

Research MAPS: Protein Docking Interfaces

by Andrew Ban, Paul L. Brown, Herbert Edelsbrunner, Jeffrey J. Headd, and Johannes Rudolph

The majority of intracellular events that affect life-processes are governed by protein-protein interactions. These include, for example, the cellular response to growth factors and hormones via the complex signaling network of protein kinases, phosphatases and G-proteins. We are interested in obtaining a better understanding of protein-protein complex formation by studying the characteristics of known protein complexes. There are no recognized universal rules that characterize these interactions at the molecular level, in contrast to much better understood systems such as gene regulation, enzymatic catalysis, and protein folding. In fact, we can compare the current knowledge of protein-protein interactions with the field of protein structure prior to the creation of the descriptors and classifications that have now become part of the standard language of protein folding. Specifically, the definition of α -helices and β -sheets and the determination of numerous protein crystal structures made it possible to visualize and classify proteins into families (e.g. β -barrels or β - α - β sandwiches). These classifications provide a set of rules that have led to important insights into protein function, protein folding, protein structure prediction, and evolutionary relationships. In an analogous manner, descriptors of protein interfaces based on composition, shape, and forces would clearly be fundamental to the protein-protein interaction community and provide a better understanding of how life-processes are regulated and how we can manipulate them for the benefit of human health. The MAPS website (<http://biogeometry.cs.duke.edu/research/docking/>) we describe here collects, analyzes, and publicizes the discovery, characterization, and visualization of such descriptors of protein interfaces.

Using tools from Computer Science and Mathematics we

have previously defined an interface surface between two or more complexed proteins [1] (see also BioGeometry News April 2004). The algorithm and software used to generate the interface surface was created with tools from computational geometry, including weighted Voronoi diagrams, weighted Delaunay triangulations, and topological persistence. Simply described, the interface surface is a wrinkly (non-flat) sheet halfway between two proteins, trimmed where it protrudes beyond the interface (Fig. 1, left panel). The advantage of working with the interface surface instead of the individual surfaces of interacting proteins is that it provides a single well-defined object that can be readily manipulated in calculations. Importantly, observations and results of further studies can be easily visualized, intuitively describing the na-

ture of protein interfaces and providing insight into existing experimental data.

In our approach, we define a hierarchy of structures on the interface surface, analogous to protein structures. We first define a primary structure that describes the adjacencies of the residues contributing to the interface. These adjacencies can be described both intra- and intermolecularly at the interface. The primary structure is best visualized in a flattened view. Next we define secondary structures, which are recurring geometric motifs. We believe that most interface surfaces consist of a few canonical motifs (equivalent to α -helices and β -sheets in protein structures). The landscape of these motifs forms the tertiary structure of an interface. Our goal is to characterize and classify protein interfaces through a deeper understanding of these three hierarchical interface structures as well as their interdependence. We also intend to study a variety of naturally defined functions on the interface surface including the distance to neighboring proteins and electrostatic potentials.

We use our Java Molecular Interface Viewer (JmiV) to analyze and disseminate our data for a large number of protein-protein complexes taken from the protein databank (PDB).

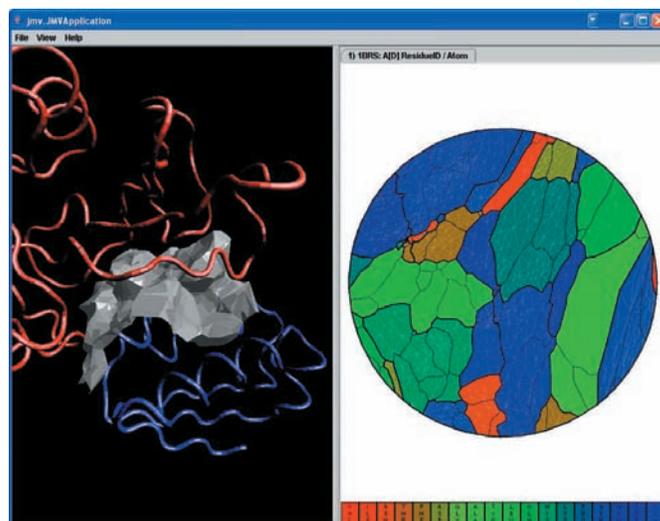


Fig. 1. Standard view of the protein interface on the MAPS website using 1BRS, the barnase/barstar protein complex as an example.

