Visualizing the Dual Space of Biological Molecules

Computational Biology and Chemistry
18 November 2005

John F. Eargle*, and Zaida A. Luthey-Schulten†,§

*Center for Biophysics and Computational Biology, Urbana, IL

†Department of Chemistry, University of Illinois, Urbana, IL

§Correspondence to:
A544 CLSL, Box 41-6
600 S. Mathews Ave.
Department of Chemistry
University of Illinois at Urbana-Champaign
Urbana, IL 61801
Phone: (217) 333-3518
Fax: (217) 244-3186

Abstract

An important part of protein structure characterization is the determiniation of excluded space such as fissures in contact interfaces, pores, inaccessible cavities, and catalytic pockets. We introduce a general tessellation method for visualizing the dual space around and within biological molecules. A three dimensional graph is constructed to provide a displayable discretization of the continuous volume. This graph structure is also used to compare the dual space of a system in two different states. Tessellator, a cross-platform implementation of the algorithm, is used to analyze the cavities within myoglobin, the protein-RNA docking interface between aspartyl-tRNA synthetase and $tRNA^{Asp}$, and the ammonia channel in the hisH-hisF complex of imidazole glycerol phosphate synthase.

1 Introduction

Protein structure characterization provides useful information for understanding the functional capabilities of proteins. An important aspect of protein structure is the dual space surrounding a protein's surface. Contact interfaces, pores, inaccessible cavities, and ligand-binding pockets are all features that help define protein folding, interaction, and catalytic activity. Most of the work in defining voids has focused on quantifying void surfaces, surface areas, and volumes, but little effort has been made to produce a simple and systematic method that visualizes the dual space with changes in its geometry between two different conformational states of the system. The most widely used protein surfaces are the solvent accessibility (SA) and molecular surfaces (MS). (Lee and Richards, 1971; Connolly, 1983) Many programs focus on the MS, which shows the limit of possible range for the solvent exterior. One problem with this representation is that separate voids may have intersecting MSs which give the impression of a single, connected void even though the solvent is too large to pass from one void to the other. An SA surface for the same system would have no overlap between voids because the surface represents the limit of possible locations for the solvent's center. Our program, tessellator, captures the irregular void shapes by providing a space-filling structure outside the bounds of the SA surface and within a specified bounding box (Fig. 1). The vertices of the tessellation are distributed thoughout void space at least one solvent radius away from atomic Van der Waals (VDW) shells. They are then connected to their nearest neighbors as defined by Delaunay triangularization. Connections are made using the Hebbian rule which was developed in neuroscience for the automated linkage of neurons to their functional neighbors. The resulting structure is a useful tool to view voids and changes in void conformations because it provides a description of changes in void space that is conceptually separate from the motions of the surrounding atoms. Subtle molecular movements can cause drastic changes to void topology, such as pore openings or interface solvation, and tessellator can track these void distortions. The bounding box allows users to focus on specific areas of interest and avoid problems caused by intervening void space that could obscure the view.

Many molecular viewers implement MS and/or SA surface representations. (Nicholls et al., 1991; Gerstein, 1992; Sayle and Milner-White, 1995; Humphrey et al., 1996; Sanner et al., 1996; Pettersen et al., 2004; Insight II, 2005; MOE, 2005) However, it can be difficult to view features like pores and cavities. Typically, the surface for an entire molecule is calculated and displayed, obscuring any internal features. The HOLE package (Smart et al., 1993) was specially developed to visualize pores within proteins. It defines a pore by traveling along a given direction vector within the pore and repeatedly positioning and expanding a sphere until a maximum-volume sphere is found that fits between

the atoms defining the pore surface. Liang et al. (Liang et al., 1998a,b) introduced atom-based Voronoi tessellation and Delaunay triangulation as tools for analytically obtaining molecular volume and atom-atom contact information. This method, through its implementation in the Alphashapes program suite, can also be used to calculate the volumes of cavities that are fully enclosed within a protein. While Alphashapes allows one to view the tessellation tetrahedra that are used to define these cavities, it does not produce a visualization that closely resembles their actual shapes.

Protein motion can also be used to identify permanent and transient internal cavities. Locally enhanced sampling (LES) considers a system with a small subsystem of particles which is replicated, thereby enhancing sampling of the subsystem. Replicated subsystems interact with a shared environment but not with one another. LES was introduced to identify pockets and escape pathways for carbon monoxide, nitric oxide, and argon in myoglobin. Locations of molecules traveling through the protein are sampled over the trajectories to identify and visualize these dynamic cavities as well as their distribution over time. (Elber and Karplus, 1990; Li et al., 1993; Ulitsky and Elber, 1994) Recently, Cohen et al. used a temperature-controlled LES (TLES) method in which protein and ligands were exposed to separate Langevin heat baths to maintain a constant temperature ensemble. Molecules of H_2 and O_2 were bubbled through a protein producing trajectories that reveal cavities. Visualizations were produced for pockets produced by thermal motions described over the course of a MD simulation. They also created volumetric solvent accessiblity maps that show all interior voids bounded by the SA surface. (Cohen et al., 2005)

Tessellator creates a volume-filling structure that defines the shape of space around and within biological molecules. Vertices are distributed throughout the dual space within a defined area so that they cover the volume. They are then connected to their nearest neighbors defined by Delaunay triangulation. This results in a three dimensional graph that approximates the volume outside of the SA surface.

