STRUCTURAL GENOMICS OF *ESCHERICHIA COLI*

M. Cygler A. Matte J. Sivaraman Y. Li F. Oullet R. Larocque G. Michel S. Raymond J.D. Schrag V. Lunin N. OToole P. Iannuzzi E.S. Rangarajan
Biotechnology Research Institute, NRC Macromolecular Structure 6100 Royalmount Avenue MONTREAL QUEBEC H4P 2R2 CANADA

We have initiated a structural genomics project based on selected protein families, with the representatives for structural studies obtained from the genome of *E. coli*. A total of 160 genes have been cloned to date, with 126 successfully overexpressing soluble protein with at least one fusion system. We are currently applying robotics methods allowing cloning in an automated manner. Target genes have been cloned as N-terminal fusions with GST, (His)6, or (His)8 affinity tags. Over 60 proteins have so far been purified to homogeneity. Purified proteins are further characterized for homogeneity and suitable solution properties using a combination of dynamic light scattering and electrophoretic methods. Purified protein samples are screened for initial crystallization conditions in 96-well format using a sparse-matrix approach. A specialized database has been developed to follow the progress of various proteins and to store all relevant experimental data (**http://sgen.bri.nrc.ca/bsgi**). Web-based software for displaying and searching the information from this local database and for tracking deposited structures in the PDB as well as the progress of other structural genomics projects has been developed. To date, crystals have been obtained for 37 of the purified proteins. of these, diffraction quality crystals were obtained for 16 proteins. Using SeMet-substituted proteins we have determined the structures of 13 of these proteins by MAD phasing. Structures determined include 2-amino-3-ketobutyrate CoA ligase, part of the threonine salvage pathway, MoeA protein, involved in molydopterin biosynthesis, histidinol phosphate aminotransferase and L-histidinol dehydrogenase, two enzymes associated with histidine biosynthesis, and RsuA, a 16S rRNA psuedouridine synthase.

Keywords: STRUCTURAL GENOMICS, PROTEIN STRUCTURE

A SYSTEM FOR THE RAPID AND SYSTEMATIC CHARACTERIZATION OF PROTEIN CRYSTALS

A. Deacon[1] L. Brinen[1] P. Kuhn[1] S. McPhillips[2] T. McPhillips[1] M. Miller[1] H. van den Bedem[1] G. Wolf[1] J. Zhong[2] Z. Zhang[2]
[1]The Joint Center for Structural Genomics Stanford Synchrotron Radiation Laboratory SLAC MS 99 2575 Sand Hill Road MENLO PARK CA 94306 USA [2]Stanford Synchrotron Radiation Laboratory, Stanford, USA

A high-throughput Crystal Screening System (CSS) has been implemented within the Distributed Control System (DCS) architecture, which is used on all the protein crystallography beam lines at the Stanford Synchrotron Radiation Laboratory (SSRL). The CSS provides a highly automated and user-friendly way of systematically collecting test data from a set of protein crystals. A database maintains a prioritized list of the crystals available at each beam line. The CSS collects crystal JPEG images and diffraction images at various crystal orientations and also measures automated fluorescence scans. Soon a robot-controlled sample mounting system, coupled with automated sample alignment software, will position each crystal in the X-ray beam. A Diffraction Image Scoring Tool (DISTL) is able to quickly analyze the quality of each image, it can detect ice-rings and it can assess the strength and shape of the diffraction spots. It can also rank each crystal processed. A prototype system has been used to screen several hundred crystals as part of the structure determination efforts of the Joint Center for Structural Genomics (JCSG). The CSS significantly improves the efficiency of the crystal screening process. The systematic approach eliminates many sources of human error and provides more reliable feedback to guide and improve the crystal preparation process.
(The JCSG is funded by the Protein Structure Initiative of the National Institutes of Health, National Institute of General Medical Sciences. SSRL operation is funded by DOE BES, and the SSRL Structural Molecular Biology program by DOE BER, NIH NCRR BTP and NIH NIGMS).

