# Structural Genomics' Role in Innovative Structure Characterization

John-Marc Chandonia[1], Steven E Brenner[2]

*[1]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, [2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Department of Plant and Microbial Biology, University of California*

***jmchandonia@lbl.gov***

The rate at which new protein structures are solved has steadily increased, and now averages over 1,000 structures per month. However, fewer than 2% represent the first solved structure from a Pfam protein family. The other 98% of solved protein structures are from families with another known structure, and would be expected to share the fold of those previously solved structures. By contrast, the first structure from a protein family reveals the structural fold of that family, and newly enables inference of ancient relationships to other proteins. A large number of structurally characterized families was needed to train deep learning algorithms that have made breakthroughs in accurate structure prediction. Structural Genomics (SG) was a worldwide effort directed at solving such structures. During the heyday of SG, between 2003 and 2007, SG centers' output peaked at around 25 first structures from a family per month, and traditional laboratories solved a similar number, with these ~50 representing ~10% of structures solved. With many key families having been characterized, it is unsurprising that most structure characterization since has focused on more detailed understanding. Yet, since SG concluded, the rate at which protein families are now structurally characterized has fallen to only about 20 per month. This is a decrease not just in the fraction of new families being characterized, but in the absolute number – representing a 2-decade low despite vastly improved technology. I will discuss the impact that directed characterization of new families had on structurally characterizing the human proteome, implications for human genetic variant interpretation, and what structural novelty might be expected to be discovered were efforts again directed at new families.