

## X-Ray Crystallography

*"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde*

Topics:

### 1. Protein Data Bank (PDB)

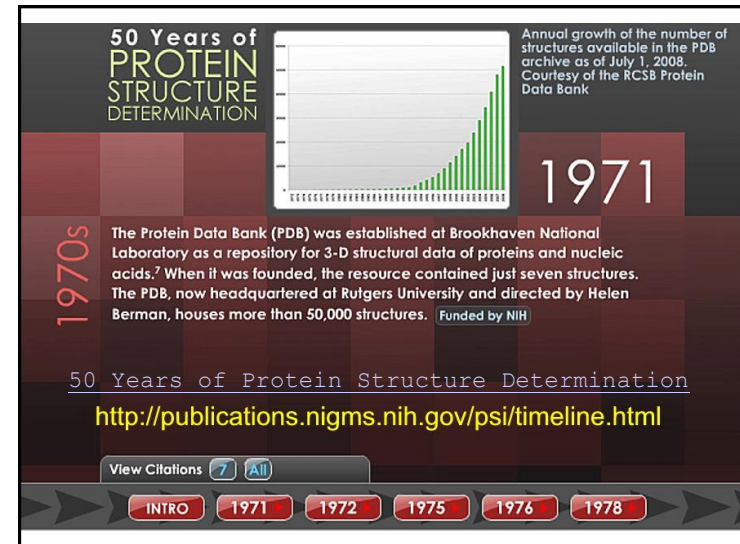
Data mining and Protein Structure Analysis Tools

### 2. Image Formation

Resolution / Wavelength (Amplitude, Phase) / Light Microscopy / EM / X-ray / (NMR)

### 3. X-Ray Crystallography

- Crystal Growth – Materials / Methods
- Crystal Lattices - Lattice Constants / Space Groups / Asymmetric Unit
- X-ray Sources – Sealed Tube / Rotation Anode / Synchrotron
- Theory of Diffraction – Bragg's Law / Reciprocal Space
- Data Collection – Methods / Detectors / Structure Factors
- Structure Solution – Phase Problem: MIR / MR / MAD
- Refinements and Models
- Analysis and presentation of results

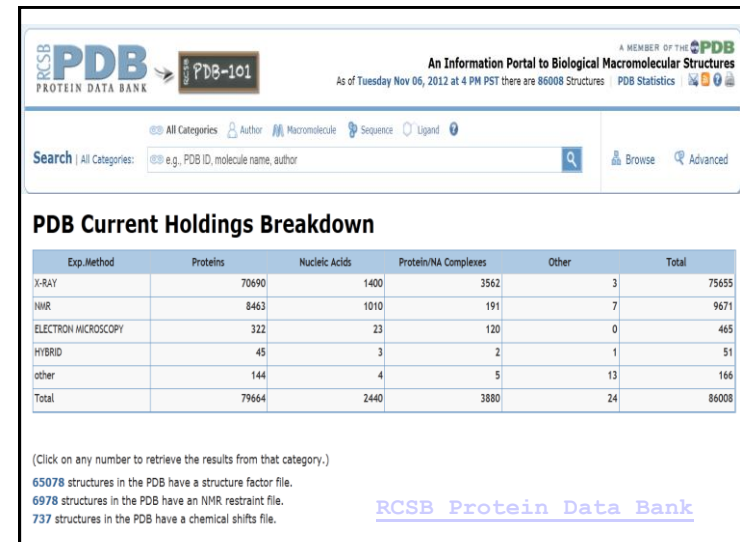
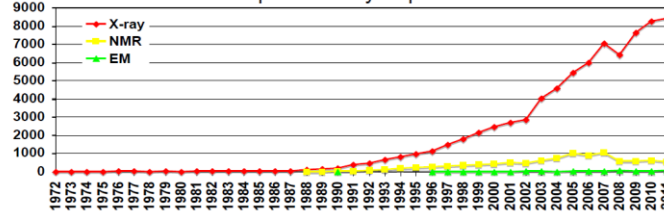


## 10,000-Fold Growth in Four Decades

<http://www.wwpdb.org/PDB40.html>

- 7 → >76,000 entries
- 2011 will see ~9,000 depositions
- Electron Microscopy is beginning to hit its stride

Calendar Year Depositions by Experimental Method



Analyze – structure (Ramachandran Plot) and biochemistry

Publish in leading biochemical or structural biology journal

Contribute results (coordinates, etc.) to PDB

\*\*\*\*\*

### Data Mining

Visualization programs (Cn3D / RasMol / SwissPDBV / etc)

SCOP – Structural Classification of Proteins

CATH – Classification / Arch / Topology

## SCOP Structural Classification of Proteins

Structural Classification of Proteins



Root: scop

Classes:

1. All alpha proteins (151)
2. All beta proteins (111)
3. Alpha and beta proteins (a/b) (117)   
Mainly parallel beta sheets (beta-alpha-beta units)
4. Alpha and beta proteins (a+b) (212)   
Mainly antiparallel beta sheets (segregated alpha and beta regions)
5. Multi-domain proteins (alpha and beta) (39)   
Folds consisting of two or more domains belonging to different classes
6. Membrane and cell surface proteins and peptides (12)   
Does not include proteins in the immune system
7. Small proteins (59)   
Usually dominated by metal ligand, heme, and/or disulfide bridges
8. Coiled coil proteins (5)   
Not a true class
9. Low resolution protein structures (17)   
Not a true class
10. Peptides (9)   
Peptides and fragments. Not a true class
11. Designed proteins (36)   
Experimental structures of proteins with essentially non-natural sequences. Not a true class

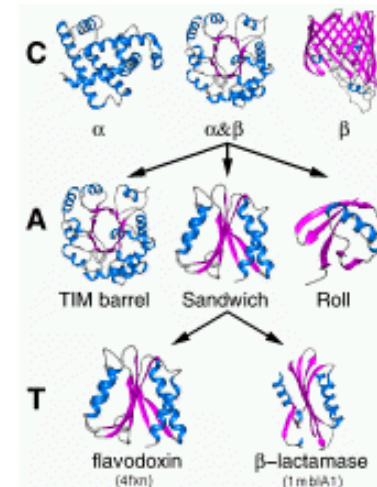
## CATH - Protein Structure Classification

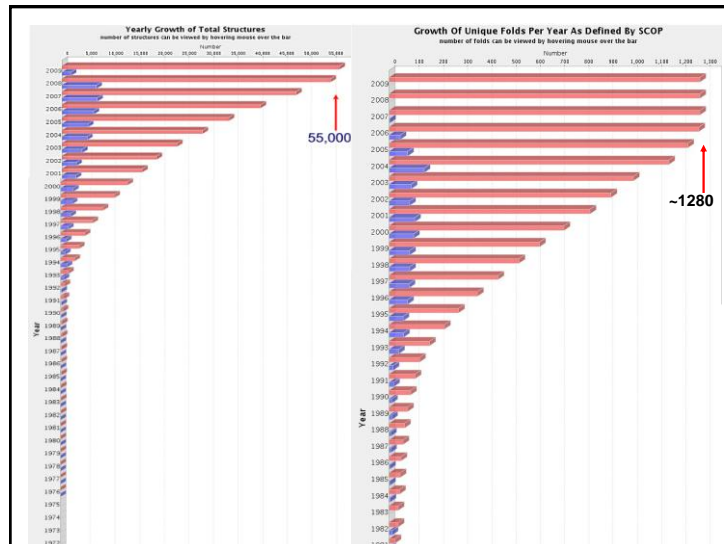
**CATH** is a novel hierarchical classification of protein domain structures, which clusters proteins at four major levels: **Class** (C), **Architecture** (A), **Topology** (T), and **Homologous** (H) Superfamily

**Class**, derived from **secondary structure** content, is assigned for more than 90% of protein structures automatically.

