X-ray Crystallography

Through X-ray crystallography, the chemical structure of thousands of organic, inorganic, organometallic, and biological compounds are determined every year.

Biochemistry and Molecular Biology

8.4.5 Protein Tertiary structure and X-ray crystallography

8.5 Proteomics and Protein Function

Macromolecular Crystallography Facility in the Dept. of Molecular Biology at Princeton University.

Overview of Crystallographic Methods and Basic Theory

同步輻射蛋白質結晶學的近期發展與應用研究 /陳俊榮 博士
物理雙月刊（廿八卷四期）2006 年8 月

X-ray crystallography

How to extend the principle of microscopy to the study of macromolecules ?

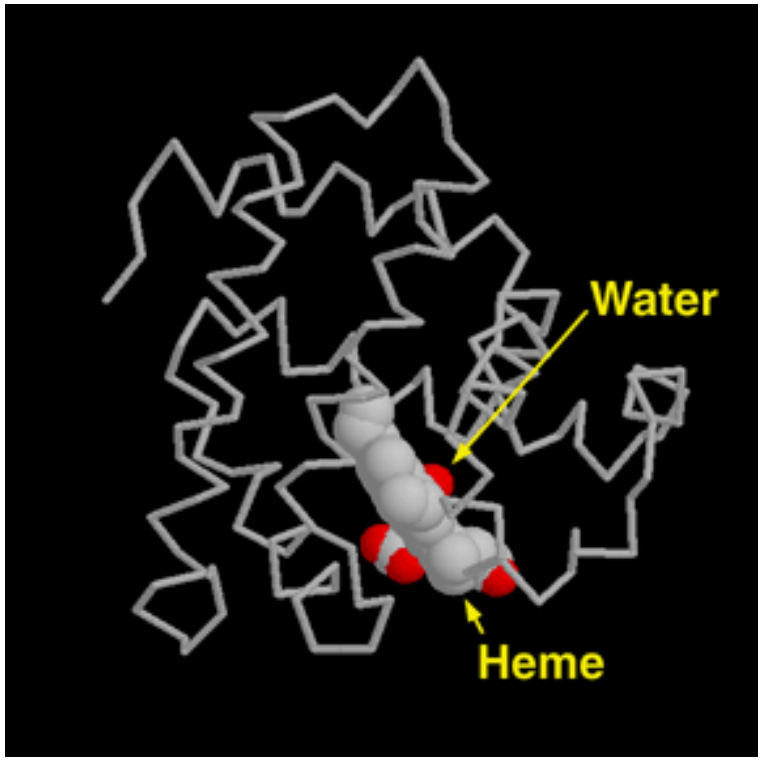
Like NMR,: x-ray crystallography provide 3-D mapping of the atoms in a crystal

Procedures:

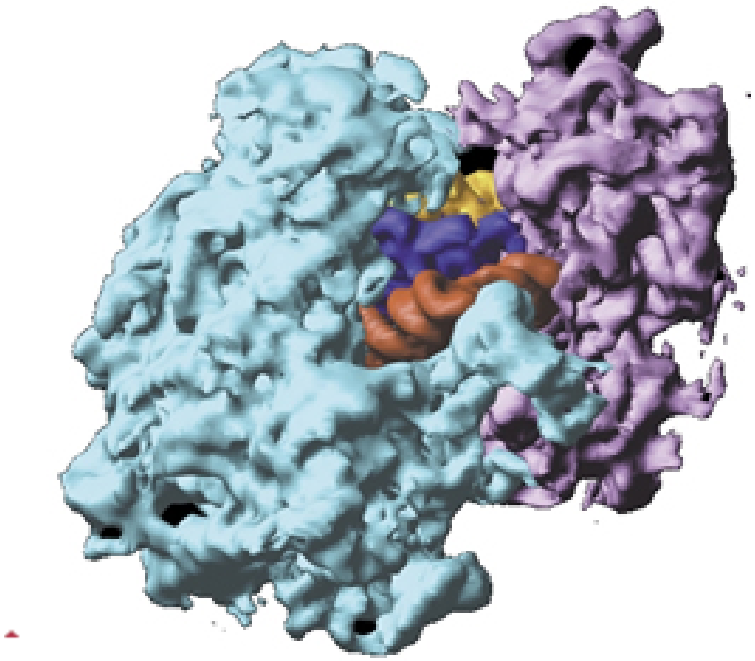
- Crystallization of molecule and Isolation of a crystal (>10mg)
- Perform X-ray diffraction experiment
- Obtain electron density map
- Determine molecular structure that agrees with the diffraction data

The First X-Ray Protein Structure: Myoglobin

The first time researchers glimpsed the complex internal structure of a protein was in 1959, when John Kendrew, working at Cambridge University, determined the structure of myoglobin using X-ray crystallography.

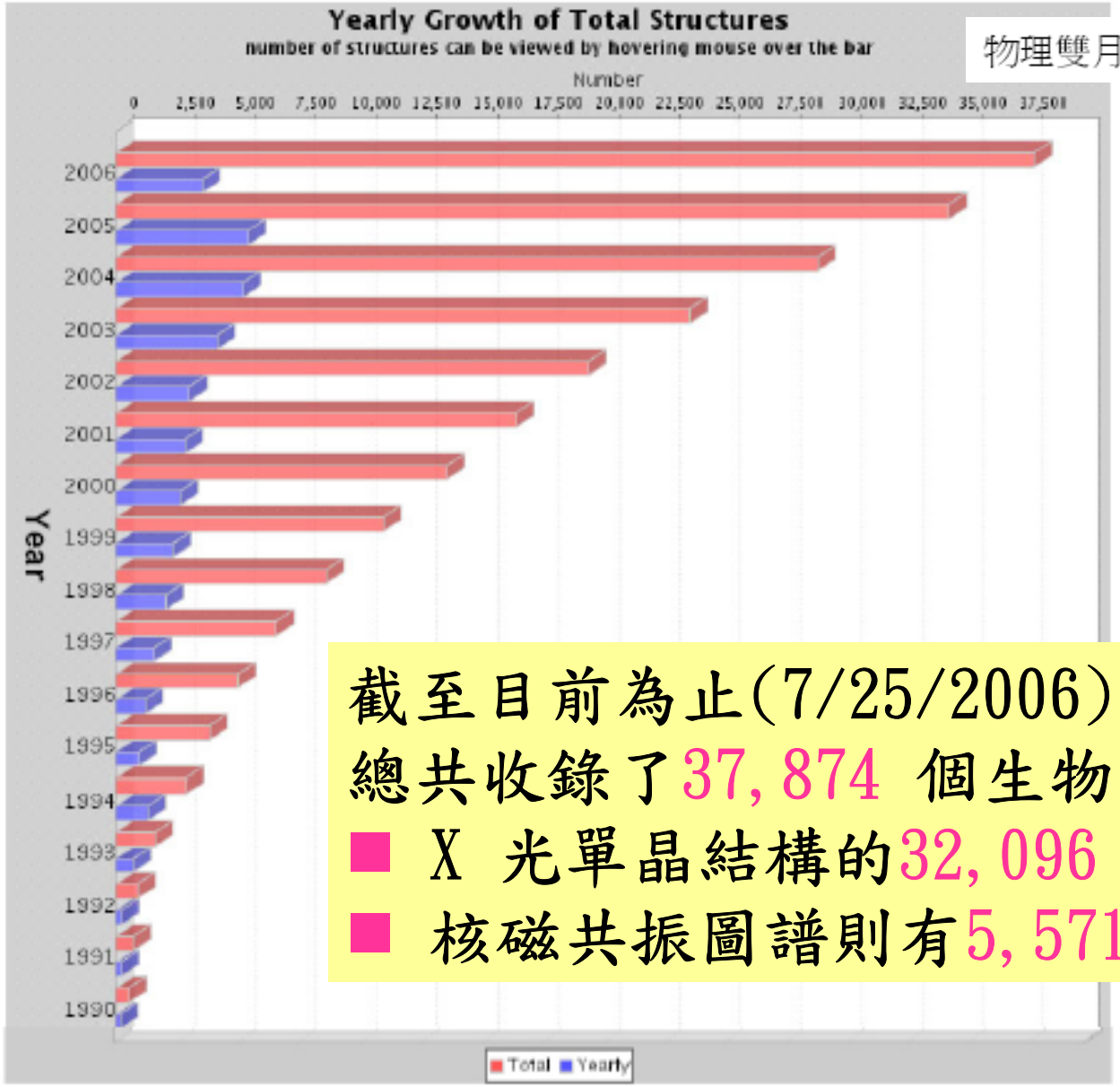


Now: *Ribosome Structure*



The first structural snapshot of an entire bacterial ribosome. The structure, which is the largest determined by X-ray crystallography to date, will help researchers better understand the fundamental process of protein production. It may also aid efforts to design new antibiotic drugs or optimize existing drugs

*Ribosome: 3 RNA, 50 proteins, **2.5 million daltons** !!!!*



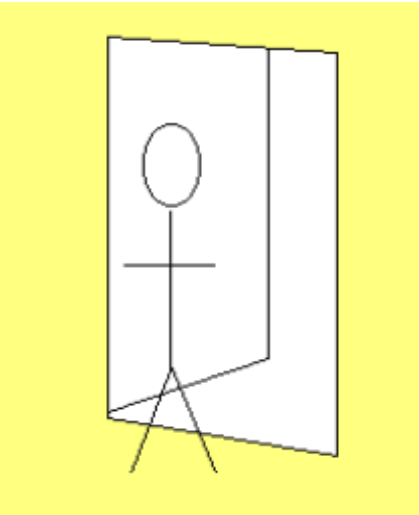
物理雙月刊（廿八卷四期）2006 年 8 月

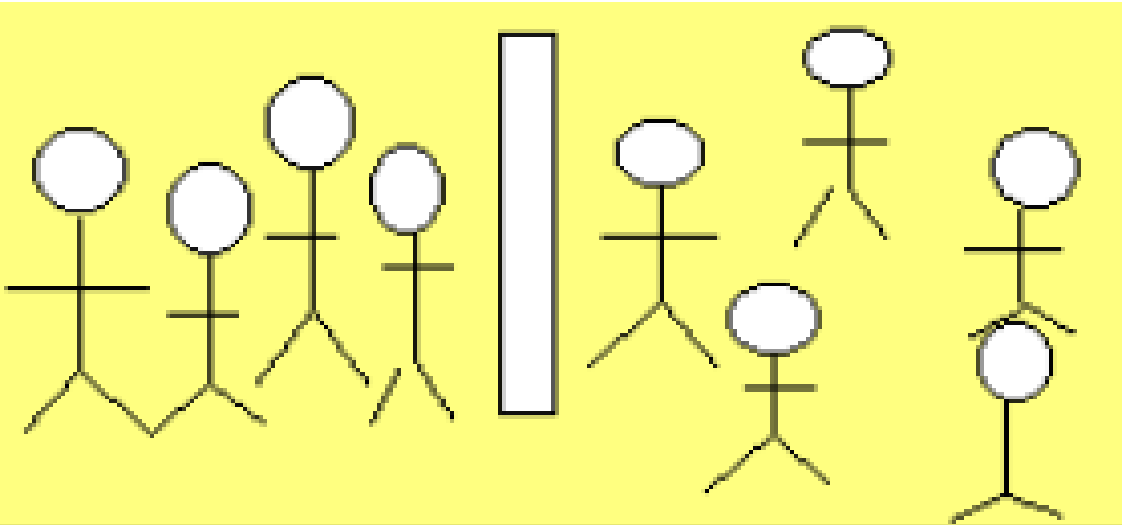
截至目前為止(7/25/2006)
總共收錄了37,874 個生物巨分子結構，

- X 光單晶結構的32,096 個(約佔84.7 %)
- 核磁共振圖譜則有5,571 個(約佔14.7 %)

圖二 近十五年(1990~2006 年)蛋白質資庫中結構發表之統計數據 (資料來源：RCSB PDB)

What is Diffraction ?

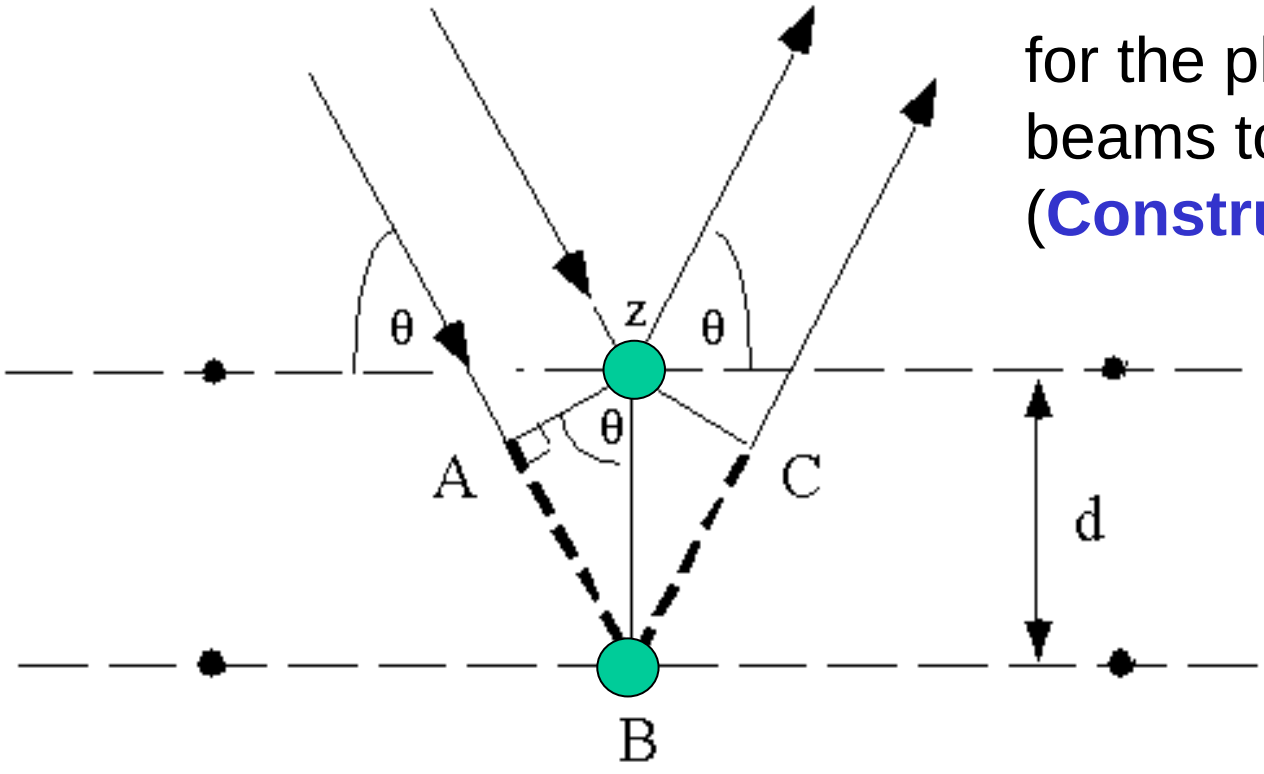




A crowd of people: Diffraction will occur (depend on the angle they came to door)

Diffraction occurs when the incident light and the size of the grating (i.e.“size of the door”) are similar.

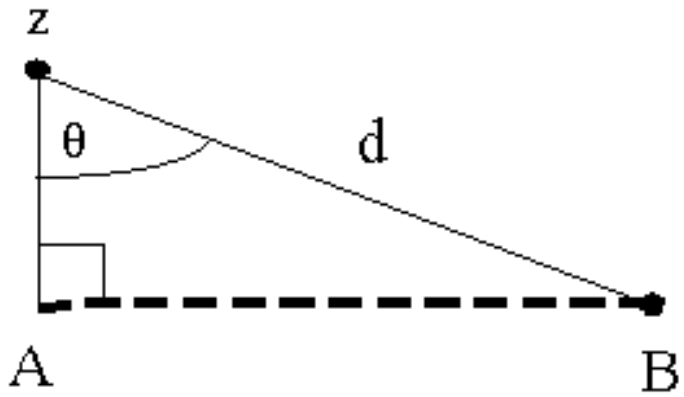
Bragg's Law



for the phases of the two beams to be the same (**Constructive interference**)