Because the tessellation uses a fixed number of vertices, and vertex motion is local, this approach is capable of identifying changes between two different snapshots of a particular void. Surface constructions typically fix the resolution of the representation beforehand. Thus, surfaces for a molecule in two different configurations tend to have different numbers of points associated with them. Even if they contain the same number of points, there is no automated procedure for mapping one set onto the other to help understand deformations of the surface. The volume tessellation computed by tessellator allows these kinds of comparisons to be made.

2 Methods

2.1 Delaunay Triangulation of 3D Void Volume

When creating a coarse-grained structure to represent a molecule, such as a protein, there are typical features one can use to guide the structure's design: amino acid center of mass, atomic coordinates, or VDW spheres. A void, however, has no such obvious internal structure. The approach we use is to represent empty space by points evenly distributed across the space. Then vertices are connected by edges to their nearest neighbors. Our concept of "nearest neighbor" is taken from Delaunay triangulation. (Delaunay, 1934) In effect, a pair of vertices will have an edge connecting them if there exists a point within the space that is closer to these two vertices than to any others. These connections produce a three dimensional graph representation that maintains the topological features of the target space (Fig. 1). The graph produced by Delaunay triangulation consists of simplices in zero to three dimensions: points, line segments, triangles, and tetrahedra. This graph is constructed by a self-organizing neural network algorithm. (Martinetz and Schulten, 1994) Self-organizing neural networks have previously been used in computationally intensive tasks such as somatosensory map simulation and visual pattern recognition. (Ritter et al., 1992) More recently, they have been used to tessellate protein volumes for fitting low-resolution electron microscopy data to x-ray crystal structures. (Wriggers et al., 1998)

2.2 Combined Neural Gas Algorithm and Hebbian Rule to Distribute Tessellation Points

The neural gas algorithm (NGA) is a general strategy to distribute points throughout a defined n-dimensional volume of any shape. Initially, the points are assigned to random positions in and around the volume defined by the bounding box. The NGA uses the stochastic selection of an input vector, v, given a probability density over the space of possible points. Here, our probability density is even, and the volume of concern is the volume excluded by all atomic VDW spheres. Vertices are placed in a list ordered by their distance from v. Then the vertices are moved towards v depending on how many other vertices in the list are closer to the input; closer vertices are moved more than those that are farther away producing a localized contraction around v. Iteration of this process initially pulls the vertices into the volume before spreading them evenly across the volume.

The competitive Hebbian rule was invented by Donald Hebb to assign con-

nections automatically from a neuron to neighboring neurons. (Hebb, 1974) As in the NGA, the Hebbian rule also employs a stochastic variable, but it is used to add edges to a set of vertices. A vector is chosen at random from the space in which the vertices lie. If the two vertices closest to this vector are unconnected, an edge is placed between them. For sufficiently large number of repetitions, all pairs of nearest neighbors are connected, and the resulting graph is the Delaunay triangulation of the vertices. (Martinetz and Schulten, 1994)

Combining the two algorithms, the NGA and the Hebbian rule, results in a procedure that simultaneously places vertices and connects them to their nearest neighbors. The output is a connected graph that fills a specified volume. This graph captures the shape of the volume while maintaining a uniform density of vertices. The same result could be achieved by completing the NGA and later connecting vertices with the Hebbian rule, but the combination allows them to be performed in parallel increasing the time-efficiency of the entire process.

2.3 Main Loop

Initially, distribute all n vertices throughout a box defining the area of interest.

The main loop of the combined algorithm:

(1) Choose an input vector v outside the SA surface based on the VDW radii of the atoms and the solvent radius.
(2) Place the vertices in a list $(w_0, w_1, ..., w_n)$ based on their distance from v.
(3) Move all vertices towards the input vector according to: $w_i^{new} = w_i^{old} + \epsilon \times e^{-k_i/\lambda}(v - w_i^{old})$
(4) Connect w_0 and w_1 with an edge of age 0 if they are not already connected. Otherwise, set the edge's age back to 0.
(5) Increment the ages of all the edges.
(6) Remove edges older than the current lifetime threshold T.

The parameter ϵ is the step size and λ is a step scaling parameter. Both of these, as well as T are time dependent and change from an initial value to a final vlue throughout the run according to $g(t) = g_i(g_f/g_i)^{t/t_{max}}$ where $t_{max} = 200n$. The initial and final parameters used in the program are $\epsilon_i = 0.3$, $\epsilon_f = 0.05$, $\lambda_i = 0.2n$, $\lambda_f = 0.01$, $T_i = 0.1n$, $T_f = 2n$. As the algorithm converges, the vertices move less and less. When the movement of w_0 is less than a set value, the process is completed.

2.4 Program Details

Tessellator was written in C++ and currently runs on MacOSX and UNIX machines. The program is freely available as open source software from the group website (http://www.scs.uiuc.edu/∼schulten/software.html). The website also contains more detailed documentation and example configuration files.

The main input files are Protein Data Bank (PDB) atomic coordinate files and previously generated tessellation (.tcl) scripts. The user specifies a 3D box in which to tessellate the dual space. This localizes the tessellation so that the user can avoid filling in large areas external to the molecule that might obscure the feature of interest. Atoms centered at coordinates within 3 Å of the box exterior are also taken into account during the tessellation to ensure the inclusion of VDW spheres that may overlap the box interior. The user can also set parameters pertaining to the tessellation run, such as number of steps, vertex radius, and how often to print output files.

