

BIOINFORMATICS OF PROTEIN-BOUND WATER

Christopher A. Bottoms

Dr. John J. Tanner, Dissertation Supervisor

ABSTRACT

Protein-bound water molecules are important components of protein structure, and therefore, protein function and energetics. Here we present a semi-automated computational approach for identifying conserved (i.e. structurally equivalent) solvent sites among proteins sharing a common three-dimensional structure. This method is tested on six protein families: (1) monodomain cytochrome *c*, (2) fatty-acid binding protein, (3) lactate/malate dehydrogenase, (4) parvalbumin, (5) phospholipase A₂, and (6) serine protease. For each family, the method successfully identified previously known conserved solvent sites. Moreover, the method discovered several novel conserved solvent sites, some of which have are more conserved than previously known sites. Our results suggest that every protein family will have highly conserved solvent sites, and that these sites should be included as defining features of protein families and folds.

Also detailed is the study of rat alpha parvalbumin at atomic resolution, including a study of conserved solvent sites among several members of the parvalbumin family.

Also detailed is a computational comparison of 101 high-resolution (≤ 1.90 Å) enzyme-dinucleotide (NAD, NADP, FAD) complexes which was performed to investigate the role of solvent in dinucleotide recognition by Rossmann fold domains. The typical binding site contains about 9-12 water molecules, and about thirty percent of the hydrogen bonds between the protein and the dinucleotide are water-mediated. Detailed inspection of the structures reveals a structurally conserved water molecule bridging dinucleotides with the well-known glycine-rich phosphate-binding loop. This water molecule displays a conserved hydrogen bonding pattern and appears to be an inherent structural feature of the classic Rossmann dinucleotide-binding domain.