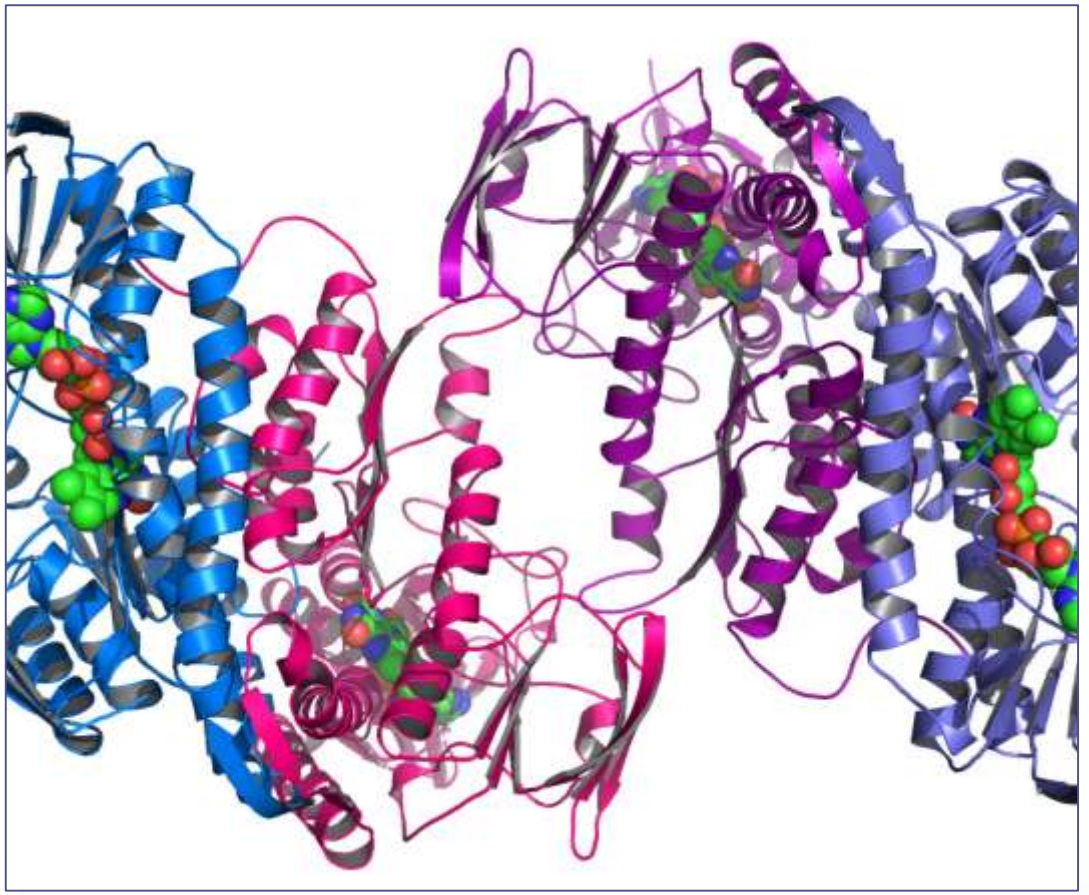
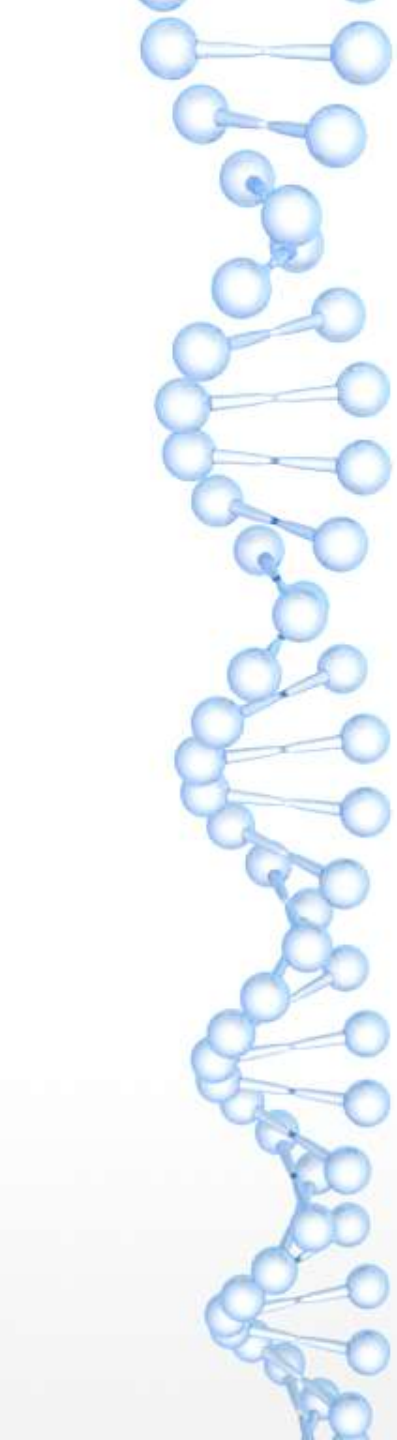
STRUCTURAL GENOMICS