The resolution of a tessellation depends upon the vertex density. Increasing the resolution, by shrinking the box or using more vertices, will allow a tessellation to model finer features of a space. To assure that the graphs produced have high enough resolution to represent continuous space, it is best to set the vertex density so that vertices are within one vertex diameter of their closest neighbors. This corresponds to the edge lengths being less than the vertex diameter. To guarantee this, one can choose the number of vertex points, n, based on the void volume, V_{tot} and the vertex radius, r. When the void volume is unknown, the box volume can be substituted to give a conservative estimate. Since the total spherical volume $V_s = n(4/3)\pi r^3$, and the highest packing density for equal-size spheres in three dimensions is $\pi/3\sqrt{2}$, $n = V_{tot}/(r^3 4\sqrt{2})$. In questionable cases, the mean edge length output by tessellator can be checked to verify this property.

2.5 Tcl Scripts for VMD

The output files from tessellator are Tcl scripts that specify the creation of 3D shapes in VMD. (Humphrey et al., 1996) These files are human readable and editable. When run in VMD, a visual representation of the tessellation appears in the OpenGL display and can be viewed in the context of the corresponding molecule from which the dual space was derived. These Tcl scripts can also be read into tessellator for further tessellation steps or image post-processing.

2.6 Post-Processing

When a 3D tessellation involves many vertices and edges, the resulting picture can be difficult to interpret due to its complexity. Using the Hebbian rule to produce a tessellation of dual space sometimes results in extraneous edges connecting voids that are not truly connected. The first post-processing step automatically prunes edges that are greater than three standard deviations above the mean edge length. When edges are pruned the mean edge length, standard deviation, and cutoff length are printed to the screen.

The tessellation can be viewed as a wireframe, but thousands of lines quickly become visually overwhelming. Another post-processing option fills in all triangles so that the resulting tessellations are opaque and appear as solid objects. This eliminates the confusing mesh of lines and allows VMD to render the structures with light and shading which clarify the 3D picture.

2.7 Visualizing Dynamical Changes

Given two different conformations of a protein, i.e. separate snapshots of a molecular dynamics simulation, the associated voids may also change shape. A tessellation from an initial state may be used as a starting point for the tessellation of successive states. This provides vertex identification for comparison between the two tessellations. Both tessellations can be used to generate representations colored by the distance traveled from one to the next. This reveals portions of the space that have undergone more conformational change. The edges for the second tessellation are recalculated by default to maintain a Delaunay triangulation. Statistics for the vertex movement are written into the Tcl file in comment lines. We use this method to show the differences between the open and closed configurations of the hisH-hisF ammonia channel (Fig. 5D).

2.8 Caveats

For complicated interfaces or fine-grained features, multiple tessellations may need to be created to capture the dual space features without obscuring the view of the interface. All representations can be displayed at once creating a 3D collage of the space of interest. Boxes in tessellator are currently defined with edges oriented in the x, y, and z directions specified by the cartesian coordinate system in the corresponding PDB file. Target void spaces may be arranged so that a single bounding box will enclose large amounts of unwanted void space that may obscure the area of interest. When investigating voids

that do not neatly fit into a single box, a helpful strategy is to divide the area into smaller pieces that each fit into a smaller box. Sometimes fine features of a space are desired within the context of flanking, large-volume spaces. Multiple tessellations may be required in these cases to allow for increased vertex density in the areas where high resolution is desired. All of our systems were analyzed with multiple tessellations (Figs. 2, 3A, 5).

3 Results and Discussion

3.1 General

Three systems were investigated with this tessellation method: the set of inaccessible cavities in myoglobin for comparison with other methods; aspartyl-tRNA synthetase (Asp-AARS) complexed with $tRNA^{Asp}$, a protein/RNA docking interface; and finally, a molecular channel in the hisH-hisF complex that transports ammonia between two active sites. Different approaches were used for the three systems in order to capture specific kinds of characteristics. The systems varied significantly in their conformations and sizes. The hisH-hisF tessellations make use of structures generated by molecular dynamics simulations to provide information about changes in the size and shape of the void space.

3.2 Inaccessible Cavities in Myoglobin

The structure of myoglobin has several interior voids that are typically unconnected to the outside solvent. There have been many crystallographic and NMR studies of xenon binding that show four specific xenon binding sites. (Schoenborn, 1965, 1969; Tilton and Kuntz, 1982) Internal cavities and ligand escape pathways were also found using LES MD simulations. (Elber and Karplus, 1990; Ulitsky and Elber, 1994) For comparison with other methods, we used PDB structure 1MBN from *Physeter catodon* including the heme heteroatoms (Watson, 1969), and the vertex radius was set to 1.2 Å. All four xenon binding sites are found as well as six other inaccessible cavities within the protein (Fig. 2). These ten cavities correspond to ten of the larger cavities computed by Alphashapes version 4.1, which finds a total of 17 cavities. (Liang et al., 1998b) Although these voids are usually wholly contained within the protein, there must be pathways that allow the entry and exit of oxygen and xenon as the molecule moves over time.

