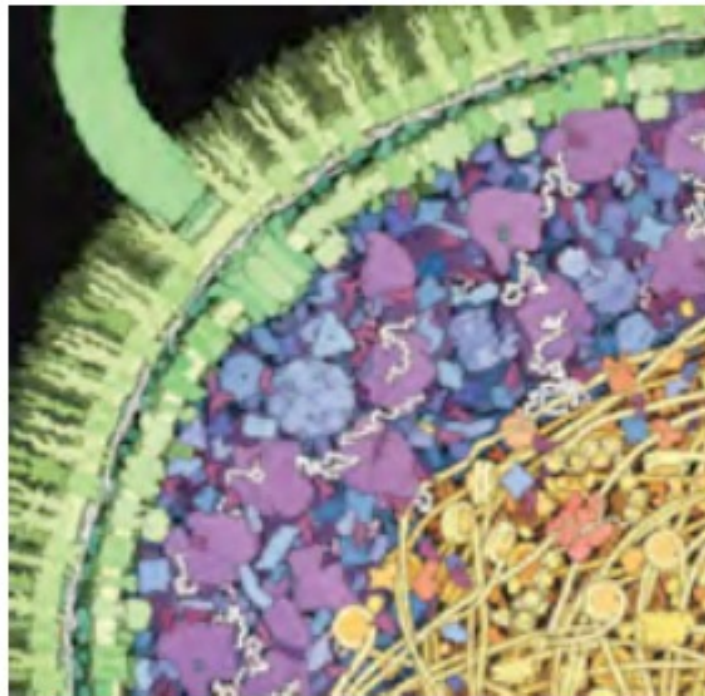


With slides from Silvia Onesti

**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.

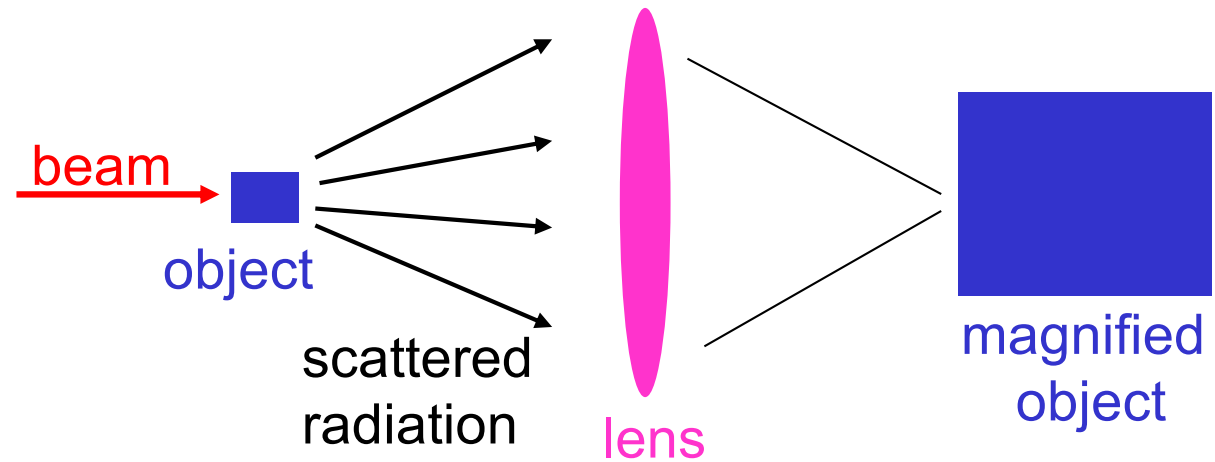
© David Goodsell, The Scripps Research Institute, La Jolla, CA, USA

Structure determination methods

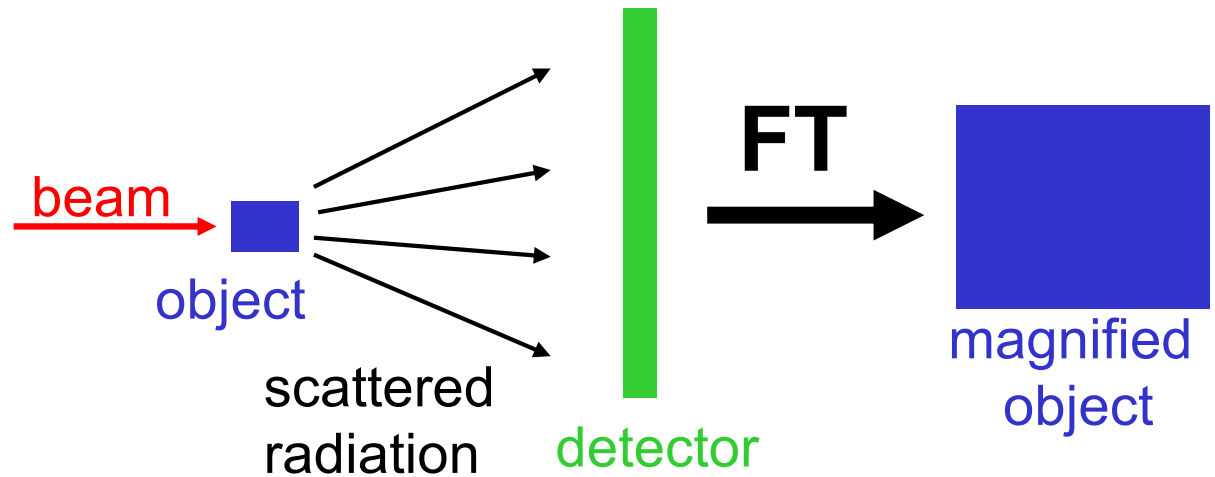
- X-ray crystallography (protein crystallography)
- X-ray fiber diffraction
- Small angle X-ray scattering (SAXS)
- Scanning electron microscopy (SEM)
- Transmission electron microscopy (TEM):
 - 2D crystals (electron diffraction)
 - Fibers/helices/tubular crystals
 - Single particle electron microscopy
- Atomic Force Microscopy (AFM)
- Nuclear magnetic resonance (NMR)

Microscopy vs diffraction

Microscopy



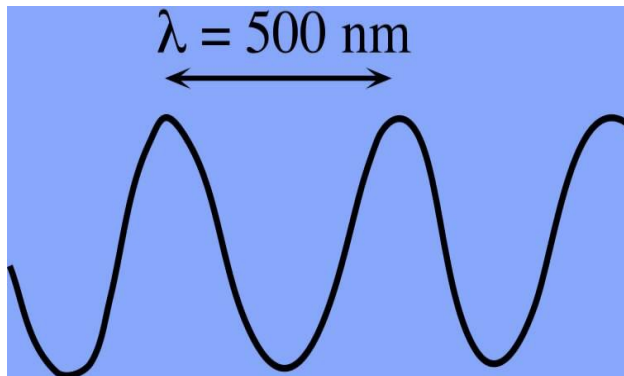
Diffraction



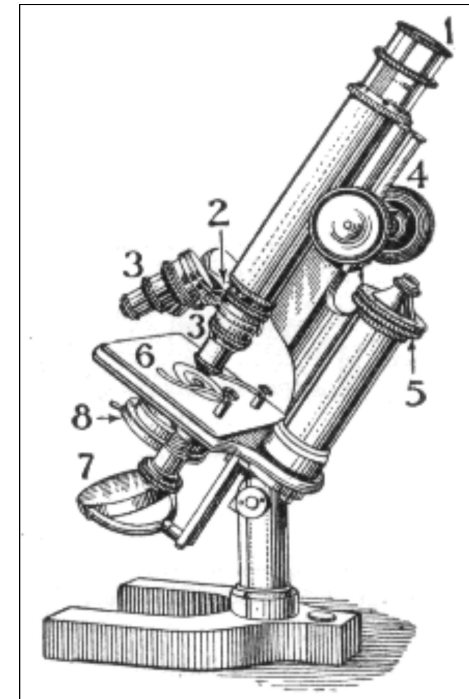
Why can't we use a microscope?

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 $\lambda=0.1 \text{ nm}$, the **right size** for looking at atoms)...

X-rays

X-rays:

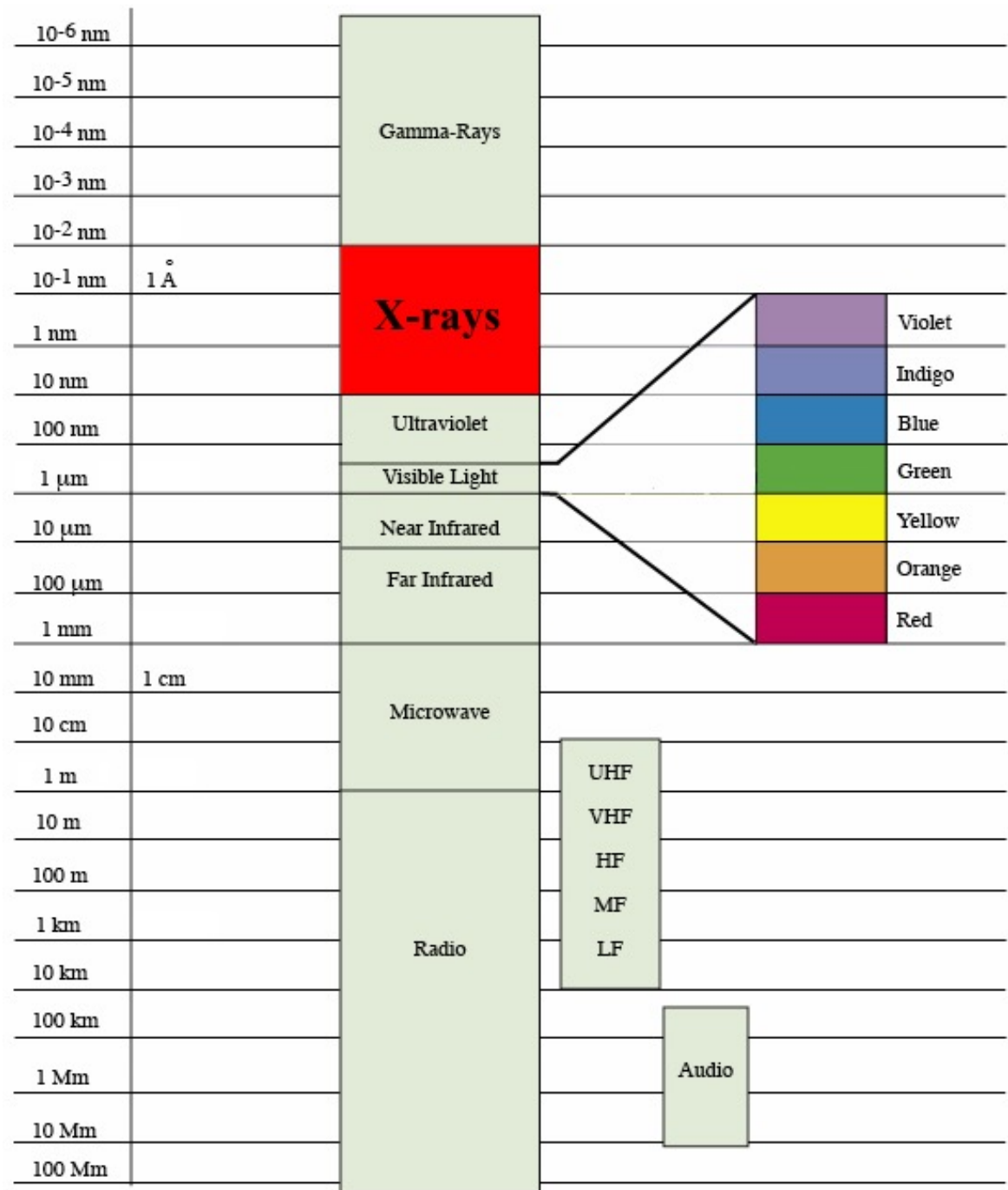
$$E = 1-100 \text{ keV}$$

$$\lambda = 10 \text{ nM}-0.01 \text{ nm}$$

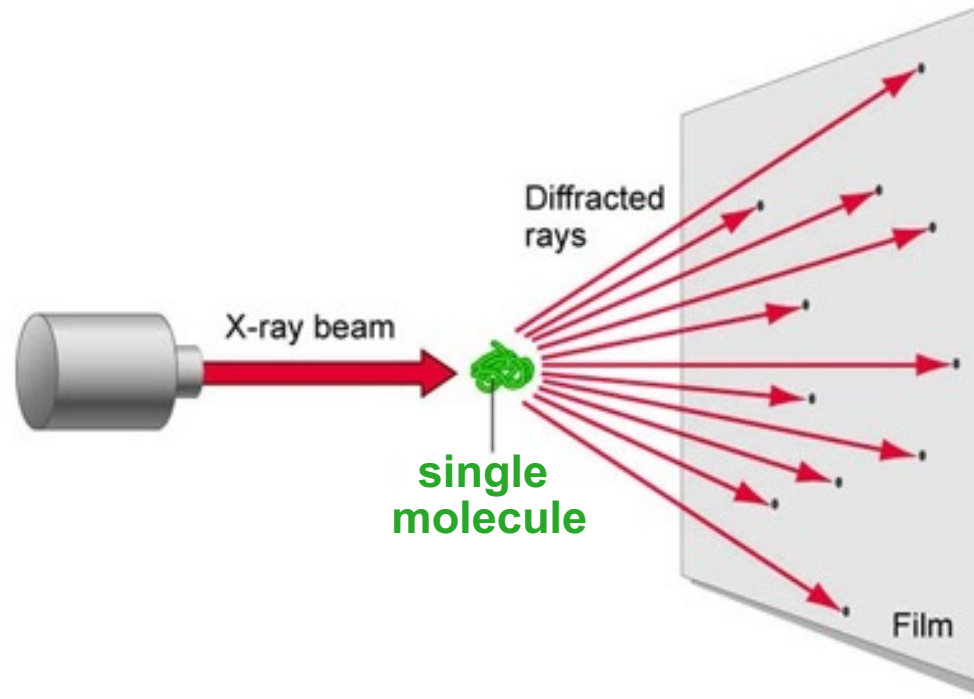
Soft X-rays (lower E) are used in medical imaging.

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.



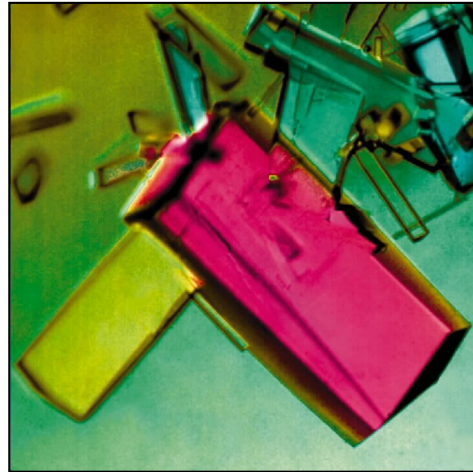
X-ray diffraction with single objects?



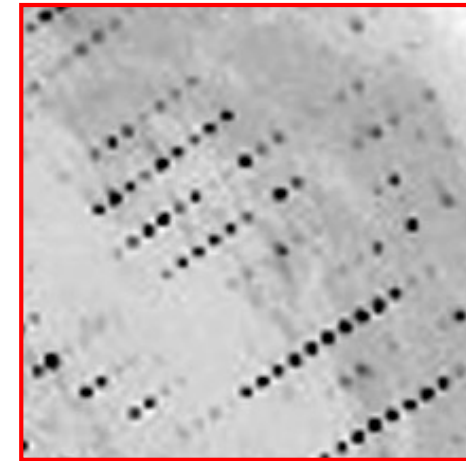
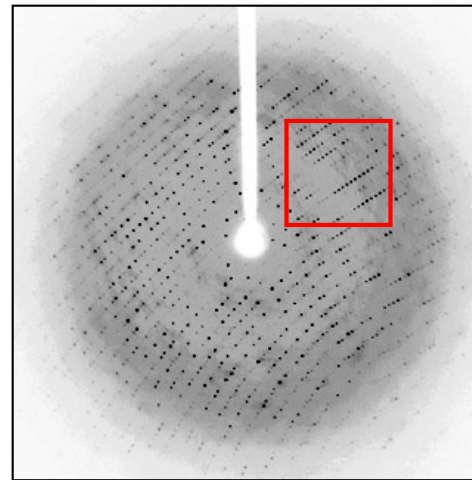
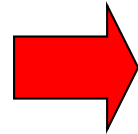
The intensity of the X-ray radiation diffracted by a single biological object (such as a protein or a cell) is very small:

➡ almost impossible to measure with current technology.

But we can use X-rays with crystals

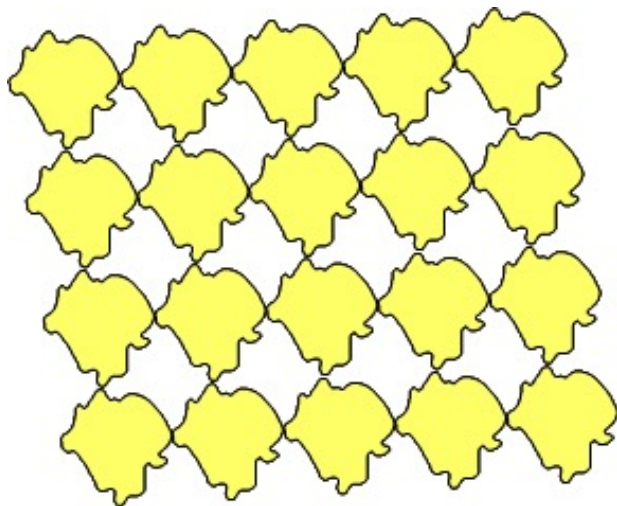


Crystal



X-ray Diffraction pattern

Crystal: an ordered array of molecules



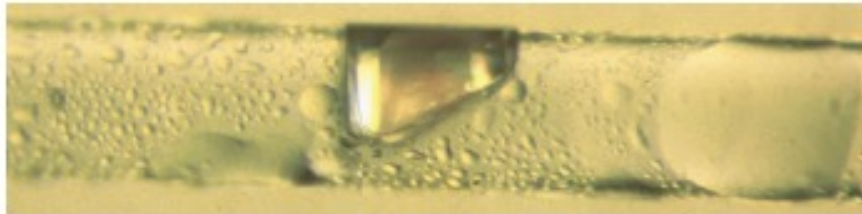
Two effects on diffraction:

- 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

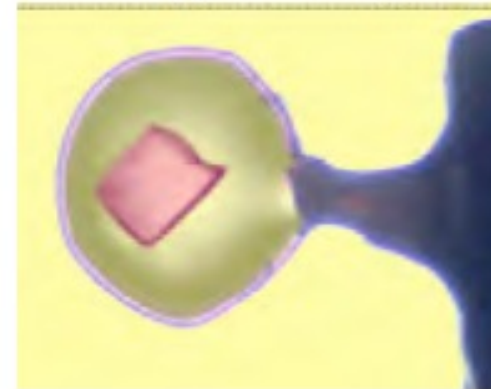
Temperatura ambiente
(293-298K)

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 N₂ o He liquidi

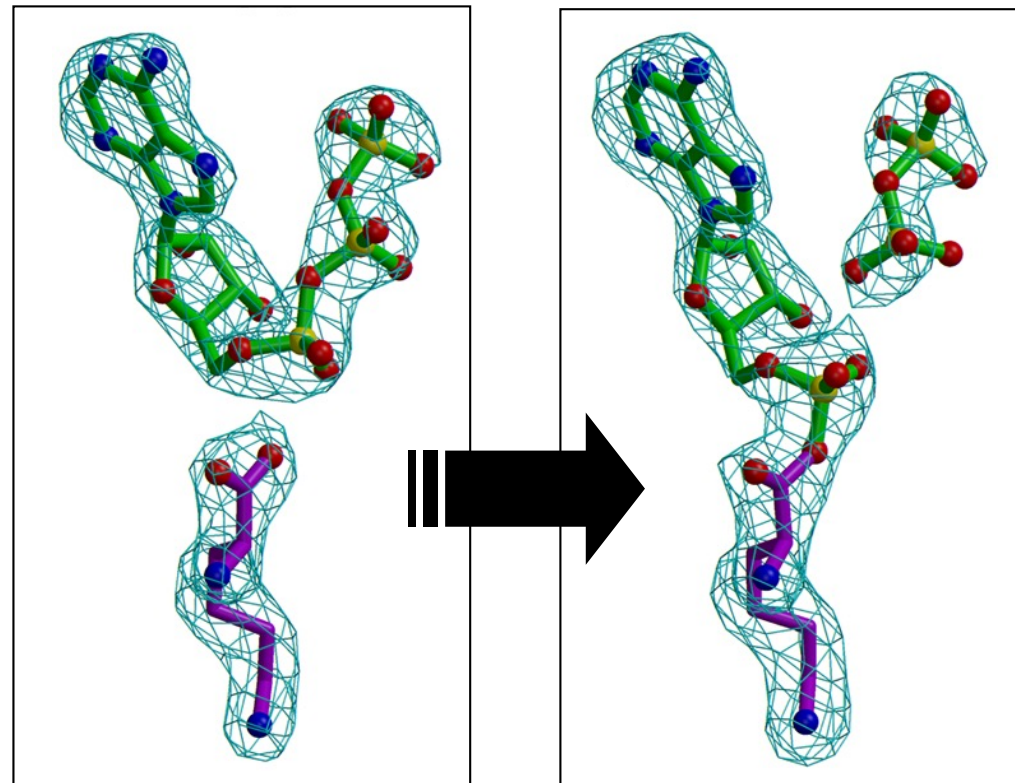


Does crystallisation affects the structure?

