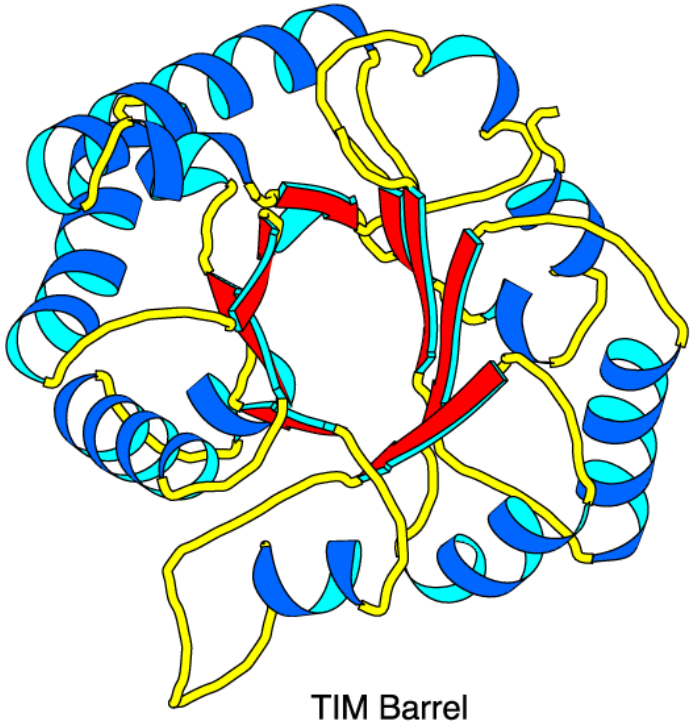
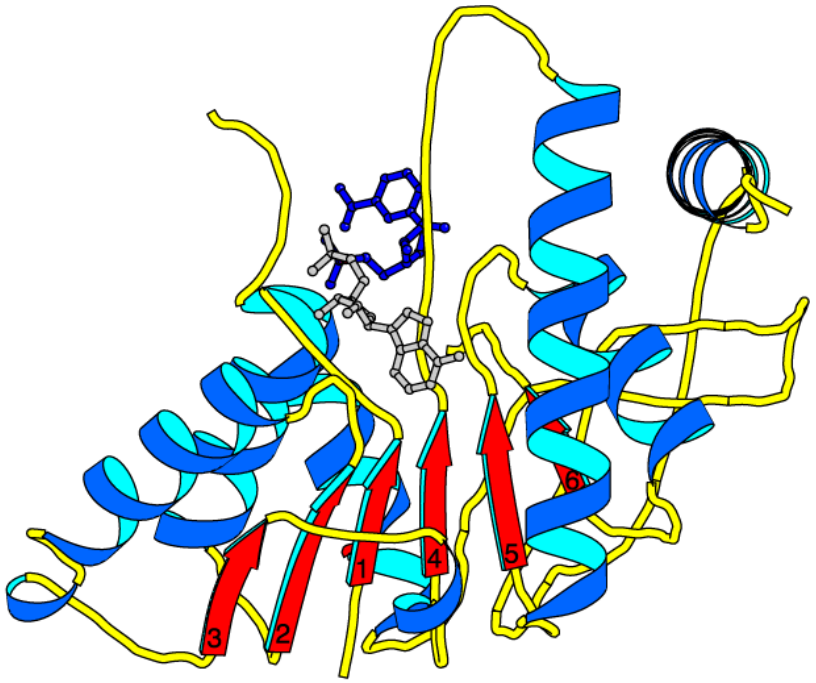
Protein Structure Determination



TIM Barrel
(7TIM)



Lactate dehydrogenase
NAD binding domain
(1A5Z)

How are these structures determined?

Why Bother With Structure?

- The amino acid sequence of a protein contains interesting information.
- A protein sequence can be compared to other protein sequences to establish its evolutionary relationship to other proteins and protein families.
- However, for the purposes of understanding protein function, the 3D structure of the protein is far more useful than the sequence.

Protein Sequences Far Outnumber Structures

- Only a small number of protein structures have been experimentally determined.

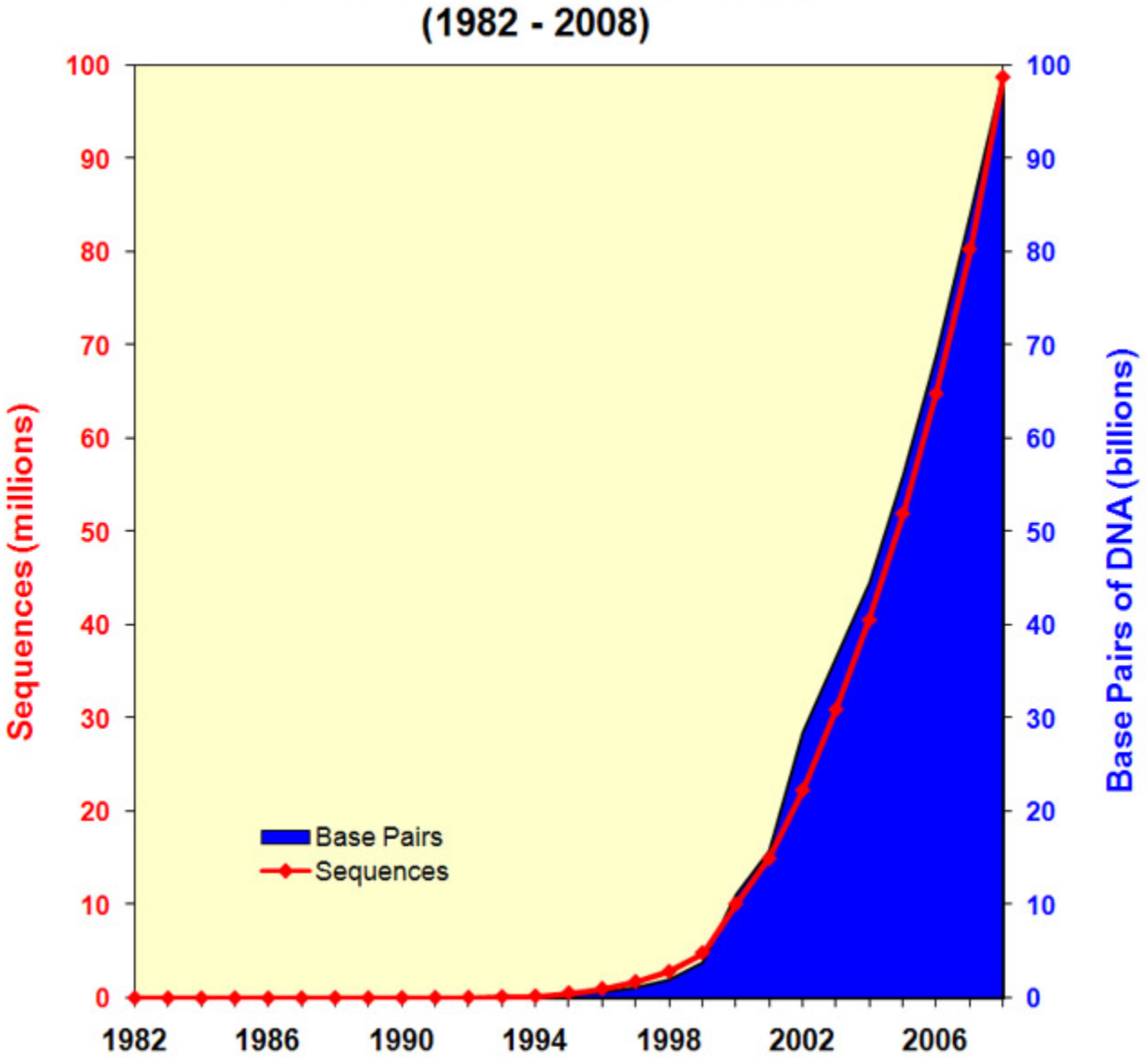
PDB	**~64,623 protein structures**
Genebank	**~61,132,599 sequences**

- Of the 64,623 structures, only **15,702** are dissimilar in sequence (<30% ID).

Growth of GenBank

Now over 100M sequences and 100B base pairs from 223K species

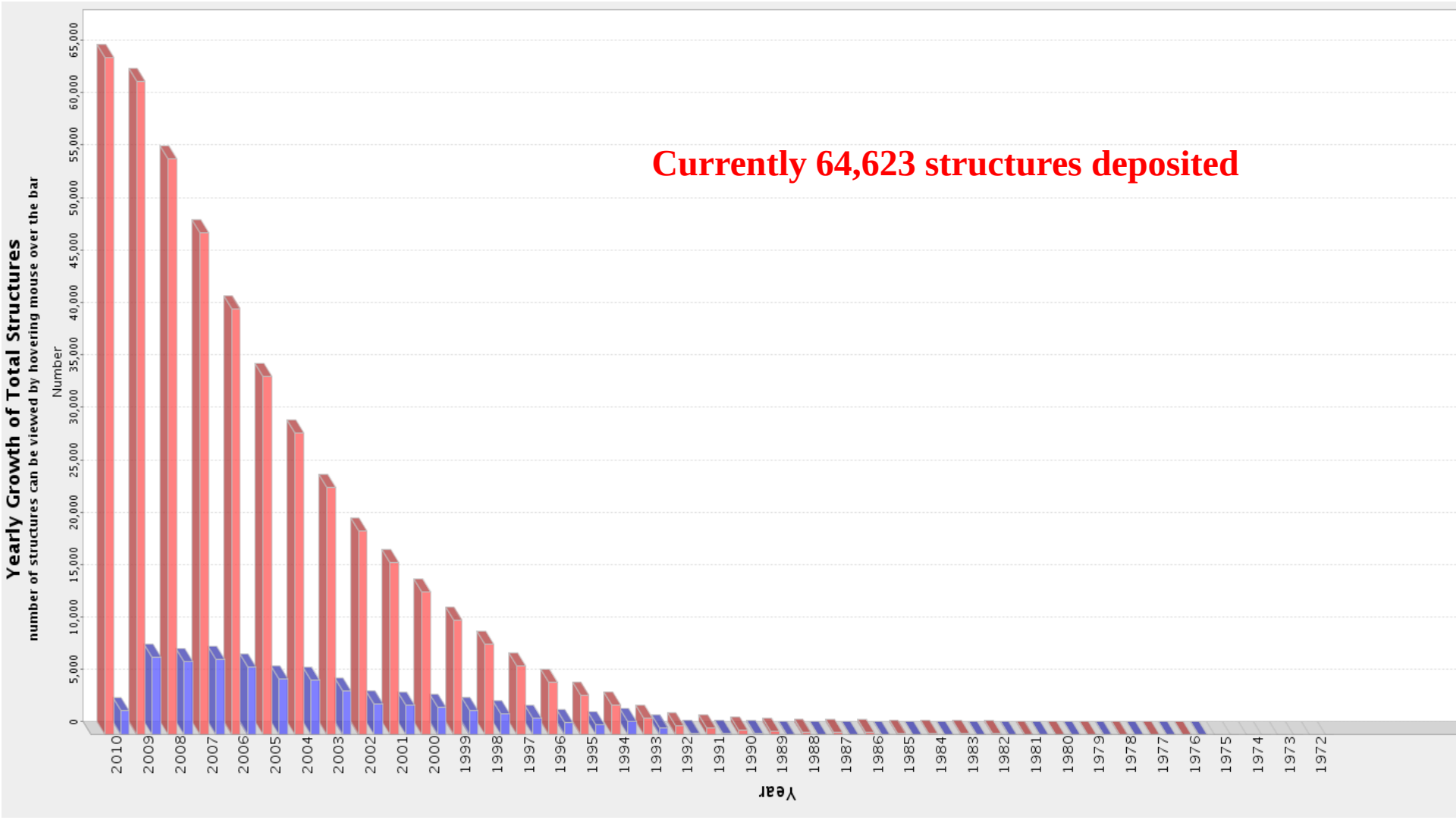
Exponential Increase



http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html

http://www.ncbi.nlm.nih.gov/Sitemap/Summary/statistics.html

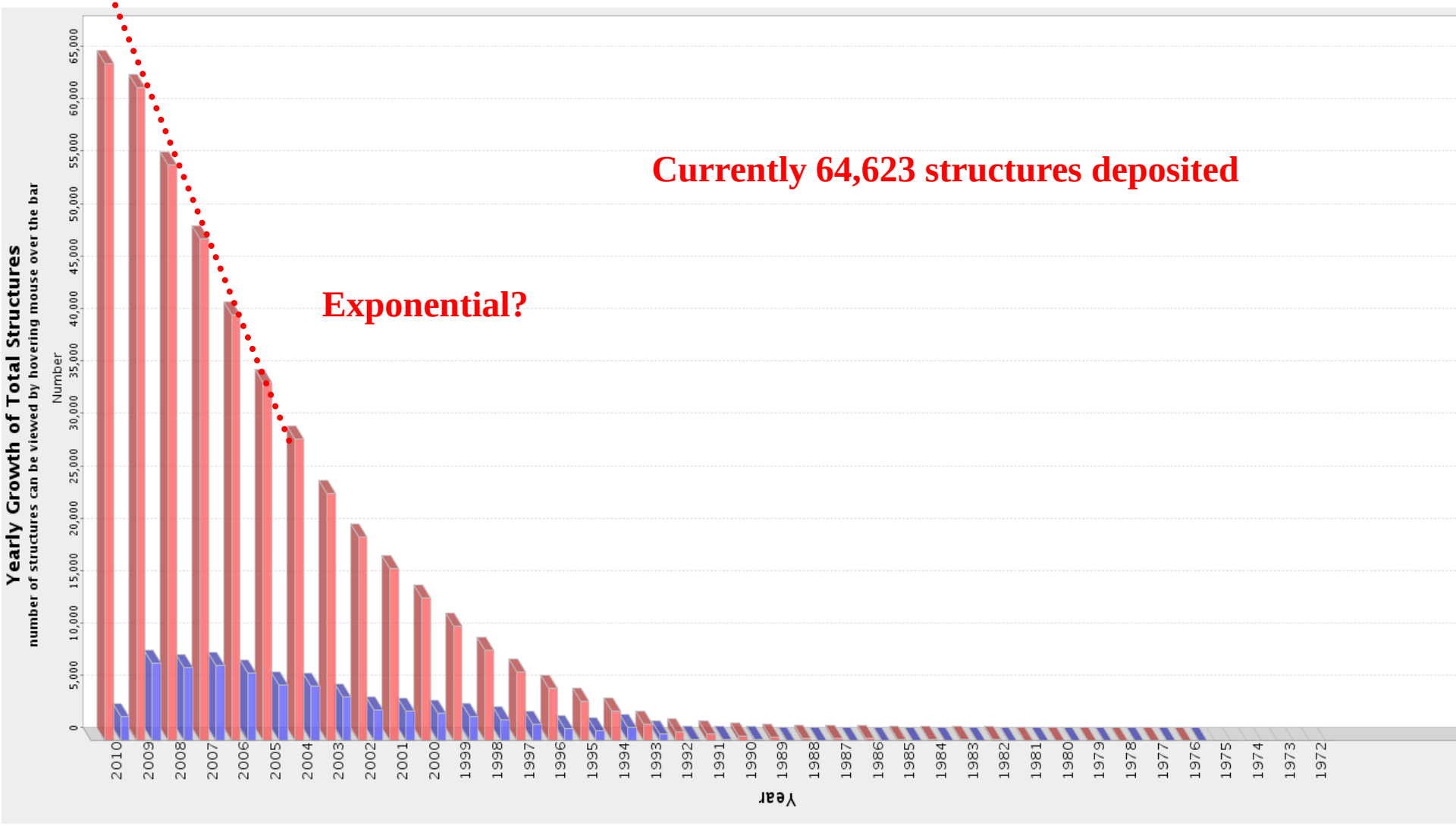
Growth of Structural Data



Currently 64,623 structures deposited

http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100

Growth of Structural Data



http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100

Structural Proteomics

- Use experimentally determined structures to model the structures of similar proteins
  - Threading
  - Homology Modeling
  - Fold recognition

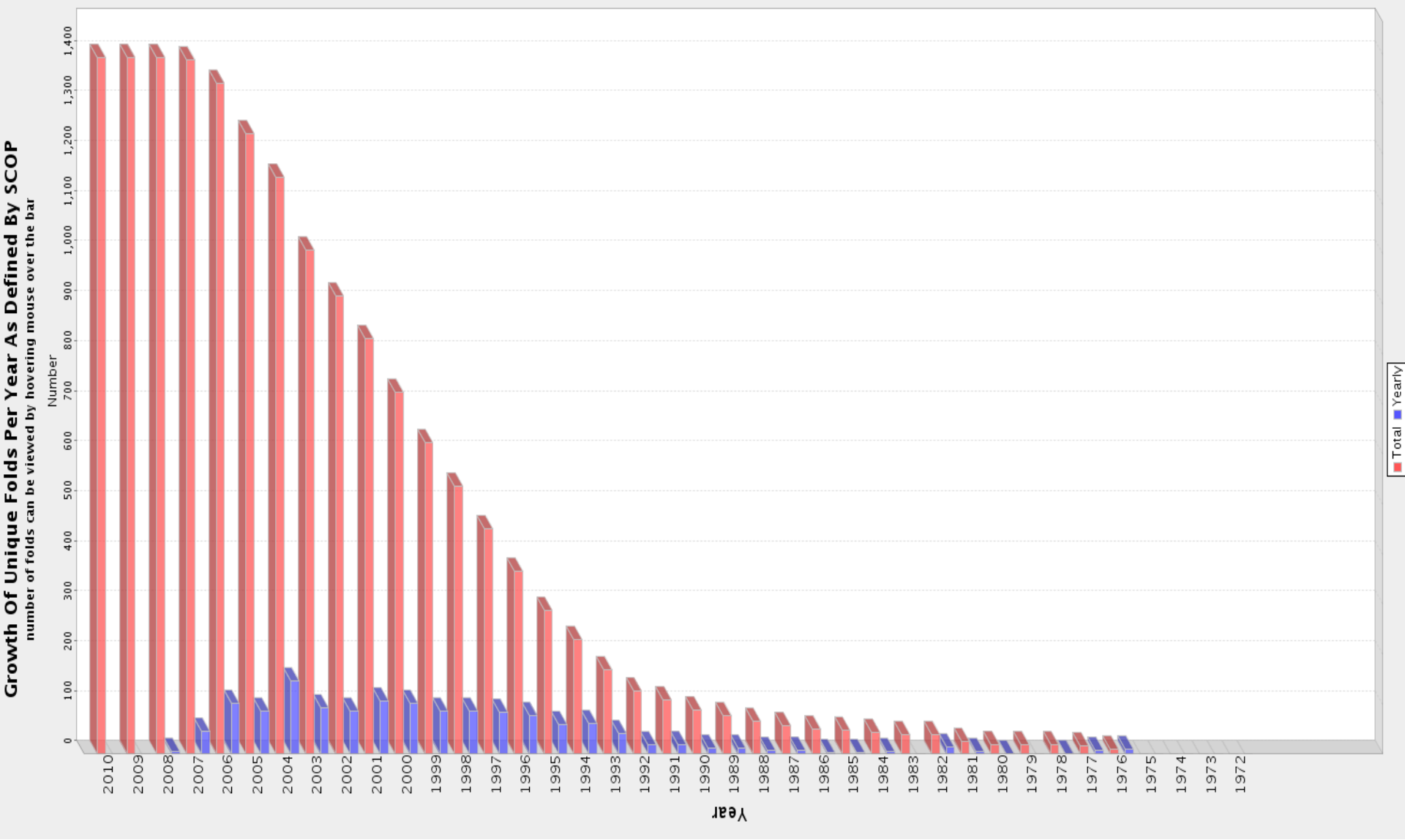
  } **Avoids *Ab initio* structure determination**

