Structural genomics programs at the US National Institute of General Medical Sciences

John C. Norvell and Alisa Zapp Machalek

National Institute of General Medical Sciences

The results of genome sequencing projects and recent advances in structure determination have ushered in structural genomics, a new field focused on the large-scale analysis of protein structures and functions based on gene sequences. In response, the National Institute of General Medical Sciences (NIGMS) announced its "Protein Structure Initiative", which is designed to organize a large, cooperative effort in structural genomics. NIGMS is a component of the US National Institutes of Health that supports basic biomedical research, including a major program in structural biology.

Through the Protein Structure Initiative, NIGMS seeks to develop a public resource to organize and analyze protein structures and fold families. This resource will link sequence, structural, and functional information and will enable the prediction of unknown structures by homology modeling. The first aim is to determine the structures of 10,000 proteins (one or more from each fold family) in 10 years.

Toward this end, the initiative supports technology development by researchers and small businesses. It also supports pilot research centers to develop high-throughput tools and strategies for every aspect of structural genomics, from target selection to structure determination and analysis.

In late September, NIGMS provided almost $30 million for the first round of research center applications: seven awards, each totaling around $4 million for the first of five years. The awards are listed below alphabetically by the name of the principal investigator.

(i) Five institutions in the New York City area have joined to form the New York Structural Genomics Research Consortium, which will develop techniques to streamline every step of structural genomics. Within five years, the consortium expects to solve several hundred protein structures from humans and a variety of of model organisms. (Stephen K. Burley, The Rockefeller University)

(ii) The Midwest Center for Structural Genomics, a consortium of seven institutions, seeks to reduce the average cost to determine a protein structure from $100,000 to $20,000. The group will select protein targets from all three kingdoms of life (Eukarya, Archaea, and Bacteria), with an emphasis on new folds and proteins from disease-related organisms. (Andrzej Joachimiak, Argonne National Laboratory)

(iii) The Structural Genomic Center aims to speed up structure determination by X-ray crystallography. It will focus on two bacteria with extremely small genomes to study proteins essential for independent life. The two bacteria, *Mycoplasma genitalium* and *M. pneumoniae*, are closely related. The former contains the smallest genome of any free-living organism and infects the human genital and respiratory tracts. The latter causes a form of pneumonia. (Sung-Hou Kim, Lawrence Berkeley National Laboratory)

(iv) Researchers in New Jersey, New York, Connecticut, Washington State, and Ontario, Canada have formed the Northeast Structural Genomics Consortium, which will target proteins from eukaryotic organisms, including *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and related human proteins. This consortium will use both X-ray crystallography and NMR spectroscopy to determine protein structures. (Gaetano Montelione, Rutgers University)

(v) A collaboration of scientists in six countries have formed the TB Strctural Genomics Consortium to determine and analyze the structures of ~400 proteins from *Mycobacterium tuberculosis*. The group seeks to optimize the technical and managerial underpinnings of high-throughput structure determination and will develop a database of structures and functions. NIH's National Institute of Allergy and Infectious Diseases, which is co-funding this project with NIGMS, anticipates that this information will also lead to the design of new and improved drugs and vaccines for tuberculosis. (Thomas Terwilliger, Los Alamos National Laboratory)

(vi) The Southeast Collaboratory for Structural Genomics, with its core in Georgia and Alabama, will analyze part of the human genome and the entire genomes of two representative organisms — the eukaryote *Caenorhabditis elegans* and an evolutionary related prokaryote, *Pyrococcus furiosus*. The group emphasizes technology development, especially for automated crystallography and NMR techniques. (Bi-Cheng Wang, University of Georgia)

(vii) The Joint Center for Structural Genomics, centered in California, is developing high-throughput methods for protein production, crystallization, and structure determination. The group will initially focus on novel structures from *C. elegans* and on human proteins implicated in cell signaling. It will also include homologous proteins from other organisms to ensure the greatest coverage of protein fold space (Ian Wilson, The Scripps Research Institute).

Worldwide momentum in structural genomics is growing rapidly. In April, 2000, NIGMS and The Wellcome Trust, a UK-based medical research charity, cosponsored the First International Structural Genomics Meeting in Hinxton, near Cambridge, UK. Participants from nine countries discussed the goals and policies of the international collaborative effort. Several task forces investigating these issues will report to the second international meeting to be held April 4-6, 2001 near Washington, DC.

More information about the NIGMS structural genomics initiative is available at *http://www.nih.gov/nigms/funding/psi.html* or from John Norvell, director of the initiative, at (301) 594-0533 or *email: norvellj@nigms.nih.gov.*