Argonne NATIONAL LABORATORY

... for a brighter future

U.S. Department of Energy

UChicago ▶ Argonne LLC

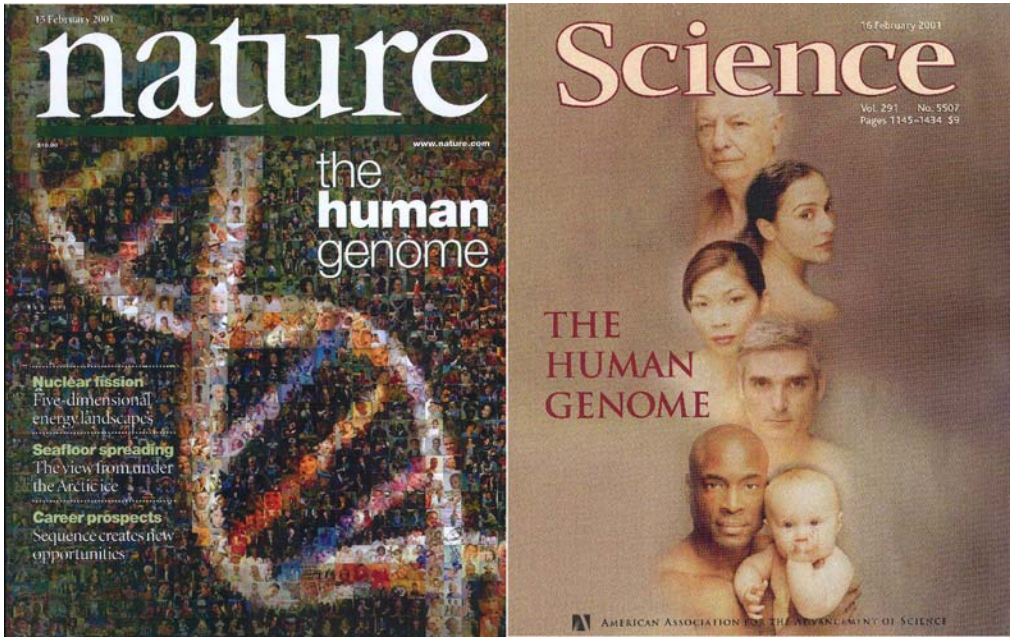
Office of Science U.S. DEPARTMENT OF ENERGY

A U.S. Department of Energy laboratory managed by UChicago Argonne, LLC

Advanced Protein Crystallization Facility at Argonne

Andrzej Joachimiak

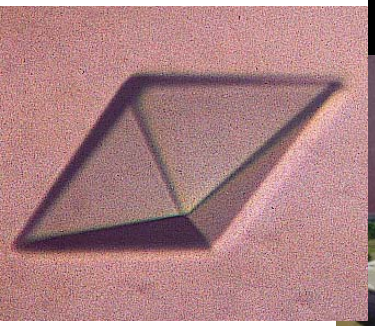
Argonne, October 24, 2007

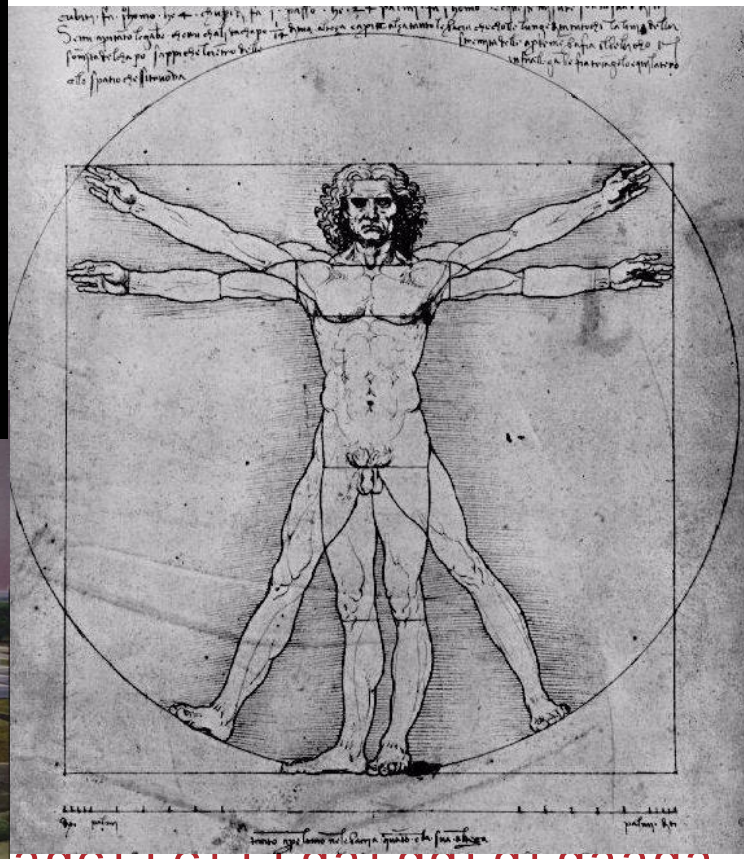


Genome Information Explosion

Challenge: to Interpret Genome Sequence in Term of Function

tgaggagggaagagacatggctaagcaagattattacgagatttaggcgtttccaaaa
cagcggaagagc cctggccatgaaataccacccg
gaccgtaaccag ttaaagagatcaaggaagctta
tgaagttctgac gatcagtatggtcatgctgcgt
ttgagcaaggtg cggcgcagacttcagcgatatt
tttggtgacgtt
cggtgctgattt
ccaaagagatcc
aaaccaggtaca
gcgccagggttt
tgatcaaagatc



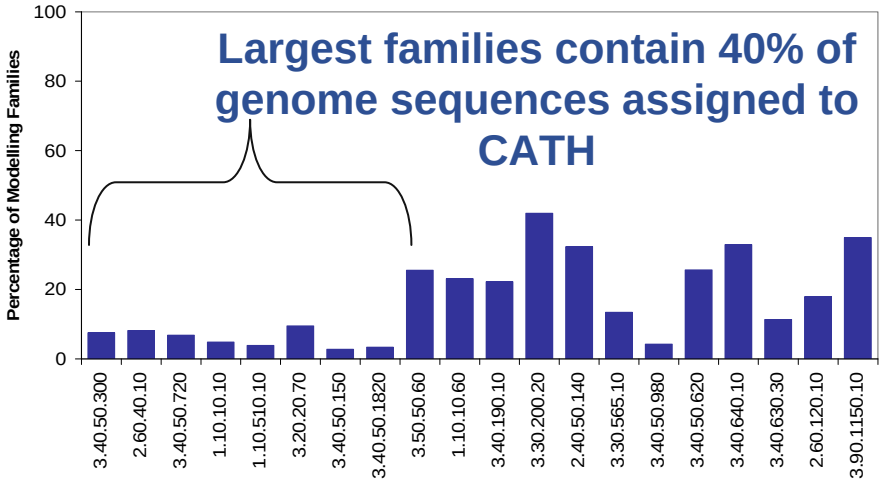


tcaagtctgtc
ccggtaggcct
tggcccaaccg agcttctttgatggtgtgaaga
agttttttgacgacctgacccgctaaaatcatgctctttctgtttt

Proteins for Structural Studies

- 527 completely sequenced genomes
- > 5 million protein gene sequences, ~100,000 protein families, ~250,000 singletons
  - ~40-60% genes have a homologues with known function
  - ~40-50% genes have a homologue with known structure
  - ~4.3 x 104 protein structures in PDB (X-ray & NMR)
  - ~9 x 10^3 structures of non-redundant proteins
  - ~0.9 x 10^3 unique folds, ~1.6 x 10^3 protein superfamilies
  - ~ 6 x 10^3/year - rate of deposition of protein structures
- Protein families annotated by CATH, Pfam and Newfam domains using HMM technology
- Integration of taxonomic data and functional data from databases: EC, GO, COG, KEGG, IntAct, MIPS

  - For the first time complete sets of genes that are required for life are available
  - Minimum set of genes required for life defined
  - Minimum set of genes identified for protein synthesis
  - This has a major impact on (bio) science and technology



In Search of the 2nd Genetic Code (May 1997)

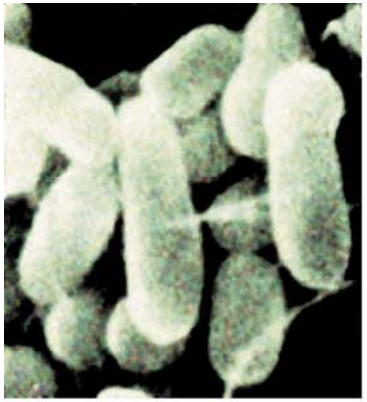
The Role of Fundamental Molecules of Life

A Combined Molecular Biology and APS Initiative

Andrzej Joachimiak, Terry Gaasterland, and Paul A. Bash

"A tractable, and potentially more important, objective would be the identification, through a computational gene sequence analysis, of a core set of genes and product proteins common to all forms of life, whose functions and 3D atomic structures could then be determined using traditional biochemical and biophysical experimental methods. This set of "fundamental" or "universal" molecules of life would form the basic components of a 2nd Genetic Code, from which insights into the structural and functional characteristics of all other proteins may be deduced."