- Need representative protein structures for the total repertoire of protein folds
- Provide 3D portraits for all proteins in an organism
- Goal: Use structure to infer function.
  - More sensitive than primary sequence comparisons

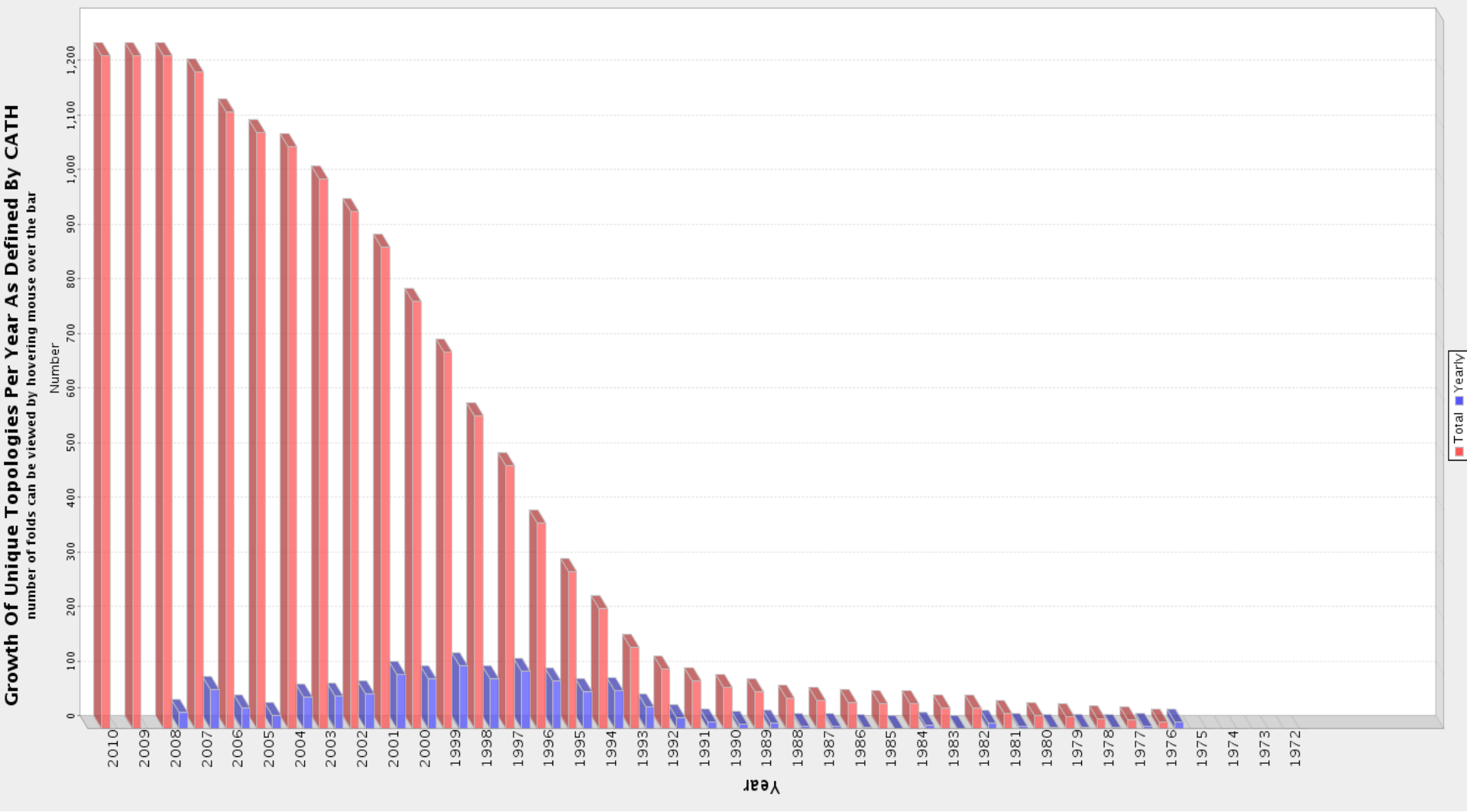
Redundancy in PDB (20 April 10)

Sequence identity	Number of non-redundant chains
90%	25615
70%	23116
50%	20306
30%	15702

Unique folds in PDB (SCOP)



Unique topologies in PDB (CATH)



New Topologies and Folds Becoming Rare

Structural Genomics

Initiated in 1999 by NIH
Phase I included 9 large centers for high throughput structure determination
Phase I ran from ~2000 - 2005

Goal

The long-range goal of the Protein Structure Initiative (PSI) is to make the three-dimensional atomic-level structures of most proteins easily obtainable from knowledge of their corresponding DNA sequences.

http://www.nigms.nih.gov/psi/mission.html

Structural Genomics

Benefits

Structural descriptions will help researchers illuminate structure-function relationships and thus formulate better hypotheses and design better experiments.

The PSI collection of structures will serve as the starting point for structure-based drug development by permitting faster identification of lead compounds and their optimization.

The design of better therapeutics will result from comparisons of the structures of proteins that are from pathogenic and host organisms and from normal and diseased human tissues.

The PSI collection of structures will assist biomedical investigators in research studies of key biophysical and biochemical problems, such as protein folding, evolution, structure prediction, and the organization of protein families and folds.

Technical developments, the availability of reagents and materials, and experimental outcome data in protein production and crystallization will directly benefit all structural biologists and provide valuable assistance to a broad range of biomedical researchers.

Structural Genomics Centers in US

PSI-1 Winners

- The Joint Center for Structural Genomics (JCSG)

  During PSI-2, the JCSG has contributed to the overall goal of maximizing structural coverage of protein families with no structural representation and has continued to developm and disseminate innovative new technologies for structural biology. The JCSG consortium theme is the “central machinery of life” — proteins that are conserved in all kingdoms of life.

- The Midwest Center for Structural Genomics (MCSG)

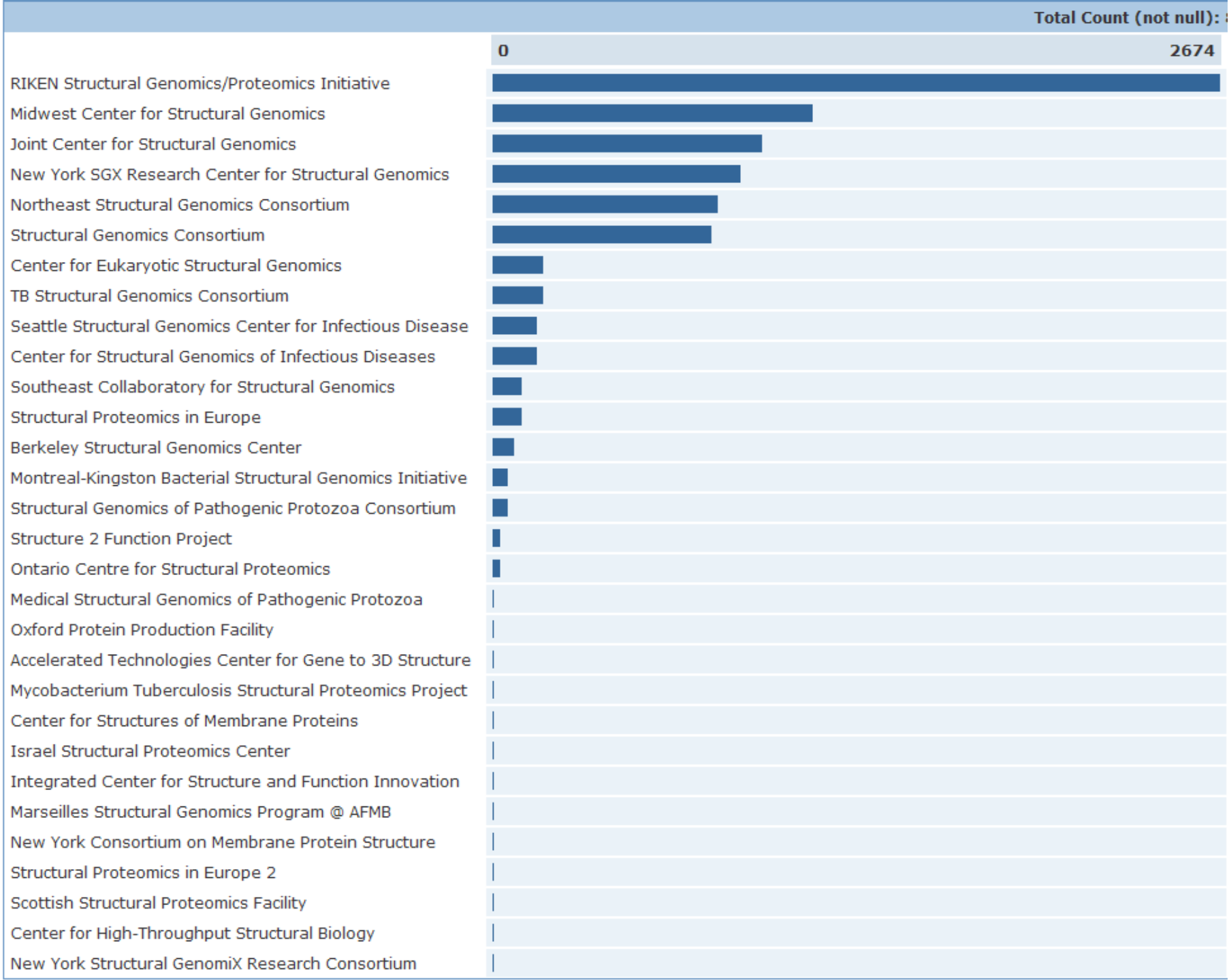
  In PSI-2, the multi-institutional consortium is rapidly determining the structures of large numbers of strategically selected proteins using x-ray crystallography both to provide structural coverage of major protein superfamilies and to elucidate the entire protein folding space.

- The New York Structural Genomics Research Consortium (NYSGRC)

  During PSI-2, the consortium’s individual project focuses on new targets, principally protein phosphatases and mul tidomain eukaryotic proteins.

- The Northeast Structural Genomics Consortium (NEGS)

  In PSI-2, the consortium is solving both prokaryotic and eukaryotic structural representatives from the major domain families constituting the eukaryotic proteome.


Total Count (not null):
0
2674
RIKEN Structural Genomics/Proteomics Initiative
Midwest Center for Structural Genomics
Joint Center for Structural Genomics
New York SGX Research Center for Structural Genomics
Northeast Structural Genomics Consortium
Structural Genomics Consortium
Center for Eukaryotic Structural Genomics
TB Structural Genomics Consortium
Seattle Structural Genomics Center for Infectious Disease
Center for Structural Genomics of Infectious Diseases
Southeast Collaboratory for Structural Genomics
Structural Proteomics in Europe
Berkeley Structural Genomics Center
Montreal-Kingston Bacterial Structural Genomics Initiative
Structural Genomics of Pathogenic Protozoa Consortium
Structure 2 Function Project
Ontario Centre for Structural Proteomics
Medical Structural Genomics of Pathogenic Protozoa
Oxford Protein Production Facility
Accelerated Technologies Center for Gene to 3D Structure
Mycobacterium Tuberculosis Structural Proteomics Project
Center for Structures of Membrane Proteins
Israel Structural Proteomics Center
Integrated Center for Structure and Function Innovation
Marseilles Structural Genomics Program @ AFMB
New York Consortium on Membrane Protein Structure
Structural Proteomics in Europe 2
Scottish Structural Proteomics Facility
Center for High-Throughput Structural Biology
New York Structural GenomiX Research Consortium

Structural Genomics Centers in US

PSI-1 Losers

Center for Eukaryotic Structural Genomics (CESG)

The CESG was founded as a collaborative effort to develop the technologies needed for economical high-throughput structure determination of biologically important eukaryotic proteins and to extend the knowledge of fold-function space. This project also aims to further the research of biologically important proteins in Arabidopsis. The protein structures are being determined via X-ray crystallography or NMR spectroscopy.

The Berkeley Structural Genomics Center (BSGC) The BSGC is pursuing an integrated structural genomics program designed to obtain a near-complete structural complement of two minimal genomes, Mycoplasma genitalium and Mycoplasma pneumoniae, two related human and animal pathogens. Both NMR spectroscopy and X-ray crystallography are being used for structural determination.

The Southeast Collaboratory for Structural Genomics (SECSG)

The objective of the SECSG is to develop and test experimental and computational strategies for high throughput structure determination of proteins by X-ray crystallography and NMR methods and to apply these strategies to scan the entire genome of an organism at a rapid pace. The eukaryotic organisms, *Caenorhabditis elegans*, *Homo sapiens* and an ancestrally-related prokaryotic microorganism having a small genome, *Pyrococcus furiosus*, have been selected as representative genomes.

Structural Genomics of Pathogenic Protozoa Consortium (SGPP)

The SGPP consortium aims to determine and analyze the structures of a large number of proteins from major global pathogenic protozoa including *Leishmania major*, *Trypanosoma brucei*, *Trypanosoma cruzi* and *Plasmodium falciparum*. These organisms are responsible for the diseases: leishmaniasis, sleeping sickness, Chagas' disease and malaria. X-ray crystallography is being used for structural determination.

The TB Structural Genomics Consortium (TB)

The goal of the TB consortium is to determine the structures of over 400 proteins from M. tuberculosis, and to analyze these structures in the context of functional information that currently exists and that is generated by the project. These structures will include about 40 novel folds and 200 new families of protein structures. The protein structures are being determined using X-ray crystallography.