➡ 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.

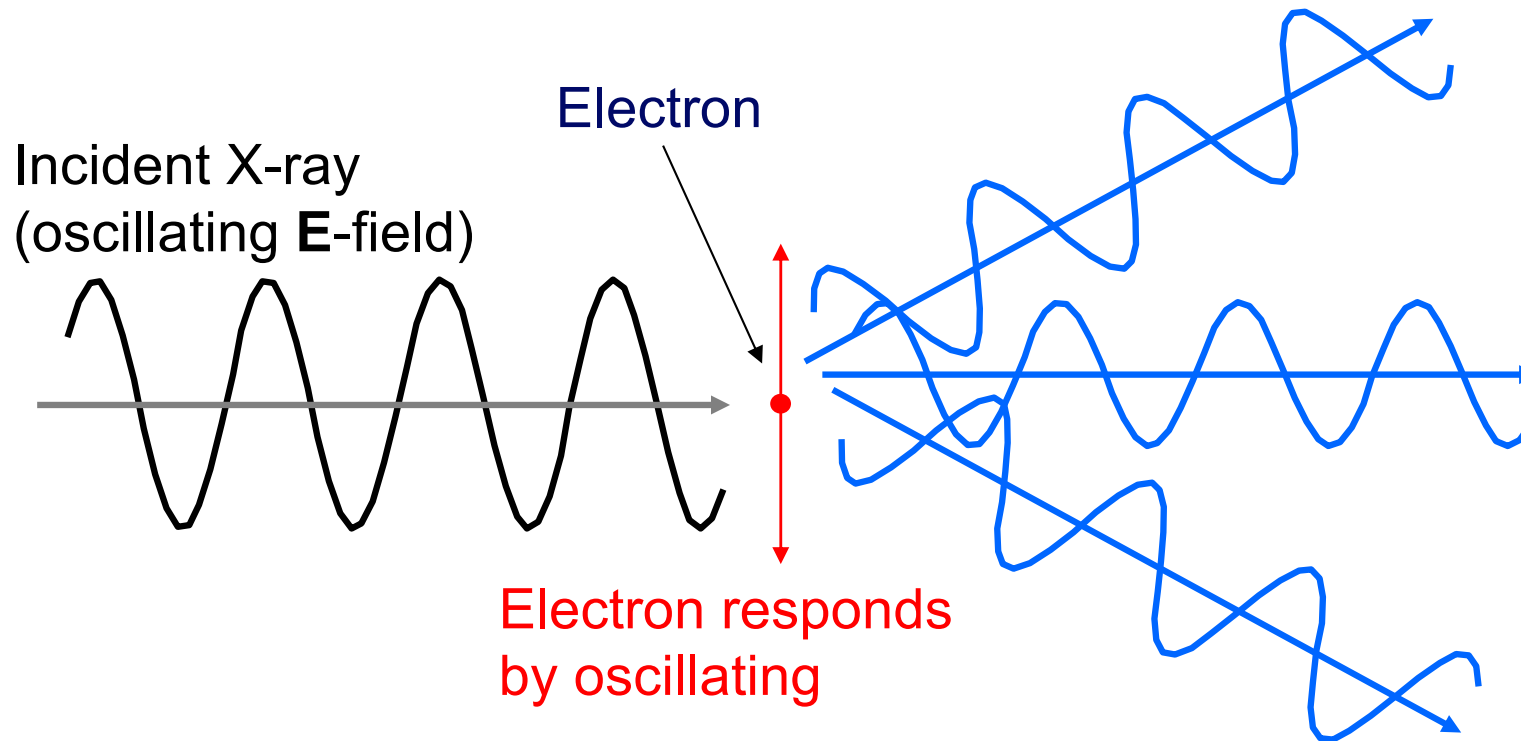
➡ Proteins crystallised in different crystal forms are often identical or almost identical.

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



X-ray scattering by a single electron

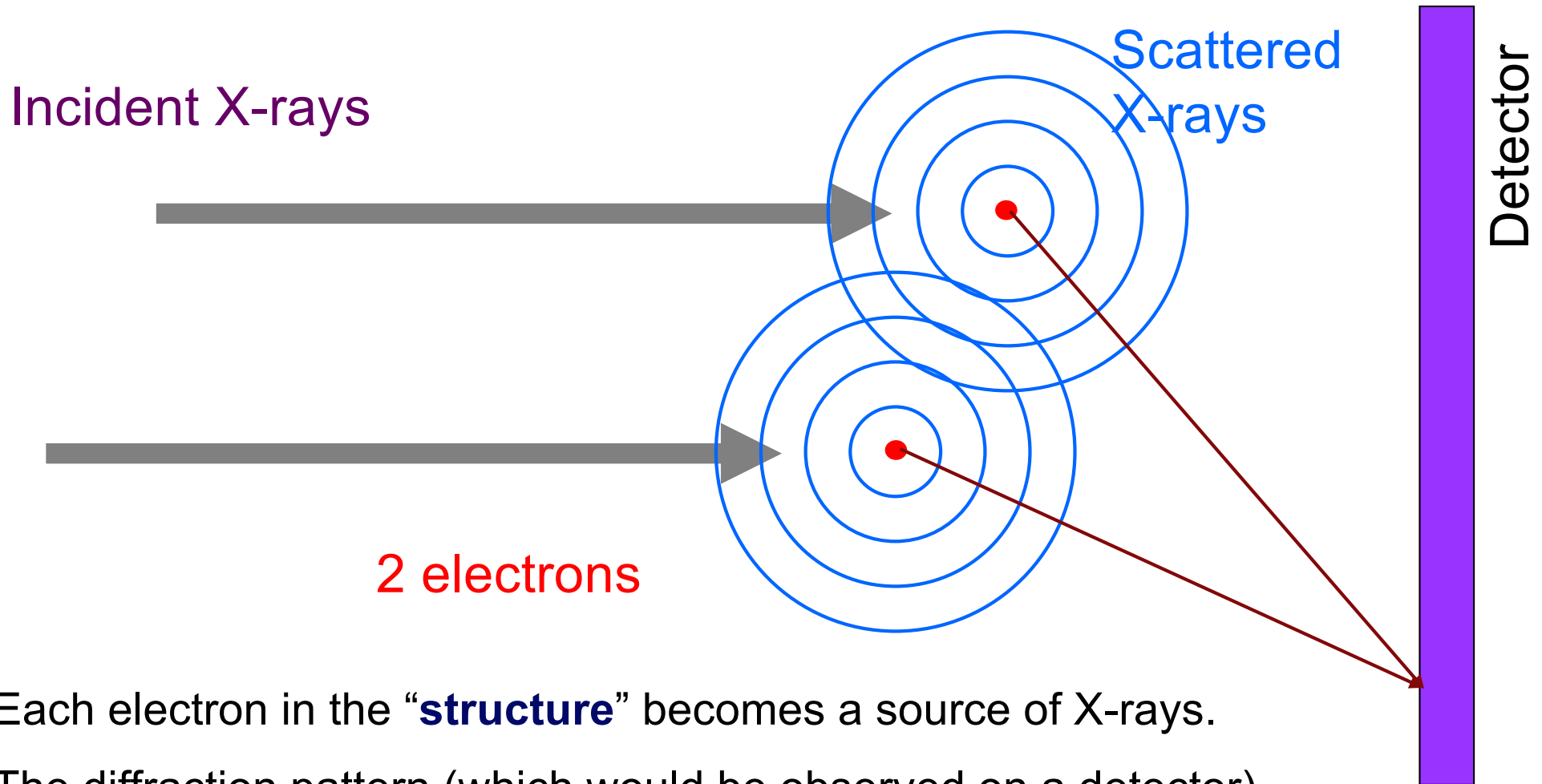
➡ the scattering of X-rays is mostly due to electrons



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.

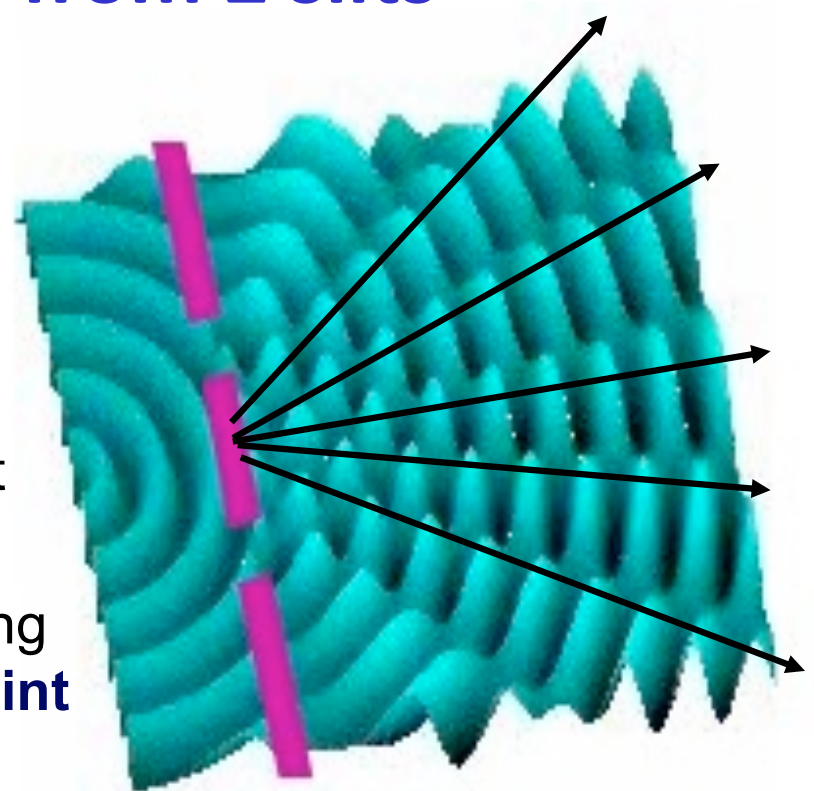
To understand diffraction, we need to know **how to add waves**, i.e. how waves interfere with each other.

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.

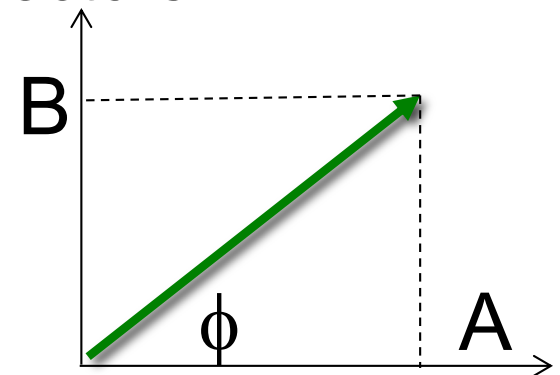
If we can work out this relationship, we can measure the diffraction pattern and figure out the “structure” that gives rise to it.

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



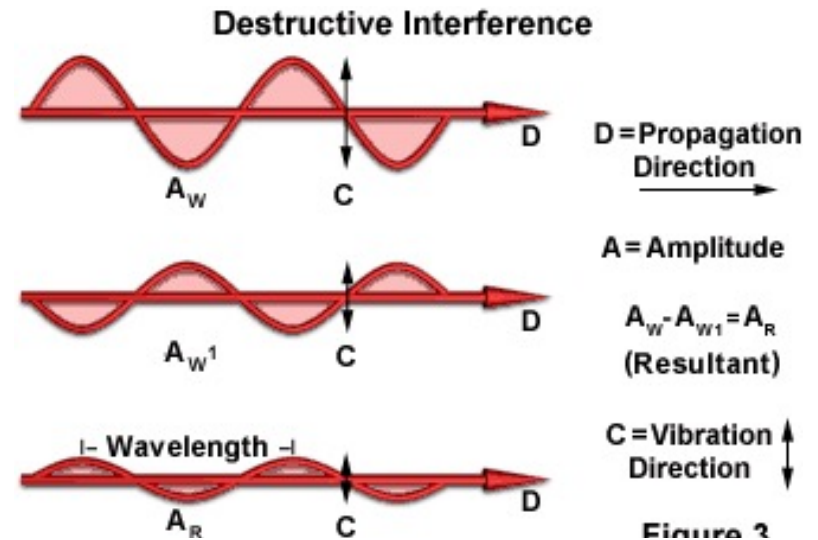
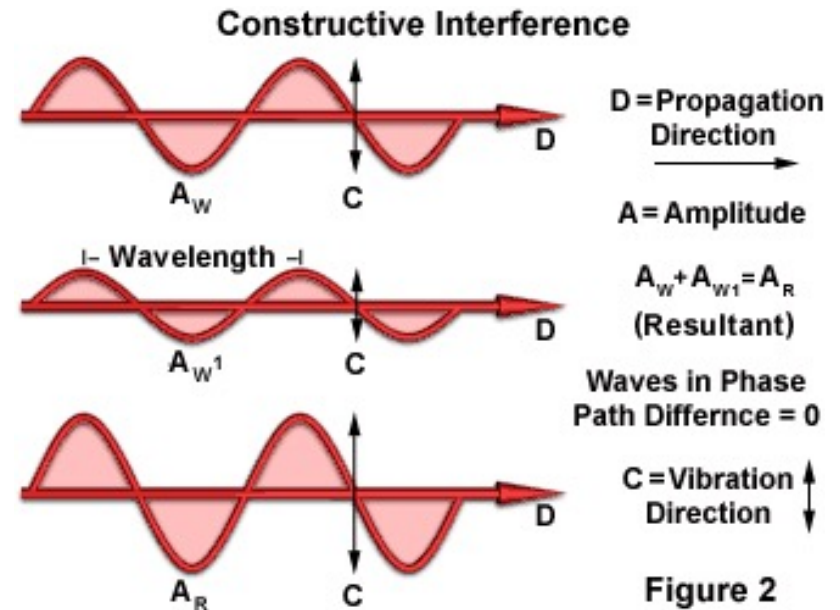
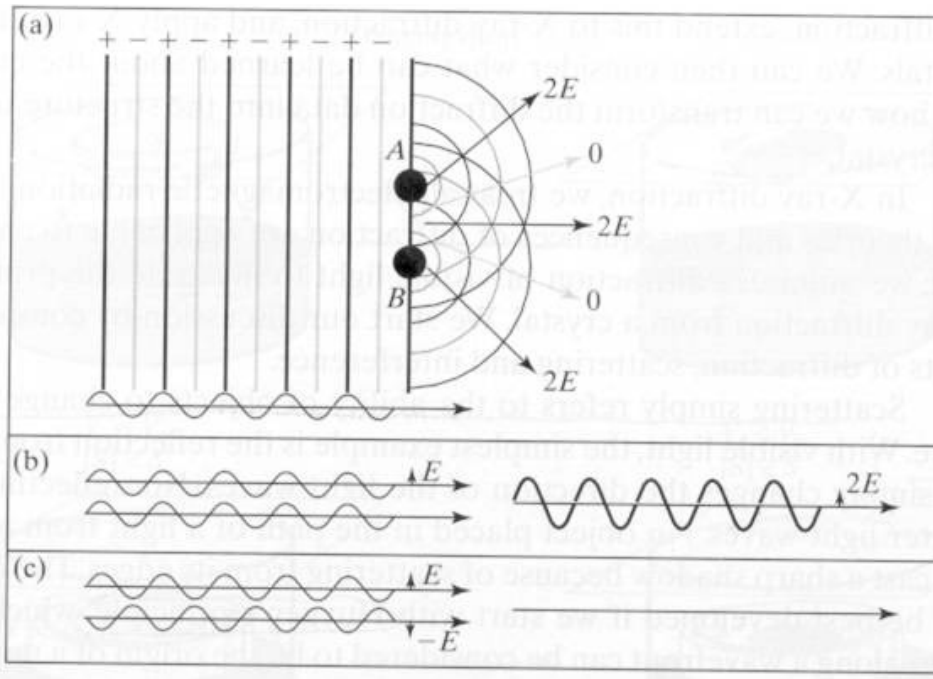
Waves can be written as complex numbers/vectors:

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

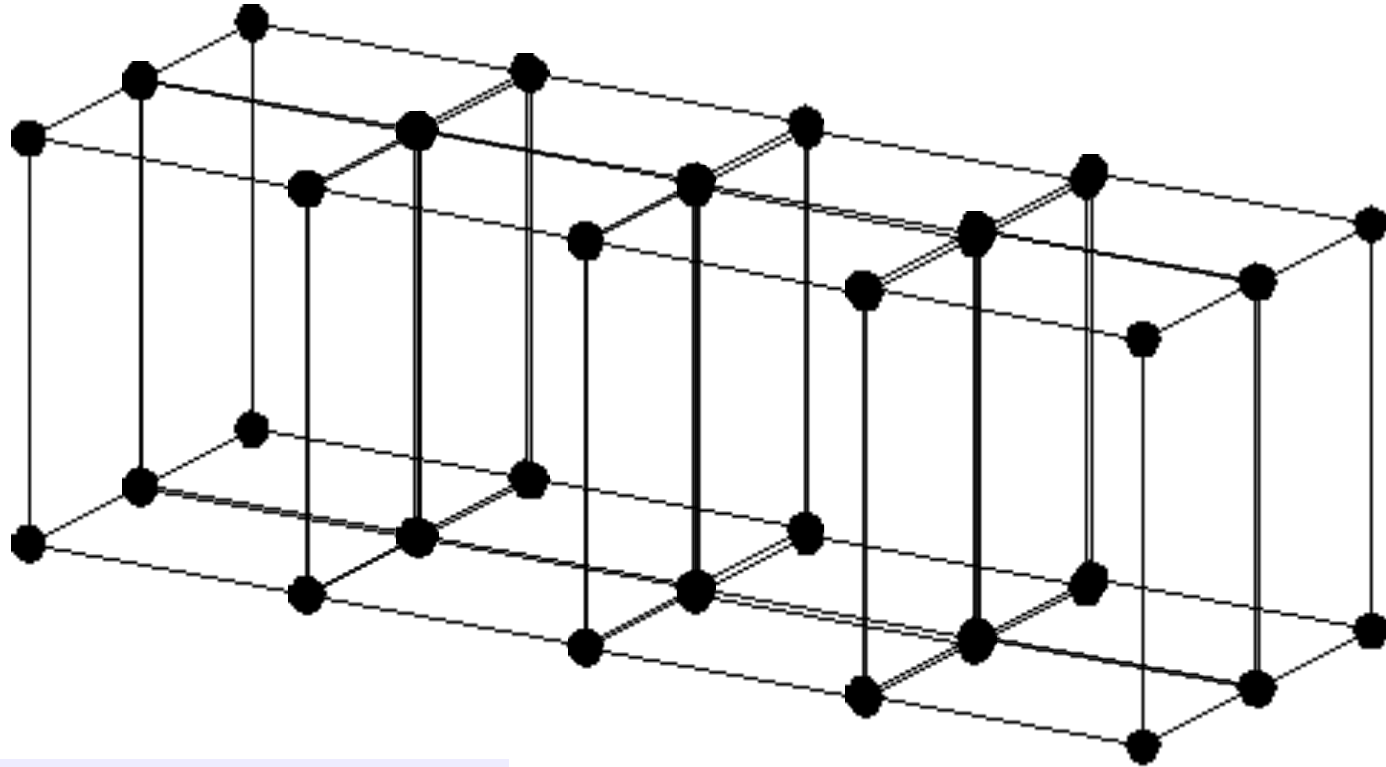


X-ray scattering: interference

- 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)



Crystals: unit cells

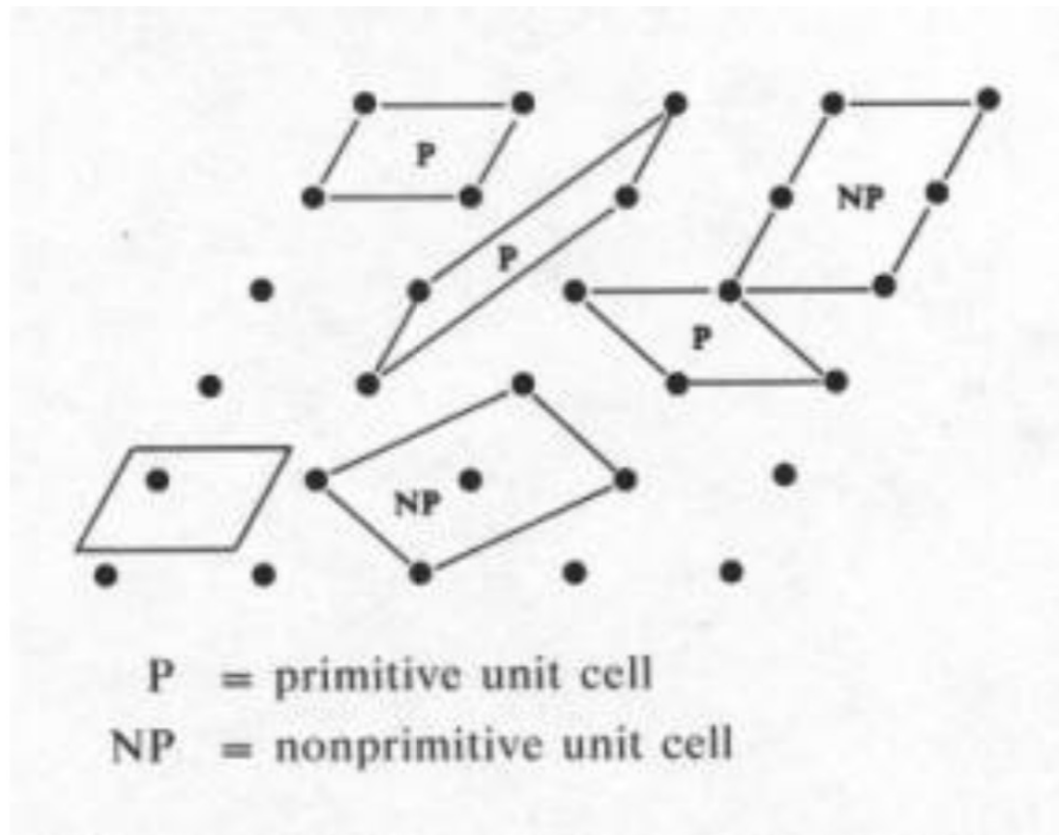


$$\vec{T} = u\vec{a} + v\vec{b} + w\vec{c}$$

u, v, w interi

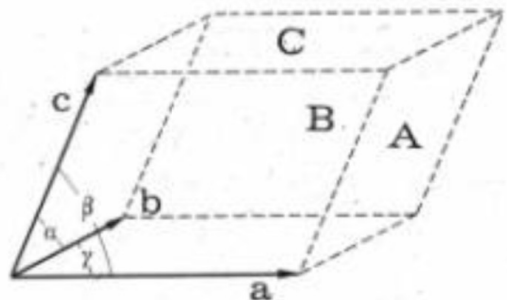
\vec{T} is a vector of the Bravais Lattice

Crystals: unit cells



Si può individuare nel motivo che si ripete, costituito da ioni e atomi, una porzione di spazio che contiene tale motivo e traslata che genera il cristallo: la **cella elementare** (quella con il volume 'minimo' e la forma associata alla massima simmetria). Una cella elementare può essere più o meno simmetrica (possedere degli elementi di simmetria).