STRUCTURAL GENOMICS



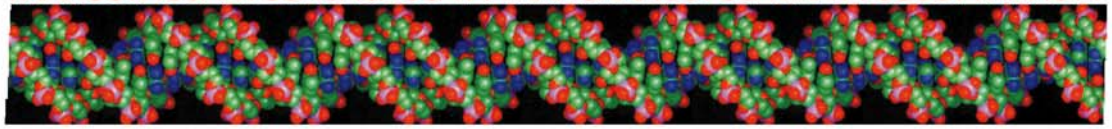
CELLS

GENOMIC DNA
EXTRACTED FROM CELLS

DETERMINE SEQUENCE OF ALL 35,000 - 40,000
GENES CODED BY THE HUMAN GENOME

DNA

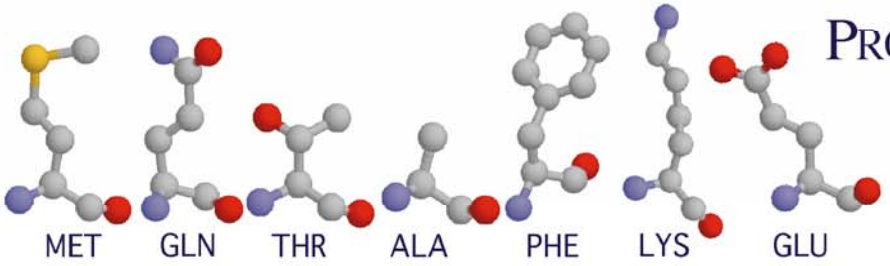
TAAAGCTAGCAATGCAGACAGCATTTAAAGAAACCTGCATAATTCGGGTTAAACTT



MET GLN THR ALA PHE LYS GLU

PROTEIN 3-D STRUCTURE

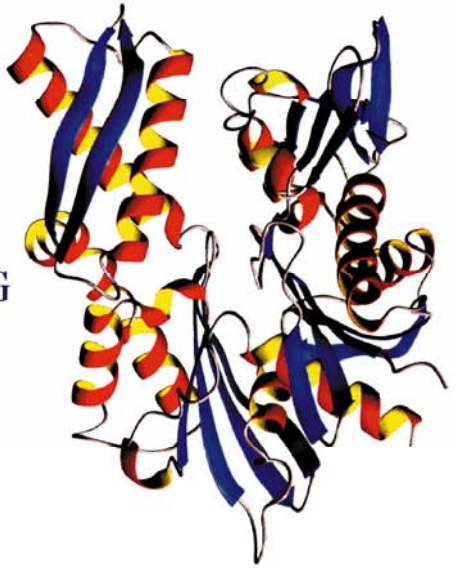
SEQUENCE OF PROTEIN GENE
PRODUCTS CAN BE DEDUCED FROM
THE SEQUENCE OF THE GENOMIC DNA



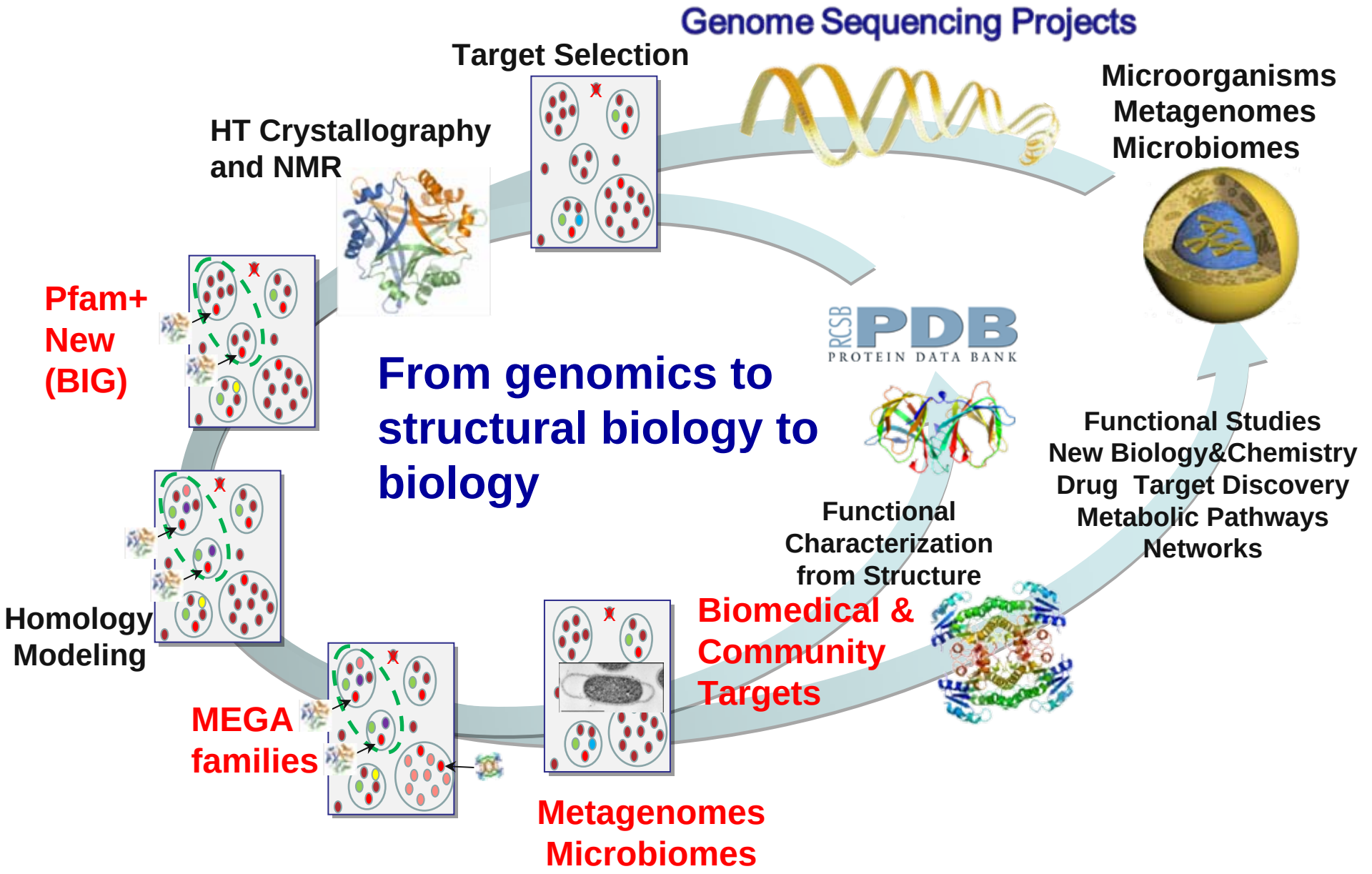
MET GLN THR ALA PHE LYS GLU

PROTEIN AMINO ACID SEQUENCE

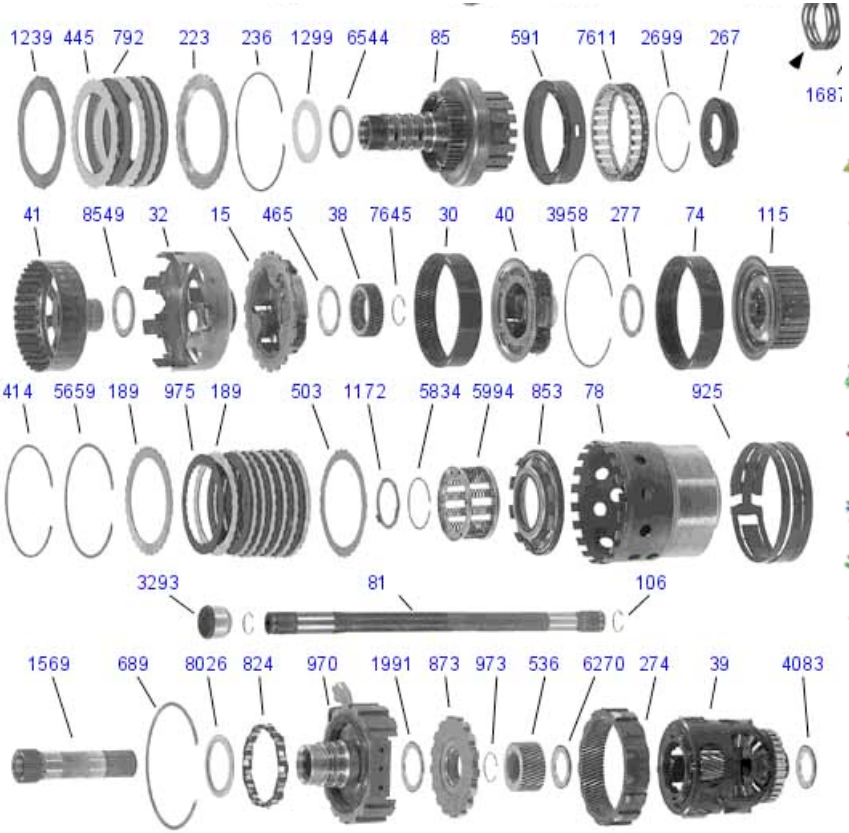
PROTEIN FOLDING

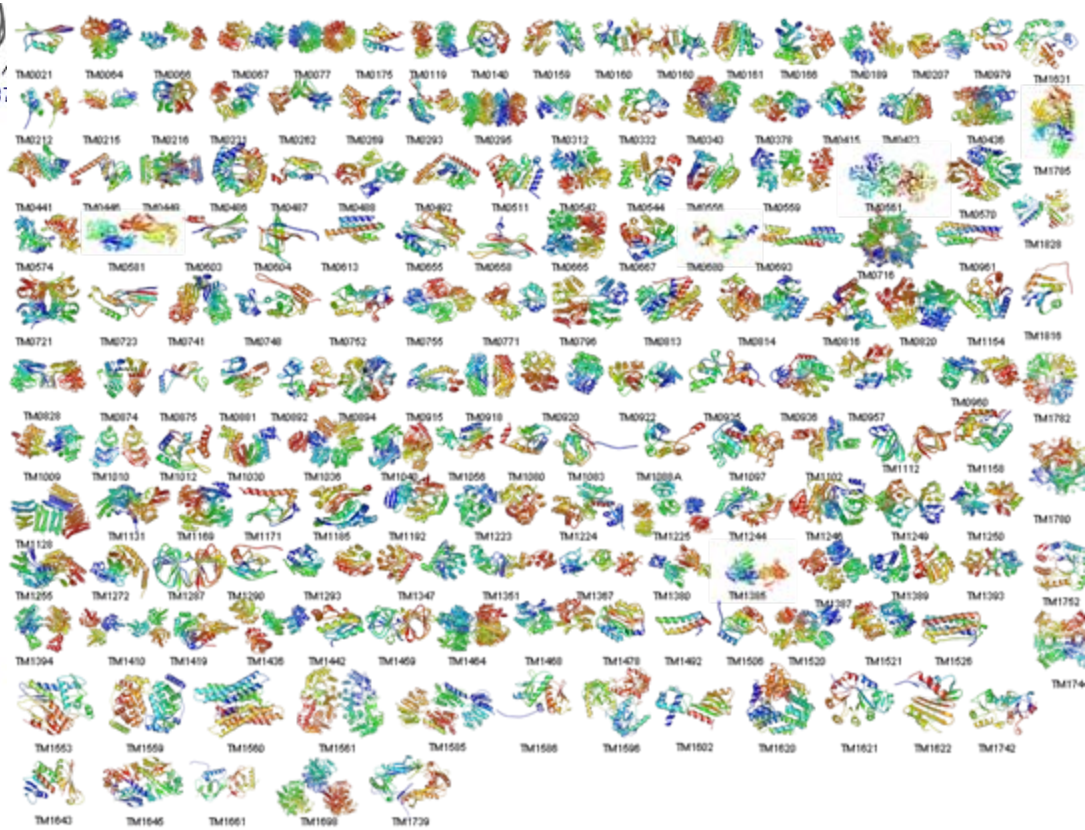


BIOCHEMICAL FUNCTION


Genome Sequencing Projects
Target Selection
HT Crystallography and NMR
Microorganisms
Metagenomes
Microbiomes
Pfam+ New (BIG)
RCSB PDB PROTEIN DATA BANK
From genomics to structural biology to biology
Functional Studies
New Biology&Chemistry
Drug Target Discovery
Metabolic Pathways
Networks
Functional Characterization from Structure
Homology Modeling
Biomedical & Community Targets
MEGA families
Metagenomes
Microbiomes