Presented By

Aqsa Javed

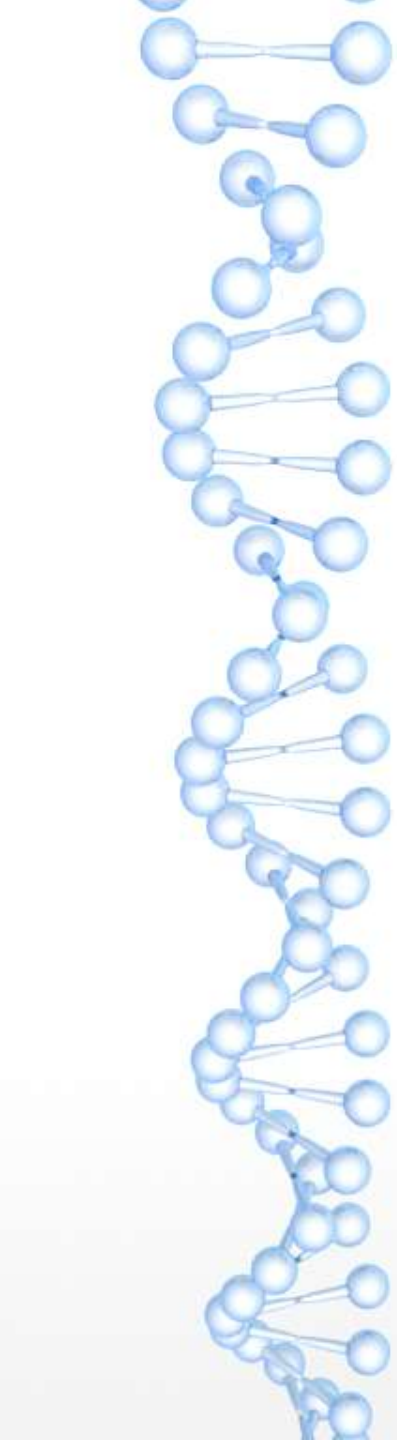
IBI2016001



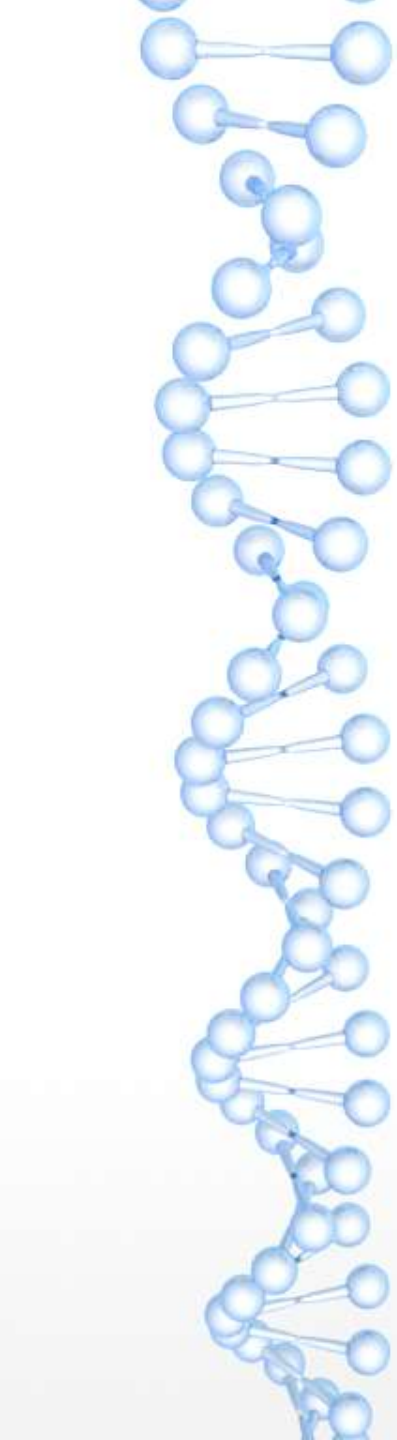
Introduction to Structural Genomics

- When the human genome was completely sequenced in 2003, researchers started thinking in how many ways it used.
- One hope was that the genome sequences would lead to a greater understanding of how genes and their encoded proteins function
- As the number of known gene sequences grew, many scientists realized they could not catch up simply by determining protein structures one by one. So a group of scientists embarked on a strategic plan to uncover the three-dimensional structures of all the proteins that these genes encode.
- This endeavor is called structural genomics.

Introduction to Structural Genomics

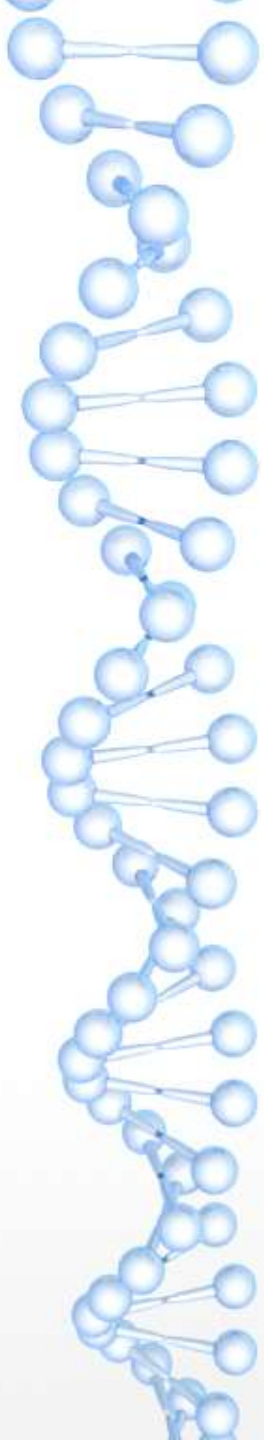


- Structural genomics is a term that refers to high-throughput three-dimensional structure determination and analysis of biological macromolecules, at this stage primarily individual protein domains.
- It mainly aims to solve the 3 dimensional structures of proteins at a rapid rate and in a cost effective manner.
- It uses a genome based approach to describe the 3- dimensional structure of every protein encoded by a given genome.
- Structural genomics has emerged as one of the most powerful approaches for defining the structure of proteins .
- Several structural proteomics groups pursue the structures of proteins that are "unique", generally ones that have less than 30% sequence identity to a protein with a known structure in the Protein Data Bank.