Keywords: AUTOMATION, STRUCTURAL GENOMICS, HIGH-THROUGHPUT

CRYSTAL STRUCTURE OF SACCHAROMYCES CEREVISIAE HOMOLOGOUS MITOCHONDRIAL MATRIX FACTOR 1 (Hmf1)

A.M. Deaconescu[1] A. Roll-Mecak[1] J.B. Bonanno[2] S.E. Gerchman[3] H Kycia[3] F.W. Studier[3] S.K. Burley[2]
[1]Rockefeller University 1230 York Avenue #381 NEW YORK NY 10021 USA [2]Rockefeller University, Howard Hughes Medical Institute [3]Brookhaven National Laboratory

The YjgF/YER057c/UK114 protein superfamily has been highly conserved among eubacteria, archaea and eukaryotes. Functionally, this protein family is remarkably diverse, with members involved in calpain activation, tumor antigenicity, translation inhibition, and various biosynthetic pathways. Here, we describe the first structure of a eukaryotic member of this large superfamily, the Saccharomyces cerevisiae Homologous Mitochondrial Matrix Factor 1 (Hmf1). Hmf1 is a cytoplasmic homolog of Mitochondrial Matrix Factor 1 Mmf1, which has been implicated in the transamination step of isoleucine biosynthesis plus maintenance of mitochondrial DNA. The crystal structure of Hmf1 (PDB ID 1JD1) was solved at 1.7Å resolution using multiple anomalous dispersion (MAD) phasing (Rwork = 0.207, Rfree=0.238) from a single Se-Met crystal, and was subsequently used for homology modeling that provided structural information for proteins from all three living kingdoms, including the YER057c superfamily (models available at **http://www.pipe.rockefeller.edu**). X-ray crystallography revealed that Hmf1 is a trimeric α/β protein, which resembles Bacillus subtilis chorismate mutase despite small pairwise sequence similarity. Analyses of the homotrimer molecular surface permitted identification of a putative active site within an inter-monomer groove, analogous to, but chemically distinct from, the active site of chorismate mutase. No motifs were identified and no template was found to have the same biochemical milieu and spatial arrangement of residues as those in the putative active site of Hmf1.

Keywords: ISOLEUCINE-BIOSYNTHESIS,CHORISMATE MUTASE, MTDNA MAINTENANCE

ANTISENSE ORFs AND CODON BIAS IN SHORT CHAIN OXIDO REDUCTASE ENZYMES AND THE EVOLUTION OF THE GENETIC CODE

W. Duax[1] A. Addlagatta[1] V. Pletnev[2] P. Yu[1]
[1] Hauptman-Woodward Medical Research Institute, Inc. Structural Biology 73 High Street BUFFALO NY 14203-1196 USA [2]Shemyakin Institute

The short chain oxidoreductase (SCOR) family of enzymes includes over 2800 members extending from bacteria to humans for which 20 have known crystal structures and 2500 have unknown function. The superfamily has not one fully conserved residue. The signatures of subgroups of the superfamily are composed of 30-40 residues conserved at ~80% identity distributed throughout the 250 aa proteins.

The conservation of all residues in contact with the cofactor suggesting strict co-evolution of fold and cofactor. D or R residues associated with NAD and NADP binding respectively predict cofactor preference of unknowns. Covariance patterns in the binding pocket correlate with substrate preference. Although no crystallographically determined structure in the SCOR family includes a disulfide, the location of cysteine residues that covary in subsets of the family suggest the potential for disulfide formation.

Nucleic acid sequence analysis reveals that 15% of the SCOR genes have an antisense open reading frame overlapping the entire sense gene, and 5% have a third ORF related by a frame shift in the sense strand. The 85 genes with double open reading frames (DORFs) and 36 with triple open reading frames (TORFs) exhibit codon bias. Over 85% of the 250 amino acids in the 121 SCOR proteins are coded for by the GC rich codons. This and other data suggest that the SCOR family of enzymes diverged from a common ancestor that evolved before the AT rich half of the genetic code was fully defined.

Supported by NIH grant NO. DK26546.

Keywords: GENOMICS, PROTEOMICS, ENZYMES