**Architecture**, which describes the **gross orientation of secondary structures**, independent of connectivities, is currently assigned manually. The **topology** level clusters structures according to their **topological connections and numbers of secondary structures**. The **homologous superfamilies** cluster proteins with **highly similar structures and functions**. The assignments of structures to topology families and homologous superfamilies are made by sequence and structure comparisons.

## CATH





## PMP | The Protein Model Portal

The Protein Model Portal (PMP) gives access to the various models that can be leveraged from PSI targets and other experimental protein structures by comparative modeling methods. The current release of the portal allows searching 7.6 million precomputed model structures provided by different partner sites, and provides access to various interactive services for template selection, target-template alignment, model building, and quality assessment.

- [CSMP](#) - Center for Structures of Membrane Proteins
- [JCSG](#) - Joint Center for Structural Genomics
- [MCSG](#) - Midwest Center for Structural Genomics
- [NESG](#) - Northeast Structural Genomics Consortium
- [NMHRCM](#) - New Methods for High-Resolution Comparative Modeling
- [NYSGXRC](#) - New York SGX Research Center for Structural Genomics

models menu  
PMP home  
advanced search  
interactive modeling  
quality estimation  
Protein Modeling 101  
CAMEO  
news and events  
documentation  
related tools  
about PMP  
contact us

Welcome to the  
**Protein Model Portal (PMP)**  
PMP gives access to various models computed by comparative modeling methods provided by different partner sites, and provides access to various interactive services for model building, and quality assessment.  
Please enter your query.  
  Examples: [UniProt AC] [UniProt ID] [RefSeq] [IPI] [PDBID] [Sequence] [Free Text]  
<http://www.proteinmodelportal.org/>  
Access all of PMP

**Interactive Modeling**  
  
**Need a model?**  
Submit your sequence to registered modeling servers and receive results by email

**Quality Estimation**  
  
**Are you aware of possible errors in a model?**  
Estimate the model accuracy by submitting to registered quality estimation servers

**Modeling highlights**  
  
**Structure-based prediction of protein-protein interactions on a genome-wide scale.**  
Zhang QC et.al. *Nature* (2012)

**Tutorials and Guides**  
  
**How reliable is a model?**  
Introduction into homology modeling.  
PMP Tutorials (video)  
Ask an Expert!

## X-Ray Crystallography

*"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde*

Topics:

### 1. Protein Data Bank (PDB)

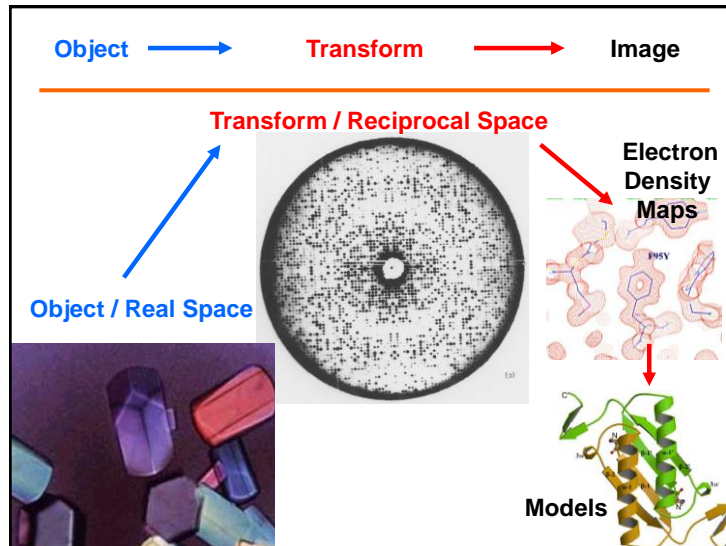
Data mining and Protein Structure Analysis Tools

### 2. Image Formation

Resolution / Wavelength (Amplitude, Phase) / Light Microscopy / EM / X-ray / (NMR)

### 3. X-Ray Crystallography

- Crystal Growth – Materials / Methods
- Crystal Lattices - Lattice Constants / Space Groups / Asymmetric Unit
- X-ray Sources – Sealed Tube / Rotation Anode / Synchrotron
- Theory of Diffraction – Bragg's Law / Reciprocal Space
- Data Collection – Methods / Detectors / Structure Factors
- Structure Solution – Phase Problem: MIR / MR / MAD
- Refinements and Models
- Analysis and presentation of results



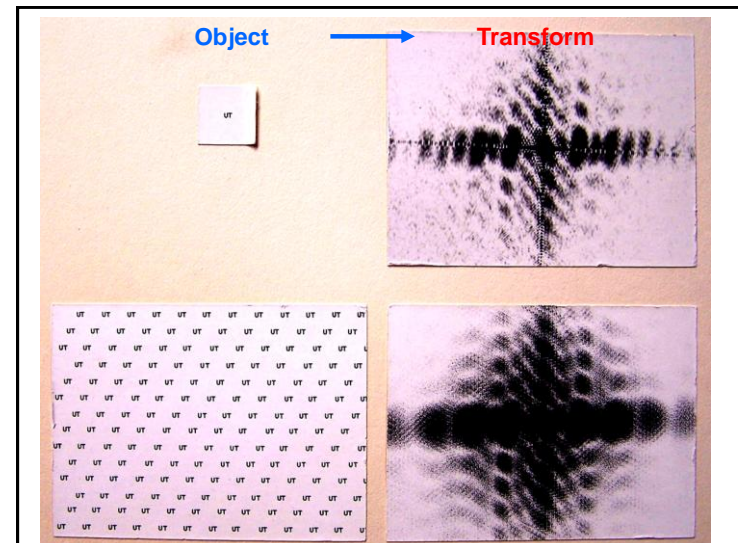
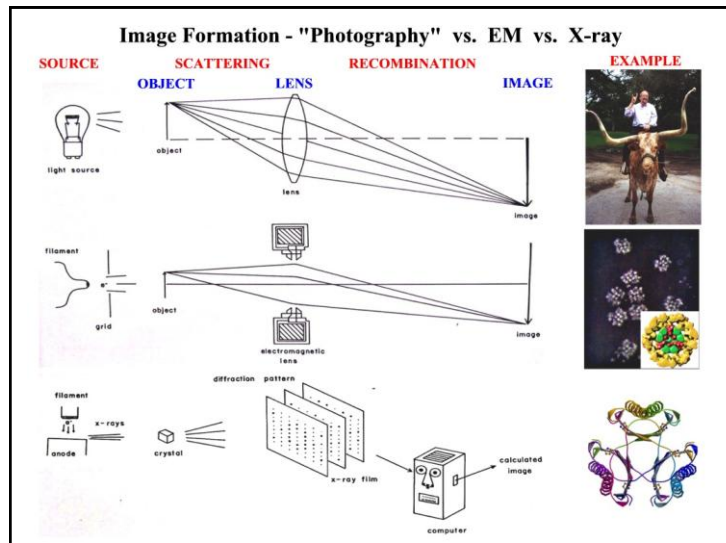
*"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde*

