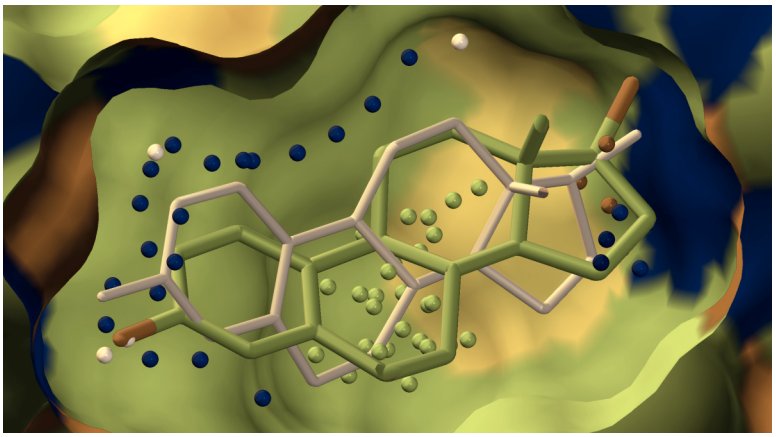
How Proteins Fold, Flex, and Bind Other Molecules

Leslie A. Kuhn, Volker Schnecke, Michael L. Raymer, & Paul C. Sanschagrin
Department of Biochemistry, Michigan State University
http://www.bch.msu.edu/labs/kuhn

Our research focuses on developing computational techniques for understanding the relationship between structure and function in proteins. Of particular interest is molecular recognition, the process by which proteins bind only a small, specific subset of the tens of thousands of molecules they encounter in their biological environments. We analyze several complementary aspects of molecular recognition: protein folding and flexibility (described in the companion abstract by Mike Thorpe), the influence of bound water molecules on protein structure and interactions, and development of combined *in vitro* and computational methods for screening large molecular databases to identify molecules that will bind to given sites on proteins.

On average, the interface between a protein and a ligand (e.g., an organic molecule it binds) contains 10 buried water molecules. This results in many water-mediated hydrogen bonds between the two molecules, with significant contributions to their free energy of binding. We have developed two approaches for identifying functionally important bound water molecules at protein surfaces. *Consolv* is a hybrid k-nearest-neighbors genetic algorithm for predicting conserved water sites at protein surfaces (developed with Bill Punch and Erik Goodman; see accompanying abstract). *Consolv* is optimized to identify those crystallographically-observed water molecules likely to remain bound to the protein when other molecules also bind. *Consolv* attains 75% accuracy in predicting water conservation/displacement, and 90% accuracy in predicting whether either a water molecule or polar atom of the ligand will bind at a site. With Shelagh Ferguson-Miller, this approach is being applied to identify water molecules important for forming the cytochrome c:cytochrome c oxidase complex. A second approach for identifying important bound water molecules is by using cluster analysis on the crystallographically-observed water sites in a series of independently-determined structures for the protein of interest. This approach, *WatCH*, identifies the most densely-occupied, non-overlapping set of water sites between the protein structures and calculates the degree of conservation for each (e.g., that a site is occupied by a water molecule in 80% of the structures). *WatCH* identified conserved water sites characteristic of serine proteases as a family, and other water sites that help particular serine proteases to discriminate between ligands.

SLIDE is a computational tool we have developed for screening large databases for structures of potential ligands that closely match a given binding site on a protein. The binding site is modeled by surface atoms, conserved bound water molecules, and a template consisting of hydrogen-bonding and hydrophobic interaction centers. During screening, distance geometry and multi-stage hashing techniques are used for mapping polar atoms and hydrophobic centers from the ligand candidates onto a subset of the template points. Ligands that match these template points are docked into the binding site and checked for steric fit. A mean-field approach allows the concurrent modeling of protein side-chain and full ligand flexibility to resolve van der Waals collisions. Potential ligands are scored on hydrogen-bond and hydrophobic contacts. *SLIDE* can screen either peptides or more general organic molecule structures, and identify and score all potential ligands to a protein within a day of computing on a typical desktop workstation. The figure above shows the solvent-accessible molecular surface calculated for the estrogen receptor crystal structure. The top-ten ligands identified by *SLIDE* included three structures of estradiol (a known ligand), and the best-scoring of these is shown by green tubes, as docked into the binding site by *SLIDE*. For comparison, the crystallographically-observed position of estradiol is shown in yellow, indicating a quite similar mode of binding to that predicted by *SLIDE*. Assays of other potential estrogen receptor ligands are being tested in collaboration with Tim Zacharewski.



See our web site, www.bch.msu.edu/labs/kuhn, for related publications. This research has been supported by MSU AURIG, the REF Center for Protein Structure, Function, and Design, a DFG fellowship to V.S., and by an NSF Early Career Award and AHA Established Investigatorship to L.K.