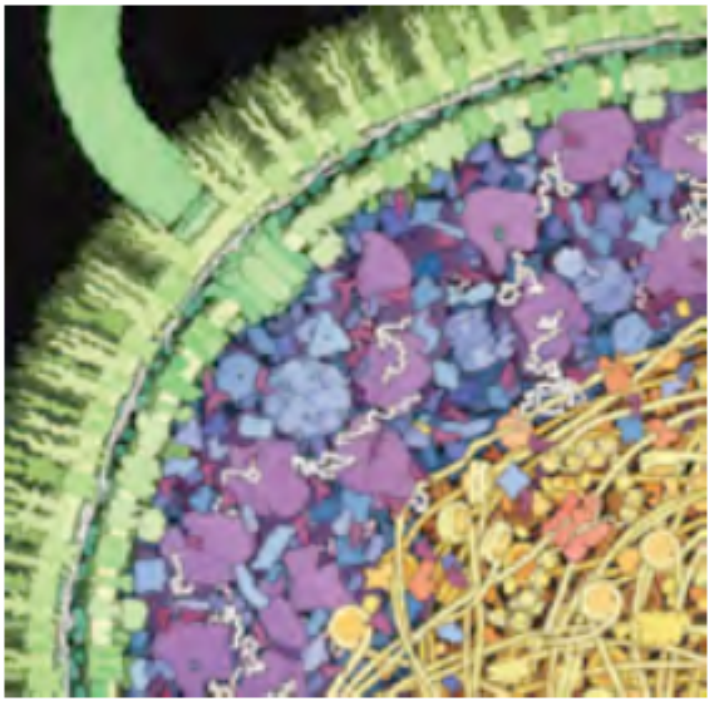
With slides from Silvia Onesti

X-ray macromolecular crystallography: an overview

Cos'è la biologia strutturale?

Una branca della scienza che si adopera per ottenere una descrizione atomica completa delle macchine molecolari in tutti i loro stati funzionali e nel loro ambiente "nativo"



Ufficialmente è nata nel 1962 con la risoluzione delle strutture atomiche del DNA e della mioglobina (la proteina dei muscoli rossi).
Da allora il lavoro dei biologi strutturali è stato riconosciuto con 45 premi Nobel per la Fisiologia e Medicina o per la Chimica.

Spaccato di una cellula di *Escherichia coli*:
proteine transmembrana e flagello in verde;
DNA genomico e nucleosomi in giallo;
proteine regolatorie in arancione;
ribosomi in viola;
proteine del citoplasma (metabolismo) in azzurro;
proteine nascenti in bianco.

Metodi sperimentali per determinare la struttura delle macromolecole

Alta risoluzione	cristallografia a raggiX | NMR | cristallografia elettronica
Media risoluzione	microscopia elettronica | diffrazione di fibre | Spettro-metria di massa | SAXS | Microscopia a Forza Atomica
Metodi Spettroscopici	NMR | dicroismo circolare | Assorbanza | Fluorescenza | Fluorescenza anisotropa | diffusione della luce
Metodi Chimici	scambio H-D | mutagenesi sito specifica | modificazioni chimiche | proteomica
Metodi termodinamici	Equilibrio di (un)folding
Metodi computazionali	predizione della struttura delle Proteine | Docking molecolare

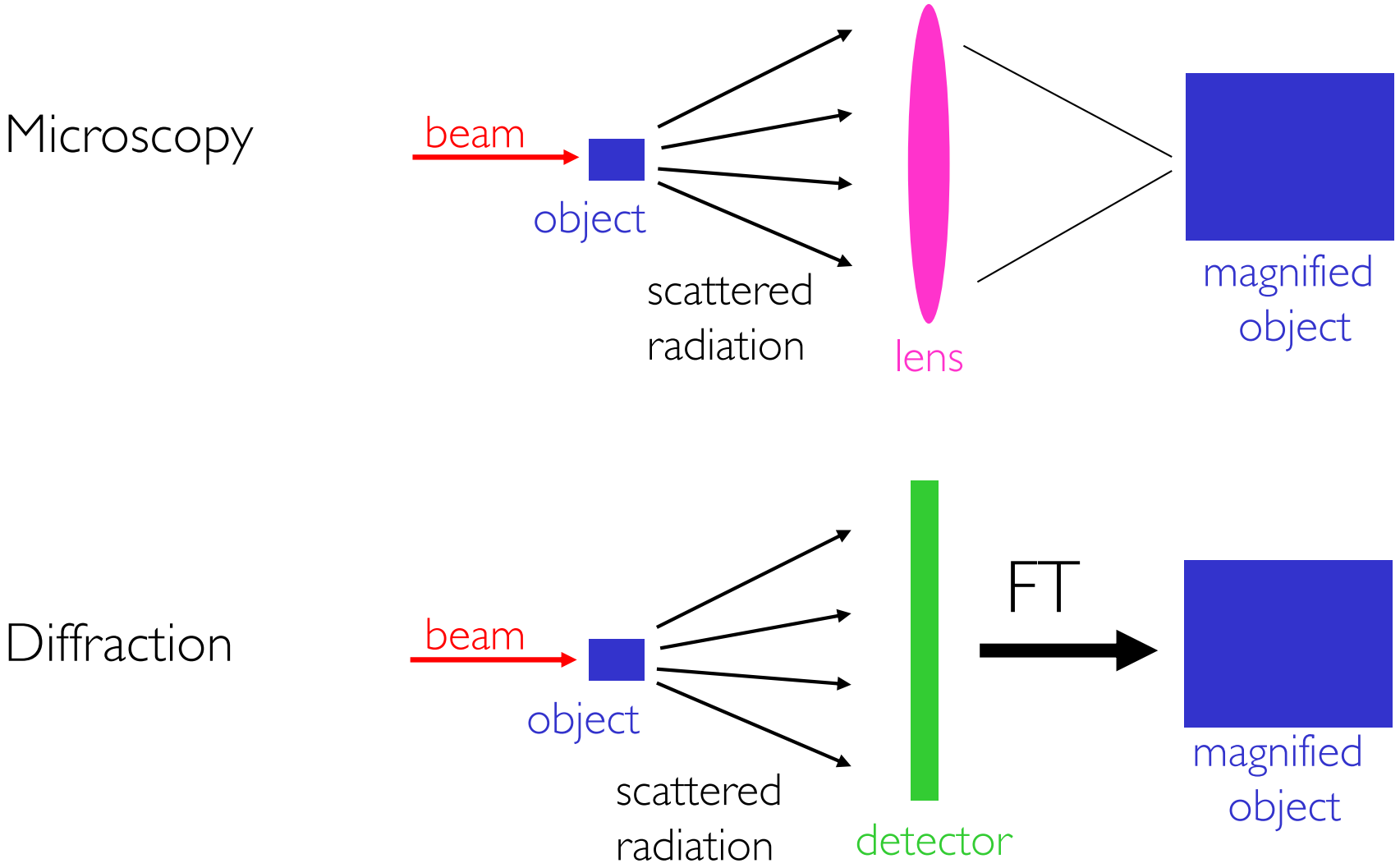
Structure determination methods

- X-ray crystallography (protein crystallography)
- X-ray fiber diffraction
- Small angle X-ray scattering (SAXS)

- Scanning electron microscopy (SEM)
- Transmission electron microscopy (TEM):
  - 2D crystals (electron diffraction)
  - Fibers/helices/tubular crystals
  - Single particle electron microscopy
- Atomic Force Microscopy (AFM)

- Nuclear magnetic resonance (NMR)

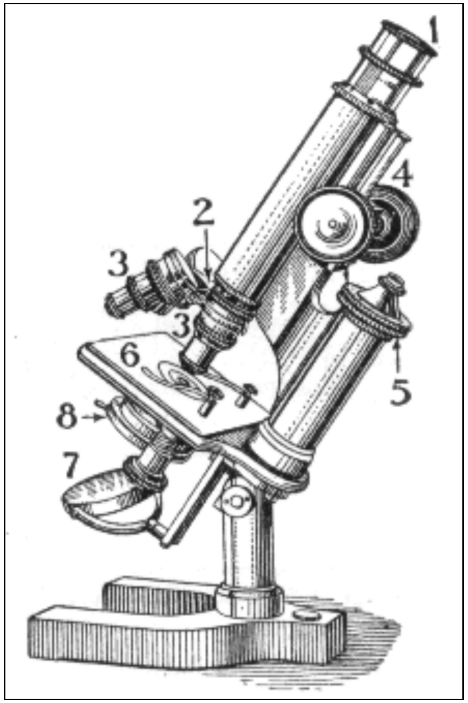
Microscopy vs diffraction

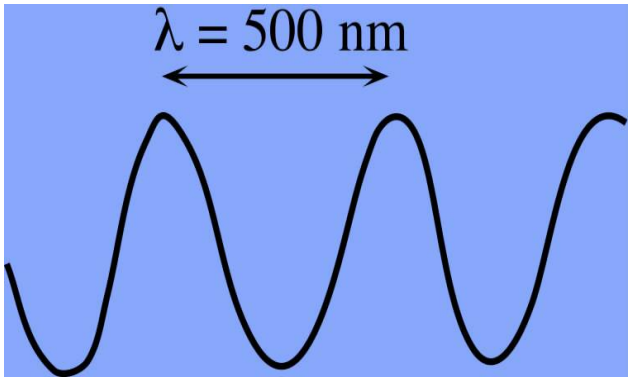


Why can't we use a microscope?

Normally, to look at small objects we use microscopes...

...but they can only provide images of things larger than the wavelength of light





Protein (10 nm)

Atom (0.1 nm=1Å)

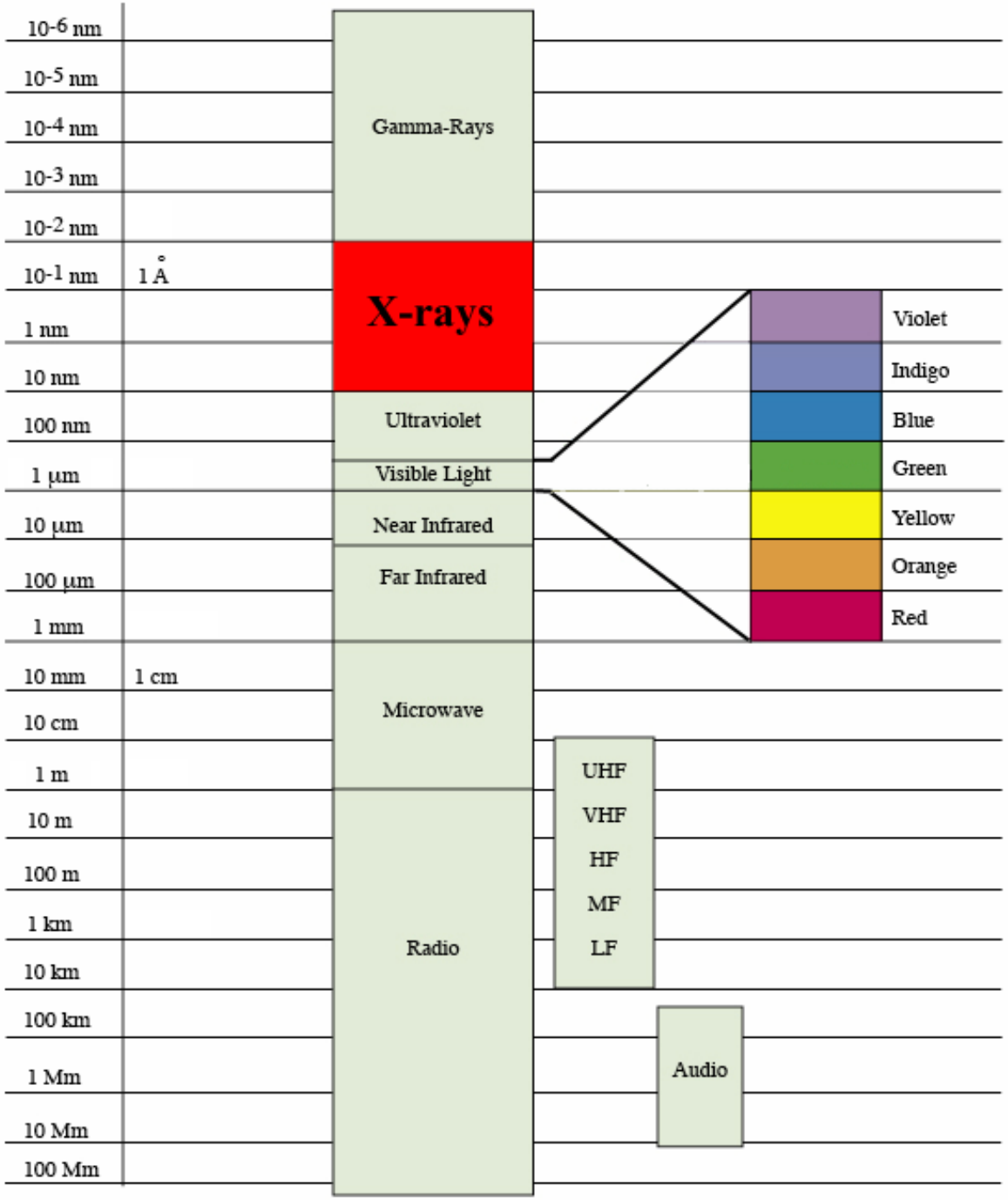
In theory, we could use X-rays (light of λ=0.1 nm, the right size for looking at atoms)...

X-rays

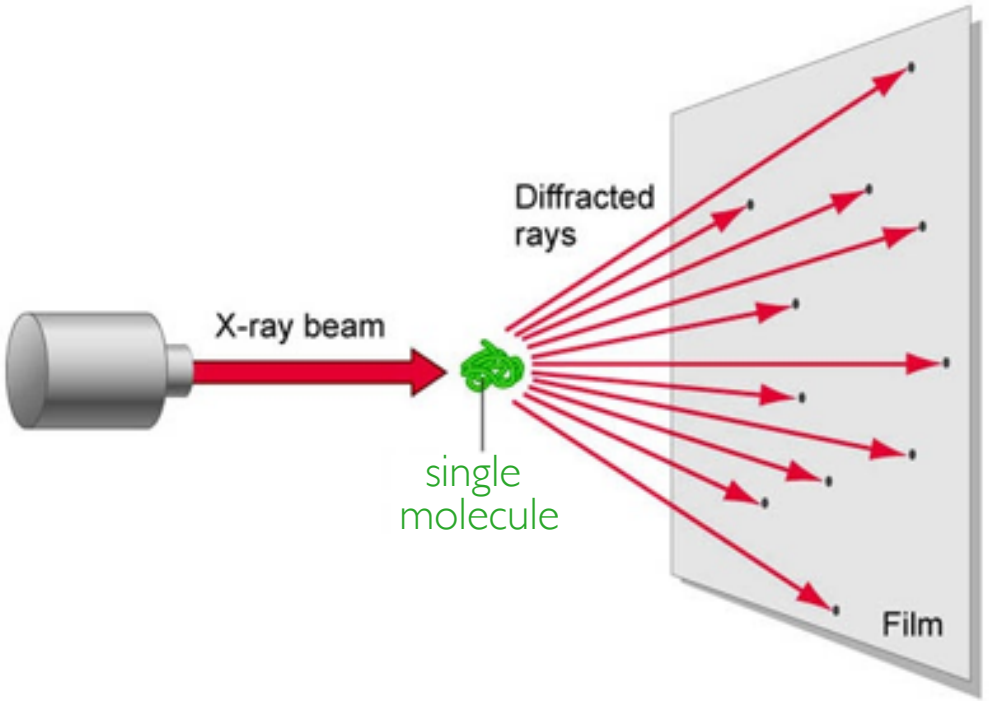
X-rays:
E = 1-100 keV
λ = 10 nM-0.01 nm

Soft X-rays (lower E) are used in medical imaging.
Hard X-rays (higher E) are used in crystallography.

...but we can't make X-ray microscopes as hard X-rays do not reflect or refract easily.



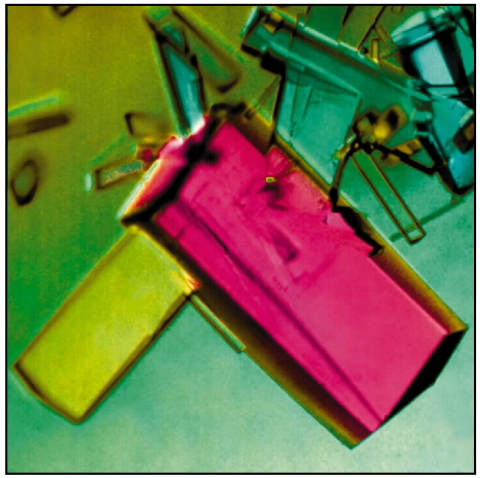
X-ray diffraction with single objects?



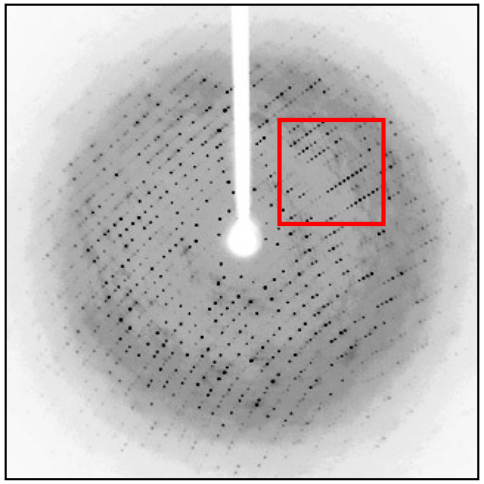
The intensity of the X-ray radiation diffracted by a single biological object (such as a protein or a cell) is very small:

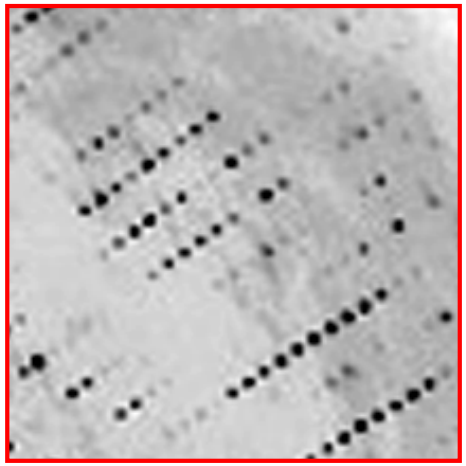
➡ almost impossible to measure with current technology.

