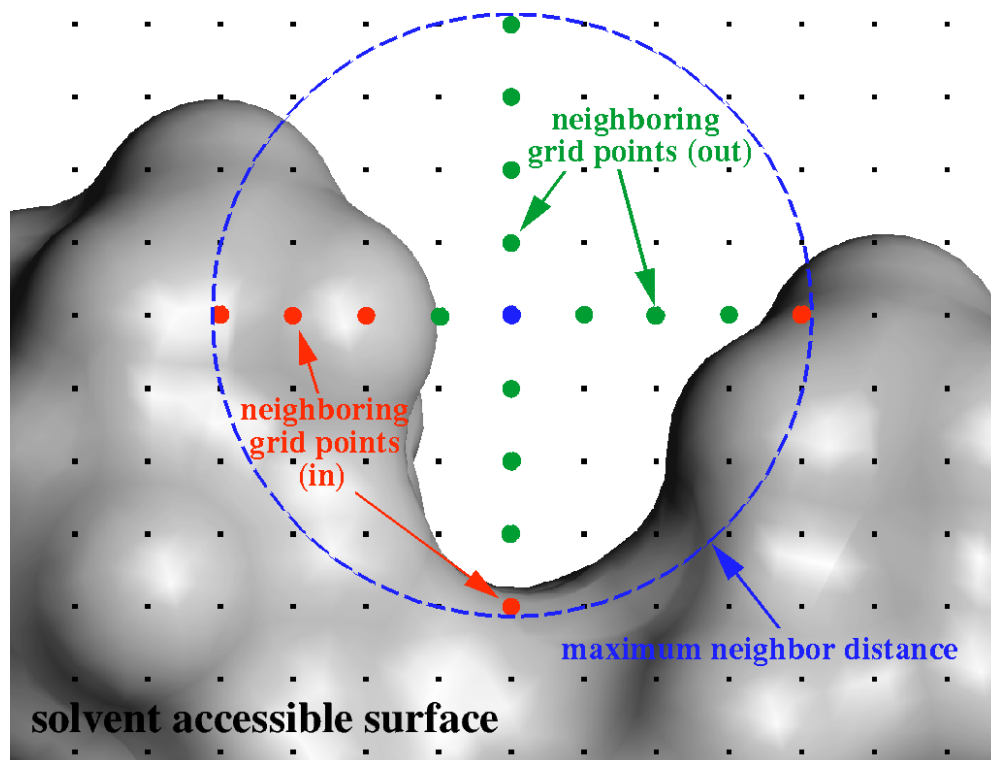
Figure 2 *Substrate channel of the cytochrome P450$_{cam}$. The cytochrome is shown in ribbon representation. The molecular lipophilicity is mapped on the solvent accessible surface building the channel*
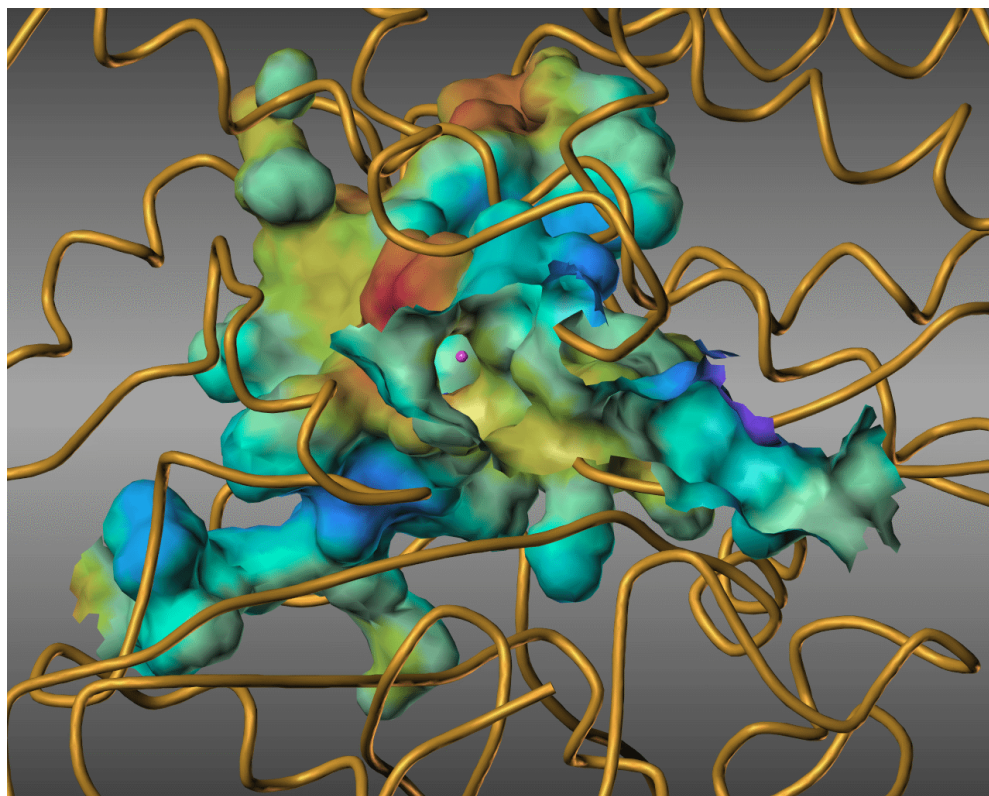
Figure 3 *Cavities of the enzyme trypsin. The molecular surface of the active site is shown in green with the amino acid residues building the active site*
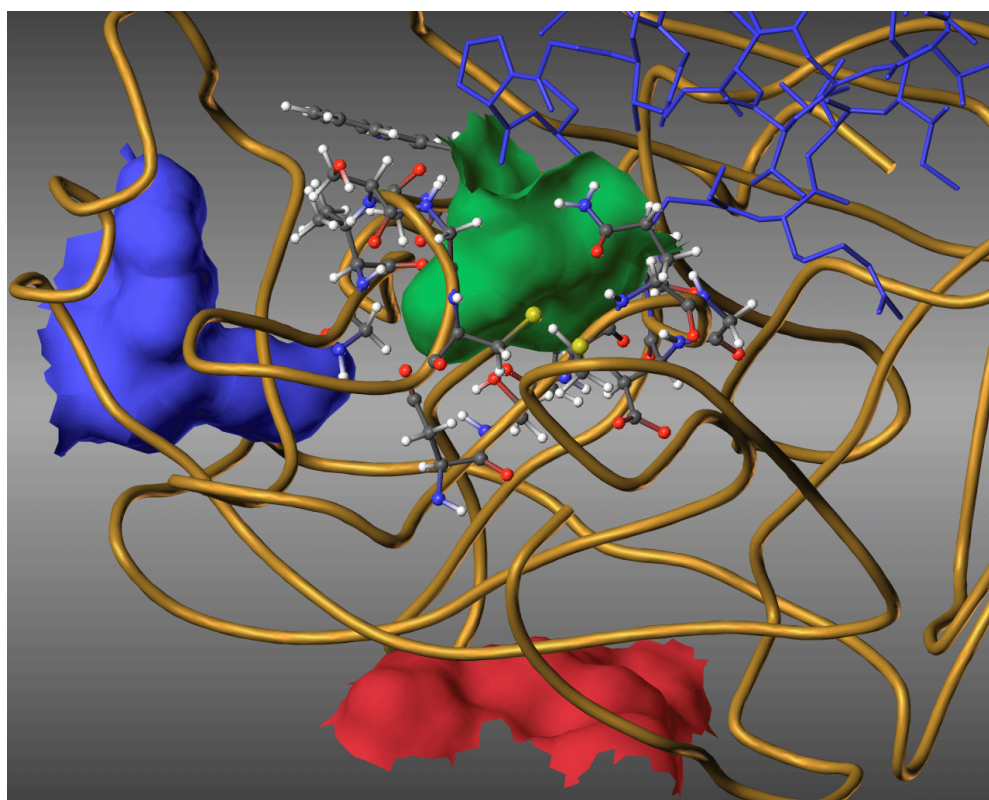
Figure 4 *Molecular lipophilicity mapped on the solvent accessible surface of the channel of the transmembrane protein gramicidine A dimer*
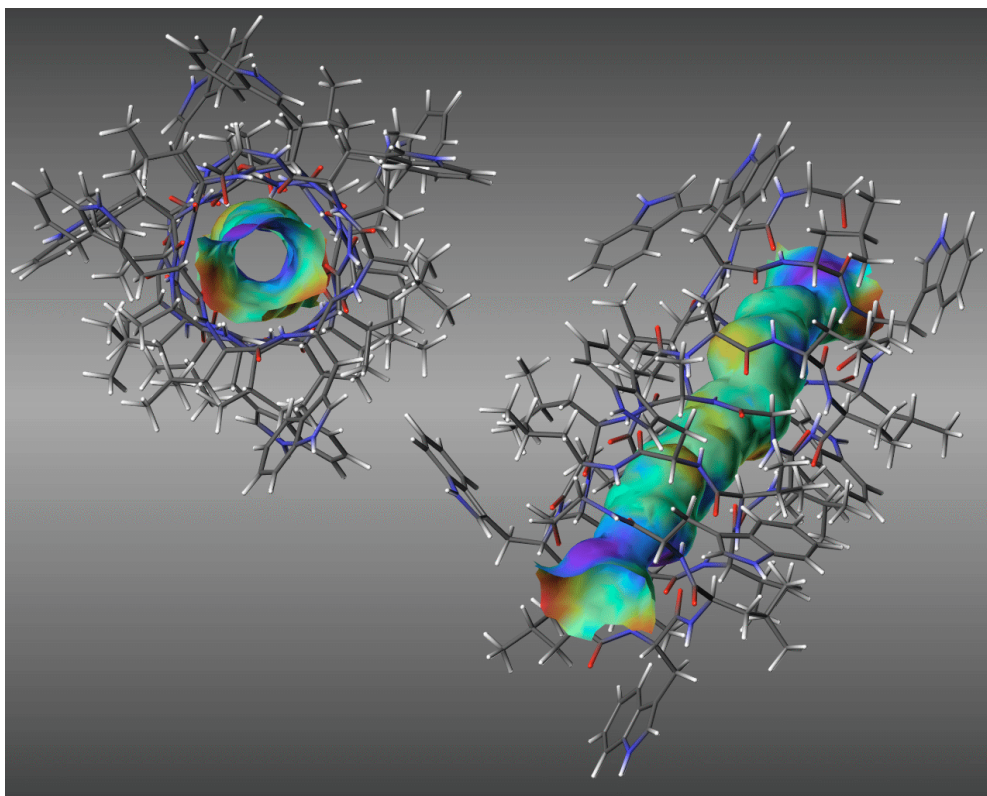
