

# Structural Genomics: Past, Present, And Future

Dr Andrzej Joachimiak<sup>1</sup>

<sup>1</sup>*Structural Biology Center, X-ray Science Division, Argonne National Laboratory Biochemistry and Molecular Biology, University of Chicago*

*andrzejj@anl.gov*

The exponential growth of genomics sequence information in the 1990s led to significant knowledge gaps in our understanding of biological systems. It was true then, and it is still true now that the sequence information bore little insights about the functions encoded in the genomes. The field of Structural Genomics (SG) arose to address these gaps. The mission of SG programs was to facilitate rapid de novo structure determination for proteins representing new protein families to provide meaningful structural coverage of the genomes. There were significant challenges to advance technologies for the preparation of thousands of proteins and for their structural and functional characterization. The SG programs quickly addressed barriers, and deficiencies, improved effectiveness, and reproducibility, and created highly integrated and cost-effective pipelines for protein production and structure determination. The improvements in experimental methods developed by the SG consortia resulted in fast progress in molecular and structural biology, enhanced structure quality, and significantly benefitted biological and biomedical research, providing insights into novel structural and functional space. The experimental three-dimensional models were promptly made public through the Protein Data Bank structure repository, facilitating the structure determination of other members of the family, and helping to understand their molecular and biochemical functions. The light sources and dedicated macromolecular crystallography beamlines, advanced software, and computing resources have contributed to SG success and expanded biology community competence in determining protein structures. Structural biology research was set to undergo a major transformation. The advancements resulted in the determination of thousands of protein structures, mostly from unique protein families, and increased structural coverage of the rapidly expanding protein universe. These structures contributed to AlphaFold/RozerTTAFold AI algorithms allowing accurate structure prediction of millions of proteins. In principle, the original goal proposed by the National Institutes of Health Protein Structure Initiative, that structures of all proteins should be available to the community experimentally or computationally, has been accomplished.

At present, x-ray crystallography remains the most powerful method capable of providing atomic information on interactions of proteins with other macromolecules and small ligands and is an essential tool for drug discovery. Existing SG protein production and structure determination pipelines contributed to tackling the molecular and structural biology of pandemic agents such as SARS-CoV-2. These resources support studies of complex diseases such as cancer, tuberculosis, and malaria. Expanded protein sequence/structure/function space allows for comprehensive approaches to studies of entire cellular systems, including many human pathogens, uncultured organisms, microbiomes, and metagenomes. SG showed that addressing large biological problems requires adequate, well-coordinated, and integrated research programs.