PREDICTING CONSERVED WATER-MEDIATED INTERACTIONS IN PROTEIN ACTIVE SITES

Michael L. Raymer[1,2], Sridhar Venkataraman[1], William F. Punch[2], Erik D. Goodman[3], and Leslie A. Kuhn[1], [1]Department of Biochemistry, [2]Department of Computer Science, and [3]Case Center for Computer-Aided Engineering and Manufacturing, Michigan State University, East Lansing, MI 48824

Water bound in protein active sites is essential for processes as diverse as degradation of penicillin by β-lactamase and recognition of DNA by *trp* repressor, yet the principles governing which waters are conserved or displaced upon ligand binding have not been characterized. While ligand-dependent steric effects may be involved, the chemistry of the water binding site is also likely to be important. Surprisingly, we show that even without explicit knowledge of the ligand, it is possible to predict most water-mediated ligand interactions from the chemical environments of bound water molecules in the free protein structure. A k-nearest-neighbor classifier coupled with a genetic algorithm was trained on 13 high-resolution protein structures' first-shell water molecules, which could be classified as conserved or displaced by comparison with a known ligand-bound structure. When this algorithm was applied to active site water molecules in free protein structures, 70% of the water sites were correctly classified as conserved or displaced upon ligand binding. Discriminant analysis, a traditional statistical technique, was also applied to predict conservation of water sites, and proved to have a similar prediction rate to the genetic algorithm outside active sites, but a random predictive level within active sites. A valuable product of both algorithms is a set of weights for the features used to characterize water-binding sites – hydrogen bonding, neighboring atom hydrophilicity, local mobility, and surface shape – reflecting the relative discriminatory ability of each feature in determining whether water is conserved or displaced. Interestingly, the features important for water conservation in active sites differ somewhat from those outside active sites. Together, our results suggest that water-mediated interactions for proteins can be predicted without knowledge of the ligand and that there is a generalizable structural chemistry for water-mediated protein:ligand interactions which can be applied to improve drug design and water modelling.