But we can use X-rays with crystals



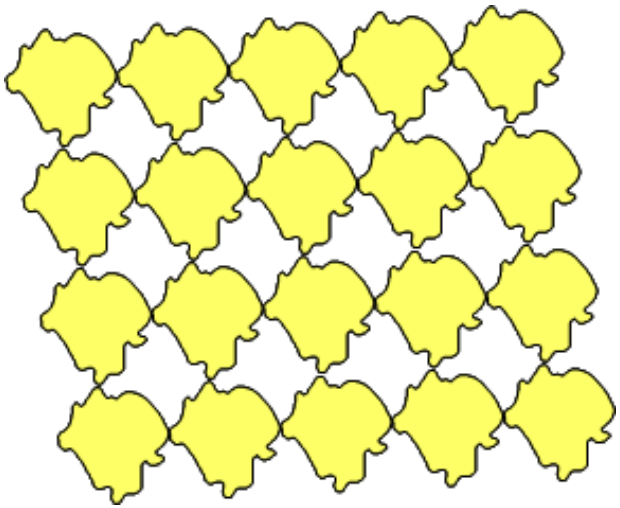
Crystal





X-ray Diffraction pattern

Crystal: an ordered array of molecules



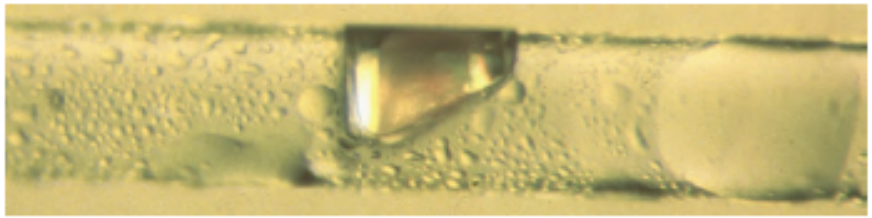
Two effects on diffraction:

- constructive interference in some directions, so that all molecules scatter in phase: strong signal
- destructive interference in most directions, so that the diffraction pattern is non-zero only at a few specific position: discrete spots

Raccolta dati

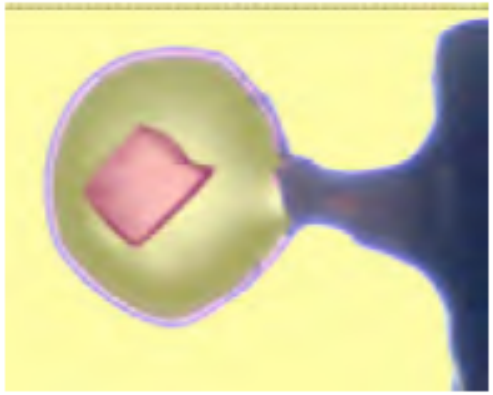
Temperatura ambiente
(293-298K)

Il cristallo viene
montato in un capillare



Temperature criogeniche
(50 - 100K)

Il cristallo, protetto con un anticongelante, viene montato in un loop di nylon e raffreddato sotto flusso di N_2 o He liquidi

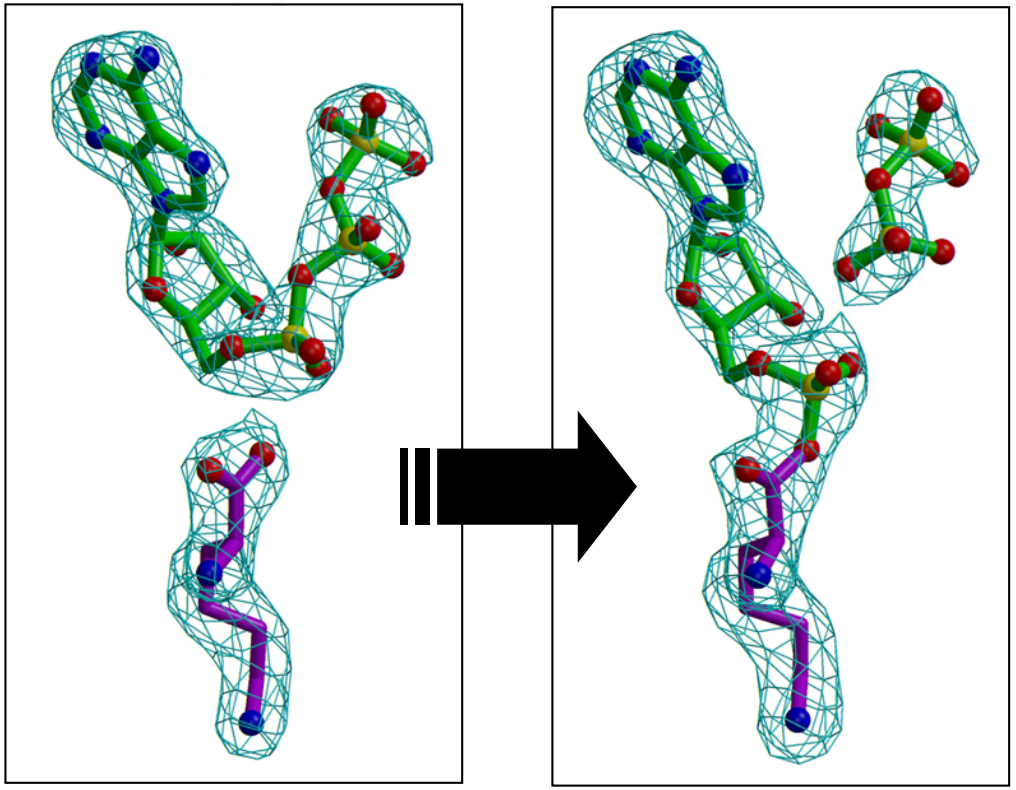


Does crystallisation affects the structure?

- The forces that hold molecules in a lattice are very much weaker than those that hold protein structures together, so gross conformational changes triggered by crystallisation are unlikely (but they do occur). However the crystal lattice may favour a conformation that is not dominant in solution.

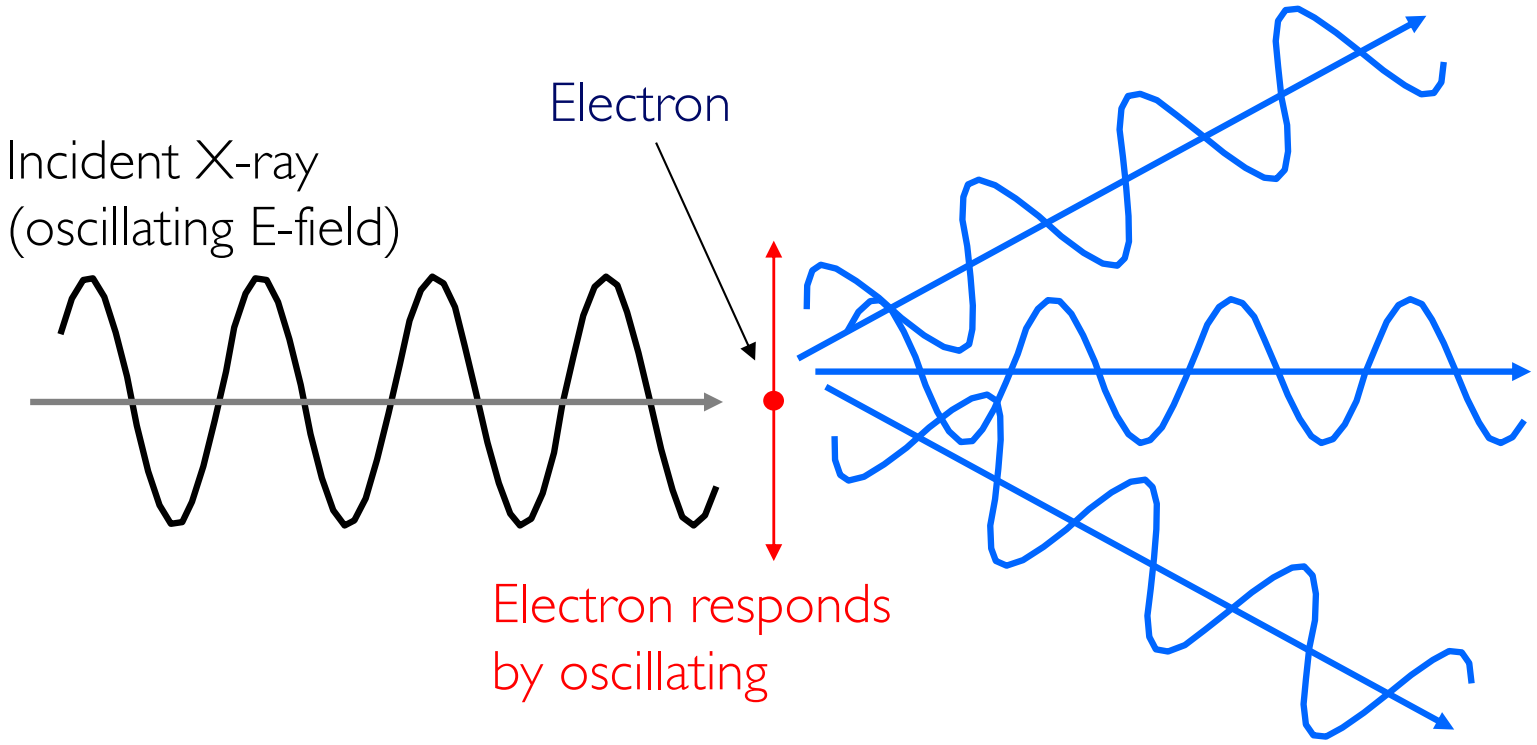
- Proteins crystallised in different crystal forms are often identical or almost identical.

- Some enzymes retain their biological activity within the crystal (a strong indication of a native-like structure).



X-ray scattering by a single electron

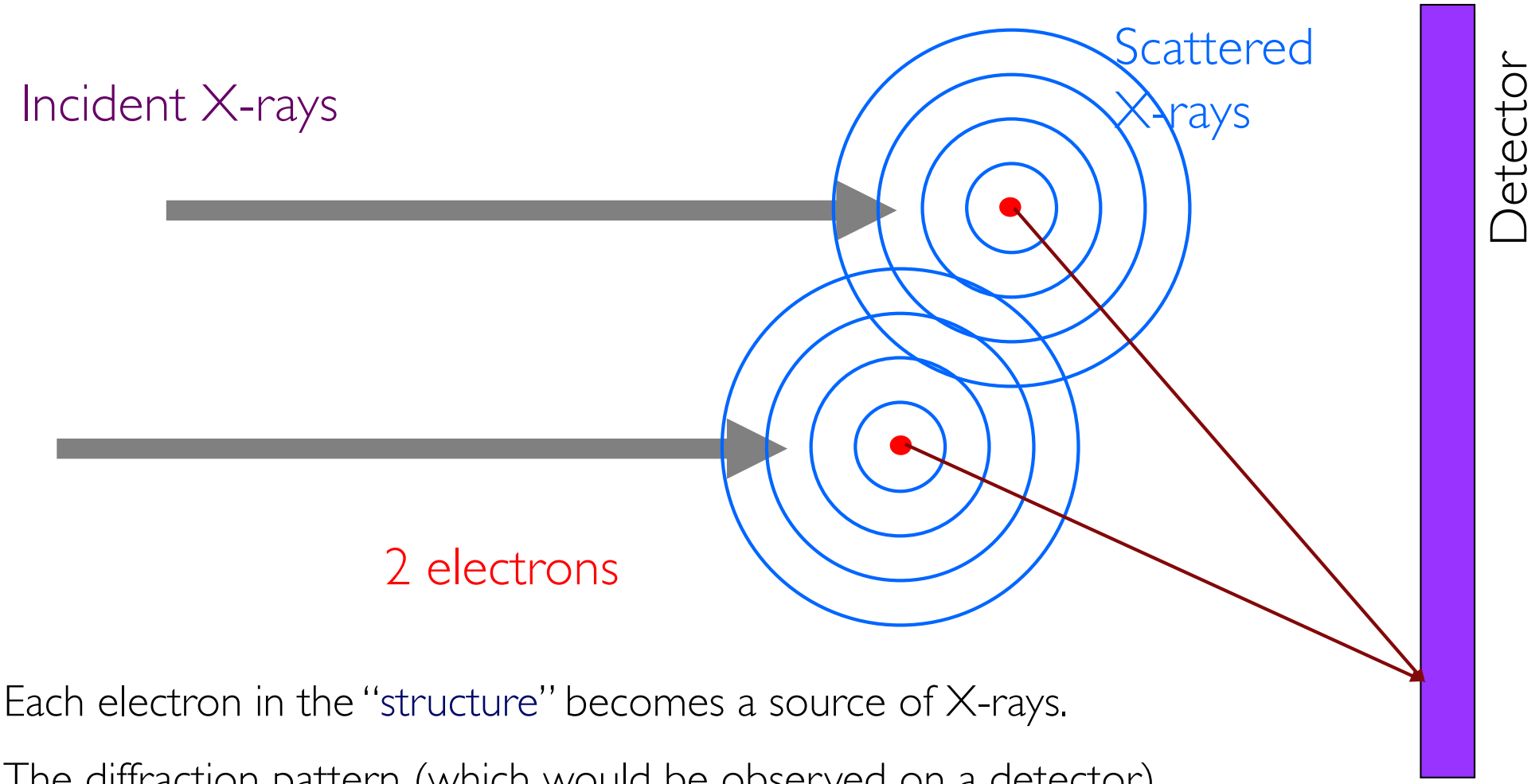
➡ the scattering of X-rays is mostly due to electrons



The electron is said to "scatter" or "diffract" the X-ray

Oscillating electron emits X-rays over a wide angle

X-ray scattering by two electrons



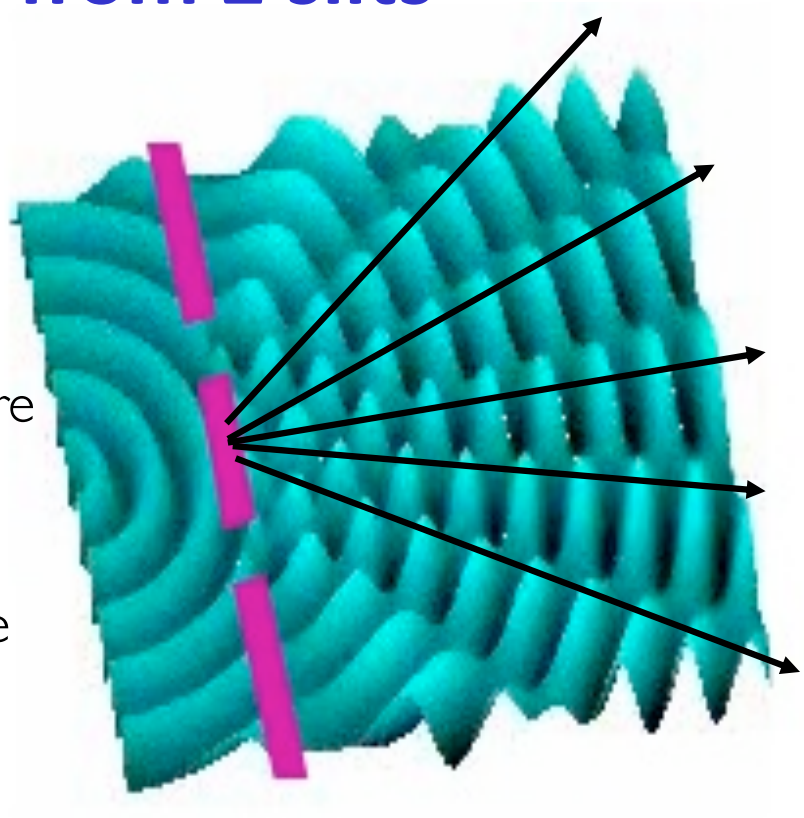
Each electron in the "structure" becomes a source of X-rays.

The diffraction pattern (which would be observed on a detector) is the resultant of adding the scattered X-ray waves.

To understand diffraction, we need to know how to add waves, i.e. how waves interfere with each other.

Example of diffraction from 2 slits

Diffraction of water waves through 2 slits gives a pattern that depends on the slit structure (width, separation) and the wavelength of the waves.

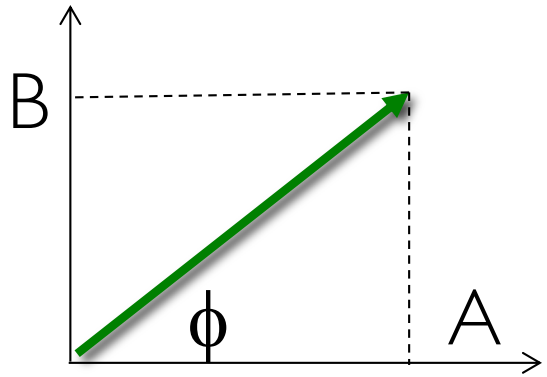


If we can work out this relationship, we can measure the diffraction pattern and figure out the "structure" that gives rise to it.

The observed pattern is the resultant of adding the wave arriving from the 2 slits at every point on the detector

Waves can be written as complex numbers/vectors:

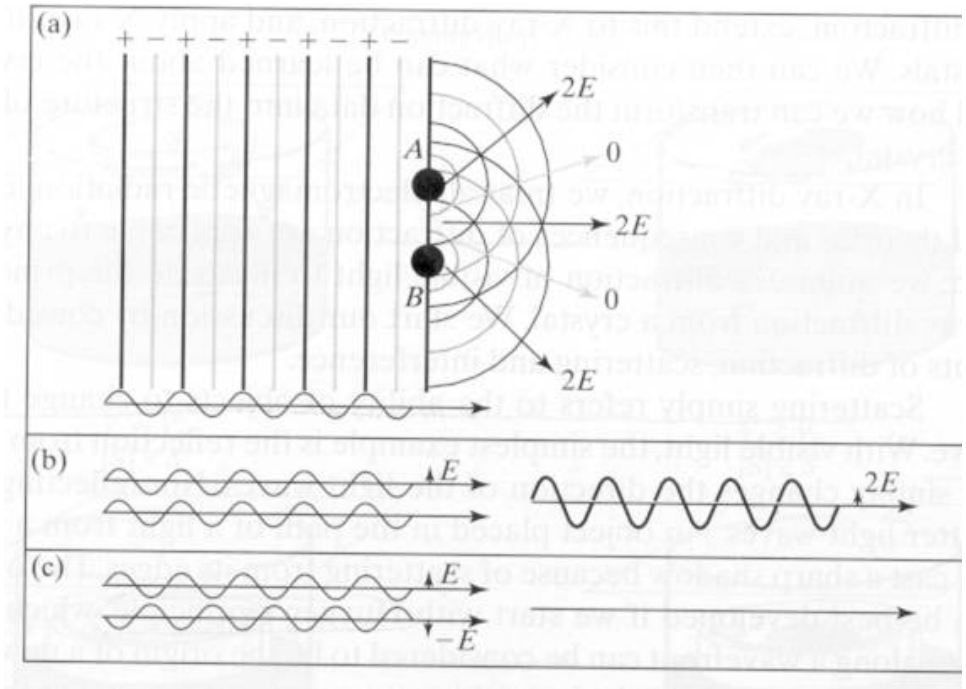
$$\mathbf{Z} = |Z| \exp(i\phi) = |Z|(\cos\phi + i\sin\phi) = A + iB$$



X-ray scattering: interference

•Scattering: ability of an object to change the direction of a wave.

•If two objects (A and B) are hit by a wave they act as a point source of a new wave with same wavelength and velocity (Huygens' principle)



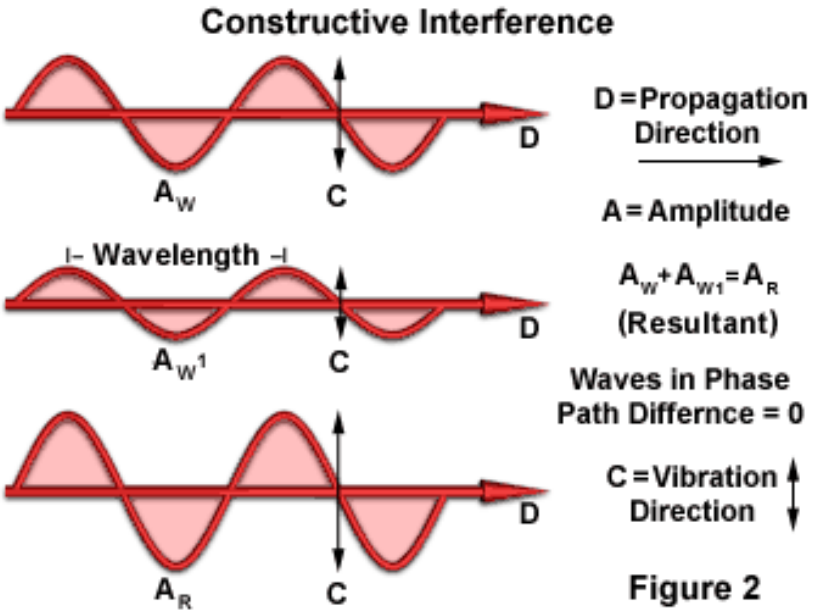


Figure 2

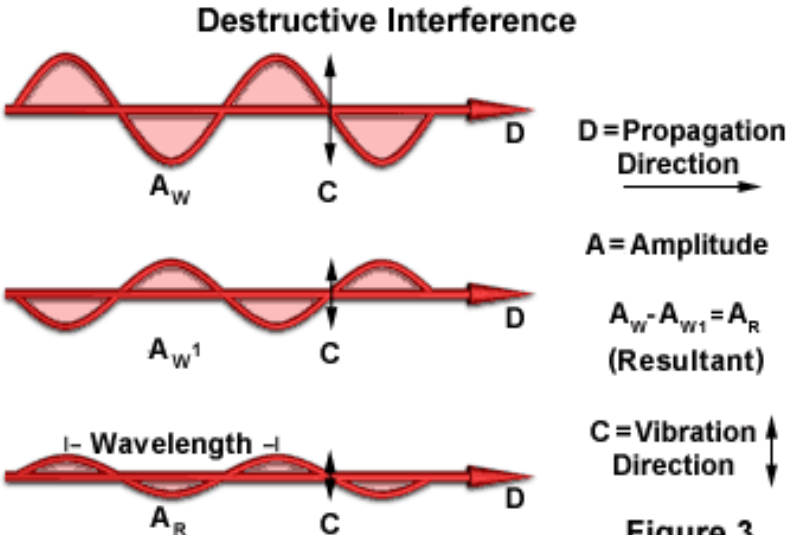
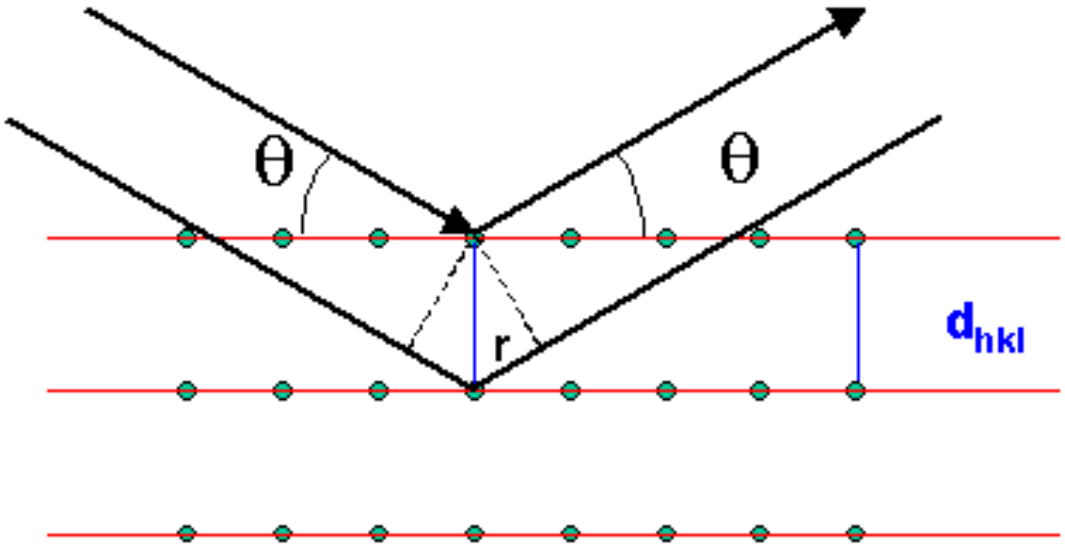