Crystals: unit cells (Primitive cells)



volume V della cella elementare
 $V = a \cdot (b \times c)$

L'orientazione dei tre assi cristallografici è scelta in modo da dare una terna destrorsa. Si indicano con A, B e C le facce della cella elementare opposte ad a, b, c. Se la cella elementare è primitiva (contiene un singolo nodo), i valori m_i sono vincolati ad essere interi per qualunque nodo del reticolo. Se è centrata allora i valori di m_i saranno in genere numeri razionali. Per la caratterizzazione della cella basterà ricordare che un nodo ai vertici della cella le appartiene per 1/8, un nodo su uno spigolo le appartiene per 1/4, un nodo su una faccia le appartiene per 1/2.

I sette sistemi cristallini

Trimetrici

$E \ 0 \ i$	Triclino	$a \neq b \neq c$
$1 \ 0 \ -1$		$\alpha \neq \beta \neq \gamma$
$C_2 \ 0 \ \sigma$	Monoclino	$a \neq b \neq c$
$2 \ 0 \ -2$		$\alpha = \beta = 90^\circ \neq \gamma$
		$\alpha = \gamma = 90^\circ \neq \beta$
$C_2 \ 0 \ \sigma$	Ortorombico	$a \neq b \neq c$
$2 \ 0 \ -2$		$\alpha = \beta = \gamma = 90^\circ$

Dimetrici

$C_4 \ 0 \ S_4$	Tetragonale	$a = b \neq c$
$4 \ 0 \ -4$		$\alpha = \beta = \gamma = 90^\circ$
$C_6 \ 0 \ S_3$	Esagonale	$a = b \neq c$
$6 \ 0 \ -6$		$\alpha = \beta = 90^\circ \ \gamma = 120^\circ$
$C_3 \ 0 \ S_6$	Trigonale	come l'esagonale
$3 \ 0 \ -3$	(Romboedrico)	$a = b = c$
		$\alpha = \beta = \gamma \neq 90^\circ$

Monometrico

4 assi ternari	Cubico	$a = b = c$
		$\alpha = \beta = \gamma = 90^\circ$

Crystals: unit cells (Non Primitive)

Centratura dei reticoli.

Si procede in modo diretto. Si aggiungono punti a ciascun reticolo dei diversi sistemi cristallini e si valuta se: (a) l'insieme dei nodi è ancora un reticolo (l'intorno di ogni nodo deve essere lo stesso), e (b) si tratta di un nuovo reticolo nello stesso sistema cristallino.

Le possibili centrature di una cella primitiva sono disposte (a partire dall'origine) in:

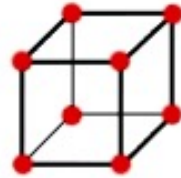
(I) a corpo centrato ($a/2 + b/2 + c/2$)

(F) a facce centrate ($a/2 + b/2$), ($a/2 + c/2$), e ($b/2 + c/2$)

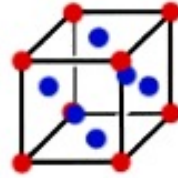
(A) A-centrata ($b/2 + c/2$) (B) B-centrata ($a/2 + c/2$) (C) C-centrata ($a/2 + b/2$)

L'analisi sistema per sistema porta ad accertare che esistono solo sette nuovi reticoli non primitivi (centrati).

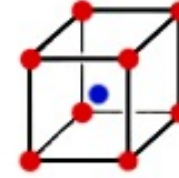
The 14 Bravais Lattices



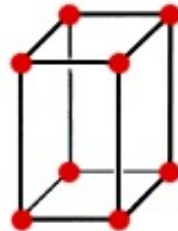
Simple cubic



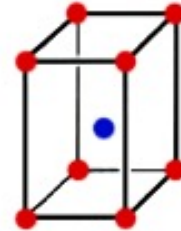
Face-centered cubic



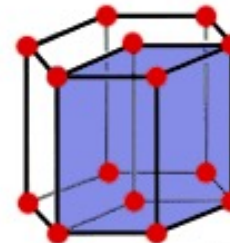
Body-centered cubic



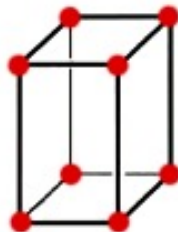
Simple tetragonal



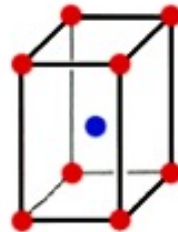
Body-centered tetragonal



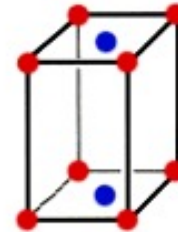
Hexagonal



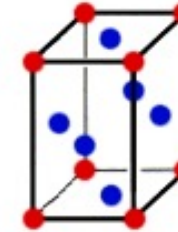
Simple orthorhombic



Body-centered orthorhombic



Base-centered orthorhombic



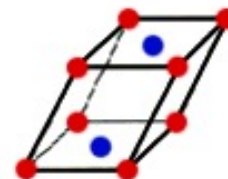
Face-centered orthorhombic



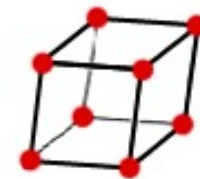
Rhombohedral



Simple monoclinic



Base-centered monoclinic



Triclinic

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	$2_1, 3_1, \text{etc.}$	—

* 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

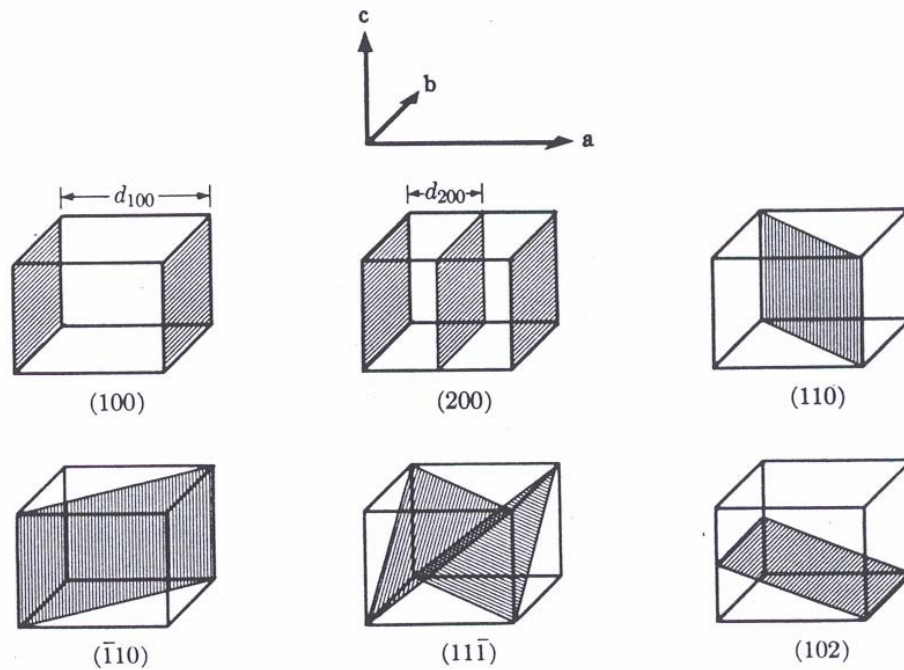


Table 7.3
THE NONCUBIC CRYSTALLOGRAPHIC POINT GROUPS^a

SCHOENFLIES	HEXAGONAL	TETRAGONAL	TRIGONAL	ORTHO-RHOMBIC	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			$n2\bar{2}$ (n even) $n2$ (n odd)
D_{nh}	D_{6h} $6/mmm$	D_{4h} $4/mmm$		D_{2h} ($m\bar{3}m$) (V_h) $2/m\bar{2}m\bar{2}$			$\frac{n}{m} \frac{2}{m} \frac{2}{m}$ ($n/m\bar{3}m$)
		D_{3h} $\bar{6}2m$					$\bar{n}2m$ (n even)
D_{nd}		D_{2d} (V_d) $\bar{4}2m$	D_{3d} ($\bar{3}m$) $\bar{3} \frac{2}{m}$				$\bar{n} \frac{2}{m}$ (n odd)

^a Table caption on p. 123.

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	<u>32</u>	<u>14</u>	<u>61</u>

The space groups

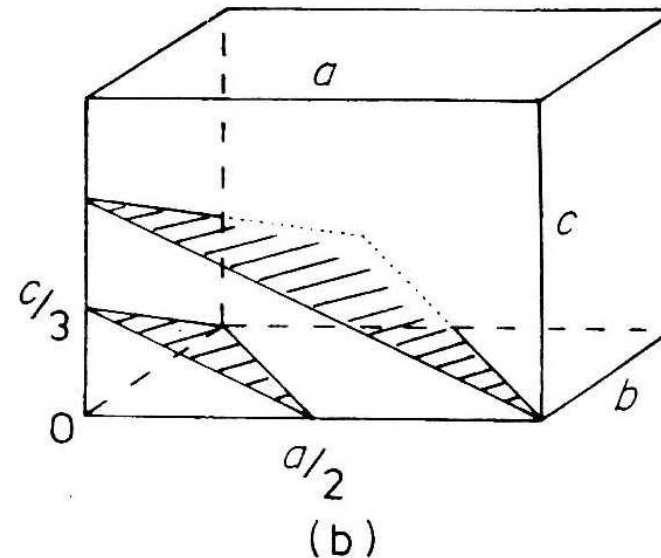
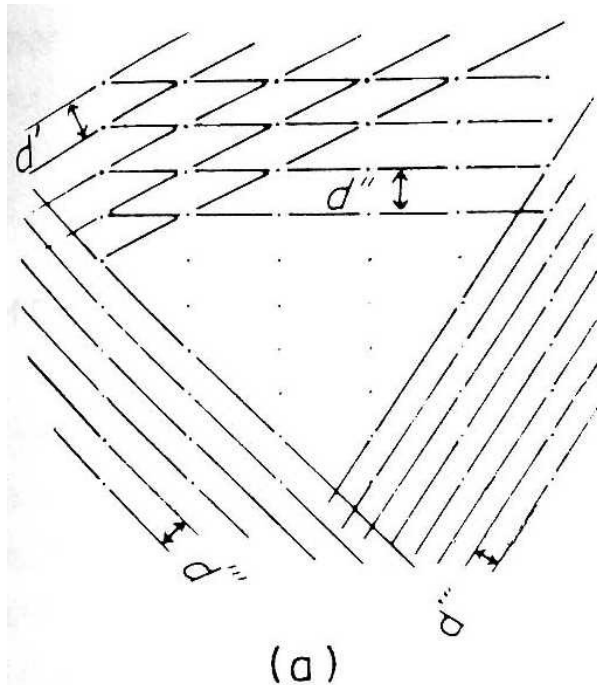
- What symmetry operations (e.g. rotation axes, (2-, 3-, 4-, 6- fold axis, mirrors, inversion axes ..., at corner, at face, ...) can be applied to the unit cell (inside crystal)? This defines the 32 point groups of the unit cell (see Table 1.4).
- The combination of the 32 symmetry types (point groups)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

The crystal is anisotropic. Can be described by different families of parallel, equidistant planes.

To define a “family” of planes (same distance) **axes intercepts** are used :

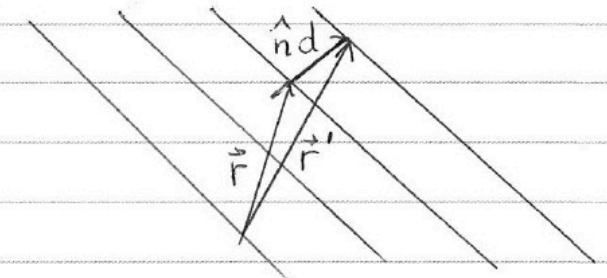
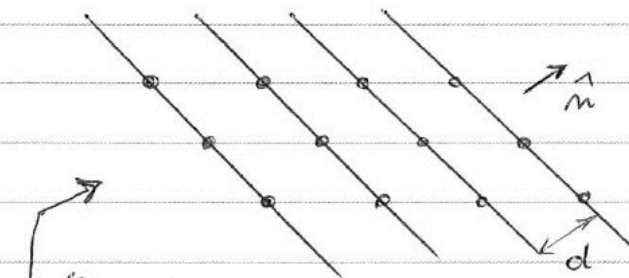
$$a/h, b/k, c/l$$



(h, k, l) are called Miller indices

FAMILY OF LATTICE PLANES and \vec{k}

(1)



● Consider a FAMILY OF LATTICE PLANES
 $\perp \hat{n}$ and equispaced by d .

The eq. of a plane (in general, in the continuum space)

→ (a) A family of lattice planes $\perp \hat{n}$ and equispaced by d are defined by:
 a vector of the reciprocal lattice:
 $\vec{k} = \frac{2\pi}{d} \hat{n}$

More precisely, it has the same value on all the planes:
 $e^{i\vec{k} \cdot \vec{r}'} = e^{i\vec{k} \cdot (\vec{r} + m\hat{n}d)} = e^{i\vec{k} \cdot \vec{r}}$

Without any loss of generality one of these planes can contain the origin and therefore $e^{i\vec{k} \cdot \vec{r}} = 1$

on the whole family of planes. Such planes are lattice planes, contain all the Bravais lattice points, therefore

$\vec{k} = \frac{2\pi}{d} \hat{n}$ satisfy $e^{i\vec{k} \cdot \vec{R}} = 1 \quad \forall \vec{R} \in \text{Bravais lattice}$

⇒ \vec{k} is a reciprocal lattice vector \vec{k}

→
(a)

A family of
lattice planes
 $\perp \hat{m}$ and
equispaced by d




are defined by:
a vector of the
reciprocal lattice:
 $\vec{k} = \frac{2\pi}{d} \hat{m}$

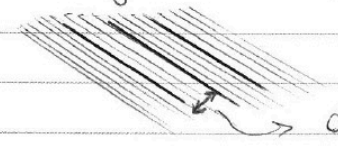
(b) Each vector $\vec{k} \in$ reciprocal lattice \Rightarrow defines a family of lattice planes $\perp \vec{k}$ and with interplanar distance $d = 2\pi/k^*$ where \vec{k}^* is the shortest rec. lattice vector $\vec{k}^* \parallel \vec{k}$

- Consider $\vec{k} \in$ reciprocal lattice
 - Consider the planes in real space with \vec{r} such that: $\boxed{e^{i\vec{k} \cdot \vec{r}} = 1}$ (A)
- For sure, one of them contains the origin, since $e^{i\vec{k} \cdot (000)} = 1$
- But the eq. (A) defines many planes which are $\perp \vec{k}$ (all the planes with \vec{r} such that $\vec{r} \cdot \vec{k} = 2\pi m$) and equispaced by $d = 2\pi/k$ (spacing)
- (since, if we consider $\vec{r}' = \vec{r} + \frac{2\pi}{k} \hat{k}$, we have:
- $$e^{i\vec{k} \cdot \vec{r}'} = e^{i\vec{k} \cdot \vec{r}} \cdot e^{i\vec{k} \cdot \frac{2\pi \hat{k}}{k}} = e^{i\vec{k} \cdot \vec{r}} = 1$$

By def of $\vec{k} \in$ rec. lattice, all the Bravais lattice vectors \vec{R} are such that $\boxed{e^{i\vec{k} \cdot \vec{R}} = 1}$



Hence, they must lie on these planes $\perp \vec{k}$ and equispaced by d . But they can be a subset of the set in the continuous space equispaced by a certain distance d'

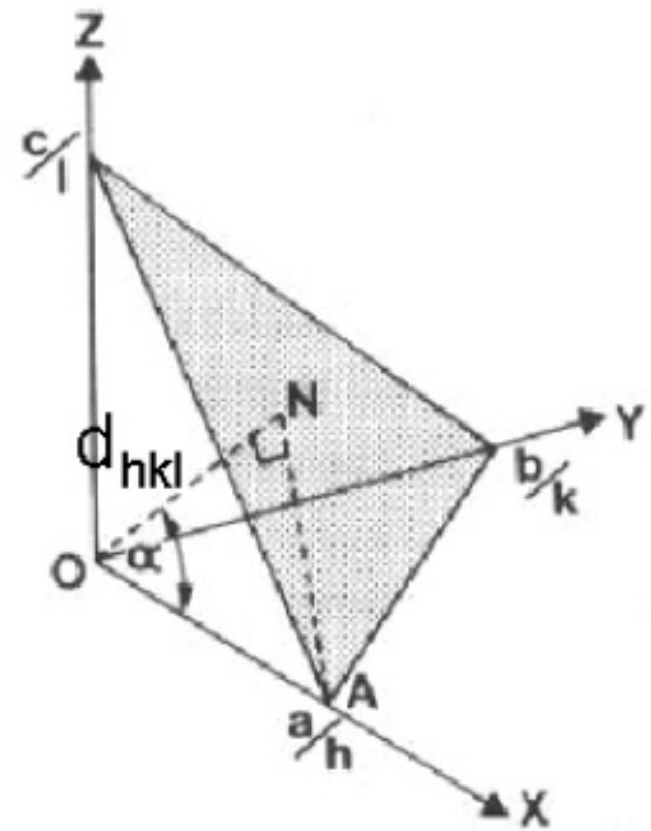
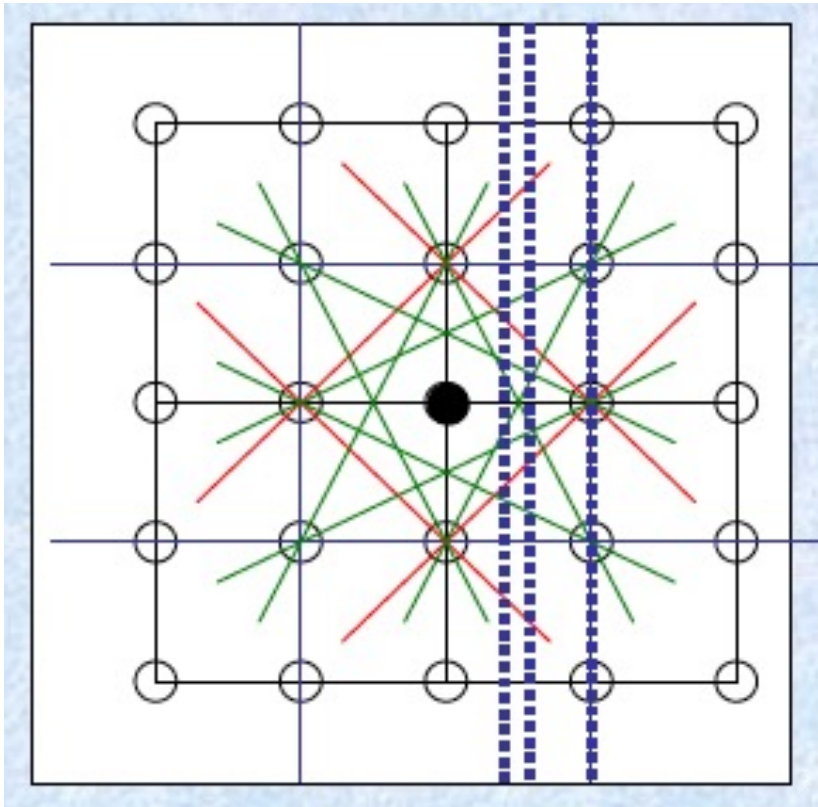


distance d'

For (a), they are defined by $\vec{k}^* = \frac{2\pi n}{d'}$

Everything is consistent if $\vec{k} = n \vec{k}^*$ i.e., \vec{k}^* is the shortest vector \in reciprocal lattice, $\parallel \vec{k}$.