**Image Formation**  
Abbe (~1873):  
Limit Res.  $\sim \lambda/2$

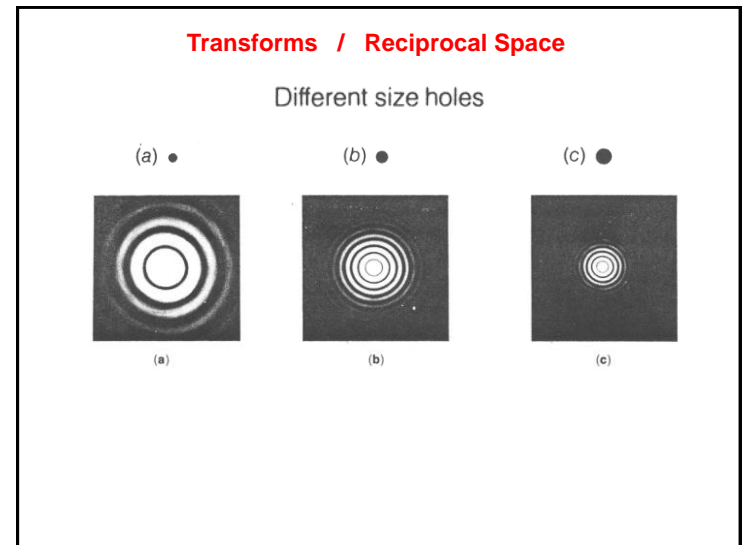
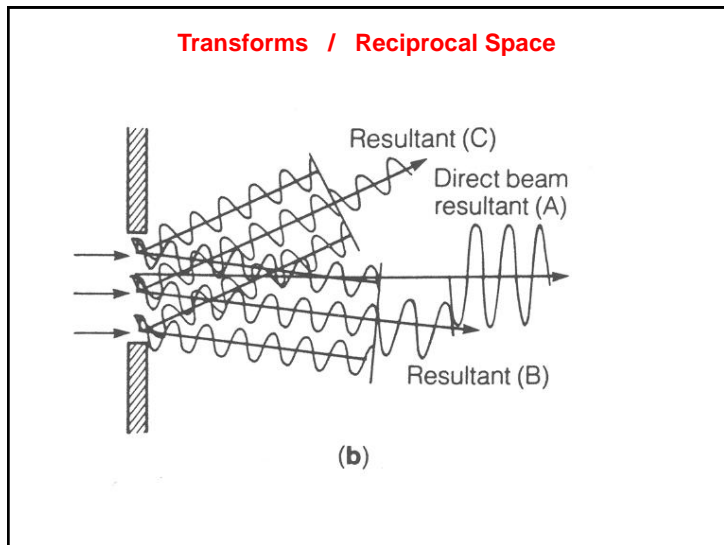
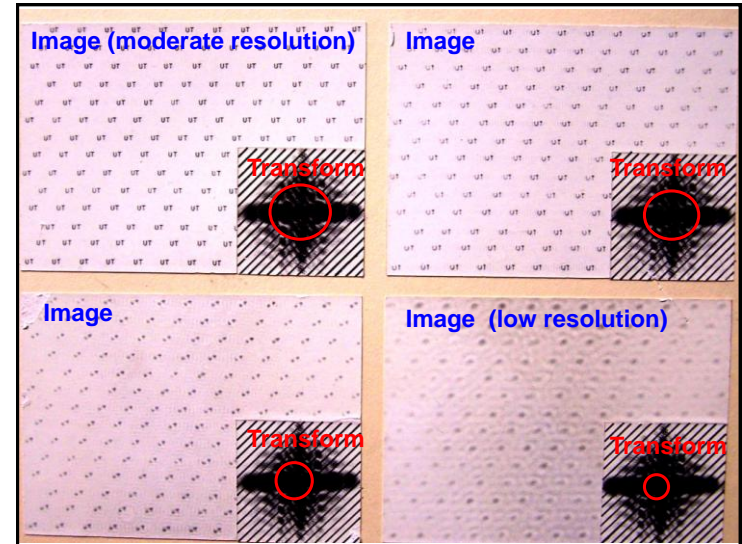
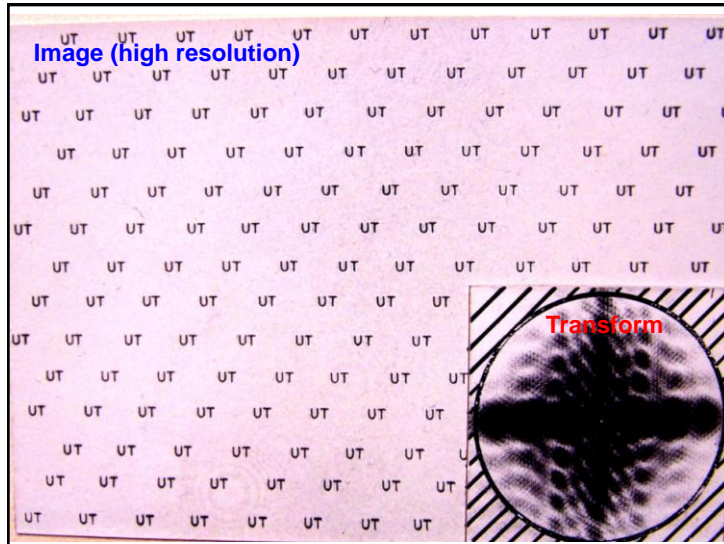
• Light Photography  
 $\lambda \sim 400 - 700 \text{ nm}$

• Electron Microscopy  
 $\lambda \sim 0.001 - 0.1 \text{ nm}$

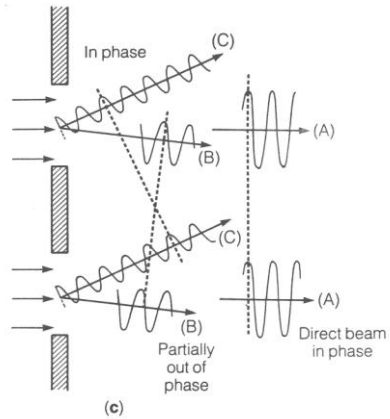
• X-Ray or NMR  
 $\lambda \sim 0.1 \text{ nm}$



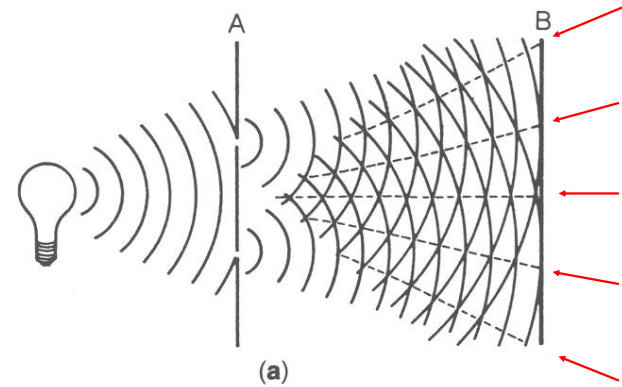




### Transforms / Reciprocal Space



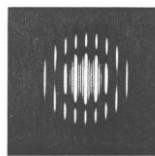
### Transforms / Reciprocal Space



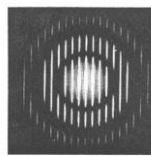
### Transforms / Reciprocal Space

Five horizontal holes  
with various spacings

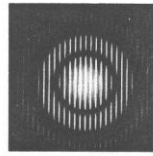
(j) ••••• (k) ••••• (l) • • • • •



(j)



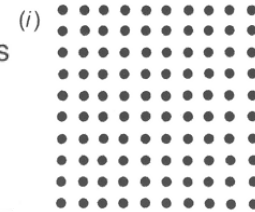
(k)