- Structural genomics is expected to yield a large number of experimental protein structures (tens of thousands) and an even larger number of calculated comparative protein structure models (millions). This enormous body of structural data will be freely available, and promises to accelerate scientific discovery in all areas of biological science.
- As the protein structure and function are closely linked, the importance of structural genomics in understanding the function of proteins is paramount. Structural genomics can also provide insight in dynamic properties such as protein folding and identify possible targets that may be used for drug discovery.

Differences between Structural Genomics and Traditional Structural Prediction



Structural Genomics

- Structural genomics attempts to determine the structure of every protein encoded by the genome, rather than focusing on one particular protein.
- Structural genomics aims to first determine the structure of proteins, and then investigate their function later.

Traditional Structural Prediction

- Traditional structure prediction focuses on one particular protein at a time.
- The function of a protein of interest is first identified and then its structure determined by means of X-ray or NMR experiments.

Structural Genomics Process

In general terms the process involves:

1. PCR amplification of the coding sequence from the target genome.
2. Cloning the coding sequence into an appropriate expression vector.
3. Expressing the protein at a sufficiently high level.
4. Sequencing the cloned gene to verify that the coding sequence was correctly amplified.

Structural Genomics Process(Contd.)

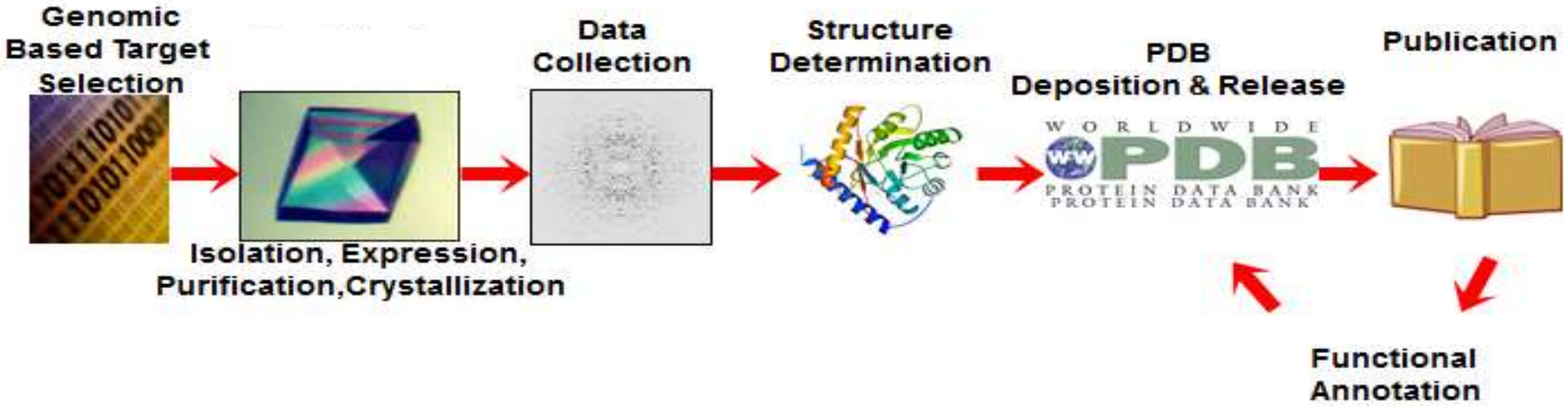
5. Confirming the identity of the expressed protein and characterizing it as a prelude to NMR or crystallographic studies.
6. Obtaining the protein in sufficient amounts and do purification for either approach.
7. Defining suitable crystallization or NMR solution conditions.
8. NMR or X-ray measurement .

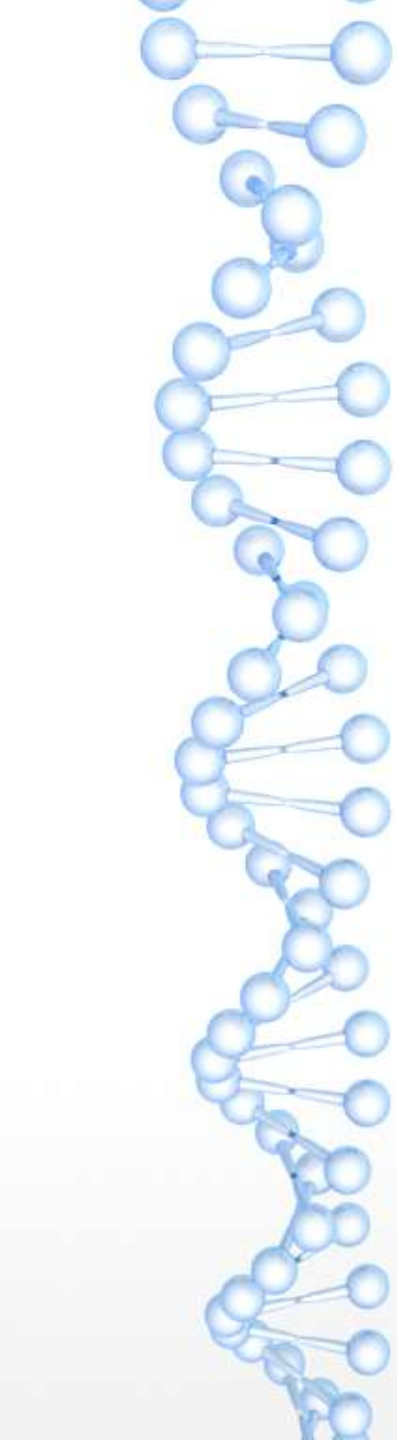
Structural Genomics Process(Contd.)

9. Determining and refining the experimental structure
10. Calculating comparative protein structure models using this new template
11. Making functional inferences from the structure
12. .Publishing the structure to PDB.