Current PSI Centers

Large-Scale Centers

- **Joint Center for Structural Genomics**
- **Midwest Center for Structural Genomics**
- **New York SGX Research Center for Structural Genomics**
- **Northeast Structural Genomics Consortium**

Specialized Centers

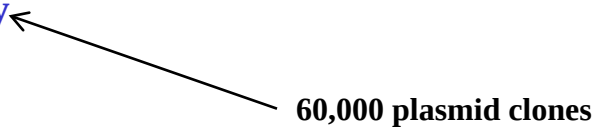
- **Accelerated Technologies Center for Gene to 3D Structure**
- **Center for Eukaryotic Structural Genomics**
- **Center for High-Throughput Structural Biology**
- **Center for Structures of Membrane Proteins**
- **Integrated Center for Structure and Function Innovation**
- **New York Consortium on Membrane Protein Structure**

Homology Modeling Centers

- **Joint Center for Molecular Modeling**
- **New Methods for High-Resolution Comparative Modeling**

Resource Centers

- **PSI-Materials Repository**
- **PSI Knowledgebase**



2008 Structural Genomics Progress

Status	Total Number of Targets	(%) Relative to "Cloned" Targets	(%) Relative to "Expressed" Targets	(%) Relative to "Purified" Targets	(%) Relative to "Crystallized" Targets
Cloned	61522	100.00	-	-	-
Expressed	39540	64.27	100.00	-	-
Soluble	18221	29.62	46.08	-	-
Purified	14031	22.81	35.49	100.00	-
Crystallized	5616	9.13	14.20	40.03	100.00
Diffraction-quality Crystals	2909	4.73	7.36	20.73	51.80
Diffraction	2429	3.95	6.14	17.31	43.25
NMR Assigned	1051	1.71	2.66	7.49	-
HSQC	1890	3.07	4.78	13.47	-
Crystal Structure	2291	3.72	5.79	16.33	40.79
NMR Structure	953	1.55	2.41	6.79	-
In PDB	2849	4.63	7.21	20.31	35.15
Work Stopped	14137	-	-	-	-
Test Target	4	-	-	-	-
Other	10	-	-	-	-

~40% of structures are from SG in Europe and Asia

http://targetdb.pdb.org/statistics/TargetStatistics.html

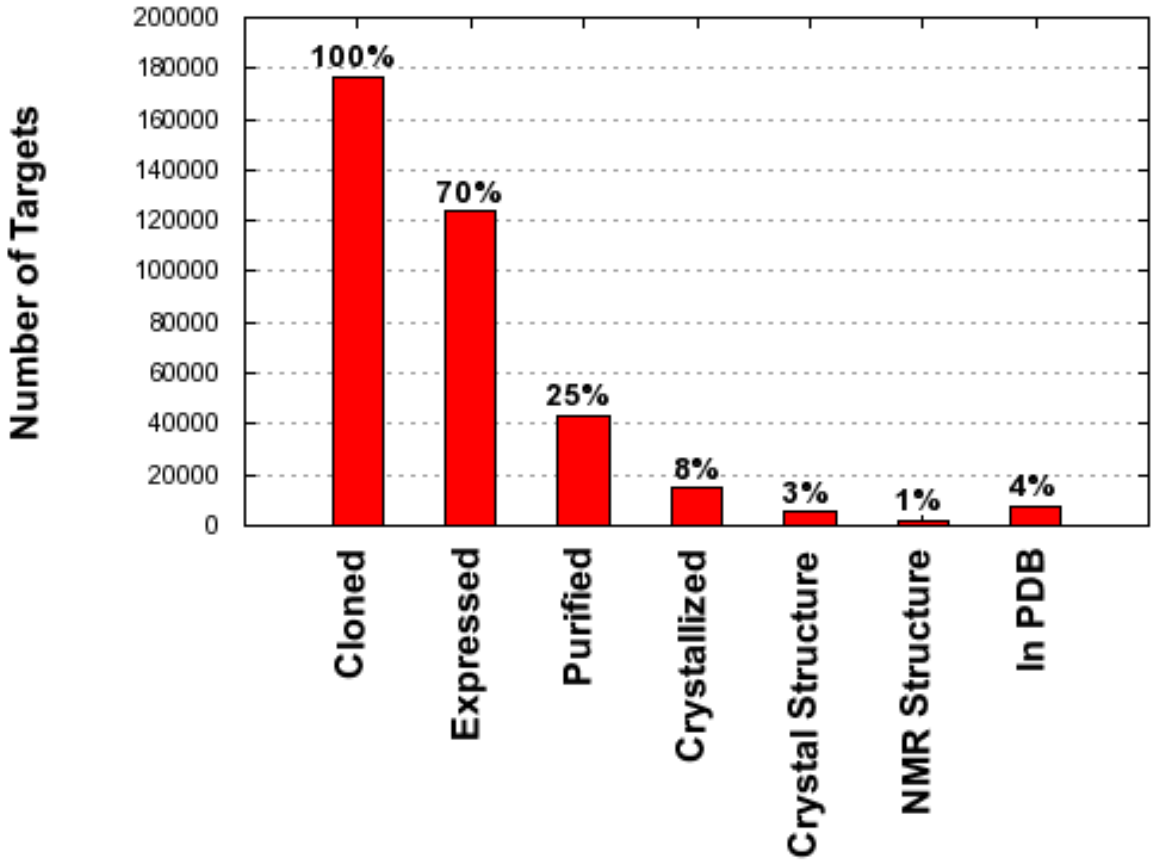
2010 Structural Genomics Progress

Status	Total Number of Targets	(%) Relative to "Cloned" Targets	(%) Relative to "Expressed" Targets	(%) Relative to "Purified" Targets	(%) Relative to "Crystallized" Targets
Cloned	176710	100.0	-	-	-
Expressed	123905	70.1	100.0	-	-
Soluble	47572	26.9	38.4	-	-
Purified	43609	24.7	35.2	100.0	-
Crystallized	14641	8.3	11.8	33.6	100.0
Diffraction-quality Crystals	7708	4.4	6.2	17.7	52.6
Diffraction	6628	3.8	5.3	15.2	45.3
NMR Assigned	2154	1.2	1.7	4.9	-
HSQC	3929	2.2	3.2	9.0	-
Crystal Structure	5215	3.0	4.2	12.0	35.6
NMR Structure	2050	1.2	1.7	4.7	-
In PDB	7569	4.3	6.1	17.4	38
Work Stopped	38962	-	-	-	-
Test Target	93	-	-	-	-
Other	8178	-	-	-	-

~36% of structures are from SG in Europe and Asia

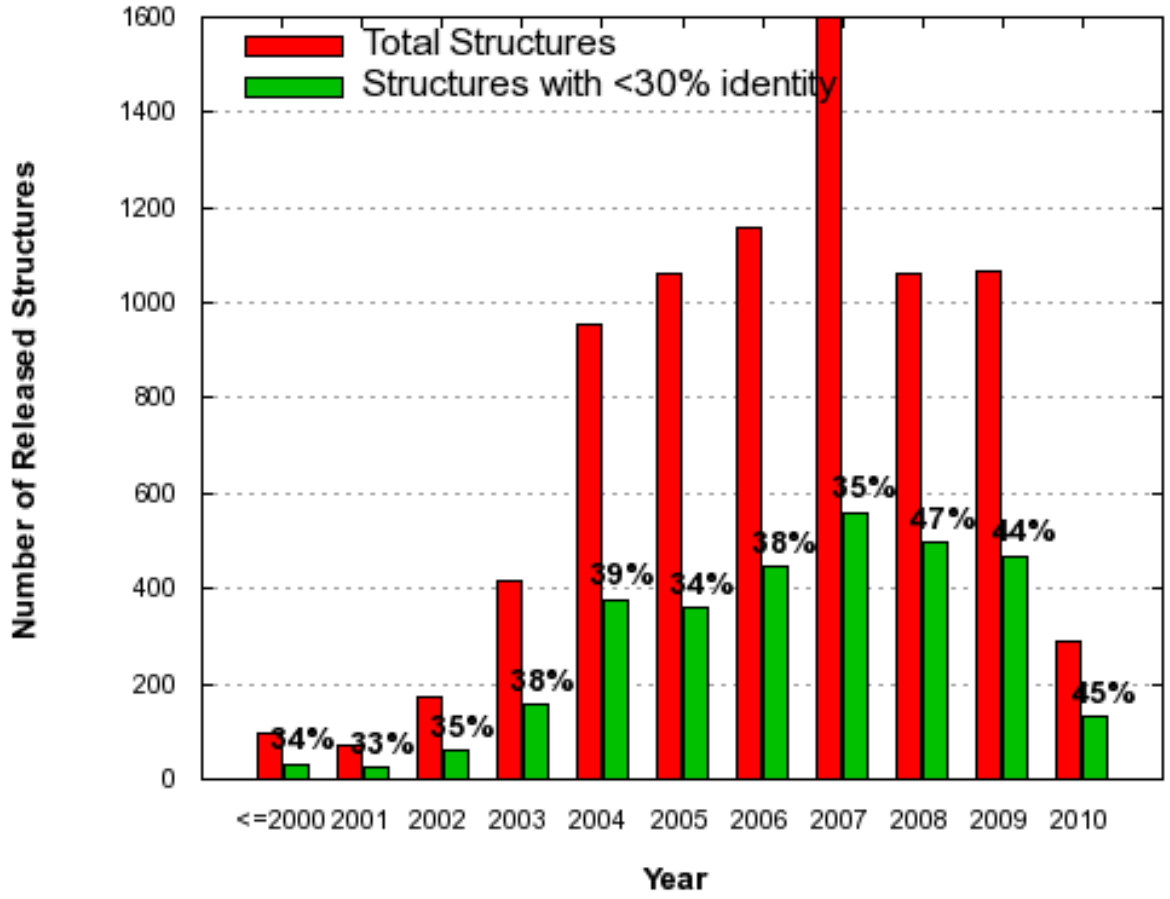
http://targetdb.pdb.org/statistics/TargetStatistics.html

Project Attrition



http://targetdb.pdb.org/statistics/TargetStatistics.html

Unique Folds?



http://targetdb.pdb.org/statistics/TargetStatistics.html

Protein Structure Databases

- **Where does protein structural information reside?**
  - **PDB**:
    - http://www.rcsb.org/pdb/
  - **MMDB**:
    - http://www.ncbi.nlm.nih.gov/Structure/
  - **FSSP**:
    - http://www.ebi.ac.uk/dali/fssp/
  - **SCOP**:
    - http://scop.mrc-lmb.cam.ac.uk/scop/
  - **CATH**:
    - http://www.biochem.ucl.ac.uk/bsm/cath_new/

Jon

Ingo

RCSB Protein Data Bank - Microsoft Internet Explorer provided by JHSPH Information Systems

http://www.rcsb.org/pdb/home/home.do

Google

Convert | Select

RCSB Protein Data Bank

Page | Tools

MyPDB Login

A MEMBER OF THE PDB

RCSB PDB PROTEIN DATA BANK

An Information Portal to Biological Macromolecular Structures

As of **Tuesday Apr 13, 2010 at 5 PM PDT** there are 64623 Structures | **PDB Statistics**

HELP | PRINT

PDB ID or Text | Search | Advanced Search

Customize This Page

Home Hide

- News & Publications
- Usage/Reference Policies
- Deposition Policies
- Website FAQ
- Deposition FAQ
- Contact Us
- About Us
- Careers
- New Website Features

Deposition Hide

- All Deposit Services
- Electron Microscopy
- X-ray | NMR
- Validation Server
- BioSync Beamline
- Related Tools

Search Hide

- Advanced Search
- Latest Release
- Latest Publications
- Sequence Search
- Chemical Components
- Unreleased Entries
- Browse Database
- Histograms

Tools Hide

- File Downloads
- FTP Services

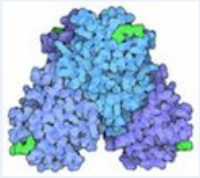
A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Hide Welcome Message

Featured Molecules (Previous Features: MOM | PSI) Hide

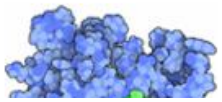


Molecule of the Month:

Concanavalin A and Circular Permutation

Evolution is a great tinkerer: once a successful plan is found, it is used again and again, often with many changes and improvements. This is easily seen in the living things around us. Most mammals have four limbs, which have evolved into all manner of arms and legs, and even into flippers and wings. Most plants are covered with leaves, which range from spiny needles to huge jungle fronds. Looking at protein sequences and structures, we see the same diversity generated through variation.

Full Article...



Protein Structure Initiative Featured Molecule:

Phytochrome

In two new NMR structures of an unusually small phytochrome, researchers

New Features Hide

New Image Collage Page

Read more about the releases:

Website Release Archive:

News Hide

Weekly | Quarterly | Yearly

Statement on Retraction of PDB Entries

2010-04-13

Exhibiting at Experimental Biology

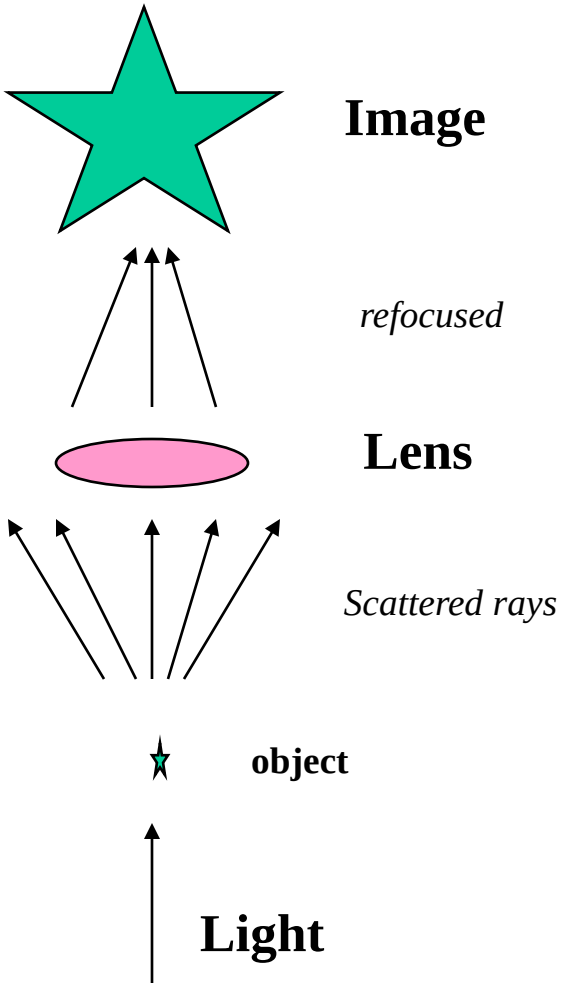
Read More

PDB Contents 20 April 2010

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	52212	1206	2401	17	55836
NMR	7279	896	154	7	8336
ELECTRON MICROSCOPY	195	17	76	0	288
HYBRID	16	1	1	1	19
Other	123	4	4	13	144
Total	59825	2124	2636	38	64623

X-ray Crystallography

Optical Microscope



Atomic Resolution

We want to resolve inter-atomic distances (~1.5 Å, 0.15 nM)

Visible light has a wavelength of ~ 500 nm (5000 Å)

Electron beam: λ_c ~ 0.001 Å (if e^- is moving at c)

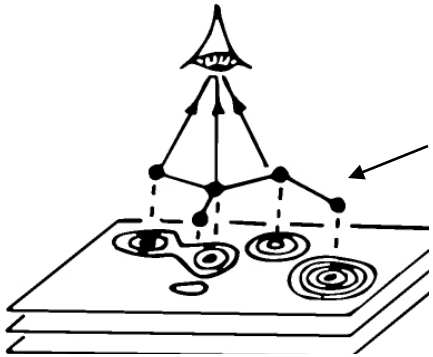
Electron velocity is less in electron microscopes

Typical resolution is ~10 Å, but can be improved

X-ray generators produce photons of λ = 0.5 – 2.5 Å

Use λ = 1.542 Å

X-ray Crystallography



Molecules

3D electron density map

Refocusing is accomplished with a computer, a crystallographer and a lot of mathematics

computer

crystallographer

mathematics

X-ray detector

Crystal size < 0.5 mm

Must use X-rays to get atomic resolution (1.5 Å = C-C bond)

X-rays

X-Ray Crystallography

1. Make crystals of your protein
   0.3-1.0mm in size
   Proteins must be in an ordered, repeating pattern.

2. X-ray beam is aimed at crystal and data is collected.

3. Structure is determined from the diffraction data.

X-Ray Crystallography

1. Make crystals of your protein
   0.3-1.0mm in size
   Proteins must be in an ordered, repeating pattern.

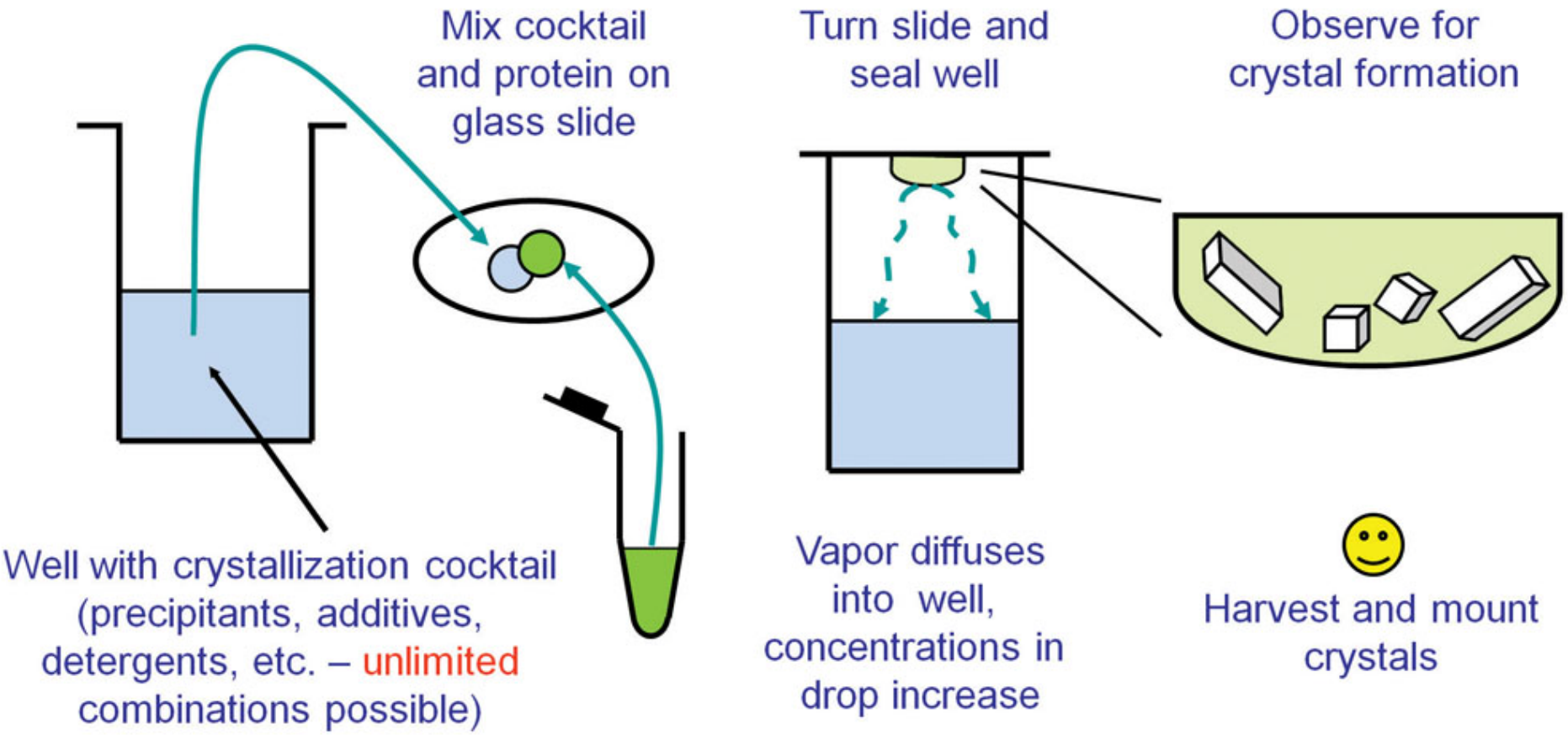
2. X-ray beam is aimed at crystal and data is collected.