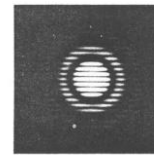
(l)

### Transforms / Reciprocal Space

Vertical holes and nets of holes

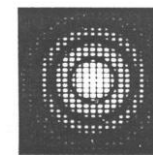


(g) ••

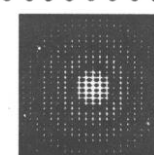


(g)

(h) ••••



(h)



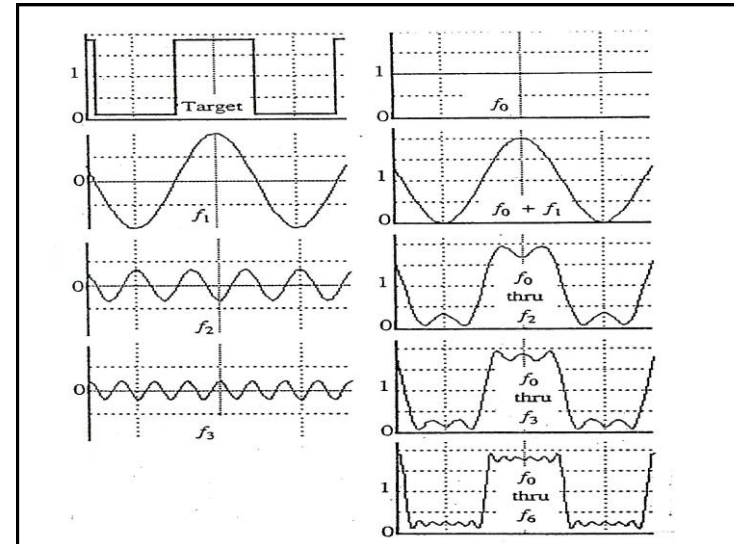
(i)

## Joseph Fourier / Fourier Series ~1808



Fourier series are named in honor of Joseph Fourier (1768-1830), who made important contributions to the study of trigonometric series, after preliminary investigations by Euler, d'Alembert, and Bernoulli. He applied this technique to find the solution of the heat equation, publishing his initial results in 1807, and publishing his *Théorie analytique de la chaleur* in 1822

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi t}{L} + \sum_{n=1}^{\infty} b_n \sin \frac{n\pi t}{L}$$



Fourier Series - a way of expressing functions in terms of an infinite series using the sum of sine and cosine functions.

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi t}{L} + \sum_{n=1}^{\infty} b_n \sin \frac{n\pi t}{L}$$

If  $f(t)$  is expanded in the range  $-L$  to  $L$  (period  $= 2L$ ) so that the range of integration is  $2L$ , i.e. half the range of integration is  $L$ , then the **Fourier** coefficients are given by

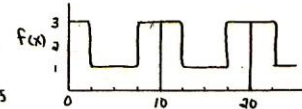
$$a_0 = \frac{1}{L} \int_{-L}^L f(t) dt$$

$$a_n = \frac{1}{L} \int_{-L}^L f(t) \cos \frac{n\pi t}{L} dt \quad b_n = \frac{1}{L} \int_{-L}^L f(t) \sin \frac{n\pi t}{L} dt$$

where  $n = 1, 2, 3 \dots$

Example:

$$f(x) = \begin{cases} 3 & 0 < x < 2.5 \\ 1 & 2.5 < x < 7.5 \end{cases}$$



$$\Rightarrow a_n = \frac{1}{10} \int_0^{2.5} (3) \cos 2\pi \frac{nx}{10} dx + \frac{1}{10} \int_{2.5}^{7.5} (1) \cos 2\pi \frac{nx}{10} dx + \frac{1}{10} \int_{7.5}^{10} (3) \cos 2\pi \frac{nx}{10} dx$$

$$a_0 = \frac{1}{10} (7.5 - 0) + \frac{1}{10} (7.5 - 2.5) + \frac{1}{10} (30 - 22.5) = 2.0$$

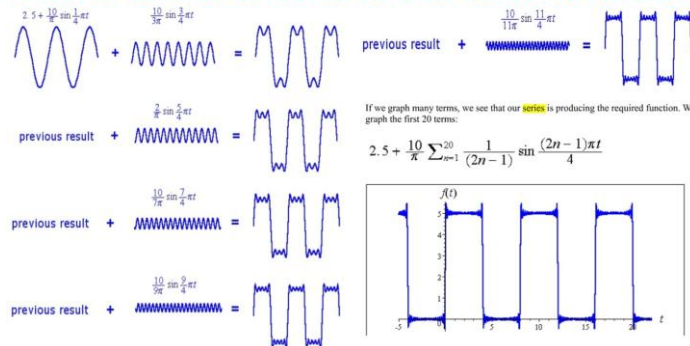
$$a_n = \frac{1}{10} \left[ \frac{3 \cdot 10}{2\pi n} \sin \frac{2\pi n x}{10} \right]_0^{2.5} + \frac{1}{10} \left[ \frac{10}{2\pi n} \sin \frac{2\pi n x}{10} \right]_{2.5}^{7.5} + \frac{1}{10} \left[ \frac{3 \cdot 10}{2\pi n} \sin \frac{2\pi n x}{10} \right]_{7.5}^{10}$$

$$\Rightarrow a_n = \frac{1}{n\pi} \left[ \sin \frac{n\pi}{2} - \sin \frac{3n\pi}{2} \right]$$

$$a_0 = 2.0$$

# Fourier Series

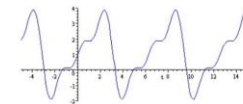
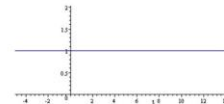
## Example - Square Wave



## Example - Saw Tooth Function

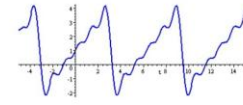
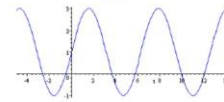
$$f(t) = 1 + 2 \sin t - \sin 2t + \frac{2}{3} \sin 3t$$

$f(t) = 1$  (first term of the series):



$$f(x) = 1 + 2 \sin t - \sin 2t + \frac{2}{3} \sin 3t - \frac{1}{2} \sin 4t + \frac{2}{5} \sin 5t + \dots$$

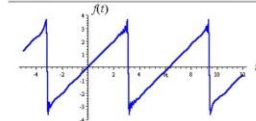
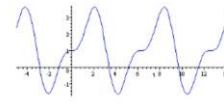
$f(t) = 1 + 2 \sin t$  (first 2 terms of the series):



The graph of the first 40 terms is:

$$\sum_{n=1}^{40} \left( \frac{2}{n} \right) (-1)^{n+1} \sin nt$$

$f(t) = 1 + 2 \sin t - \sin 2t$  (first 3 terms of the series):



Fourier Series Applet <http://www.falstad.com/fourier/>

Kevin Cowtan's Picture Book of Fourier Transforms - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop

Bookmarks Location <http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html> What's Related

## Kevin Cowtan's Book of Fourier

This is a book of pictorial 2-d Fourier Transforms. These are particularly relevant to my own field of X-ray crystallography, but should be of interest to anyone involved in signal processing or frequency domain calculations.

**Contents:** <http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html>

- Introduction
- Book of Crystallography
- Duck Tales and missing data
- A little Animal Magic and cross phasing
- A Tale of Two Cats and image restoration
- Animal Liberation and free-sets
- The Gallery Other interesting pictures.