- Failures are anticipated at every step, making the process somewhat akin to a funnel.

Structural Genomics Pipeline

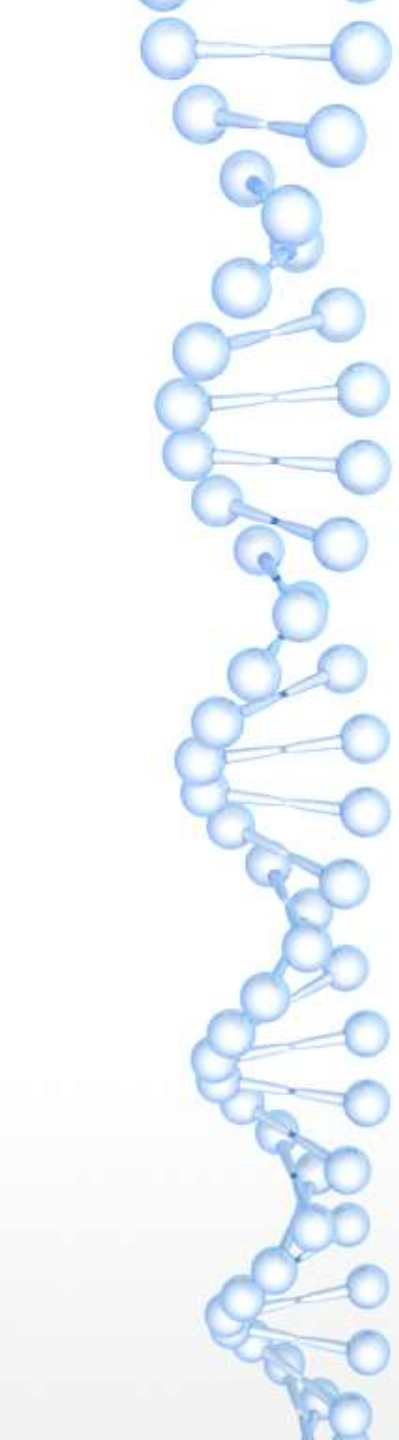




Strategic considerations

- Target selection is the most important strategic issue confronting the structural genomics pilot studies. Their respective performances will be measured in terms of the number of structures determined, what fraction contain novel folds, their impact on biology, and the cost per structure.

- Before selecting as to which structure to target , one needs to be convinced of the social and medical benefits of the structural genomics initiative.



Strategic considerations(Contd.)

- Projects have been made from exhaustive studies of all proteins found in a model organism (*Methanococcus jannaschii, Mycobacterium tuberculosis*) to selectively chosen targets from a large number of different organisms.
- Another important strategic issue concerns money. To date, no clear picture has emerged as to how much it will cost to determine 10,000–20,000 experimental structures..
- Time is also an issue because it is not a simple matter to decide when a structural genomics initiative will be complete.