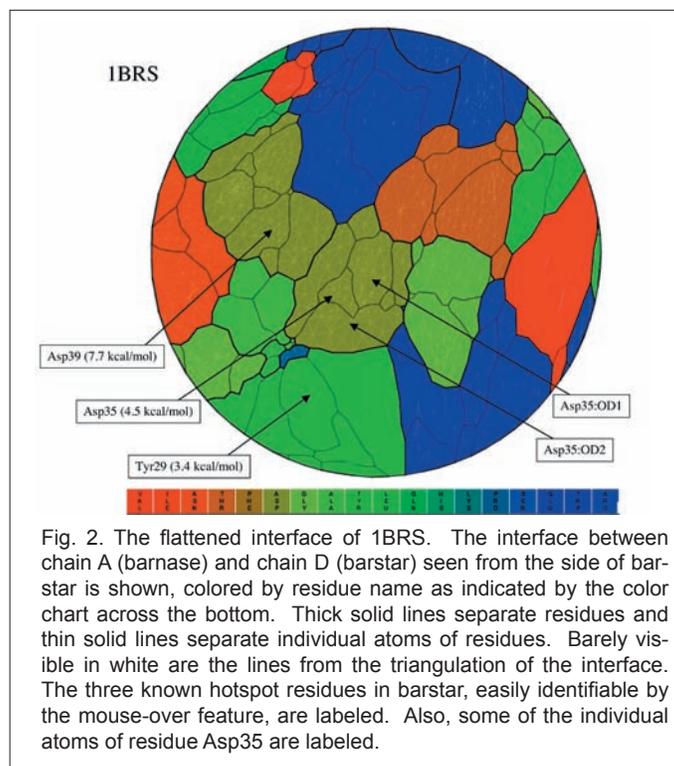


Fig. 2. The flattened interface of 1BRS. The interface between chain A (barnase) and chain D (barstar) is shown, colored by residue name as indicated by the color chart across the bottom. Thick solid lines separate residues and thin solid lines separate individual atoms of residues. Barely visible in white are the lines from the triangulation of the interface. The three known hotspot residues in barstar, easily identifiable by the mouse-over feature, are labeled. Also, some of the individual atoms of residue Asp35 are labeled.

JmiV is an extension of Java Molecular Viewer (JMV) originally developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. The default view of a protein interface shows two panels (Fig. 1). On the left, the interface is represented in 3D space imbedded within the original JMV visualization. On the right, the interface is shown in a planar parameterization that, although somewhat distorted, preserves the connectedness at the docking interface. Below we describe some of the features of our flattened view of the interface and encourage the reader to visit our website to explore this novel tool.

Tiling Paradigm. The interface surface consists of Voronoi polygons, each one being associated with a pair of atoms - one from each "side" of the docking. The parameterization preserves this atom-pairing association and thus makes it possible to colorize the visualization based on characteristics directly or indirectly related to atomic structure at the protein docking interface. By convention, contiguous polygons having the same characteristic value will be drawn with the same color and the boundary between differing characteristic values will be denoted with a black line. Polygons with identical characteristics are fused to form *tiles*. Usually, this means the polygons have the same color, but same color polygons can be in different tiles, e.g. if they are not contiguous.

Two levels of tiles can be visualized simultaneously - this is called Primary Tiling and Secondary Tiling. Colorization applies only to the characteristics chosen for the primary tiles. To help distinguish between primary and secondary tile boundaries, the thickness of the

tile-boundary lines differs - a primary tile boundary is thicker than a secondary tile boundary. In Fig. 2 the colorization is by residue type (default) with a palette of

Want to find a residue of interest? Pausing the mouse-pointer above the flattened interface will invoke the appearance of a tool-tip giving the characteristic values of the tile beneath. Also, this will always include information as to residue and atom beneath the pointer. Alternatively, click on one of the hotspot residues in the table below the graphics to select a residue for which biochemical data has been generated.