**Other topics:**

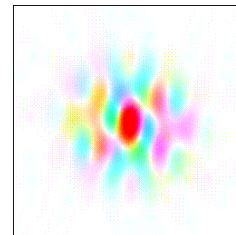
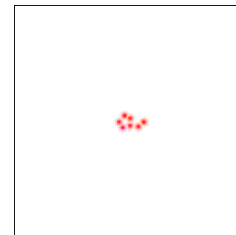
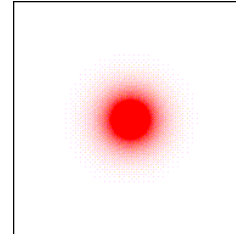
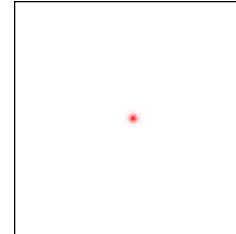
[The Interactive Structure Factor Tutorial](#): Learn about structure factors and maps.

An introduction to crystallographic [Fourier transforms](#). The mathematical link between [Scattering theory](#) and Fourier theory. An explanation of the [convolution theorem](#).

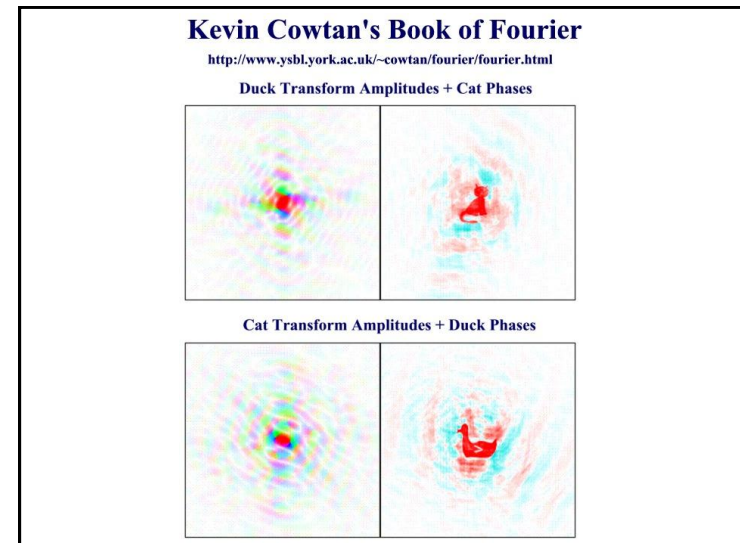
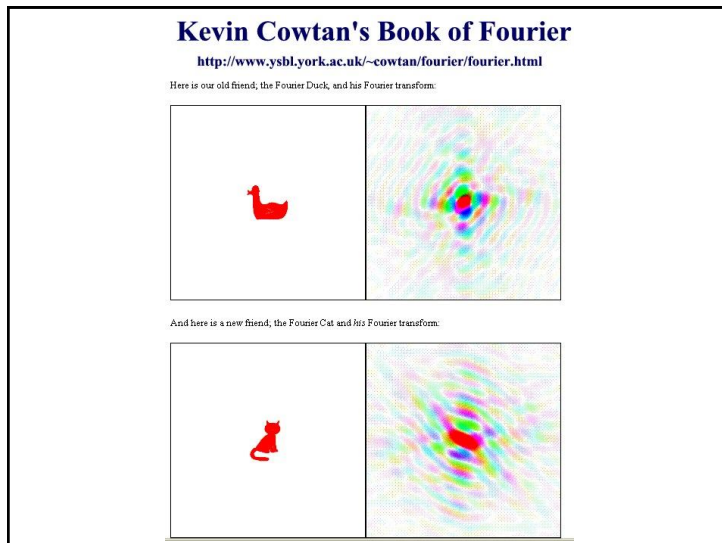
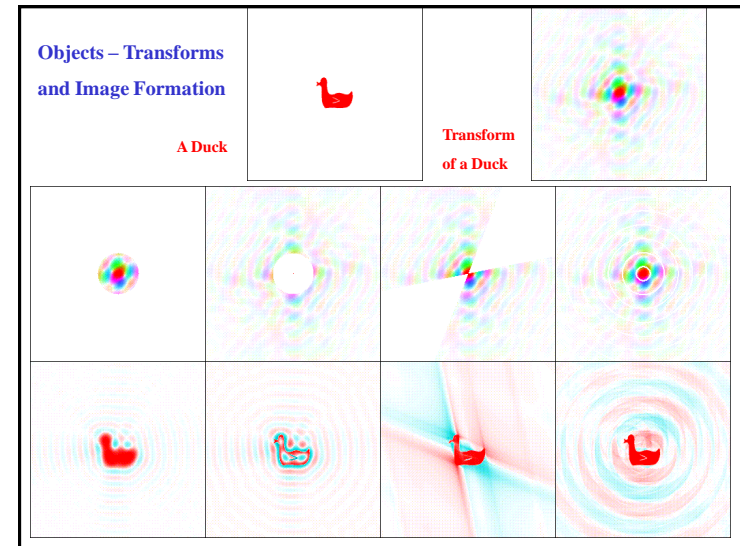
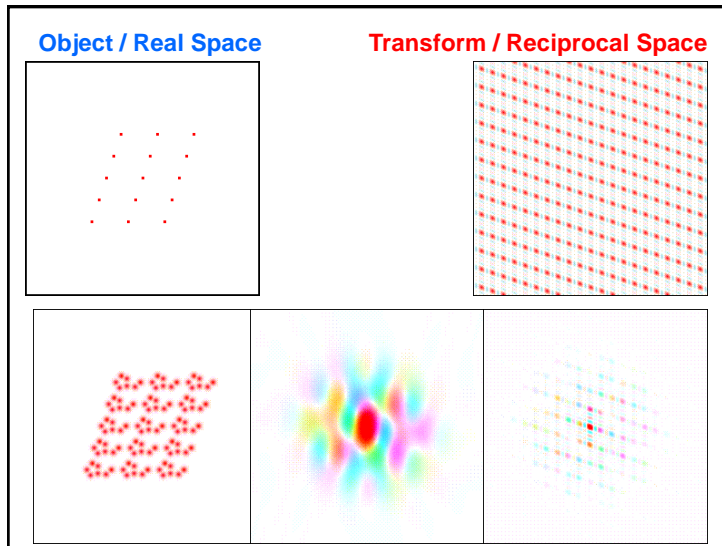
**Teaching materials elsewhere**