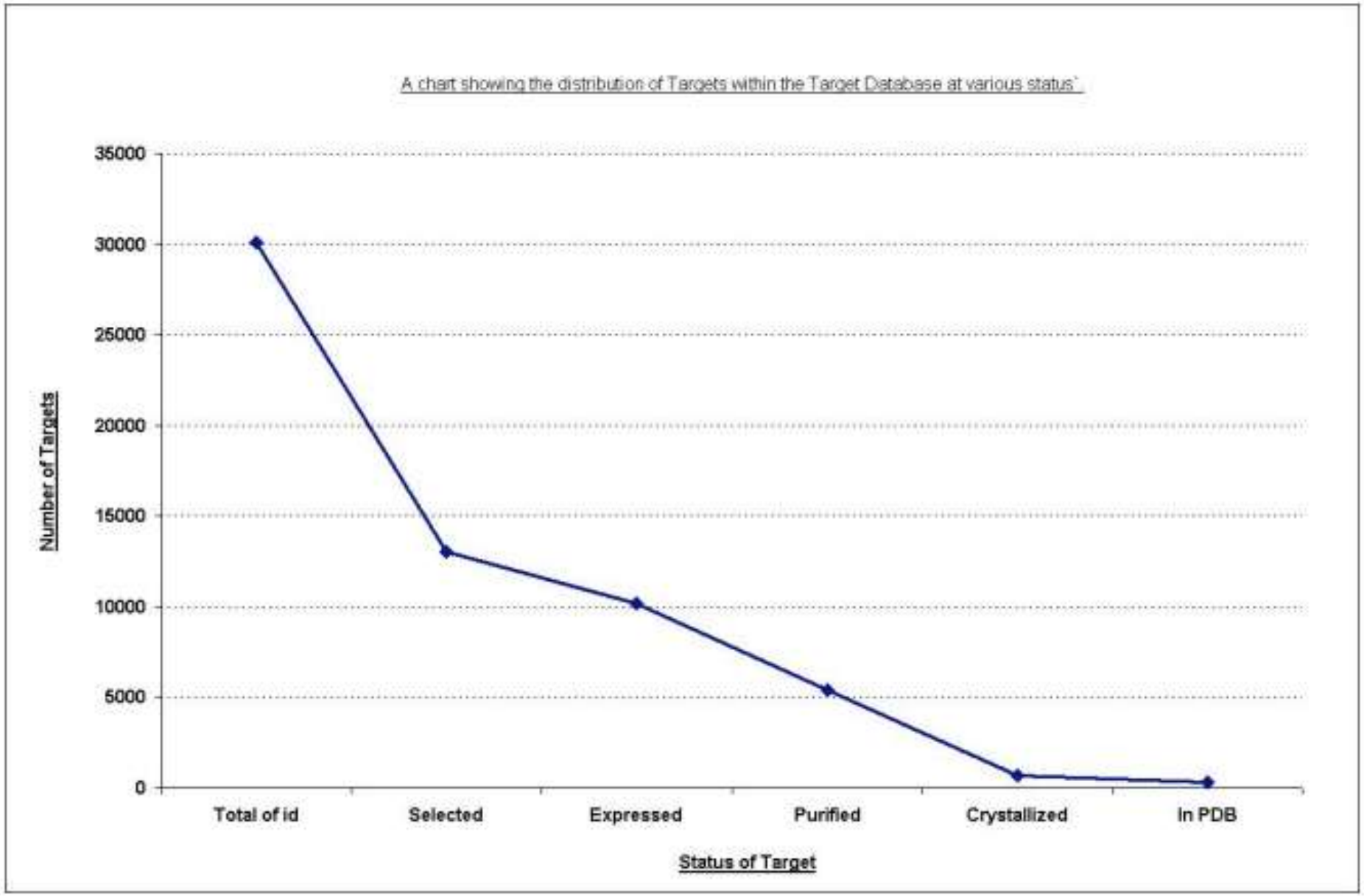
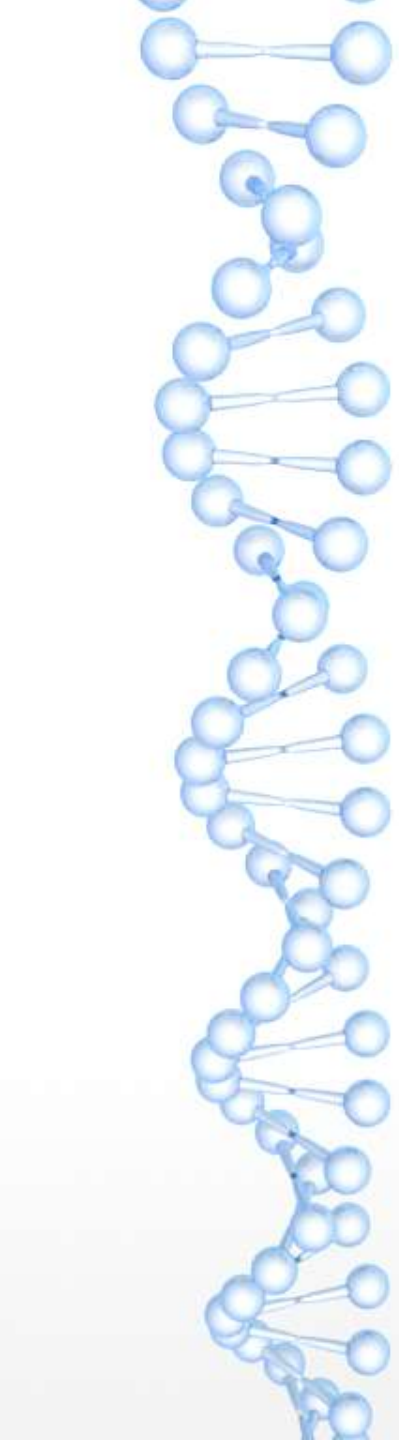


Figure 1 Structural Genomics Targets at Different Stages of Solution (April 1, 2003)

- In the year 2002-2003,314 structures resulting from structure genomics were reported by TargetDB.
- During the same period, a total of 3324 structures were deposited with the PDB
.
- Thus structure genomics is currently contributing approximately 10% of structures to the field of structural biology.

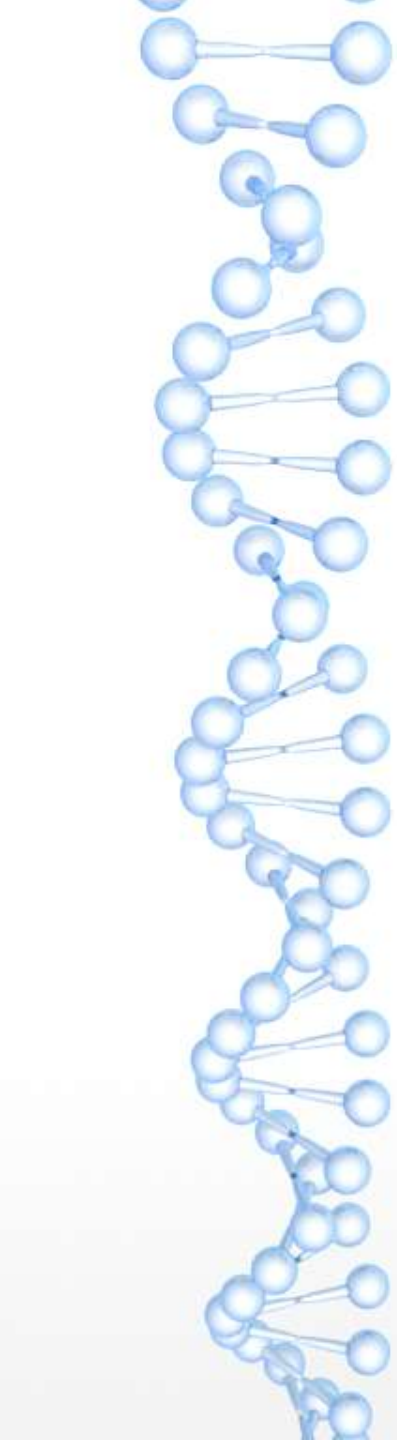


Structural Genomics Initiatives

- The flood of new genomic sequence information together with technological innovations in protein structure determination have led to worldwide structural genomics (SG) initiatives.
- The goals of SG initiatives are to accelerate the process of protein structure determination, to fill in protein fold space and to provide information about the function of uncharacterized proteins. In the long-term, these outcomes are likely to impact on medical biotechnology and drug discovery, leading to a better understanding of disease as well as the development of new therapeutics