3.3 AARS-tRNA^{Asp} Interface

In translation of mRNA to proteins, AARSs enforce the genetic code by loading amino acids onto their cognate tRNA molecules. In yeast, Asp-AARS is comprised of two domains: the catalytic domain where the amino acid is loaded onto the tRNA, and the anticodon-binding (ACB) domain which helps recognize the correct tRNA through interactions with its anticodon loop (AC loop). The 5' and 3' ends of tRNA meet in a basepaired helix referred to as the acceptor stem. There are also three hairpin loops in tRNA called the D, T, and AC loops. The pattern of recognition is known to be at several sites. Important recognition elements on the protein are located in and around the active site of the catalytic domain and at the section of the ACB domain that binds to the tRNA AC loop. Different AARS-tRNA systems vary their contact regions to guarantee binding specificity because each system has to load an amino acid and recognize the anticodon in tRNA. The contact sites are well conserved (O'Donoghue and Luthey-Schulten, 2003), and it has been shown that few changes in these contact sites and the AARS active site can result in a system that loads an unnatural amino acid onto a tRNA. (Wang et al., 2001)

The structure used was PDB code 1ASY, the AARS-tRNA^{Asp} complex from *Saccharomyces cerevisiae*. (Ruff et al., 1991) Only protein chain B and RNA chain S were used to generate the tessellations. The protein structure begins at residue 68 so there is a significant segment missing form the N-terminus, but there are no available AARS-tRNA^{Asp} structures with this portion resolved. Asp-AARS is a dimeric protein, but a tRNA molecule makes almost all of its contacts with the AARS monomer to which it binds. The only close-range contact with the other monomer is between U1 and Lys293. Since the interface has a large and complex shape, the tessellation consists of twelve smaller tessellations combined to reveal the space around the entire interface. The solvent radius was set to 1.5 Å so vertex points represent volumes large enough to accomodate ions or water molecules.

This interface tessellation (Fig. 3A) reveals the partitioning of contacts into four main sites of protein/RNA interaction, each significantly different from the rest. The first two are associated mainly with the catalytic domain, the third spans both protein domains, and the last is the contact interface between the tRNA AC loop and the ACB domain of Asp-AARS. Aside from the contacts, it is also worthwhile to investigate the three areas between the tRNA and the synthetase that are large enough to accomodate intervening small molecules such as waters and ions. These crossing voids separate the four interaction sites from each other. Bioinformatic analysis was performed alongside the tessellation to focus on evolutionarily important residues in the interface. Evolutionary profiles (Sethi et al., 2005; Woese et al., 2000) were developed

to highlight interface residues that have been conserved across the *Eucarya* and across all three domains of life: *Eucarya*, *Archaea*, and *Bacteria*. Contact conservation in *Eucarya* is shown in Figure 4. More detailed pictures and an animation of the system with conservation shown through the tessellation are available at the group website (http://www.scs.uiuc.edu/∼schulten/software.html).

In site 1, the CCA strand and the last few bases in the acceptor stem make contacts with predominantly charged and polar residues in and around the AARS active site pocket. The ATP and Asp substrates were not present in this structure so their typical locations in the active site were treated as void space. When compared to another structure of the same system with ATP bound (PDB code 1ASZ) the CCA strand is in its bound state and not invading the ATP binding pocket. There is, however, a small perturbation in the loop around Gly282 which is in contact with A76. (Cavarelli et al., 1994) U1, the only 5' nucleotide member of this site interacts with Lys553, Asn328, and Asn330 at the mouth of the active site. Almost all of the active site amino acids are well conserved across all three domains of life. This is understandable as the last four bases of the $tRNA^{Asp}$'s 3' acceptor stem are also completely conserved. The first crossing void is between contact sites 1 and 2. It lies within the major groove of the acceptor stem, the larger of two grooves running down the length of basepaired RNA. It also winds under the A72, G73, and C74, the only nucleotides involved in both sites 1 and 2. At interaction site 2, nucleotides 67-74 on the 3' end of the acceptor stem coil around to make backbone contacts with polar and charged amino acids. Asp229, highly conserved in the *Eucarya*, is involved in a repulsive electrostatic interaction with the backbone phosphate of G68. This could counter the attractive forces involved in binding to allow faster separation of the tRNA and synthetase. By far the largest intervening space in the interface appears between sites 2 and 3, near the T-stem and slightly 5' of the D-arm. This is the space beneath the crook in the characteristic L-shaped bend in the tRNA. At interaction site 3, the basepaired residues of the D-stem rest against a predominantly hydrophobic patch of residues on the synthetase. Here, the protein contacts bases or sugars of the tRNA nucleotides 10-13 and 24-39. Pro205, in contact with the G10 sugar, is conserved across all of life and may play a role in stabilizing the positioning of the two protein domains with respect to each other. The third large space is between sites 3 and 4 and lies across the ACB domain of the synthetase and within the major groove of the tRNA helix leading to the anticodon of the tRNA. Interaction site 4, the ACB site, is the largest and is characterized on the protein side by largely polar residues. The nucleotides in the AC loop are not basepaired so there are more opportunities for hydrogen bonding between the tRNA and the protein. Three amino acids (Arg119, Phe127, and Gln138) are conserved throughout all of life and are in direct contact with the anticodon bases of the tRNA. Phe127 is pi-stacked with U635, the middle anticodon base, which is also hydrogen bonded to Gln138. Gly122, which is highly conserved and involved in a beta hairpin loop, also makes

VDW contacts with the modified base ψ32. 1MG37 is the only nucleotide in the AC loop that makes contact with the catalytic domain of the synthetase. It is notable that the T-loop side of the tRNA makes no contacts with the protein from nucleotides 39 to 71. Apart from these broad patches of interaction, there is an isolated attractive electrostatic contact, surrounded by void, between Lys155 and the G30 phosphate. Lys155 is also highly conserved among the *Eucarya*. Nucleotide contact information is shown in Figure 3B.

The conserved amino acid contacts suggest locations for mutational studies. Mutations of Lys155 and Asp229 would be expected to alter the synthetase-tRNA binding affinity, and mutation of Pro205 may destabilize the orientation of the two protein domains with respect to each other.