## Object / Real Space

## Transform / Reciprocal Space

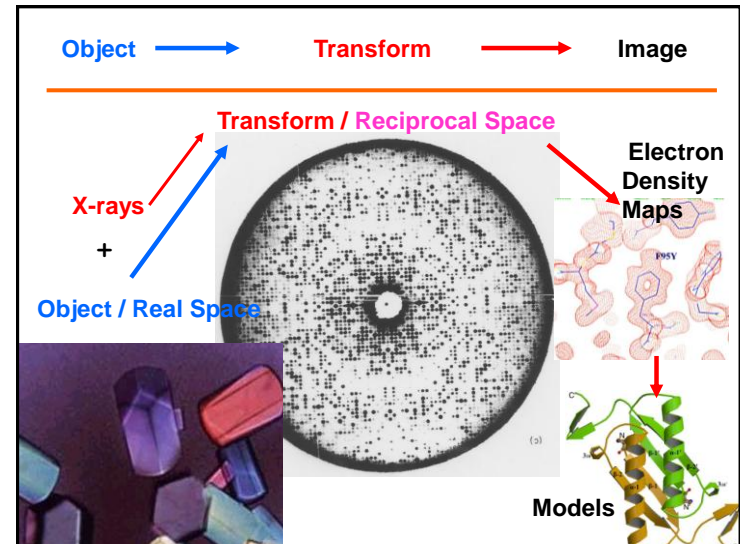
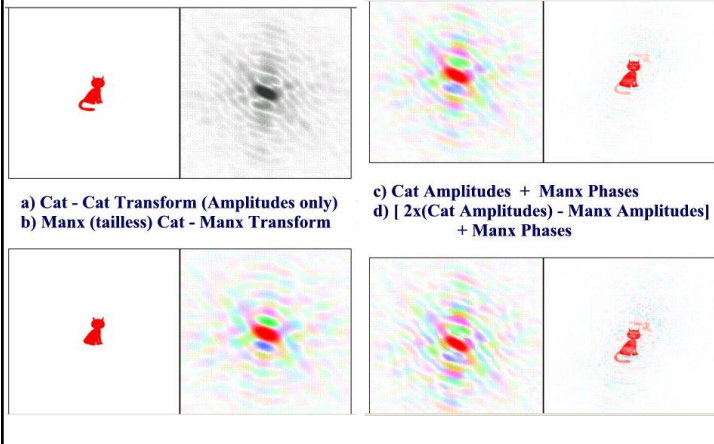






## Kevin Cowtan's Book of Fourier

<http://www.yzbl.york.ac.uk/~cowtan/fourier/fourier.html>



## X-Ray Crystallography

*"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde*

Topics:

### 1. Protein Data Bank (PDB)

Data mining and Protein Structure Analysis Tools

### 2. Image Formation

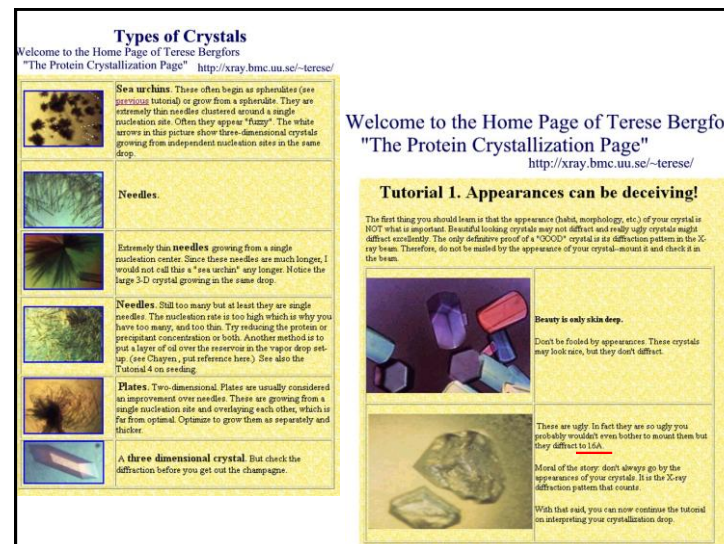
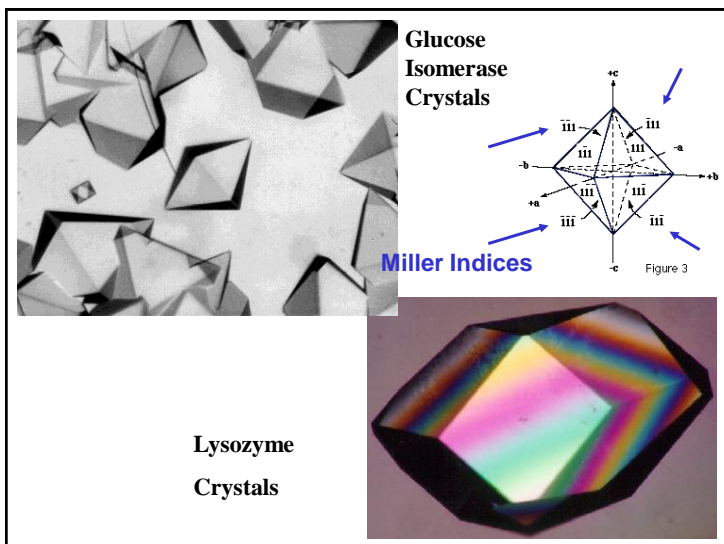
Resolution / Wavelength (Amplitude, Phase) / Light Microscopy / EM / X-ray / (NMR)

### 3. X-Ray Crystallography (after NMR)

- a) Crystal Growth – Materials / Methods
- b) Crystal Lattices - Lattice Constants / Space Groups / Asymmetric Unit
- c) X-ray Sources – Sealed Tube / Rotation Anode / Synchrotron
- d) Theory of Diffraction – Bragg Equation (take 1) / Reciprocal Space
- e) Data Collection – Methods / Detectors / Structure Factors
- f) Structure Solution – Phase Problem: MIR / MR / MAD
- h) Refinements and Models
- i) Analysis and presentation of results







## Variables that influence crystal growth

1. **Nature of macromolecule** – Purity and concentration of macromolecule
2. **Nature and concentration of precipitant**
3. **pH / Temperature / Pressure**
4. **Level of reducing agent or oxidant**
5. **Substrates, coenzymes, and ligands / Metal ions**
6. **Preparation and storage of macromolecule / Proteolysis and fragmentation**
7. **Age of macromolecule / Degree of denaturation**
8. **Vibration and sound**
9. **Volume of crystallization sample**
10. **Seeding**
11. **Amorphous precipitate**
12. **Buffers**
13. **Cleanliness**
14. **Organism or species from which the macromolecule was isolated**
15. **Gravity, gradients and convection**

## Common Compounds used in Crystallization

**Ammonium sulfate** / or sodium

Sodium or ammonium citrate

Sodium or ammonium acetate

Magnesium sulfate

Cetyltrimethyl ammonium salts

**Polyethylene glycol** 400, 1000, 2000, 4000, 6000, 8000, 15,000 M

## Methods for protein crystallization

**Batch crystallization** (simply dump reagents together)

**Liquid-liquid diffusion** in a capillary tube

**Vapor diffusion**-the most successful method (**hanging drop**, **sitting drop**), typically using a Limbro plate. Equilibration occurs between the liquid and vapor phase.

**Dialysis**

## Hanging Drop Method - Crystal Screening

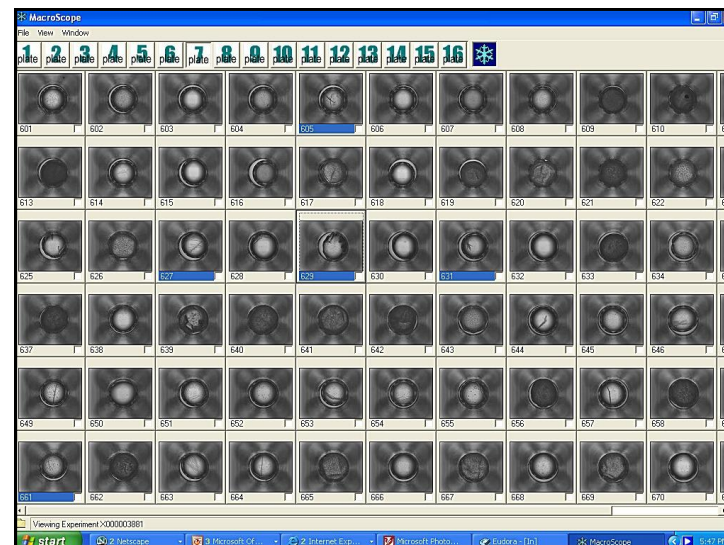
### The Experimental Setup

In order to obtain a crystal, the protein molecules must assemble into a periodic lattice. One starts with a solution of the protein with a fairly high concentration (2-30 mg/ml) and adds reagents that reduce the solubility close to spontaneous precipitation. By slow further concentration, and under conditions suitable for the formation of a few nucleation sites, small crystals **may** start to grow. Often very many conditions have to be tried to succeed. This is usually done by **initial screening**, followed by a systematic optimization of conditions. Crystals should be a few tenths of a mm in each direction to be useful for diffraction experiments.