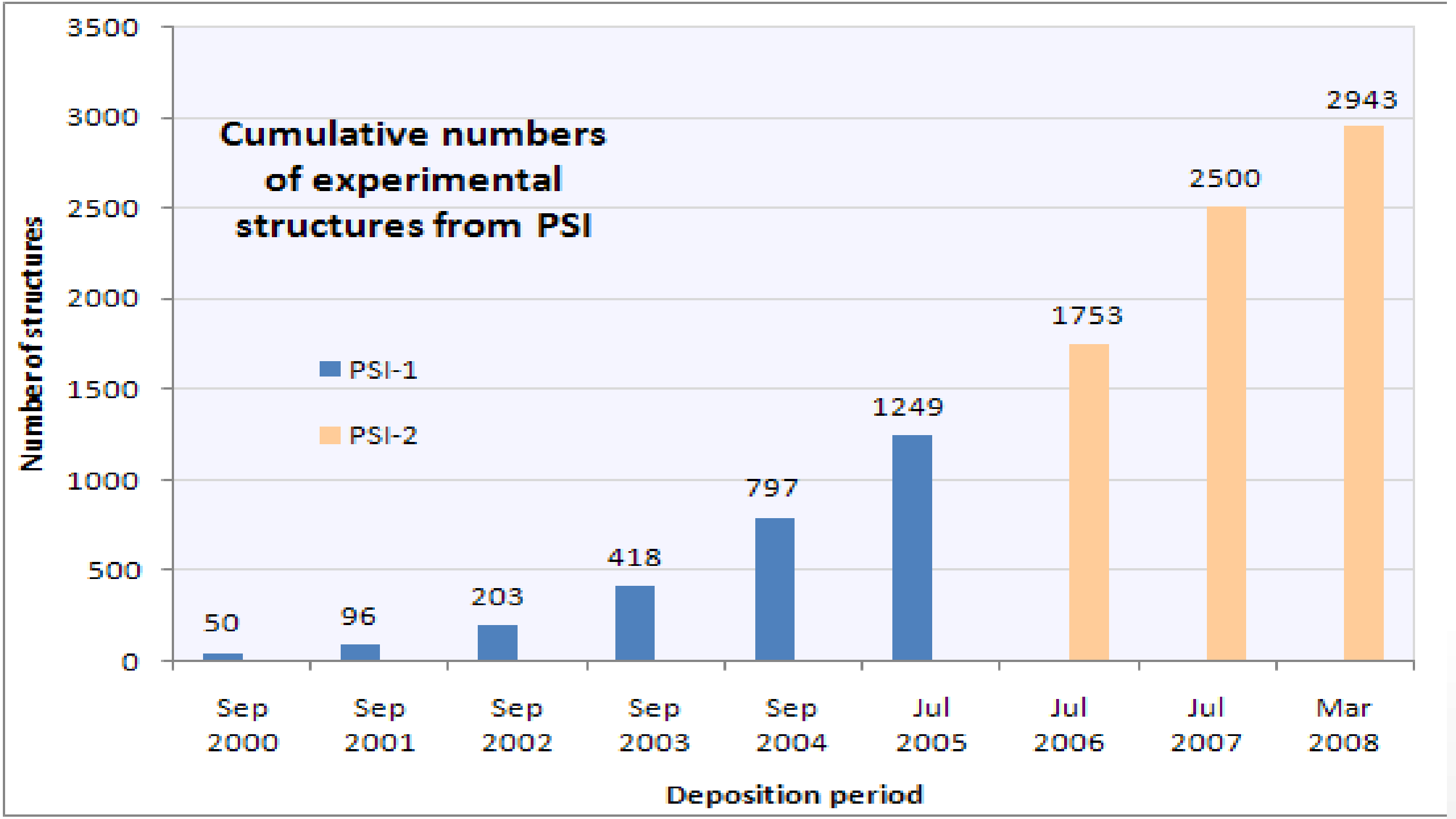
Protein Structure Initiative

- Prominent Structural Genomics projects can be roughly lumped into three groups: the Japan-based program, led by RIKEN, called Protein 3000; the Protein Structure Initiative (PSI) and the efforts from the European research community.
- The most important among them was Protein Structure Initiative(PSI).It was a USA based project established in 2000 by the National Institute of Health (NIH) and National Institute of General Medical Sciences(NIGMS) that aimed at accelerating discovery in structural genomics.



Protein Structure Initiative

- It was the most productive of the initiatives taken for structural genomics, having solved around 3000 new structures between 2000 and 2008.
- The project was organized into three separate phases.
- The first phase of the Protein Structure Initiative (PSI-1) spanned from 2000 to 2005, and was dedicated to demonstrating the feasibility of high-throughput structure determination, solving unique protein structures, and preparing for a the second phase.
- The second phase, PSI-2, focused on implementing the high-throughput structure determination methods developed in PSI-1, as well as addressing bottlenecks like modeling membrane proteins.
- The third phase, PSI : Biology, began in 2010 and consisted of networks of investigators applying high-throughput structure determination to study a broad range of biological and biomedical problems.
- PSI program ended in 2015


Cumulative numbers of experimental structures from PSI
Number of structures
Deposition period
PSI-1
PSI-2
3500
3000
2500
2000
1500
1000
500
0
50
96
203
418
797
1249
1753
2500
2943
Sep 2000
Sep 2001
Sep 2002
Sep 2003
Sep 2004
Jul 2005
Jul 2006
Jul 2007
Mar 2008