3.4 HisH-HisF Pore

The third system studied was the pore of the hisH-hisF complex from *Thermotoga maritima* (PDB code 1GPW). (Douangamath et al., 2002) SMD studies showed that ammonia could be transported from the glutaminase active site of hisH through the TIM barrel of hisF to its cyclase site. Tessellations of the pore in closed and open states provide a fast and easy way to visualize the subtle changes that occur within the protein. Hydrogens were added for the SMD runs and were included in the two tessellations. A vertex radius of 1.5 Å was too large and did not reveal the pore. A radius of 1.0 Å successfully found the pore without picking up extraneous, non-pore cavities within hisF. Figure 5A provides a view of a homologous system from *Saccharomyces cerevisiae* (PDB code 1OX5) where the closed pore has been visualized by HOLE using 0.25 Å increments along its sampling vector. (Amaro et al., 2005) While HOLE constructs the correct path, it requires prior knowledge of the pore's location and orientation to set the input sampling vector. It also does not reveal the constricted areas where the pore is closed off.

The first tessellation was created from the equilibrated protein structure in the closed state. All substrates and intermediates were removed before tessellation of the pore. In this configuration (Fig. 5B) the pore is separated into three segments by two sets of intervening residues. The four conserved putative gate residues, Arg5, Glu46, Lys99, and Glu167, form salt bridges that close the top of the barrel. In the open configuration, the pore does not proceed through the putative gate residues because it is energetically unfavorable to break these salt bridges. (Amaro et al., 2003) Very little movement is required, a slight rotamer change in Lys99, in order to open the pore wide enough to allow the passage of an ammonia.

The second tessellation is a snapshot taken of the system at a later state to re-

veal areas of relative movement or stability within the void. It shows the pore in its open configuration (Fig. 5C). This underlying protein structure was generated in the course of an SMD simulation in which an ammonia was pulled through the pore. (Amaro et al., 2005) Due primarily to a conformational change in Lys99, the previously segmented channel is now fully connected. The four charged putative gate residues connect across the top of the barrel, and the channel winds down to the side of this structure beneath Lys99 and Glu46. Figure 5D shows the difference between the two conformations. Tessellation vertices have streamed into the two newly open connections within the pore, and the rest of the structure remains stable. The triangles in the resulting structure are colored by the rmsd of their vertices, blue for little to no movement and red for greater movement.

4 Conclusion

The construction and display of dual space representations provides important information for the study of molecular biological systems. While previous approaches have focused on full protein surfaces or specific types of dual space, we have developed an application capable of creating volumetric structures corresponding to any space excluded by atomic VDW spheres. These structures, having a fixed number of points, can be compared with structures corresponding to the system in different configurations, elucidating the changes in voids between the structures. The focus of this presentation has been void space, but tessellator is equally capable of tessellating non-void shapes such as the volume defined by the VDW shell of a protein. Future work may employ the use of tessellation for low-resolution molecular models.

Acknowledgments

We would like to thank Klaus Schulten for suggesting the original class project on coarse-graining large proteins, Rommie Amaro for the SMD trajectories and HOLE pore determination of the hisH-hisF complex, and Carl Woese for the Asp-AARS sequence alignments used in the evolutionary profiles.

References

Amaro, R., Myers, R., Davisson, V., Luthey-Schulten, Z., 2005. Structural Elements in IGP Synthase Exclude Water to Optimize Ammonia Transfer. Biophys. J. 89, 475–487.

Amaro, R., Tajkhorshid, E., Luthey-Schulten, Z., 2003. Developing an energy landscape for the novel function of a (beta/alpha)8 barrel: Ammonia conduction through HisF. Proc. Natl. Acad. Sci. USA 100, 7599–7604.

Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J. C., Moras, D., 1994. The active site of yeast aspartyl-tRNA synthetase: Structural and functional aspects of the aminoacylation reaction. EMBO J. 13, 327–337.

Cohen, J., Kim, K., King, P., Seibert, M., Schulten, K., 2005. Finding gas diffusion pathways in proteins: O2 and H2 gas transport in CpI [Fe]-hydrogenase and the role of packing defects. Structure.

Connolly, M. L., 1983. Solvent-accessible surfaces of proteins and nucleic acids. Science 221, 709–713.

Delaunay, B., 1934. Sur la sphère vide. Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk 7, 793–800.

Douangamath, A., Walker, M., Beismann-Driemeyer, S., Vega-Fernandez, M., Sterner, R., Wilmanns, M., 2002. Structural Evidence for Ammonia Tunneling across the $(\beta/\alpha)_8$ Barrel of Imidazole Glycerol Phosphate Synthase Bienzyme Complex. Structure 10, 185–193.

Elber, R., Karplus, M., 1990. Enhanced Sampling in Molecular Dynamics: Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon Monoxide Diffusion through Myoglobin. J. Am. Chem. Soc. 112, 9161–9175.

Gerstein, M., 1992. A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. Acta Cryst. A 48, 271–276.

Hebb, D., 1974. Organization of Behavior. Wiley, New York.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD – Visual Molecular Dynamics. J. Mol. Graphics 14, 33–38.

Insight II, 2005. accelrys, San Diego.

Lee, B., Richards, F. M., 1971. The Interpretation of Protein Structures: Estimation of Static Accessibility. J. Mol. Biol. 55, 379–400.

Li, H., Elber, R., Straub, J. E., 1993. Molecular dynamics simulation of NO recombination to myoglobin mutants. J. Biol. Chem. 268, 17908–17916.

Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V., Subramaniam, S., 1998a. Analytical Shape Computation of Macromolecules: I. Molecular Area and Volume Through Alpha Shape. PROTEINS: Struc., Func., and Genetics 33, 1–17.

Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V., Subramaniam, S., 1998b. Analytical Shape Computation of Macromolecules: II. Inaccessible Cavities in Proteins. PROTEINS: Struc., Func., and Genetics 33, 18–29.

Martinetz, T., Schulten, K., 1994. Topology Representing Networks. Neural Networks 7, 507–522.

MOE, 2005. Molecular Operating Environment. Chemical Computing Group Inc., Montreal.

Nicholls, A., Sharp, K. A., Honig, B., 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.

PROTEINS: Struc., Func., and Genetics 11, 281–296.

O'Donoghue, P., Luthey-Schulten, Z., 2003. Evolution of Structure in Aminoacyl-tRNA Synthetases. Microbiol Mol Biol Rev 67, 550–573.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., Ferrin, T. E., 2004. Ucsf chimera - a visualization system for exploratory research and analysis. J. Comp. Chem. 25, 1605–1612.

Ritter, H., Martinetz, T., Schulten, K., 1992. Neural Computation and Self-Organizing Maps - An Introduction. Addison-Wesley, New York.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J., Moras, D., 1991. Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). Science 252, 1682–1689.

Sanner, M. F., Olson, A. J., Spehner, J. C., 1996. Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 38, 305–320.

Sayle, R. A., Milner-White, E. J., 1995. RASMOL: biomolecular graphics for all. Trends Biochem Sci 20, 374–374.

Schoenborn, B. P., 1965. Binding of xenon to horse haemoglobin. Nature 208, 760–762.

Schoenborn, B. P., 1969. Structure of alkaline metmyoglobin-xenon complex. J. Mol. Biol. 45, 297–303.

Sethi, A., O'Donoghue, P., Luthey-Schulten, Z., 2005. Evolutionary profiles from the QR factorization of multiple sequence alignments. Proc. Natl. Acad. Sci. USA 102, 4045–4050.

Smart, O., Goodfellow, J., Wallace, B., 1993. The Pore Dimensions of Gramicidin A. Biophys. J. 65, 2455–2460.

Tilton, Jr., R. F., Kuntz, Jr., I. D., 1982. Nuclear magnetic resonance studies of xenon-129 with myoglobin and hemoglobin. Biochemistry 21, 6850–6857.

Ulitsky, A., Elber, R., 1994. Application of the Locally Enhanced Sampling (LES) and a Mean Field with a Binary Collision Correction (cLES) to the Simulation of Ar Diffusion and NO Recombination in Myoglobin. J. Phys. Chem. 98, 1034–1043.

Wang, L., Brock, A., Herberich, B., Schultz, P. G., 2001. Expanding the genetic code of Escherichia coli. Science 292, 498–500.

Watson, H., 1969. The stereochemistry of the protein myoglobin. Prog. Stereochem. 4, 299.

Woese, C. R., Olsen, G. J., Ibba, M., Söll, D., 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process 64, 202–236.

Wriggers, W., Milligan, R., Schulten, K., McCammon, J., 1998. Self-organizing neural networks bridge the biomolecular resolution gap. J. Mol. Biol. 284, 1247–1254.

Figure Legends

Figure 1. 2D depiction of a tessellation structure. The thick line is the SA surface defined by rolling the solvent (small, black circles) along the VDW surface, the rectangle is the bounding box for the tessellation run, and the tessellation is the Delaunay triangulation of points placed between the two sets of atoms. There is a dashed circle around one of the vertices representing the vertex radius used to create the tessellation.

Figure 2. Tessellation of myoglobin inaccessible cavities. Xenon binding sites are labeled.

Figure 3. Tessellation of the $tRNA^{Asp}$-synthetase interface. A) The catalytic domain of the synthetase is shown in blue, and the ACB domain is shown in green. The $tRNA^{Asp}$ is colored by structural composition: acceptor stem (orange), D-loop (white), AC loop (red), and the variable region and T-loop (purple). The tessellation in gray shows the areas within the interface that are potentially accessible to H_2O. The four interaction sites are labeled. B) Secondary structure of $tRNA^{Asp}$ colored by synthetase interaction site: site 1 (red), site 2 (blue), site 3 (orange), site 4 (green), and G30 (purple). Black is used for non-interacting bases.

Figure 4. Amino acids within 5 Å of the tRNA are shown in surface representation and colored from high (blue) to low (red) conservation across the *Eucarya*. Nucleotides within 5 Å of the protein are shown in purple except for the anticodon bases, which are yellow.

Figure 5. Tessellation of the ammonia pore of the hisH-hisF complex. A) The hisH-hisF complex of *S. cerevisiae*. HisH is shown with red α-helices and orange β-strands, and hisF is shows with blue α-helices and purple β-strands. The pore shown was created using HOLE and depicted with spheres of uniform 2.0 Å radius. (Smart et al., 1993) B) Tessellation of the closed pore. C) Tessellation of the open pore. D) Tessellation of the open pore colored by difference from the closed pore tessellation. The most stable regions are colored blue, and the highest moving regions are red.

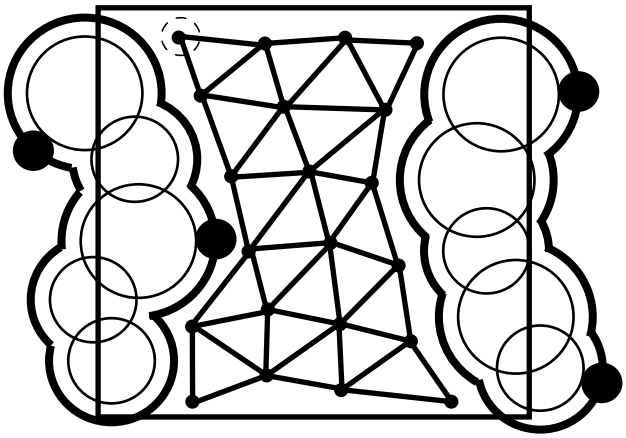

Fig. 1.