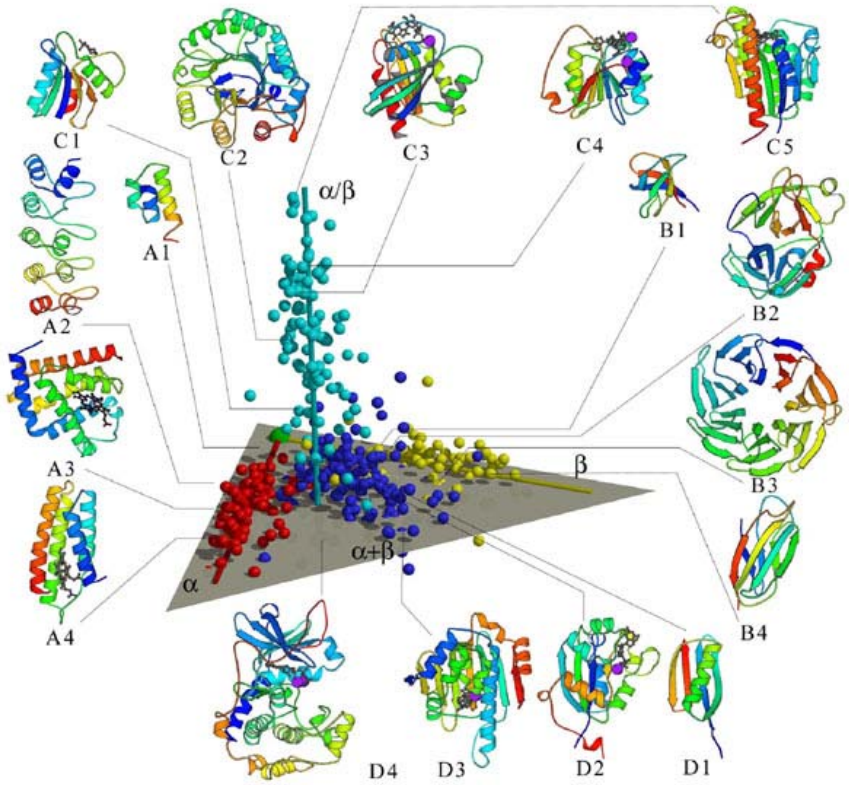
Genomes (and structure galleries) are really lists of parts…





Converting genome sequences into the protein universe to give new biological insights

- How many folds or topologies?
- How many sequences can adopt the same fold?
- How does function evolve within a family?
- Are there still Kingdom-specific families?
- Can we determine function from structure?
- How diverse are metabolic pathways and networks?
- How many novel drug targets are in these pathways & networks ?



Realizing the Potential of the Genome Revolution: The Genomes to Life Program

Marvin E. Frazier,[1] Gary M. Johnson,[2] David G. Thomassen,[1]* Carl E. Oliver,[2] Aristides Patrinos[1]

- Systems Biology
  - ***To understand living organisms, we must characterize the structure and function of all the proteins of a cell***
- Why proteins?
  - Proteins do virtually all the work of the cell
  - The genome sequence can provide us with the identity of a protein - but not its function
  - Sequence homology to well characterized proteins provides some information
  - For the rest, experiments are necessary
  - To carry out experiments, you need to produce the proteins
- Why protein structure?
  - Protein function is associated with protein structure
  - Therefore structure can provide critical information about protein function
  - Structural information is important in biology, medicine and biotechnology

Protein Structure Initiative (PSI) Mission:

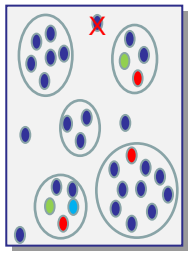
To make the three-dimensional atomic level structures of most proteins easily available from knowledge of their corresponding DNA sequences

PSI Goals:

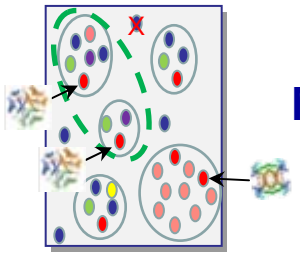
- **Increase the number of sequence families with structural representatives**
- **Make homology models available for most sequenced genes**
- **Continue methodology and technology development**
- **Increase biological impact of structures**

PSI Stages:

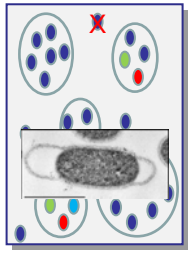
- **2000-2005 Pilot Phase**
- **2006-2010 Production Phase**



BIG Families
Pfam+, New



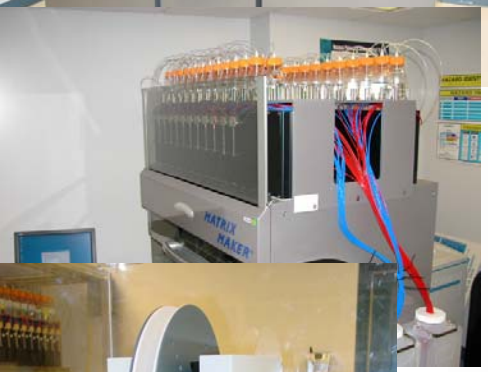
MEGA Families

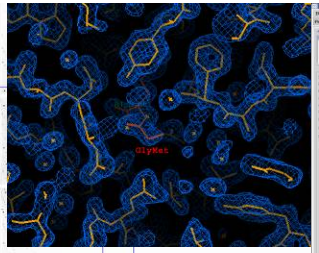


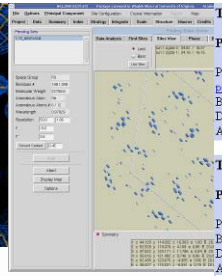
META families
Metagenomes
Microbiomes

Midwest Center for Structural Genomics

PSI Protein Structure Initiative







Target APC24875

PDB 1T20 PDBsum

Protein BC2969 (111 aa) COG2329 related proteins

Bacillus cereus

Deposited: 09 Jul 2004

Authors: Y. Kim, L. Lezondra, A. Joachimiak

1T20

Target APC24929

PDB 1T06 PDBsum

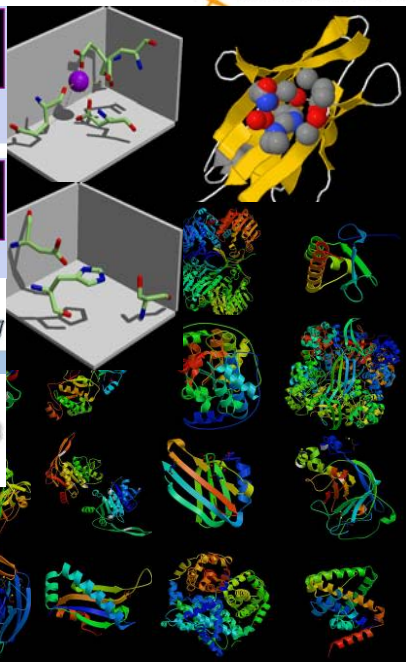
Protein BC3264 (235 aa) related proteins

Bacillus cereus

Deposited: 07 Apr 2004

Authors: R. Zhang, R. Wu, S. Moy, A. Joachimiak

1T06

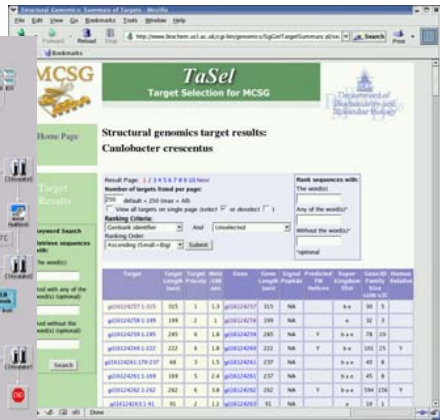


MCSG Comparative Protein Structure Modeling

A collaboration between the Midwest Center for Structural Genomics and the CATH project at University College London

ProFunc

ATGAGAATGAAGCGATTTTTAA
TTTGGTTACAAAATTGTTCAAA
AAATTACAACAAAAATTTCAAA
TTTGAATCACTTGAAAAAATAA
TCATTAAATTATCCAGTACTAC
GAGCGAGAACATCGACGTAAAG



APC35529	RESTP0061		-							Active
APC35531	RESTP1011	+	+	+	+					Active
APC35532	RESTP2755	+	+	+	+	-				Stopped (failed trials)
APC35534	RESTP1017	+	+	+						Active
APC35536	RESTP2758	+	+	+	+	+	+	SAD	1XA0	In PDB
APC35540	RESTP2765	+	+	+	+					Suspended
APC35541	RESTP1026	+	+	+	+					Stopped-homolog solved by others
APC35544	RESTP1038		-							Active
APC35545	RESTP1039	+	-/+	+	+					Active
APC35547	RESTP0099	+	+	+						Active

HTP Protein Crystallization

Commercial screens

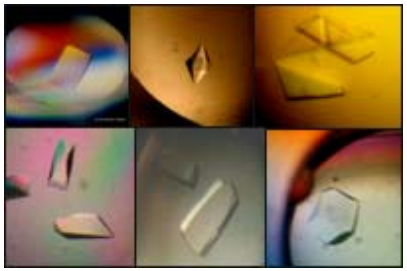
Custom screens making

Dispensing of reservoirs

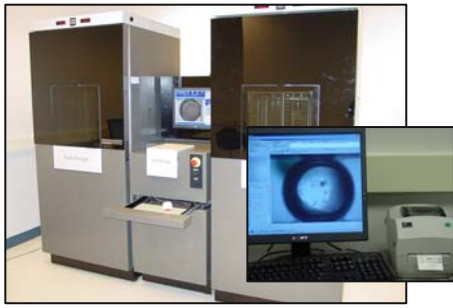
Set up crystallization

- Initiation & optimization of nanoliter crystallization for structural biology.
  - Volume optimization.
  - Robotic crystallization using nanoliter technology.
- PSI contributed to commercial crystallization screens refinement.
- Optimized crystallization screens tested on thousands of proteins.
- New crystallization screens development.
- Cryo-condition databases.
- Ligand screening.
- Affinity tag as a variable in crystallization.