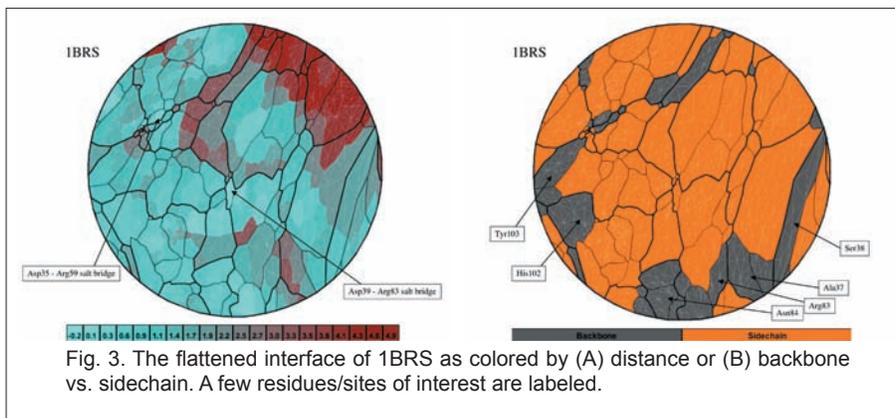


Fig. 3. The flattened interface of 1BRS as colored by (A) distance or (B) backbone vs. sidechain. A few residues/sites of interest are labeled.

color assignment provided at the bottom of the window and the secondary tiling is by atom type (no coloring).

Two Sides. The docking interface is two sided by nature and by default only one side is visualized at a time. It is easy to view the other side choosing "reverse" from the pop up menu. To view the two sides simultaneously, either clone the view you have and then change the side you are viewing, or view as merged.

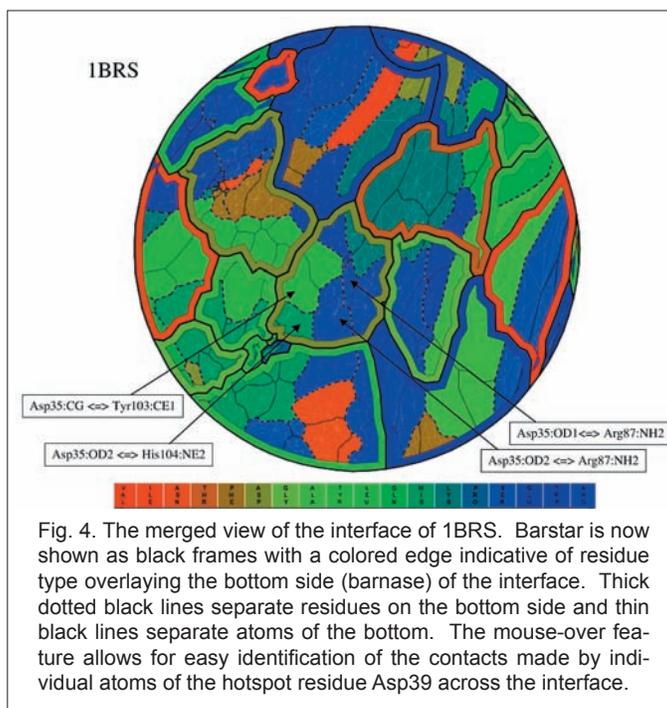


Fig. 4. The merged view of the interface of 1BRS. Barstar is now shown as black frames with a colored edge indicative of residue type overlaying the bottom side (barnase) of the interface. Thick dotted black lines separate residues on the bottom side and thin black lines separate atoms of the bottom. The mouse-over feature allows for easy identification of the contacts made by individual atoms of the hotspot residue Asp39 across the interface.

We allow the user to view many different colorings of the flattened interface, including residue name, atom type, electrostatics, hydrophobicity, backbone vs. sidechain, secondary structure, distance, temperature factors, and rank (Fig. 3). To change the coloring scheme, use the pop-up menu accessible with a right button mouse-click. If you want to see more than one coloring scheme at a time, "clone" your flattened interface by selecting "duplicate" from the pop-up menu.

The merged view allows for both sides of the flattened interface to be visualized simultaneously (Fig. 4). When merged, the "back" side remains essentially unchanged while the "front" side is reduced to frames that outline the tiles from the "front" side. These frames have the black boundary-edge plus a colored inner frame that corresponds to the tile coloring that would be used in non-merge mode. On the "back" side, primary tile boundaries will be rendered as dashed lines.

We welcome comments and recommendations from all users of our flattened interface view. Also, we are happy to add additional protein complexes, particularly if you know of any biochemical data to help us in our future structure/function analyses of protein-protein interfaces.

[1] Y.-E. A. Ban, H. Edelsbrunner, and J. Rudolph. Interface surfaces for protein-protein complexes. Proc. 8th Ann. Intl. Conf. Res. Comput. Mol. Biol., 2004, 205-212.