$n\lambda = AB + BC$

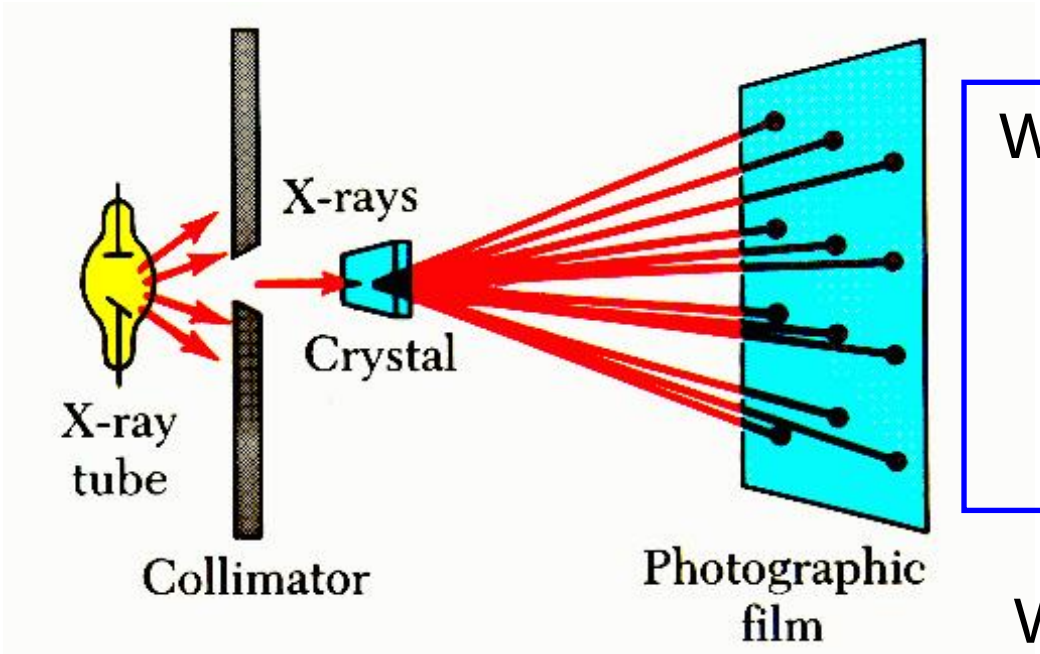
$AB = d \sin\theta$



cause AB = BC

$n\lambda = 2AB$

$\lambda = 2\, d \sin\theta_{h.k.l}$

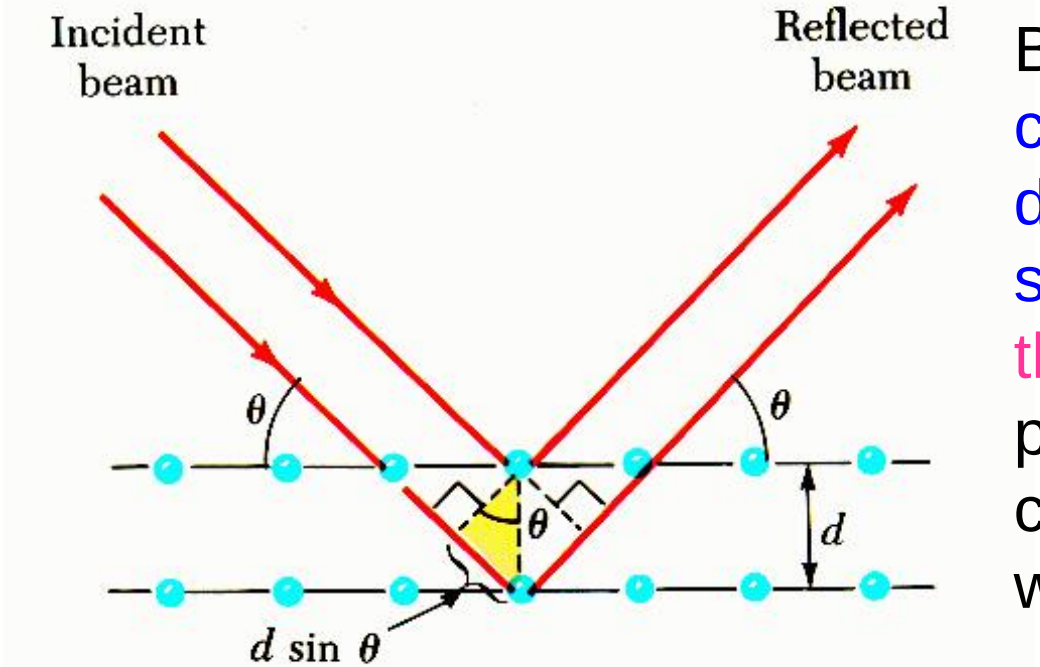


Waves add in some directions

=

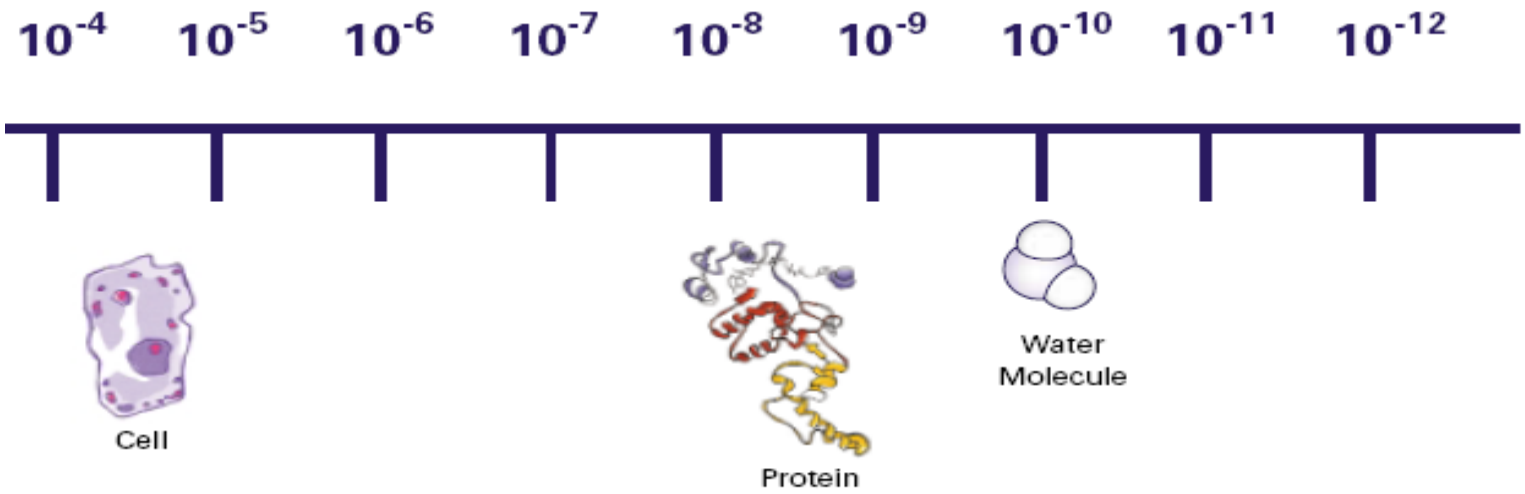
They cancel out in others

$\Rightarrow$ Array of spots



When we put a crystal into an X-ray beam, some of the Bragg planes will be in the correct orientation to show diffraction, and we will see spots for them. If we rotate the crystal, other sets of planes will come into the correct orientation, and we will see new diffraction spots.

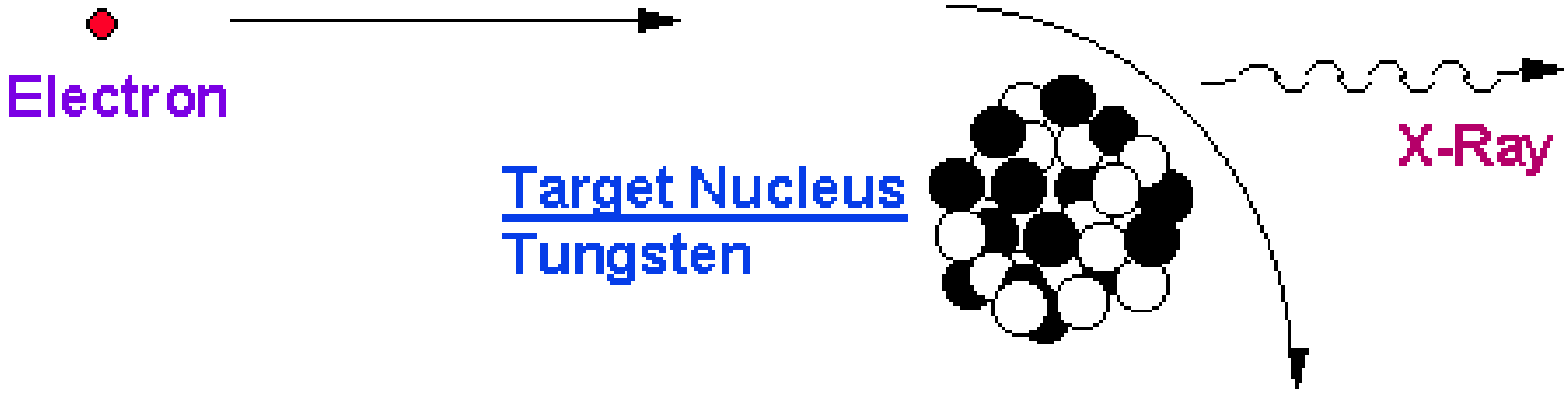
Principles—Why X-rays?



0.5 to 1.5 angstroms (1Å=10^{-10} meters).

– very similar to distance between bonded carbon atoms

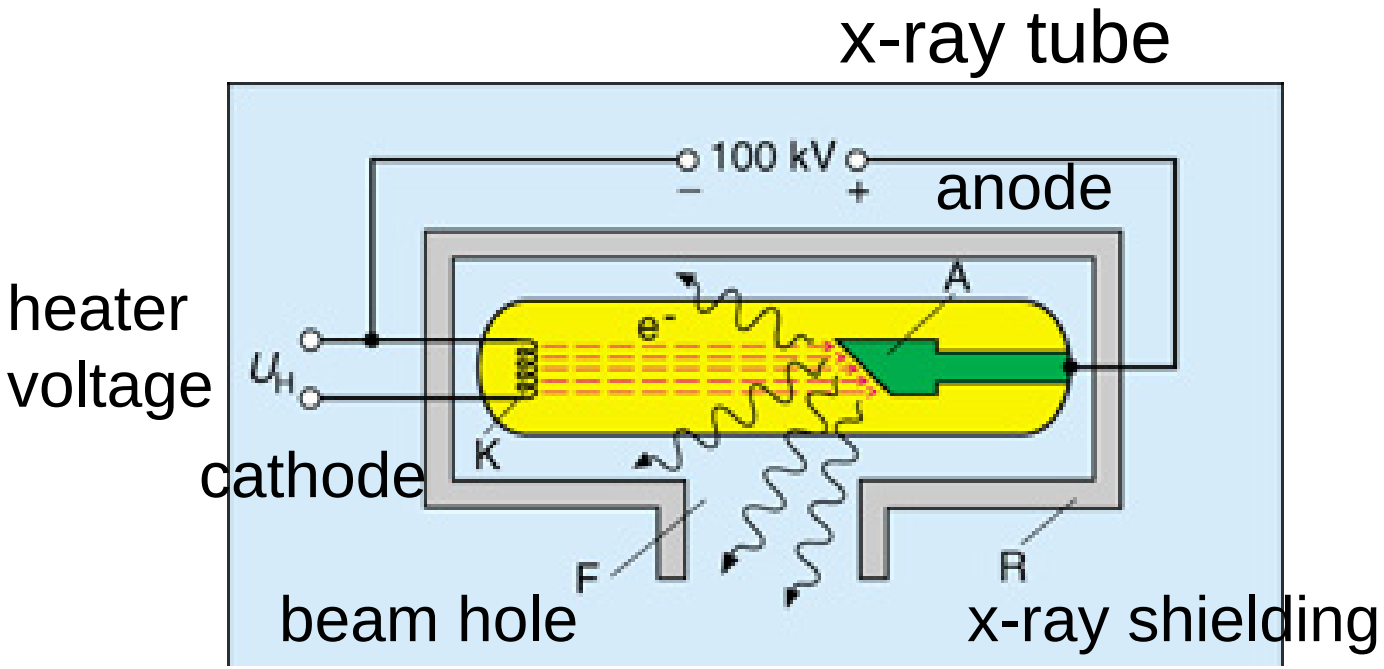
X-Ray Radiation



x-ray is a packet (or photon) of electromagnetic radiation emitted from an atom

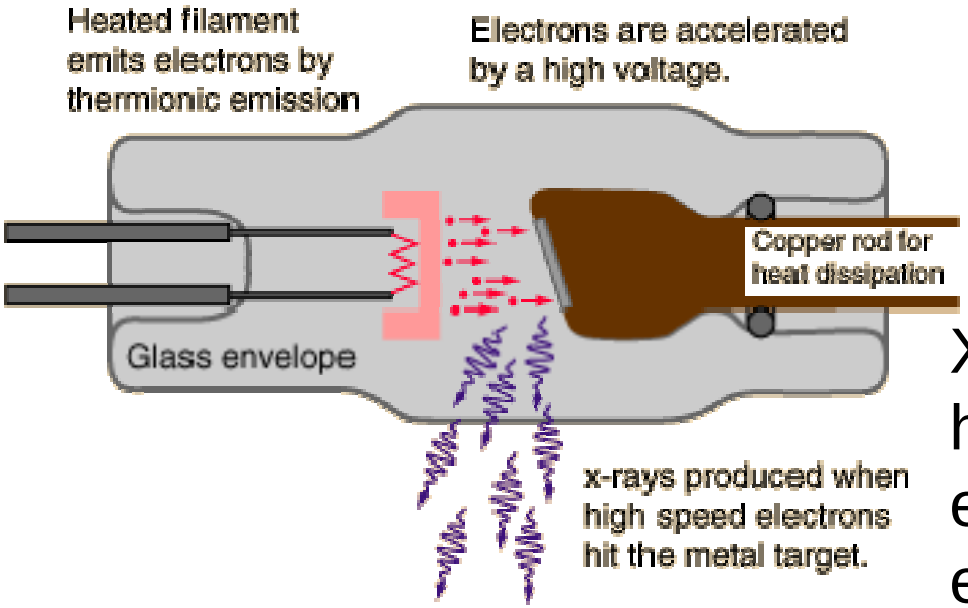
x-rays produced when high-energy electrons strike a target made of a heavy metal, such as tungsten or copper. As electrons collide with this material, some have their paths deflected by the nucleus of the metal atoms. This deflection results in the production of x-rays as the electrons lose energy

How to generate X-Ray Radiation

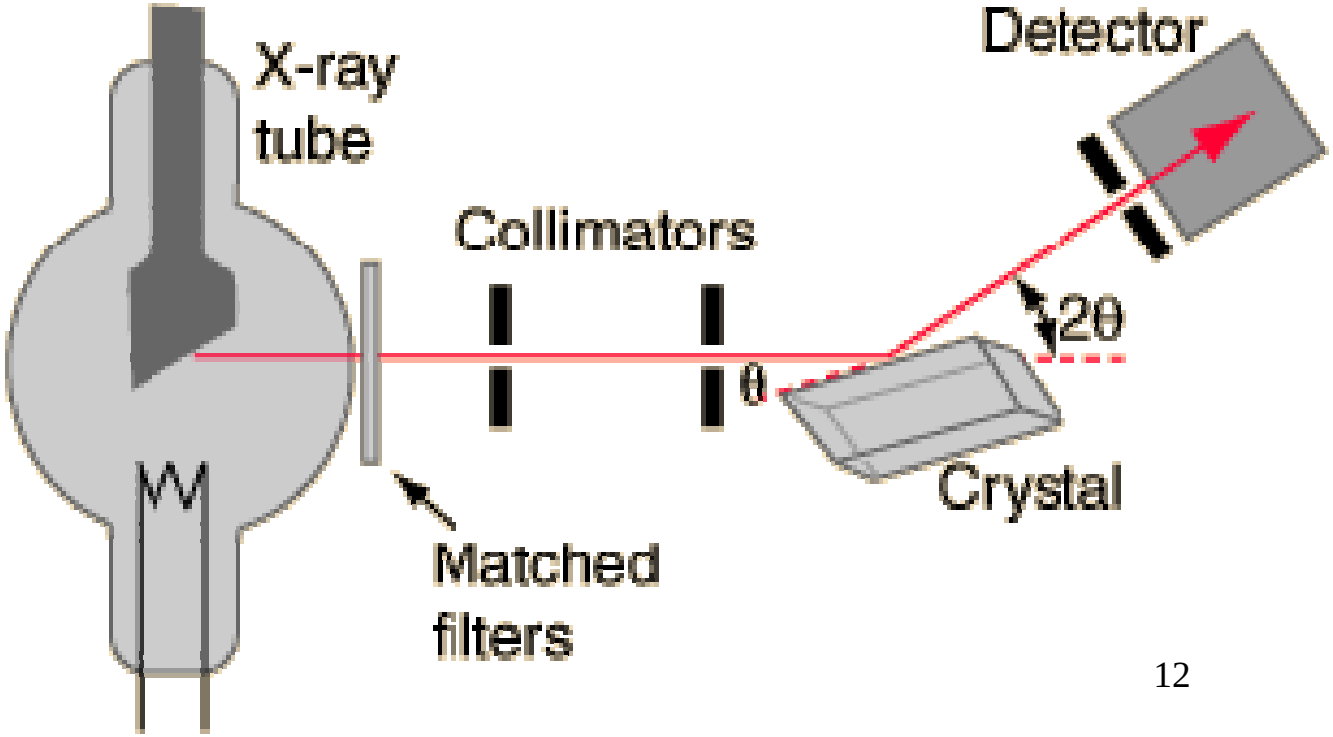


e: electrons exiting the cathode and accelerated to the anode,

Bragg Spectrometer



X-ray was the name given to the highly penetrating rays which emanated when high energy electrons struck a metal target



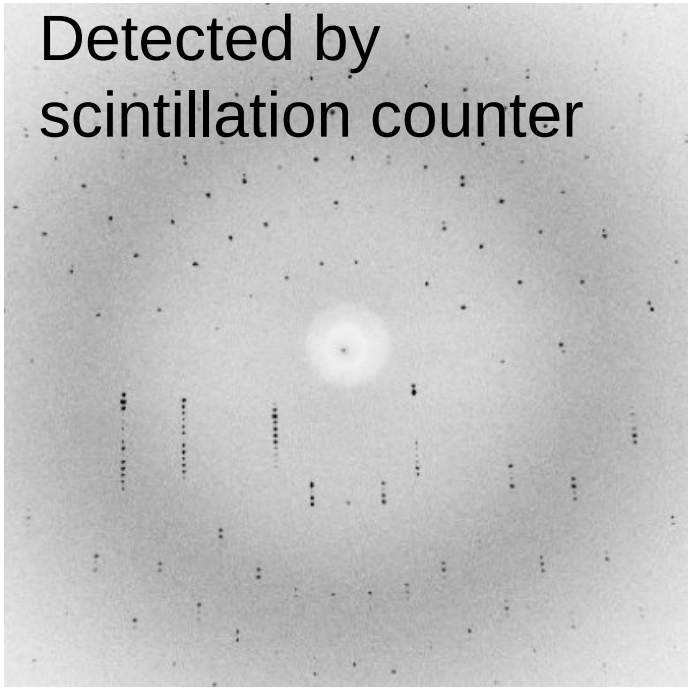
Why Electron Density?

- charge/mass: electrons >>atomic nuclei or protons
- electrons acceleration $\Rightarrow$ electromagnetic radiation

Why Crystals?

- Huge number of molecules in same orientation (diffraction pattern for $>10^{15}$ molecules)
- Sufficient intensity

Diffraction from a Real Crystal

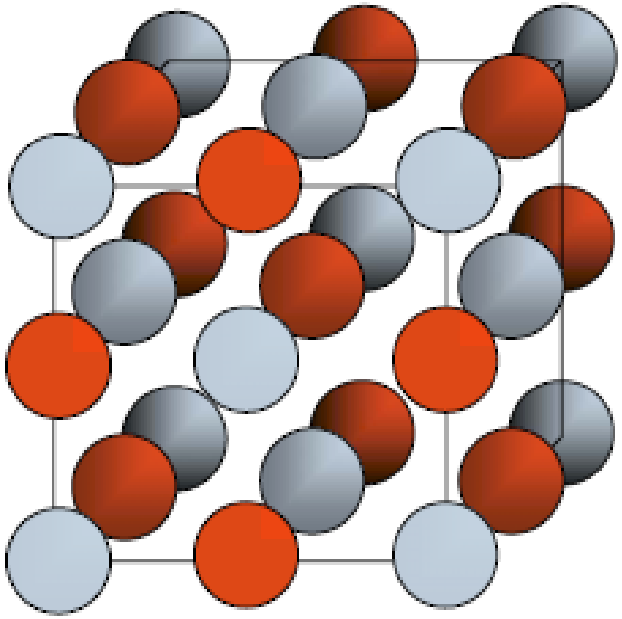


- Electrons diffract x-rays – the **amplitude of the diffracted x-ray is proportional to its number of electrons**.
e.g. C diffracts 6 times more strongly that H

- The diffracted waves recombine constructively if they are in phase and destructively if they are out of phase

- The way in which the diffracted waves recombine depends on the arrangement of the atoms.