X-ray test & Data collection



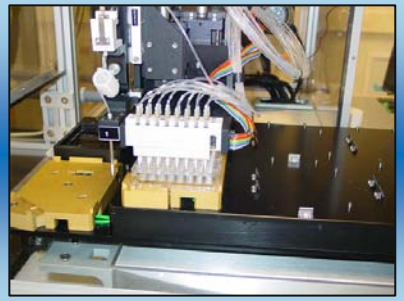
Crystal freezing



Visualization *Manual and robotic*

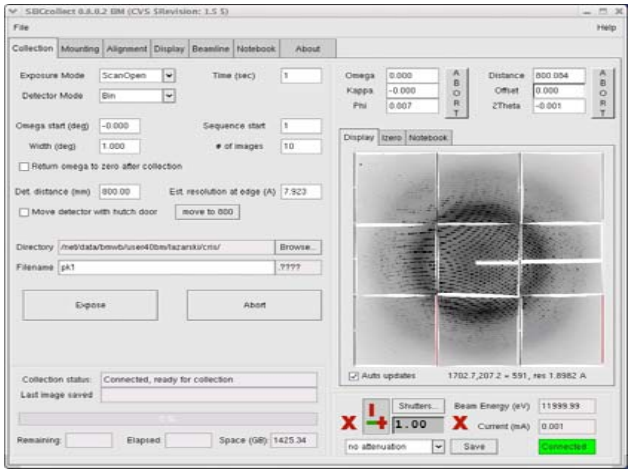
Crystal optimization

DB



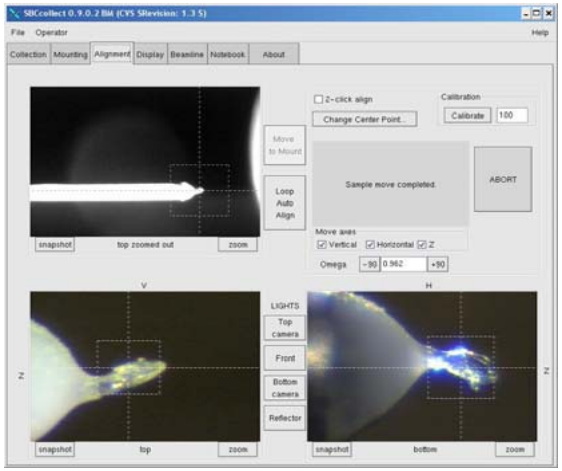
APS Synchrotron Beamlines for Macromolecular Crystallography

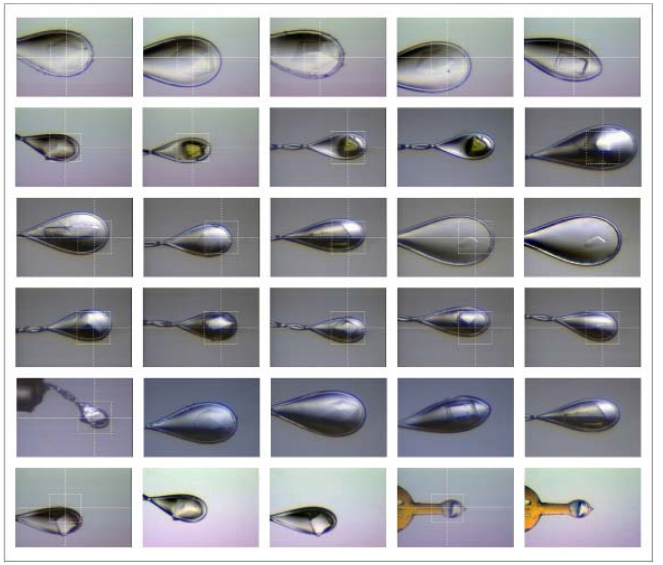
SBC beamlines contributed data to 1719 deposits and 769 publications



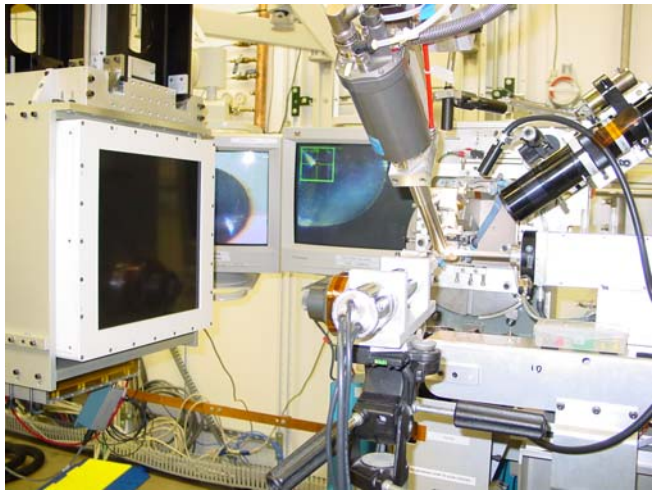
19ID kappa goniostat and Q315 detector

SBCcollect loop auto-centering Crystal “point-and-click” centering

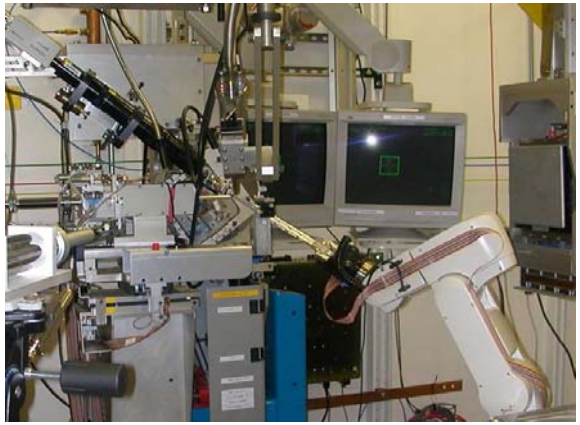




19BM kappa goniostat, SBC3 detector and ROBAC



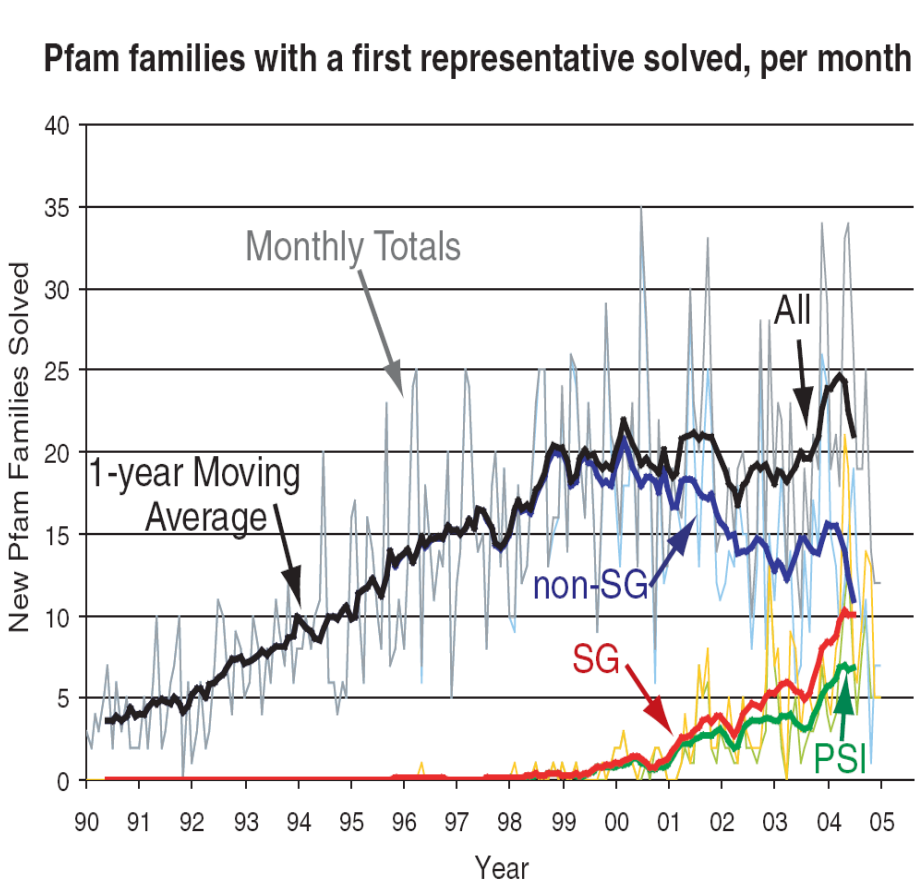
HKL3000 semi-automated structure determination