3. Structure is determined from the diffraction data.

Protein Crystals

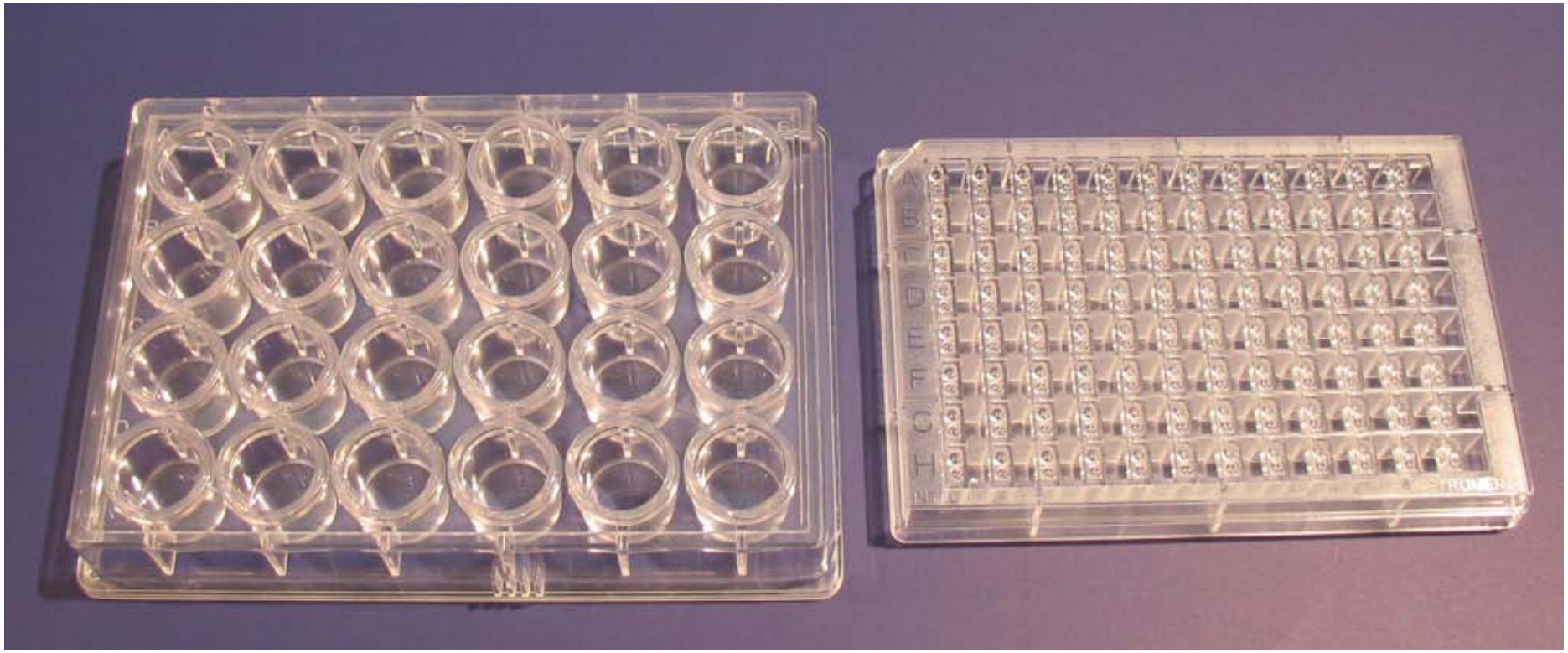
Schmid, M. Trends in Microbiolgy, **10**:s27-s31.

Protein Crystals

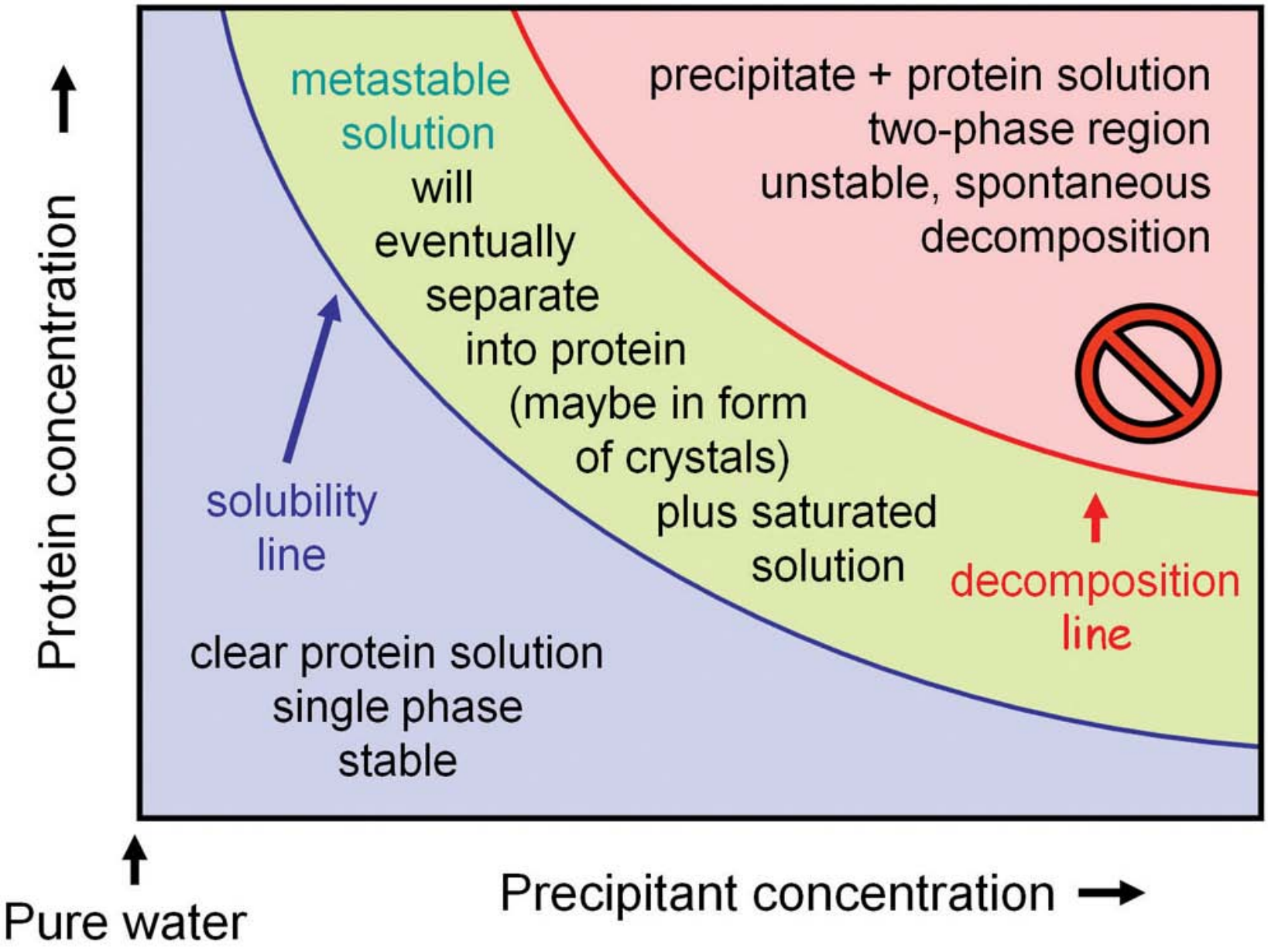


http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5

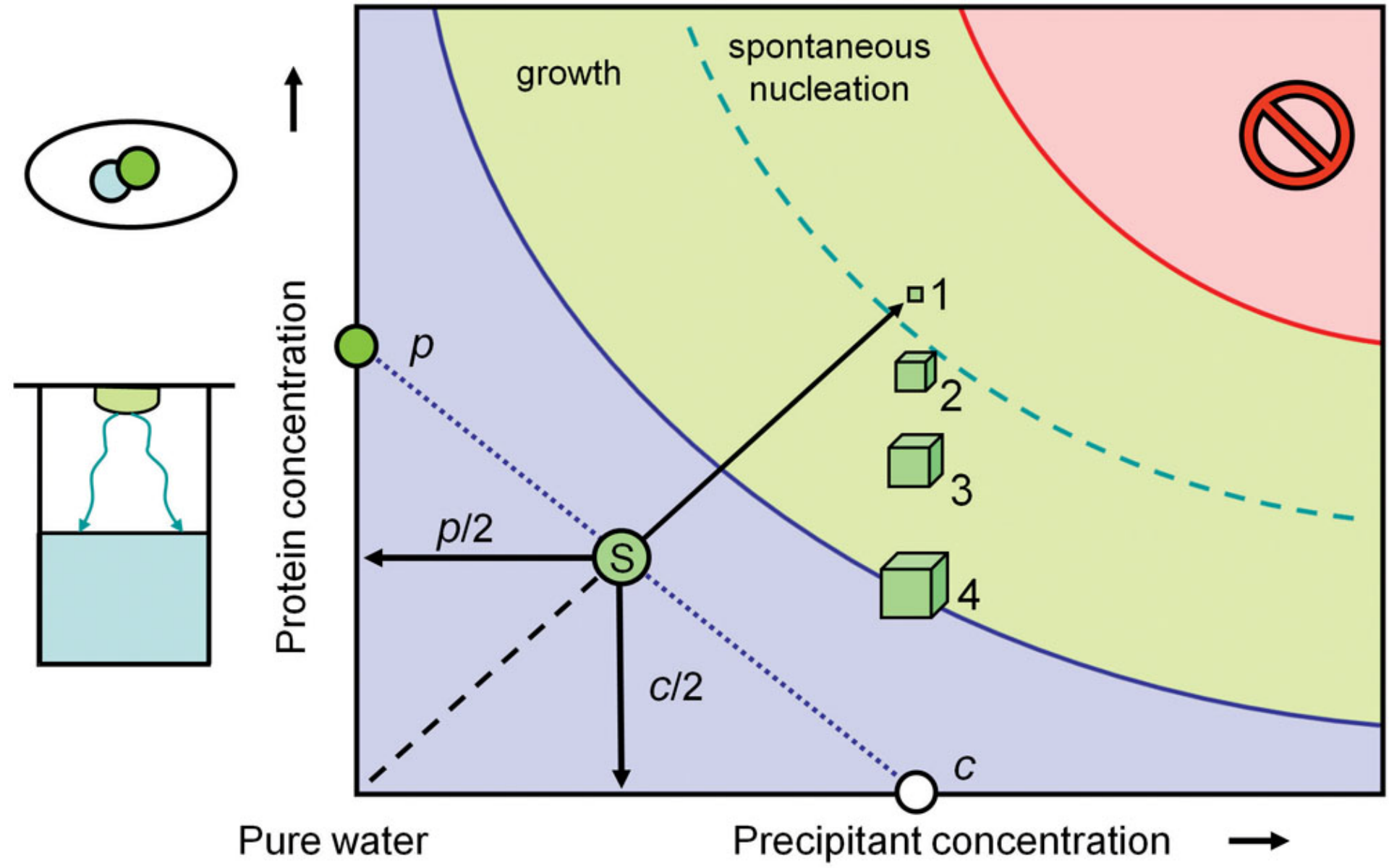
Protein Crystals



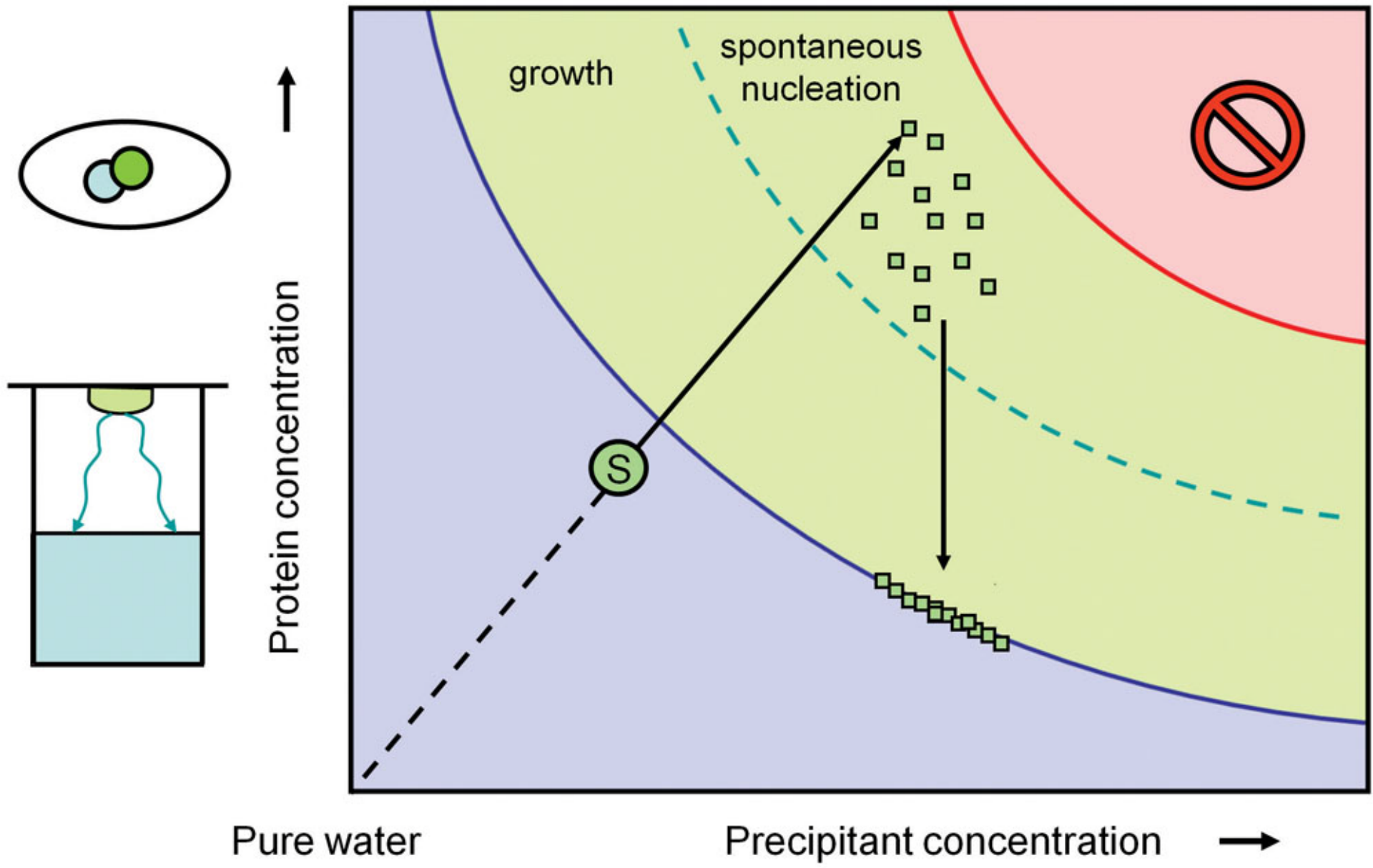
http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5



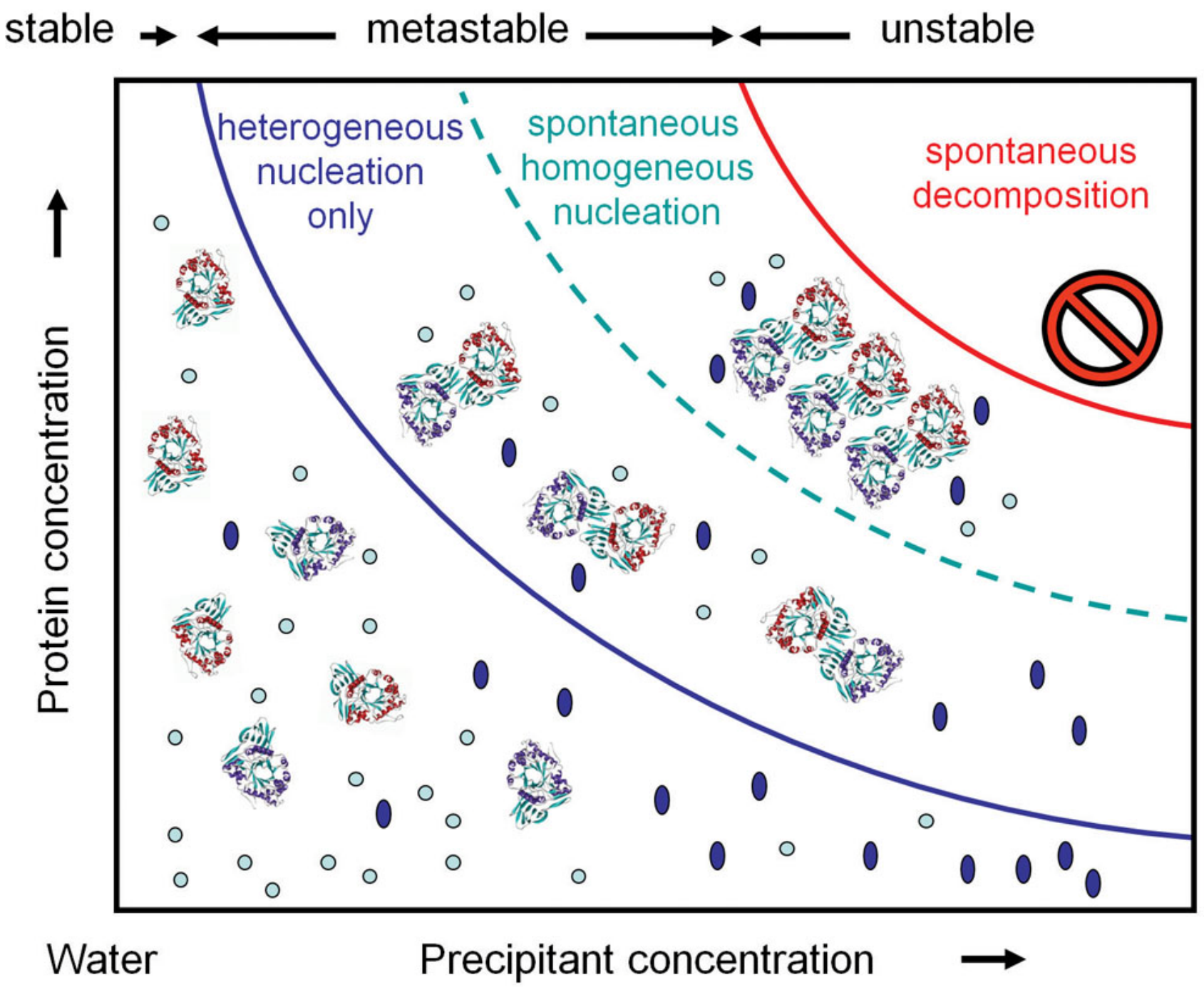
http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5