Tools and Databases for Structural Genomics

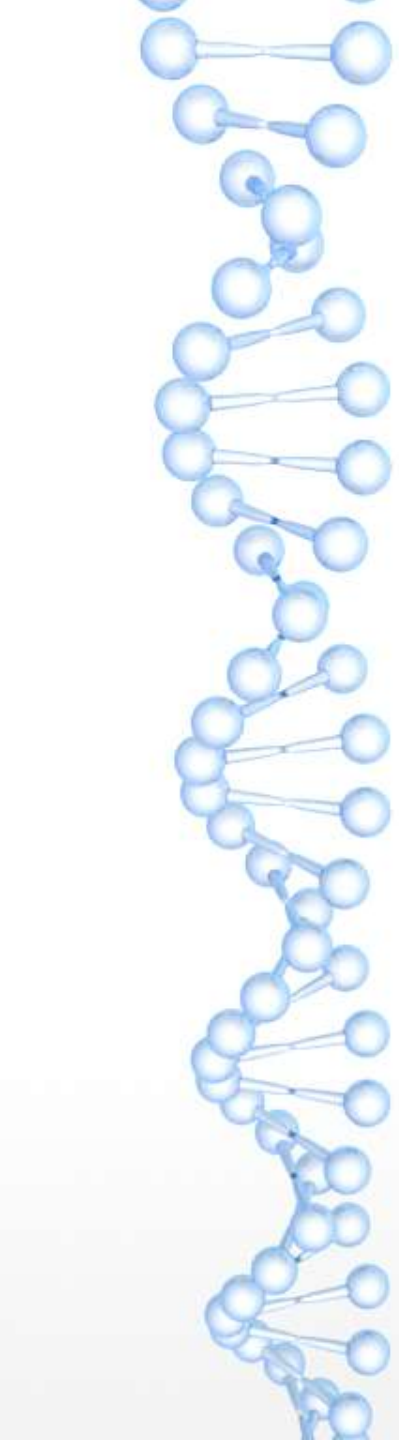
Resource	Description	URL
Tools		
Eisenberg group	Threading tools	http://www.doe-mbi.ucla.edu/People Eisenberg/Projects/
Expasy	Swiss-Prot site contains many sequence and structure searching tools	http://www.expasy.ch/
Gerstein group	Structure prediction of eight genomes comparative genomics	http://bioinfo.mbb.yale.edu/genome/
National Center for Biotechnology Information (Bethesda, MD)	BLAST sequence similarity search tool	http://www.ncbi.nlm.nih.gov/BLAST/
Sali group	Tools for protein structure modeling, including MODELLER	http://guitar.rockefeller.edu/sub-pages/programs.html
Skolnick–Kolinski group	Threading tools, *ab initio* folding tools, FFF library	http://bioinformatics.danforthcenter.org
Thornton group	Library of three-dimensional active site motifs	http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html
Databases		
Protein Data Bank	Database of solved protein structures	http://nist.rcsb.org/pdb/
Expasy	Swiss-Prot protein sequence and structure database	http://www.expasy.ch/
CATH	Protein structure classification database	http://www.biochem.ucl.ac.uk/bsm/cath/
SCOP	Murzin's database of protein structure classification	http://scop.mrc-lmb.cam.ac.uk.scop/

Examples of Structural Genomics

1. *Thermotogo maritima* proteome

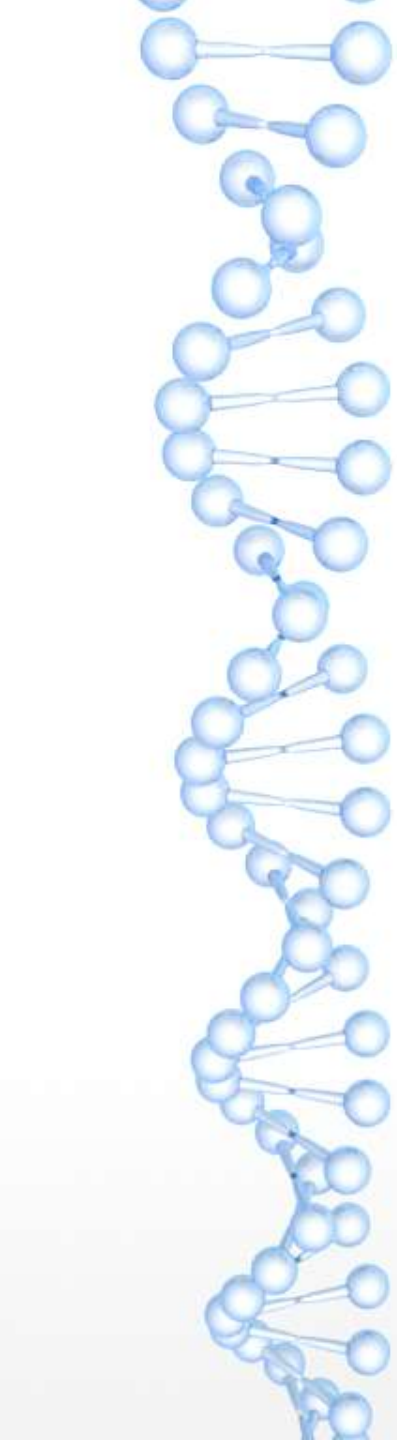
- The Protein Structure Initiative (PSI) solved the structures for all the proteins in *Thermotogo maritima*, a thermophillic bacterium.
- *T. maritima* was selected as a structural genomics target based on its relatively smal genome consisting of 1,877 genes and the hypothesis that the proteins expressed by a thermophilic bacterium would be easier to crystallize.
- E.coli was used to express all the open-reading frames (ORFs) of *T. martima*. These proteins were then crystallized and structures were determined for successfully crystallized proteins using X-ray crystallography.
- Among other structures, this structural genomics approach allowed for the determination of the structure of the TM0449 protein, which was found to exhibit a novel fold(a tetramer with four interconnected active sites, each containing a flavin adenine dinucleotide molecule).