Crystal lattice



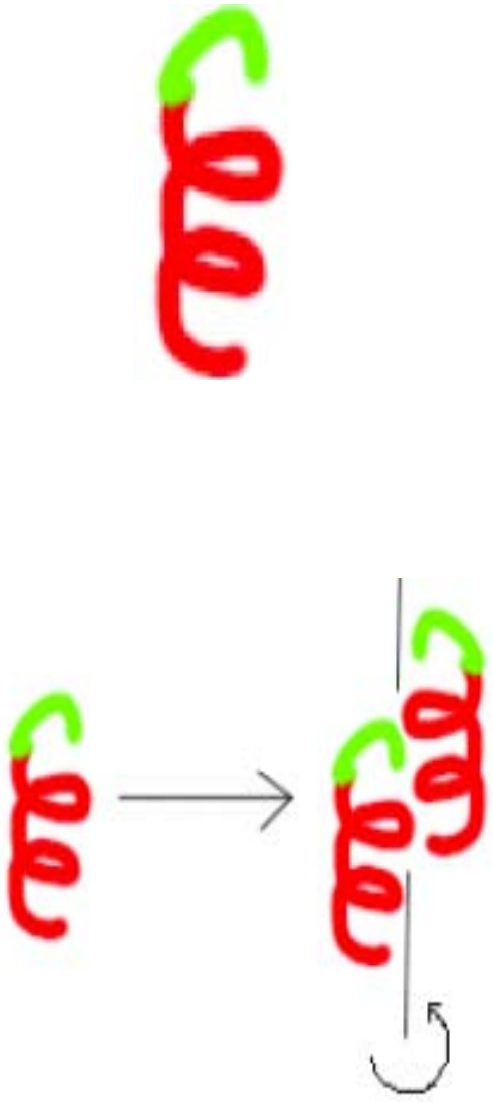
NaCl

- both types of atoms are charged
- ionic interaction

give pattern of electron density

How about biomolecule?

Motif v.s. Lattice

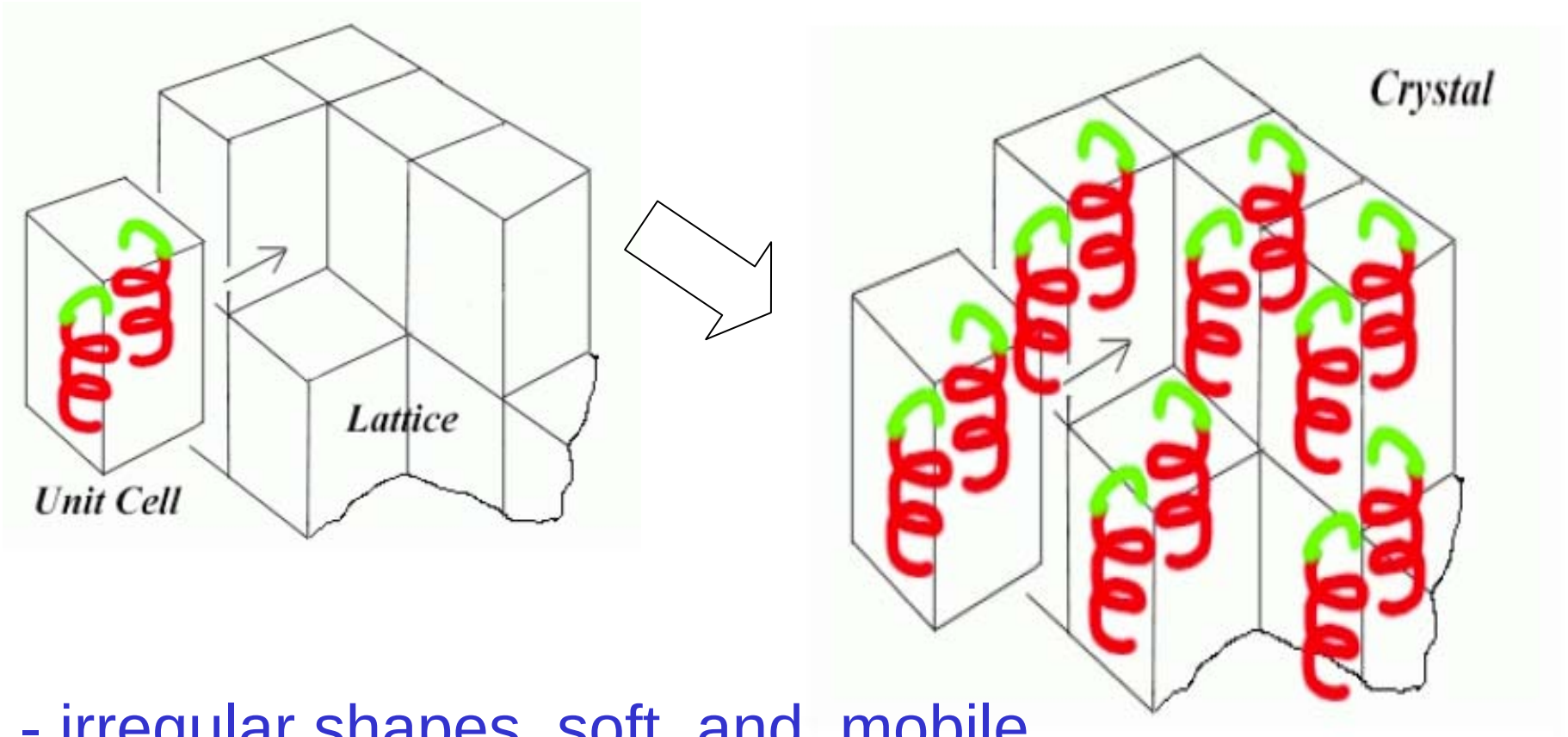


A **motif** can be a single atom, a small molecule, a protein or any combination.

We can simply repeat this motif in three dimensions.
This will result in a simple crystal, like the NaCl crystal.

Often, however, motifs can arrange in differently oriented copies (this of course doesn't hold for atoms!).

Formation of Crystal



- irregular shapes, soft, and mobile

- Homogeneity of the protein sample
- Crystallization methods
- Screening for crystals

Can Protein Crystals be Formed?

Usually >10 mg, 0.2 mm

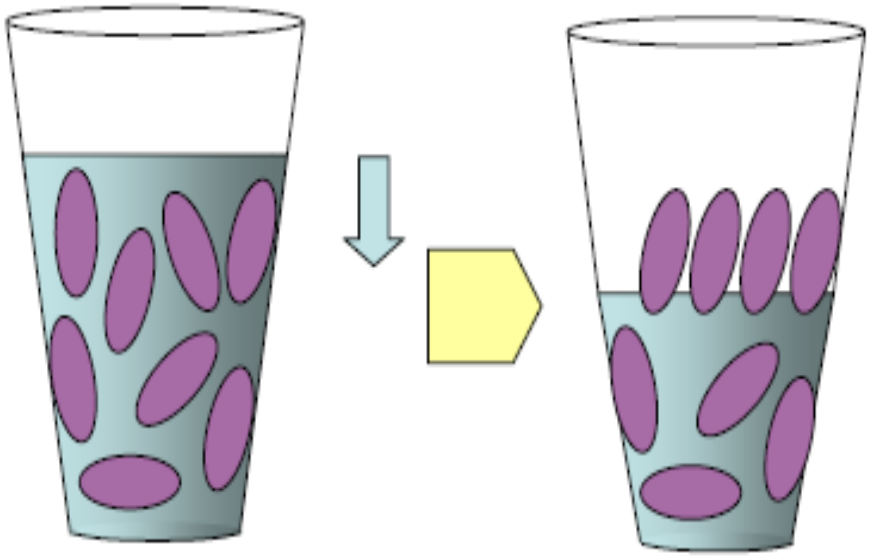
- Protein crystals contain on average **50% solvent**, mostly in large channels between the stacked molecules on the crystal. --→ **Like jelly**
- The interactions holding proteins together in a crystal are weak: **hydrogen bonds, salt bridges, and hydrophobic interactions**.

A crystal that is **0.5 millimeters** on each side contains around 1,000,000,000,000,000 (or 10^{15}) medium-sized protein molecules (50% water).

Can Protein Crystals be Formed?

To form protein crystals, we have to find conditions so that the proteins assemble in a periodic lattice.

This is done by having a highly concentrated solution of the protein and adding reagents to reduce the solubility.

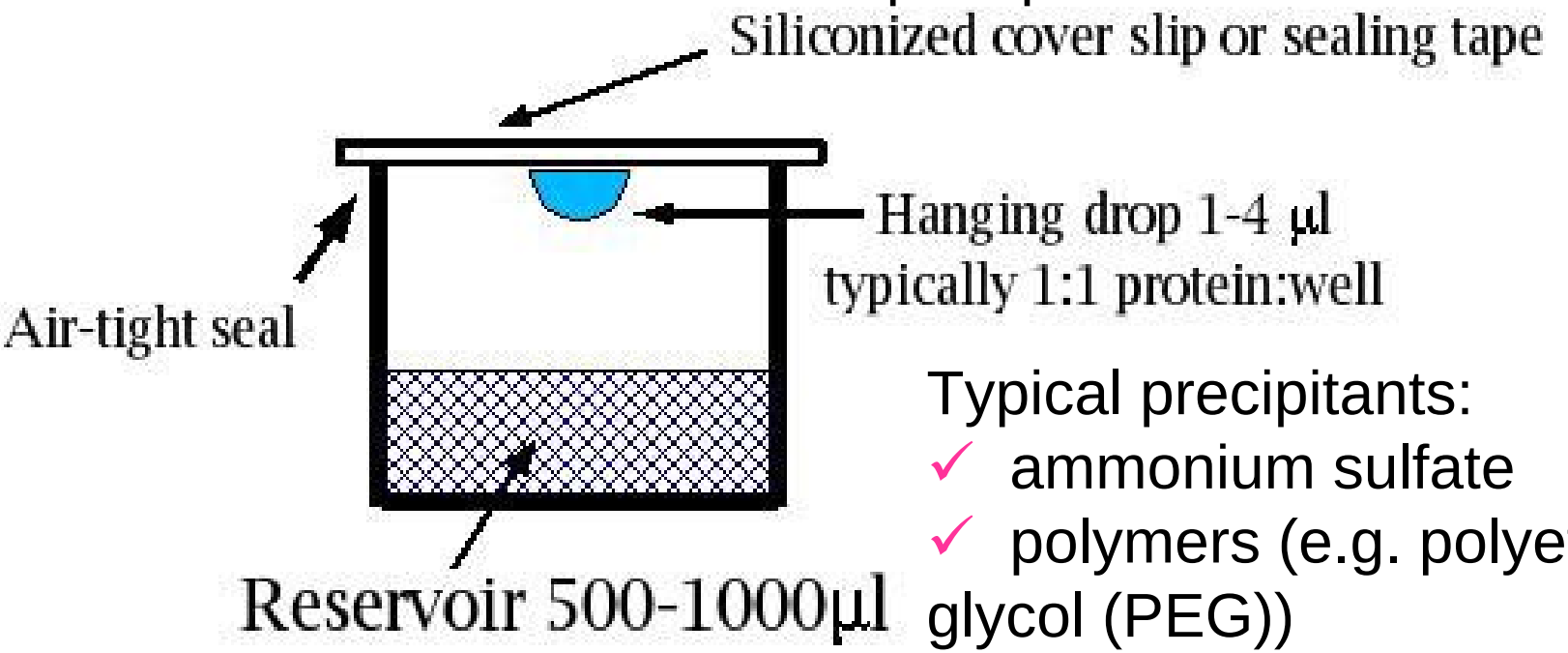


Growing Quality Crystals

Vapor Diffusion Method— hanging drop vapor diffusion

Low $[(NH4)_2SO_4]$

mixture of protein solution
and precipitant solution

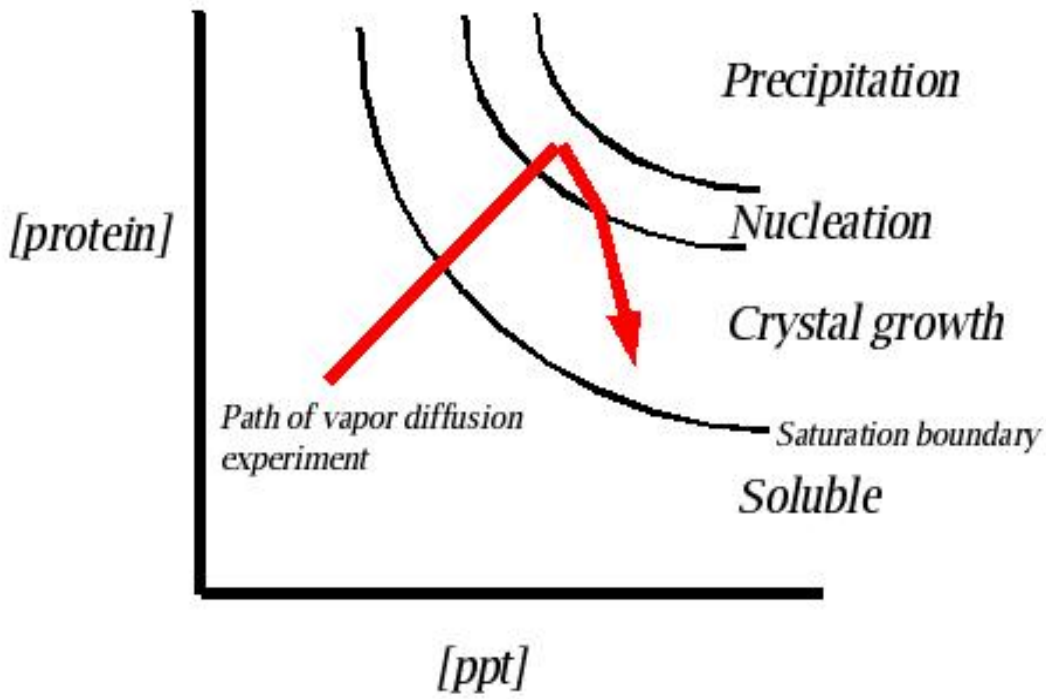


High $[(NH4)_2SO_4]$

Typical precipitants:

- ✓ ammonium sulfate
- ✓ polymers (e.g. polyethylene glycol (PEG))
- ✓ polyalcohols
- ✓ organic solvents
- ✓ saccharides