http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5

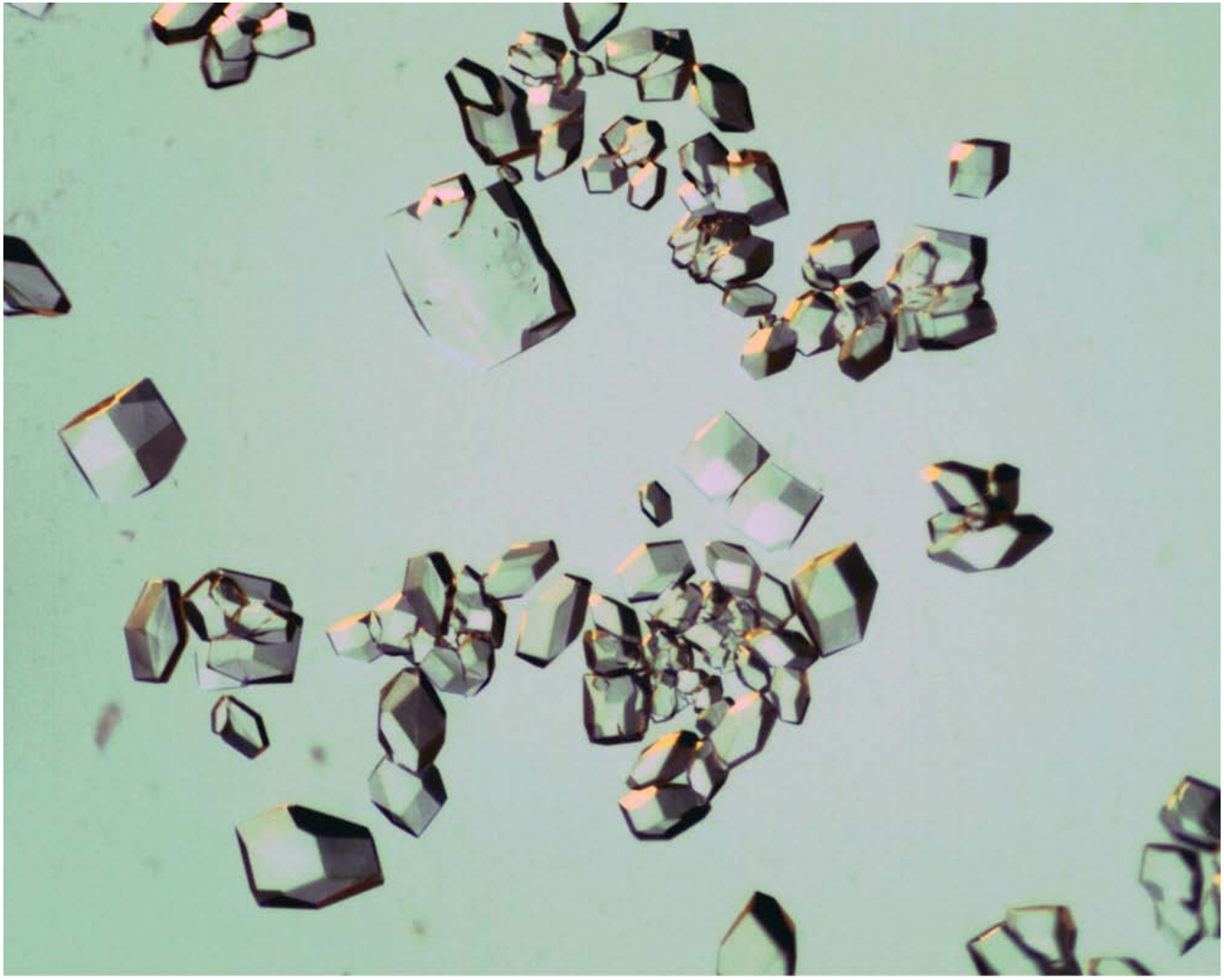


http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5



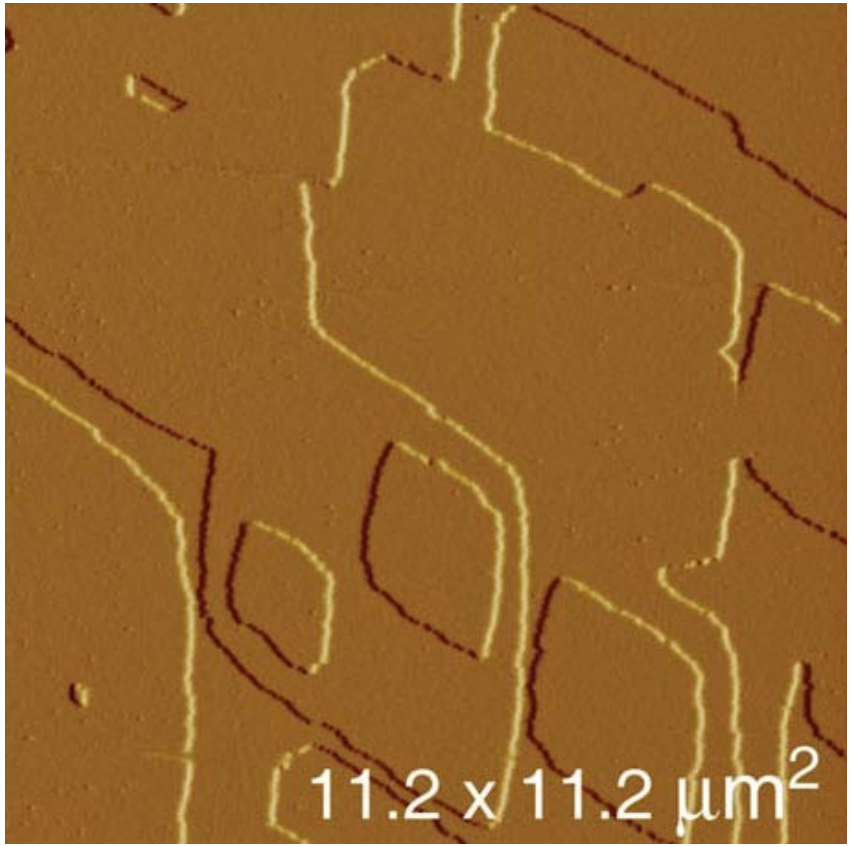
http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5

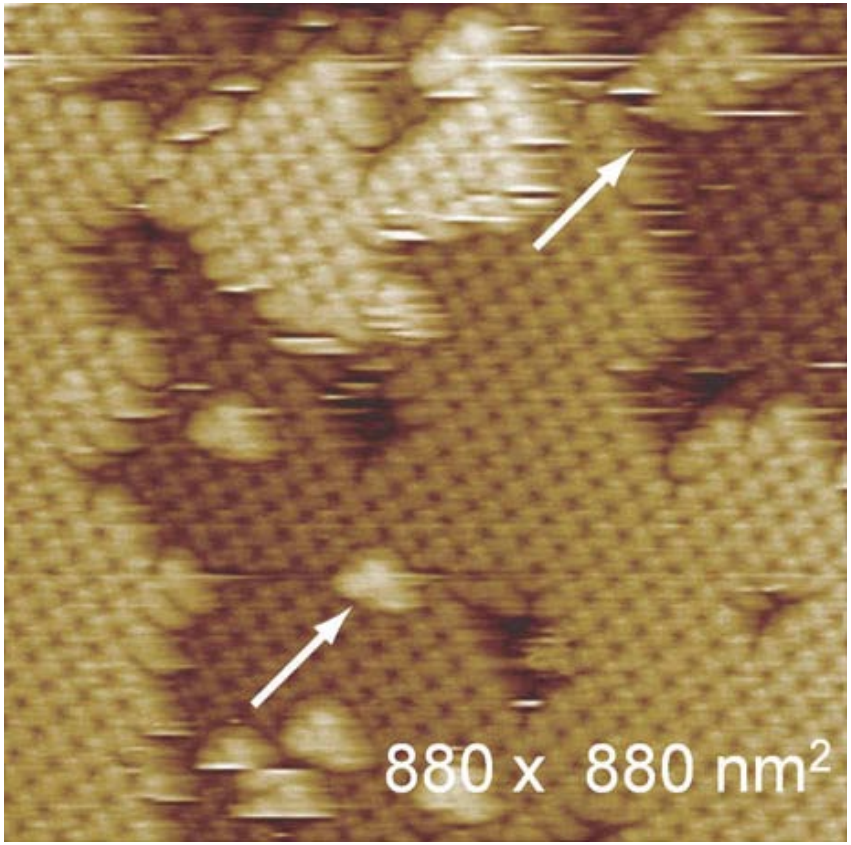
Protein Crystals



http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5

Protein Crystals





http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5

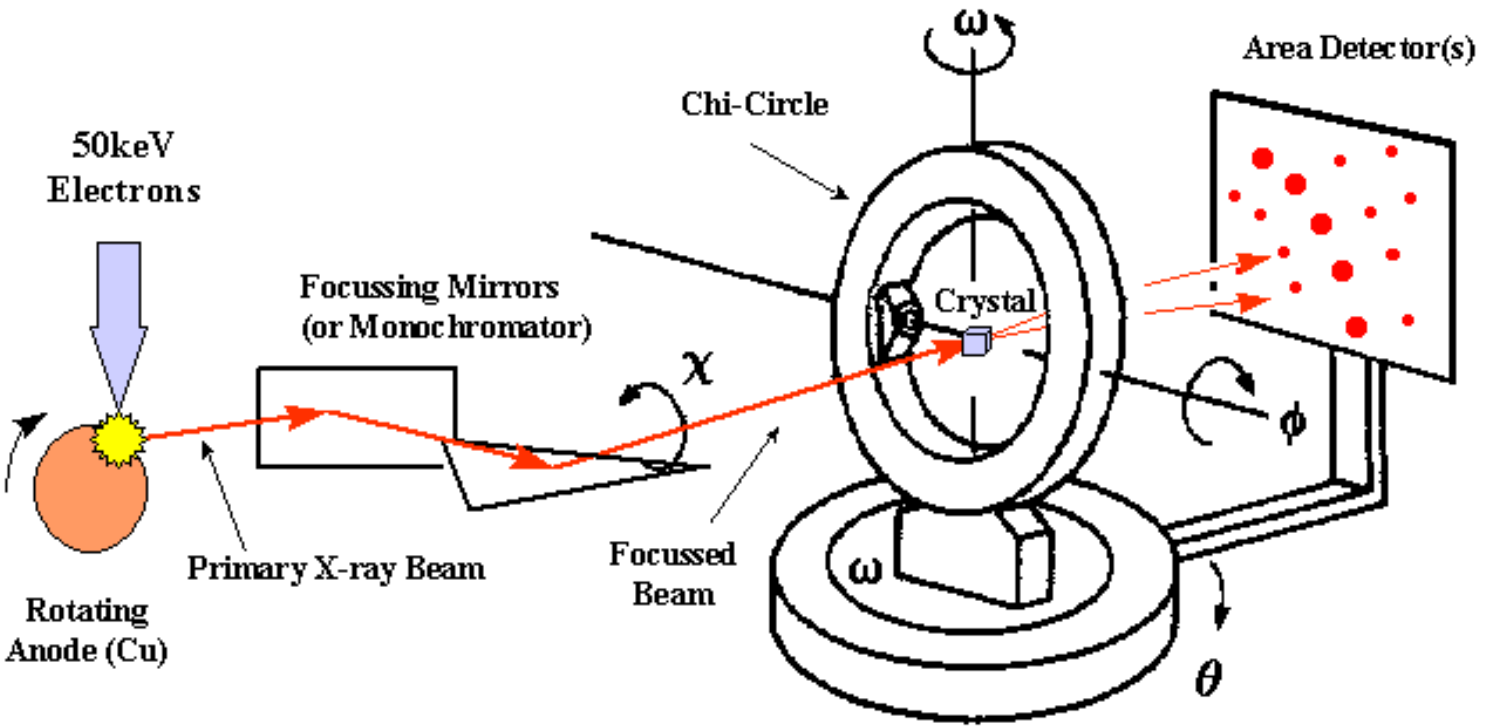
X-Ray Crystallography

1. Make crystals of your protein
   0.3-1.0mm in size
   Proteins must be in an ordered, repeating pattern.

2. X-ray beam is aimed at crystal and data is collected.

3. Structure is determined from the diffraction data.

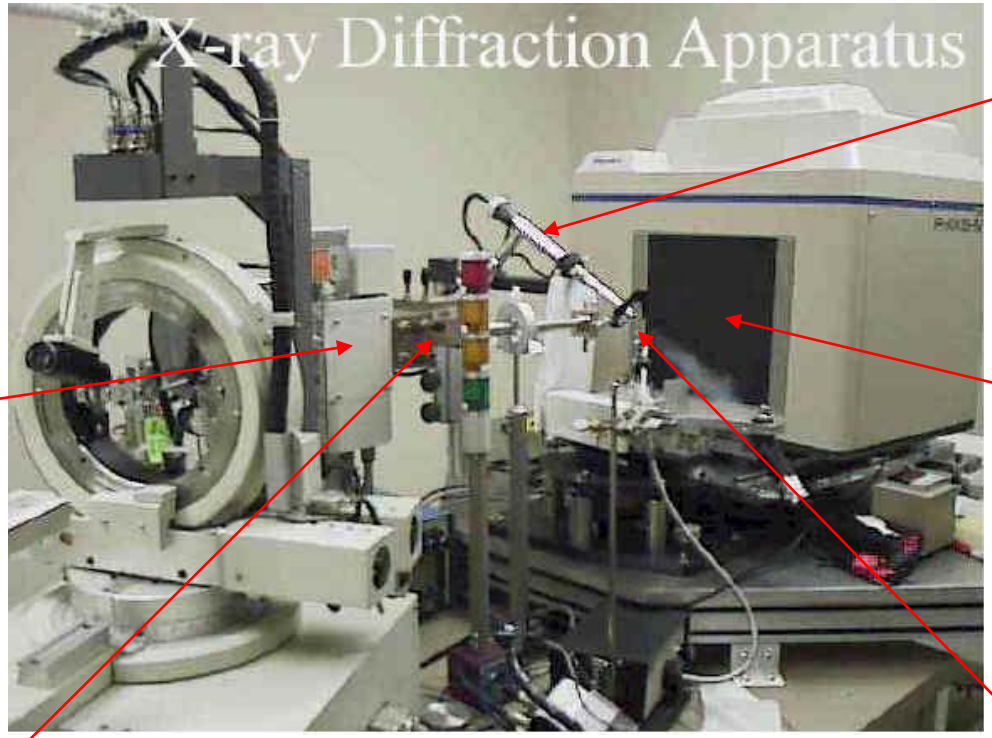
X-Ray Diffraction Experiment



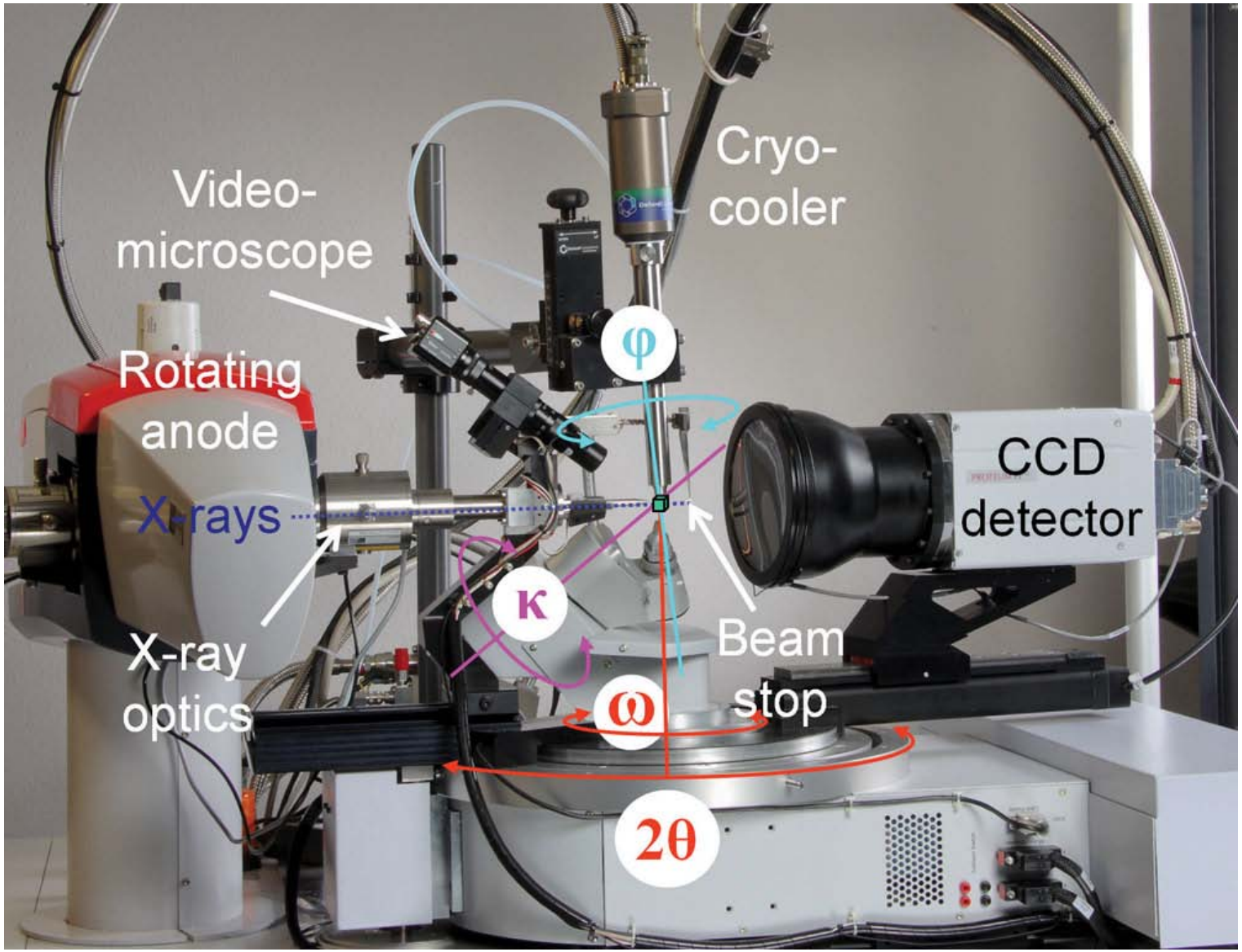
Generator **Optics** **Goniometer** **Detector**

