Functional Characterization of Structural Genomics Proteins in the Crotonase Superfamily

Caitlyn L. Mills, Pengcheng Yin, Mary Jo Ondrechen, Penny J. Beuning

Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115

Abstract

In 2000, the Protein Structure Initiative (PSI) was started to determine three-dimensional structures of proteins so that each family has structural coverage. Structural Genomics (SG) protein structures are solved by high-throughput techniques and deposited in the Protein Data bank (PDB). As of today, there are over 12,500 SG proteins in the PDB, but most lack a functional assignment. Many SG proteins have only been assigned a putative function based on their global structural similarities with proteins of known function and many such assignments are incorrect. The Crotonase Superfamily (CS) contains at least 60 SG proteins and so is ideal to test predictions of protein function. Our focus is to determine if these SG proteins are functionally similar to the known proteins in this superfamily. First, computational analysis by POOL\(^2\) and SALSA\(^3\) is used to compare the catalytically active residues of the known members in the superfamily to those of the SG proteins. Next, biochemical assays will be developed and used to test these predictions. The main outcome of this project will be to successfully classify these SG proteins based on their local structure at the predicted active sites, and to provide a starting basis for the classification of the remaining SG proteins within the PDB.

Computational Analysis of Superfamily

- Before the SG proteins can be analyzed, the proteins of well-characterized and known function must be sorted into subgroups based on biochemical function.
- Once these catalytic active site residues are located via computational methods and literature searches, the structural genomic proteins can be compared to them.

Partial Order Optimum Likelihood (POOL)

POOL\(^2\) is able to predict the residues within the well-characterized proteins that are important for catalysis. Next, POOL is used to predict the residues within SG proteins that are catalytically important. These residues are compared to the well-characterized proteins by the method below.

Structurally Aligned Local Sites of Activity (SALSAs)

The SALSA\(^3\) method first uses POOL predicted residues and literature references for proteins of known function to develop consensus signatures for each functional subfamily. The POOL results from the SG proteins are then compared to the set of consensus signatures to gain insight into the possible functions of these SG proteins.

Currently, the SG proteins in this family are classified as having hydratase activity.

References


Background

The Crotonase Superfamily

- The enzymes in the Crotonase Superfamily catalyze a variety of metabolic reactions, but share the ability to stabilize the enolate anion intermediate formed from the substrate acyl-CoA.\(^4\)
- Founding member of superfamily is enoyl-CoA hydratase from Rattus norvegicus, (PDB ID 1DUB)\(^5\)
- Conserved quaternary structure consists of a trimer or a dimer of trimers as shown. Each monomer is shown in a different color.\(^6\)
- Each monomer consists of three ββ α repeating units that fold so the pairs of β-sheets are almost perpendicular to each other.\(^7\)
- The functional subgroups of this superfamily represent different types of reactivity, including hydratase, isomerase, and dehalogenase activities.
- Members of this superfamily have potential applications in areas as diverse as agriculture, environmental remediation, medicine, and green chemistry.

Northeastern University

Ondrechen Research Group – Computational Biology

Funding and Acknowledgements

- NSF: CHE-1305655 (MJO, PJB)
- Dr. Srinivas Somaradhulu for his computational work on the crotonase superfamily members
- NU Office of the Provost: Ph. D. Graduate Research (CLM)