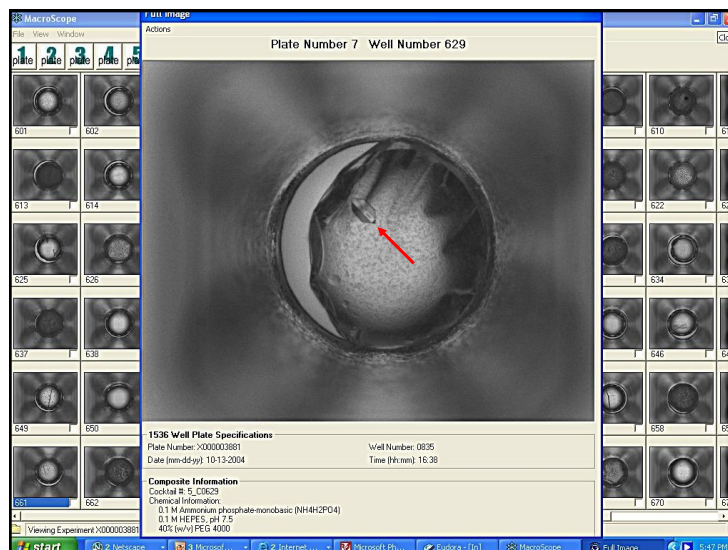


### Hampton Crystal Screen Solutions

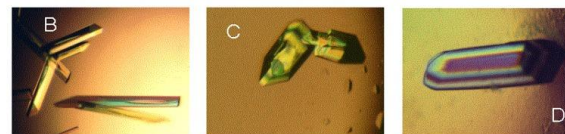
Tube #	SALT	BUFFER	Precipitant	Miniscreen	Tube #
1	0.02M Calcium Chloride	0.1M Na Acetate pH 4.6	30% w/v 2-methyl-2,4-pentanediol	Y	1
2	None	None	0.4M K <sub>2</sub> Na Tartrate tetrahydrate		2
3	None	None	0.4M Ammonium dihydrogen phosphate		3
4	None	0.1M Tris-HCl pH 8.5	2.0M Ammonium Sulfate	Y	4
5	0.2M tri-sodium citrate	0.1M Na HEPES pH 7.5	30% w/v 2-methyl-2,4-pentanediol		5
6	0.2M Magnesium chloride	0.1M Tris-HCl pH 8.5	30% w/v PEG 4000		6
7	None	0.1M Na Cacodylate pH 6.5	1.4M Sodium acetate trihydrate		7
8	0.2M tri-sodium citrate	0.1M Na Cacodylate pH 6.5	30% v/v 2-propanol		8
9	0.2M Ammonium acetate	0.1M Na Citrate pH 5.6	30% w/v PEG 4000	Y	9
10	0.2M Ammonium acetate	0.1M Na Acetate pH 4.6	30% w/v PEG 4000	Y	10



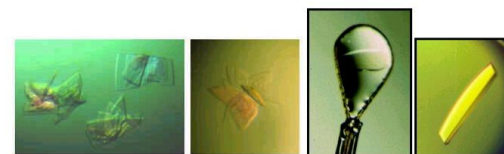




HinJ - His complex:

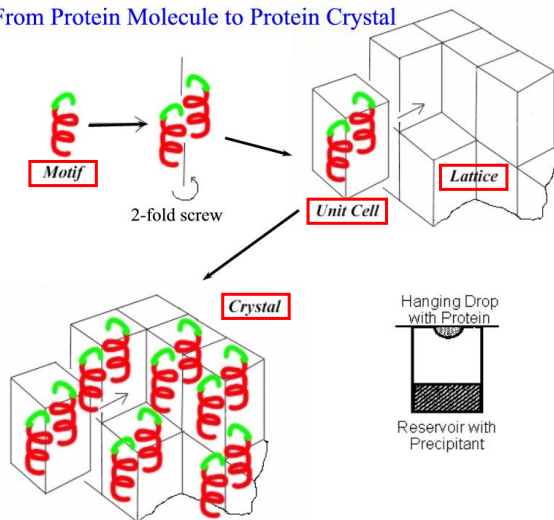


Tetanus Toxin C Fragment:



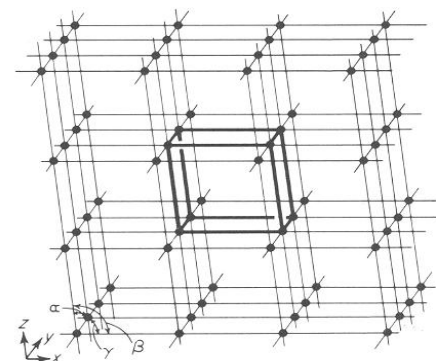
[http://www.ccp14.ac.uk/ccp/web-mirrors/lnlrupp/crystal\\_lab/Crys\\_lab.html](http://www.ccp14.ac.uk/ccp/web-mirrors/lnlrupp/crystal_lab/Crys_lab.html)

### From Protein Molecule to Protein Crystal



A **unit cell** is defined by its lattice constants:

**a, b, c** and  **$\alpha, \beta, \gamma$**

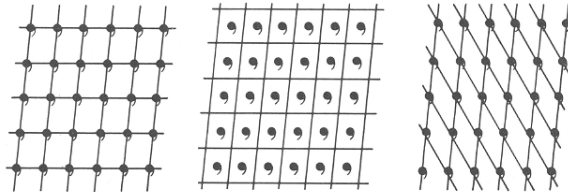


Three-dimensional lattice, showing unit cell (heavy lines).

## How to identify “the” unit cell ?

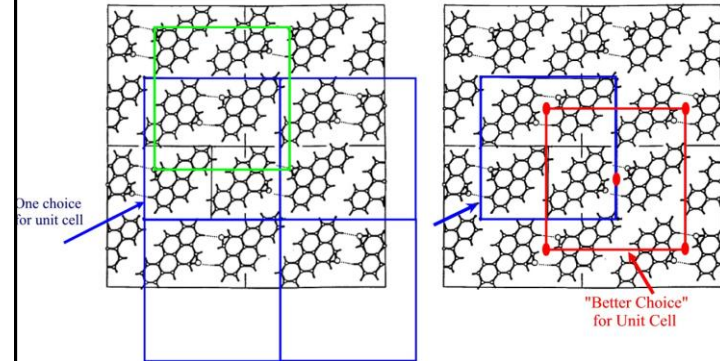


Regular two-dimensional array.



Three different grid systems referred to the array same array.