Figure 3

A practical approach: Bragg diffraction



X-rays reflected by crystal planes families

This is not physically correct: the real process is a scattering + interference
But helps in visualizing the 3D space

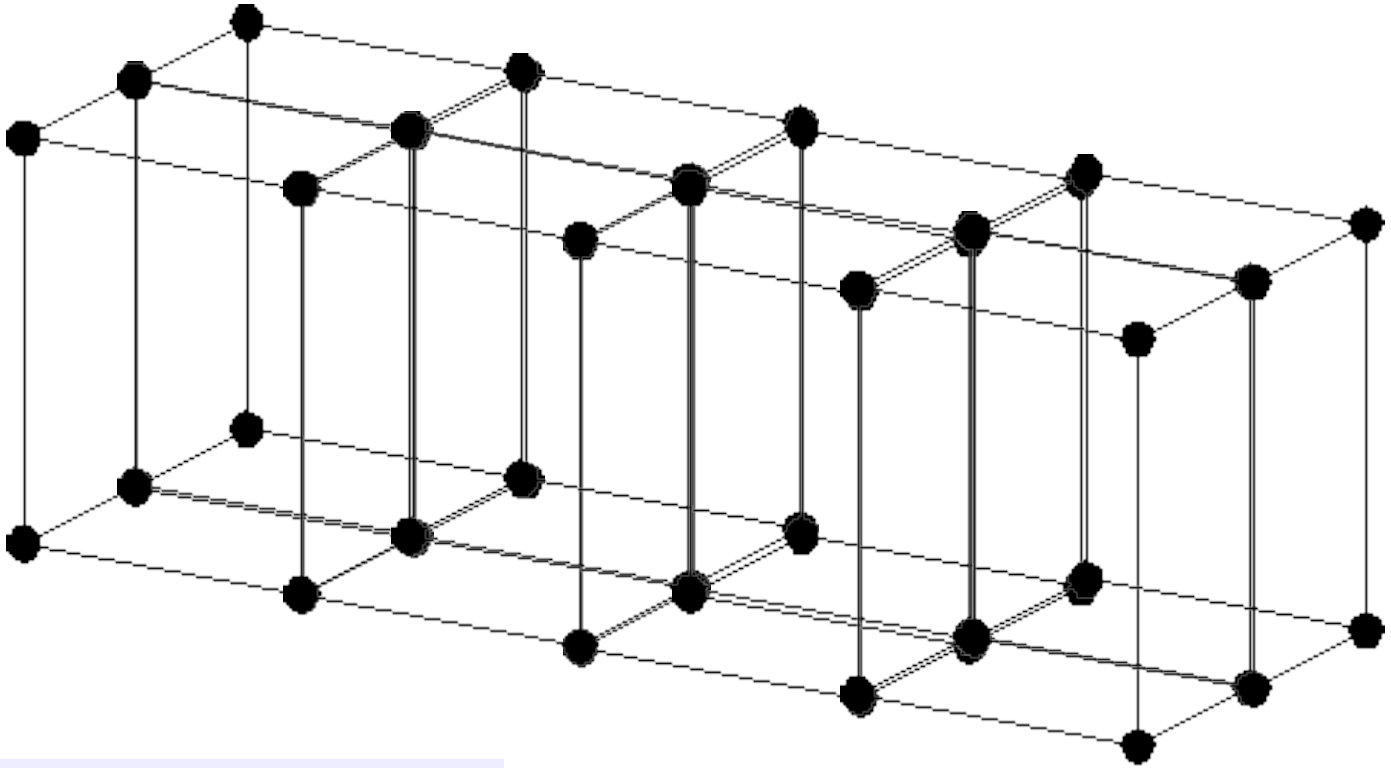
$$2d_{hkl} \sin(\theta) = n \lambda$$

Bragg's law

$1/d = (2/\lambda) \sin\theta$
$1/d \propto \sin\theta$

When d is large, diffraction pattern is compressed

Crystals: unit cells

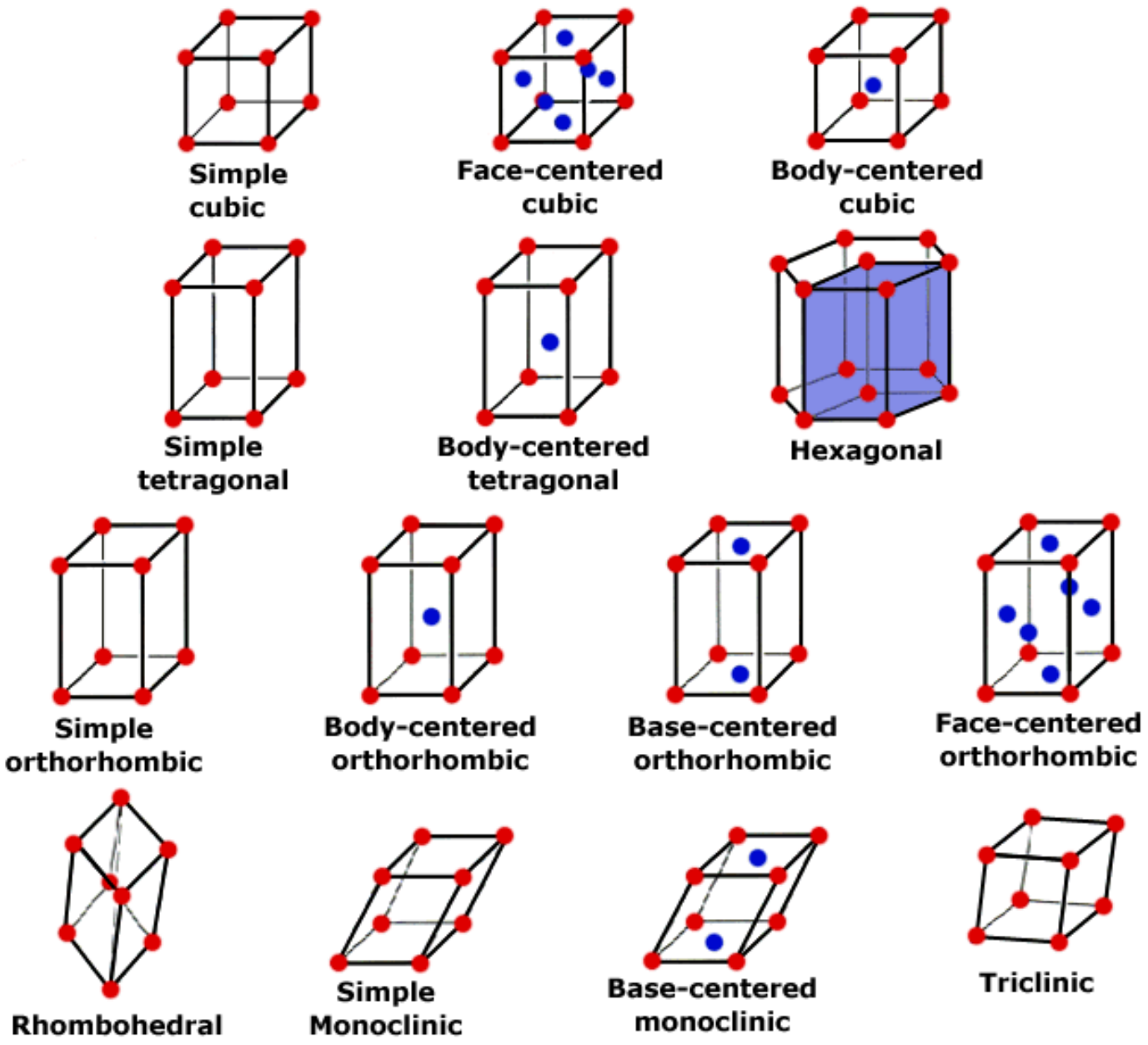


$$\vec{T} = u\vec{a} + v\vec{b} + w\vec{c}$$

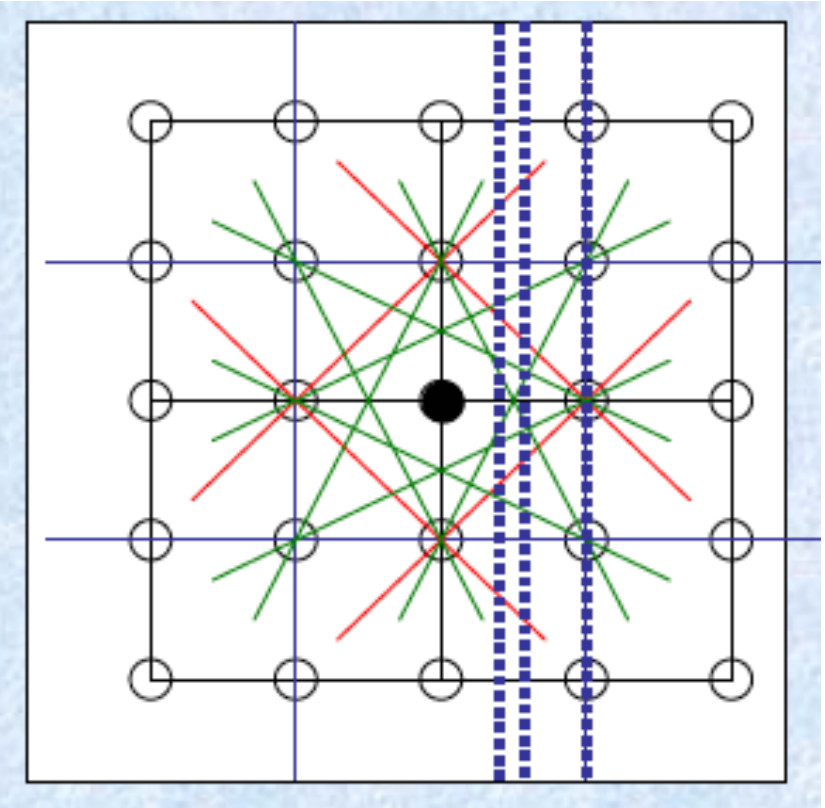
u, v, w interi

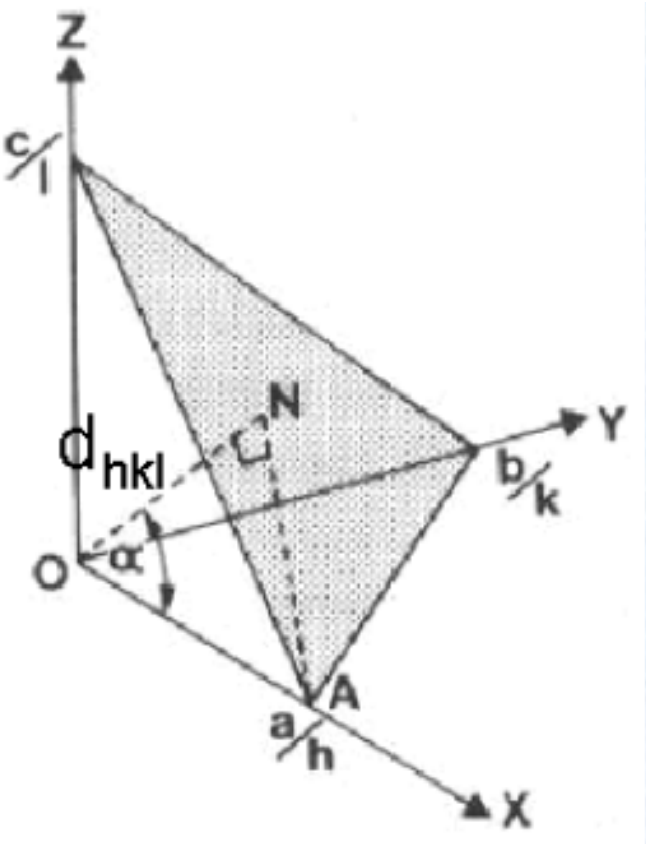
$\vec{T}$ is a vector of the Bravais Lattice

The 14 Bravais Lattices



The reciprocal space



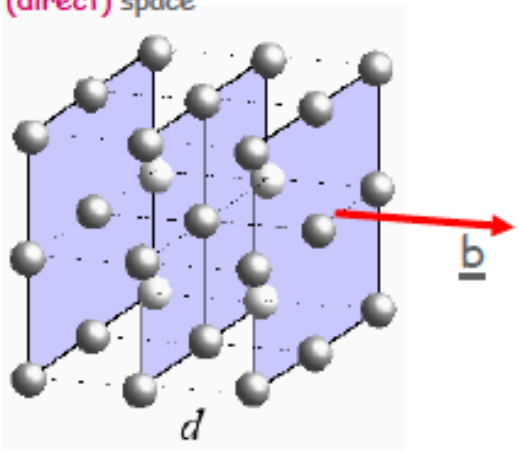


To design all the possible plane families is in fact too complex.
Instead, a family of planes is represented by the d^*_{hkl} vector,
i.e. a unique point in the reciprocal space!

$$d_{hkl}^* = K / d_{hkl}$$

The reciprocal space

Crystallographic planes
in real (direct) space

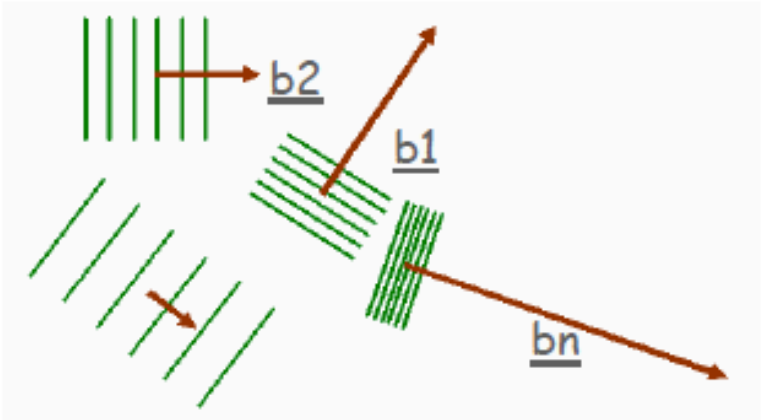


The red arrows indicates the direction orthogonal to the family of planes (vector b).

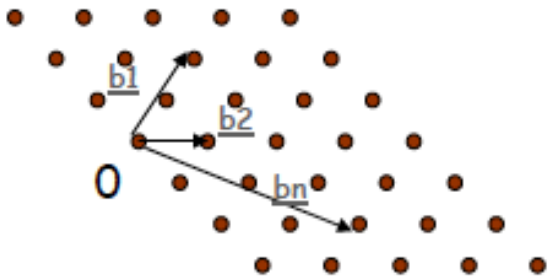
All planes of a family has in common the othogonal direction

A family of planes can be described by a vector b

It is possible to define a new space that contains vectors b1, b2…bn, that are orthogonal to crystallographic planes.

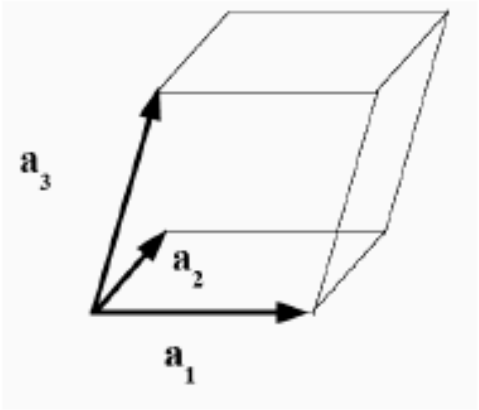


Real (direct) space



Reciprocal space

The reciprocal space (more rigorous)



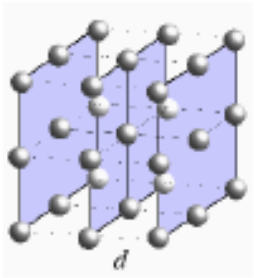
$$\mathbf{b}_1 = \frac{2\pi}{V}\mathbf{a}_2 \times \mathbf{a}_3$$

$$\mathbf{b}_2 = \frac{2\pi}{V}\mathbf{a}_3 \times \mathbf{a}_1, \quad \mathbf{b}_3 = \frac{2\pi}{V}\mathbf{a}_3 \times \mathbf{a}_1$$

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\,\delta_{ij}$$

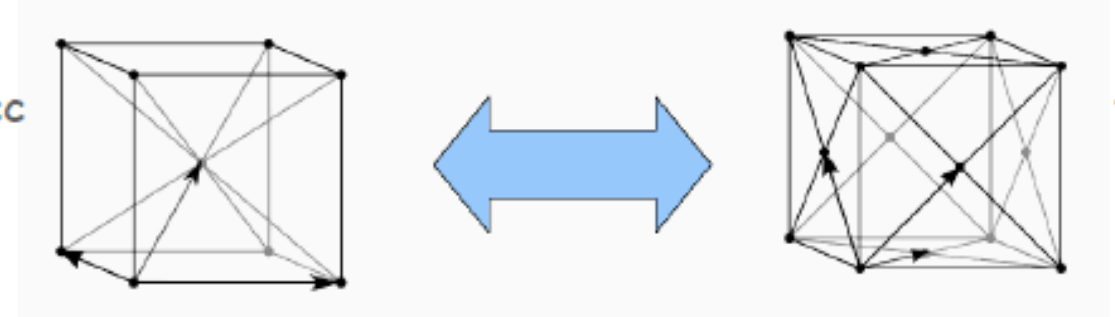
A family of planes is fully described in reciprocal space by a vector $\underline{K}$

- $\underline{K}$ is orthogonal to the planes
- $|\underline{K}| = 1/d_{hkl}$



Example:

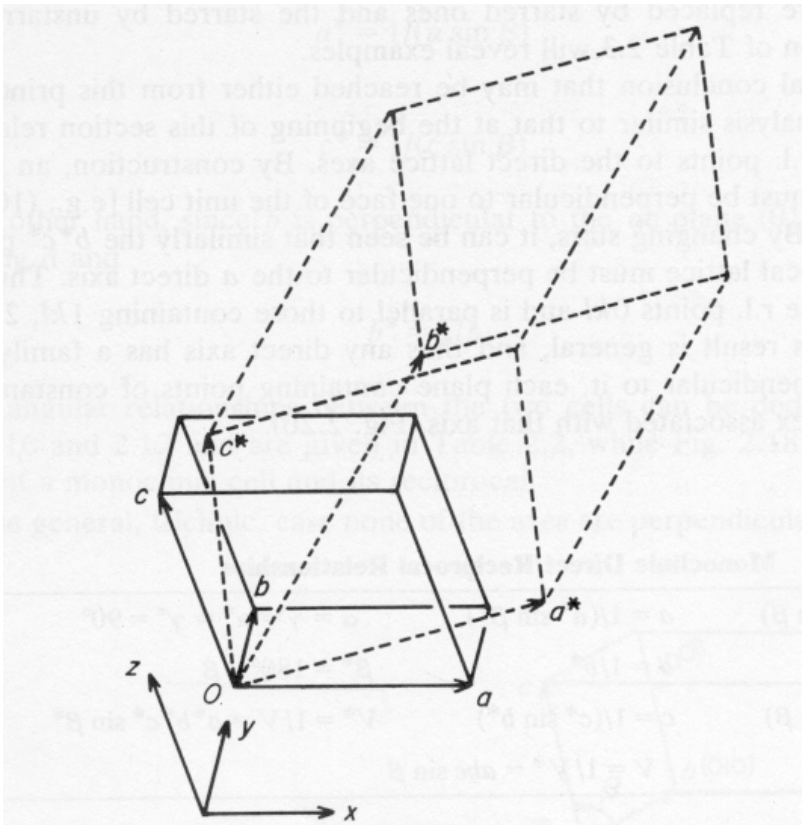
bcc



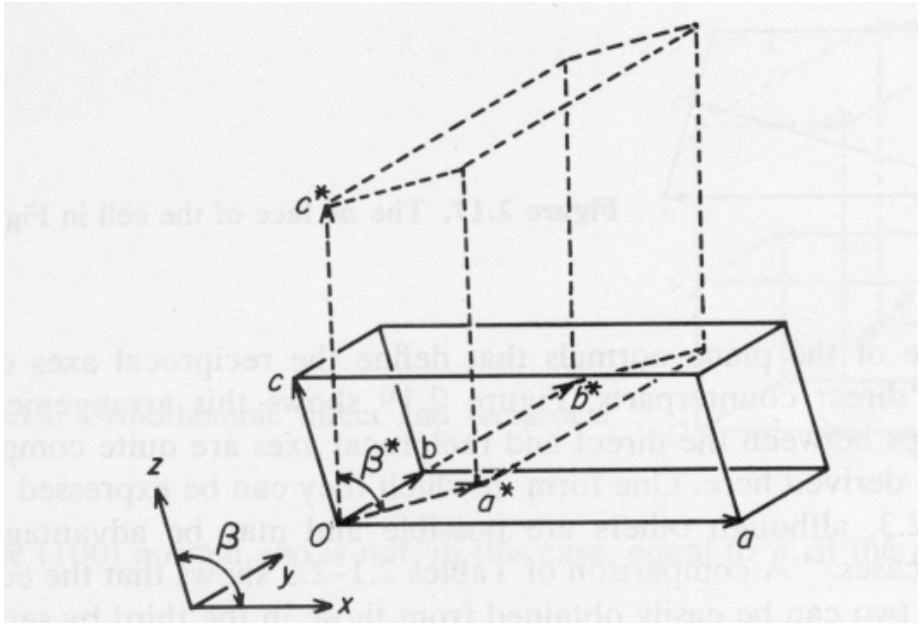
fcc

Examples of direct and reciprocal lattices

(Figures from Jensen and Stout "X-Ray structure determination. A pratical guide"

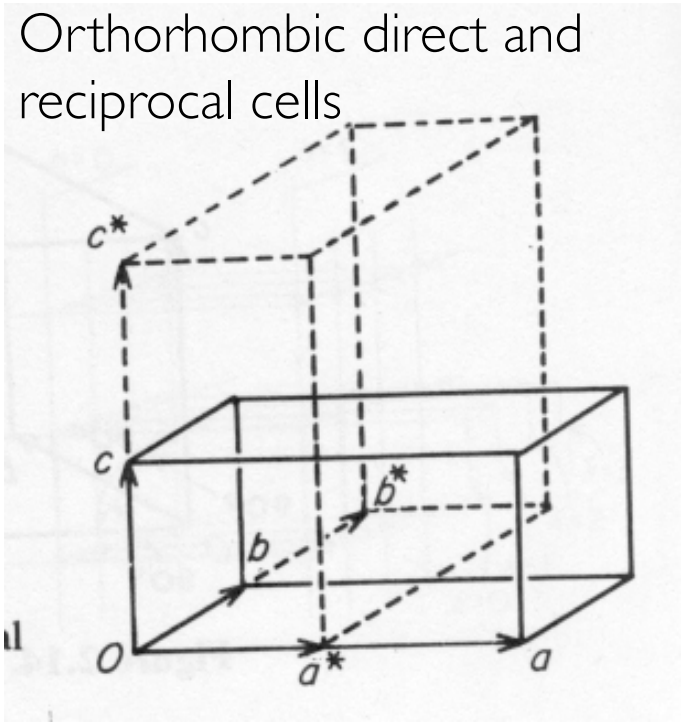


Triclinic direct and reciprocal cells

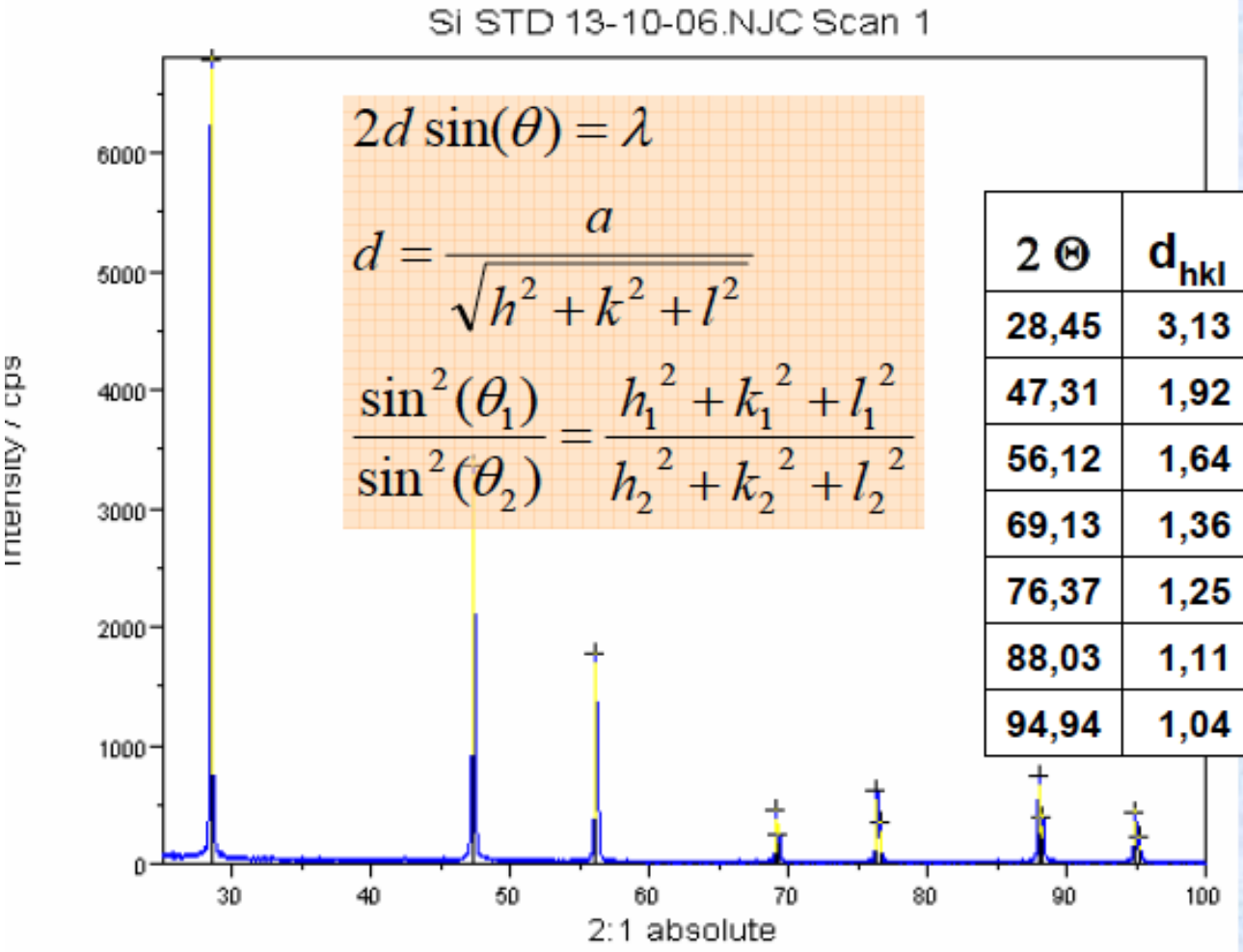


Monoclinic direct and reciprocal cells

Orthorhombic direct and reciprocal cells



Dai valori angolari a cui si osservano i riflessi di Bragg è possibile ottenere le informazioni sulla forma e dimensione della cella unitaria



$$2d\sin(\theta) = \lambda$$

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

$$\frac{\sin^2(\theta_1)}{\sin^2(\theta_2)} = \frac{h_1^2 + k_1^2 + l_1^2}{h_2^2 + k_2^2 + l_2^2}$$

2 Θ	d_{hkl}	sen²	sen²/ sen²			hkl
28,45	3,13	0,06	1,00	2,00	3,00	111
47,31	1,92	0,16	2,66	5,33	7,99	220
56,12	1,64	0,22	3,66	7,33	10,99	311
69,13	1,36	0,32	5,33	10,66	15,99	400
76,37	1,25	0,38	6,33	12,65	18,98	331
88,03	1,11	0,48	7,99	15,99	23,98	422
94,94	1,04	0,54	8,99	17,98	26,97	333

Morphology of a crystal

- By observing the **spacing and pattern of reflections on the diffraction pattern**, we can determine the **lengths, and angles** between each side of the **unit cell**, as well as the **symmetry or space group** in the unit cell.

- Still, how do we find out what's **inside** the unit cell? (i.e. the interesting stuff, like proteins).

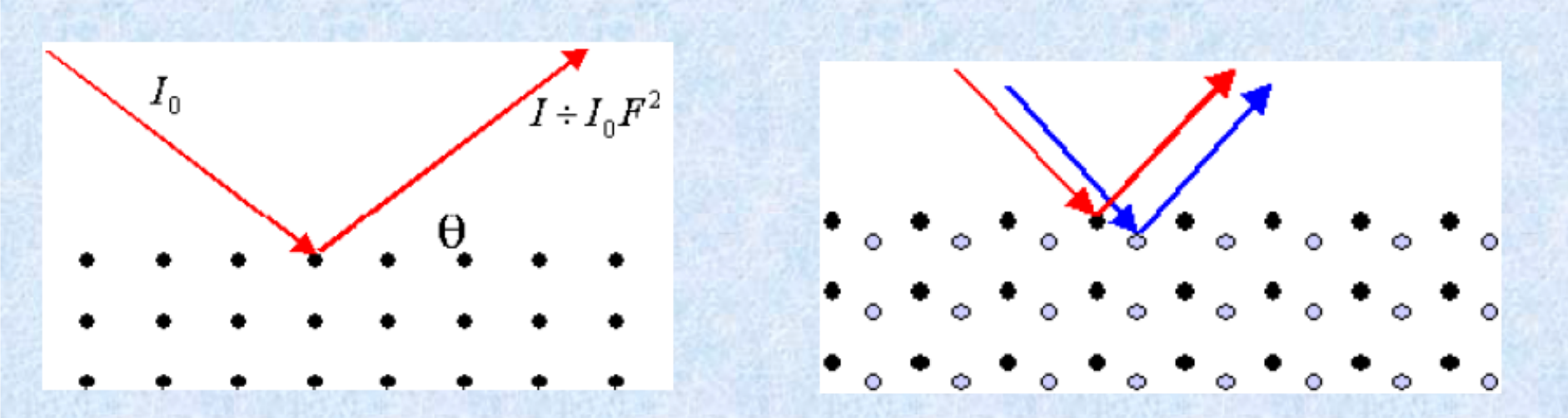
Solving macromolecular structures by X-ray diffraction

- Now assume that we have more than one atom in a unit cell→ the diffraction angle θ will remain the same!! The diffraction angle is still related to the unit cell dimension by Bragg's law.

- **However**, with one atom, the phase and amplitude of the resultant wave from each plane (atom) was the same. With more than one atom, the phase and amplitude coming from each plane of different atoms may be different, resulting in **different intensities** of the reflections.
- → Deconvolute each reflection into the phase and amplitude contributions from each of the individual reflections from each atom in the molecule.

# of atoms in unit cell does NOT affect reflection angle (Bragg's law).
# of atoms (electron density) in unit cell DOES affect intensity of reflection spots.

Is Bragg's law still valid for two or more atoms in a unit cell?



Two atoms in a unit cell (reflect) waves from their respective planes.

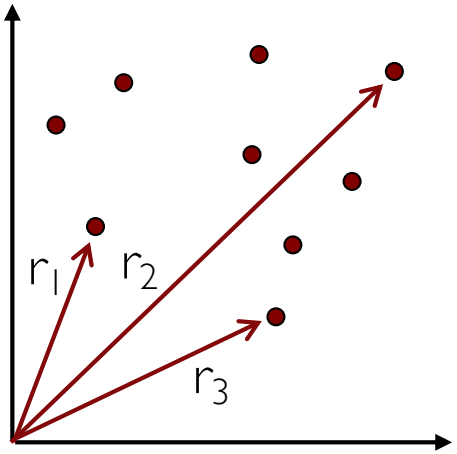
The waves combine and form a resultant wave, that looks like it has been reflected from the original unit cell lattice plane.

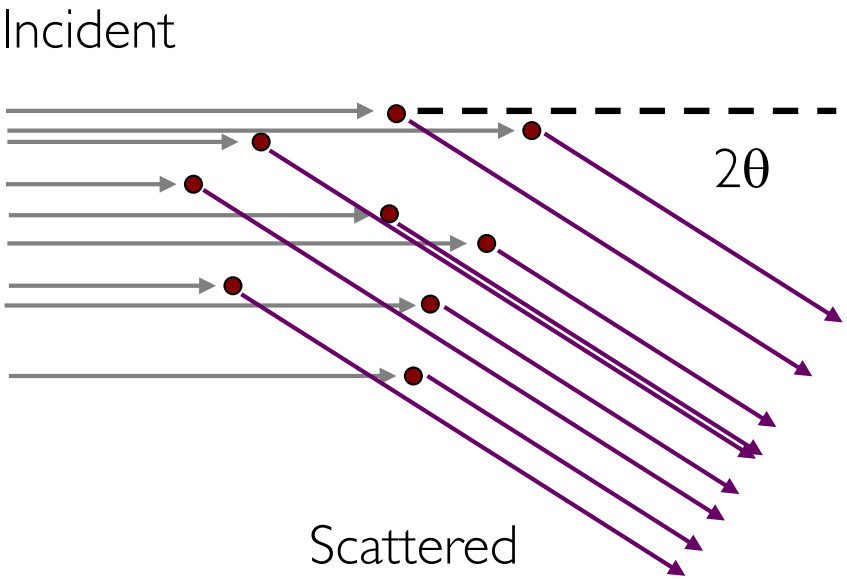
The 2 lattice are the same, but shifted one with respect to the other

Diffraction spot is in the same place, but has different intensity (intensity of resultant wave).

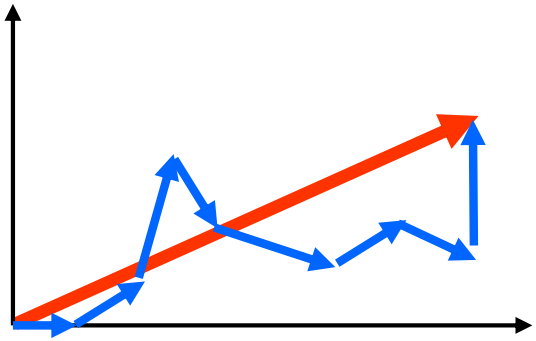
We assumed the electron density is in planes. In reality it is spread throughout the unit cell. Nevertheless, the derivation is still valid, since it can be shown that waves scattered from electron density not lying in a plane P, can be added to give a resultant as if reflected from the plane.

Scattering from a molecule:





$$f(S) = \sum_{j=1}^{N} A_j e^{i 2\pi r_j \cdot s}$$



It is like adding many waves/vectors with magnitude A_j and phase (rj.s) - I can calculate this from the wavelength, angle and the positions of the scatterers.

In practise I can consider all the electrons in one atom as a "block" – I sum together the contributions of the atoms.

Moving from fixed points to electron density inside a cell

- → The sum is replaced by an integral over the **volume of the unit cell, V.**
- → the amplitude f_i is replaced by the **electron density ρ(r)** (the more density the larger the amplitude).

$$F(hkl) = F\left(\vec{S}\right) \sim \int_V \rho\left(\vec{r}\right) \cdot e^{2\pi i \vec{S}\cdot\vec{r}}\, dV$$

This is very nice now: If we know the electron density r(r) (i.e. positions of atoms) inside unit cell, we can predict the intensity, of each reflection spot (h, k, l) on the film

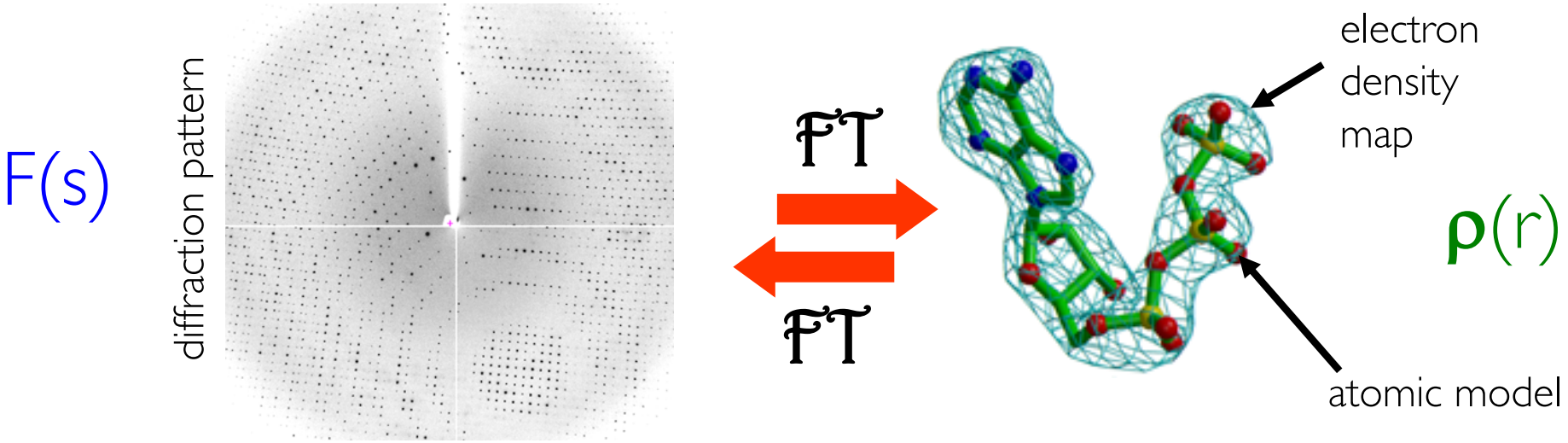
$$I\left(\vec{S}\right) = I(h,k,l) \sim \left|F(h,k,l)\right|^2$$

Knowing the size of the unit cell, and knowing the electron density, we can calculate (predict) the location and intensity of each diffraction spot.

Diffraction and Fourier synthesis

$F(s) = \mathcal{FT}[\rho(r)]$ the diffraction pattern is the Fourier transform of the electron density

$\rho(r) = \mathcal{FT}[F(s)]$ the electron density is the Fourier Transform of the diffraction pattern



We can always go back and forward using Fourier transforms:

- if we know the electron density we can calculate the diffraction pattern
- if we know the diffraction pattern we can calculate the electron density

Moving from the structure factor to the electron density (final goal).

- We need to do the 'Fourier transform' of the structure factor to get the electron density out.

$$\rho(\vec{r}) \sim \frac{1}{V} \int_{V^*} F(\vec{S}) \cdot e^{-2\pi i \vec{S} \cdot \vec{r}} \, dV^*$$

Now, the diffraction pattern is actually made up of points, so the integral becomes a sum (Fourier series) over all diffraction spots:

$$\rho(\vec{r}) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) \cdot e^{-2\pi i \vec{S} \cdot \vec{r}}$$

And that is what we want!!

… but wait, there is a problem

…the phase problem

- We cannot measure F(hkl), only |F(hkl)|.
- We can measure the intensity I(hkl) of the spots

$$I(\vec{S}) = |F(hkl)|^2 = |F(\vec{S})|^2 = F(\vec{S}) \cdot F(\vec{S})^* = F_{re}^{\ 2} + F_{im}^{\ 2}$$

$$F(hkl) = |F(hkl)| \cdot e^{i\alpha_{hkl}}$$

$$\rho(\vec{r}) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \cdot e^{-2\pi i \vec{S} \cdot \vec{r}} e^{i\alpha_{hkl}}$$

We only know the magnitude but not the direction (phase α) of F(S).