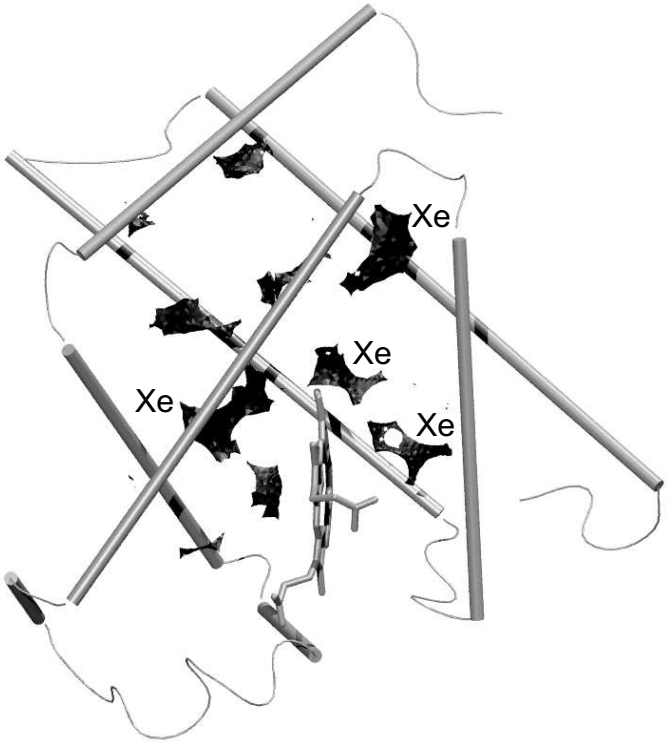

Fig. 2.

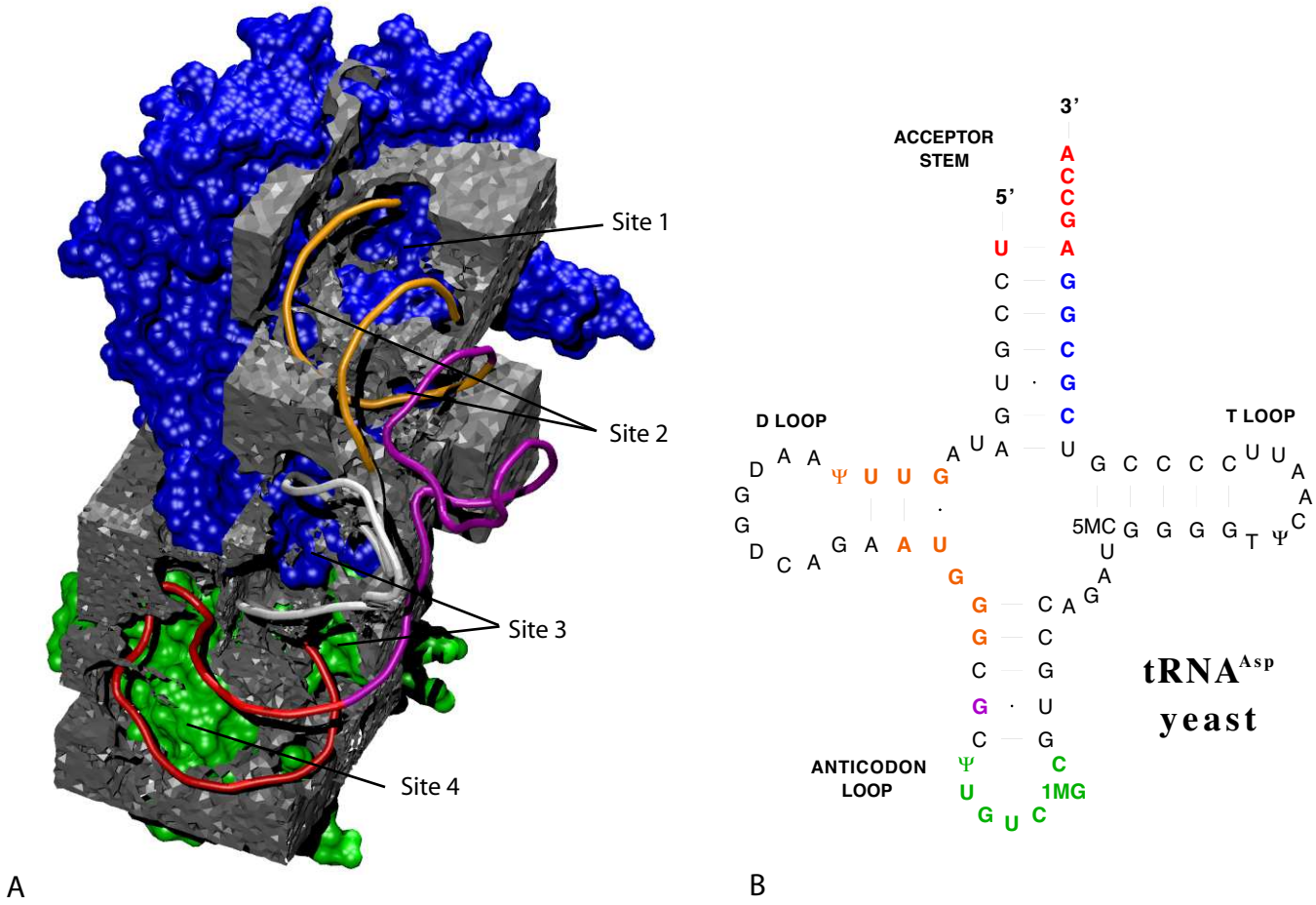
Site 1
Site 2
Site 3
Site 4
A
B
3'
ACCEPTOR STEM
5'
D LOOP
T LOOP
ANTICODON LOOP
tRNA^Asp
yeast

Fig. 3.