## Unit Cell Selection is Based on Symmetry



## Crystal Systems

Crystal System	Bravais Type(s)	External Minimum Symmetry	Unit Cell Properties
Triclinic	P	None	$a \neq b \neq c$ , $\alpha \neq \beta \neq \gamma$
Monoclinic	P, C	One 2-fold axis, parallel b (b unique)	$a \neq b \neq c$ , $\alpha = \beta = \gamma = 90^\circ$
Orthorhombic	P, I, F	Three perpendicular 2-folds	$a \neq b \neq c$ , $\alpha = \beta = \gamma = 90^\circ$
Tetragonal	P, I	One 4-fold axis, parallel c	$a = b \neq c$ , $\alpha = \beta = \gamma = 90^\circ$
Trigonal	P, R	One 3-fold axis	$a = b \neq c$ , $\alpha = \beta = \gamma = 120^\circ$
Hexagonal	P	One 6-fold axis	$a = b \neq c$ , $\alpha = \beta = \gamma = 120^\circ$
Cubic	P, F, I	Four 3-folds along space diagonal	$a = b = c$ , $\alpha = \beta = \gamma = 90^\circ$

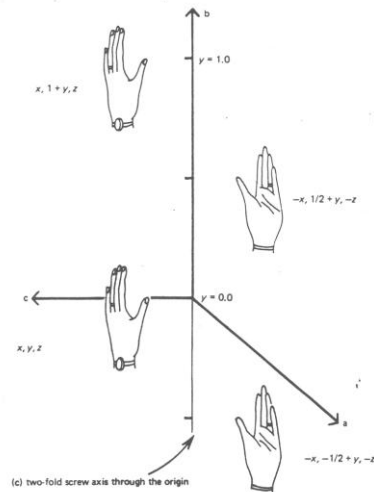
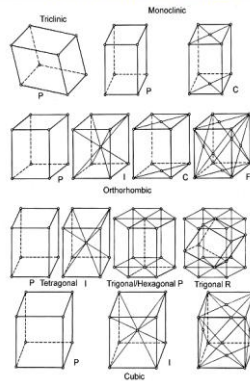
Symmetry operations : 1, 2, 3, 4, 6, -1, -2, -3, -4, -6, m

Crystal System	Point groups	Lauze Class	Patterson Symmetry
Triclinic	1, -1	-1	P-1
Monoclinic	2, m, 2/m	2/m	P2/m, C2/m
Orthorhombic	222, mm2, mmm	mmm	Pmm, Cmm, Fmm, Immm
Tetragonal	4, -4, 4/m, 422, 4mm, -42m, 4/mmm	4/m, 4/mmm	P4/m, 4/m, P4/mmm, 4/mmm
Trigonal	3, -3, 32, 3m, -3m	-3, -3m	P-3, R-3, P-3m, P-31m, R-3m
Hexagonal	6, -6, 6/m, 622, 6mm, -62m, 6/mmm	6/m, 6/mmm	P6/m, P6/mmm
Cubic	23, m-3, 432, -43m, m3m	m-3, m-3m	Pm-3, Im-3, Fm-3m, Pm-3m, Im-3m

Notes **Note: Only 2-, 3-, 4-, and 6-fold rotations allowed**

- Lauze class corresponds to symmetry of reciprocal space (diffraction pattern)
- Patterson symmetry is Lauze class plus allowed Bravais centering, i.e. centrosymmetric and symmetric

## The 14 Bravais Lattices



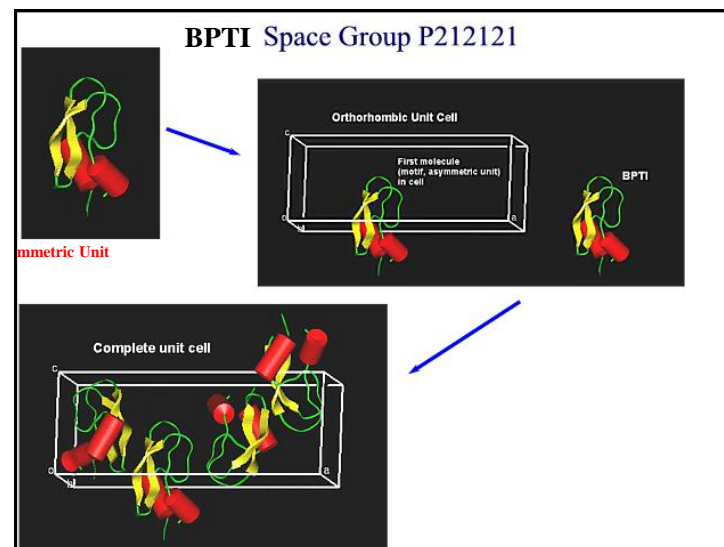
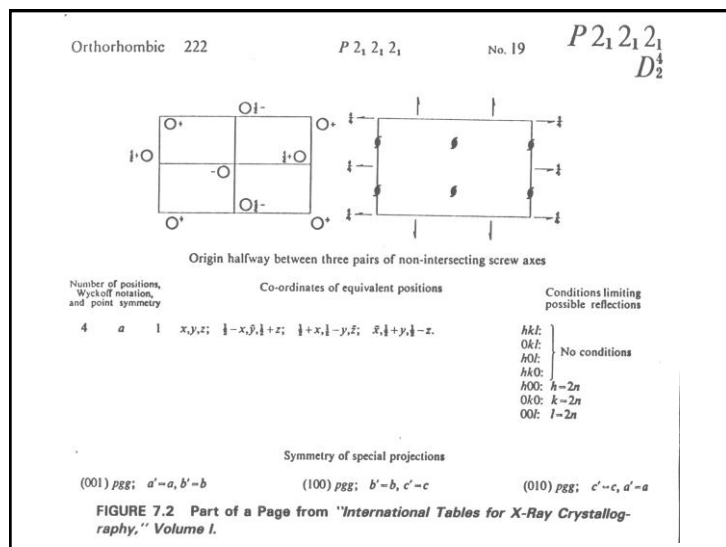


TABLE 16-5 The 65 "Biological" Space Groups

CRYSTAL SYSTEM	LAT- TICE	MINIMUM SYMMETRY OF UNIT CELL	UNIT CELL EDGES AND ANGLES <sup>a</sup>	DIFFRACTION PAT- TERN SYM- METRY <sup>b</sup>	SPACE GROUPS <sup>c</sup>
Triclinic	$P$	None	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$	$\bar{1}$	$P1$
Monoclinic	$P$ $C$	2-fold axis parallel to $b$	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ$ $\beta \neq 90^\circ$	$2m$	$P2, P2_1, C2$
Orthorhombic	$P$ $C$ $I$ $F$	3 mutually perpendicular 2-fold axes	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	$mmm$	$P222, P2_12_12_1, P222_1, P2_12_12_1, C222, C222_1, [P222, P2_12_12_1], F222$
Tetragonal	$P$ $C$ $I$	4-fold axis parallel to $c$	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	$4m$ $4/mmm$	$P4, (P4_1, P4_2), P4_3, P4_4, I4, I4_1, P422, (P4_122, P4_222), P4_322, P42_12, (P4_12_12, P4_22_12), P4_22_12, I422, I4_122, R3$
Trigonal/rhombohedral	$R^h$ $P6$	3-fold axis parallel to $c$	$a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$	$\bar{3}$ $\bar{3}m$	$P3, (P3_1, P3_2), R32, [P321, P312], [(P3_121, P3_221), (P3_12, P3_22)]$
Hexagonal	$P$	6-fold axis parallel to $c$	$a = b \neq c$ $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$	$6/m$ $6/mmm$	$P6, (P6_1, P6_2), P6_3, (P6_4, P6_5), P6_6, P622, (P6_122, P6_222), P6_322, (P6_422, P6_522)$
Cubic	$P$ $I$ $F$	3-fold axes along cube diagonals	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$	$m\bar{3}$ $m\bar{3}m$	$P2_1, P2_3, [P2_12_1, P2_32_1], P2_3, P432, (P4_132, P4_232), P4_322, I432, I4_132, F432, F4_32$