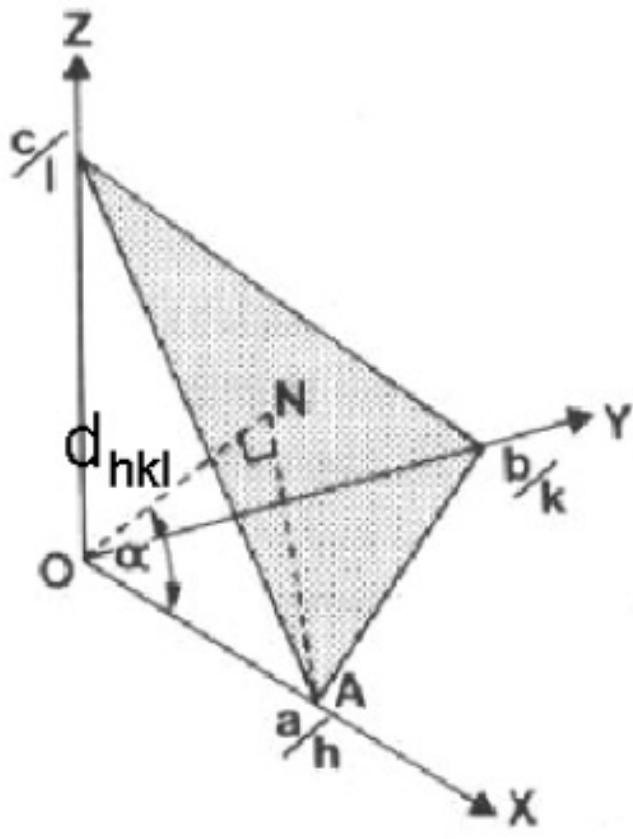
The reciprocal space



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



$$(a/h) \cos \alpha = d_{hkl}$$

$$\cos \alpha = (h/a) d_{hkl}$$

$$\cos \beta = (k/b) d_{hkl}$$

$$\cos \gamma = (l/c) d_{hkl}$$

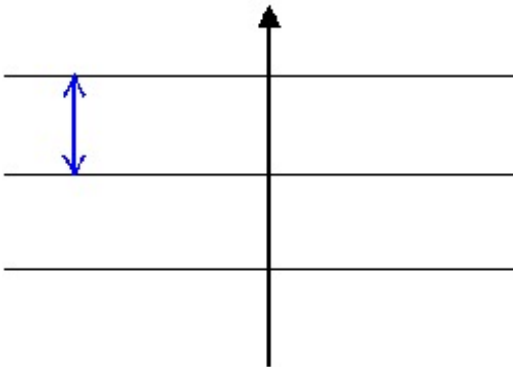
For orthorombic lattice:

$$(\cos \alpha)^2 + (\cos \beta)^2 + (\cos \gamma)^2 = 1$$

$$(h/a)^2 d_{hkl}^2 + (k/b)^2 d_{hkl}^2 + (l/c)^2 d_{hkl}^2 = 1$$

For a cubic crystal:

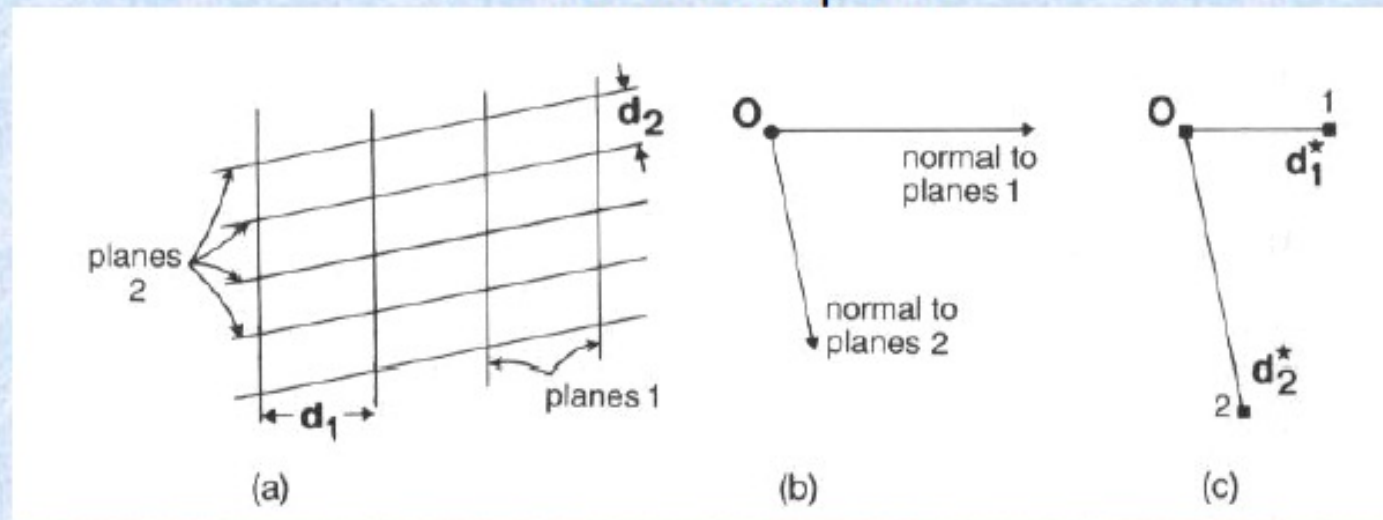
$$1/d_{hkl}^2 = 1/a^2 * (h^2 + k^2 + l^2)$$



The reciprocal space

Regole di costruzione del reticolo reciproco:

Si prende un'origine comune, da cui partono i vettori che caratterizzano le normali ai piani.



Si troncano i vettori a valori d_1^* e d_2^* , lunghi K/d_1 e K/d_2 (K è una costante arbitraria, normalmente pari a 1)

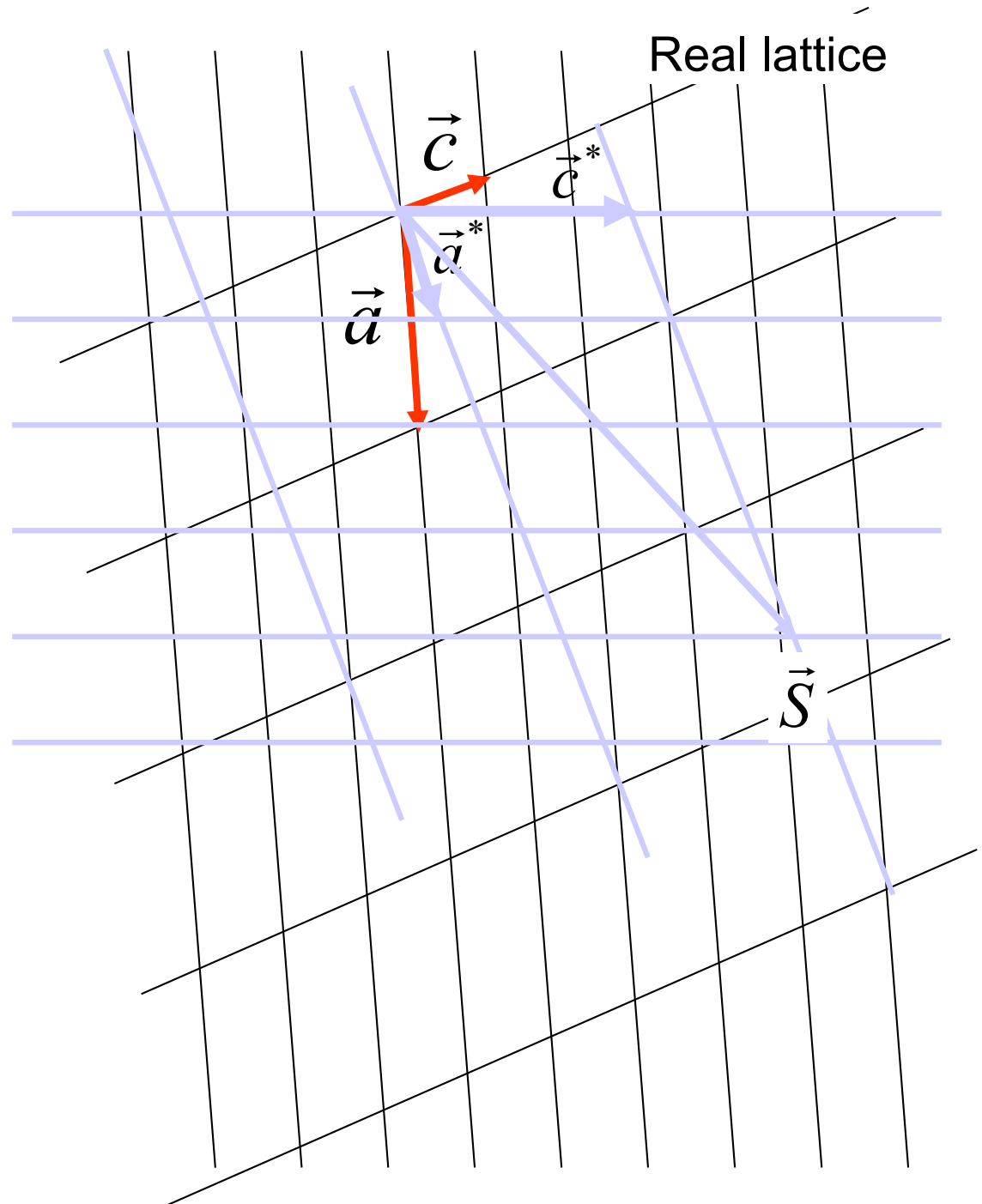
d_1^* e d_2^* sono vettori del reticolo reciproco, si misurano in $[\text{Lunghezza}]^{-1}$, ad esempio \AA^{-1}

Ne consegue che se $d_1 = 0.5 \text{\AA}$ e $K = 1$, $|d_1^*| = 1/0.5 \text{\AA} = 2.0 \text{\AA}^{-1}$

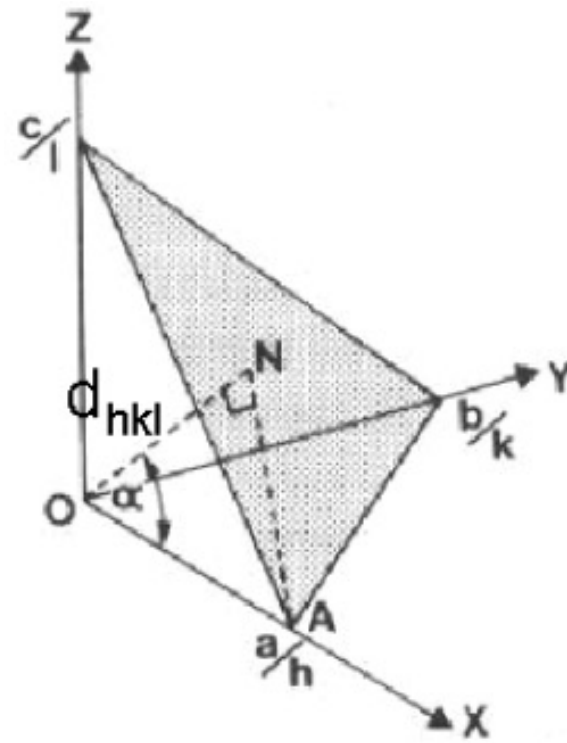
Reciprocal lattice

$$\begin{aligned}\vec{a} \cdot \vec{a}^* &= 1 \\ \vec{b} \cdot \vec{b}^* &= 1 \\ \vec{c} \cdot \vec{c}^* &= 1 \\ \vec{a} \cdot \vec{b}^* &= 0 \\ \vec{a} \cdot \vec{c}^* &= 0 \\ \vec{b} \cdot \vec{a}^* &= 0 \\ \vec{b} \cdot \vec{c}^* &= 0 \\ \vec{c} \cdot \vec{a}^* &= 0 \\ \vec{c} \cdot \vec{b}^* &= 0\end{aligned}$$

\vec{a}^* is perpendicular to bc plane
 \vec{b}^* is perpendicular to ac plane
 \vec{c}^* is perpendicular to ab plane



The reciprocal space



Nello spazio diretto i punti del reticolo sono definiti dai vettori

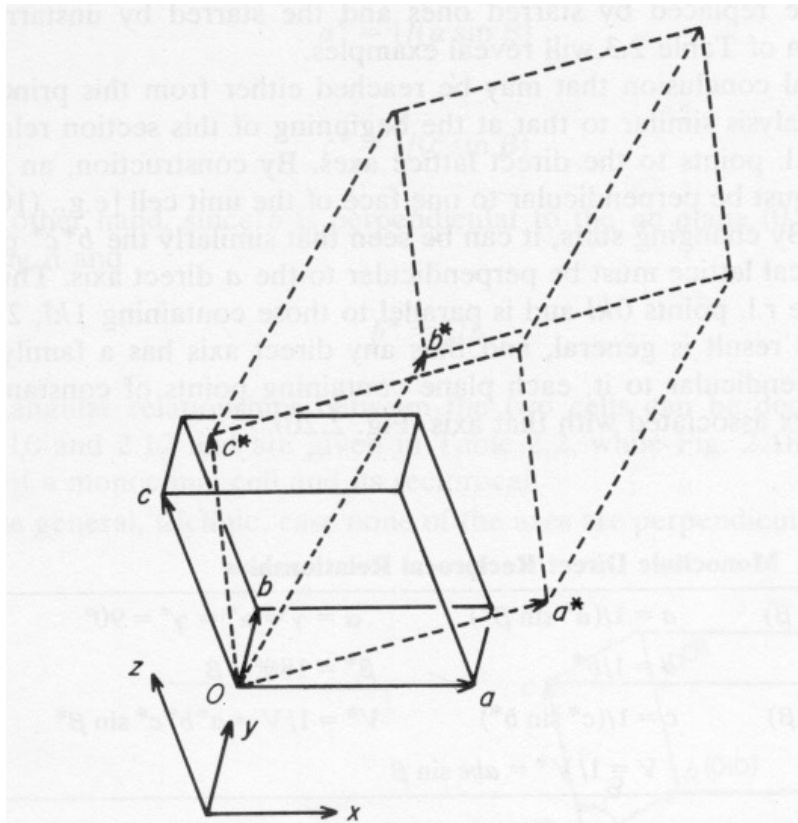
$$\mathbf{r}_{uvw} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$$

Nello spazio reciproco, $\mathbf{d}_{hkl}^* = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$,

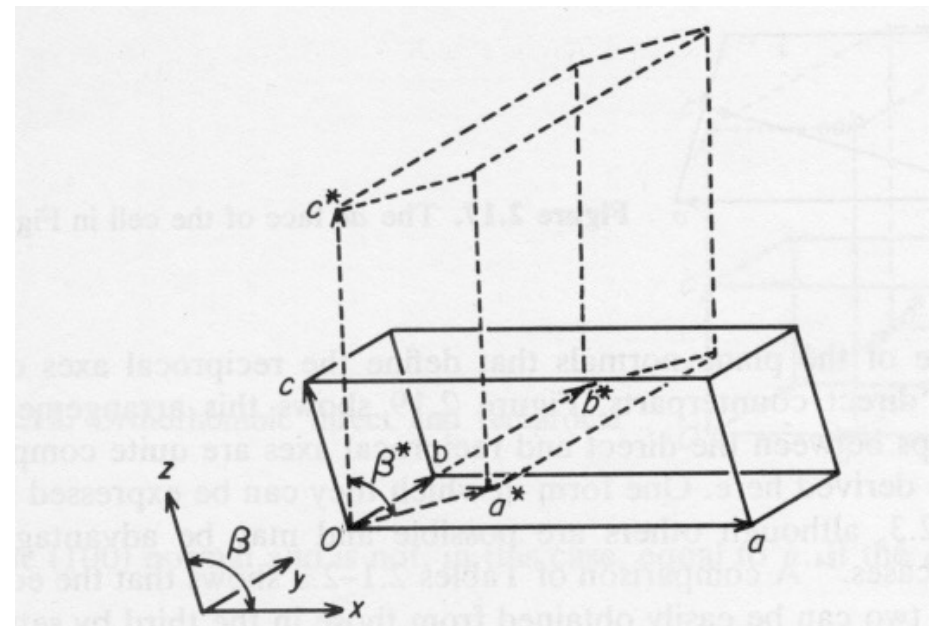
le componenti del vettore **reciproco** sono gli indici di Miller (hkl)

Examples of direct and reciprocal lattices

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

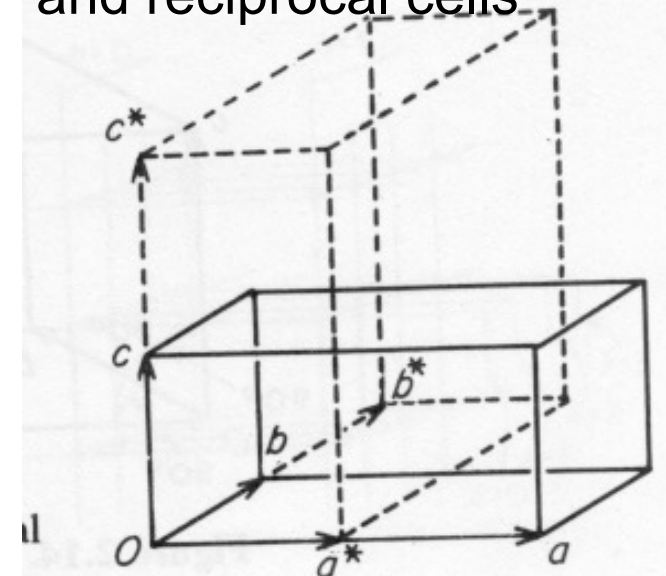


Triclinic direct and reciprocal cells

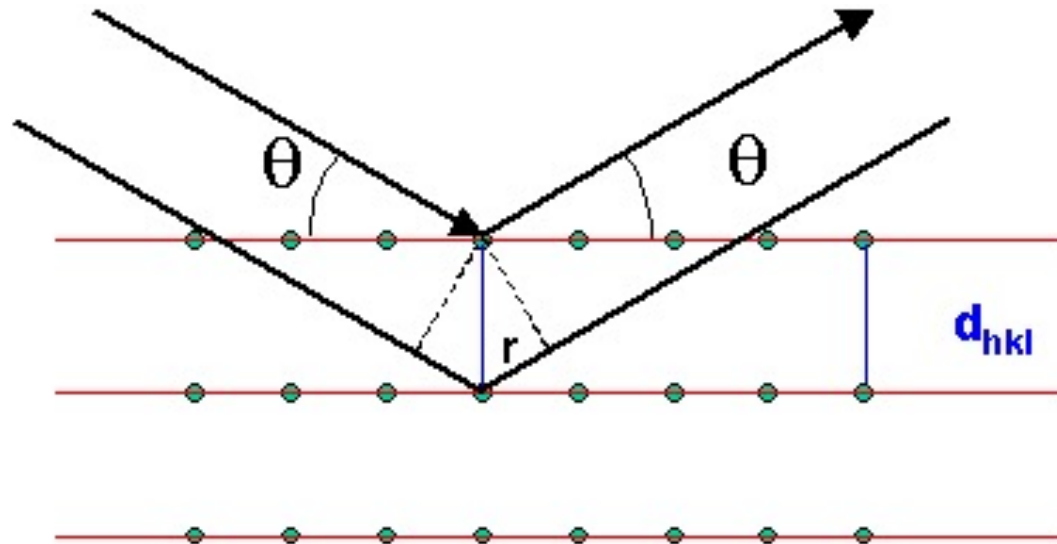


Monoclinic direct and reciprocal cells

Orthorhombic direct and reciprocal cells



Scattering by electrons: 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

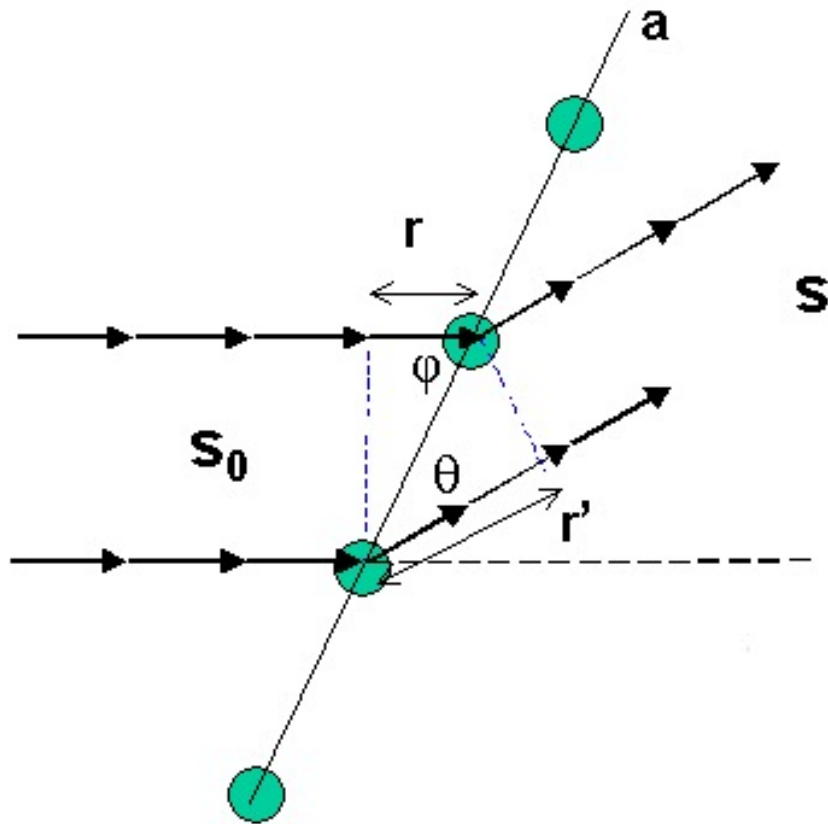
$$2d_{hkl} \sin(\theta) = n \lambda$$

Bragg's law

$$\begin{aligned} 1/d &= (2/\lambda) \sin\theta \\ 1/d &\propto \sin\theta \end{aligned}$$

