# Computational Approaches for Structural and Functional Genomics

**Steven E. Brenner**
Department of Plant & Microbial Biology
111 Koshland Hall #3102
University of California,
Berkeley, CA 94720-3102, USA.
brenner@compbio.berkeley.edu

## Abstract

Structural genomics aims to provide a good experimental structure or computational model of every tractable protein in a complete genome. Underlying this goal is the immense value of protein structure, especially in permitting recognition of distant evolutionary relationships for proteins whose sequence analysis has failed to find any significant homolog. A considerable fraction of the genes in all sequenced genomes have no known function, and structure determination provides a direct means of revealing homology that may be used to infer their putative molecular function. The solved structures will be similarly useful for elucidating the biochemical or biophysical role of proteins that have been previously ascribed only phenotypic functions. More generally, knowledge of an increasingly complete repertoire of protein structures will aid structure prediction methods, improve understanding of protein structure, and ultimately lend insight into molecular interactions and pathways.

We use computational methods to select families whose structures cannot be predicted and which are likely to be amenable to experimental characterization. Methods to be employed included modern sequence analysis and clustering algorithms. A critical component is consultation of the presage database for structural genomics, which records the community s experimental work underway and computational predictions. The protein families are ranked according to several criteria including taxonomic diversity and known functional information. Individual proteins, often homologs from hyperthermophiles, are selected from these families as targets for structure determination. The solved structures are examined for structural similarity to other proteins of known structure. Homologous proteins in sequence databases are computationally modeled, to provide a resource of protein structure models complementing the experimentally solved protein structures.

**References**

Brenner SE, Levitt M. 2000. Expectations from structural genomics. Protein Sci. 9:197-200.

Brenner SE. 1999. Errors in genome annotation. Trends Genet 15:132-133.

Brenner SE, Barken D, Levitt M. 1999. The PRESAGE database for structural genomics. Nucleic Acids Res 27:251-253.

Brenner SE, Chothia C, Hubbard TJP. 1998. Assessing sequence comparison methods with reliable structurally-identified distant evolutionary relationships. Proc Natl Acad Sci USA 95:6073-6078.

Brenner SE, Chothia C, Hubbard TJP. 1997. Population statistics of protein structures. Curr Opin Struct Biol 7:369-376.

Brenner SE, Chothia C, Hubbard TJP, Murzin AG. 1996. Understanding protein structure: Using SCOP for fold interpretation. Meth Enzymol 266:635-643.

Brenner SE, Hubbard T, Murzin A, Chothia C. 1995. Gene duplications in the H. influenzae genome. Nature 378:140.

Brenner SE. 1995. World wide web and molecular biology. Science 268:622-623.

## Curriculum Vitae

**Steven E. Brenner**

Assistant Professor, Department of Plant and Microbial Biology

Faculty Scientist, Physical Biosciences, Lawrence Berkeley National Laboratory
Affiliated Assistant Professor, Department of Molecular & Cellular Biology
Affiliated Assistant Professor, Department of Bioengineering
Member, Biophysics Group
Member, UCSF/UCB Bioengineering Group

A.B. 1992. Harvard University
M.Phil. 1994. University of Cambridge
Ph.D. 1997. University of Cambridge
Visitor. 1997. Japan National Institute of Bioscience
Sloan/DOE Postdoctoral Fellow. 1999. Stanford University

**Research Interests**

We develop computational methods for the analysis and integration of molecular sequence, structure, and expression data. Our aim is to understand organismal biology by interpreting the information encoded in complete genomes. This work is presently focused on the areas of structural and functional genomics.

Structural genomics projects attempt to provide an experimental structure or a good theoretical model for every tractable protein in all completed genomes. Our work involves organizing proteins into families according to homology; classifying proteins and RNA according to structure; predicting structure from homology and constructing atomic coordinate models; providing information resources for structural genomics; developing methods for selection of proteins for experimental characterization; and analyzing solved structures to detect homology and functional information.

We study computational functional genomics by creating algorithms using molecular sequence, structure, phylogeny, and expression information to infer the functions of genes. This work includes the use of gene genealogies to trace gene histories and functional divergences;

reverse-genomics comparison of multiple complete genomes to locate genes associated with characterized cellular or biochemical functions; creation of databases of genomic information; and continued refinement of sequence comparison methods. We also combine sequence comparison with expression and other experimental data to improve molecular and cellular functional characterization.

**Selected Publications**


Brenner SE, Levitt M. 2000. Expectations from structural genomics. *Protein Sci. 9*:197-200.

Brenner SE. 1999. Errors in genome annotation. *Trends Genet 15*:132-133.

Brenner SE, Barken D, Levitt M. 1999. The PRESAGE database for structural genomics. *Nucleic Acids Res 27*:251-253.

Brenner SE, Chothia C, Hubbard TJP. 1998. Assessing sequence comparison methods with reliable structurally-identified distant evolutionary relationships. *Proc Natl Acad Sci USA 95*:6073-6078.

Brenner SE, Chothia C, Hubbard TJP. 1997. Population statistics of protein structures. *Curr Opin Struct Biol 7*:369-376.

Brenner SE, Chothia C, Hubbard TJP, Murzin AG. 1996. Understanding protein structure: Using SCOP for fold interpretation. *Meth Enzymol 266*:635-643.

Brenner SE, Hubbard T, Murzin A, Chothia C. 1995. Gene duplications in the H. influenzae genome. *Nature 378*:140.

Brenner SE. 1995. World wide web and molecular biology. *Science 268*:622-623.