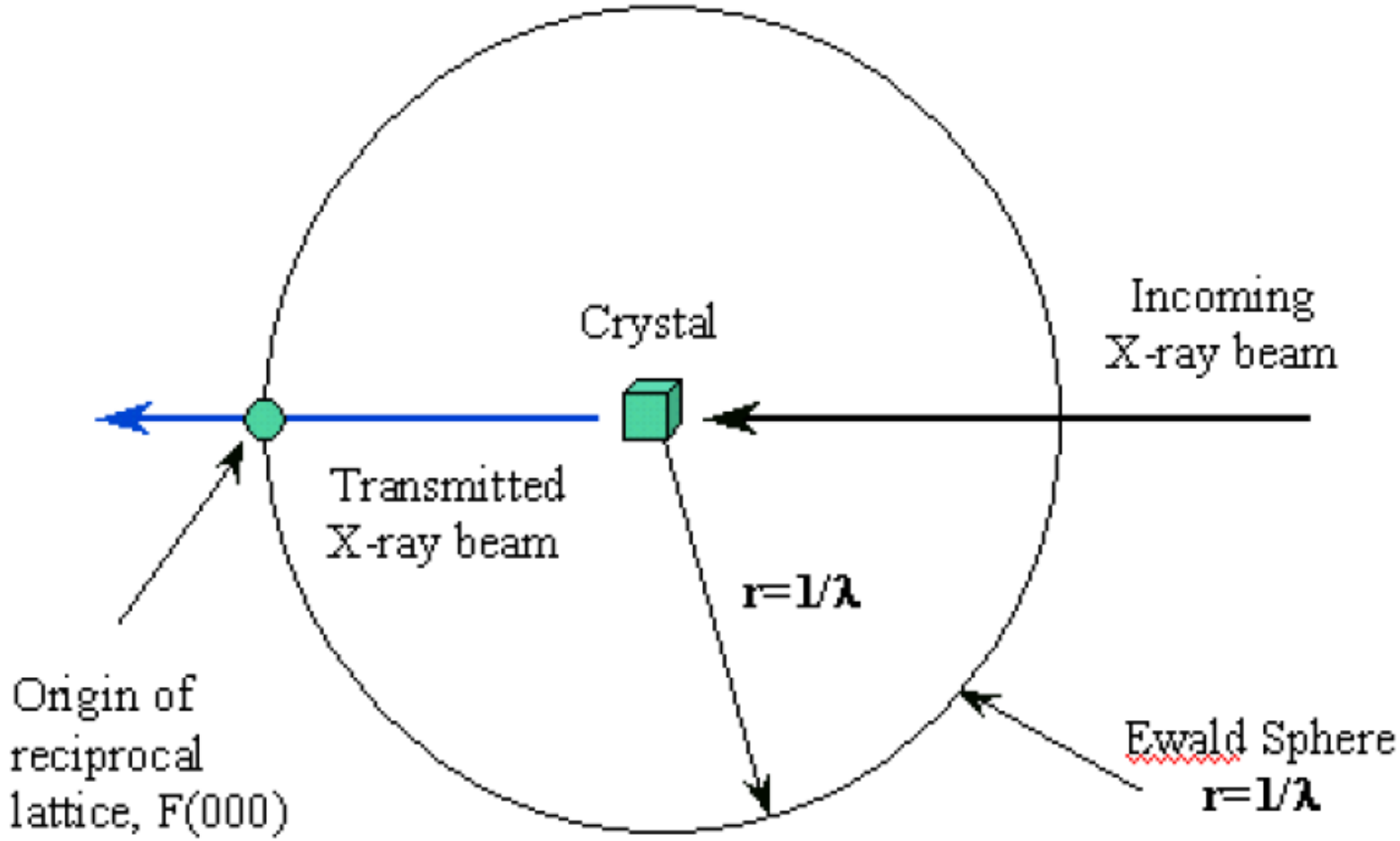
Formation of Crystal

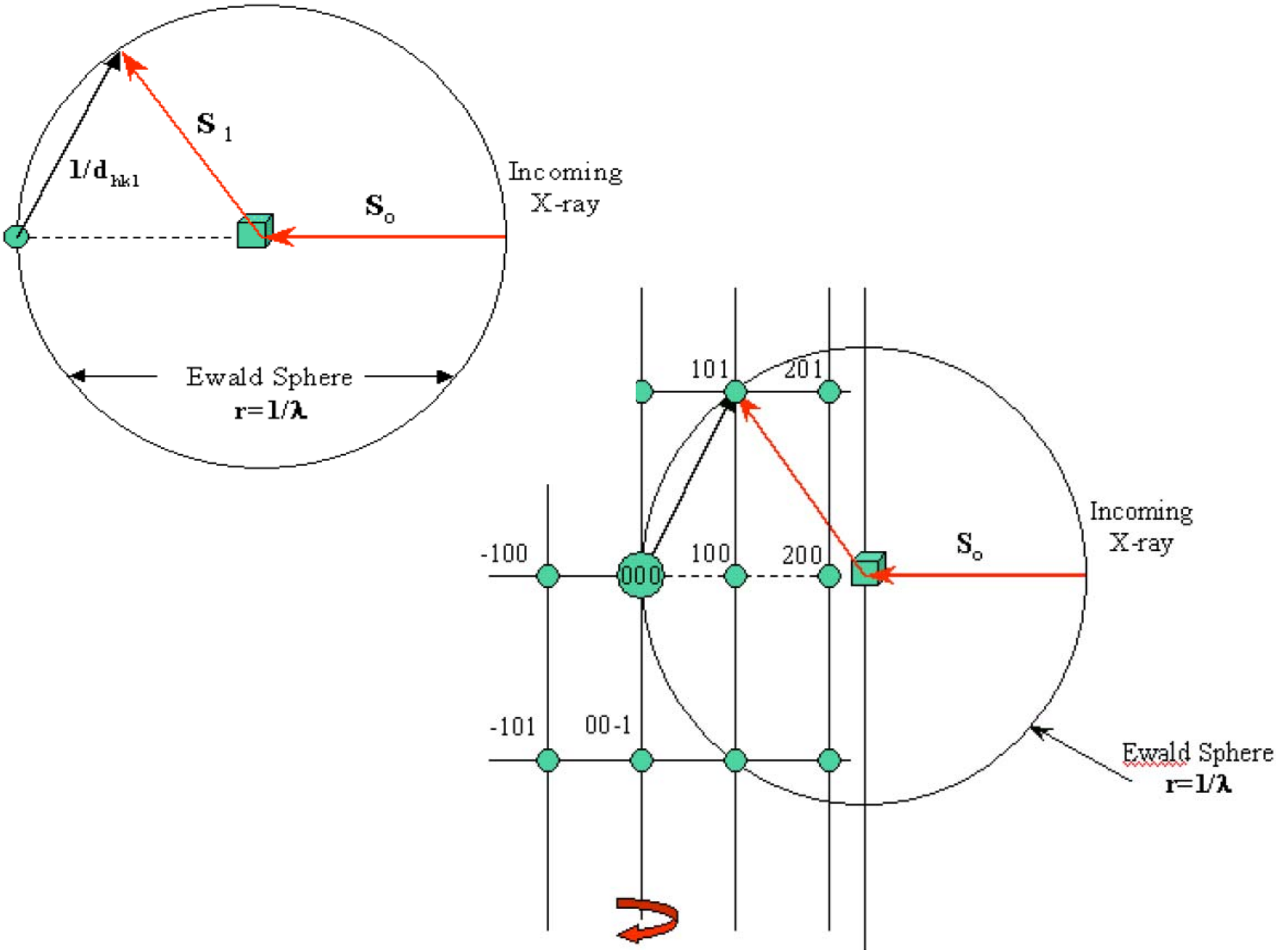


All crystal growth and nucleation occurs beyond the saturation point.

Formation of crystal can be done by highly concentrated solution of the protein and adding reagents to **reduce the solubility.**

First we need to construct a reciprocal lattice. This is done using the Ewald construction.

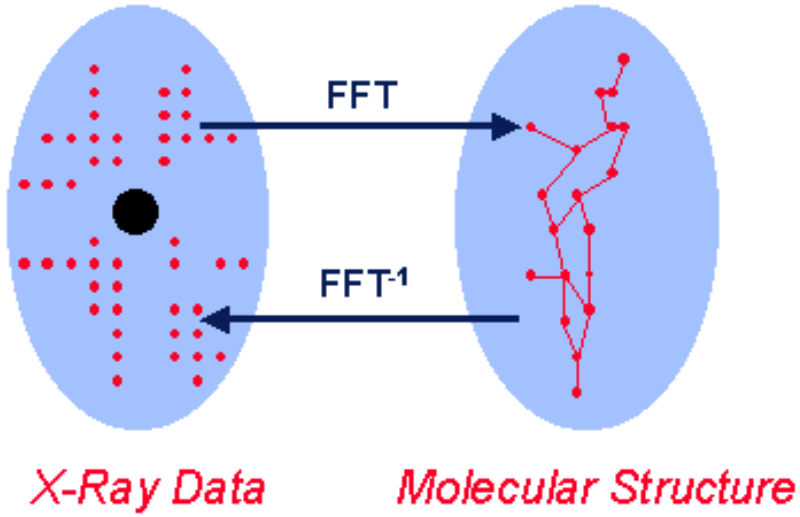



S 1
1/d hkl
S o
Incoming
X-ray
Ewald Sphere
r=1/λ
101
201
-100
000
100
200
S o
Incoming
X-ray
-101
00-1
Ewald Sphere
r=1/λ

Fourier Theory (mathematical lens)

- Diffraction pattern related to object
- Mathematical operation called *Fourier Transform*
- Can be inverted to give pattern of electron density
- Requires amplitude and phase of diffracted waves

The Phase Problem



The Phase Problem

- What we need:
  - Phase and Amplitude of diffracted waves
- What we have:
  - Number of X-ray photons in each spot
- What we can get:
  - Number of Photons $\Rightarrow$ Intensity $\Rightarrow$ Amplitude^2
- What we miss:
  - Phase has been lost for different spots

Finding the Phase

The phases we obtain from the Fourier transformations are **not referenced to a specific point in space**….

- Perturb the system
  - – Isomorphous replacement

(Position of few heavy atoms)

✓ Single-wavelength anomalous dispersion (SAD)

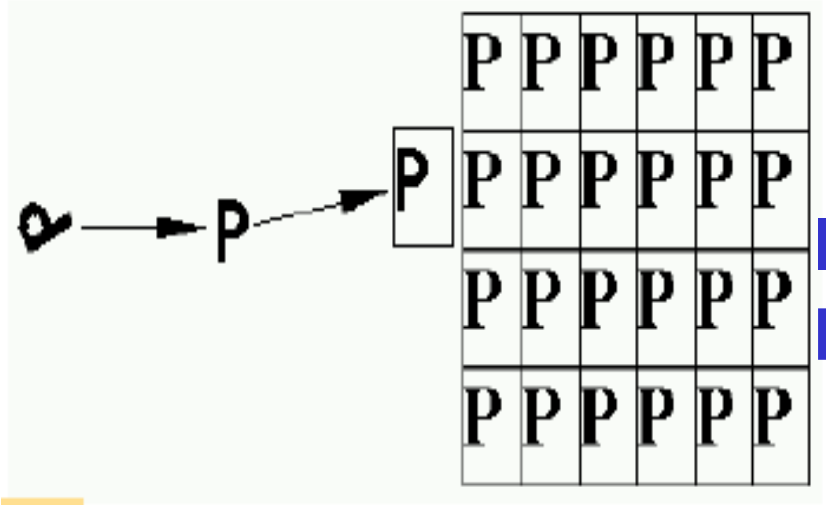
✓ Multiple-wavelength anomalous dispersion (MAD)

(Anomalous scatterers)

- Guess the phase
  - – Molecular Replacement

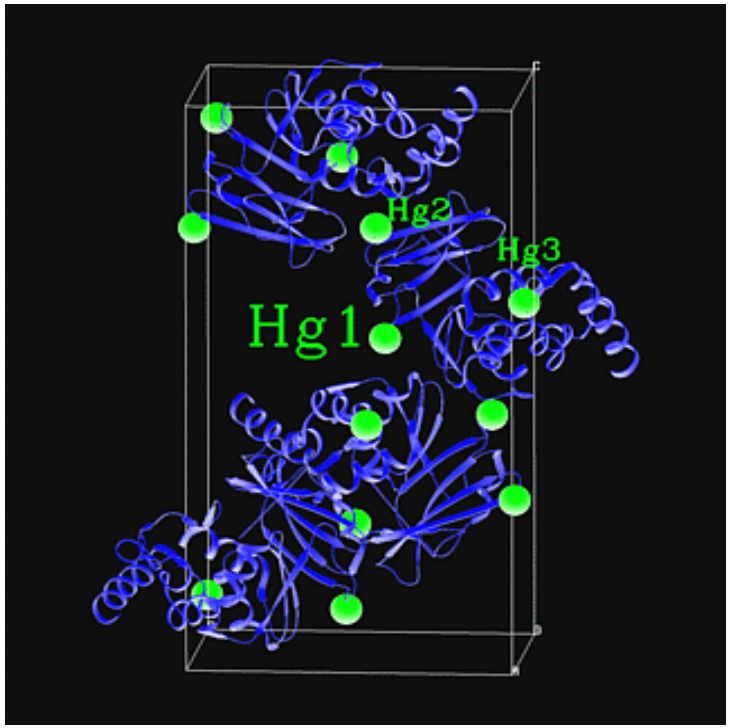
Based on known structure

- Good model
- 6-D problem
- Mathematical Function



Finding Phase

Isomorphous replacement Position of Few Heavy Atoms

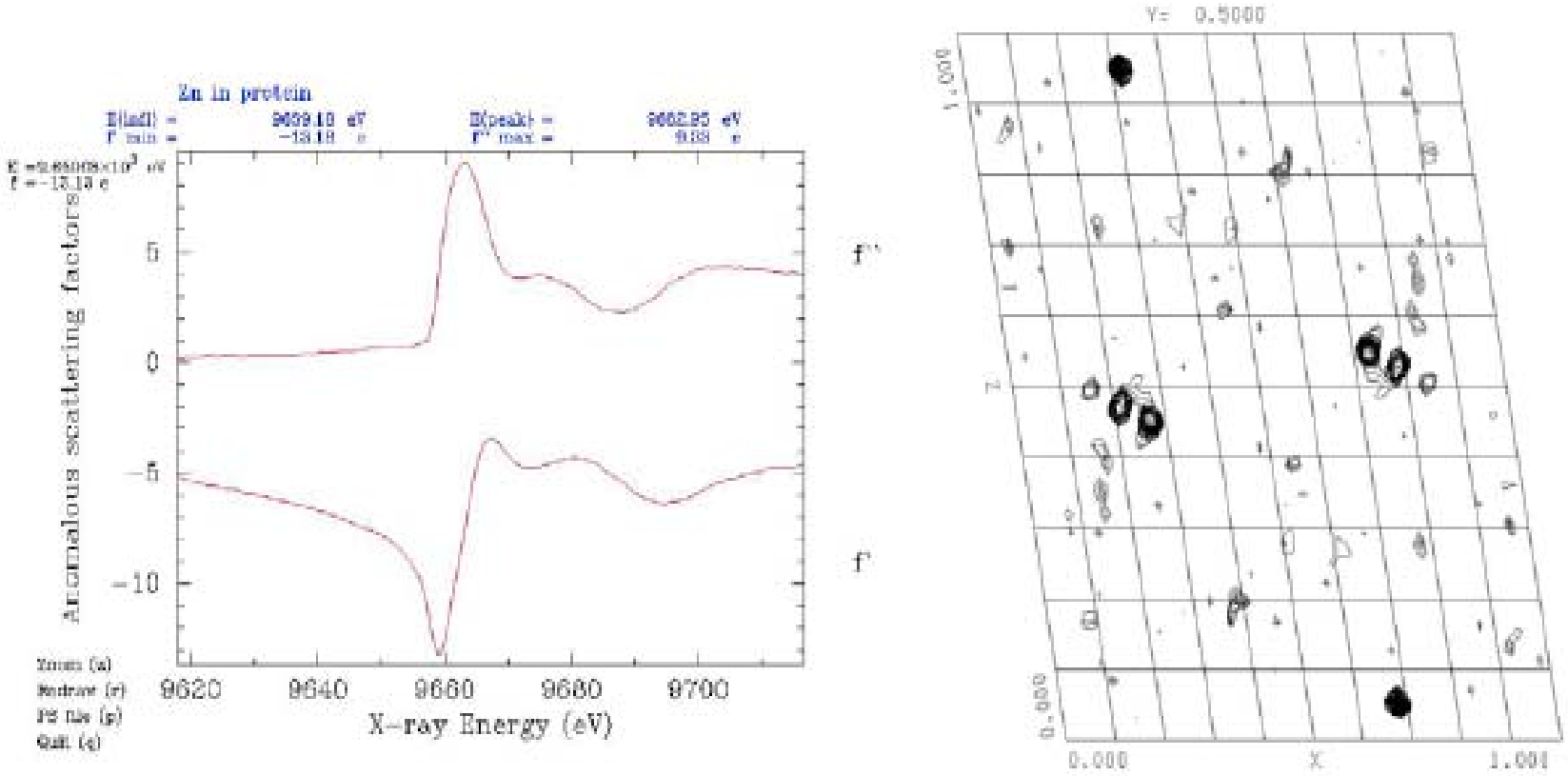


To solve the problem, we introduce a "heavy atom" such as uranium (U) or mercury (Hg) at specific sites in the protein – for example, by soaking the protein crystals. The diffraction patterns for these "heavy atoms" are well known, so they can be used as references.

Multiple-wavelength anomalous dispersion (MAD)

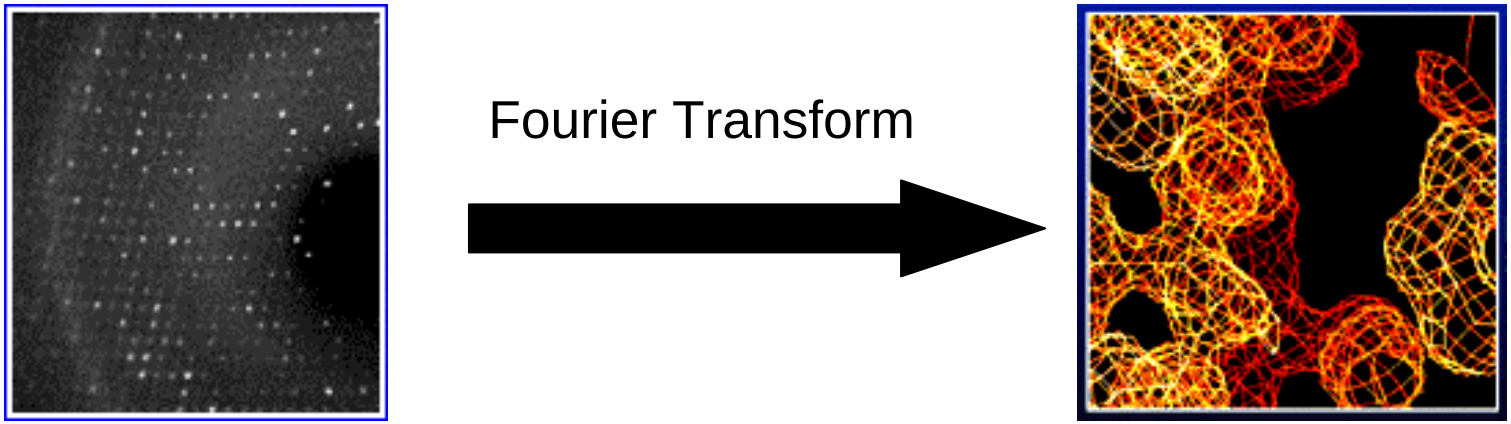
a technique used in X-ray crystallography that facilitates the determination of the structure of proteins or other biological macromolecules (e.g. DNA) by allowing the solution of the phase problem.[1] This is possible if the structure contains one or more atoms that cause significant anomalous scattering from incoming X-rays at the wavelength used for the diffraction experiment. Atoms in proteins which are suitable for this purpose are sulfur or heavier atoms, for example metal ions in metalloproteins. The most commonly used atom for phase determination via MAD, however, is selenium, since it is usually possible to replace the natural sulfur containing amino acid methionine by selenomethionine. The use of the MAD technique in an experiment utilizing different wavelengths of X-rays generated at a synchrotron relieves the crystallographer from the traditional method of phase determination via **Multiple isomorphous replacement** (MIR), which involves the preparation of heavy atom derivatives in a trial-and-error approach.