When d is large, diffraction pattern is compressed

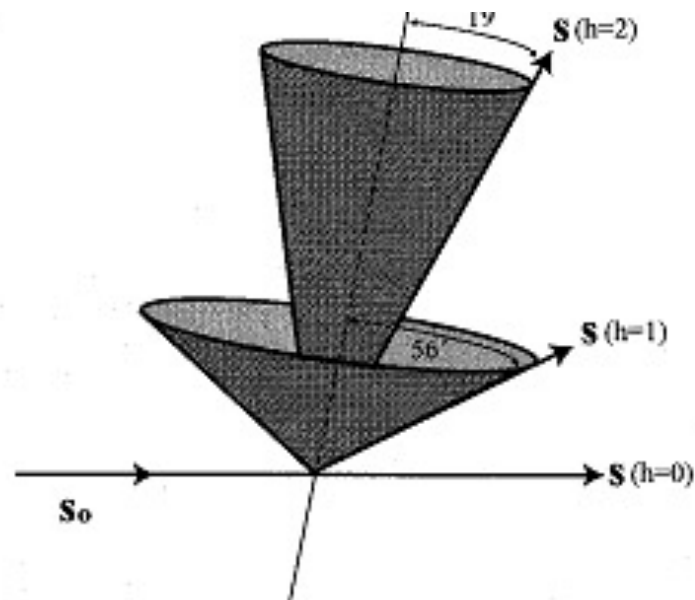
von Laue condition for diffraction



$$r' - r = a \cos(\theta) - a \cos(\phi) = h \lambda$$

h integer. In vectorial terms:

$$\mathbf{r}' - \mathbf{r} = \mathbf{a} \cdot (\mathbf{S} - \mathbf{S}_0) = h \lambda$$



(b)

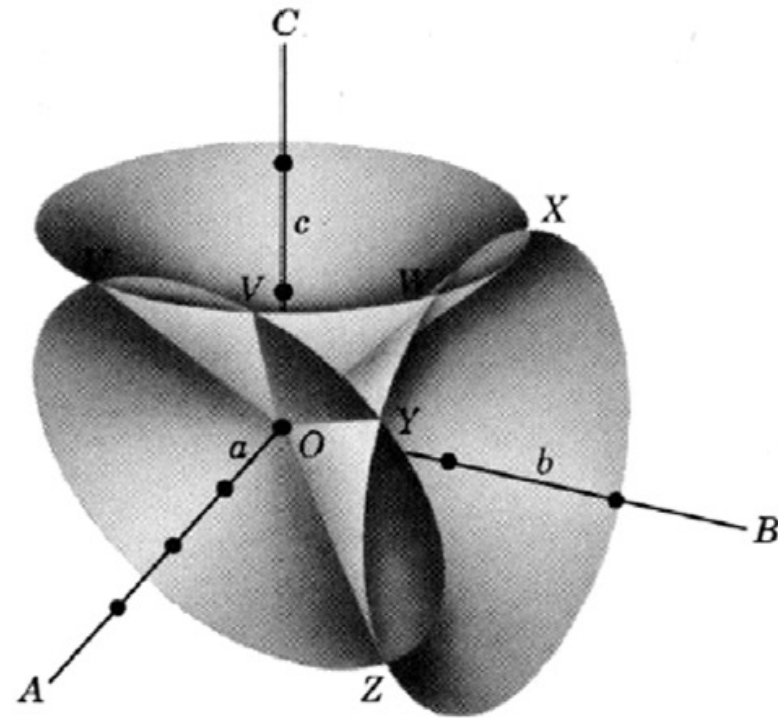
von Laue condition for diffraction

In 3D:

$$a \cdot (\mathbf{S} - \mathbf{S}_0) = h \lambda$$

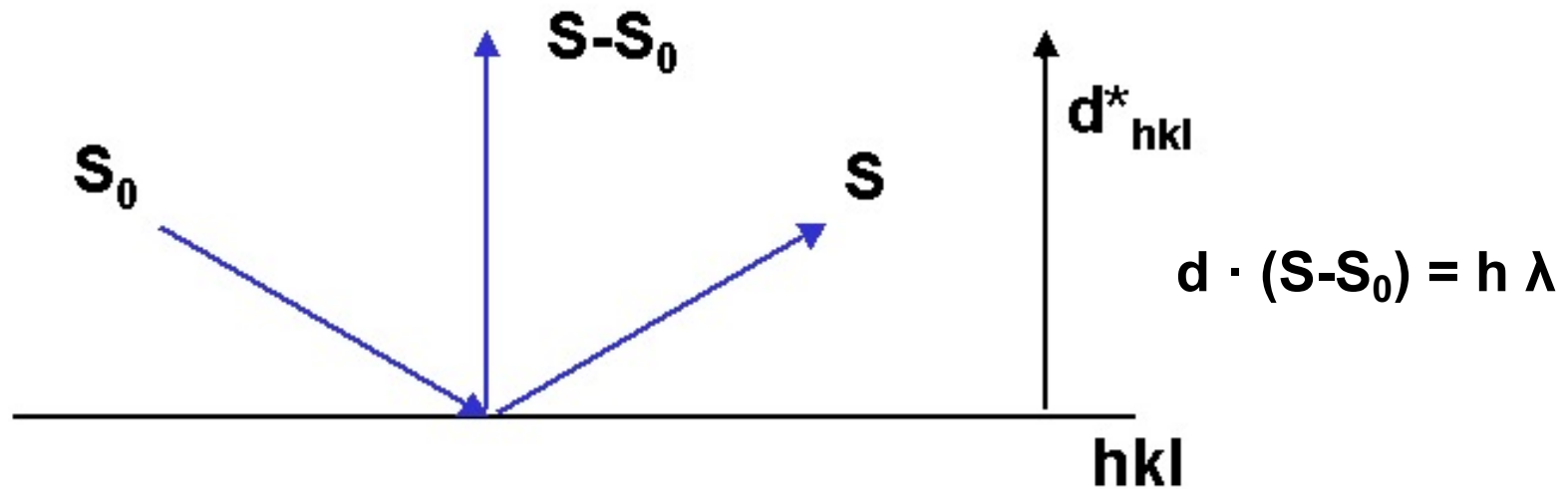
$$b \cdot (\mathbf{S} - \mathbf{S}_0) = k \lambda$$

$$c \cdot (\mathbf{S} - \mathbf{S}_0) = l \lambda$$



Diffraction is at the intersection between the 3 Laue cones

von Laue Bragg diffraction equivalence



$$|d^*_{hkl}| = 1/d_{hkl}$$

$$\lambda = 2d_{hkl} \sin(\theta) = |S-S_0| / |d^*_{hkl}|$$

From which Bragg's law become:

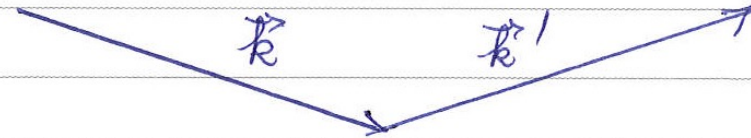
$$(S-S_0) / \lambda = d^*_{hkl} = ha^* + kb^* + lc^*$$

Bragg's law is valid when $(S-S_0)/\lambda$ is the vector of the reciprocal lattice corresponding to the diffracting planes

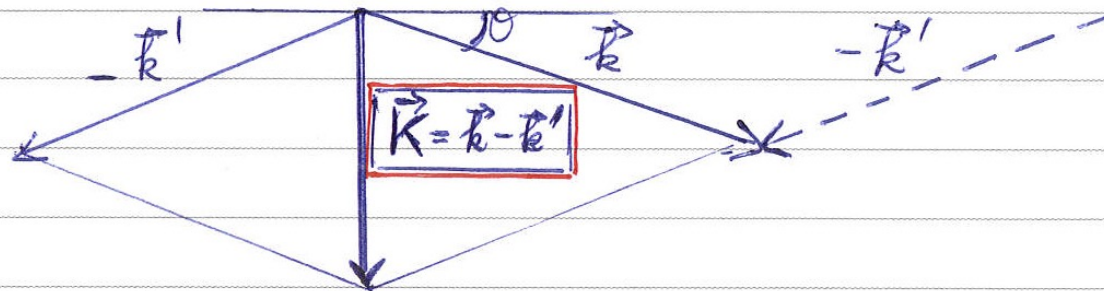
FROM VON LAUE TO BRAGG

DESCRIPTION OF THE CONDITION OF
CONSTRUCTING INTERFERENCE

Consider \vec{k}, \vec{k}' (IN, OUT)



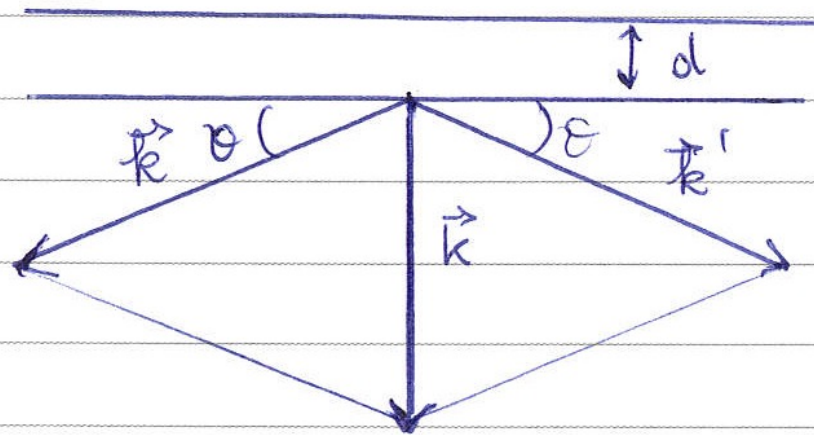
Draw $\vec{k} - \vec{k}'$, which, according to the Von Laue condition, is \vec{K} ;
to do this, it is convenient to consider $\vec{k} + (-\vec{k}')$
translating $-\vec{k}'$ on the same origin of \vec{k} .



\vec{K} defines a plane \perp to it; let ϑ be the angle formed with
this plane by the vector \vec{k} .

By construction: $|\vec{k} \sin \vartheta = \frac{K}{2}|$ (*)

\vec{K} (like any $\vec{K} \in$ reciprocal lattice) corresponds to a
family of lattice planes $\perp \hat{m} = \vec{K}/K$ and equispaced
by $d = n \frac{2\pi}{K}$ with n such that $\vec{K} = n \vec{K}^*$ \vec{K}^*
being the shortest vector $\vec{K}^* \parallel \vec{K}$.



← This is therefore the family of lattice planes. Note that θ is also the incident angle of \vec{k} w.r.t. those planes.

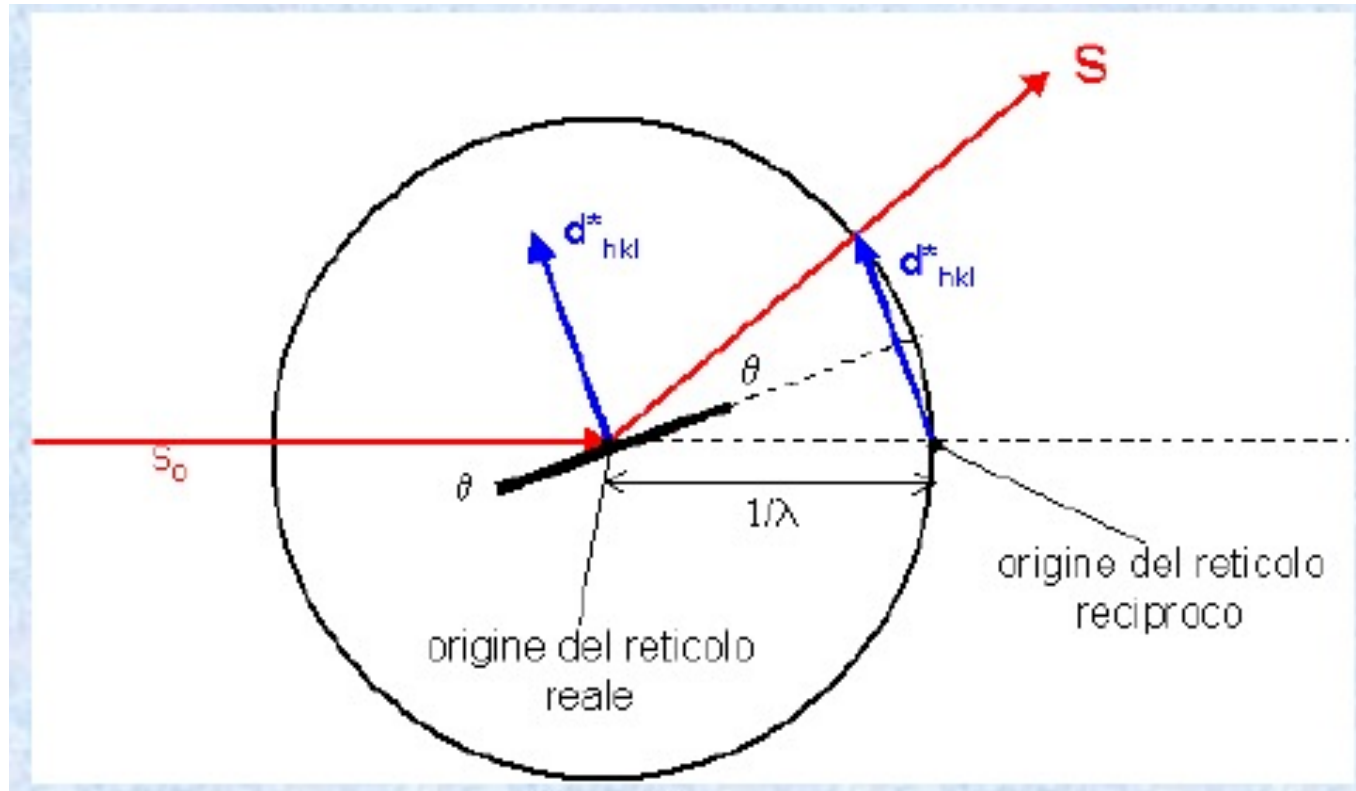
From $\textcircled{\star}$ $\Rightarrow \frac{2\pi}{\lambda} \sin \theta = n \frac{2\pi}{d} \Rightarrow$

and $\textcircled{\star\star}$

$$d \sin \theta = n \lambda$$

which is the Bragg condition.

The Ewald's sphere



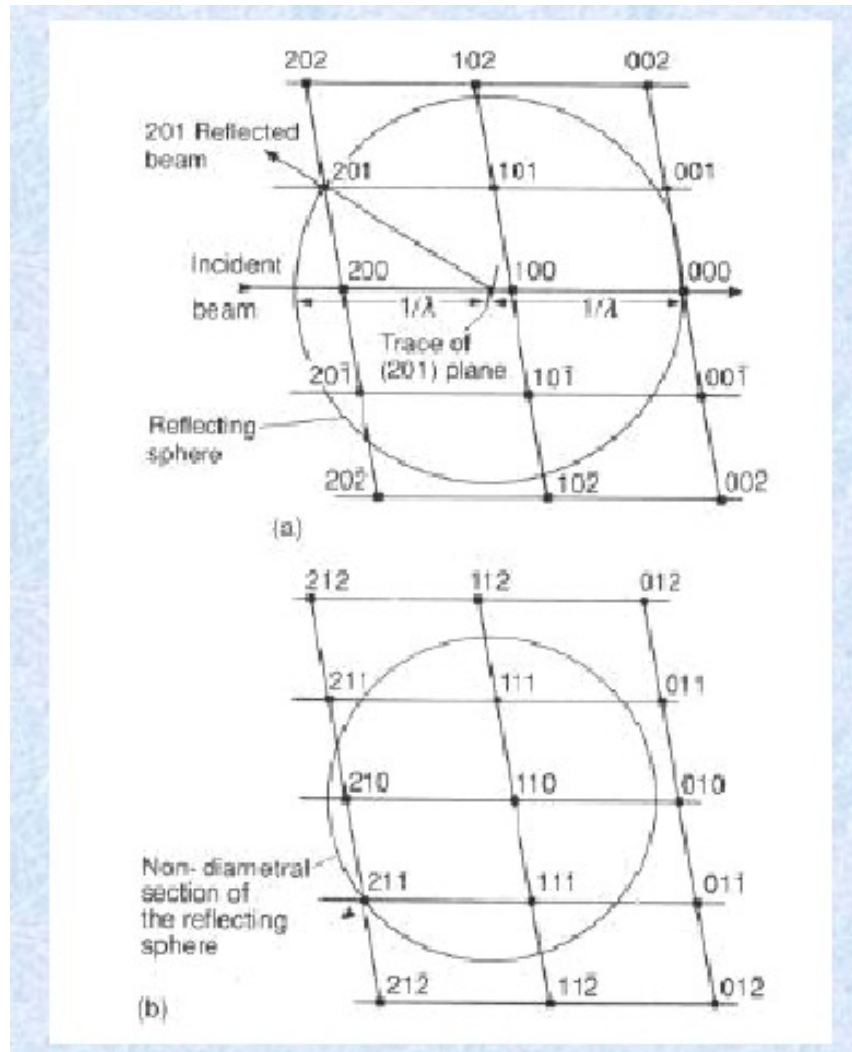
How to describe a diffraction experiment:

Draw a sphere, centered on the crystal, with radius equal to beam energy

The origin of the RL is the intersection between S_0 and the sphere.

The intersection of S with the sphere is a point of the RL corresponding to the hkl family planes

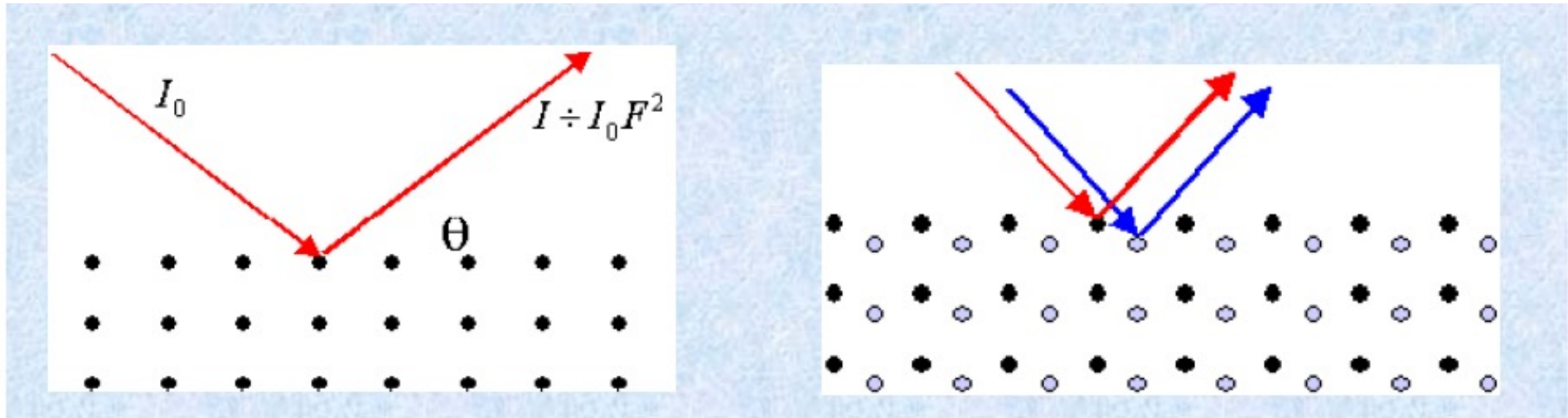
The Ewald's sphere



All the point on the sphere are diffraction points. To explore all the possible reflections:

- Rotate the crystal
- Change beam energy

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).

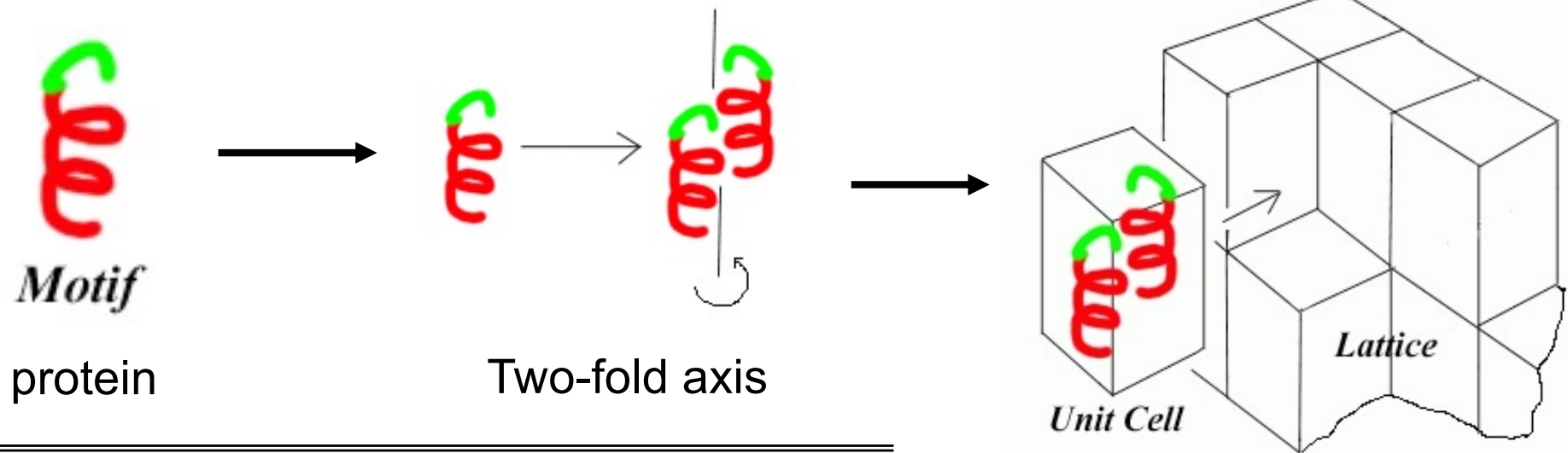
We assumed the electron density is in planes. In reality it is spread throughout the unit cell. Nevertheless, the derivation is still valid, since it can be shown that waves scattered from electron density not lying in a plane P, can be added to give a **resultant as if reflected from the plane**.