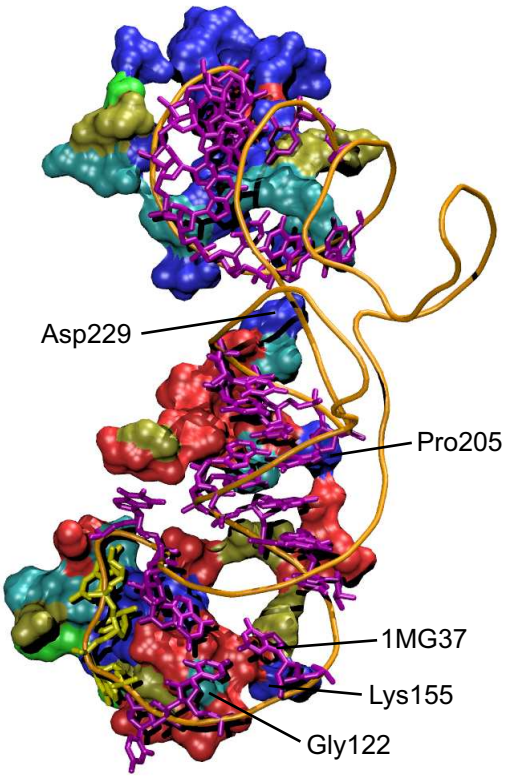

Fig. 4.

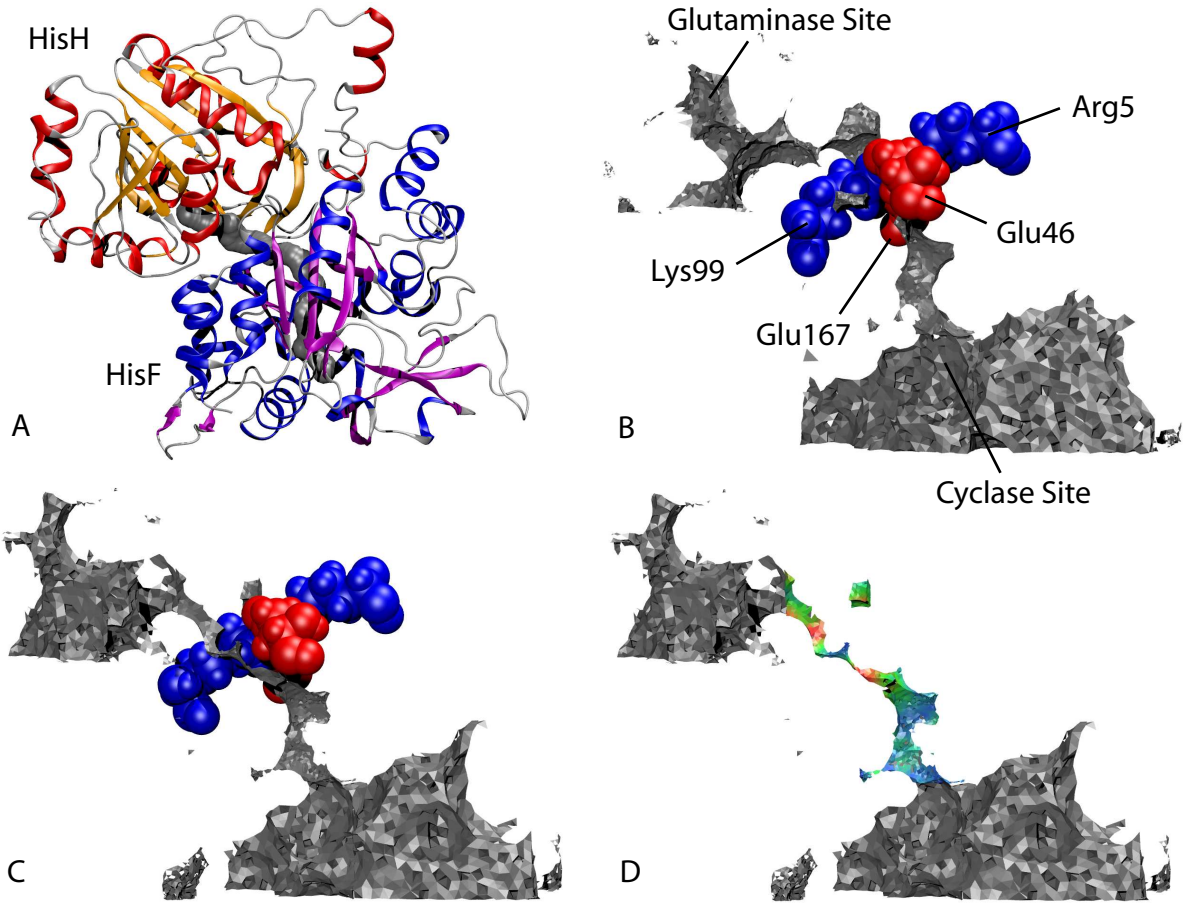
HisH
HisF
A
Glutaminase Site
Arg5
Glu46
Lys99
Glu167
Cyclase Site
B
C
D

Fig. 5.