FUNCTIONALLY RELEVANT CLUSTERING OF THE ARSENATE REDUCTASE (ArsC) SUPERFAMILY

Mikaela R. Rosen, Janelle B. Leuthaeuser, Cooper A. Taylor, Carol A. Parish, Jacquelyn S. Fetrow
Department of Chemistry, Gottwald Center for the Sciences, University of Richmond, Richmond, Virginia 23173, USA

We are developing computational methods that can expand our understanding of the atomic-level detail of the active site of proteins, focusing on the Arsenate Reductase (ArsC) superfamily. Arsenic concentrations have been rising, especially due to the increased pollutants, resulting from mining and agricultural activities. ArsC proteins are vital within arsenic redox microorganisms and are important in the bioremediation strategies that prevent arsenic, a highly toxic and carcinogenic metalloid, from reaching alarming levels in the environment. Our research utilizes a method called Multi-level Iterative Sequence Searching Technique (MISST), which uses the functional site profiles from structurally known proteins to computationally characterize members of groups and then uses DASP (Deacon Active Site Profiler) tool to expand the groups using data from GenBank. MISST further subdivides known groups, searching the second sphere (10 A around key residues selected using the literature) of a protein active site. One goal of the project is to validate the MISST process by confirming the divisions of the superfamily with published groups. Our project then focuses on active site geometry that is related to the protein's functional characteristics observed in molecular dynamics (MD) simulations and that allows us to hypothesize the enzyme reaction mechanism of, and the differences between, each isofunctional family. Using Amber 16, we ran apo-dynamics of 6 representative proteins of our groups to determine their structural differences and have found conserved mechanisms within each group. We are also running pHMD to find changes in side chain pKa for polarizable residues at snapshots during the simulations. We hypothesize that the ArsC superfamily, previously thought of as two or three ambiguous groups, should in fact be six isofunctional groups. Our computational methods can be applied to any subfamily within a superfamily, providing an efficient means of identifying functionally and structurally similar proteins.