Multiple-wavelength anomalous dispersion + Zn atom



圖六 X-光能量對應蛋白質單晶中重金屬鋅原子之異常振幅因子(Δf' 和Δf")(左)；異常繞射差異的帕特森圖(Anomalous Difference Patterson Map)可決定鋅原子位置進而計算出蛋白質晶體的結構因子相位角(右)

From Diffraction to Electron Density Map



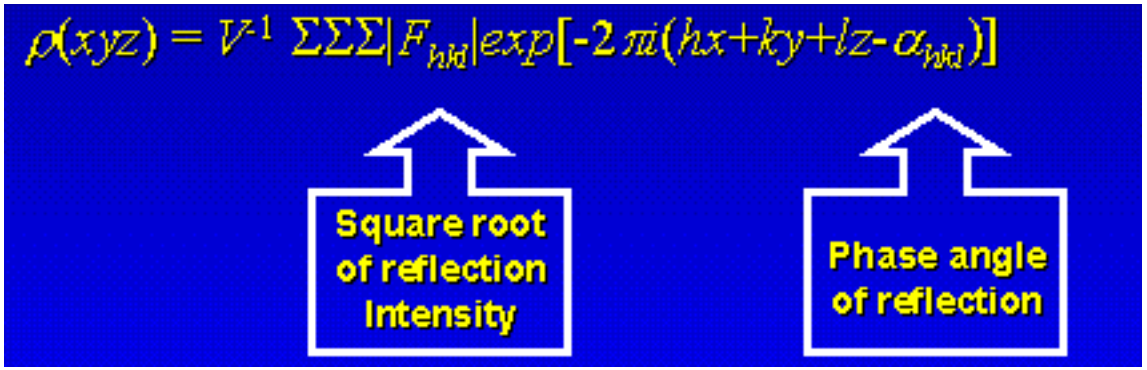
To get from the diffraction pattern to the electron density, you have to use a Fourier Transform.

$$\rho(xyz) = V^{-1} \sum\sum\sum |F_{hkl}| exp[-2\pi i(hx+ky+lz-\alpha_{hkl})]$$

Square root of reflection Intensity

Phase angle of reflection

Solving the Phase Problem



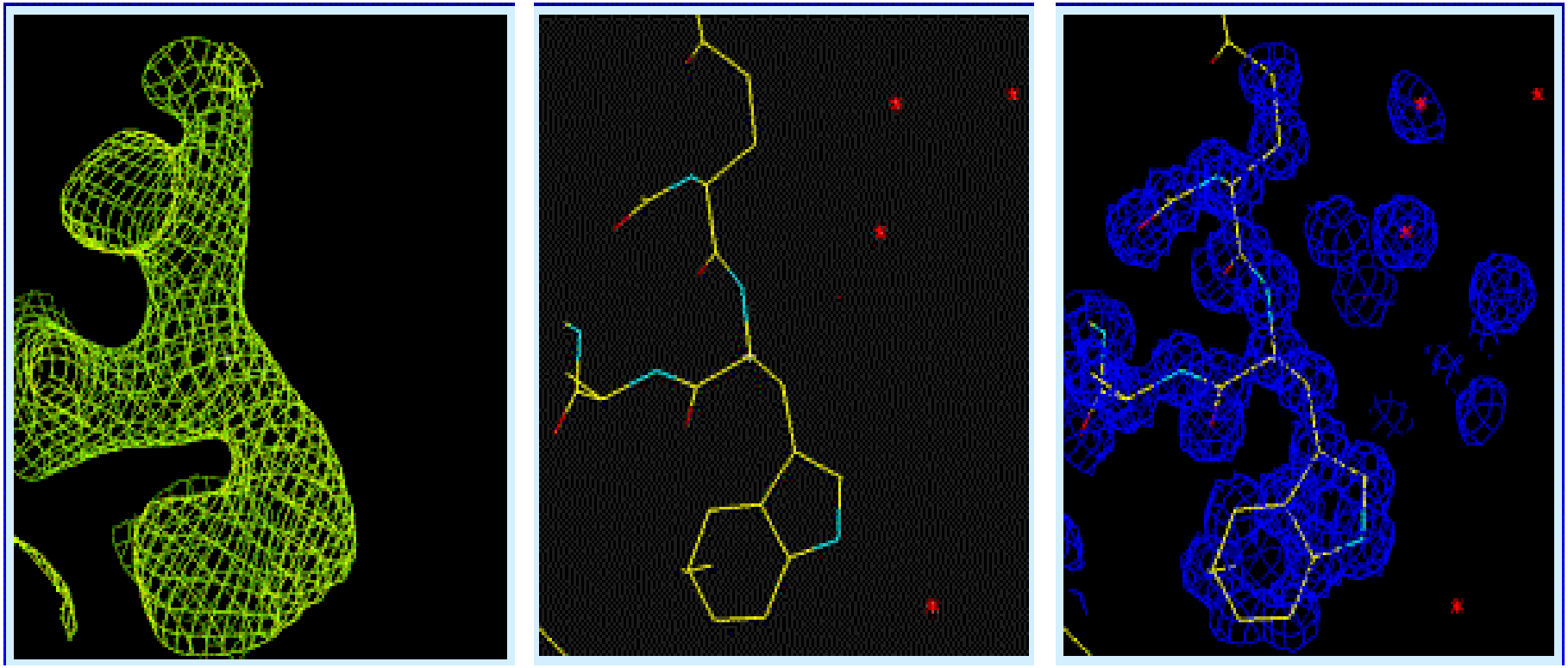
- The diffraction data does not give the phase angle that is needed to calculate the electron density map.
- Have to get the phase angle through other methods.

Isomorphous Replacement: Insert a heavy metal atom into crystal protein, and locate in diffraction pattern and in the cell. Use the location of metal ion to find the phase angle for the other protein atoms.

- Requirements:
  - Add atom with same unit cell size.
  - Cannot disturb protein structure.

Often use Hg, Pt, Au.

Electron Density Map Amplitude + Correct phase



Once you have an electron density map, you can begin to fit models to it.

Fitting and Refinements

- Fitting Model and Experiment
  - Computer graphic Programs
  - Phase problem $\Rightarrow$ poor initial model
- Improvement process: Refinement
  - Atomic model adjusted to improve with measured diffraction data
  - standard crystallographic R-factor

$$R = \frac{\sum_{hkl} (|F_{obs}| - |F_{calc}|)}{\sum_{hkl} |F_{obs}|}$$

0.6-VERY BAD

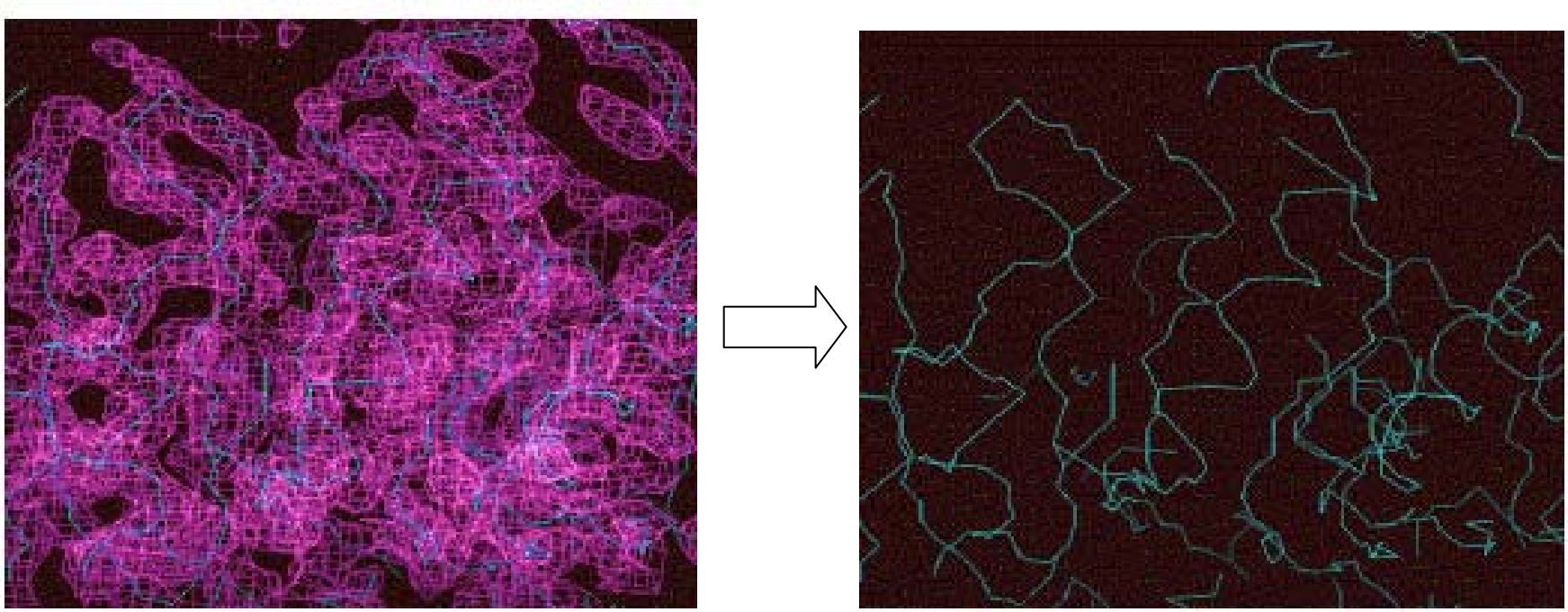
0.5 -BAD

0.4-Recoverable

0.2-Good for Protein

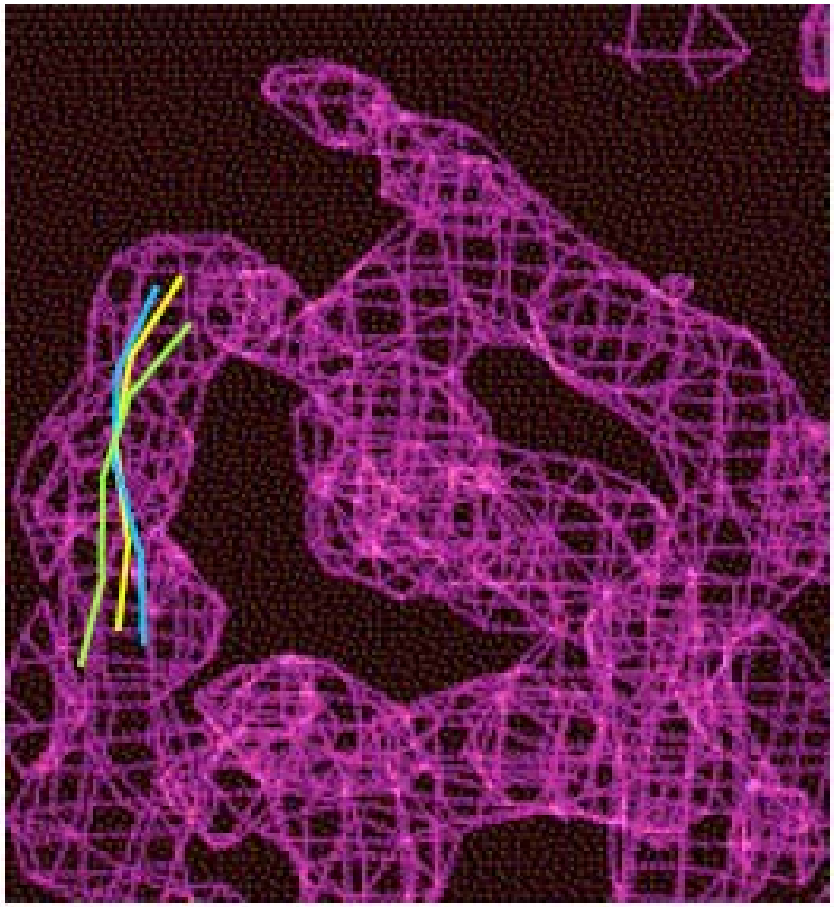
0.05-Good for small organic models

0-PERFECT FIT



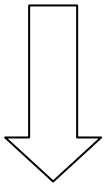
Given that you know the sequence of the protein, you can generate a model and fit it within the density (using a number of computer programs).
First, you start with the backbone.
Then you can fit the side-chains

Poor Initial Model



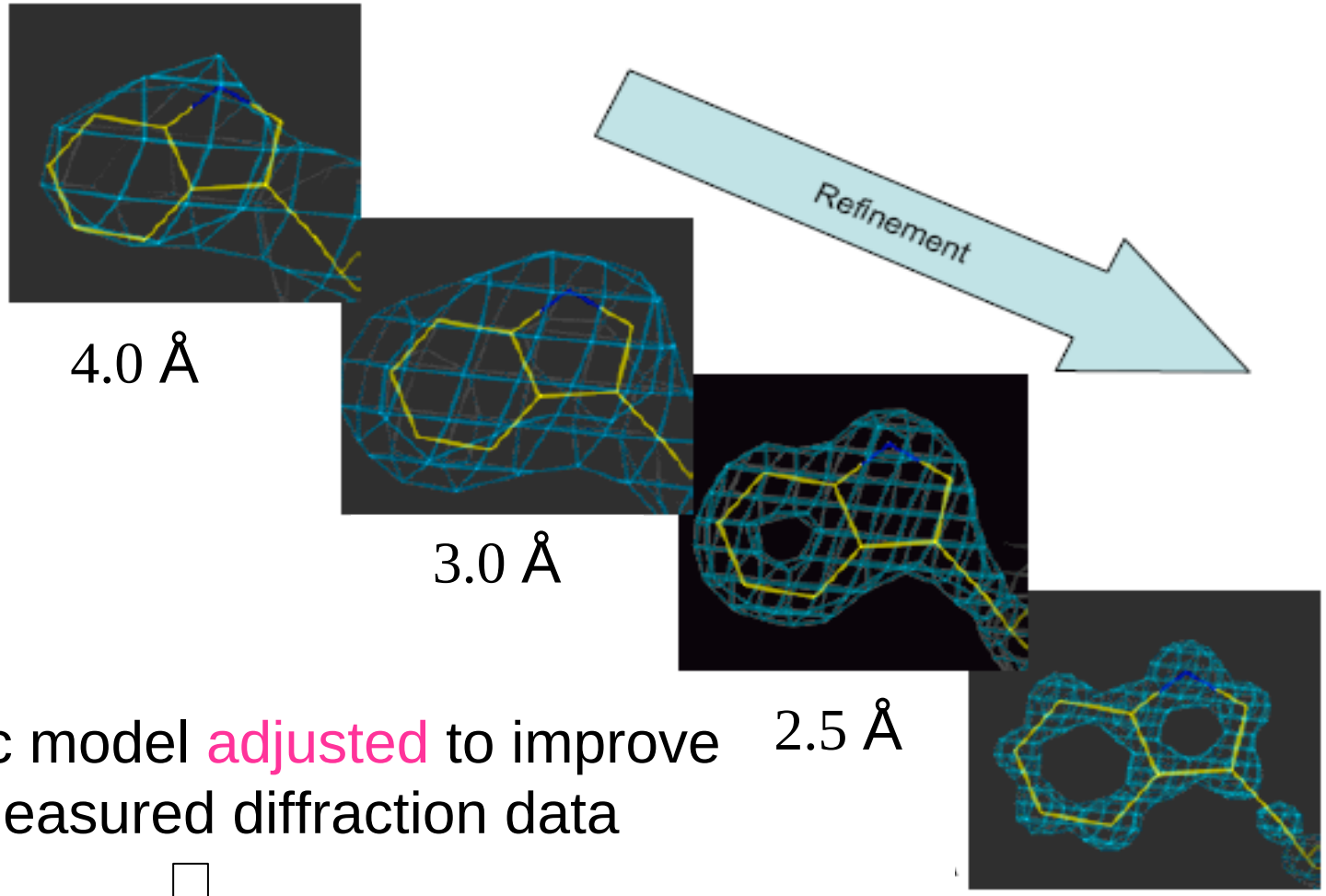
At this stage, the structure is poorly resolved.

There may be a number of possible solutions.



Since we want atomic resolution (i.e. on the order of 1 Å), we need to *refine* the structure.

Refinement of Structure

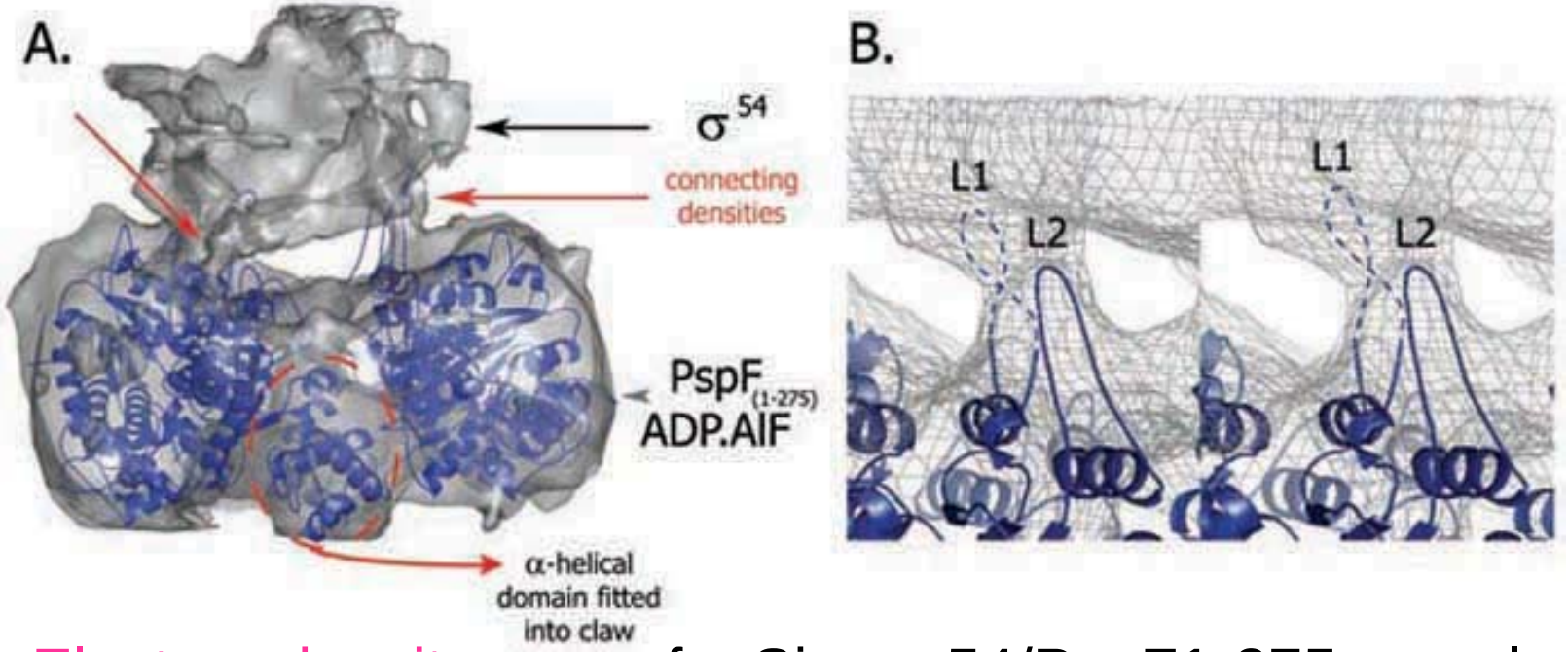


Atomic model adjusted to improve with measured diffraction data

⇩

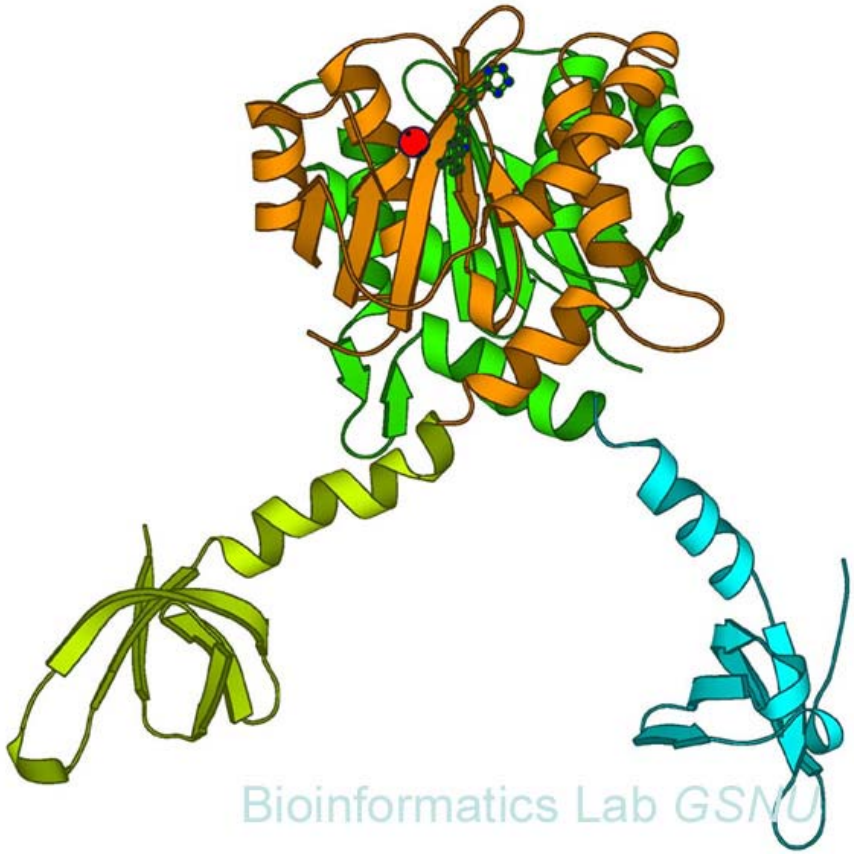
Require a lot of time !!!