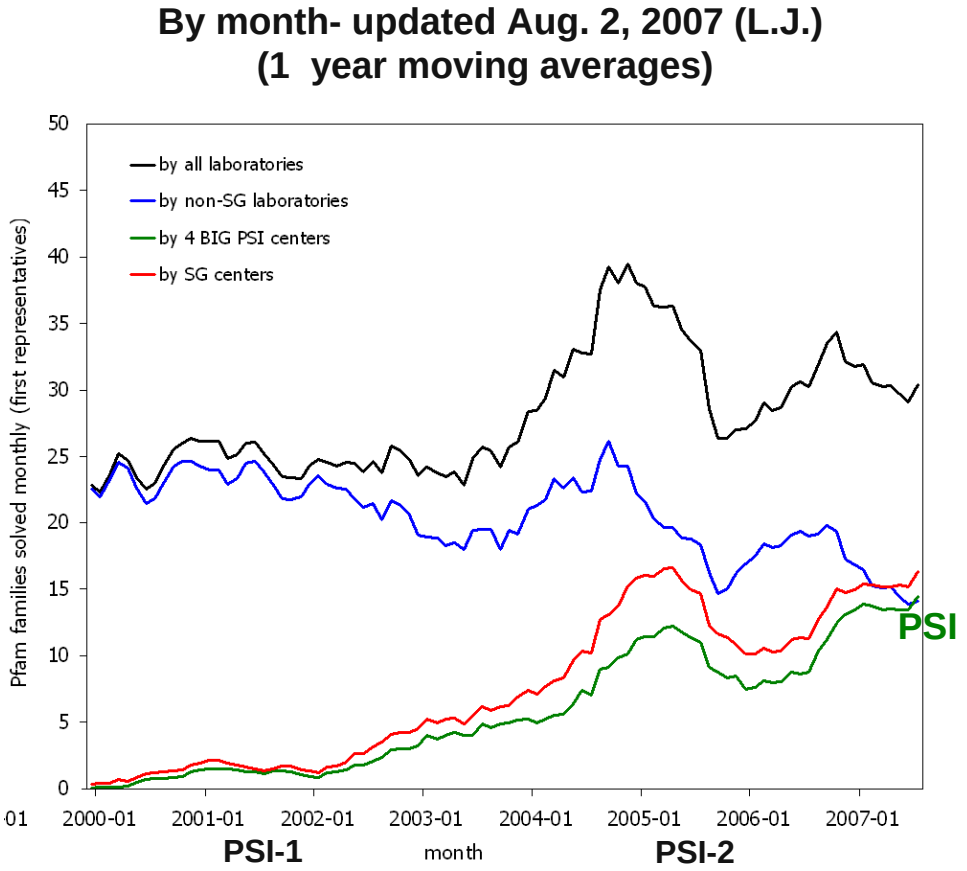


The PSI is making an impact on protein folds and families

The Impact of Structural Genomics: Expectations and Outcomes

John-Marc Chandonia and Steven E. Brenner, Science 311:347-351(2006)



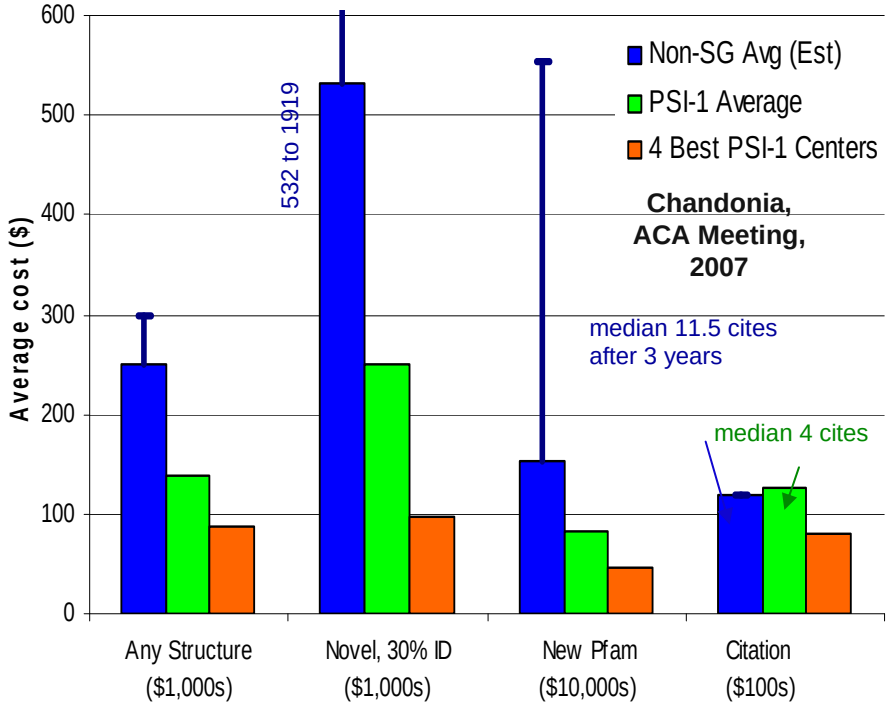


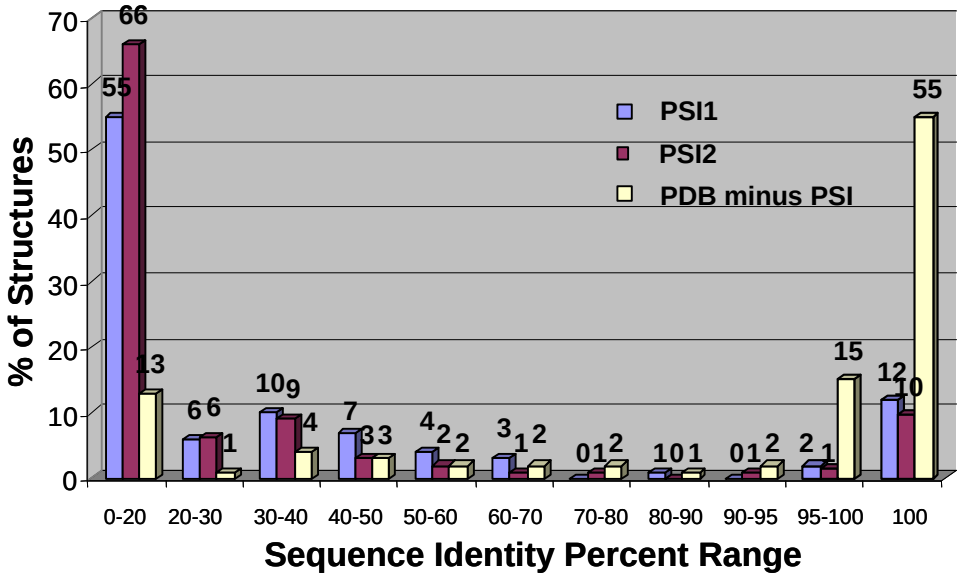
PSI is Driven by Technology, Process Integration and Parallel Processing

Source: PDB 08/14/2007

• Cloned Targets	89,814	
• Expressed	53,255	(59%)
• Purified Soluble	18,327	(23% =100%)
• Crystallized	6,249	(34%)
• HSQC	1,349	(7.5%)
• Diffracting Crystals	2,839	(15.4%)
• Target Structures in PDB	2,182	(12%)
• Total Structures in PDB	2,541	
• NMR structures	221	(8.7%)

- **Technologies for HT and for improving success rates**
- **Target selection emphasizes protein families**
- **Application to a proteome – *T. maritima, human pathogens***
- **Inclusion of metagenome targets**





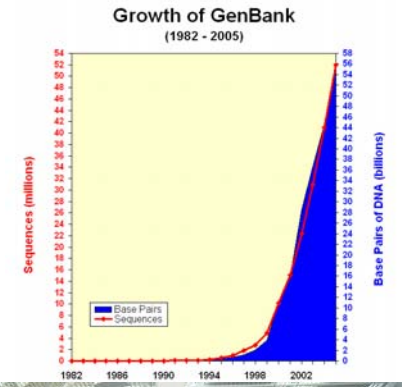
World-wide Structural Genomics Programs

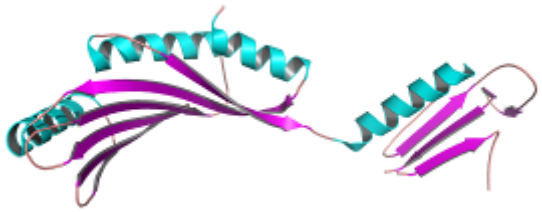
- **May 2006** Protein Structure Initiative - 2000th structure in PDB
- **Nov.2006** Midwest Center for Structural Genomics - 500 X-ray structures deposited to PDB
- **July 2005** Start of Phase-2 of Protein Structure Initiative
- **Mar.2005** Midwest Center for Structural Genomics - 250 X-ray structures deposited to PDB
- **Feb 2005** Protein Structure Initiative - 1000th structure solved
- **Dec 2004** Midwest Center for Structural Genomics - 112 x-ray structures deposited to PDB in a single year
- **Nov 2004** ISGO International Conference on Structural Genomics (ICSG 2004) (Washington DC, USA)
- **Jan 2004** RFA for next generation structural genomics centers in USA
- **Nov 2003** Midwest Center for Structural Genomics - 100 th structure deposited in the PDB
- **Oct 2003** Joint Center for Structural Genomics - 100 th structure solved
- **Early 2003** RIKEN - 100 th structure deposited in PDB
- **Oct 2002** ISGO International Conference on Structural Genomics (ICSG 2002) (Berlin, Germany)
- **Apr 2002** Start of the National Project on Protein Structural and Functional Analyses in Japan
- **Mar 2002** Start of the European drive for post-genome research, Structural Proteomics in Europe (SPINE)
- **Sep 2001** Start of the new two centers for NIGMS Protein Structure Initiatives in USA
- **Jun 2001** Formation of Plexxikon
- **May2001** Presentation of NIGMS Structural Genomics Initiative at the BERAC meeting, Washington DC
- **Apr 2001** Start of International Structural Genomics Organization (ISGO)
- **Apr 2001** Second International Structural Genomics Meeting (Airlie House, USA) - Start of ISGO
- **Jan 2001** OECD/CSTP/GSF Further Study on Structural Genomics (Paris, France)
- **Nov 2000** OECD/GSF Contact Group Meeting (Yokohama, Japan)
- **Nov 2000** International Structural Genomics Task Forces Meeting, (Yokohama, Japan)
- **Nov 2000** International Structural Genomics Task Forces Meeting, (Yokohama, Japan)
- **Nov 2000** International Conference on Structural Genomics 2000 (ICSG 2000) (Yokohama, Japan)
- **Sep 2000** Start of the NIGMS Protein Structure Initiatives in USA with seven Centers
- **Sep 2000** Structural Genomics: From Gene to Structure to Function (Cambridge, UK)
- **Aug 2000** Formation of Affinium Pharmaceuticals (formerly Integrative Proteomics)
- **Jun 2000** OECD/Global Science Forum, Structural Genomics Workshop (Florence, Italy)
- **Apr 2000** First International Structural Genomics Meeting (Hinxton, UK)
- **Jan 2000** OECD Committee for Scientific and Technological Policy proposal of initiating study of structural genomics
- **Dec 1999** Formation of Astex Technology
- **Dec 1999** Formation of Structural Genomix (formerly Protarch)
- **Jun 1999** Call for grant applications for NIGMS/NIH pilot projects
- **Feb 1999** Formation of Syrrx (formerly Agencour)
- **Feb 1999** O'Hare meeting on integrated approach to determining structures of a fundamental protein structures
- **Oct. 1998** Structure-Based Functional Genomics meeting at Avalon in USA
- **1998** Start of the initial pilot projects in Germany, Canada, and USA
- **1997** Start of the New Jersey Initiative in Structural Genomics and Bioinformatics
- **Jan 1998** The workshop on Structural Genomics (Argonne, IL, USA)
- **1997** Initiating study of structural genomics at DOE and NIGMS/NIH in USA
- **1997** Start of the New Jersey Initiative in Structural Genomics and Bioinformatics
- **May 1997** Start of "The periodic table of fundamental folding units" project (Argonne, IL, USA)
- **Apr 1997** Start of structural genomics pilot project at RIKEN Institute
- **Feb 1995** LBNL structural genomics expression/crystallization technology development initiated
- **1995** Proposal of structural genomics projects in Japan