Morphology of a crystal

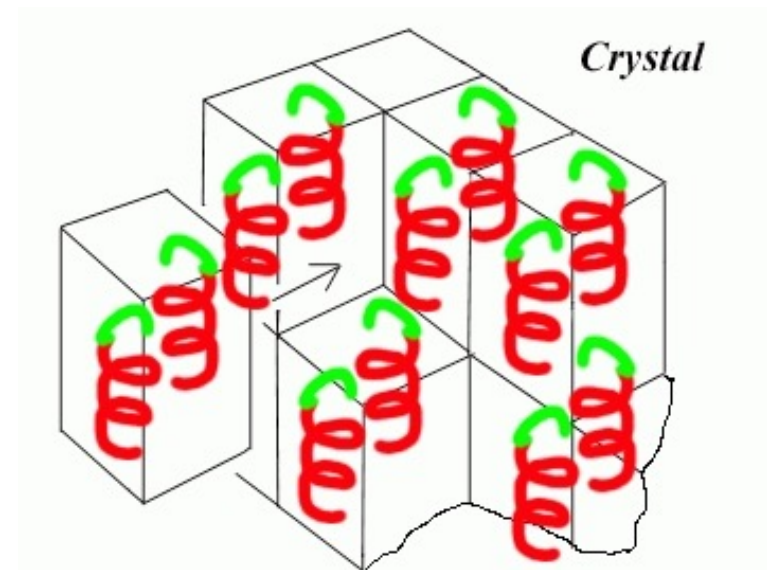
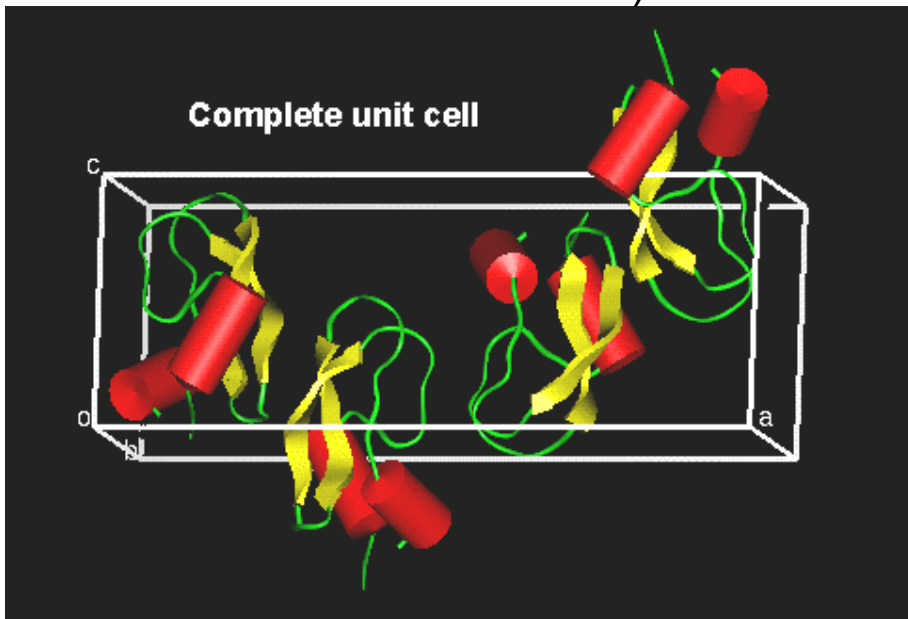
- By observing the **spacing and pattern of reflections on the diffraction pattern**, we can determine the **lengths, and angles** between each side of the **unit cell**, as well as the **symmetry or space group** in the unit cell.
- Still, how do we find out what's **inside** the unit cell? (i.e. the interesting stuff, like proteins).

Examples

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



Bovine Pancreatic Trypsin Inhibitor P $2_12_12_1$
(Primitive, orthorhombic unit cell with three two-fold screw axes)



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).

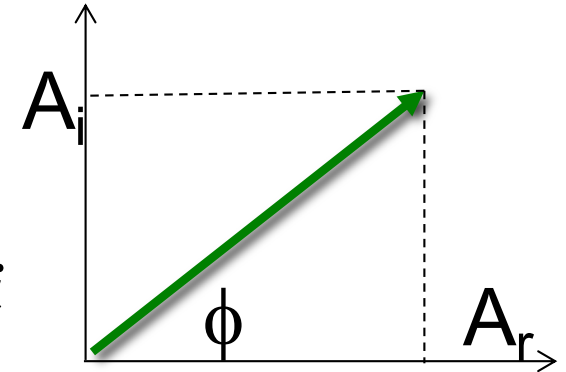
of atoms (electron density) in unit cell DOES affect **intensity** of reflection spots.

Diffraction from a multi-atom unit cell

Waves can be written as complex numbers/vectors:

$$A = A_0 e^{i\varphi}$$

$$A = |A| \exp(i\varphi) = |A| (\cos \varphi + i \sin \varphi) = A_r + iA_i$$



$$\vec{F}_{hkl} = \sum_{j=1}^n f_j \exp(i\varphi_j) \quad \text{Structure Factor}$$

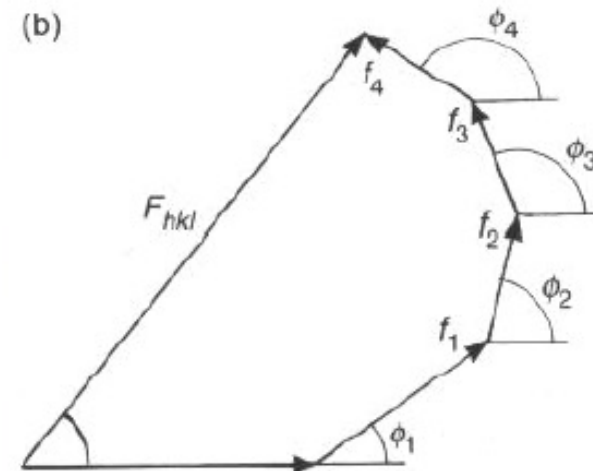
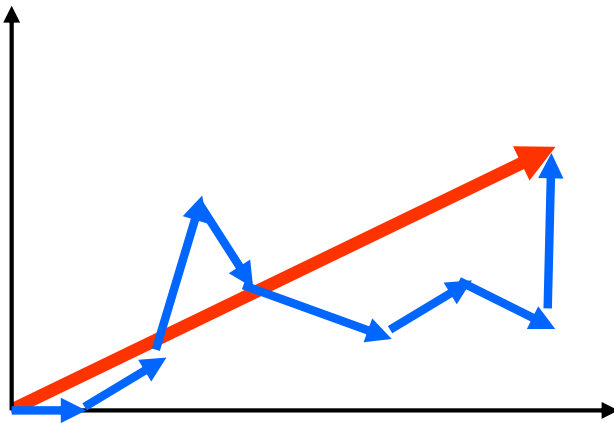
$$\varphi_j = 2\pi(hx_j + ky_j + lz_j)$$

Gives the total amplitude of the wave scattered by all the atoms in the UC.

Diffraction from a multi-atom unit cell

$$\vec{F}_{hkl} = \sum_{j=1}^n f_j \exp(i\varphi_j) = |F_{hkl}| \exp(i\varphi)$$

Structure Factor



The total intensity of the diffraction from hkl planes is proportional to the **square of the module of the structure factor** relative to that plane

Structure Factor for many atoms inside unit cell

$$F(hkl) = F(\vec{S}) = \sum_{j=1}^N f_j e^{2\pi i \vec{S} \cdot \vec{r}_j}$$
$$\vec{S} \cdot \vec{r}_j = hx_j + ky_j + lz_j$$

- $F(hkl)$ is electric field amplitude for a certain scattering direction (reflection spot h, k, l on film)
- $x, y, z \dots$ *fractional coordinates* describing the position of each atom *inside* unit cell.
- f_j is the amplitude of the electric field coming from each atom j inside unit cell (amplitude is proportional to electron density).
- The sum is over all N atoms inside a unit cell.
- $S = (ha^*, kb^*, lc^*)$ is the scattering vector for each diffraction spot

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 $\rho(\mathbf{r})$** (the more density the larger the amplitude).

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

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

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

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

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.

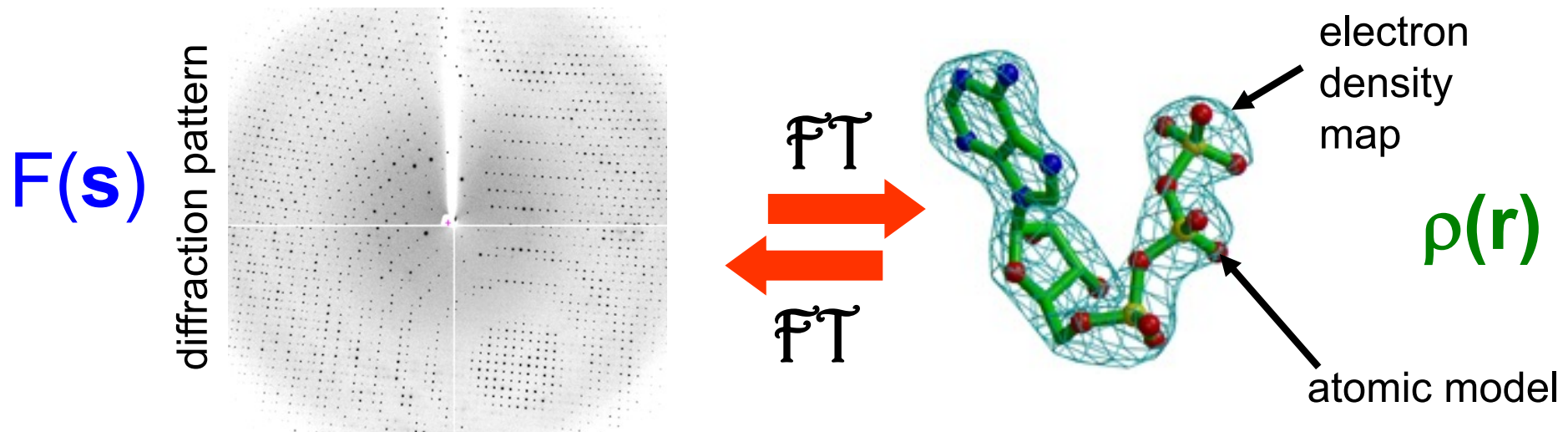
How bad is that? Pretty bad, because we can calculate the “magnitude” of the electron density, but we don't know where it is in the unit cell.

→ What can we do??

Diffraction and Fourier synthesis

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

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



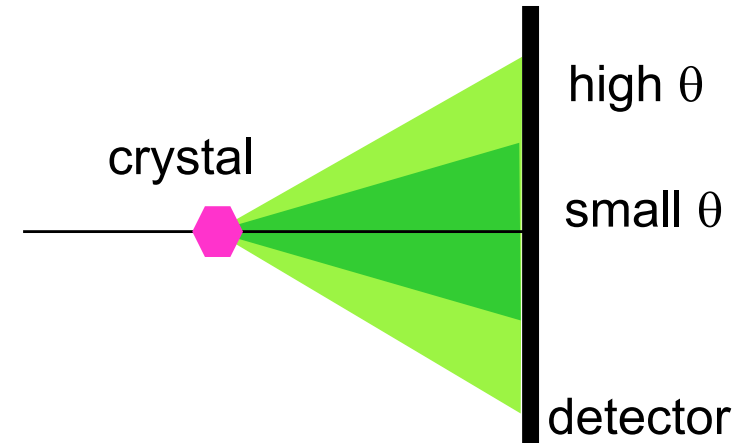
We can always go back and forward using Fourier transforms:

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

The electron density equation: problem # 1: resolution limits

Theoretical limit

There are theoretical limits that depends on the wavelength of the X-ray radiation. Typically one uses a wavelength of roughly 1 Å, which would allow for resolution of 0.5 Å.



Practical limit

In practice for macromolecules the resolution limit is usually set by the intrinsic degree of order of the crystal typically one sees diffraction to 2.0-3.5 Å – this is even more true of membrane protein crystals which tend to be more disordered due to the less directional nature of hydrophobic interactions.

Why is resolution important?

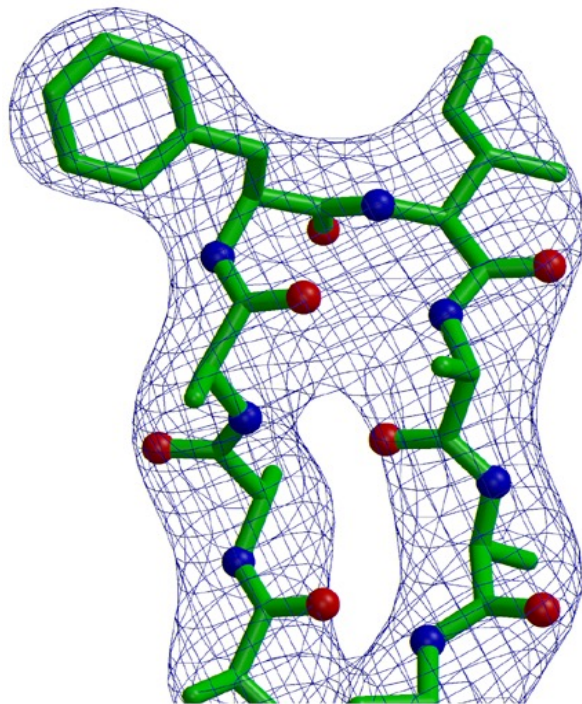
“Resolution” is related to the level of details that can be visualised: a high resolution map gives a far more detailed picture of the atomic structure:

low resolution

low s , small θ , large d

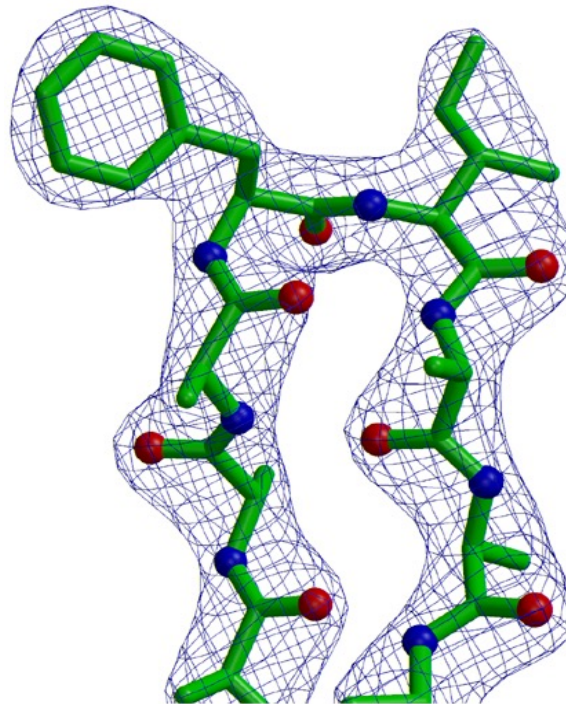
high resolution

high s , large θ , small d

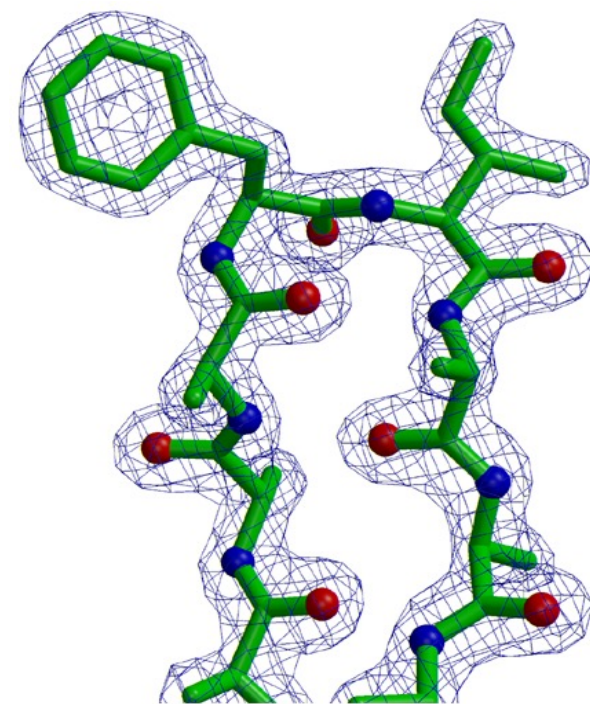


Here is a 4 Å map

things which are less than 4Å apart tend to be blurred



Here is a 2.9 Å map

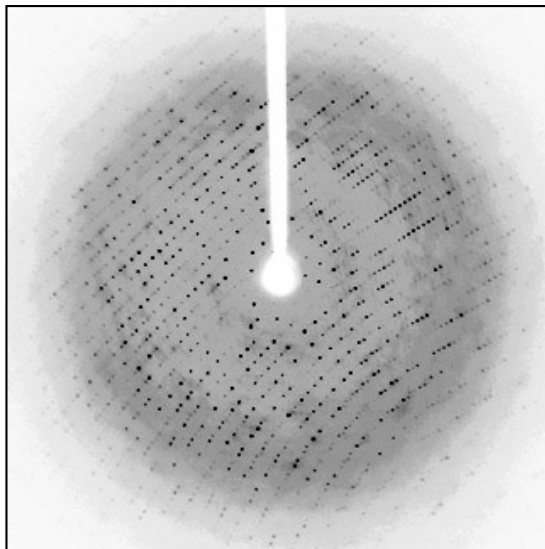
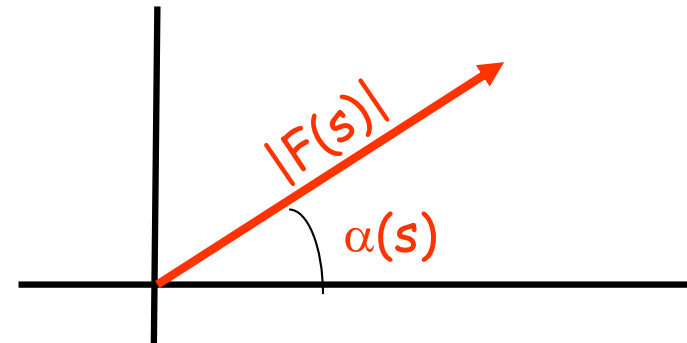


Here is a 2 Å map

almost atomic detail

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

◆ $F(\mathbf{s})$ is a complex number with modulus $|F(\mathbf{s})|$ and phase α



In a diffraction experiment we measure the intensity of each spot $I(\mathbf{s})$

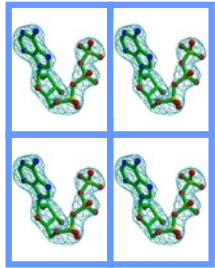
$$I(\mathbf{s}) = |F(\mathbf{s})|^2$$

We can derive the amplitude $F(\mathbf{s})$ but we have lost the information about the relative phase

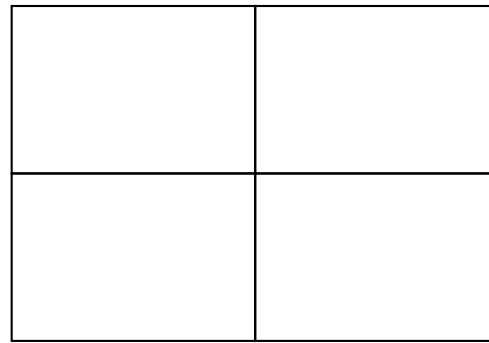
➡ ***the phase problem!!***

Universal problem in crystallography – also for small molecules.