Phage Shock Protein F (PspF)



Electron density map of a Sigma 54/PspF1-275 complex in the presence of ADP.AlFx, an ATP hydrolysis transition state analogue. The crystal structure of PspF1-275 has been fitted into the density. The model shows that PspF contacts Sigma 54 through a flexible loop (L1 loop) at the point of ATP hydrolysis. This contact is crucial for the transfer of energy to the closed promoter complex.

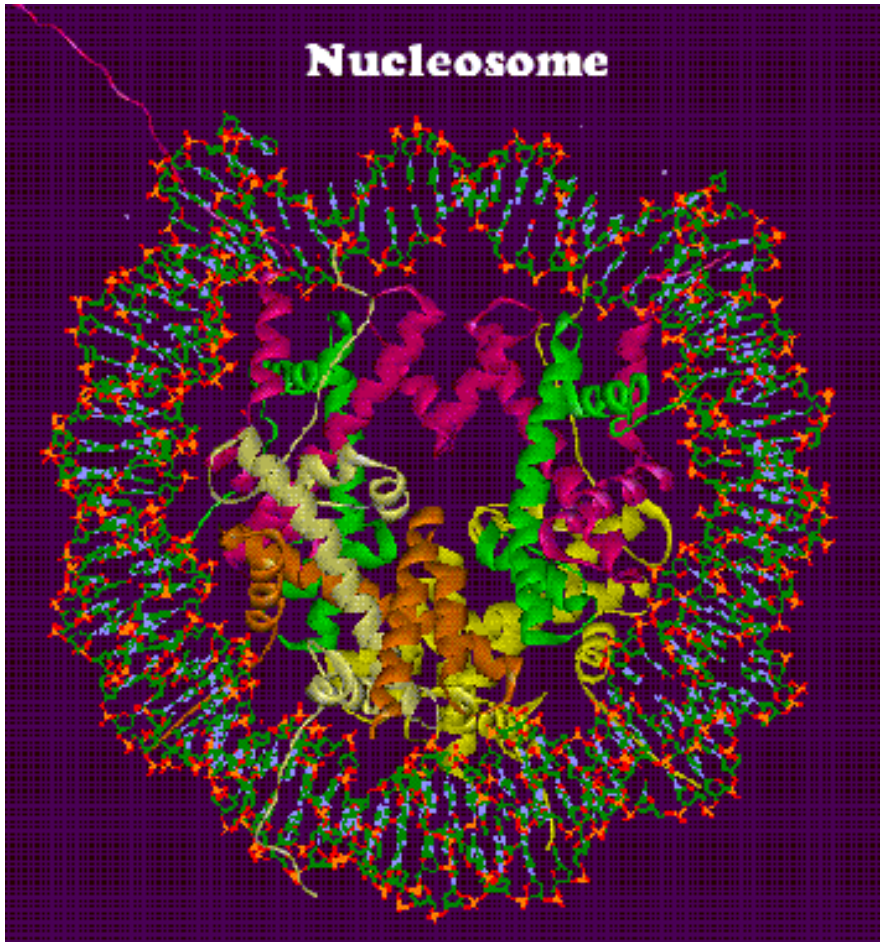
Examples in Biomedical Applications



HIV1-IN (Human Immunodeficiency Virus 1-INtegrase)

E. coli Leucyl-tRNA Synthetase (LeuRS)

Protein-protein interaction and protein complex



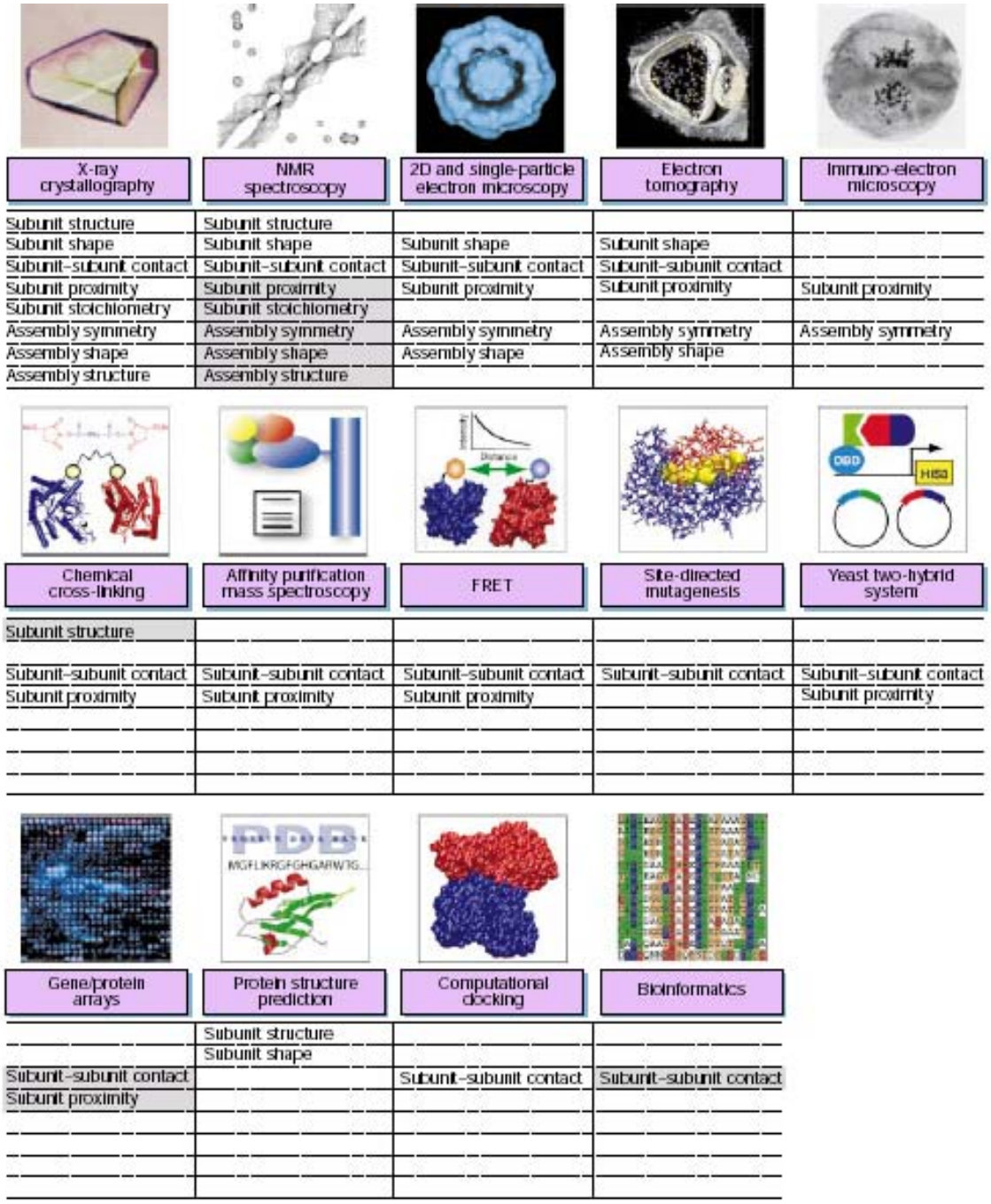
A chromosome contains five types of histones: H1 (or H5), H2A, H2B, H3 and H4. H1 and its homologous protein H5 are involved in higher-order structures. The other four types of histones associate with DNA to form nucleosomes. H1 (or H5) has about 220 residues. Other types of histones are smaller, each consisting of 100-150 residues.

Journal of Biochemistry and Molecular Biology, Vol. 37, No. 1, January 2004, pp. 45-52

Some limitations

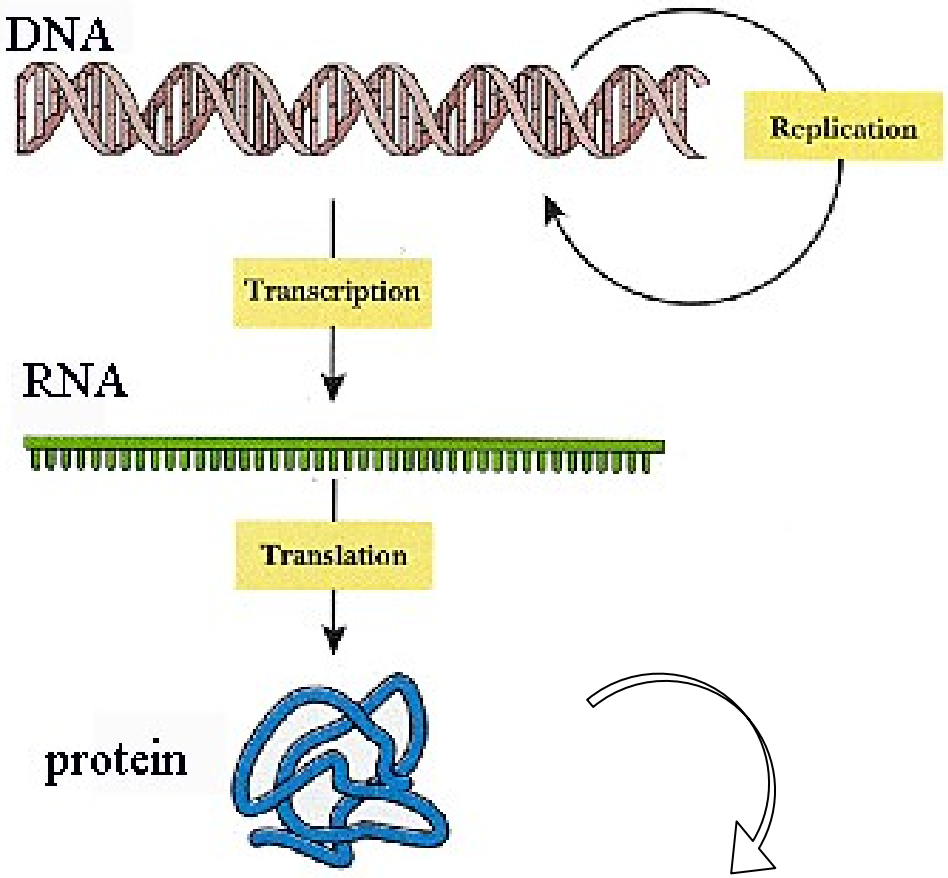
- Resolution
- Cannot resolve the positions of hydrogen atoms
- Cannot reliably distinguish nitrogen from oxygen from carbon
- Chemical identity of terminal side-chain atoms is uncertain for Asp, Gln and Thr

Structural Proteomics



Nature,422, 216 (2003)

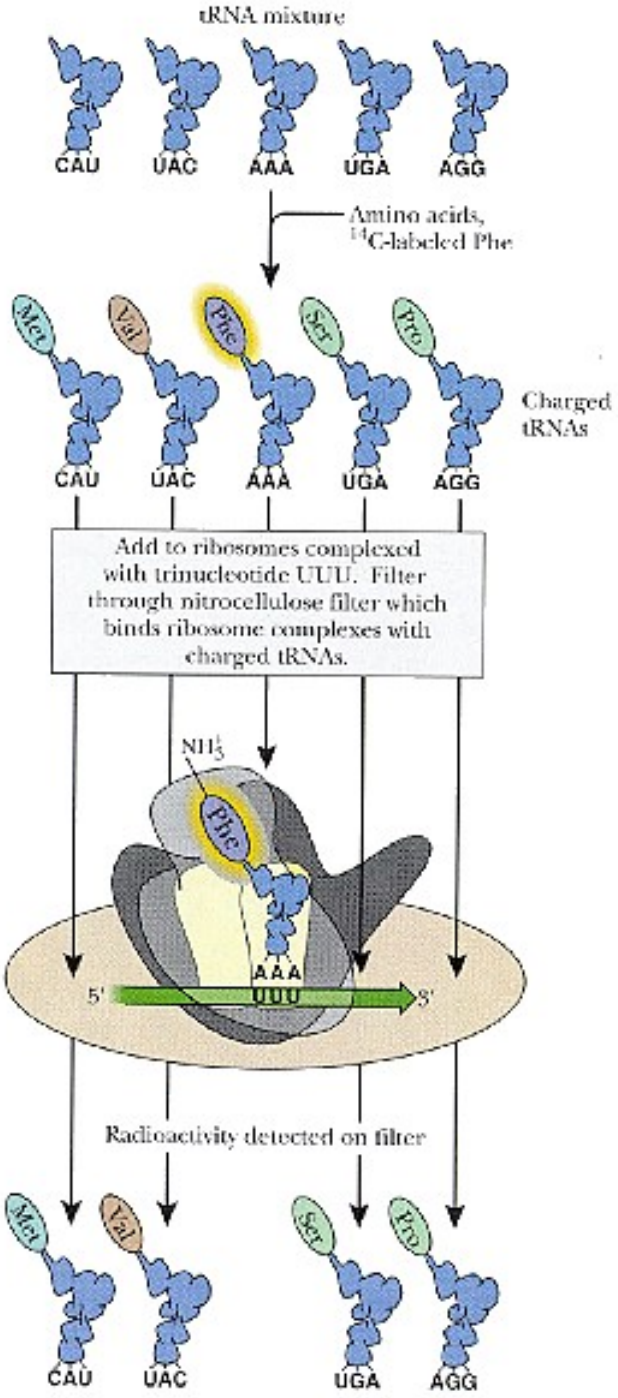
Role of the Ribosome



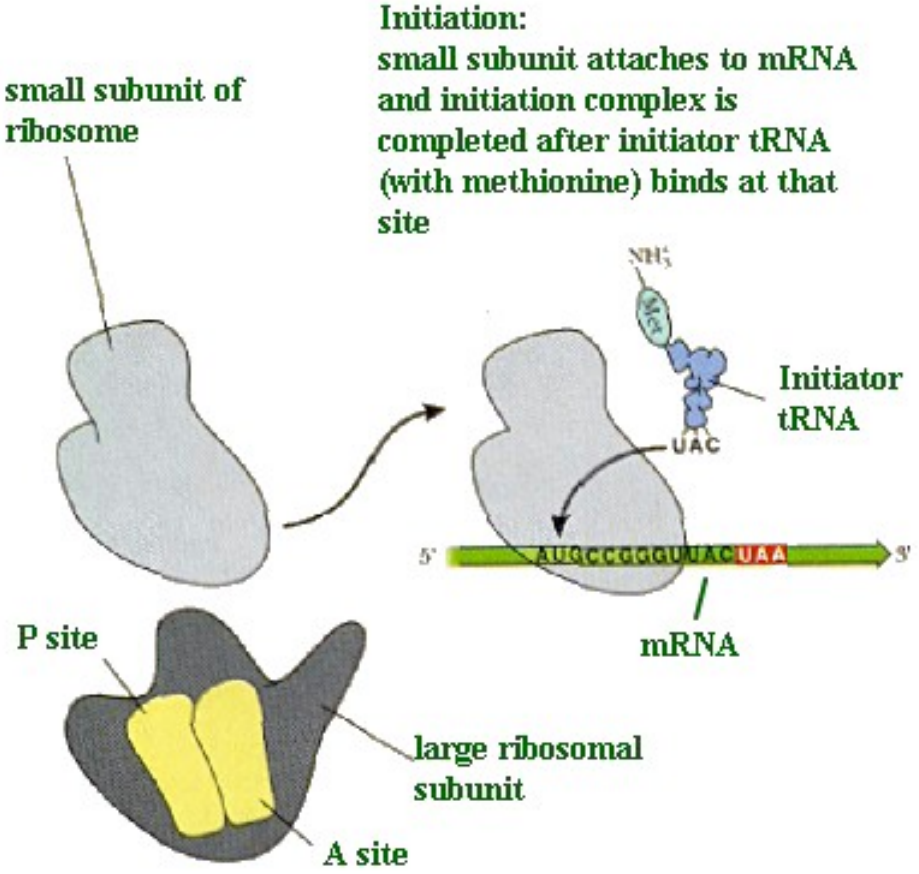
Ribosomes are tiny organelles responsible for protein synthesis in the cell. They receive genetic information from messenger RNA molecules, which are copies of the gene sequence, and use this to specify the assembly of amino acids, brought by transfer RNA

clusters of ribosomes, called "polyribosomes".

cellbio.utmb.edu/cellbio/ribosome.htm

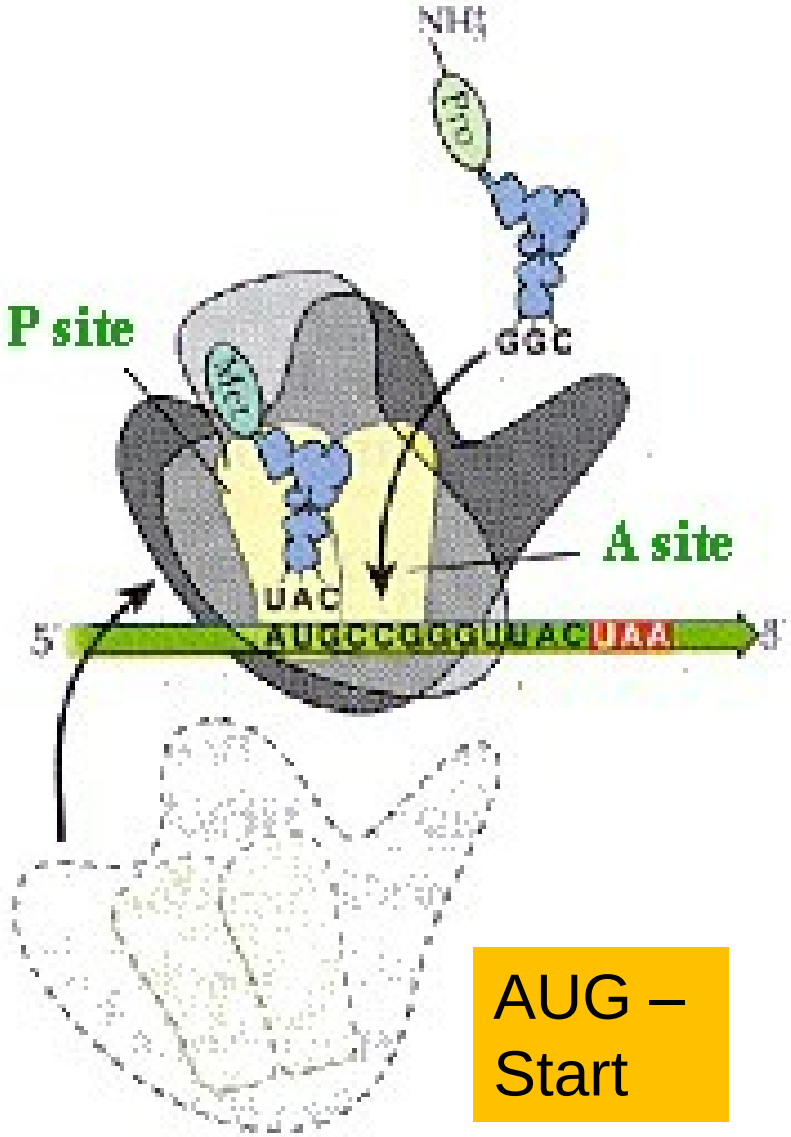

tRNA mixture
CAU
UAC
AAA
UGA
AGG
Amino acids, 14C-labeled Phe
Met
Val
Phe
Ser
Pro
Charged tRNAs
Add to ribosomes complexed with trinucleotide UUU. Filter through nitrocellulose filter which binds ribosome complexes with charged tRNAs.
NH3+
5'
3'
Radioactivity detected on filter