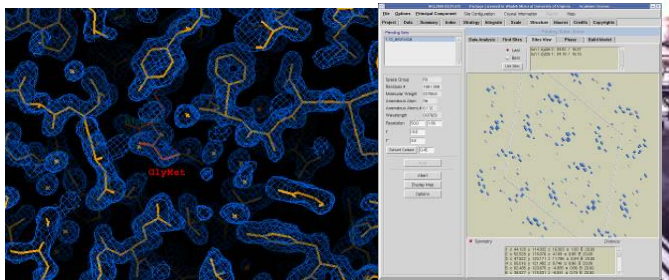
Shapiro L, Lima CD. The Argonne Structural Genomics Workshop: Lamaze class for the birth of a new science. Structure. 1998 6:265-7.

Major Changes in Structural Biology

- We in a process of conversion structural biology from “cottage industry” to “large-scale science”.
- What factors contributed to these changes?
  - Rapid advances in genome sequencing – significantly increased the number of potential targets for structure determination
  - Maturation of molecular biology and proteomics technologies – made more samples are available for structural studies
  - Development dedicated synchrotron facilities with insertion devices – increase in beam flux and brilliance (small crystals, large assemblies, membrane proteins)
  - Cryo-crystallography – effectively reduced radiation damage
  - Phasing using anomalous signal – *in vivo* incorporation of seleno-methionine into proteins and chemical incorporation of Br into nucleic acids
  - Advances in software and computing – increased success rate and reduced time to analyze data, determine and refine structures.

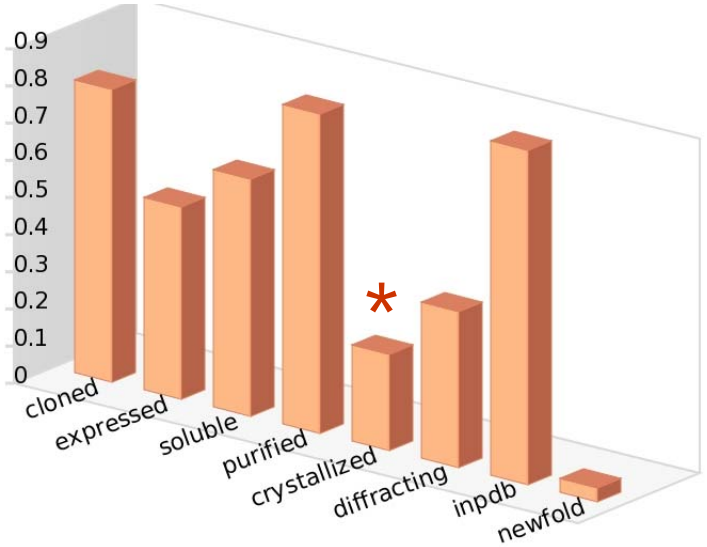






Major Bottlenecks in Structural Biology

- Efficient production of high-quality proteins for structural studies
- Efficient production of high-quality protein crystals
- High-throughput and effective screening for crystals of protein complexes with inhibitors
- Fast feedback from synchrotron experiments to protein cloning, production and crystallization facility
- Several classes of proteins are not compatible with current technologies
- New state-of-the-art facility is needed near the APS and BIO to take advantage of existing facilities, expertise and technology and provide opportunities for the future



APCF Project

- The APCF will establish a state-of-the-art, highly automated laboratory and scientific-collaboration facility to produce proteins and protein crystals needed to take full advantage of ANL's capacity for determining the three-dimensional structures of proteins.
- The APCF will be more efficient than any other facility in the world.
- The APCF will allow the Argonne's structural genomics and structural biology projects to establish new specialized laboratory space devoted to structure determination of more challenging classes of proteins and assemblies that are currently not being actively pursued.
- The APCF will provide the necessary laboratories and modern computer space to support system biology projects.
- The APCF will also support plans to pursue additional programmatic support and new equipment.
- The APCF will provide advanced user facilities at the APS.

APCF Project

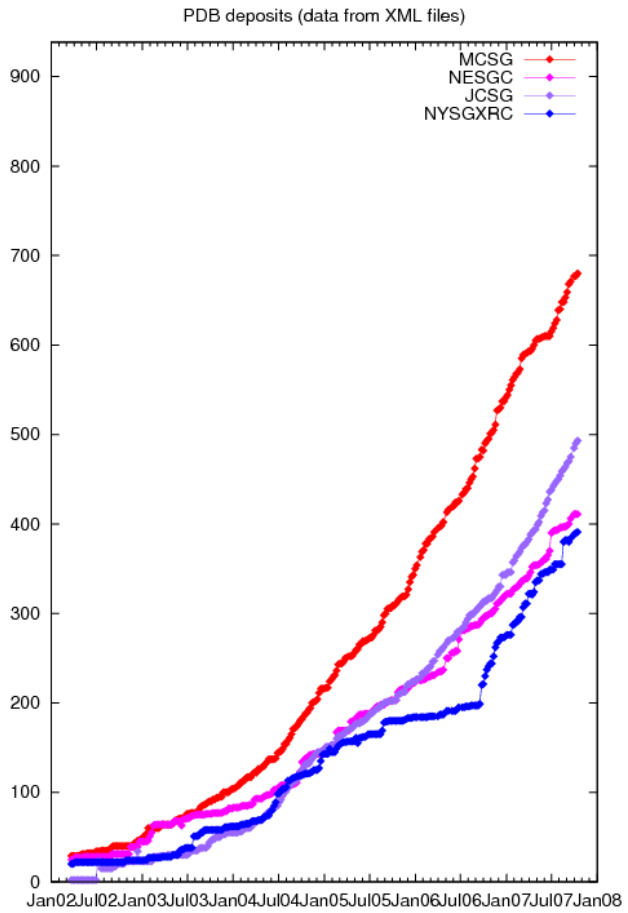
- Production of high-quality protein crystals for structure determination is currently the limiting factor in the protein structure determination pipeline.
- The existing BIO facility's infrastructure is not compatible with the requirements and goals of tomorrow's high-throughput protein crystallization technologies.
- The existing BIO facility, designed and constructed in the 1950s, does not have the required temperature, humidity and vibration controls conducive to high-throughput protein crystallization.
- Current and future high-throughput crystallization technologies (such as nano- and pico-liter crystallization screening) depend on advanced laboratory space with the capability to adjust environmental settings such as humidity and temperature.

APCF - Structural Genomics Component

- **The APCF laboratory space will host integrated robotic workstations for preparation of crystallization formulations, setting up crystallizations, crystal detection and visualization systems, crystal extraction and preservation, and computer and network environments for data storage and analysis.**
- **HTP depend on the application of thousands of simultaneous experiments using controlled laboratory environmental settings and specialized laboratories for structural studies.**
- **The APCF will provide a state-of-the-art facility with the space and environmental controls required to increase throughput and maximize the MCSG's potential. It will allow the MCSG to streamline the process from protein to structure, improve efficiency and safety, and reduce time and cost.**
- **It is expected that the new facility will increase crystal production by at least an order of magnitude. This is not currently possible with ANL's existing facilities.**
- **Additional space is also required to allow the MCSG to establish new specialized laboratories devoted to classes of proteins and assemblies that are currently not being actively pursued. These classes include membrane proteins and protein/protein and protein/nucleic acid assemblies. This additional space will allow work on many new types of protein crystallization technologies, thus expanding the structural coverage of protein space and enhance biomedical and biotechnological capabilities.**
- **Additional space will also support research work of recently funded Center for Structural Genomics of Infectious Diseases.**

Structural Biology Resources at Argonne

- Advanced Photon Source (APS) – a state-of-the-art 3rd generation synchrotron facility – a major impact on biological research
  - Structural Biology Center – DOE/OBER
    - *national user facility (APS, Bioscience)*
  - Midwest Center for Structural Genomics – NIH/NIGMS
    - *regional genomic center (Bioscience, APS)*
  - Center for Structural Genomics of Infectious Diseases - NIH/NIAID
    - *regional genomic center (Bioscience, APS)*
  - GMCA-CAT – NIH/NIGMS/CA
    - *national user facility (APS, Bioscience)*



APCF - Systems Biology Component

- Argonne and the University of Chicago are addressing the challenges of twenty-first century biology through the formation of the joint Institute for Genomics and Systems Biology (IGSB).
- This Institute will recruit new investigators and establish computational and experimental tools at both Argonne and the University.
- The efforts of this Institute will be interfaced with the biological sciences at Argonne to build novel research programs that address the problems of understanding biological processes in complex, multi-scale systems.
- Argonne's ongoing systems biology work, carried out by the IGSB and the BIO Division, includes both experimental and computational work.
- The APCF's state-of-the-art laboratories will house a portion of the systems biology at Argonne.
- The APCF project will provide the necessary modern computer space to support the work.