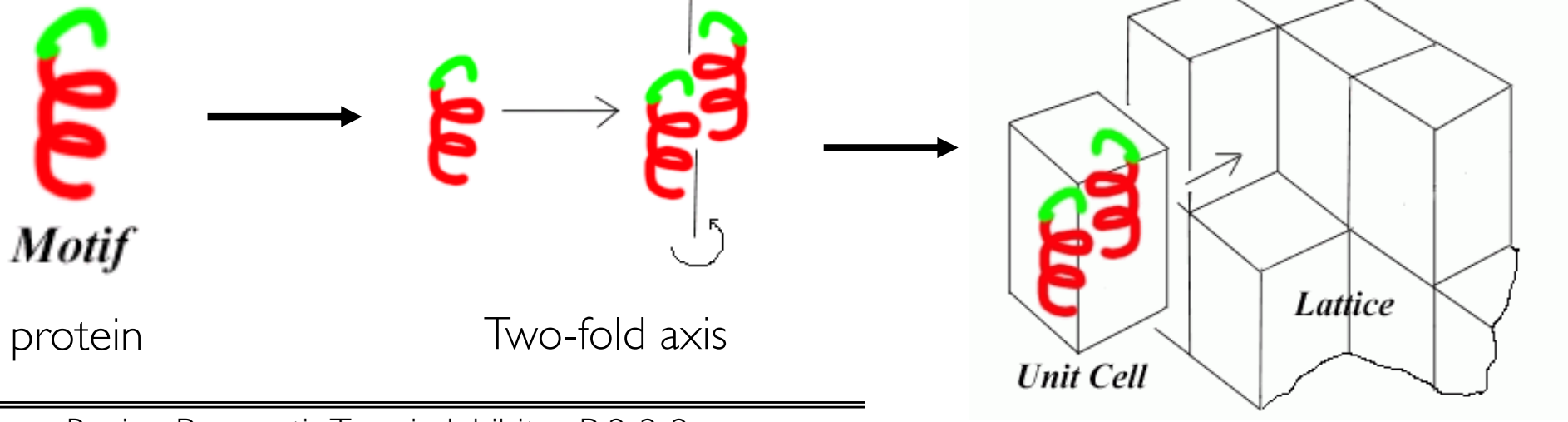
We have lost the phase.

How bad is that? Pretty bad, because we can calculate the "magnitude" of the electron density, but we don't know where it is in the unit cell.

→ What can we do??

Examples

(adapted from Bernhard Rupp, University of California-LLNL)
http://www-structure.llnl.gov/Xray/tutorial/Crystal_sym.htm

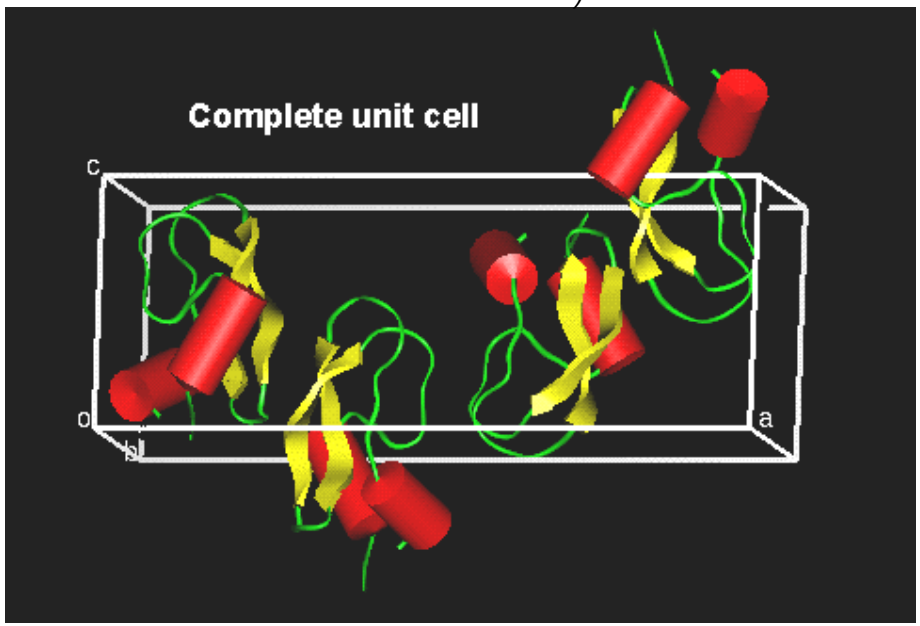


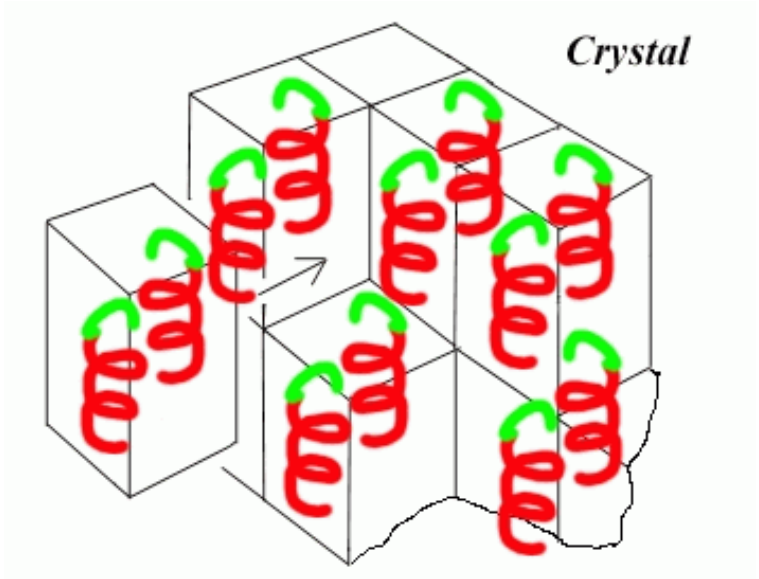
protein

Two-fold axis

Bovine Pancreatic Trypsin Inhibitor P $2_12_12_1$

(Primitive, orthorhombic unit cell with three two-fold screw axes)





The space groups

- What **symmetry** operations (e.g. rotation axes, (2-, 3-, 4-, 6- fold axis, mirrors, inversion axes …, at corner, at face, …) can be applied to the unit cell (inside crystal)? This defines the **32 point groups** of the unit cell.

Table 1.2 *Symmetry elements*

	Symmetry element	Hermann–Mauguin symbols (crystallography)	Schönflies symbols (spectroscopy)
Point symmetry	Mirror plane	m	σ_v, σ_h
	Rotation axis	$n = 2, 3, 4, 6$	$C_n (C_2, C_3, \text{etc.})$
	Inversion axis	$\bar{n} (= 1, 2, \text{etc.})$	—
	Alternating axis*	—	$S_n (S_1, S_2, \text{etc.})$,
	Centre of symmetry	$\bar{1}$	i
Space symmetry	Glide plane	a, b, c, d, n	—
	Screw axis	$2_1, 3_1$, etc.	—

* The *alternating axis* is a combination of rotation (n-fold) and reflection perpendicular to the rotation axis. It is little used in crystallography.

The space groups

- What **symmetry** operations (e.g. rotation axes, (2-, 3-, 4-, 6- fold axis, mirrors, inversion axes …, at corner, at face, …) can be applied to the unit cell (inside crystal)? This defines the **32 point groups** of the unit cell.

- The combination of the 32 symmetry types (point groups) with the 14 Bravais lattice, yields **230 distinct space groups**.

Table 7.4
ENUMERATION OF SOME SIMPLE SPACE GROUPS

SYSTEM	NUMBER OF POINT GROUPS	NUMBER OF BRAVAIS LATTICES	PRODUCT
Cubic	5	3	15
Tetragonal	7	2	14
Orthorhombic	3	4	12
Monoclinic	3	2	6
Triclinic	2	1	2
Hexagonal	7	1	7
Trigonal	5	1	5
Totals	32	14	61

The space groups

- What **symmetry** operations (e.g. rotation axes, (2-, 3-, 4-, 6- fold axis, mirrors, inversion axes …, at corner, at face, …) can be applied to the unit cell (inside crystal)? This defines the **32 point groups** of the unit cell (see Table 1.4).
- The combination of the 32 symmetry types (point groups) with the 14 Bravais lattice, yields **230 distinct space groups**.

- In biological molecules, there are really only 65 relevant space groups (no inversion axes or mirrors allowed, because they turn L-amino acids into D-amino acids.
- → The **space group** specifies the **lattice type** (outside crystal morphology) and the **symmetry** of the unit cell (inside).
- Different crystals that have identical unit cell lengths and angles and are in the same space group are **isomorphous**.

The space groups

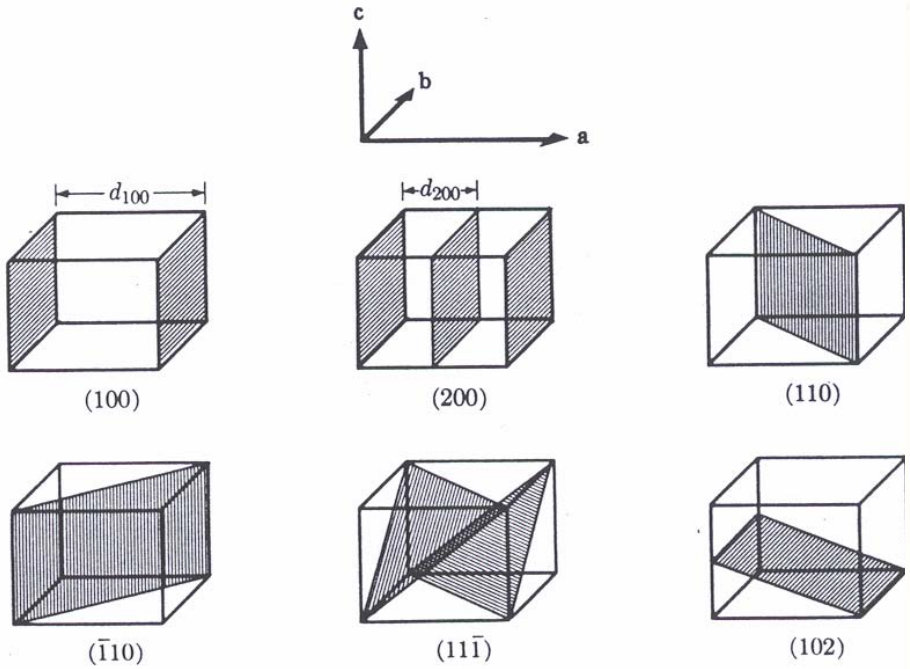


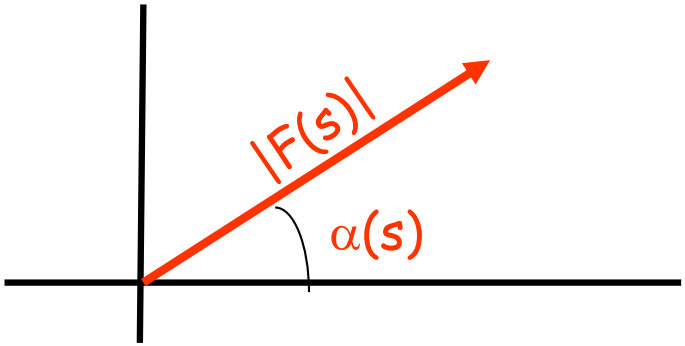
Table 7.3
THE NONCUBIC CRYSTALLOGRAPHIC POINT GROUPS[a]

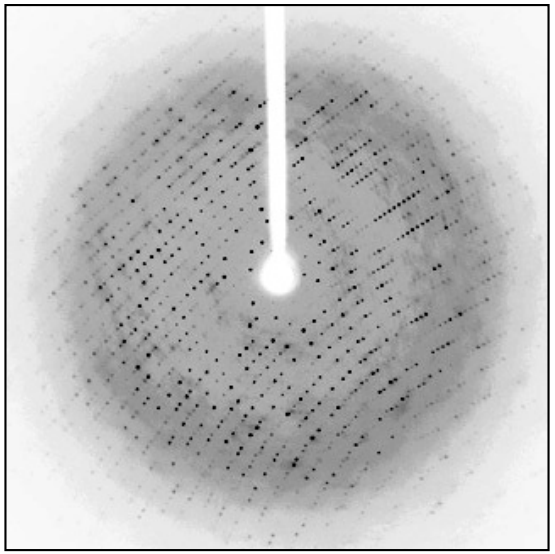
SCHOENFLIES	HEXAGONAL	TETRAGONAL	TRIGONAL	ORTHORHOMBIC	MONOCLINIC	TRICLINIC	INTERNATIONAL
C_n	C_6 6	C_4 4	C_3 3		C_2 2	C_1 1	n
C_{nv}	C_{6v} $6mm$	C_{4v} $4mm$	C_{3v} $3m$	C_{2v} $2mm$			nmm (n even) nm (n odd)
C_{nh}	C_{6h} $6/m$	C_{4h} $4/m$			C_{2h} $2/m$		n/m
	C_{3h} $\bar{6}$				C_{1h} $(\bar{2})$ m		$\bar{n}$
S_n		S_4 $\bar{4}$	S_6 (C_{3i}) $\bar{3}$			S_2 (C_i) $\bar{1}$	
D_n	D_6 622	D_4 422	D_3 32	D_2 (V) 222			$n22$ (n even) $n2$ (n odd)
D_{nh}	D_{6h} $6/mmm$	D_{4h} $4/mmm$		D_{2h} (mmm) (V_h) $2/mmm$			$\frac{n}{m}\frac{2}{m}\frac{2}{m}$ (n/mmm)
	D_{3h} $\bar{6}2m$						$\bar{n}2m$ (n even) $\bar{n}\frac{2}{m}$ (n odd)
D_{nd}		D_{2d} (V_d) $\bar{4}2m$	D_{3d} $(\bar{3}m)$ $\bar{3}\frac{2}{m}$				

[a] Table caption on p. 123.

The electron density equation: problem # 2:the phase problem

- F(s) is a complex number with modulus |F(s)| and phase α





In a diffraction experiment we measure the intensity of each spot I(s)

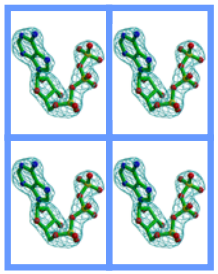
$I(s) = |F(s)|^2$

We can derive the amplitude F(s) but we have lost the information about the relative phase

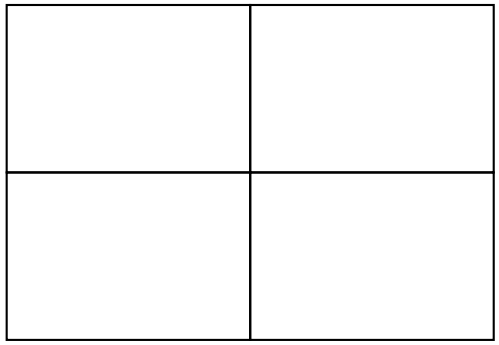
⇒ the phase problem!!

Universal problem in crystallography – also for small molecules.

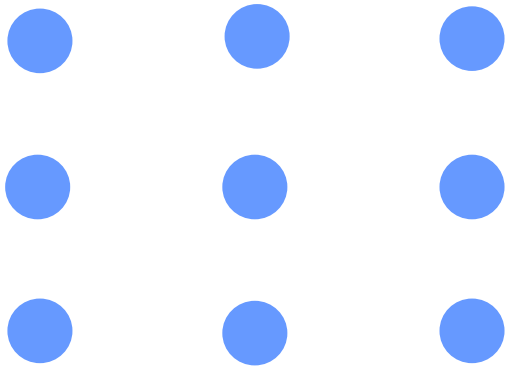
Small molecule crystal



small unit cell

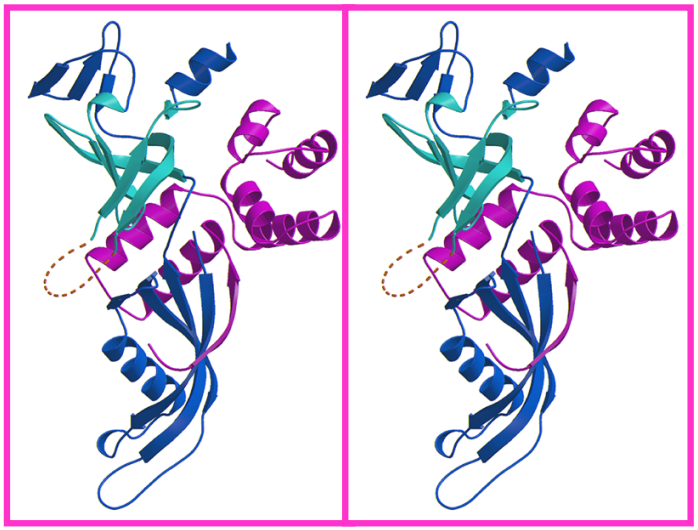


large "reciprocal lattice"
(coarse fringe function)

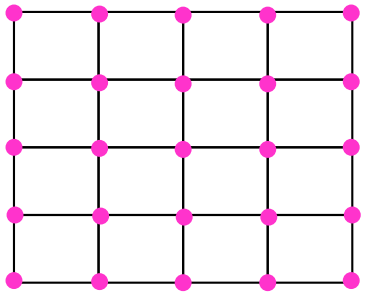


a few, very strong reflections;
diffract to high resolution

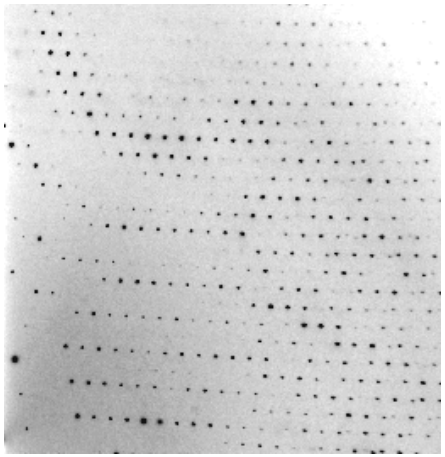
Protein crystal



large unit cell



small "reciprocal lattice"
(fine fringe function)



many, very weak reflections;
diffract to low resolution

How to solve macromolecular structures:

MIR (multiple isomorphous replacement)

Older method (Cambridge, 60') – relies on binding "heavy" atoms to the crystal and compare he diffraction pattern to the native. Trial and error search for good heavy atoms, it may take longer to get it right

MR (molecular replacement)

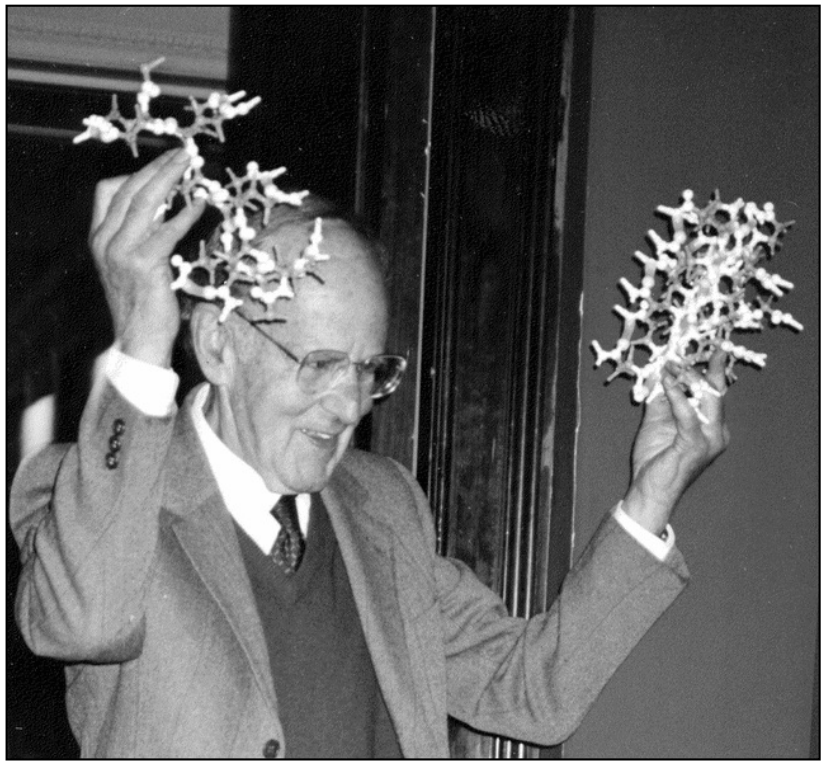
Older method (Cambridge, 70'-80') – relies on the expected similarity between the protein and another whose structure is known. Cannot solve de novo structures. Requires high homology (30% sequence identity?)

MAD (multiwavelength anomalous dispersion)

Relies on the absorption of specific wavelengths due to electronic transitions within the atom core. Similar to MIR but generally far quicker and more accurate. Requires high specification synchrotron radiation.