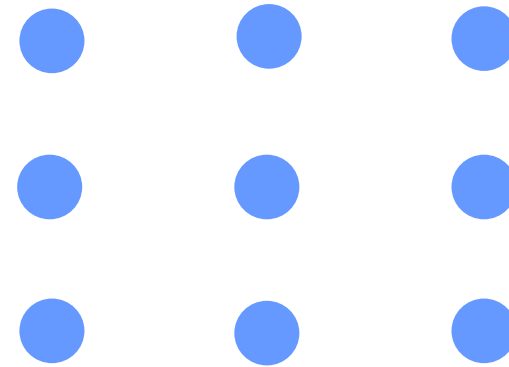
Small molecule crystal



small unit cell

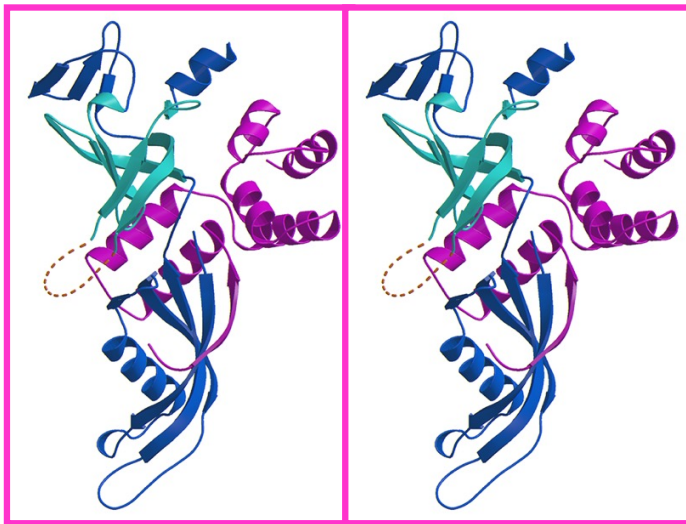


large “reciprocal lattice”
(coarse fringe function)

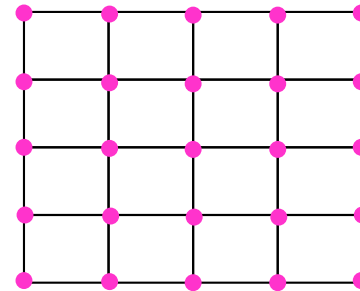


a few, very strong reflections;
diffract to high resolution

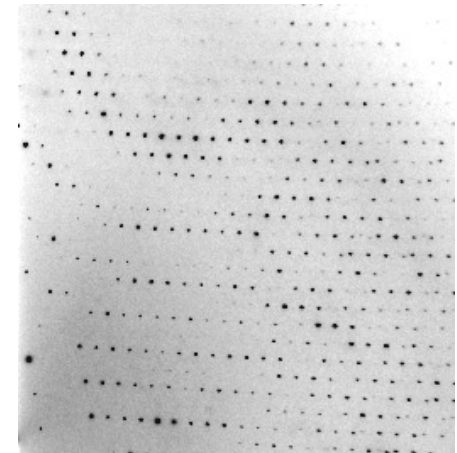
Protein crystal



large unit cell



small “reciprocal lattice”
(fine fringe function)



many, very weak reflections;
diffract to low resolution

X-ray fiber diffraction

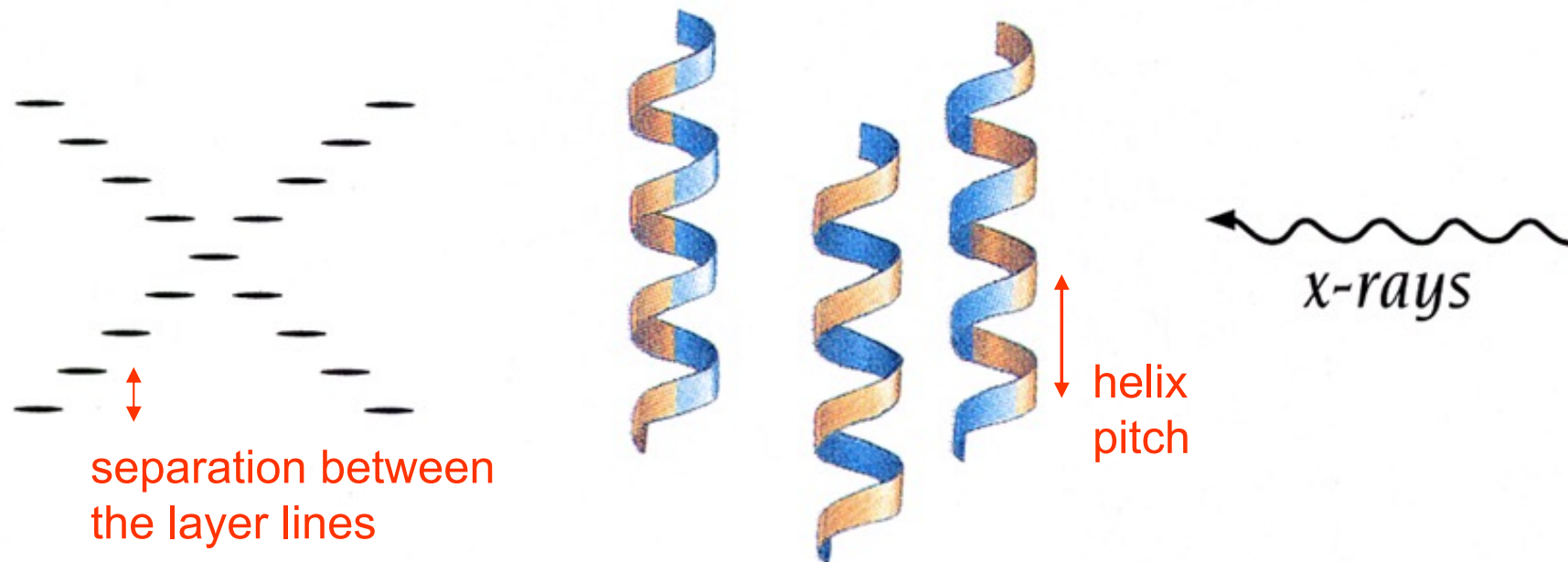
- If there is inherent symmetry in a molecule (i.e. long helical biopolymers) X-ray diffraction patterns can be obtained from non-crystalline samples.
- Like one-dimensional crystal along helix axis.
- We can treat an exact repeat of helix as a crystalline unit cell.
- Can get information about helical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



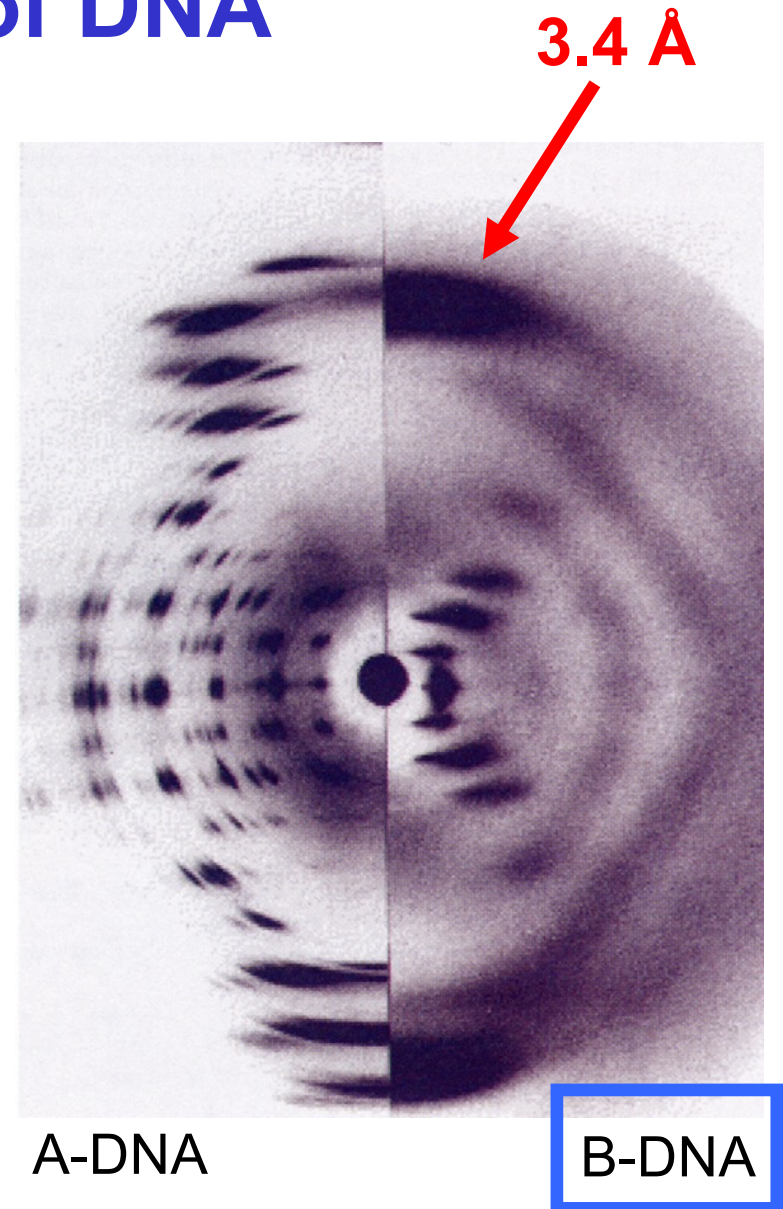
The separation between the layer lines is determined by the helix pitch: as the helix pitch increases the layer lines move closer together.

X-ray fiber diffraction: the structure of DNA

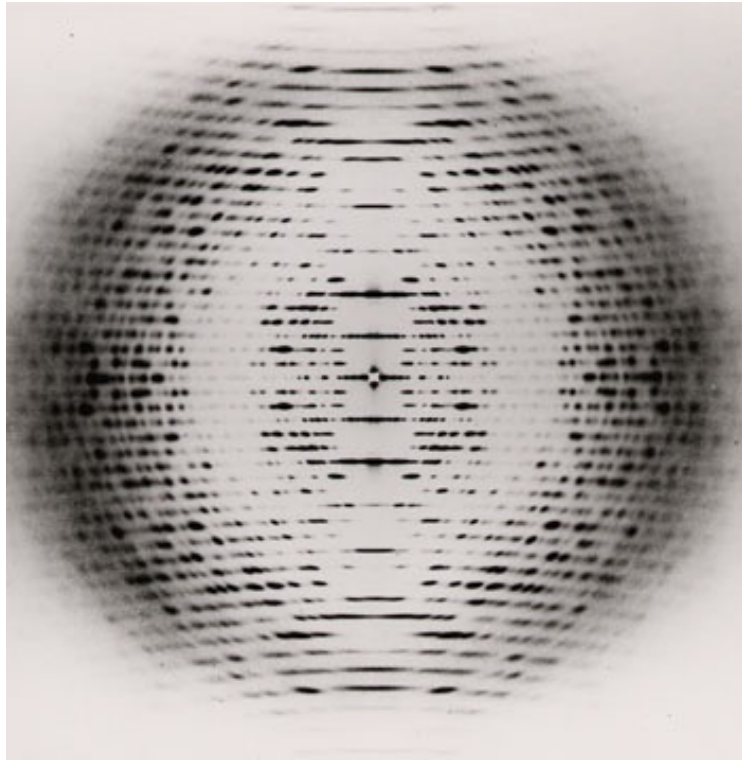
Here is the diffraction pattern from two forms of DNA: the pattern from the A-DNA indicates a higher degree of regularity in the way the fibers pack together (more 'spots').

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 to 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



***X-ray fiber diffraction:
the structure of tobacco mosaic virus***

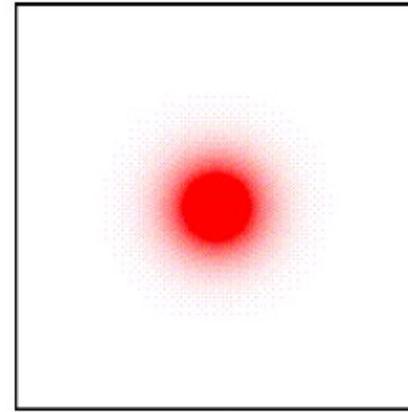
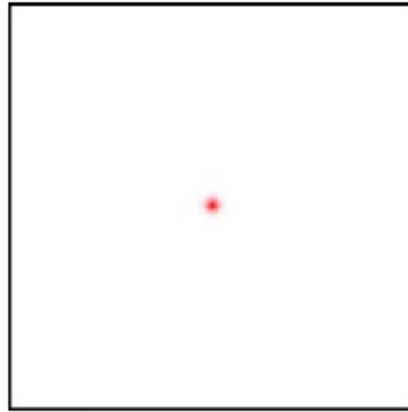


Fiber diffraction pattern

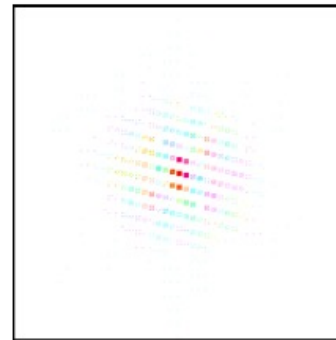
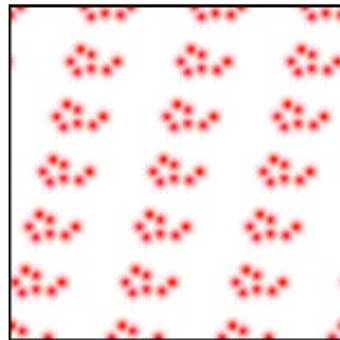


Model of TMV virion

- An atom, and its Fourier Transform:

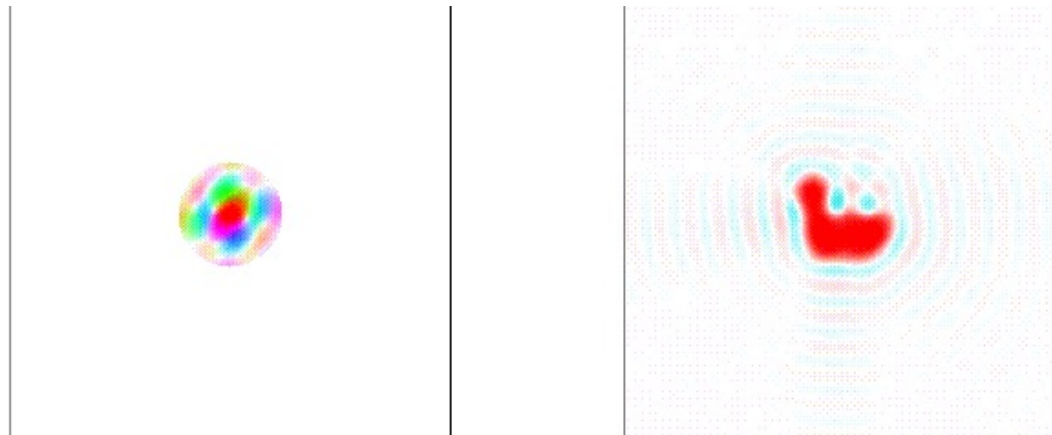
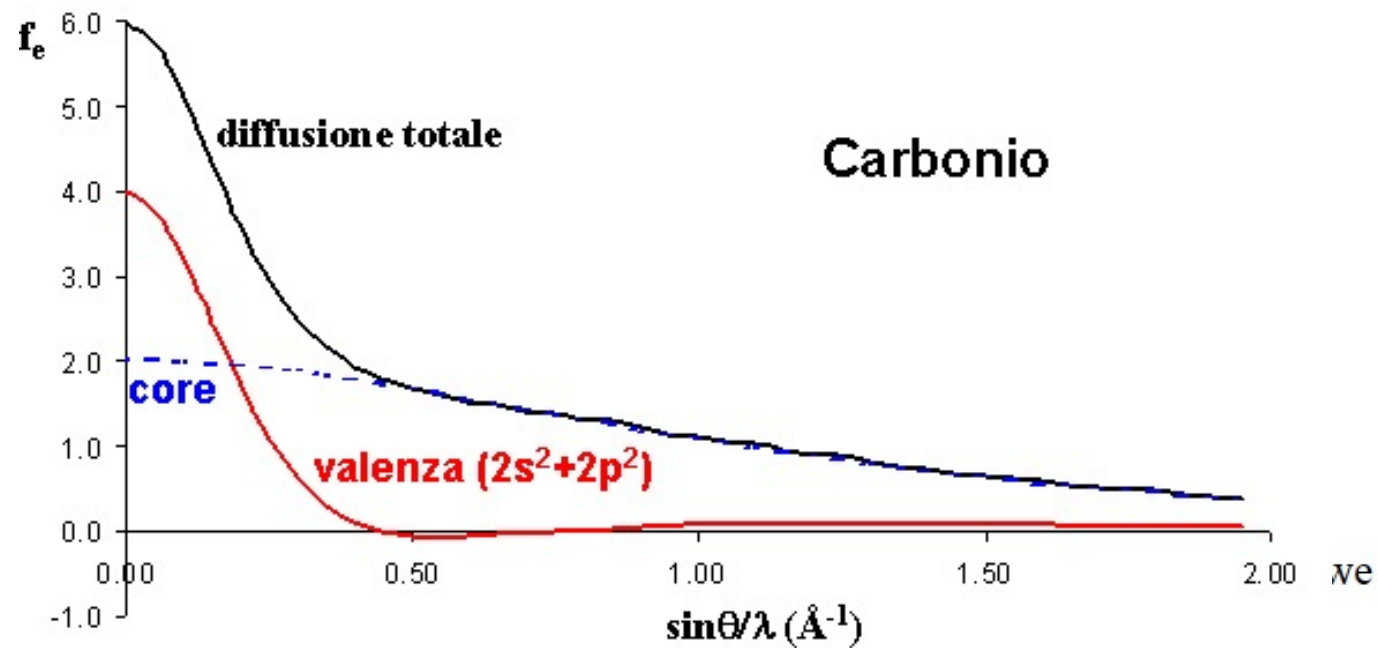


- Note the both functions have circular symmetry. The atom is a sharp feature, whereas its transform is a broad smooth function. This illustrates the reciprocal relationship between a function and its Fourier transform.
- A crystal, and its Fourier Transform:



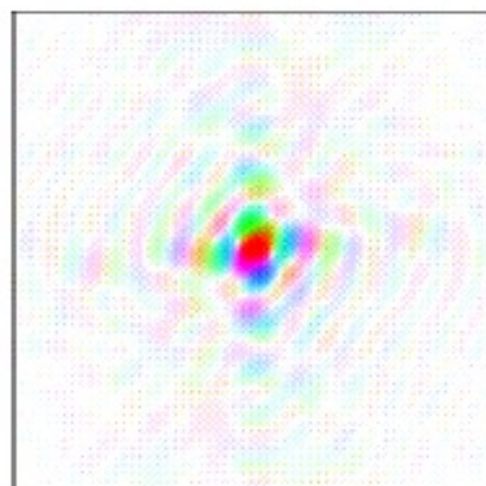
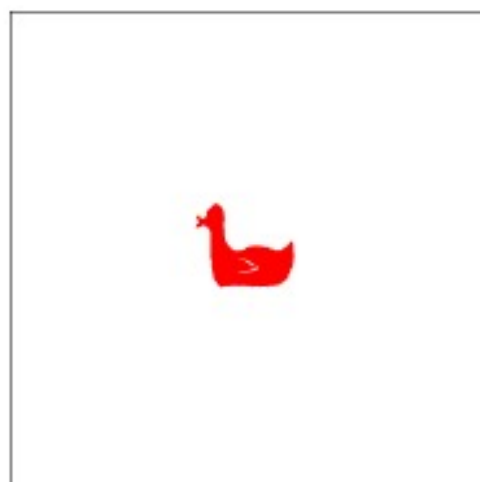
- Finally, we build up a crystal by convoluting the *molecule* with the *grid*. The result is a crystal structure. The Fourier transform of the crystal is thus the product of the *molecular transform* and the *reciprocal lattice*. This is the *diffraction pattern*.

- A duck and its Fourier Transform

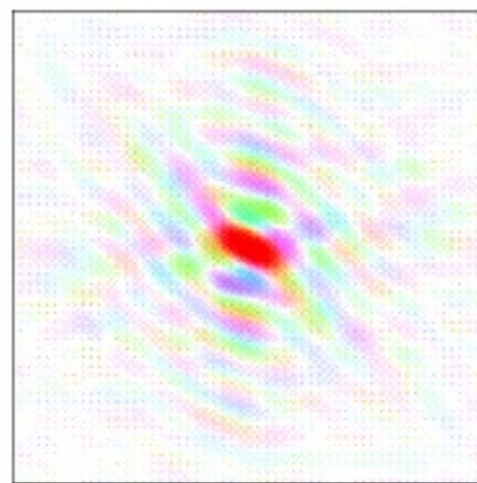


- **Crystallographic Interpretation:**
- There is considerable loss of detail. At low resolution, your atomic model may reflect more what you expect to see than what is actually there.