Initiation


Initiation:
small subunit attaches to mRNA and initiation complex is completed after initiator tRNA (with methionine) binds at that site
small subunit of ribosome
Met
UAC
Initiator tRNA
5'
AUGCCGGGUUACUAA
3'
mRNA
P site
large ribosomal subunit
A site

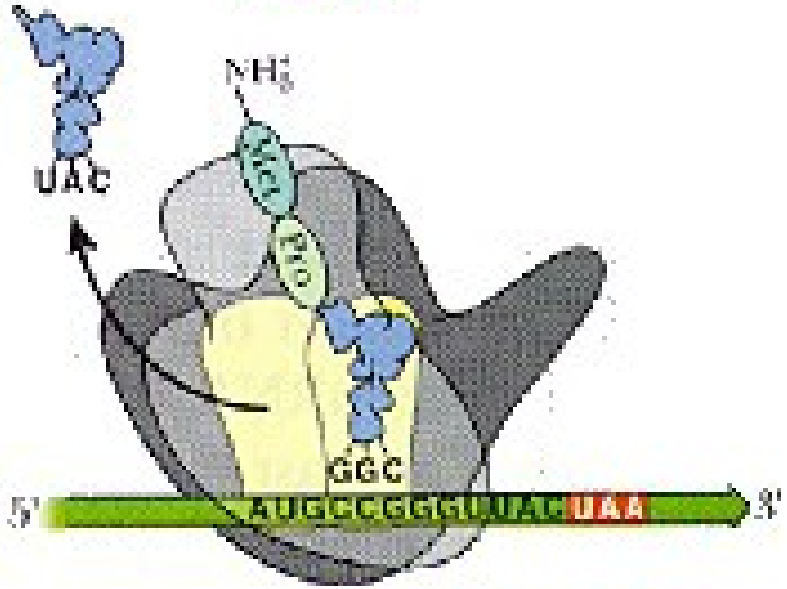
Elongation

large ribosome binds and
then the next tRNA binds at A site



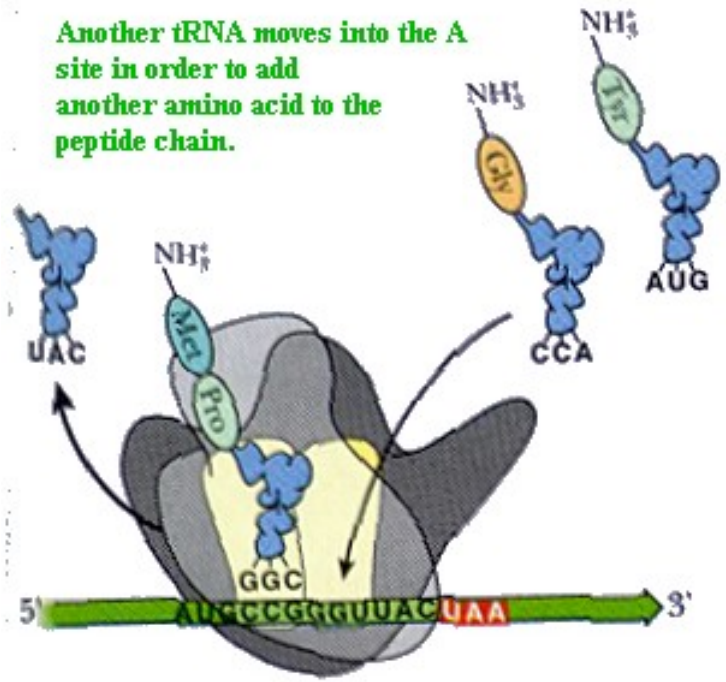
AUG – Start

peptide bond forms;
complex then moves
to P site

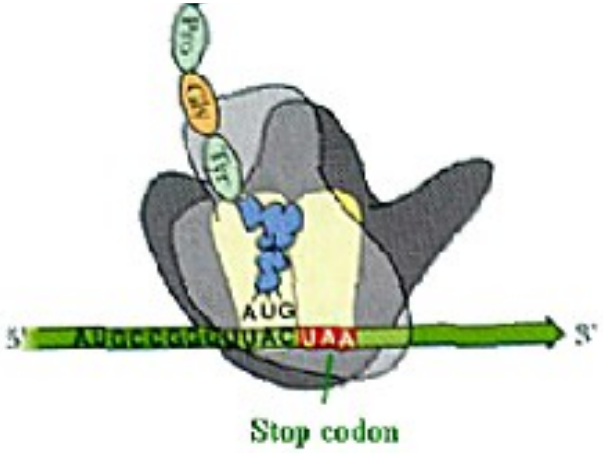


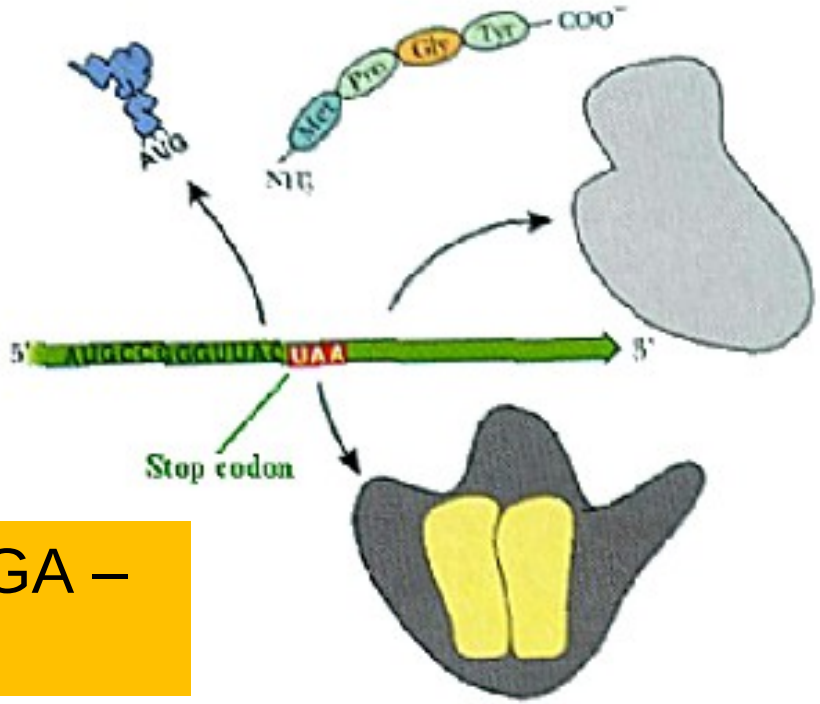
ribosome moves
from 5' to 3'

End of Translation



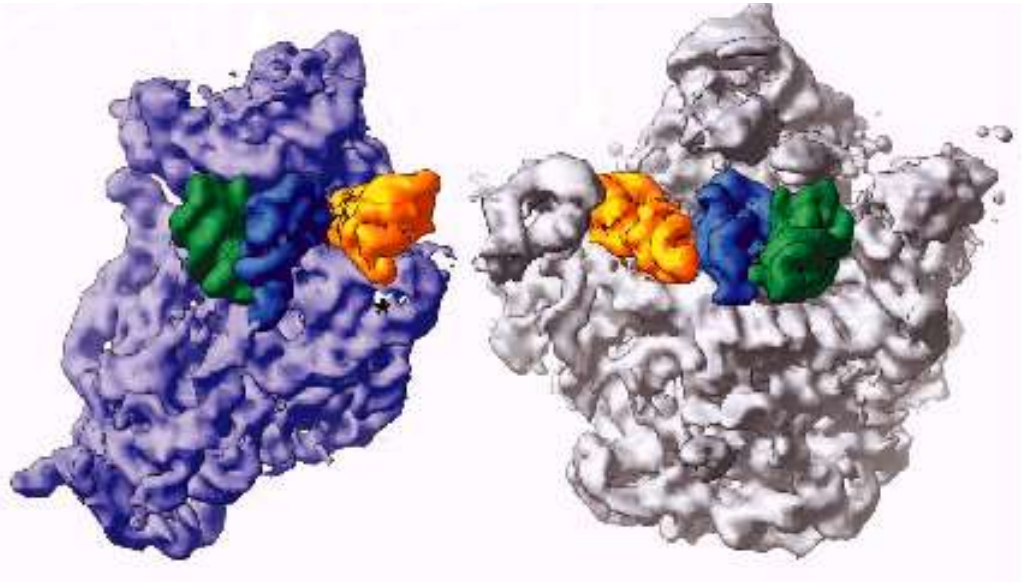
unters a stop codon (shown as the red triplet), there
d the ribosome separates and leaves the mRNA.





UAA, UAG, UGA – Stop

Structure of Ribosome

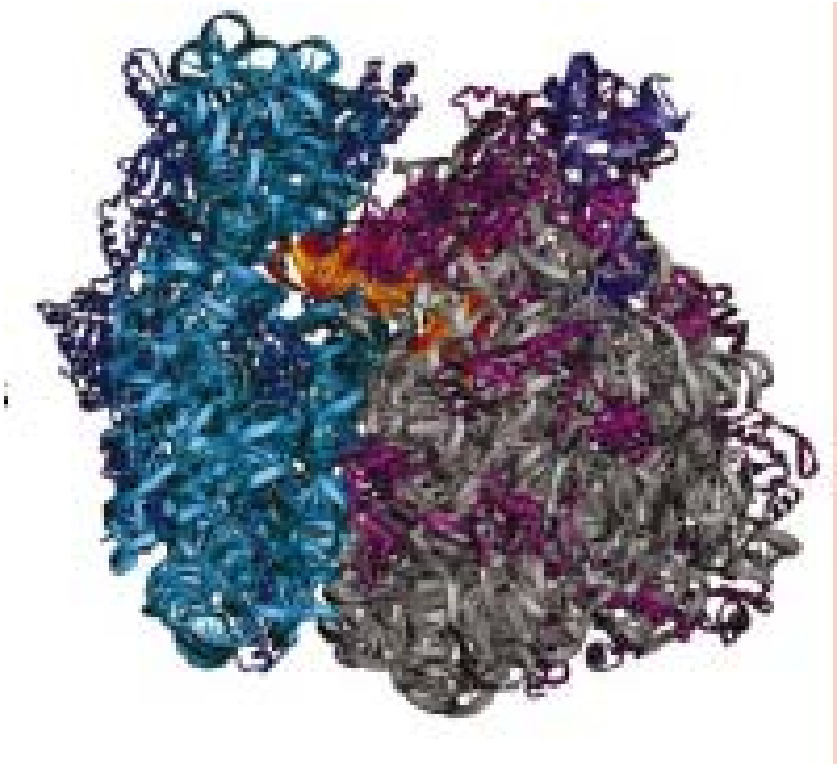


Structure of the largest of the two subunits of the ribosome, the "protein factory" of the cell.

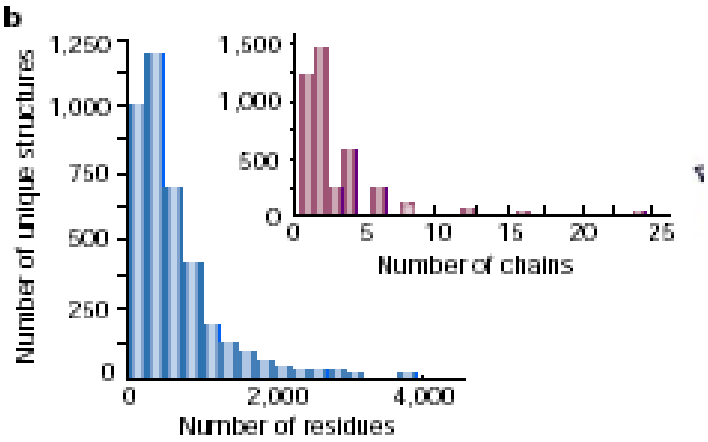
Transfer RNAs (green, blue and yellow) occupy cavity between two ribosomal subunits

www.lbl.gov/.../ribosome-crystallography.html

70S ribosome Determined by X-ray Crystallography



70S ribosome, which consists of 52 proteins and 3 RNA molecules, and has a relative molecular mass of ~2,500,000 (*M*r 2,500K).



Hybrid Approaches to Structure Determination

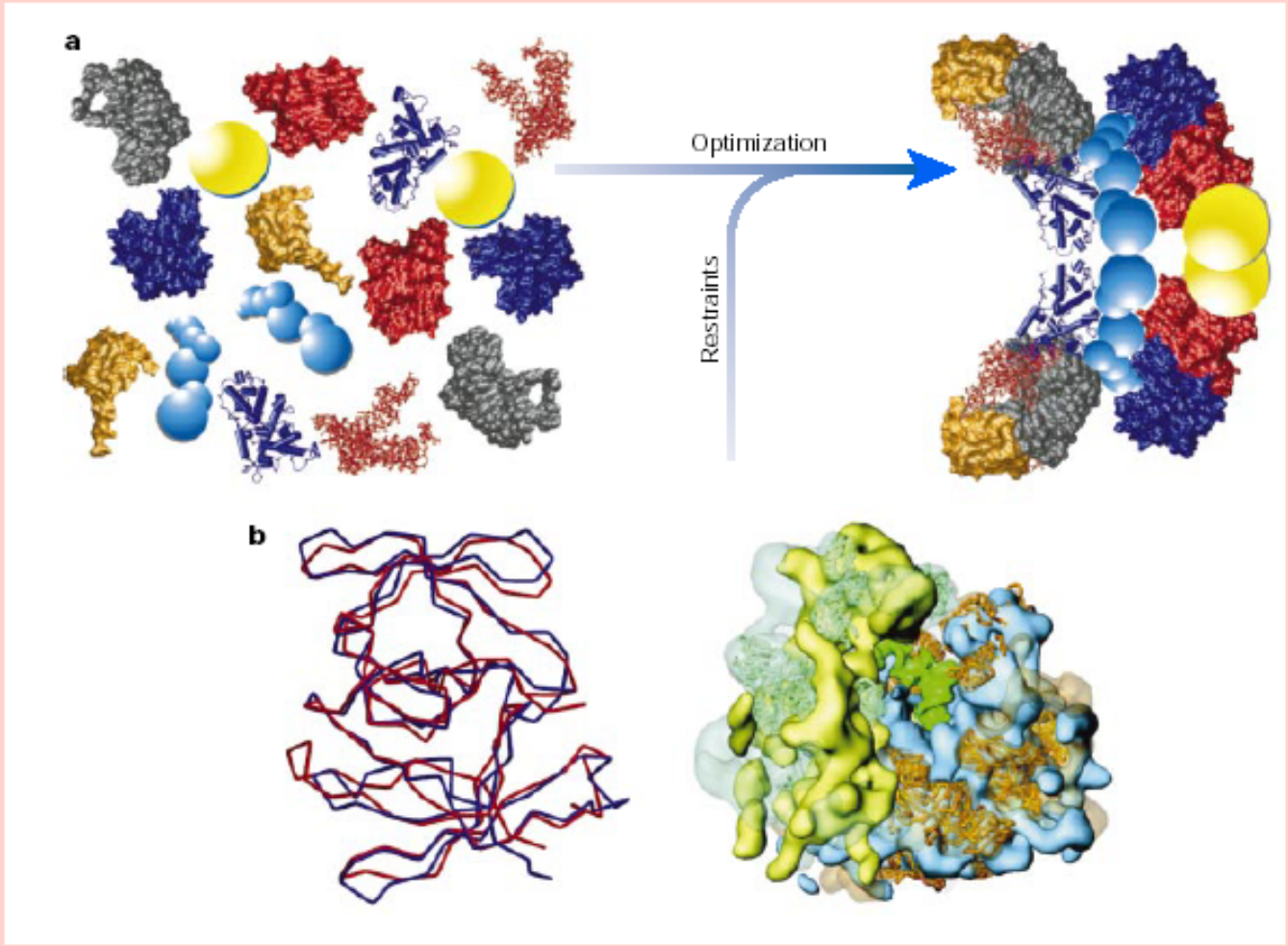
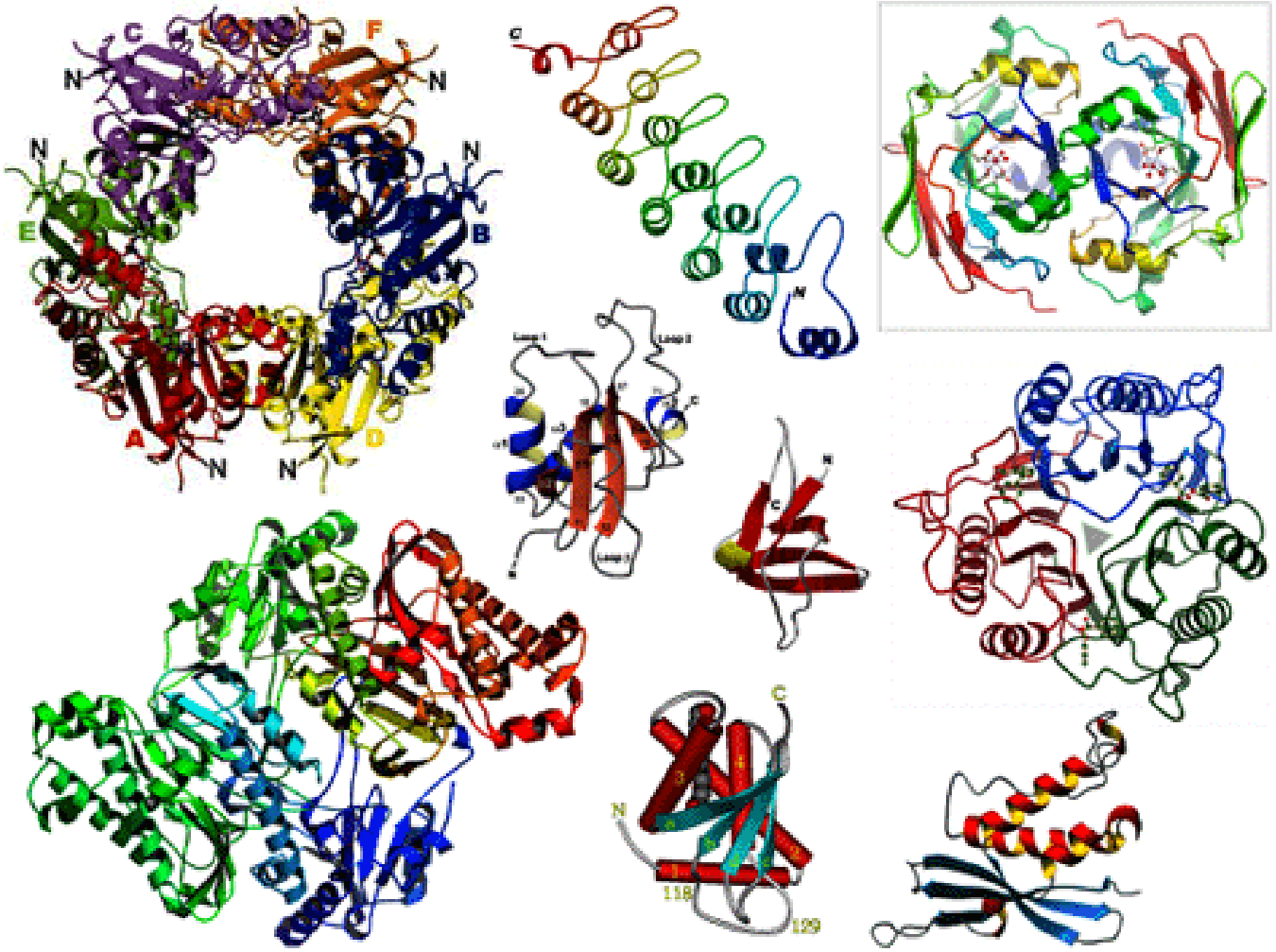


Figure 5 Hybrid approaches to structure determination of macromolecular complexes. **a**, Scheme illustrating the integration of a diverse set of structures varying in reliability and resolution into a hypothetical hybrid assembly structure. **b**, Hybrid assembly of the 80S ribosome from yeast[34]. Superposition of a comparative protein structure model for a domain in protein L2 from *Bacillus stearothermophilus* with the actual structure (1RL2) (left). A partial molecular model of the whole yeast ribosome (right) was calculated by fitting atomic rRNA (not shown) and comparative protein structure models (ribbon representation) into the electron density of the 80S ribosomal particle.