Some of the players...

Max Perutz (1914-2002), the inventor of MIR



In 1953, Perutz showed that the diffracted X-rays from protein crystals could be phased by comparing the patterns from crystals of the protein with and without heavy atoms attached.
In 1959, he employed this method to determine the molecular structure of hemoglobin. This work resulted in his sharing with John Kendrew the 1962 Nobel Prize for Chemistry.

Error treatment in MIR

The real breakthrough in using MIR for the determination of protein structures came when people learned how to deal with errors.

Blow D.M. & Crick F.H.C. (1959) "The treatment of errors in the isomorphous replacement method". Acta Crystallogr. 12, 794-802

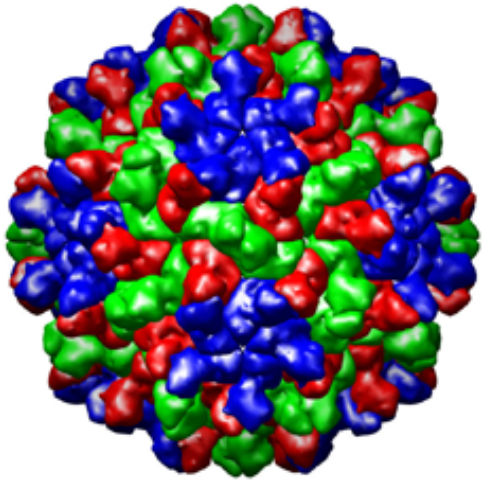
David Blow
(1931-2004)

Francis Crick
(1916-2004)

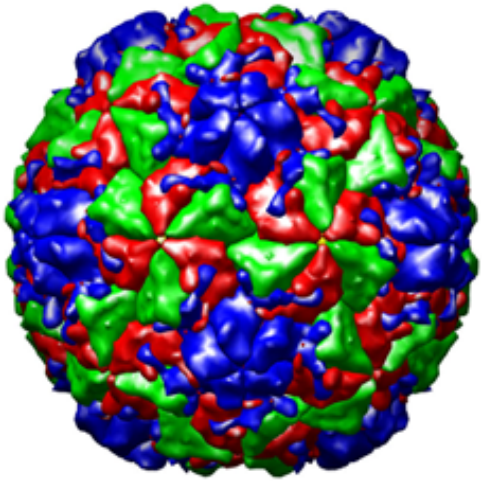
Error treatment is very complex – but it was an essential step in solving protein structures.

Molecular Replacement

The Molecular replacement method was mostly developed by Michael Rossmann. He used the structure of Tomato Bushy Stunt Virus, a plant virus, to determine the crystal structure of the Human Rhinovirus 14 (the common cold virus) in the early 80's.



Tomato Bushy Stunt Virus



Human Rhinovirus 14

However the theoretical basis were developed much earlier:

Rossman, M. G. and Blow, D. M. (1962). ActaCryst. 15:24-31.

Phase problem solved?
Calculating an electron density map

Once we have decent experimental estimates of the phases (from MIR, MR or MAD) we can obtain an electron density map by calculating a weighted Fourier transform with coefficients F_{obs} (i.e. the amplitudes of the diffracted rays and phases $\alpha_{MIR}/\alpha_{MR}/\alpha_{MAD}$.

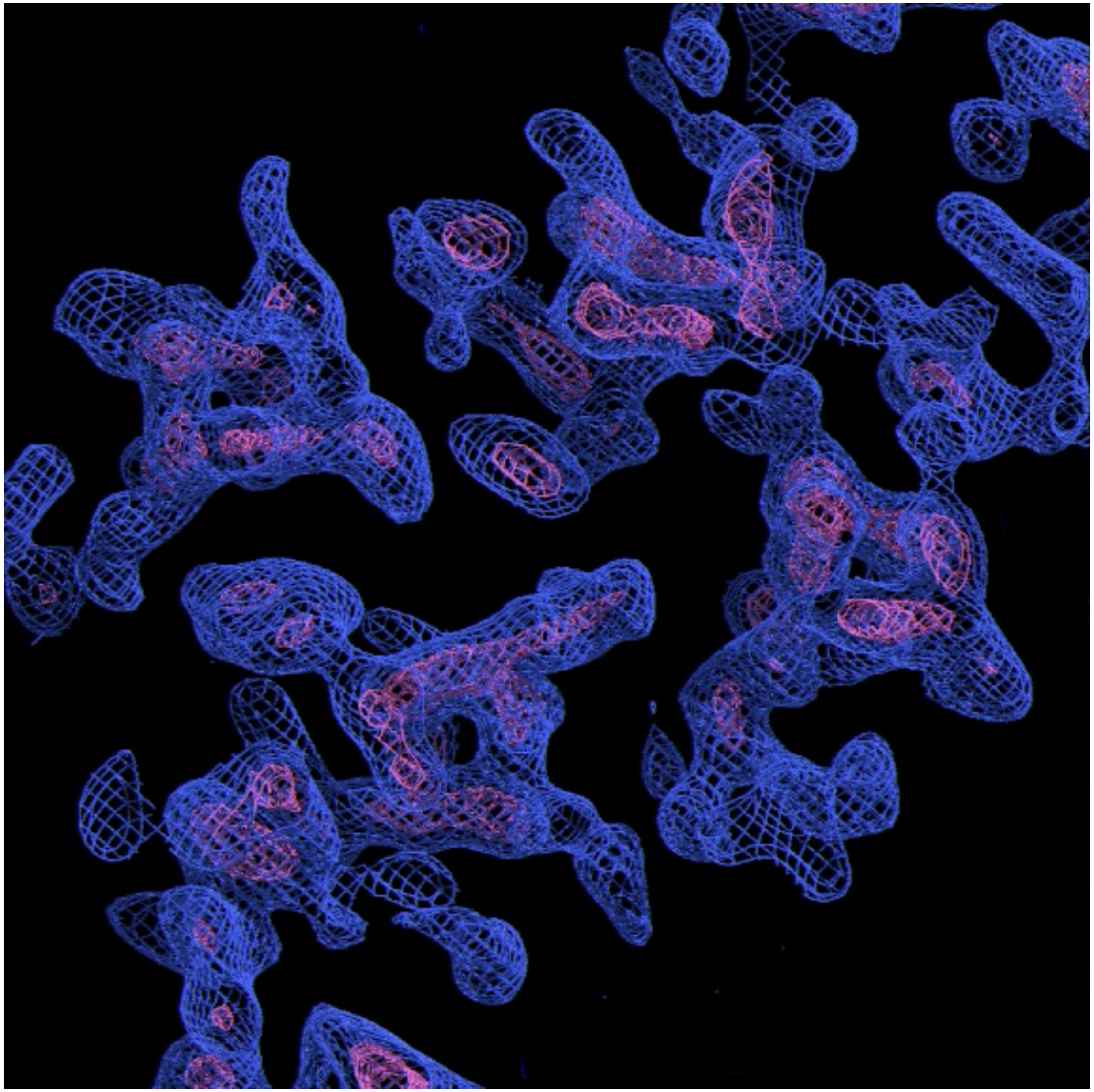
$$\rho(xyz) = \sum_{hkl} W(hkl) * |F_P(hkl)| \exp[i\,\alpha(hkl)] \exp[-2\pi\, i(hx + ky + lz)]$$

At the end of the day, the only criterion for determining how good is your MIR/MR/MAD solution is whether the map is interpretable.

Can you build a polypeptide chain?

Electron density maps

After all this effort, we have a 3D map showing the shape of the protein:

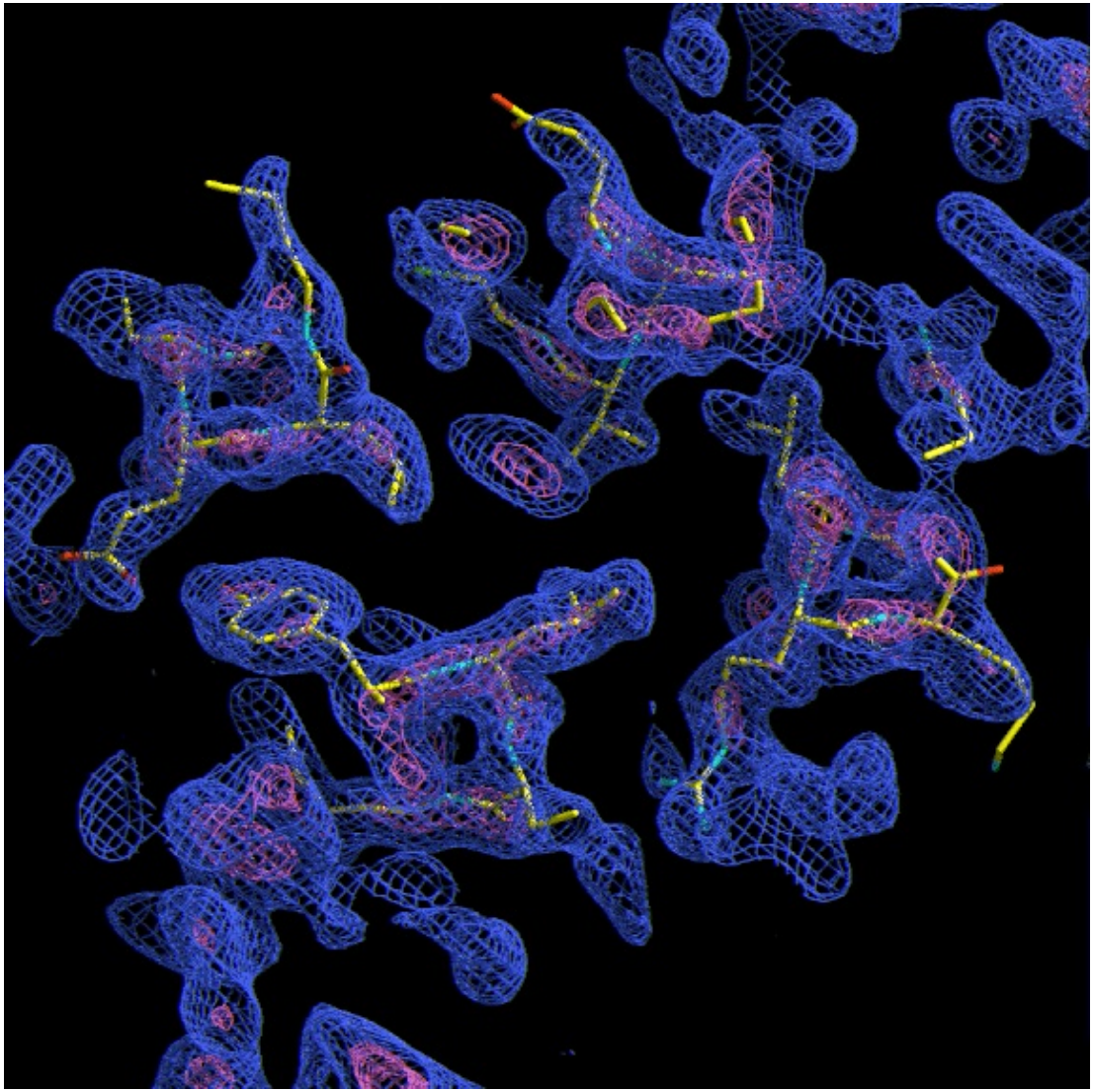


Electron density map displayed at two contour levels:

blue = 1 r.m.s (1σ);
magenta = 3 r.m.s (3σ)

Electron density maps

The task now is to try to fit an atomic model of the protein to the map…



Electron density map displayed at two contour levels:

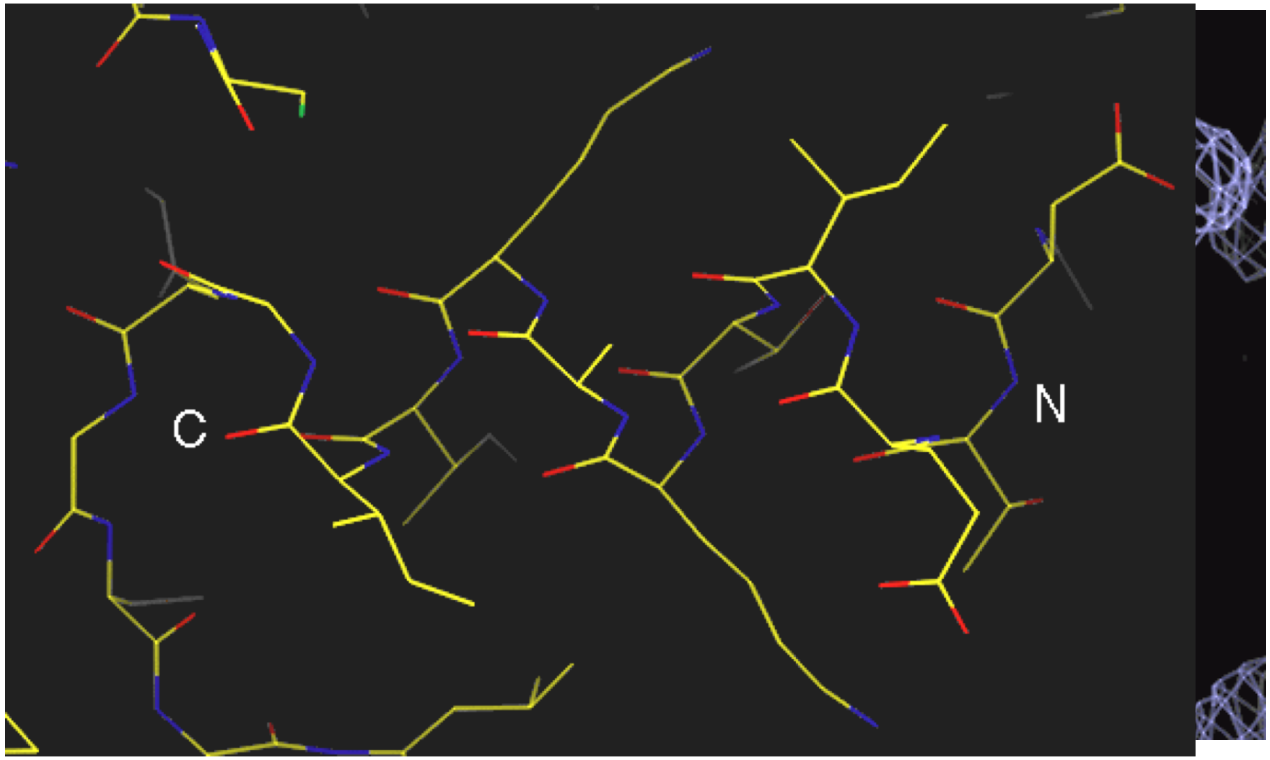
blue = 1 r.m.s (1σ);
magenta = 3 r.m.s (3σ)

Atomic model fitted to the map (in yellow)

Displaying the electron density map

Visualising a 3D electron density map over the entire unit cell can be daunting.

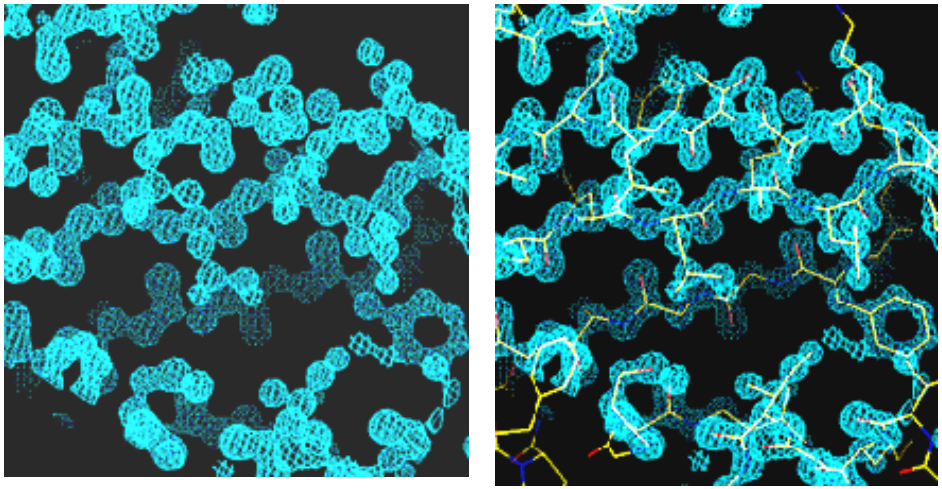
To visualise electron density maps we use high resolution interactive graphics, which display the map as a chicken wire. A map is a 4-dimensional object (each point x,y,z has an associated value) and to display it in 3D we have to choose a contour level that allows us to see about 70 % of the van der Waals radius of the atoms (typically one contours at the 1 σ level)



Here is a rather good bit of density (much better than the density you are likely to obtain from the initial phases) and still it is not easy for the "untrained eye" to see what is going on...

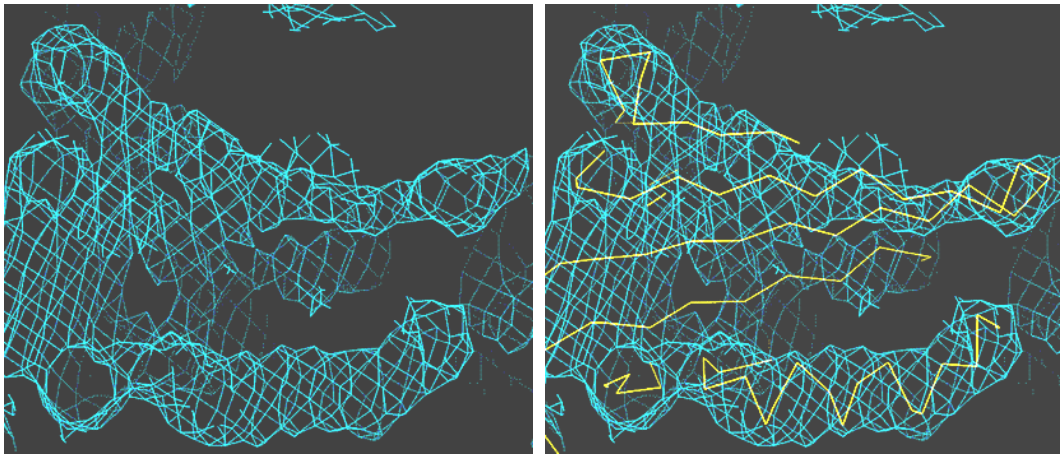
Maps and resolution

The task of model building is to interpret the electron density maps in light of chemical knowledge, basic stereochemistry, chemical sequence, etc...
The level of interpretation depends on the resolution of the map:



Here is a 1 Å map

At very high resolution, individual atoms can be seen and fitted in the electron density blobs: the problem therefore is reduced to 'join-the-dots'



Here is a 6 Å map

At very low resolution only large features can be seen - for example helices look like rods and β-sheets can barely be detected.

Refinement

Manual model building is not sufficient to build a completely accurate model but is required to get to the starting point for refinement.
Refinement is a process of optimisation of the atomic model to match the observed data and to conform to ideal stereochemistry

Successful structure determination usually requires several alternated rounds of model-building and refinement.

During refinement one calculate the expected diffraction pattern from the current model, compare it to the experimental diffraction data, and minimise the square of the differences.

Two problems:

- non linearity and presence of multiple minima
- low ratio observation parameters (especially at low res.); compensate by using stereochemical constraints.

The crystallographic R-factor

During the cycles of refinement, we calculate "R factors" to assess the progress. These are similar to the Q values used in minimisation.

$$Q = \sum \left[F_{obs} - F_{calc} \right]^2$$

$$R_{factor} = \sum \frac{\left| \left| F_{obs} \right| - \left| F_{calc} \right| \right|}{\left| F_{obs} \right|}$$

R = "residual" = fractional difference between observed and calculated diffraction - a sort of "fractional error"

To monitor the refinement, we calculate R after each cycle (the summation is over all reflections). If things are going well, R reduces.