tein. A ligand could be placed in these cavities using a docking algorithm to generate possible enzyme-ligand-complexes.

The third example is the channel through the transmembrane protein gramicidine A dimer. The surface of the channel with the molecular lipophilicity is shown in figure 4

Supplementary Materials Available The cytochrome P450, the enzyme trypsin, and the transmembrane protein gramicidine A dimer are available in PDB-format as supplementary materials. The 3D-structure and the channel surface of P450 can also be inspected interactively with a VRML browser. By selecting the next viewpoint of the VRML-scene, a flight through the substrate channel to the active side of P450 can be started.

References

1. Levitt, D.G. and Banaszak, L.J. *J.Mol.Graphics* **1992**, *10*, 229-234.
2. Meng, E.C.; Shoichet, B.K.; and Kuntz, I.D. *J.Comp. Chem.* **1992**, *13*, 505-525.
3. Laskowski, R.A. *J.Mol.Graphics* **1995**, *13*, 323-330.
4. Delaney, J.S. *J.Mol.Graphics* **1992**, *10*, 174-177.
5. Richards, F.M. *Annu.Rev.Biophys.Bioeng.* **1977**, *6*, 151-176.
6. Connolly, M.L. *Science* **1983**, *221*, 709-713.
7. Moeckel, G.; Exner, T.; Keil, M.; and Brickmann, J. *Proceedings of the Pacific Symposium on Biocomputing 1998*, **1997**, 327-338.