APCF Scope

- The APCF will satisfy the facility requirements of the Argonne's structural genomics and structural biology research.
- The APCF will satisfy a portion of the facility requirements of Argonne's systems biology work. Preliminary facility requirements currently include:
  - Laboratory space for robotic, automated production, purification, and crystallization of proteins.
  - Laboratory space for structural biology procedures involving classes of proteins and assemblies that represent new avenues of research.
  - Laboratory space for systems biology work.
  - Bio-Safety Level 2 (BSL-2) laboratory space.
  - Computer support space.
  - An electron microscopy suite outfitted to accommodate an intermediate voltage electron microscope and ancillary equipment for cryo-electron microscopy.
  - Compliance with environmental control criteria necessary to support the science.

APCF Infrastructure Elements

- Highly flexible infrastructure
- Support for HTP robotic environment
- Wireless environment
- High throughput for information/databases
- Electricity and Water
- High pressure air/natural gas/helium/vacuum
- BLS2+ capability
- Water purification systems
- Centralized Liquid Nitrogen
- Vibration control
- One radioactive exhaust system in research area
- Chemical exhaust (air quality related to contamination)
- MEPFP systems typical for bio labs w/air sensors
- Constant air quality [temp/humidity]
- Central water system for cooling
- Emergency back-up for freezers/multiple generators

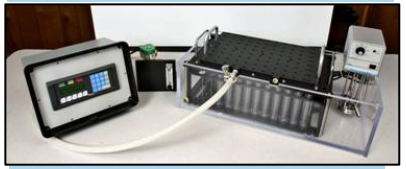
APCF Intelligent Laboratory Infrastructure

- Requirements
  - Automate routine and common laboratory procedures
  - Provide intelligent support for laboratory workers
  - Improve efficiency of lab workers
  - Improve reliability and repeatability of processes
- Deliverables
  - Unified interface for controlling instruments and LIMS
  - Automation of protocols
  - Handheld interfaces

Integrated Custom and Commercial Instrumentation

- Cloning robotics
- Gene design
- Parallel fermentation
- Protein expression vectors
- Affinity purification
- Secondary purification
- Automated protein production
- Nano- and picodrop crystallization
- Fine screening
- Plate imaging
- HT diffraction screening
- Automated sample changer
- Automated structure determination

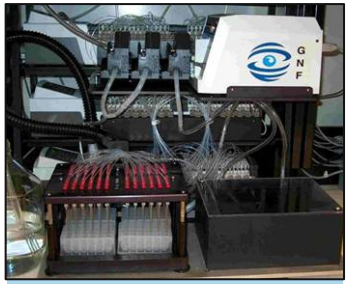
cloning robotics



parallel fermentation

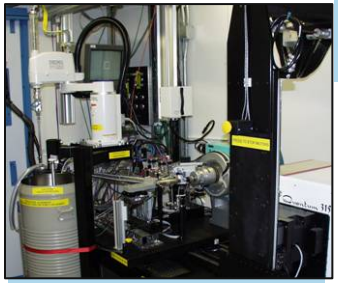
parallel affinity purification

crystal plate setup



fine screen setup

plate imager

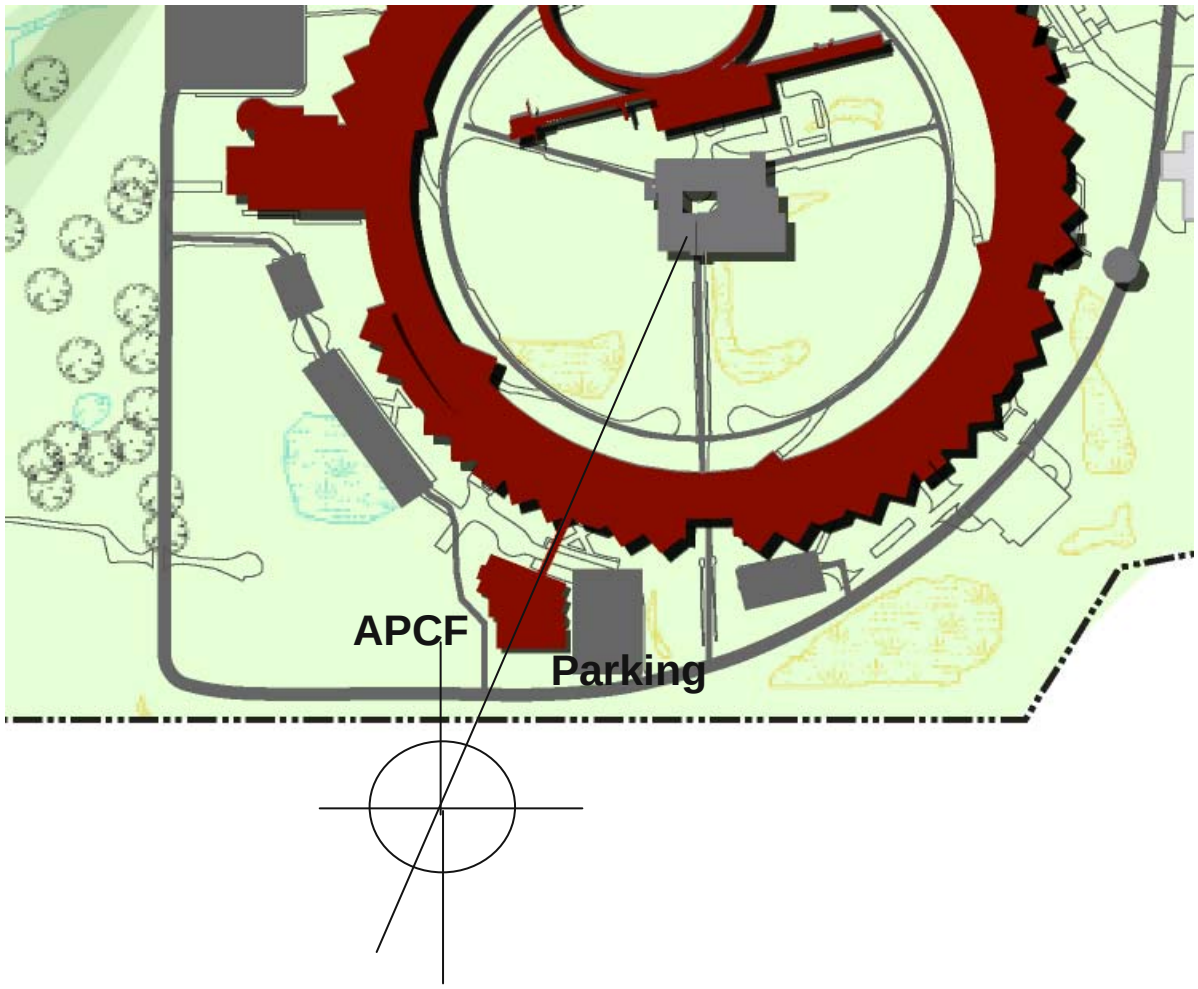


beamline robotics

APCF Funding

- Design and construction
  - State
- Operation funding will come from various sources
    - *Federal (DOE, NIH)*
    - *Biotechnology Interests*
- ANL advantage – science/prior experience/track record/partnerships/national user facilities/research networks
- Huge advantage to be next to APS at ANL