Structural Proteomics

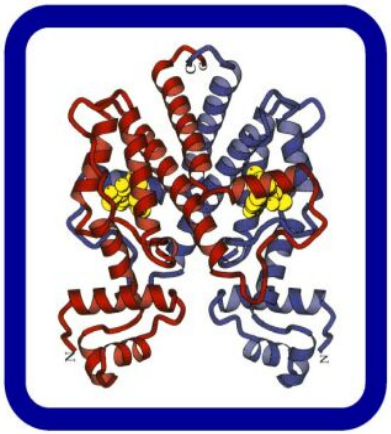


Structures of human proteins determined within the Protein Structure Factory.

Protein Structure Factory, Berlin

sequence

gene



structure

protein

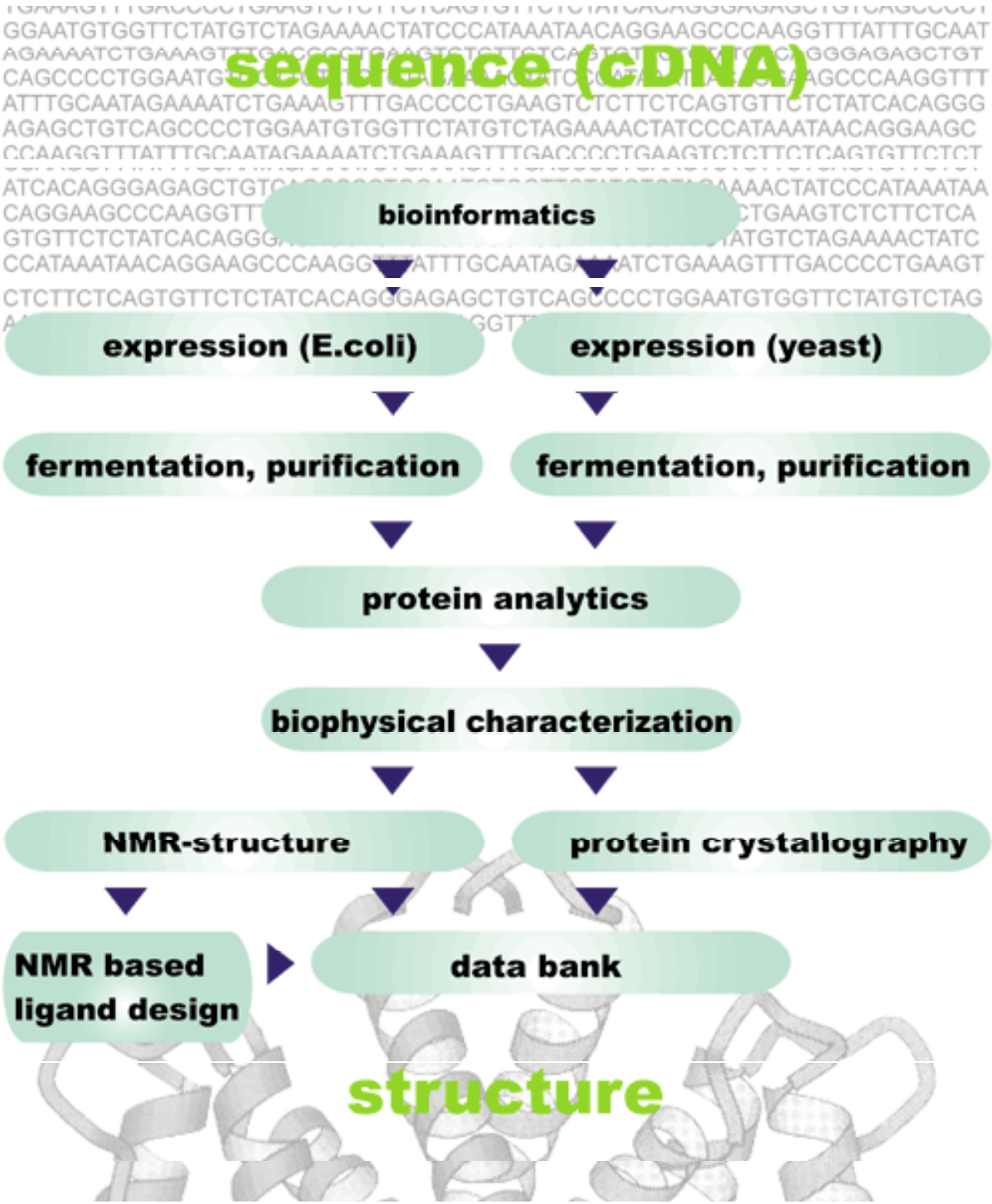
drug

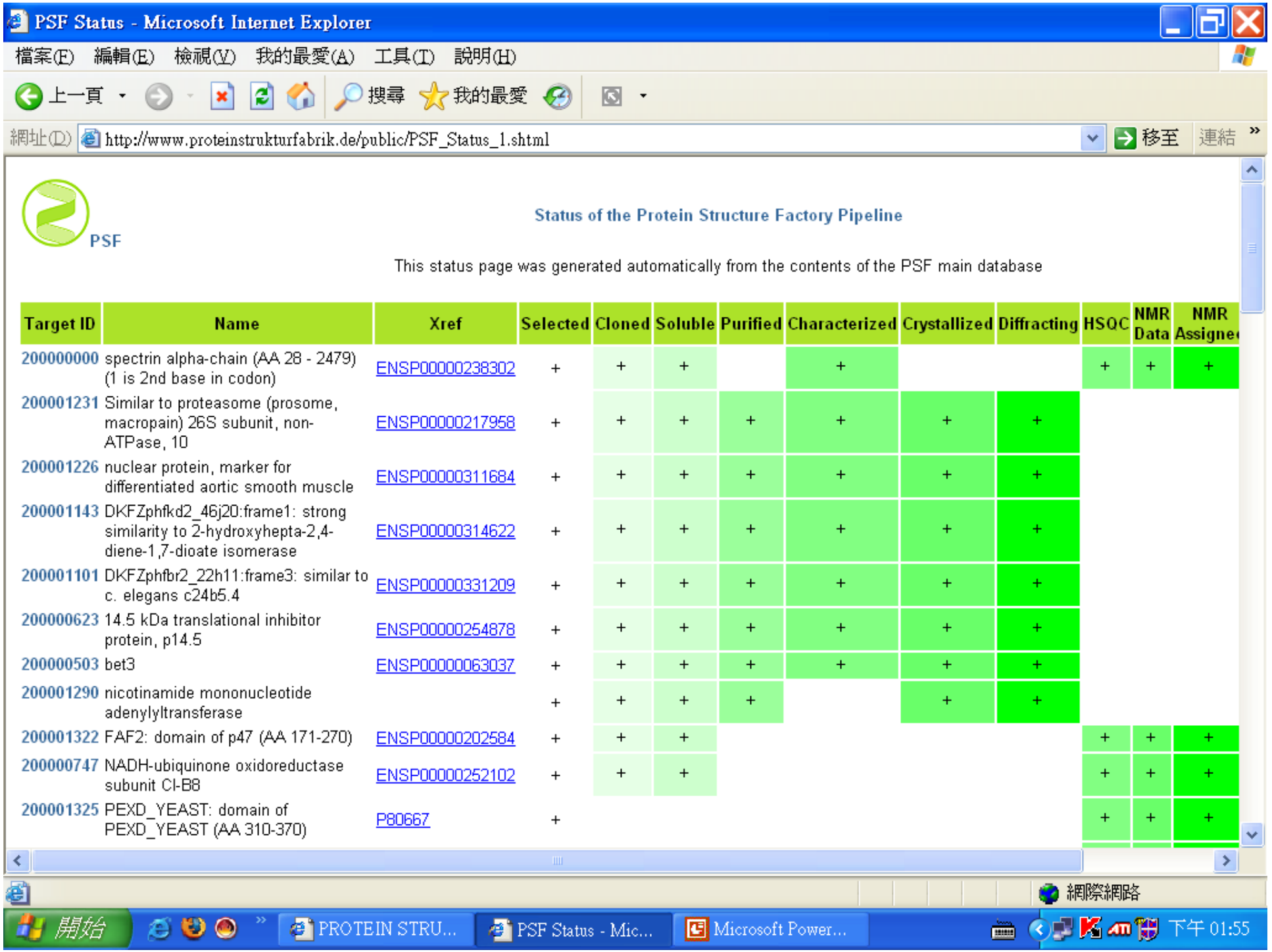
new pharmaceuticals

Protein Structure Factory, Berlin

It represents an integrative approach to structure analysis combining the computer-based analysis of genes by bioinformatics techniques, automated gene expression and purification of gene products, generation of a biophysical fingerprint of the proteins and the determination of their three-dimensional structures either in solution by NMR spectroscopy or in the crystalline state by X-ray Diffraction.

At a later stage, it may analyze various sets of input proteins selected by criteria of potential structural novelty or **medical or biotechnological usefulness**.


sequence (cDNA)
bioinformatics
expression (E.coli)
expression (yeast)
fermentation, purification
fermentation, purification
protein analytics
biophysical characterization
NMR-structure
protein crystallography
NMR based ligand design
data bank
structure

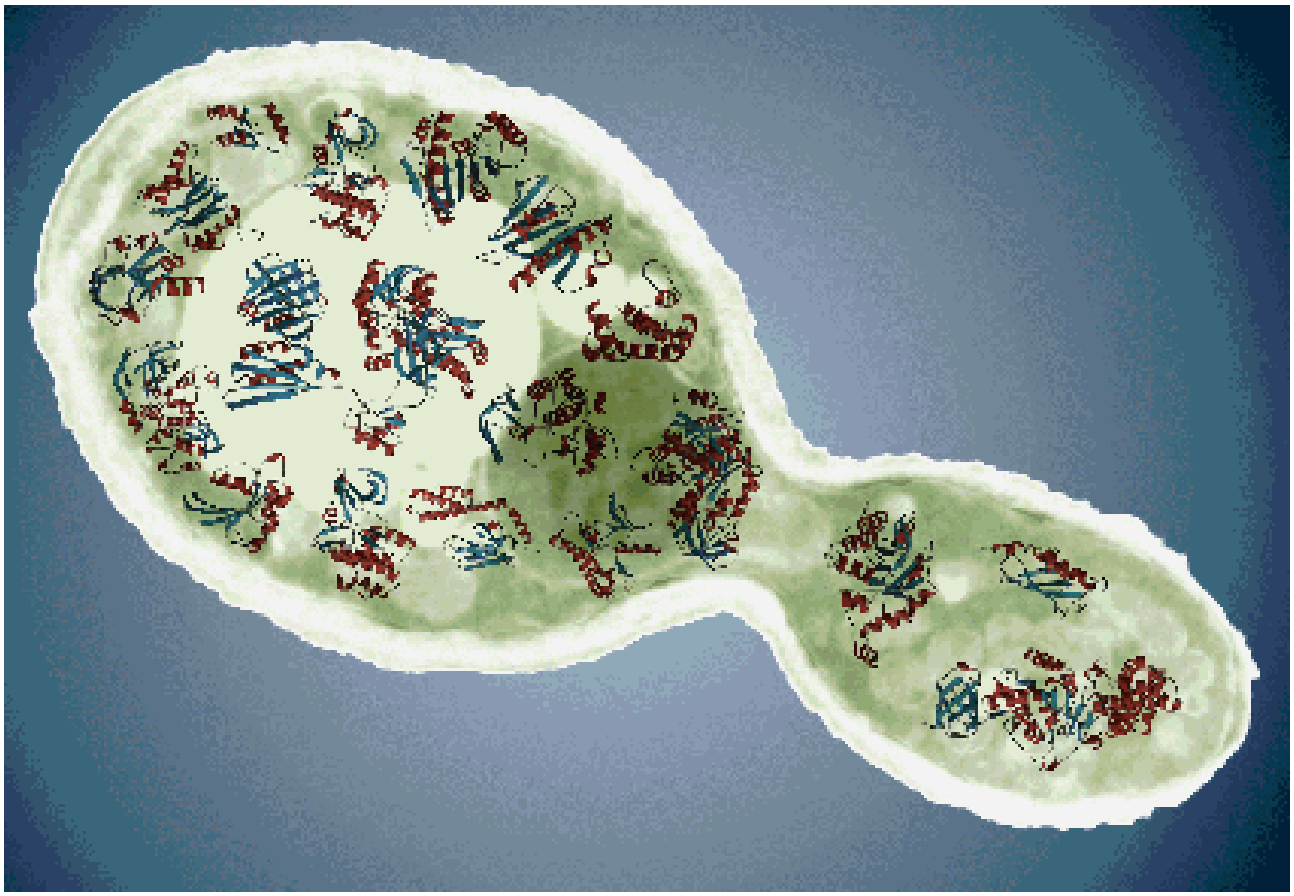


Target ID	Name	Xref	Selected	Cloned	Soluble	Purified	Characterized	Crystallized	Diffracting	HSQC	NMR Data	NMR Assigne
200000000	spectrin alpha-chain (AA 28 - 2479) (1 is 2nd base in codon)	ENSP00000238302	+	+	+		+			+	+	+
200001231	Similar to proteasome (prosome, macropain) 26S subunit, non-ATPase, 10	ENSP00000217958	+	+	+	+	+	+	+			
200001226	nuclear protein, marker for differentiated aortic smooth muscle	ENSP00000311684	+	+	+	+	+	+	+			
200001143	DKFZphfkd2_46j20:frame1: strong similarity to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	ENSP00000314622	+	+	+	+	+	+	+			
200001101	DKFZphfbr2_22h11:frame3: similar to c. elegans c24b5.4	ENSP00000331209	+	+	+	+	+	+	+			
200000623	14.5 kDa translational inhibitor protein, p14.5	ENSP00000254878	+	+	+	+	+	+	+			
200000503	bet3	ENSP00000063037	+	+	+	+	+	+	+			
200001290	nicotinamide mononucleotide adenylyltransferase		+	+	+	+		+	+			
200001322	FAF2: domain of p47 (AA 171-270)	ENSP00000202584	+	+	+					+	+	+
200000747	NADH-ubiquinone oxidoreductase subunit CI-B8	ENSP00000252102	+	+	+					+	+	+
200001325	PEXD_YEAST: domain of PEXD_YEAST (AA 310-370)	P80667	+							+	+	+

Target ID	Name	Xref	Selected	Cloned	Soluble	Purified	Characterized	Crystallized	Diffracting	HSQC	NMR Data	NMR Assigned	Structu
200000595	SRB7	ENSP00000282892	+	+	+	+							
200000548	general transcription factor	ENSP00000274408	+	+	+	+							
200000487	ornithine decarboxylase antizyme 2	ENSP00000316264	+	+	+	+							
200000485	'human homologue of rat ribosomal protein L9'	ENSP00000330879	+	+	+	+							
200000251	ubiquitin fusion-degradation 1 protein	ENSP00000263202	+	+	+	+							
200000203	transaldolase	ENSP00000321259	+	+	+	+							
200000166	phosphate carrier	ENSP00000228318	+	+	+	+							
200001464	domain of 200001460 AA(1-36): domain of domain of 200001241 AA(532-622) (AA 1-36)		+	+		+							
200001918	RAD51 homolog protein isoform 2		+	+	+								
200001916	cyclin-dependent kinase 4		+	+	+								
200001903	ras-related nuclear protein		+	+	+								
200001873	nuclear receptor coactivator												

Target ID	Name	Xref	Selected	Cloned	Soluble	Purified	Characterized	Crystallized	Diffracting	HSQC	NMR Data	NMR Assigned	Structu
200000921	DKFZphmel2_4i19:frame2: unknown protein	ENSP00000295240	+	+									
200000919	DKFZphfbr1_14j1:frame1: putative protein	ENSP00000328280	+	+									
200000918	DKFZphfbr1_10p15:frame1: putative protein		+	+									
200000916	DKFZphtes3_2k14:frame2: putative protein	ENSP00000254772	+	+									
200000915	DKFZphfbr1_16d9:frame2: putative protein		+	+									
200000914	DKFZphtes3_15l12:frame1: similarity to c-terminus of putative dtdp-glucose4-6-dehydratase	ENSP00000283148	+	+									
200000913	DKFZphfbr1_9b13:frame2: putative protein	ENSP00000328653	+	+									
200000912	DKFZphtes3_8o12:frame3: unknown	ENSP00000331417	+	+									
200000911	DKFZphtes3_20n2:frame2: unknown	ENSP00000327125	+	+									
200000877	uridine phosphorylase												

Incomplete Work!



A small sample of the 1,100 comparative models calculated for the proteins in the yeast genome is displayed over an image of a yeast cell