Examples of Structural Genomics(Contd.)



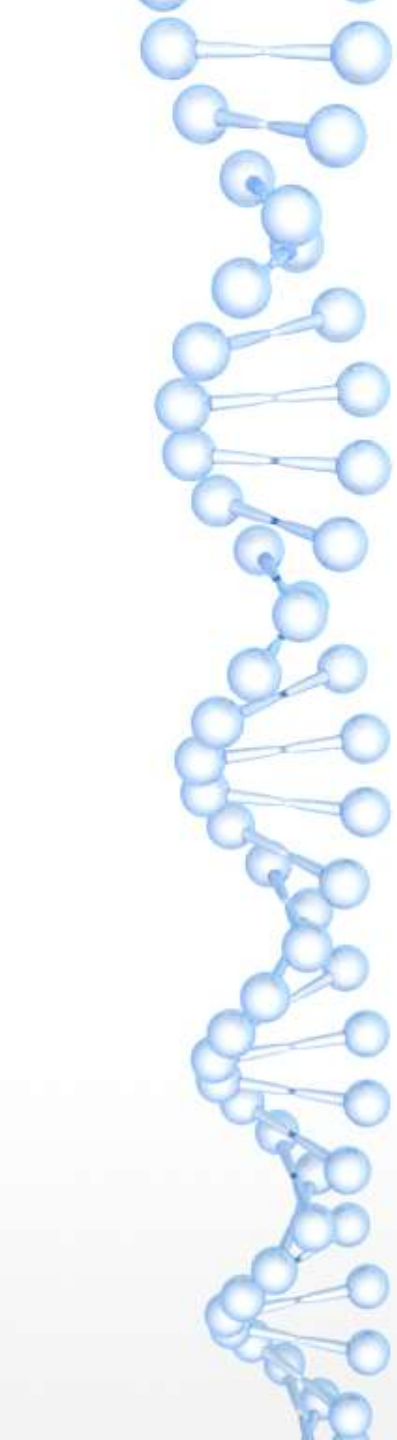
2. *Mycobacterium tuberculosis* proteome

- The structures of potential drug targets in *Mycobacterium tuberculosis*, the bacterium that causes tuberculosis were determined by TB Structural Genomics Consortium.
- The development of novel drug therapies against tuberculosis are particularly important given the growing problem of multi-drug-resistant tuberculosis.
- The fully sequenced genome of M. tuberculosis allowed scientists to clone many of these protein targets into expression vectors for purification and structure determination by X-ray crystallography.
- So far, structures have been determined for 708 of the proteins encoded by *M. tuberculosis.*



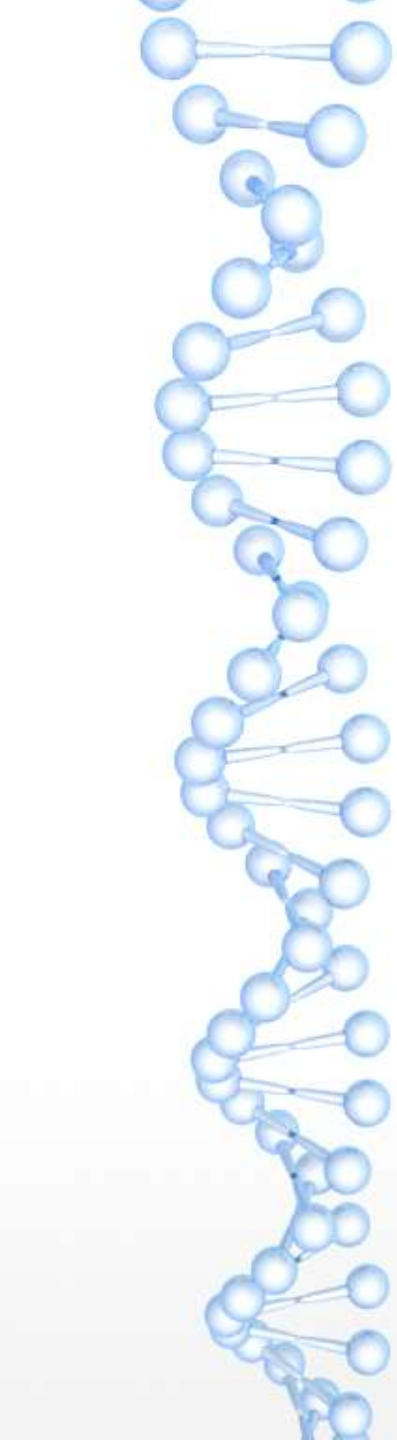
Limitations and challenges to Structural Genomics

- As the structural genomics efforts continues to mature and structure determination for the easy proteins or ‘low hanging fruit’ is completed, attention will naturally turn to the more difficult problems.
- The future challenges include expression and purification of proteins from eukaryotic organisms containing post-translational modifications, membrane proteins and large multi-domain protein complexes.
- There are some regions of protein structure space that will not succumb immediately to either NMR or X-ray methods.



Limitations and Challenges to Structural Genomics(Contd.)

- Novel high throughput, cost-effective expression protocols will need to be developed .
- Membrane protein crystallization continues to represent a considerable technical challenge, but advances in protein purification and crystallization may ease these difficulties.



References

1. Goldsmith-Fischman Sharon,Honig B. 2003 “Structural genomics: Computational methods for structure analysis” Protein Sci. 2003 Sep; 12(9): 1813–1821
2. Burley SK “An overview of structural genomics” Nat Struct Biol. 2000 Nov;7 Suppl:932-4.
3. Bourne PE et al. “The status of structural genomics defined through the analysis of current targets and structures” Pac Symp Biocomput. 2004:375-86.

THANK YOU