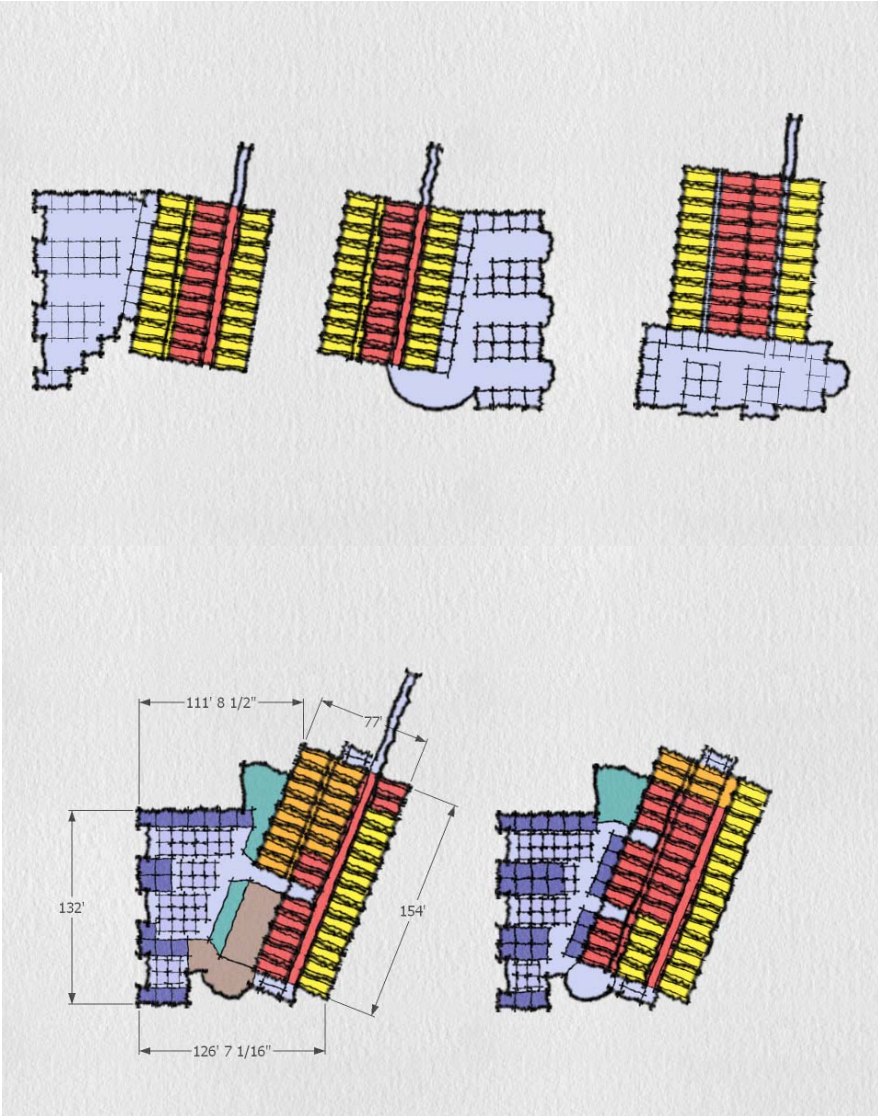
APCF Proposed Location



APCF Project Elements

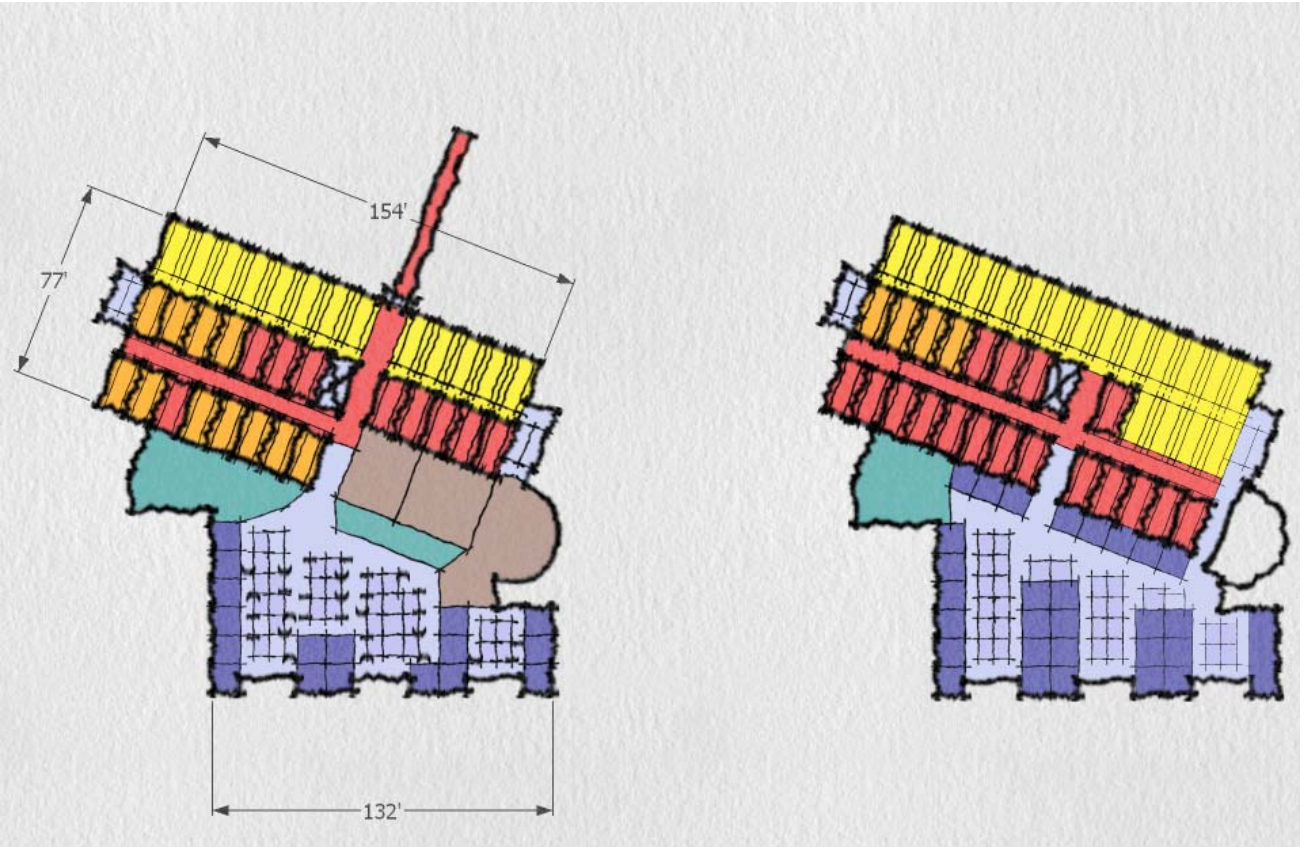
- **Human Elements**
  - **Flexibility for reorganization labs and offices**
  - **Interactive teaching environment (people and processes)**
  - **Interactive conferencing (small to medium)**
  - **Auditorium to seat 20 - 120 – lectures (flexible)**
  - **Exhibit areas**
  - **Production showcase spaces**
  - **Break rooms**
  - **Multiple interactive spaces for brainstorming**
- **Environmental Elements**
  - **Vibration isolation (Crystallization, EM)**
  - **Temperature and humidity control (Crystallization)**
  - **No significant electrical/magnetic fields**
  - **Systems integration**
  - **Non-sensitive information**
  - **Non-infectious (max. BSL2+ laboratory)**
  - **Open environment – open facility**
  - **As much visibility as possible and light**

Floor Plate Studies and Key Considerations



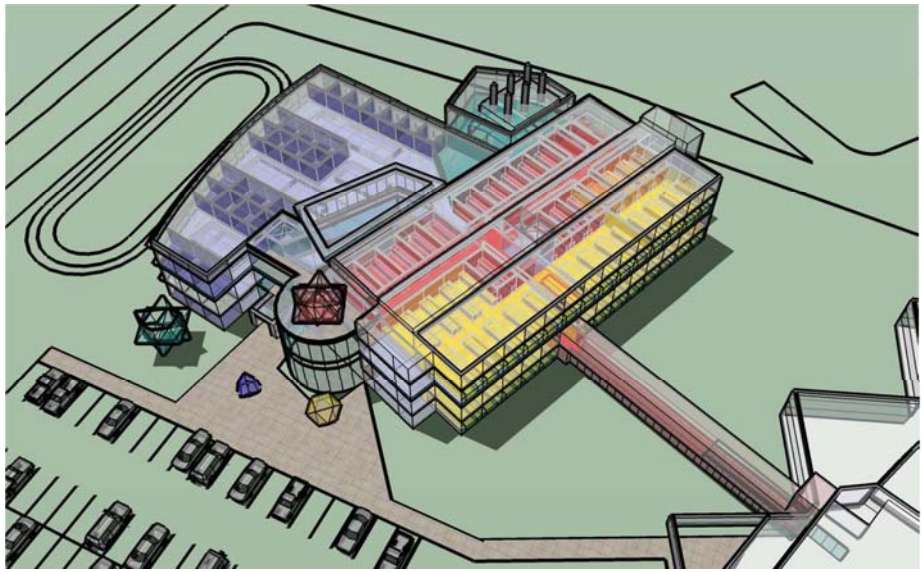
- Modular Framework 20’x 11’ Basic Module
- Open / Closed Labs
- Open / Closed Offices
- Relationship Between Labs and Offices
- Front Door / Back Door
- Relationship to LOM Entrance and Parking
- Service Point
- Fan Room Penthouse?

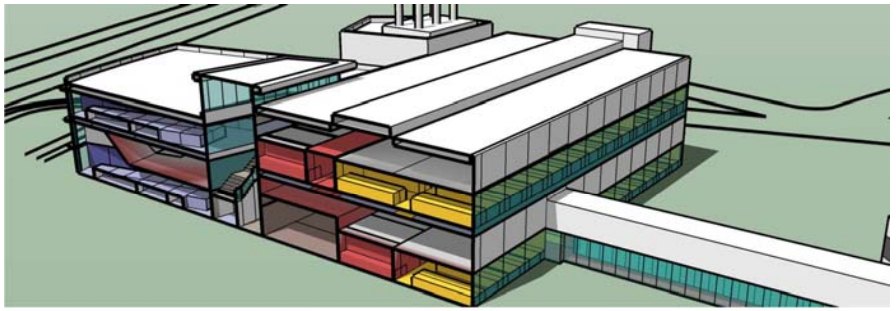
Conceptual Floor Diagrams

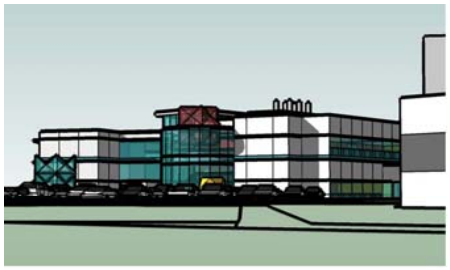


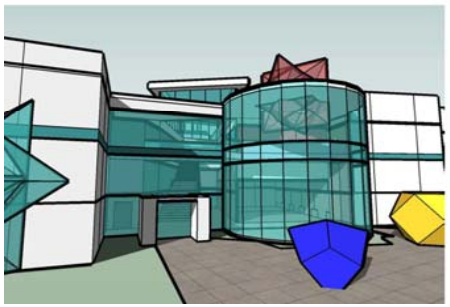
- Modular Framework 20’x 11’ Basic Module
- Open / Closed Labs
- Open / Closed Offices
- Relationship Between Labs and Offices
- Front Door / Back Door
- Relationship to LOM Entrance and Parking
- Service Point
- Fan Room Penthouse?

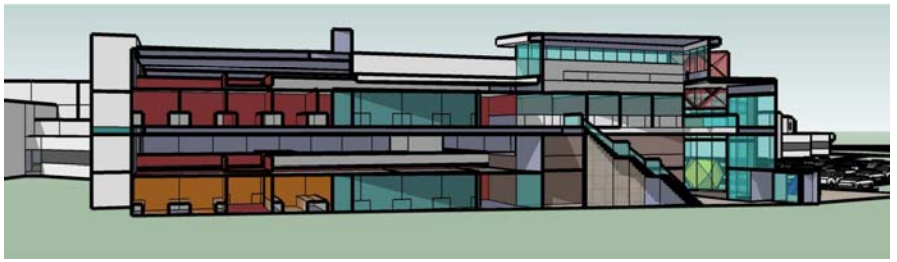
APCF Preliminary Design











Acknowledgments

Argonne BIO

C. Giometti
L. Makowski

MCSG, SBC and BIO staff memebers

Argonne

K. Hellman

Jacobs Consultancy
Advanced Planning Group
GPR Lab Planners

P. Hogan
S. Clepper
S. Weber

Argonne/UoC IGSB

K. White
I. Hurley

Funding: State of Illinois