However, with complicated refinements like these, it is possible to "over-fit" the data – for cross-validation we take away 5% of the data (which we do not use in the refinement, to monitor the agreement. This is known as R_{free}

X-ray fiber diffraction

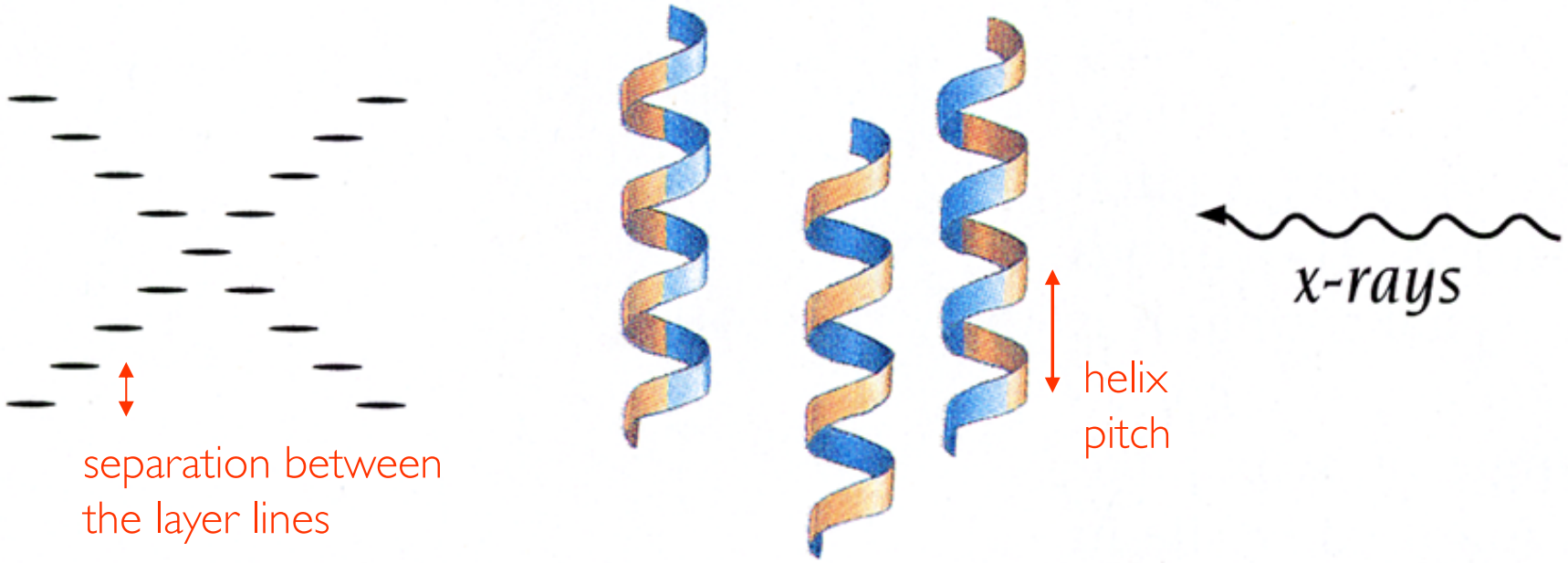
- If there is inherent symmetry in a molecule (i.e. long helical biopolymers) X-ray diffraction patterns can be obtained from non-crystalline samples.
- Like one-dimensional crystal along helix axis.
- We can treat an exact repeat of helix as a crystalline unit cell.
- Can get information about helical symmetry, pitch and radius.
- Meridianal reflections: information on pitch and symmetry
- Equatorial reflections: radius of helix

Many biological polymers and assemblies are long, helical, filamentous structures. Examples include the filamentous viruses, cytoskeletal filaments, bacterial flagellae, chromatin, components of the extra-cellular matrix, and many simple polymers such as nucleic acids and polysaccharides.
The components of these assemblies are difficult to crystallize, since their natural tendency is to form filaments.

X-ray fiber diffraction in biology

DNA fibers or fibrous proteins tend to pack side by side in an ordered manner, with the axis of their fibers lining up in a parallel fashion.

The polymers in fibers typically assume helical structures; this gives a characteristic cross-like variation in the diffraction pattern.



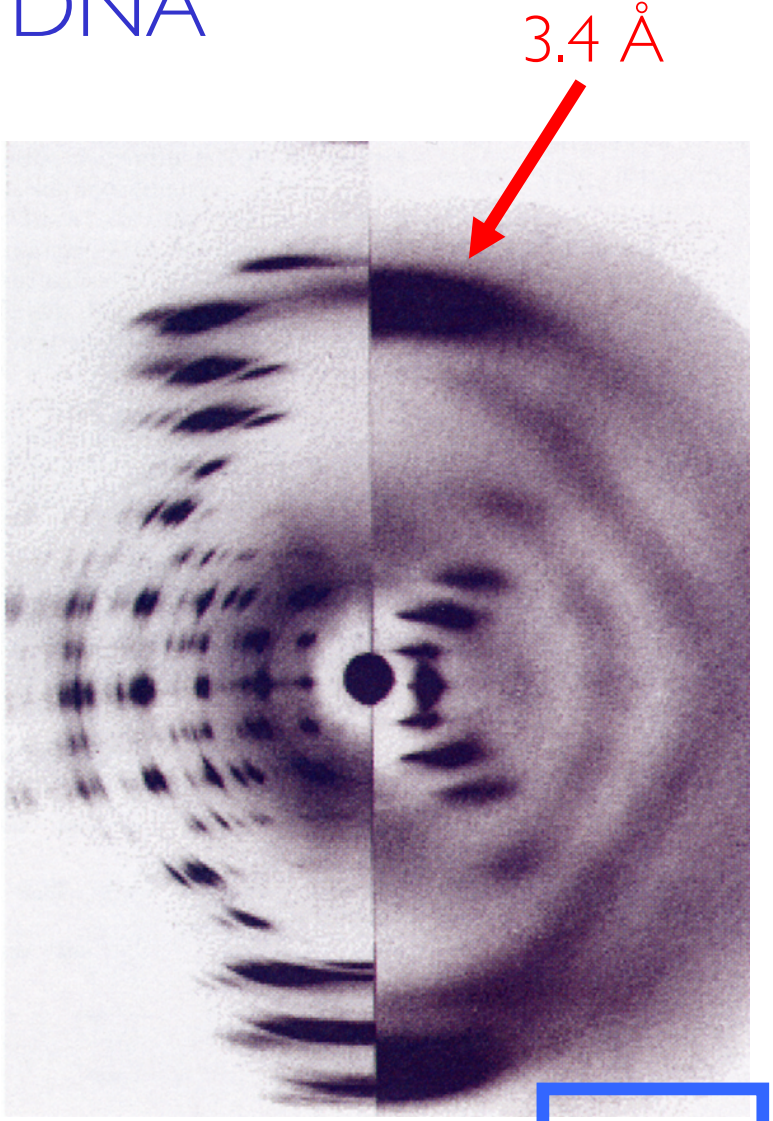
The separation between the layer lines is determined by the helix pitch: as the helix pitch increases the layer lines move closer together.

X-ray fiber diffraction: the structure of DNA

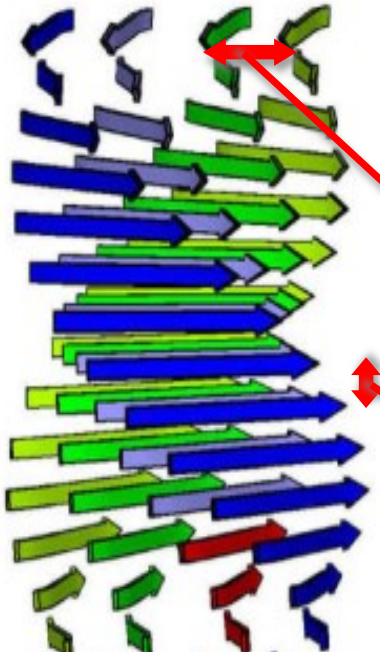
Here is the diffraction pattern from two forms of DNA: the pattern from the A-DNA indicates a higher degree of regularity in the way the fibers pack together (more 'spots').

The pitch of the A-form is 28 Å while for the B form is 34 Å (smaller spacing). Knowing the helix pitch, it is possible determine an approximate value for the radius of the helix from the inclination of the arms of the cross.

Arguments such as these, based on the X-ray fiber diffraction patterns obtained by M. Wilkins and R. Franklin, were used by J. Watson and F. Crick in the construction of the double-helical model for DNA

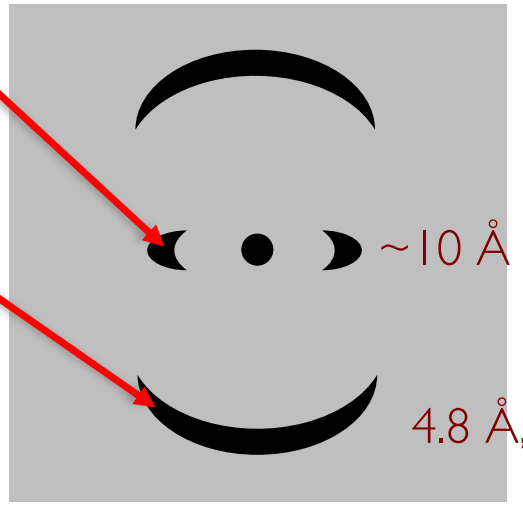


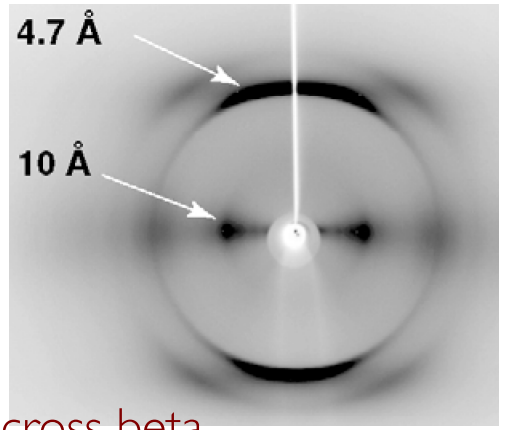
Amyloid fibers



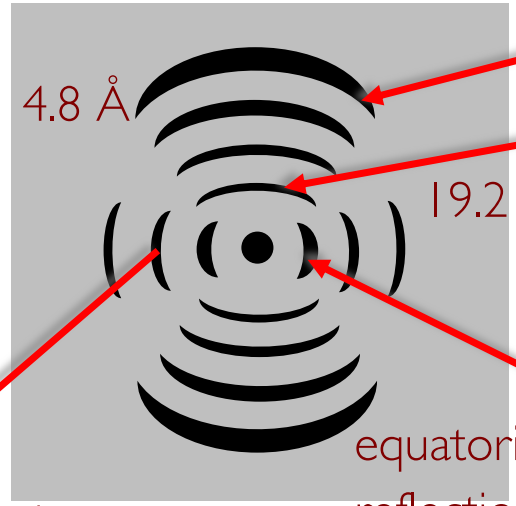
Stacked

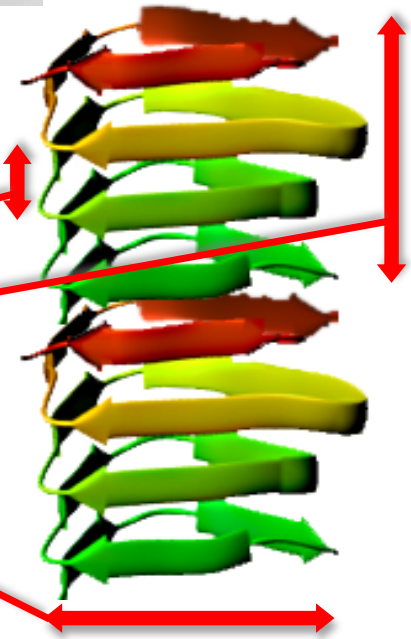
ß-sheets





IAPP amyloid,
X –ray fiber diffraction,
Makin et al., 2006





Parallel beta-helix

equatorial reflections represent structure and packing of fibrils

Natural vs recombinant sources

Most of the effort needed to determine a crystal structure involves preparing your protein and obtaining good crystals.

The large majority of protein structures have been solved using recombinant proteins.

- Can over-express protein in a variety of different cell systems (e.g. E. coli, yeast, insect cells, mammalian cells)
- Easier to handle than large animals
- Easier to change specie to increase chances of getting crystals
- Easier purification via tag technology
- Easier to obtain the protein sequence which we need for model building
- Allows to modify the protein to make it less floppy or more soluble
- Possible to co-express more than one protein to make a complex

Natural vs recombinant sources

Recombinant DNA technology makes it easier to produce large quantities of purified protein.

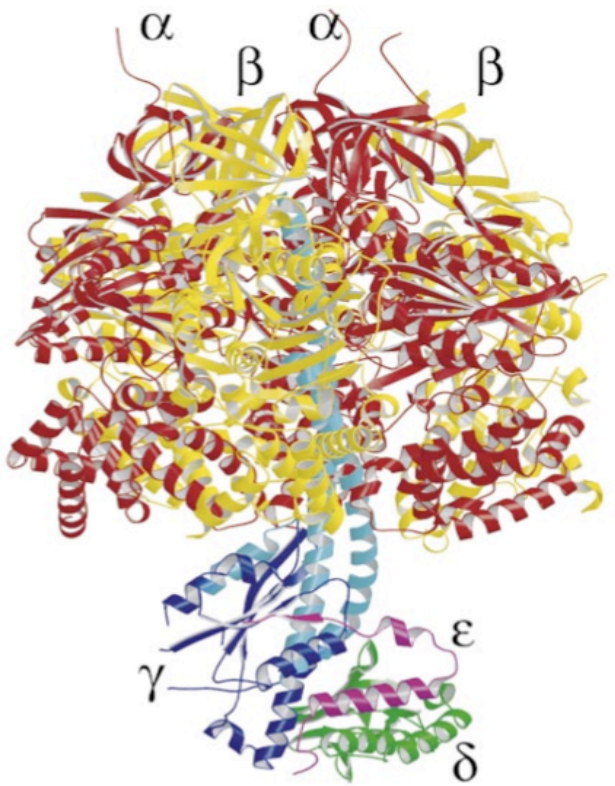
However "natural" sources still useful especially for big complexes, since these cannot be easily re-constituted by recombinant expression.

Examples:

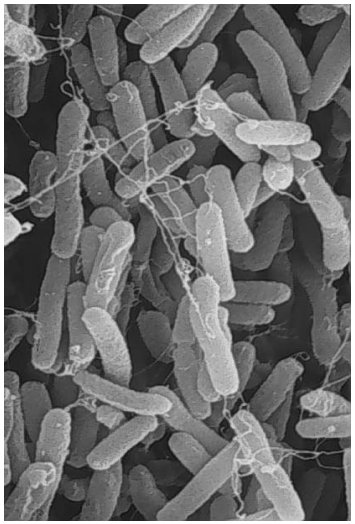
F_oF_1 ATP synthase - purified from
bovine heart muscle/yeast cells

ribosomes - purified from archaea

RNAPII - purified from yeast cells



Expression in Escherichia coli



First choice organism:

- cloning and handling is easy
- grows rapidly in liquid cultures (doubling time < 30 min)
- 1-6 L of culture can yield 5-50 mg of purified protein
- well characterised organism

Not foolproof…

- Difficult to make proteins with disulphide bridges
- Protein may over-express well but be unfolded
- Control folding with temperature and/or chaperones
- May be able to "refold" proteins by controlled denaturation/renaturation during purification - but tricky!
- Proteins may be toxic for bacteria - try to get around this by tight control of expression

Easy/difficult to crystallise?

A lot of interesting proteins are difficult to crystallise.

Best cases:

- single proteins
- rigid domain structure
- one dominant conformation

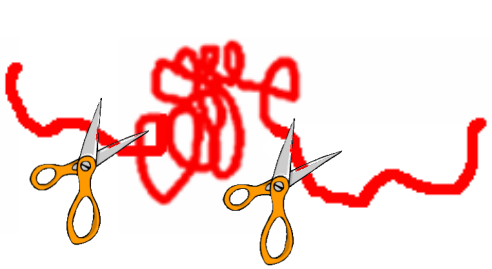
Worst cases:

- multi-domain proteins with flexible unstructured linkers
- proteins with flexible N- and/or C-termini
- proteins that are part of large macromolecular complexes
- presence of posttranslational modifications

"Optimise" proteins for crystallisation

Use bioinformatics (database searches, sequence alignments) to identify "core domains that can be expressed in a soluble form.

Use limited proteolysis to identify compact domains.



beware:
you can
cut flexible
loops).

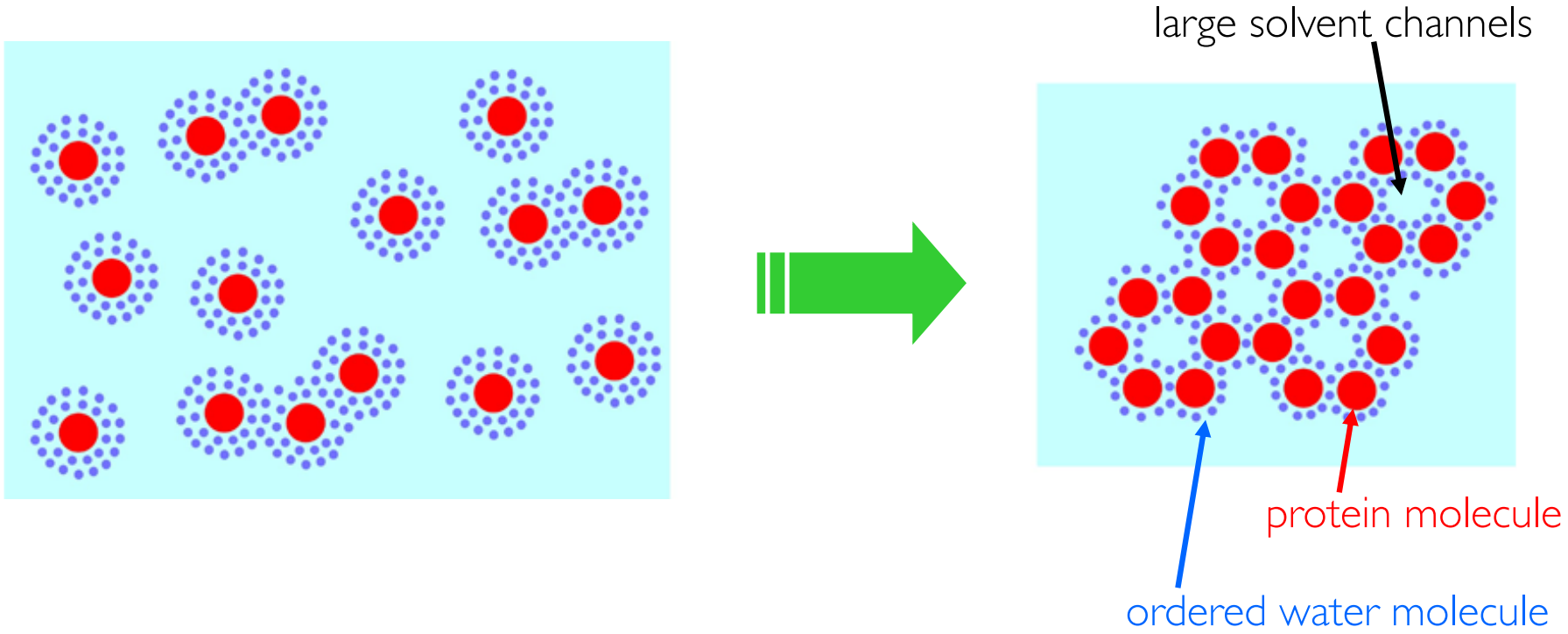
Co-express proteins that are part of the same complex.

Add ligands/inhibitors/cofactors/metals
to stabilise one conformation.

Avoid/encourage post-translational modifications (mutagenesis of target residues, mimic modifications by mutagenesis, change expression system/cell lines, so that the process does/does not occur).

Crystallisation

Protein crystallisation is the transfer of protein molecules from an aqueous solution to an ordered solid phase.