Optional: Cryo for protein samples

X-ray Crystallography Equipment



X-ray Crystallography Equipment



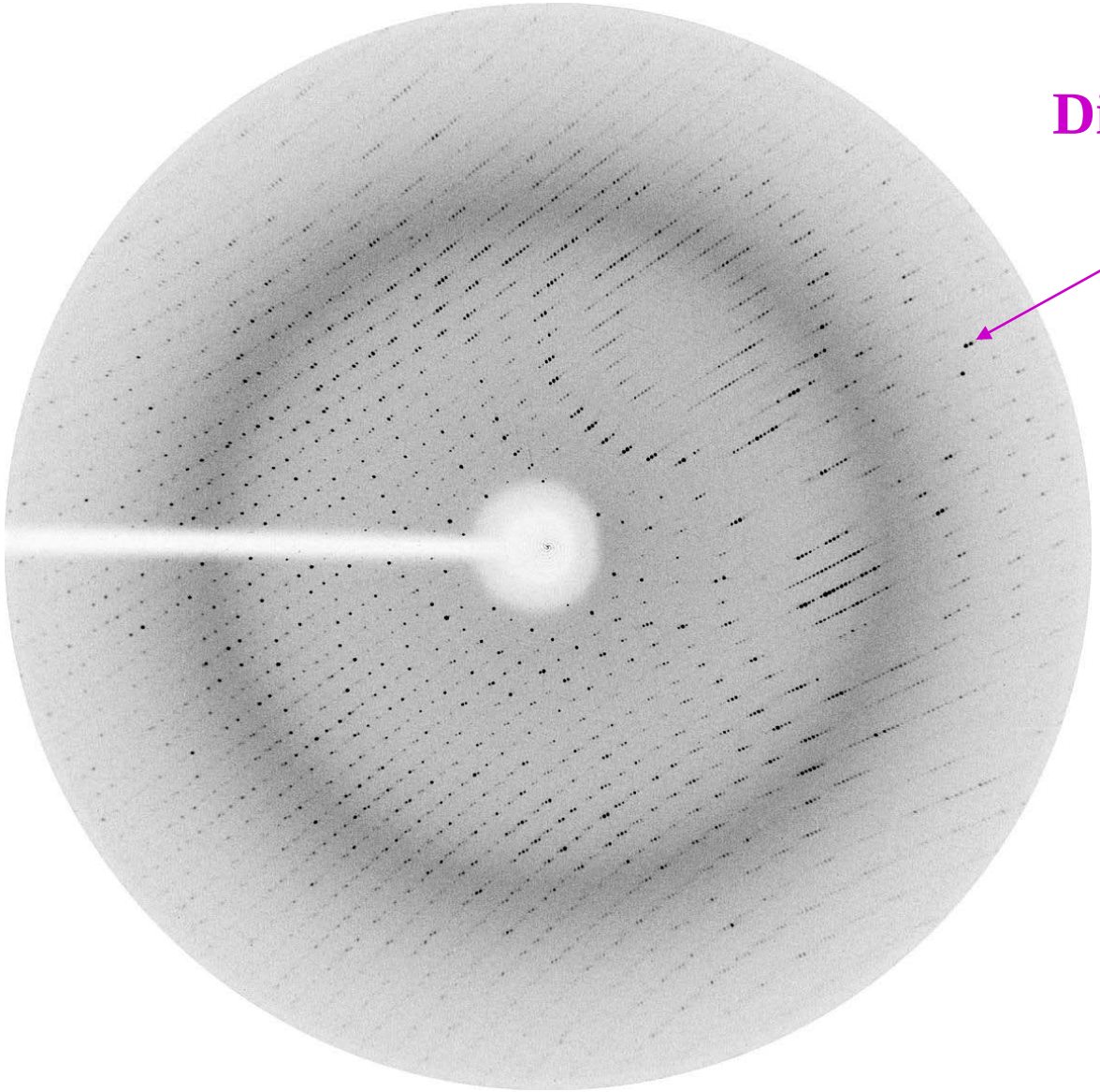
X-Ray Crystallography

1. Make crystals of your protein
   0.3-1.0mm in size
   Proteins must be in an ordered, repeating pattern.

2. X-ray beam is aimed at crystal and data is collected.

3. Structure is determined from the diffraction data.

Protein Diffraction Image



Why Spots?

X-ray scattering from individual proteins is diffuse

Spots arise from a phenomenon called diffraction that is based on the crystal lattice

Location of reflections indicates how an object crystallized

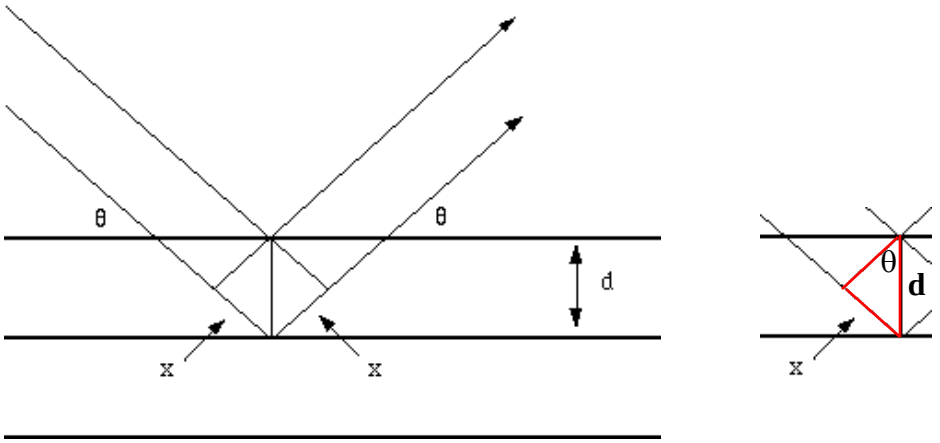
230 possibilities

Intensity of reflections contains information about the structure of the object in the crystal

Bragg's Law

Why do we get spots (reflections) and not a diffuse pattern of scattered x-rays?

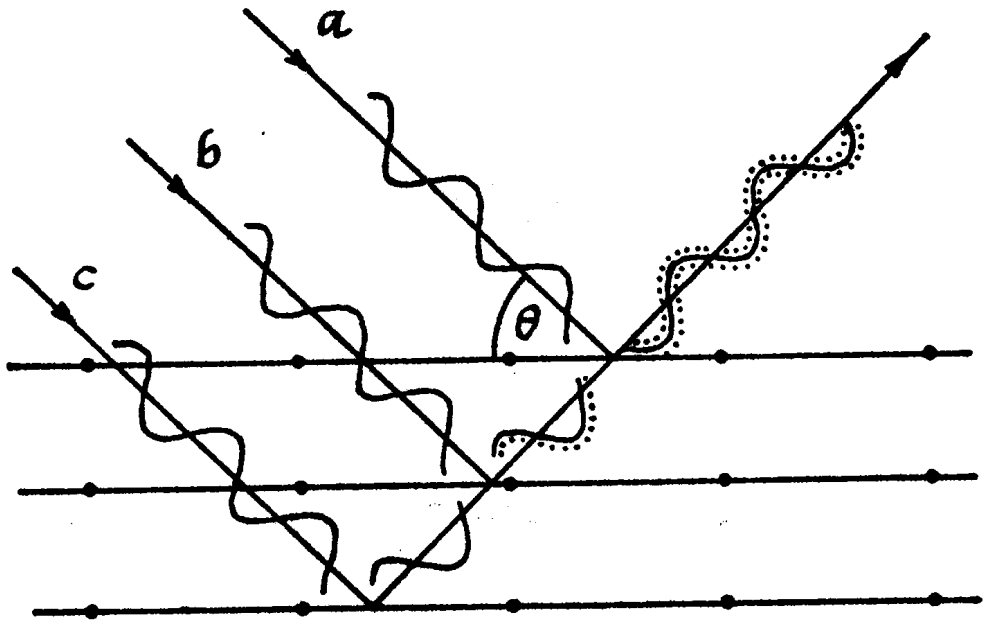
$$2\boldsymbol{d}_{\boldsymbol{hkl}} \sin\theta = \boldsymbol{n}\lambda$$



sin θ = x/d
2d sin θ = 2x

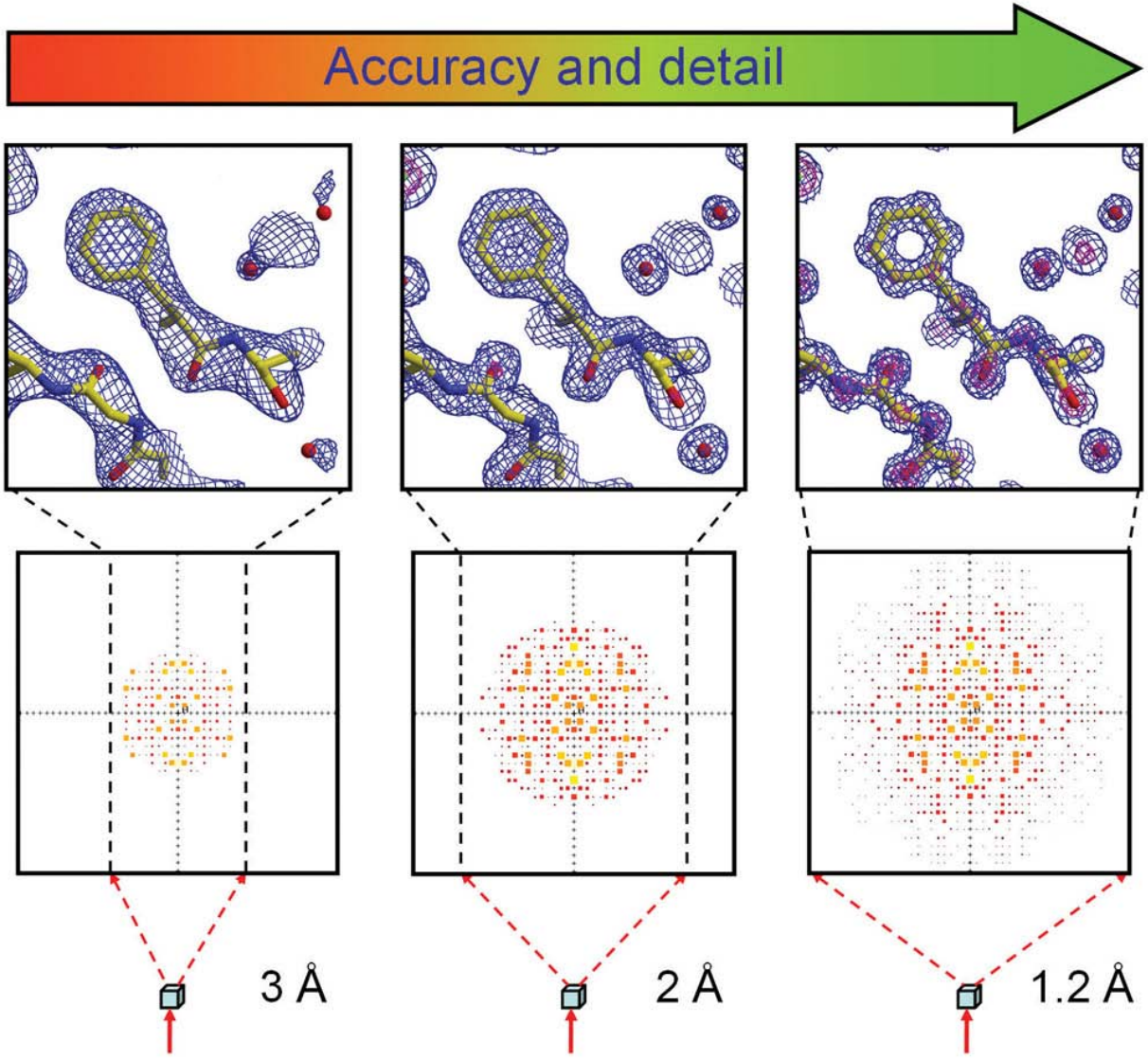
Difference in path (2x) must equal integral number of wavelengths (nλ)

Constructive Interference



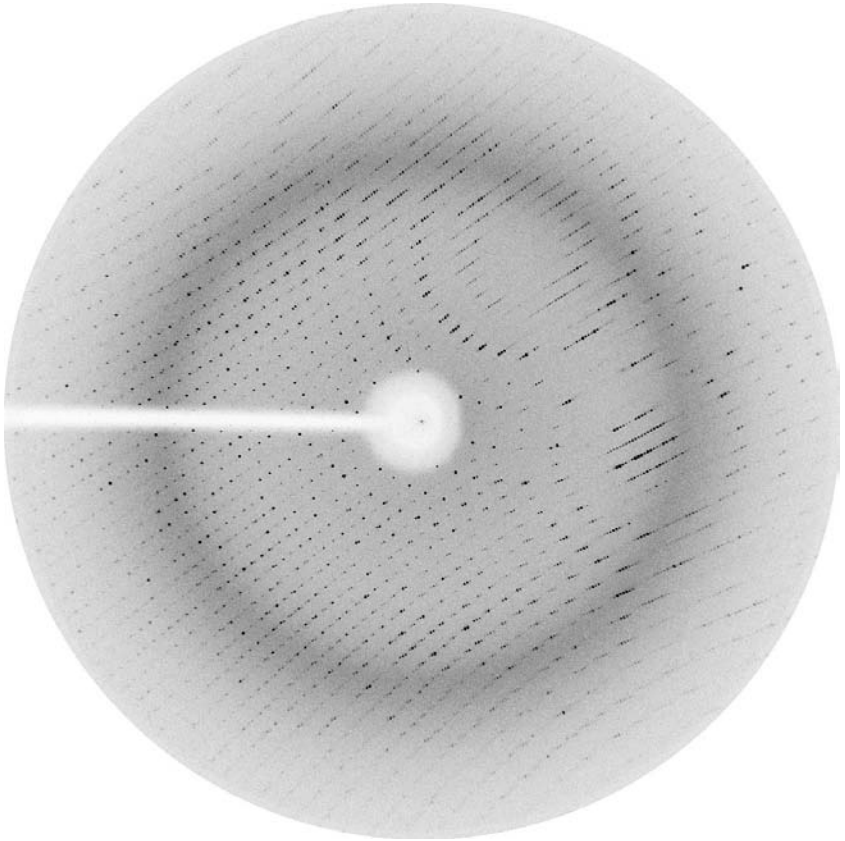
- Condition for reflection

Resolution



http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5

Phase Problem



Every diffraction spot (reflection) has a phase and intensity

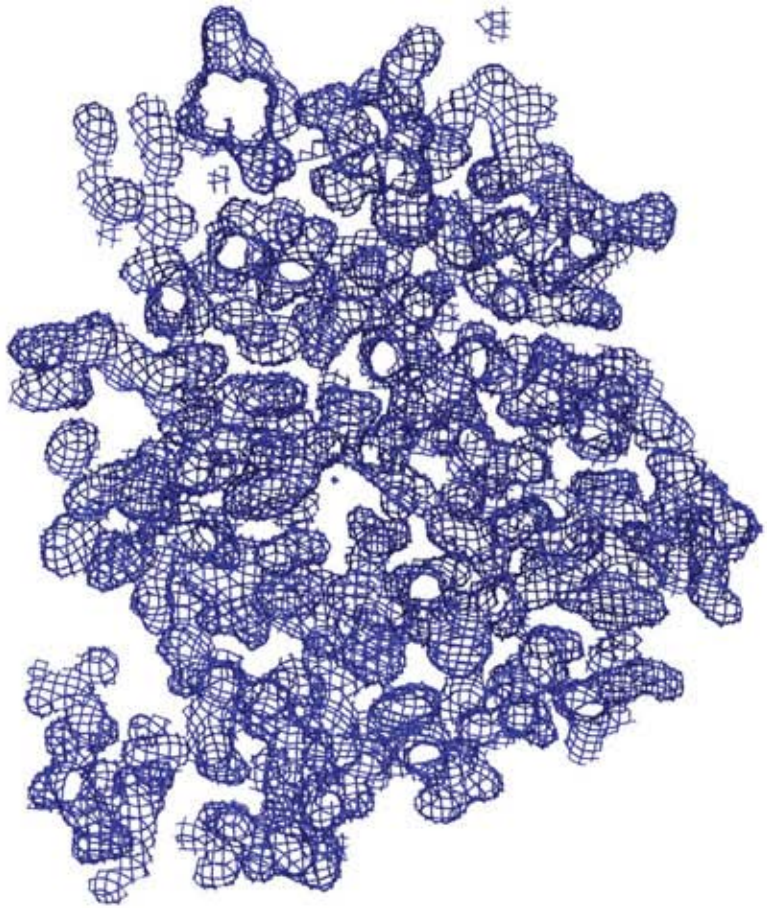
-The intensities are recorded by the detector

