

## **A Surface Patch Ranking Method Identifies Correlated Substrate Specificity Residues in Highly Homologous Enzymes**

**Gong-Xin Yu<sup>\*</sup>, Byung-Hoon Park, Praveen Chandramohan, Rajesh Munavalli, George Ostrouchov, Al Geist, Nagiza F. Samatova<sup>\*</sup>**

Computational Biology Group,  
Computer Science and Mathematics Division  
Oak Ridge National Laboratory

<sup>\*</sup>Corresponding authors: [yug@ornl.gov](mailto:yug@ornl.gov) and [samatovan@ornl.gov](mailto:samatovan@ornl.gov)

### **ABSTRACT**

The binding between an enzyme and its substrate is highly specific, despite the fact that many different enzymes show significant sequence and structure homology. There must be, then, substrate specificity residues that enable different enzymes to recognize their unique substrate. We reason that a coordinated, not independent, action of such residues determines enzymatic activity and specificity. Here we develop a computational procedure to discover putative correlated substrate specificity residues (CSSRs) in protein families that ranks the importance of surface patches to enzyme-substrate recognition. A family of highly homologous proteins is split into groups of proteins, where proteins in the same group have the same functional specificity and proteins from different groups have different specificities. As case studies we investigate two highly homologous enzymatic protein pairs: Guanylyl cyclases (GCs) vs. adenylyl cyclases (ACs) and Lactate dehydrogenase (LDH) vs. malate dehydrogenases (MDH). Without using experimental data, we predict a cluster of CSSRs for each pair, which covers multiple overlapping surface patches. We compare our predictions with current experimental results and obtain considerable agreement with them. Specifically, the GC/AC cluster determines multiple functionally linked amino acids that determine functional specificity in a coordinated manner. In previous mutagenesis experiments, only a simultaneous substitution at those residues can shift substrate specificity from GC to its homologous counterpart AC, whereas any individual mutation abolishes their enzymatic functions. These results demonstrate that CSSRs may be accurately identified from sequence and structure. This should help select target residues for mutagenesis experiments and, thus, focus rational drug design, protein engineering, and functional annotation to the relevant regions of a protein.

Keywords: bioinformatics, substrate specificity, surface patches, molecular recognition, protein function, mutagenesis

The submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.