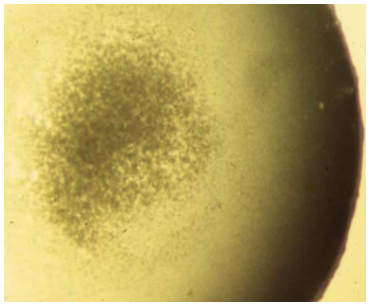
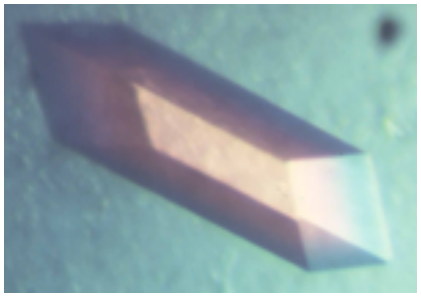
Protein crystallisation is controlled by:

- thermodynamic factors governing the solubility
- kinetic factors governing nucleation and growth

Solubility and crystals

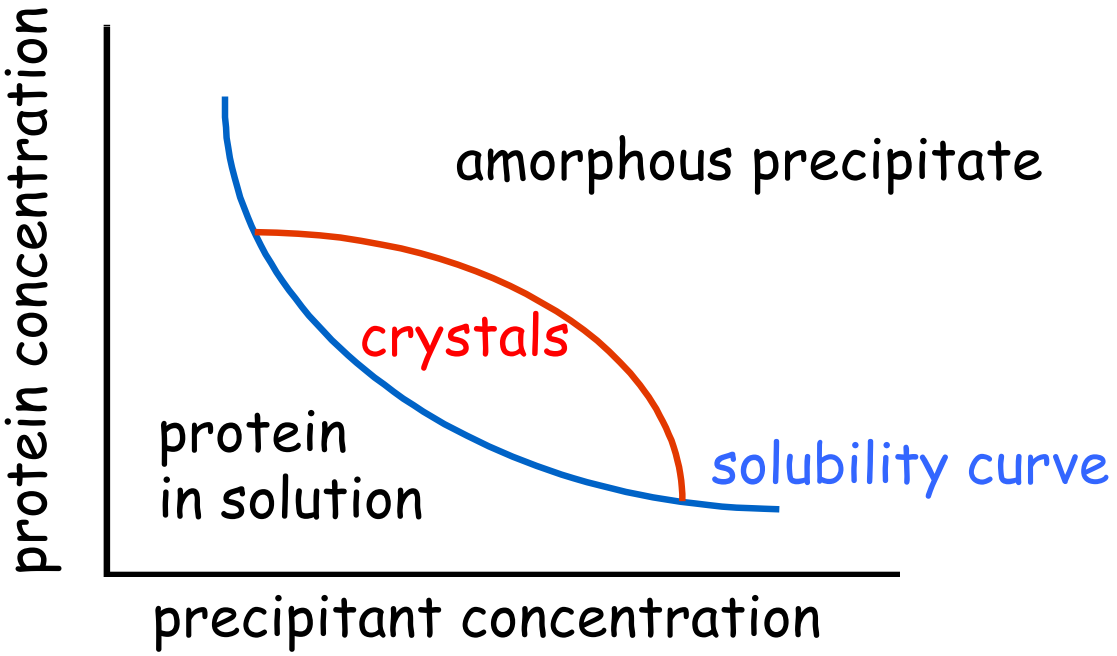
To coax a protein into forming crystals, one has to decrease its solubility to the point where the solution becomes saturated. This is done by changing pH, ionic strength or temperature, or by adding organic solvents.

In a saturated solution the protein in solution is in thermodynamic equilibrium with one or more solid states:

- amorphous precipitate → 
- crystalline state
  - micro-crystals
  - needles
  - plates
  - large single crystals → 

Phase diagram

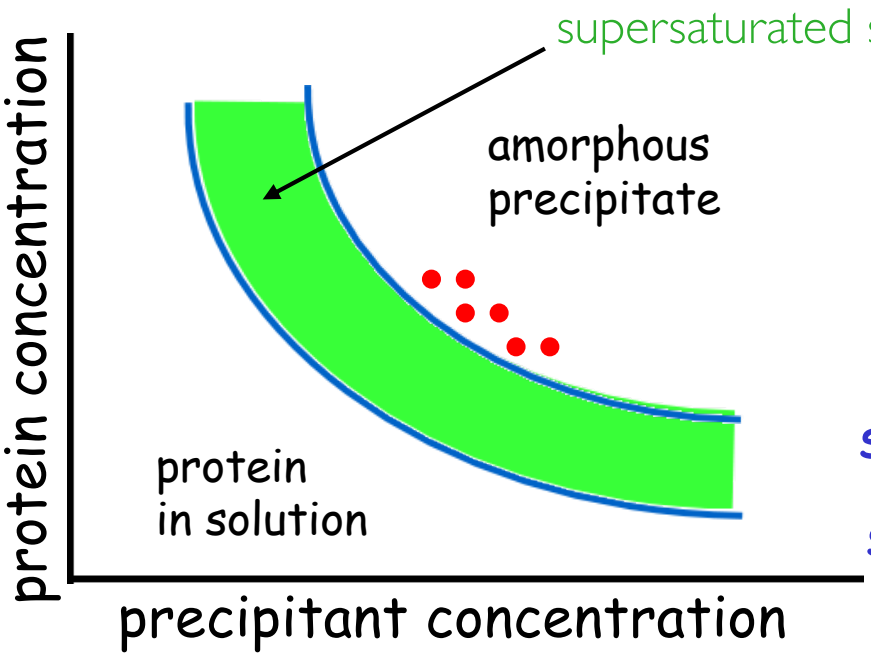
The process of crystal formation can be understood by using a phase diagram. Because several factors influence protein solubility, the phase diagram has many dimensions. A very important factor is the precipitant concentration, and a two-dimensional diagram is shown here:



Supersaturation

Above the solubility curve the solid phase is thermodynamically favoured; yet, because of kinetic barriers to the formation of aggregates, the protein may remain in solution, forming a supersaturated solution.

A supersaturated solution is thermodynamically metastable: its conversion to a stable solid state is kinetically controlled and can be triggered by the presence of nucleation centres.



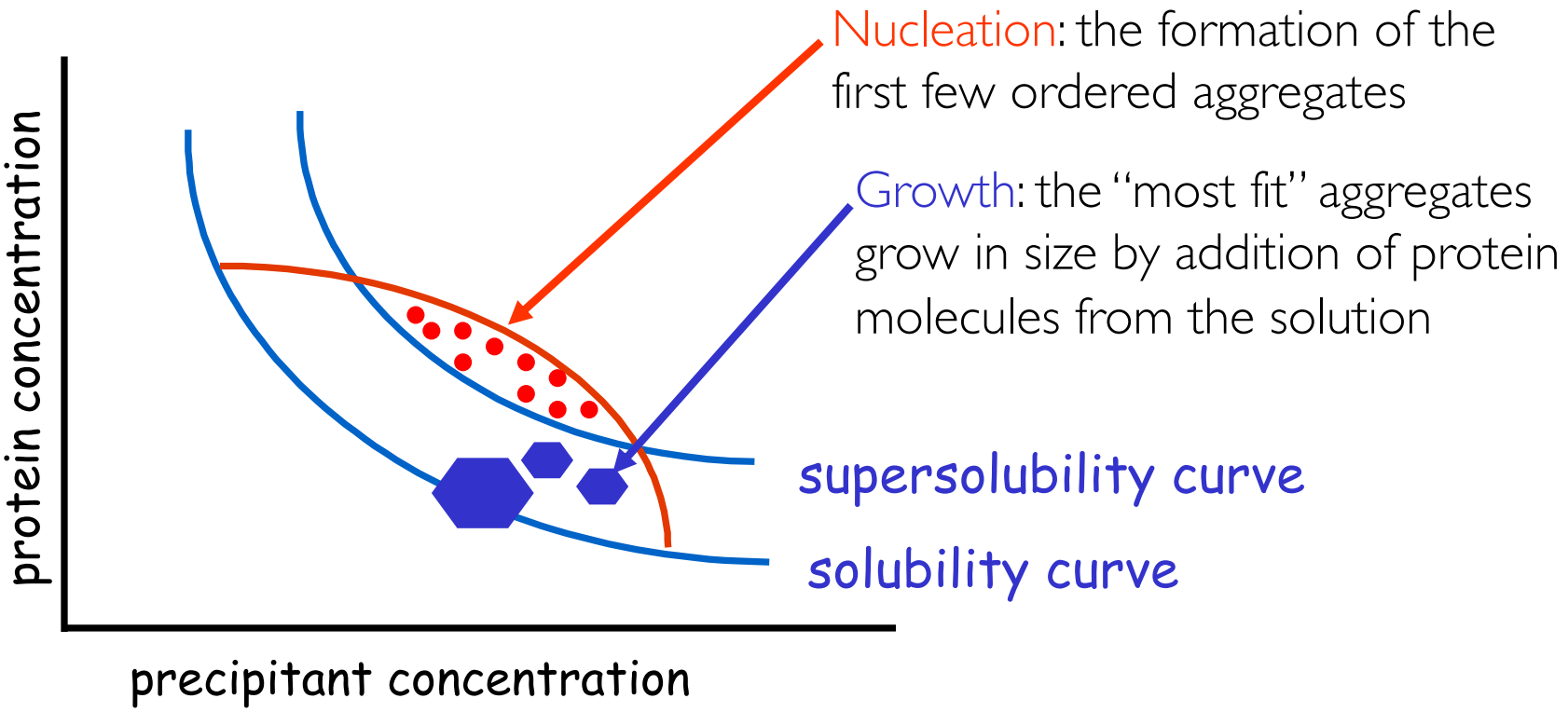
Above a certain level a supersaturated solution will spontaneously give rise to the formation of nuclei, without any need of 'external' nucleation centres.

supersolubility curve

solubility curve

Nucleation and growth

The formation of crystals requires aggregates of a critical size (nuclei). Aggregates smaller than the critical size are unstable and will dissolve, while aggregates larger than the critical size will grow.
The process of crystallisation can be divided into two distinct stages:



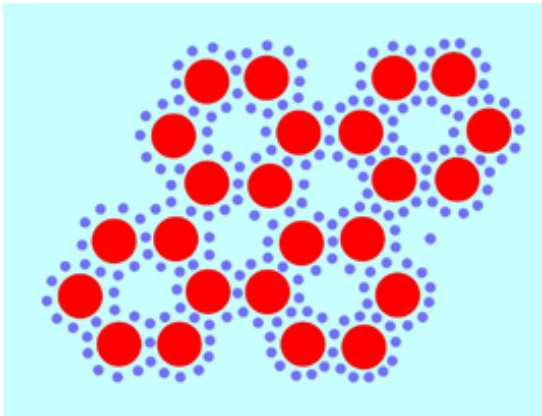
Crystallising proteins

Why proteins are difficult to crystallise:

- most proteins are labile and easily denatured
- large proteins often exist in multiple conformations
- complex behaviour -> polymorphism
- many proteins are difficult to obtain in large amounts
- proteins needs to be highly purified for crystallisation

Why protein crystals are difficult to handle:

- high solvent content (30-80%)
- mechanically fragile
- not well ordered -> resolution limits
- sensitive to radiation damage



Crystallisation of membrane proteins may presents additional problems such as homogeneity of the purified samples, choice of detergents, presence of micelles, tendency to form hydrophobic interactions which are less directional and ordered, etc..

Crystallising proteins

Purity is not an absolute requirement for crystallisation, but often crystals obtained from impure solutions are small or of poor quality.

The protein solution has to be homogeneous not only in terms of composition, but also in terms of structure and conformation.

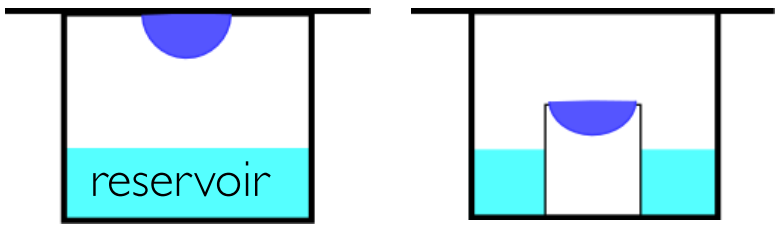
There are many factors to consider. Some of the more important ones are:
- presence of ligands (substrates, cofactors, metal ions etc)
- flexible domains and loops
- post-translational modifications (glycosylation, phosphorylation etc)
- aggregation

Highly homologous proteins will typically crystallise in very different conditions.

One needs large amounts of very pure protein that can be concentrated to several mg/ml without aggregation.
Robotic screening allows testing of ~1000 conditions/mg of protein.

Vapour diffusion

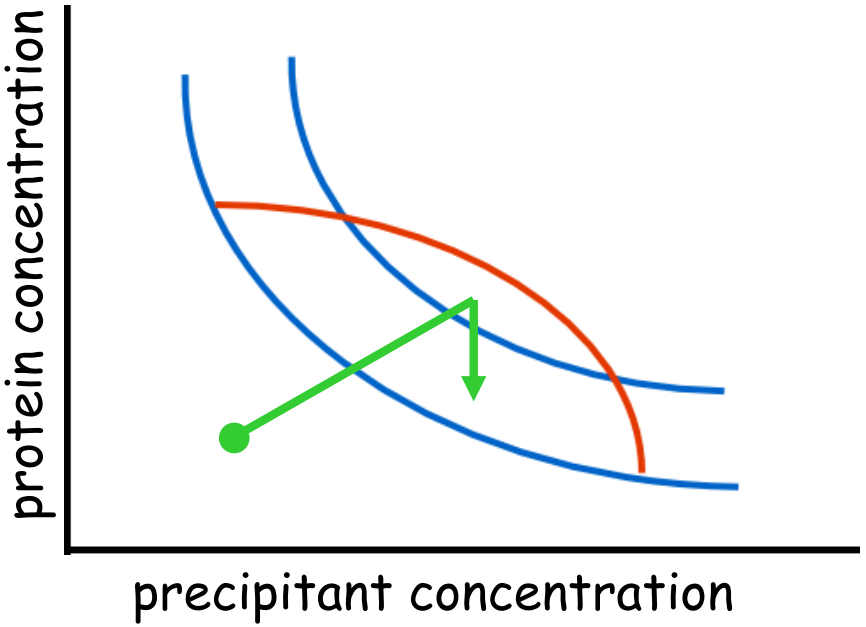
A droplet containing the protein and the precipitating agent is equilibrated against a reservoir containing the precipitant at higher concentration. Equilibrium is reached through the vapour phase, usually by loss of water from the droplet.



hanging drop

sitting drop

Typically the drop is made by mixing equal volumes of protein and reservoir solution.

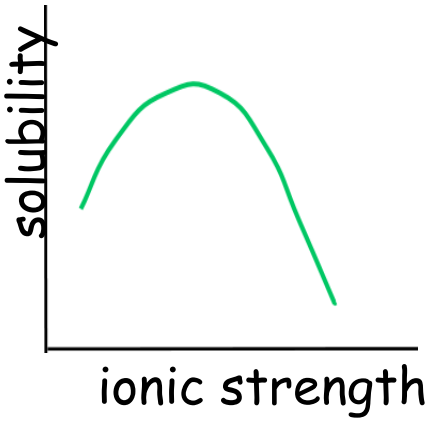


Vapour diffusion is best suited for robotic nanolitre crystallisation.

Precipitants

SALTS

Solubility changes with the ionic strength of the solution, causing salting in/salting out effects
Common ionic precipitants include $(NH4)_2SO_4$, various phosphates, NaCl, LiCl, and many others.



ORGANIC SOLVENTS

Reduce the dielectric constant of the medium, thereby enhancing the electrostatic interactions between the protein molecules and lowering the solubility. May denature proteins.
Commonly used organic solvents include ethanol and isopropanol.

POLYETHYLENE GLYCOLS (PEGs)

Believed to act by a combination of effects: volume exclusion, modification of the dielectric properties of the solution, electrostatic interaction with proteins, depletion of the hydration sphere of the protein. Very effective

Salts and PEGs are the most common precipitants

High throughput crystallisation

Automated methods for crystallisation (and crystal visualization) are now routinely used by most labs. These development are driven by the needs of pharmaceutical companies and structural genomics projects.

Robotic crystallization systems relies on the same principles described so far, but carry out each step of the procedure quickly, accurately, in smaller volumes.

Manual crystallisation:

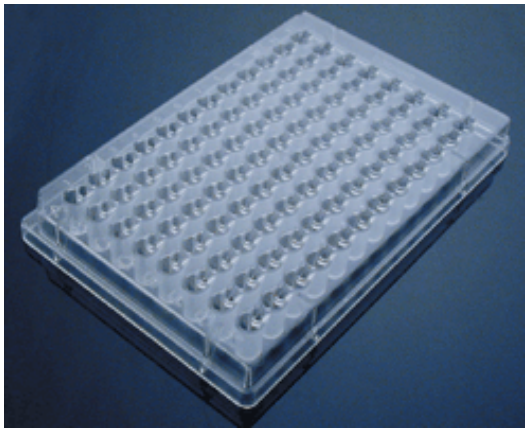
- slow and time consuming
- error prone and not always reproducible
- expensive in terms of amount of purified protein (drops: 1-2 μl)

Robotic crystallisation:

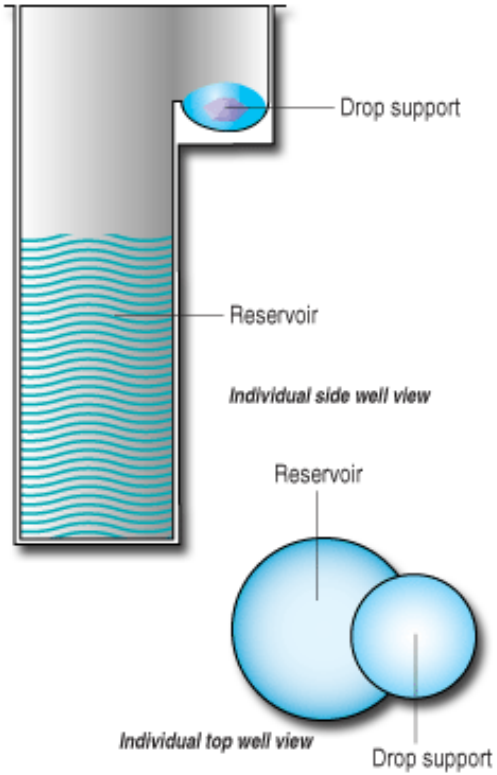
- faster and more efficient
- more accurate and reproducible
- smaller sample sizes (down to 100 nl drops) cut down on expenditure of purified protein

96-wells plates

The standard format for automated crystallisation are plates with 96 reservoir wells. These plates are designed for the sitting drop vapor diffusion method, and are sealed with clear tape.



8 x 12 matrix

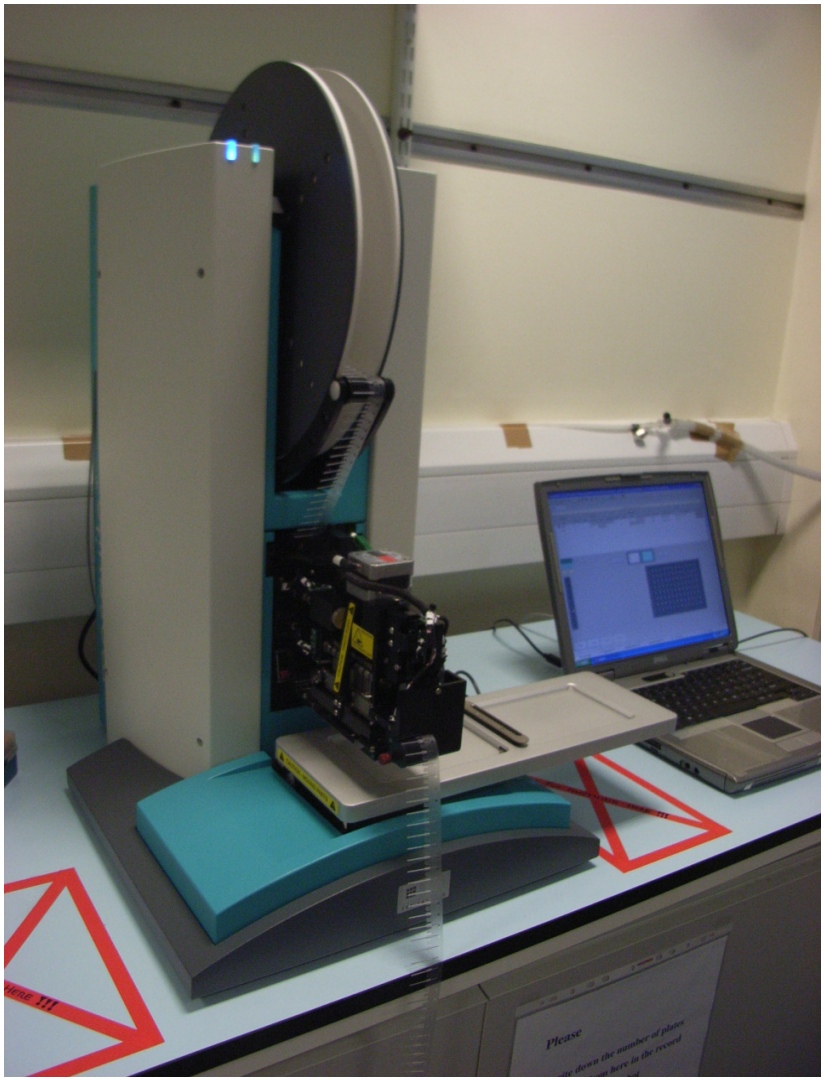


Mosquito crystallisation robot

To set up the crystallisation drops containing nanolitre volumes of protein and well solution.

Employs disposable tips and pipettes 100-200 nl drops.

Takes only 2 min to set up 96 drops.



Use of specialised 96-well plates