-The phases are lost

-Must have both to reconstruct the image (structure)

Phase Problem

h	k	l	F(hkl)	φ(hkl)
2	0	0	228.0	180.0
1	0	1	10.4	90.0
2	0	1	901.8	270.0
1	1	1	367.0	332.1
1	2	3	149.3	37.8
8	9	1	97.9	255.1
7	7	2	111.5	139.7

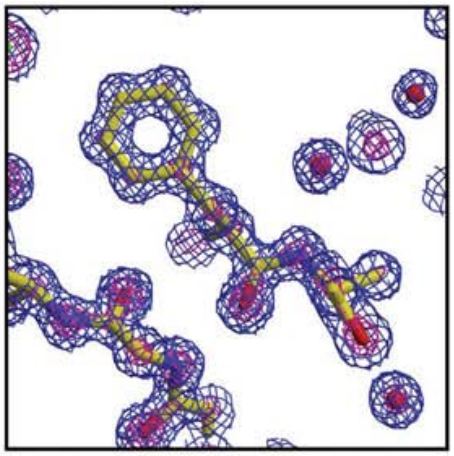


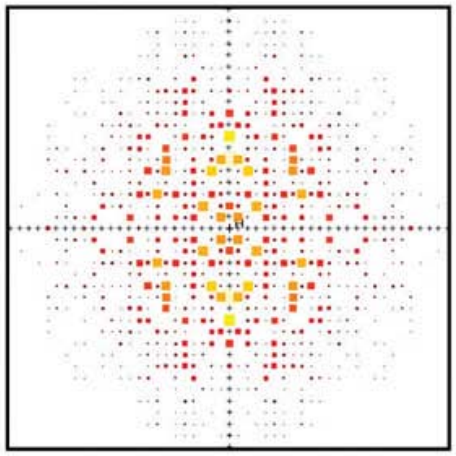
Electron Density

http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5

Phase Problem

$$\rho(x,y,z) = \frac{1}{V}\sum_{-h}^{h}\sum_{-k}^{k}\sum_{-l}^{l} F_{hkl} \cdot \exp[-2\pi i(hx + ky + lz - \alpha_{hkl})]$$





The crystallographic phase problem

Electron Density **Intensities** **Phases**

http://www.ruppweb.org/Garland/gallery/Ch1/index.htm#5

Solutions to the Phase Problem

Molecular replacement

-Use known structure of close homologue

-Rotational and translational search for solution

Heavy atom labeling

-Label the protein with electron dense atoms (Hg)

-Compare independent datasets collected from native and labeled protein

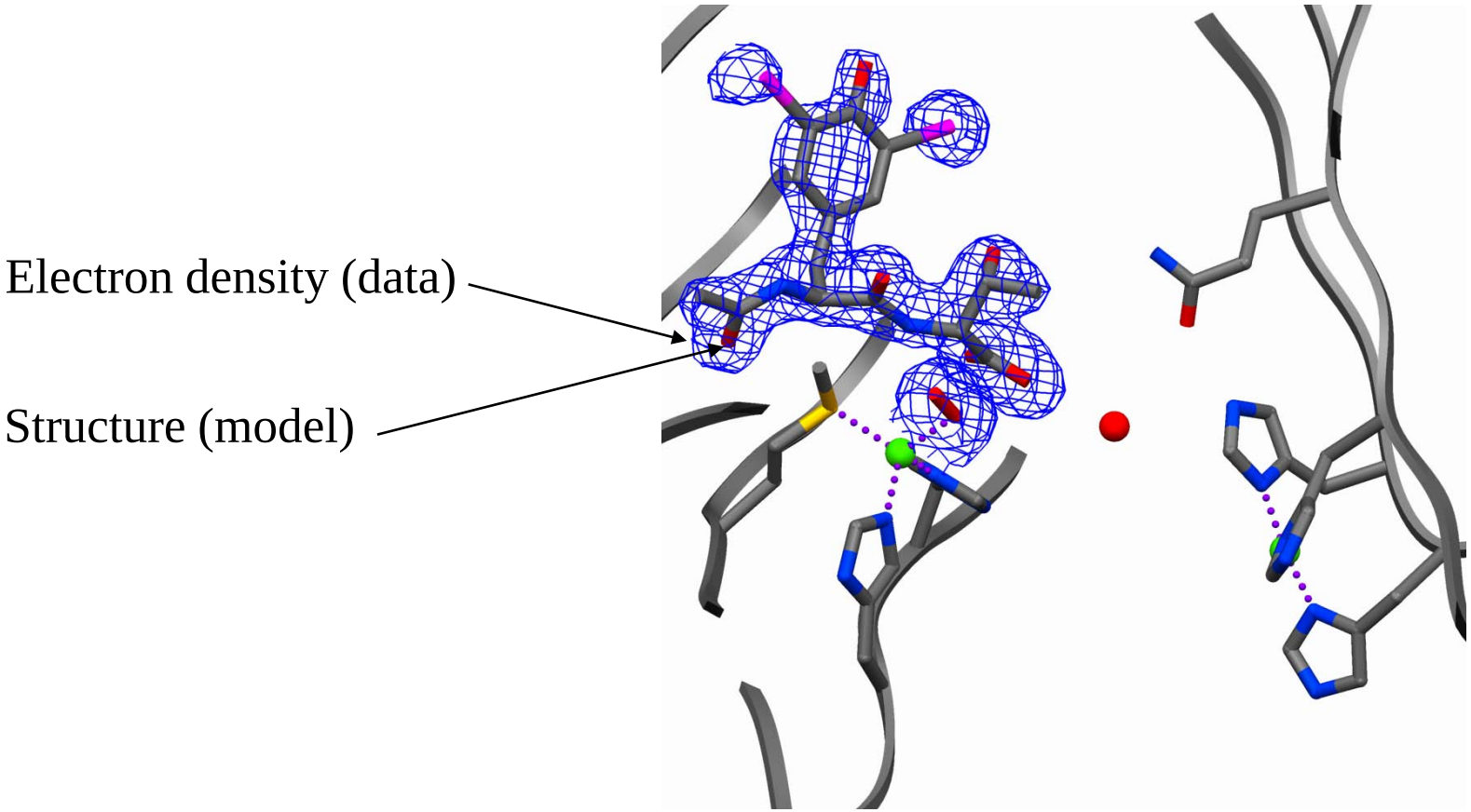
-Heavy atom substructure provides initial phases

Anomalous diffraction

-Crystal must contain atoms with absorption edges between 0.5 and 2.5 Å

-Compare independent datasets collected at pre-edge and post-edge x-ray energies

Model Building



Crystallography Pros/Cons

Advantages

-can be ”fast” – down to a few months
-large structures possible (ribosome)
-very low resolution (down to 0.5 Å)
-observables typically > refinement parameters

Disadvantages

-requires crystal formation
-non-physiological conditions
-crystal contacts can limit protein motion

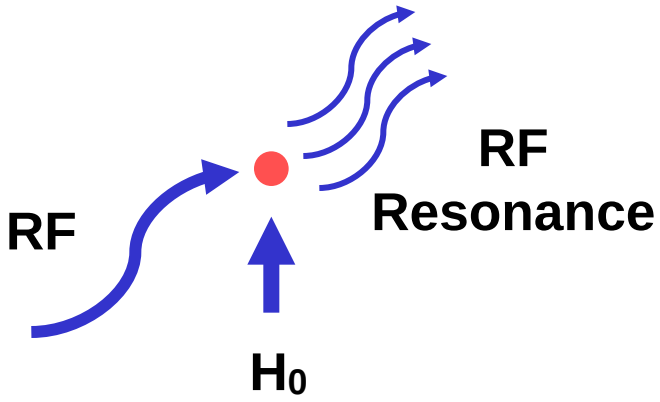
Nuclear Magnetic Resonance

Nuclear Magnetic Resonance

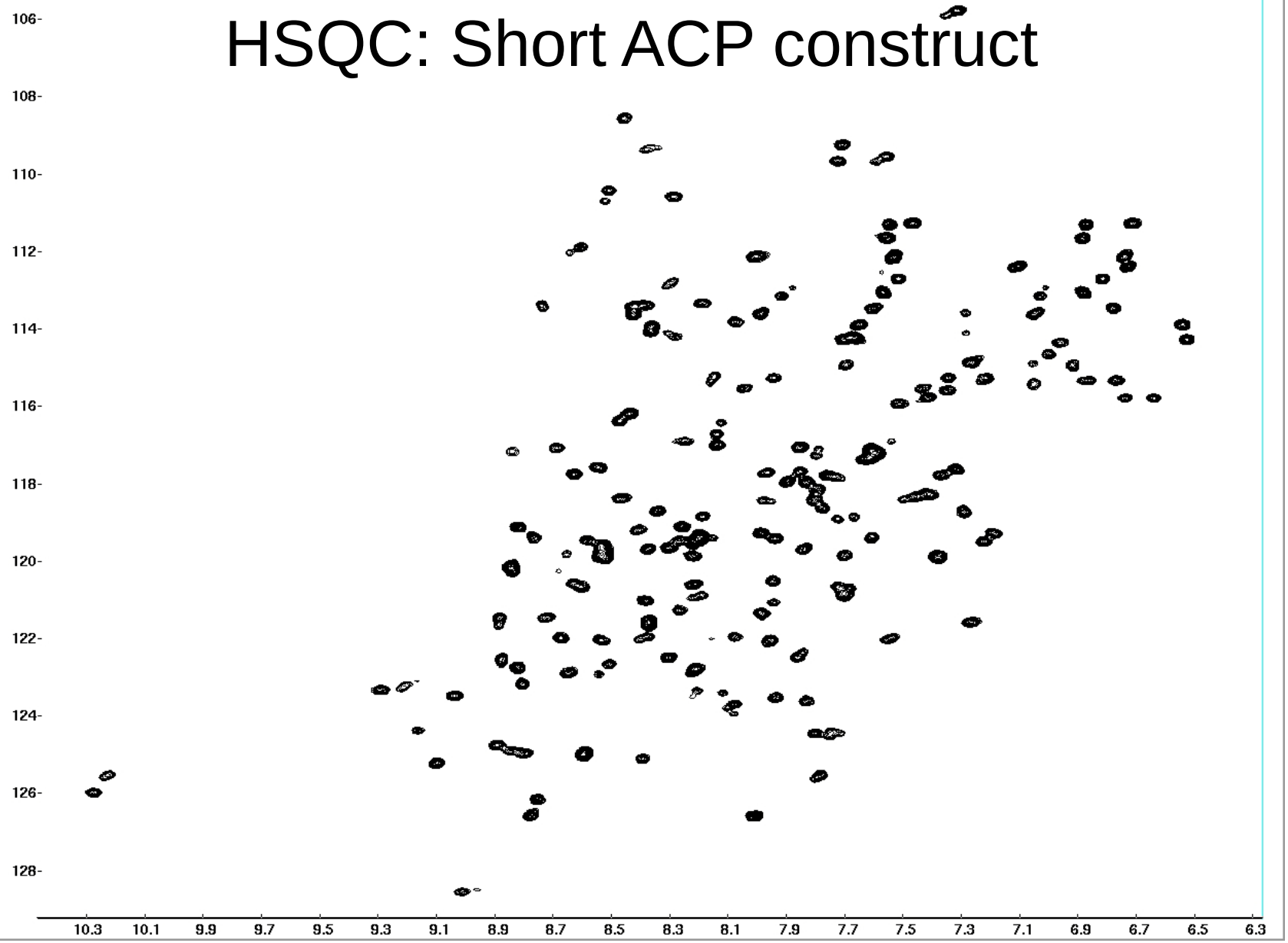
Magnetically align unpaired proton spins (H_0)

Probe with radio frequency (RF)

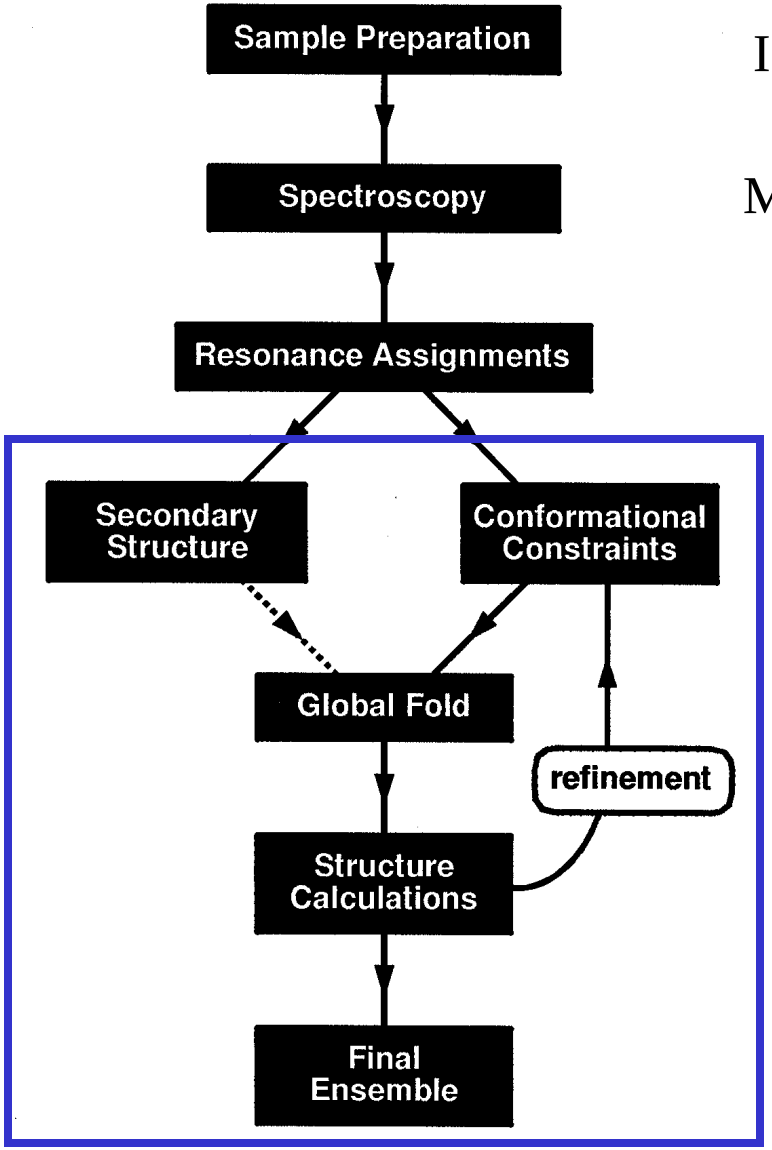
Observe resonance



HSQC: Short ACP construct



NMR Overview



Isotopic labeling (^{15}N, ^{13}C)

Multiple experiments (pulse sequences)

Positional refinement typically not possible. Dihedral angles used.

As many as 20 structures produced

NMR Experimental Observables

- **Backbone conformation from chemical shifts (Chemical Shift Index- CSI)**
- **Distance constraints from NOEs**
- **Hydrogen bond constraints**
- **Backbone and side chain dihedral angle constraints from scalar couplings**
- **Orientation constraints from residual dipolar couplings**

NMR Pros/Cons

Advantages

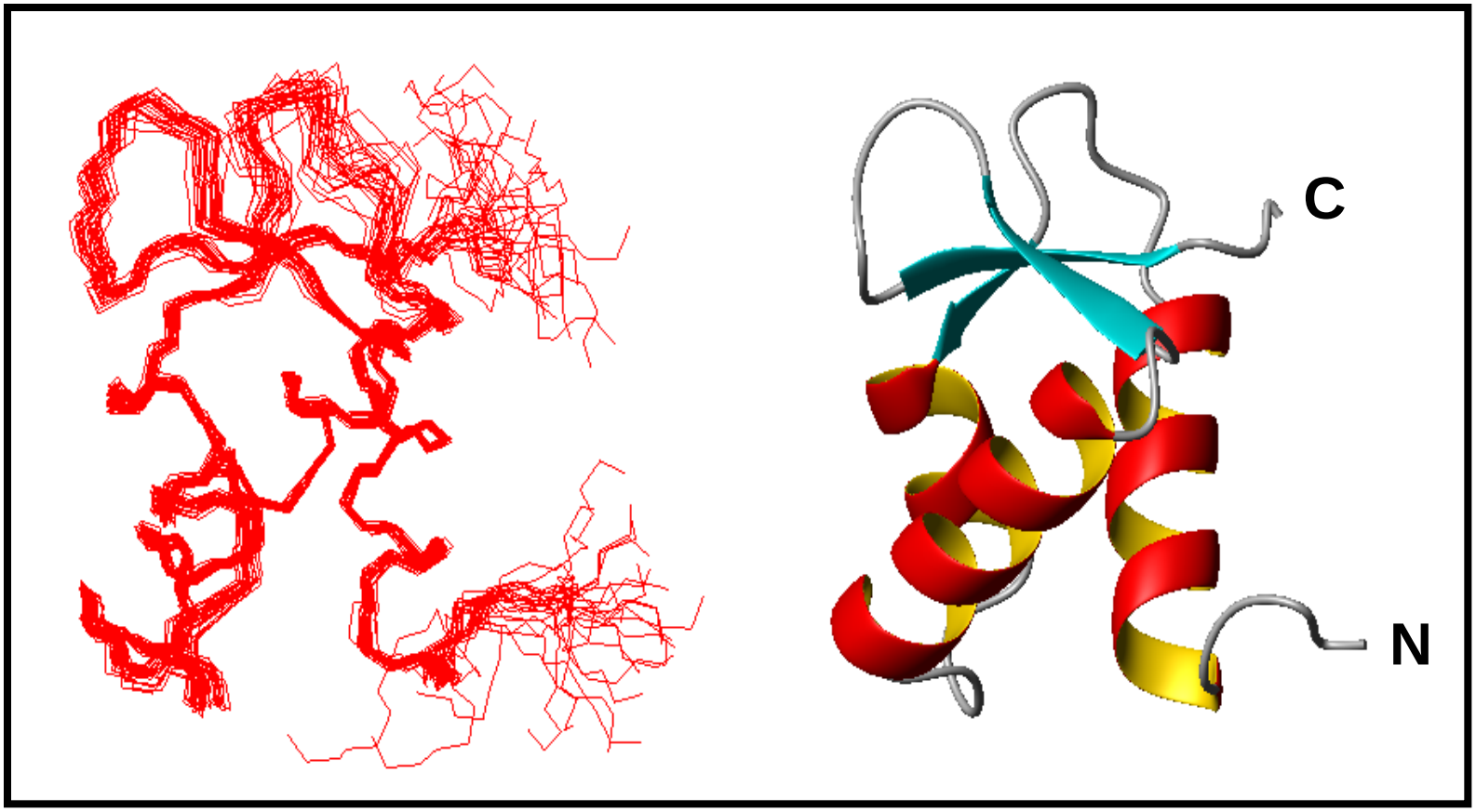
-no crystal formation needed
-more physiological conditions

Disadvantages

-results in a set of models that are compatible with data
-size limitation to 200-300 residues (extended recently)
-must label protein with ^{15}N and ^{13}C
-observables typically < refinement parameters

Precision

NMR vs. X-ray



RMSD of the ensemble **Mean coordinate error**

A PDB File

Header contains information about protein and structure
date of the entry, references, crystallographic data,
contents and positions of secondary structure elements

```
HEADER    OXIDOREDUCTASE                          03-OCT-02   1MXT
TITLE     ATOMIC RESOLUTION STRUCTURE OF CHOLESTEROL OXIDASE
TITLE    2 (STREPTOMYCES SP. SA-COO)
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: CHOLESTEROL OXIDASE;
COMPND   3 CHAIN: A;
COMPND   4 SYNONYM: CHOD;
COMPND   5 EC: 1.1.3.6;
COMPND   6 ENGINEERED: YES;
COMPND   7 OTHER_DETAILS: FAD COFACTOR NON-COVALENTLY BOUND TO THE
COMPND   8 ENZYME
```

A PDB File

Header contains information about protein and structure
date of the entry, references, crystallographic data,
contents and positions of secondary structure elements

```
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: STREPTOMYCES SP.;
SOURCE   3 ORGANISM_COMMON: BACTERIA;
SOURCE   4 GENE: CHOA;
SOURCE   5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   6 EXPRESSION_SYSTEM_COMMON: BACTERIA;
SOURCE   7 EXPRESSION_SYSTEM_STRAIN: BL21(DE3)PLYSS;
SOURCE   8 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE   9 EXPRESSION_SYSTEM_PLASMID: PCO202
```

A PDB File

Header contains information about protein and structure
date of the entry, references, crystallographic data,
contents and positions of secondary structure elements

```
AUTHOR    A.VRIELINK,P.I.LARIO
REVDAT   1   25-FEB-03 1MXT    0
JRNL       AUTH   P.I.LARIO,N.SAMPSON,A.VRIELINK
JRNL       TITL   SUB-ATOMIC RESOLUTION CRYSTAL STRUCTURE OF
JRNL       TITL 2 CHOLESTEROL OXIDASE: WHAT ATOMIC RESOLUTION
JRNL       TITL 3 CRYSTALLOGRAPHY REVEALS ABOUT ENZYME MECHANISM AND
JRNL       TITL 4 THE ROLE OF FAD COFACTOR IN REDOX ACTIVITY
JRNL       REF    J.MOL.BIOL.                V. 326  1635 2003
JRNL       REFN   ASTM JMOBAK  UK ISSN 0022-2836
```

A PDB File

Header contains information about protein and structure
date of the entry, references, crystallographic data,
contents and positions of secondary structure elements

```
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 0.95
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 28.00
REMARK   3   DATA CUTOFF            (SIGMA(F)) : 0.000
REMARK   3   COMPLETENESS FOR RANGE       (%) : 94.1
REMARK   3   CROSS-VALIDATION METHOD          : FREE R
REMARK   3   FREE R VALUE TEST SET SELECTION   : RANDOM
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT (NO CUTOFF).
REMARK   3   R VALUE   (WORKING + TEST SET, NO CUTOFF) : 0.110
REMARK   3   R VALUE          (WORKING SET, NO CUTOFF) : 0.110
REMARK   3   FREE R VALUE                (NO CUTOFF) : 0.132
REMARK   3   FREE R VALUE TEST SET SIZE (%, NO CUTOFF) : 5.000
REMARK   3   FREE R VALUE TEST SET COUNT   (NO CUTOFF) : 13180

REMARK   3   TOTAL NUMBER OF REFLECTIONS   (NO CUTOFF) : 263551
```

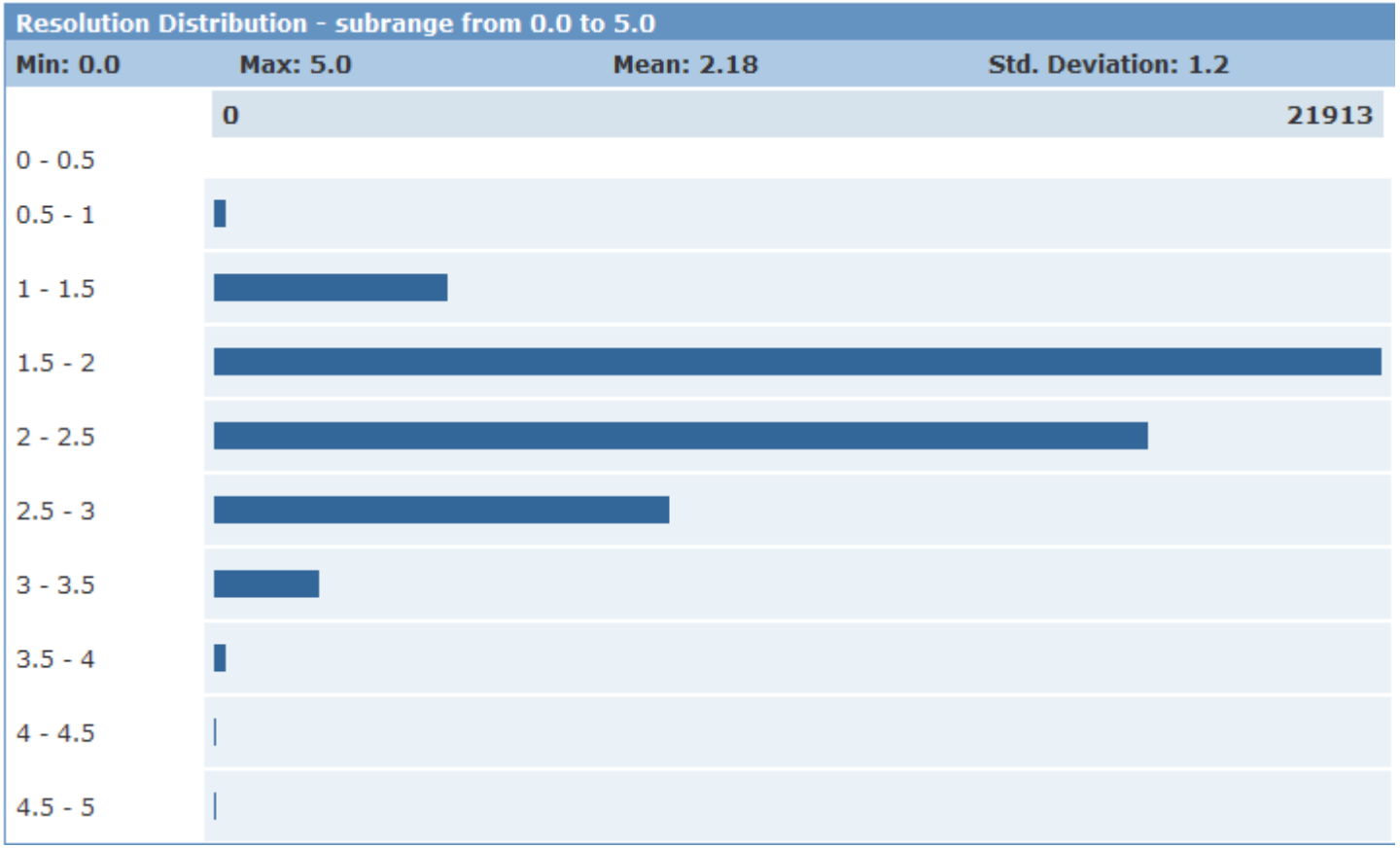
Resolution:

Low > 3 Å
Mid 2-3 Å
High 1.5-2 Å
Very High < 1.5 Å

R factor (residual):

Low resolution ~ 27%
Mid resolution ~ 22 %
High resolution ~ 29 %
Very High res ~ 15%

Resolution


Resolution Distribution - subrange from 0.0 to 5.0
Min: 0.0
Max: 5.0
Mean: 2.18
Std. Deviation: 1.2
0
21913
0 - 0.5
0.5 - 1
1 - 1.5
1.5 - 2
2 - 2.5
2.5 - 3
3 - 3.5
3.5 - 4
4 - 4.5
4.5 - 5

A PDB File

Header contains information about protein and structure
date of the entry, references, crystallographic data,
contents and positions of secondary structure elements

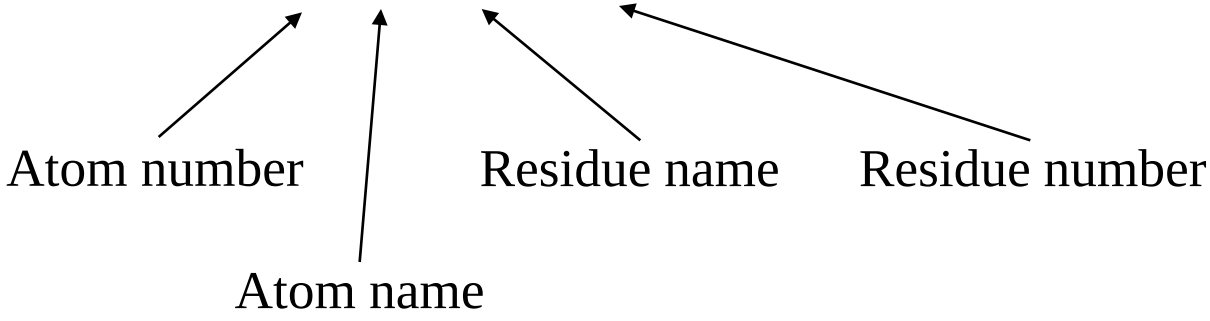
```
HELIX   14  14 ALA A  289  THR A  304  1                                   16
HELIX   15  15 THR A  402  GLN A  405  5                                    4
HELIX   16  16 ASN A  406  GLY A  425  1                                   20
HELIX   17  17 ASP A  474  ILE A  478  5                                    5
HELIX   18  18 PRO A  486  VAL A  506  1                                   21
SHEET    1   A 6 HIS A 248  GLN A 255  0
SHEET    2   A 6 TYR A 261  LYS A 268 -1  O  GLU A 266   N  GLN A 249
SHEET    3   A 6 LEU A 274  LEU A 287 -1  O  LEU A 275   N  GLN A 267
SHEET    4   A 6 TYR A  10  ILE A  16  1  N  VAL A  14   O  PHE A 286
SHEET    5   A 6 THR A  36  GLU A  40  1  O  LEU A  37   N  VAL A  15

SHEET    6   A 6 VAL A 242  THR A 246  1  O  THR A 243   N  MET A  38
```

A PDB File

Body of PDB file contains information about the atoms in the structure

```
ATOM     76  N   PRO A  12      31.129  -4.659  43.245  1.00  9.00           N
ATOM     77  CA  PRO A  12      32.426  -4.662  42.542  1.00  9.00           C
ATOM     78  C   PRO A  12      32.423  -4.009  41.182  1.00  8.02           C
ATOM     79  O   PRO A  12      33.267  -3.177  40.892  1.00  8.31           O
ATOM     80  CB  PRO A  12      32.791  -6.126  42.592  1.00 10.02           C
ATOM     81  CG  PRO A  12      32.190  -6.663  43.857  1.00 10.12           C
ATOM     82  CD  PRO A  12      30.850  -5.927  43.925  1.00  9.87           C
ATOM     90  N   ALA A  13      31.485  -4.468  40.316  1.00  8.06           N
ATOM     91  CA  ALA A  13      31.357  -3.854  39.004  1.00  7.28           C
ATOM     92  C   ALA A  13      29.947  -3.309  38.814  1.00  7.21           C
ATOM     93  O   ALA A  13      28.969  -3.932  39.200  1.00  7.56           O
ATOM     94  CB  ALA A  13      31.636  -4.879  37.897  1.00  8.54           C
```

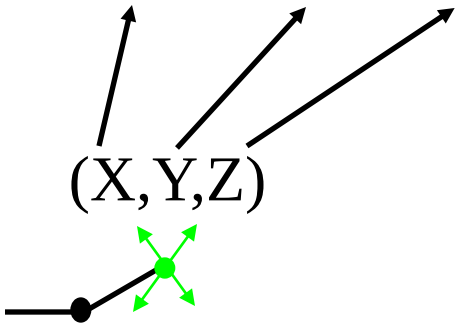


A PDB File

Body of PDB file contains information about the atoms in the structure

```
ATOM     76  N   PRO A  12     31.129  -4.659  43.245  1.00  9.00           N
ATOM     77  CA  PRO A  12     32.426  -4.662  42.542  1.00  9.00           C
ATOM     78  C   PRO A  12     32.423  -4.009  41.182  1.00  8.02           C
ATOM     79  O   PRO A  12     33.267  -3.177  40.892  1.00  8.31           O
ATOM     80  CB  PRO A  12     32.791  -6.126  42.592  1.00 10.02           C
ATOM     81  CG  PRO A  12     32.190  -6.663  43.857  1.00 10.12           C
ATOM     82  CD  PRO A  12     30.850  -5.927  43.925  1.00  9.87           C
ATOM     90  N   ALA A  13     31.485  -4.468  40.316  1.00  8.06           N
ATOM     91  CA  ALA A  13     31.357  -3.854  39.004  1.00  7.28           C
ATOM     92  C   ALA A  13     29.947  -3.309  38.814  1.00  7.21           C
ATOM     93  O   ALA A  13     28.969  -3.932  39.200  1.00  7.56           O
ATOM     94  CB  ALA A  13     31.636  -4.879  37.897  1.00  8.54           C
```

Coordinates in Å



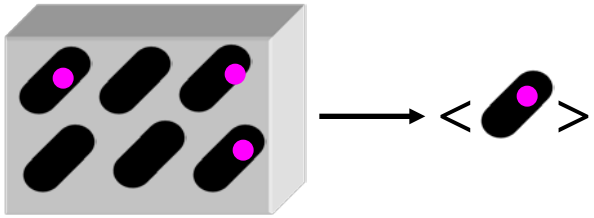
Mean coordinate error:

Low > 3 Å	.4 Å
Mid 2-3 Å	.3 Å
High 1.5-2 Å	.2 Å
Very High < 1.5 Å	.1 Å

A PDB File

Body of PDB file contains information about the atoms in the structure

```
ATOM     76  N   PRO A  12      31.129  -4.659  43.245  1.00  9.00           N
ATOM     77  CA  PRO A  12      32.426  -4.662  42.542  1.00  9.00           C
ATOM     78  C   PRO A  12      32.423  -4.009  41.182  1.00  8.02           C
ATOM     79  O   PRO A  12      33.267  -3.177  40.892  1.00  8.31           O
ATOM     80  CB  PRO A  12      32.791  -6.126  42.592  1.00 10.02           C
ATOM     81  CG  PRO A  12      32.190  -6.663  43.857  1.00 10.12           C
ATOM     82  CD  PRO A  12      30.850  -5.927  43.925  1.00  9.87           C
ATOM     90  N   ALA A  13      31.485  -4.468  40.316  1.00  8.06           N
ATOM     91  CA  ALA A  13      31.357  -3.854  39.004  1.00  7.28           C
ATOM     92  C   ALA A  13      29.947  -3.309  38.814  1.00  7.21           C
ATOM     93  O   ALA A  13      28.969  -3.932  39.200  1.00  7.56           O
ATOM     94  CB  ALA A  13      31.636  -4.879  37.897  1.00  8.54           C
```



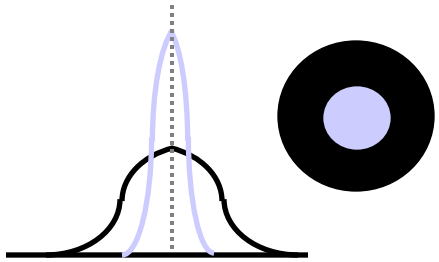
- Occupancy of 0.5

Fractional occupancy

A PDB File

Body of PDB file contains information about the atoms in the structure

```
ATOM     76  N   PRO A  12      31.129  -4.659  43.245  1.00  9.00           N
ATOM     77  CA  PRO A  12      32.426  -4.662  42.542  1.00  9.00           C
ATOM     78  C   PRO A  12      32.423  -4.009  41.182  1.00  8.02           C
ATOM     79  O   PRO A  12      33.267  -3.177  40.892  1.00  8.31           O
ATOM     80  CB  PRO A  12      32.791  -6.126  42.592  1.00 10.02           C
ATOM     81  CG  PRO A  12      32.190  -6.663  43.857  1.00 10.12           C
ATOM     82  CD  PRO A  12      30.850  -5.927  43.925  1.00  9.87           C
ATOM     90  N   ALA A  13      31.485  -4.468  40.316  1.00  8.06           N
ATOM     91  CA  ALA A  13      31.357  -3.854  39.004  1.00  7.28           C
ATOM     92  C   ALA A  13      29.947  -3.309  38.814  1.00  7.21           C
ATOM     93  O   ALA A  13      28.969  -3.932  39.200  1.00  7.56           O
ATOM     94  CB  ALA A  13      31.636  -4.879  37.897  1.00  8.54           C
```



B-factor Å^2