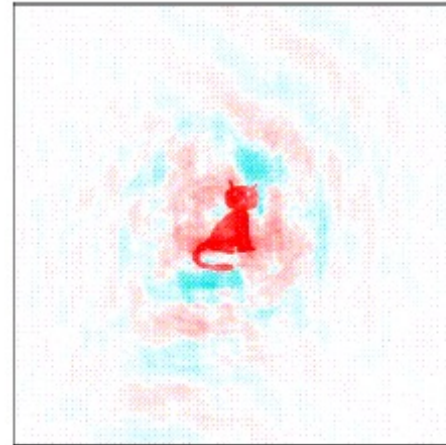
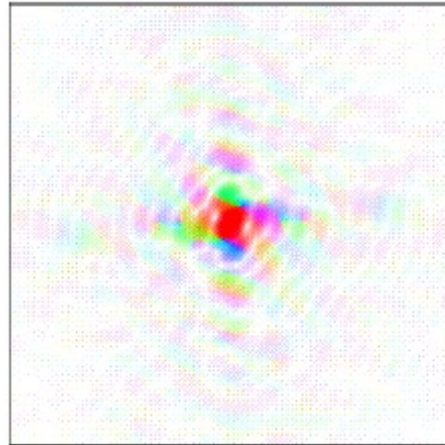
- Here is our old friend; the Fourier Duck, and his Fourier transform:



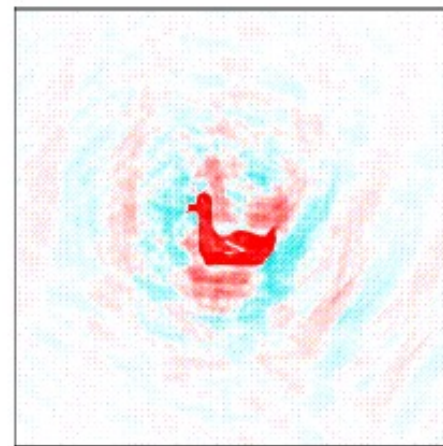
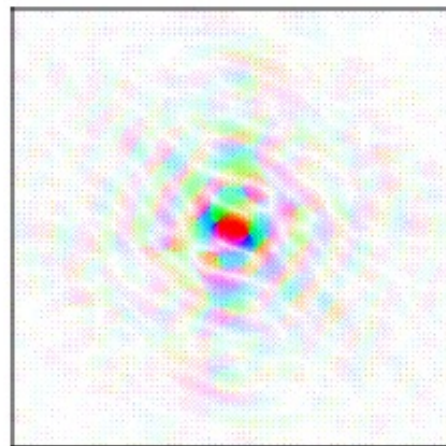
- And here is a new friend; the Fourier Cat and *his* Fourier transform:



- Now we will mix them up. Let us combine the the magnitudes from the Duck transform with the phases from the Cat transform. (You can see the brightness from the duck and the colours from the cat). If we then transform the mixture, we get the following:



- We can do the same thing the other way round. Using the magnitudes from the Cat transform and the phases from the Duck transform, we get:



How to solve macromolecular structures:

MIR (multiple isomorphous replacement)

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

MR (molecular replacement)

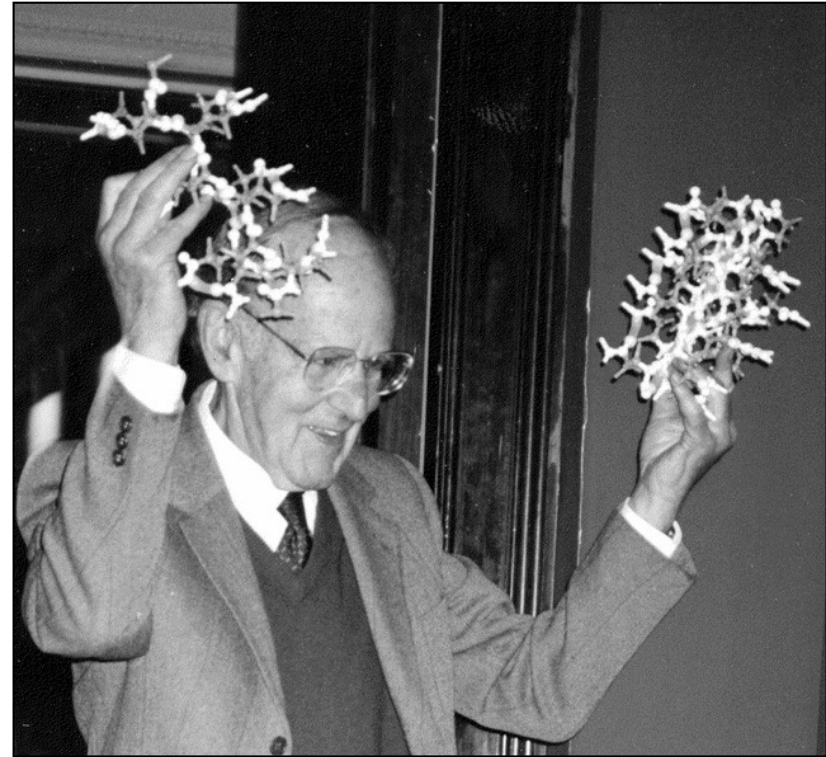
Older method (Cambridge, 70'-80') – relies on the expected similarity between the protein and another whose structure is known. Cannot solve de novo structures. Requires high homology (30% sequence identity?)

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)



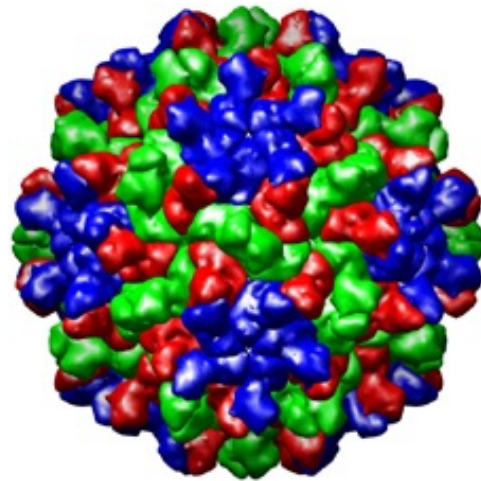
Francis Crick
(1916-2004)



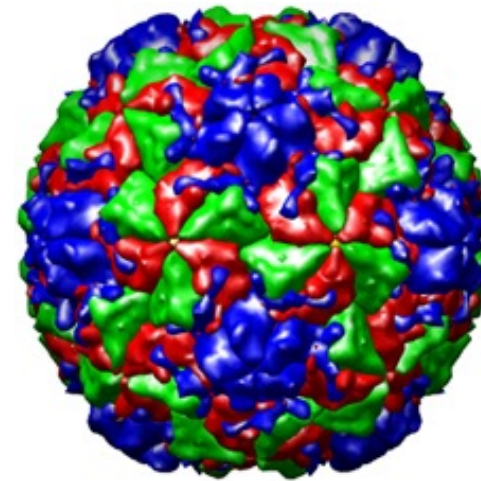
Error treatment is very complex – but it was an essential step in solving protein structures.

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:

Rossmann, M. G. and Blow, D. M. (1962). *Acta Cryst.* 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** $\alpha_{\text{MIR}}/\alpha_{\text{MR}}/\alpha_{\text{MAD}}$.

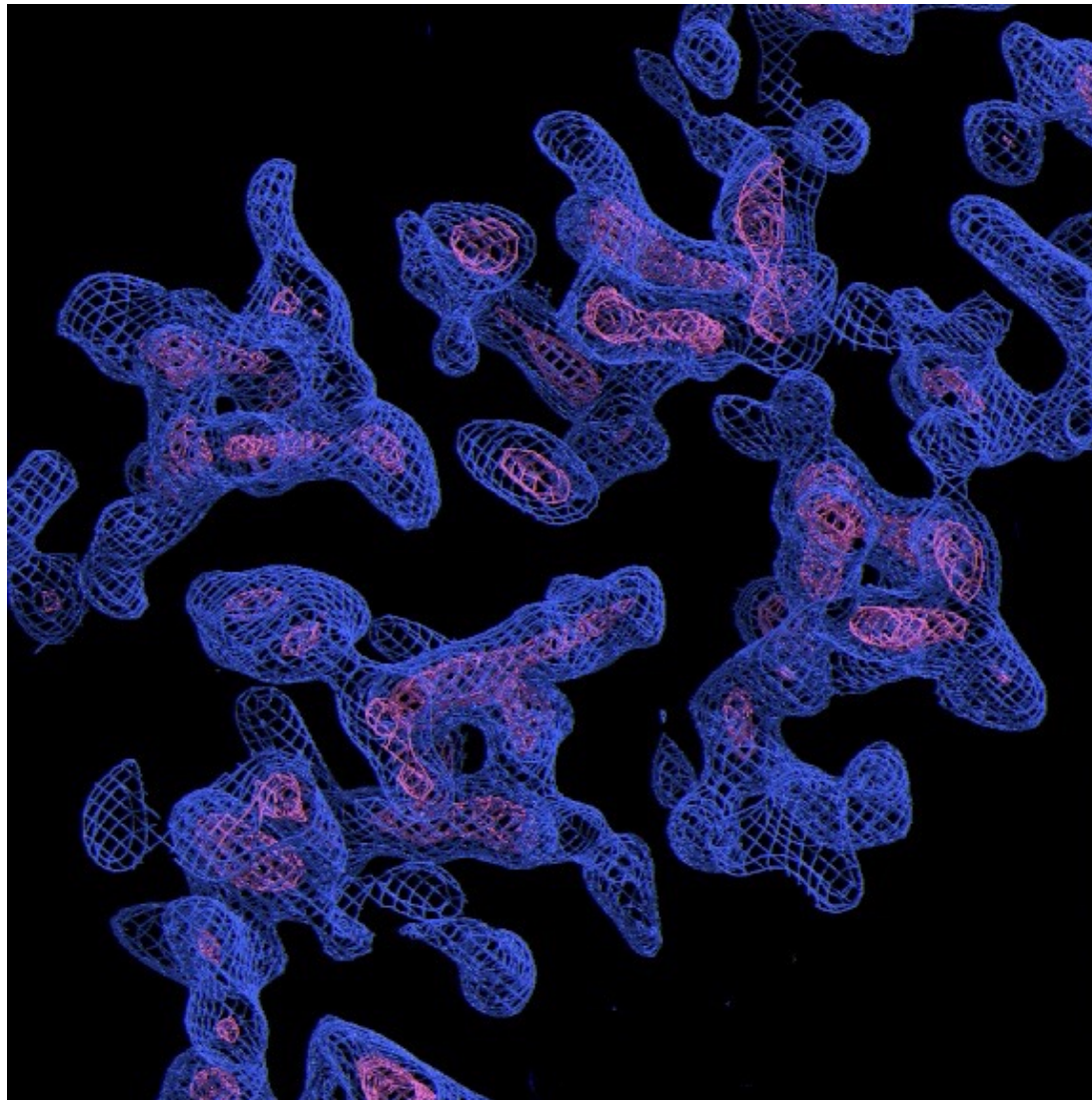
$$\rho(xyz) = \sum_{hkl} W(hkl) * |F_P(hkl)| \exp[i \alpha(hkl)] \exp[-2\pi i(hx + ky + lz)]$$

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:



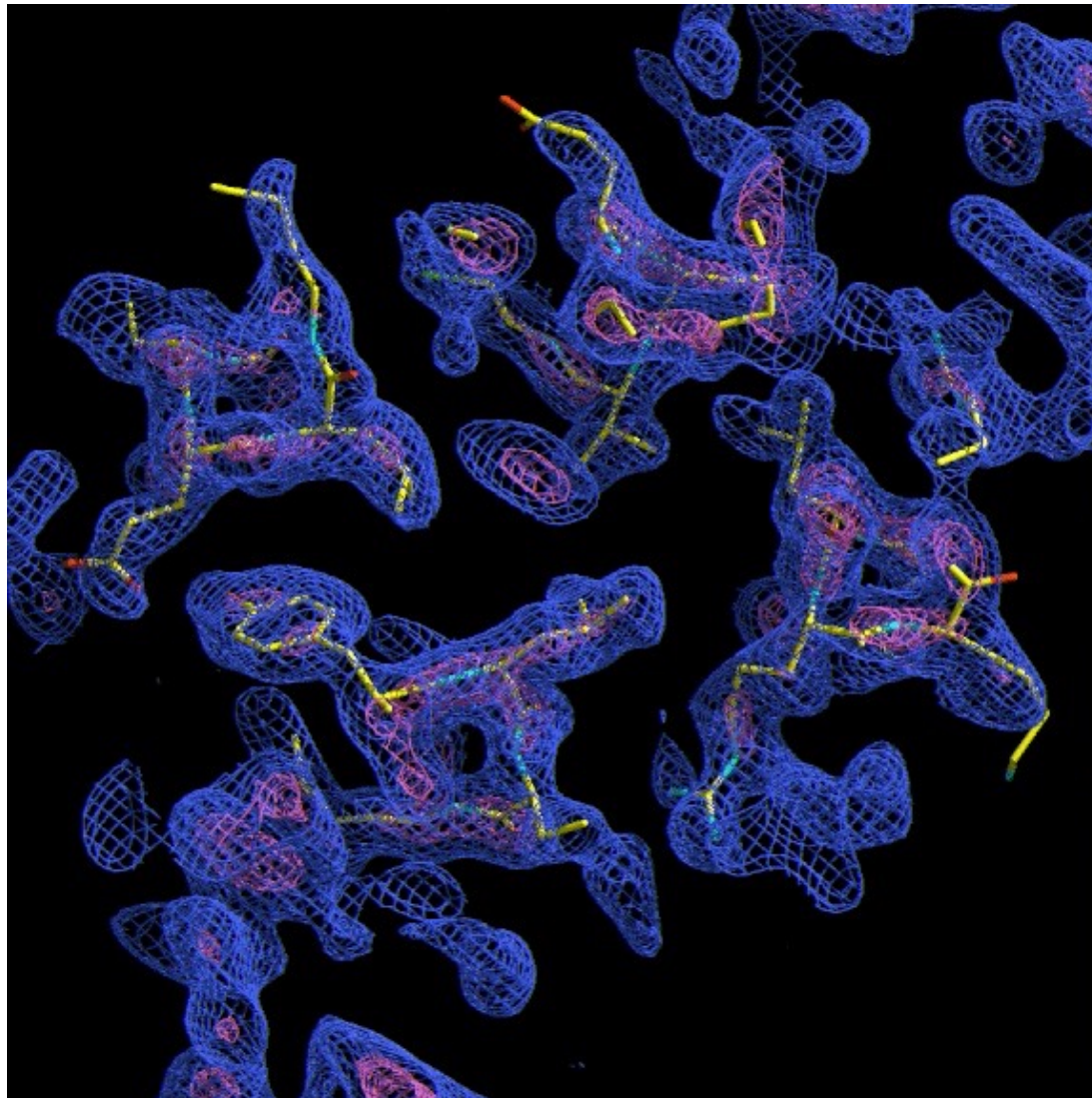
Electron density map
displayed at two
contour levels:

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);

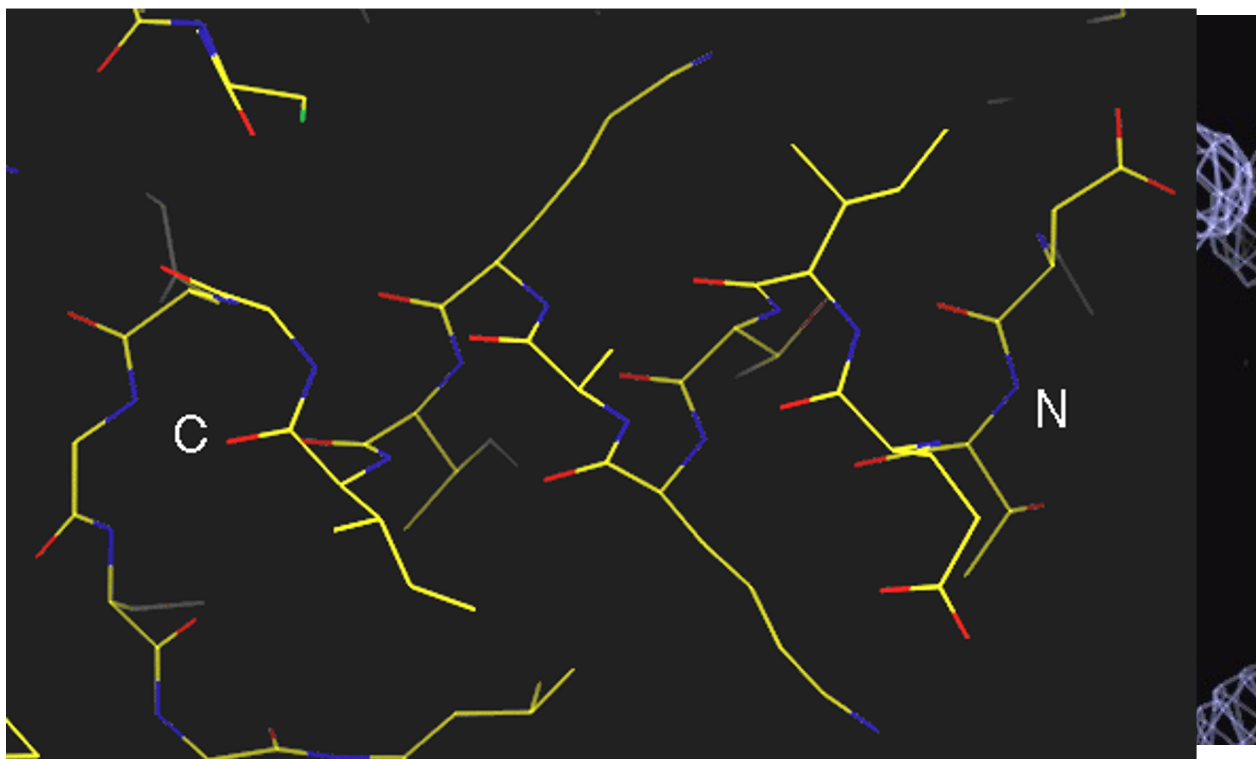
magenta = 3 r.m.s (3σ)

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 **1σ level**)

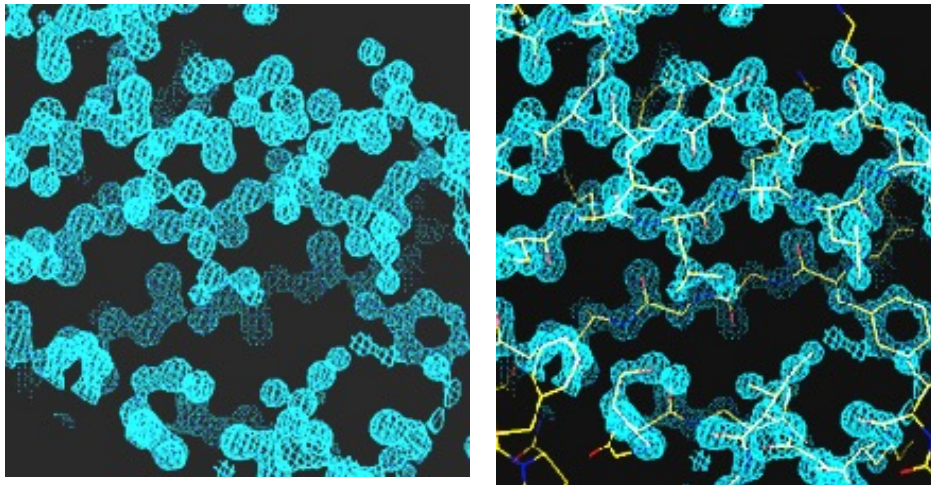


Here is a rather good bit of density (much better than the density you are likely to obtain from the initial phases) and still it is not easy for the “untrained eye” to see what is going on...

Maps and resolution

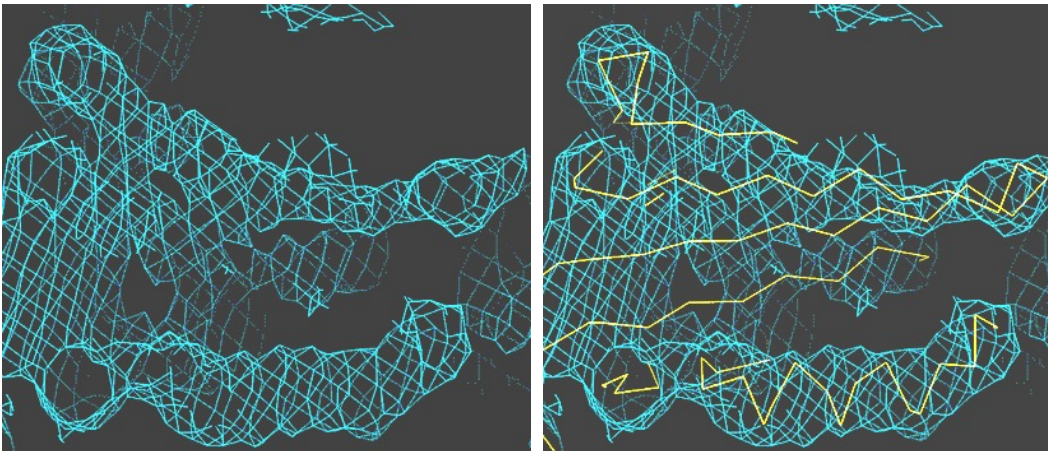
The task of model building is to interpret the electron density maps in light of chemical knowledge, basic stereochemistry, chemical sequence, etc...

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.

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

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

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}**

Practical considerations

Protein production for crystallography

- Natural sources
- Recombinant technology

Crystallisation

- The theory
- Techniques
- Crystallisation robots

X-rays generation and data collection

How to read a PX paper

How to use a PDB

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:

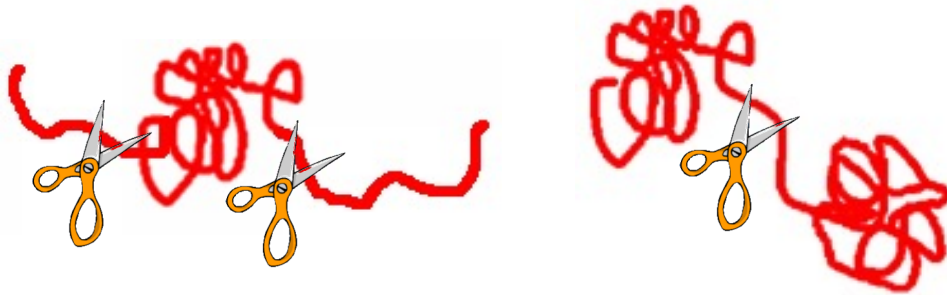
- multi-domain proteins with flexible unstructured linkers
- proteins with flexible N- and/or C-termini
- proteins that are part of large macromolecular complexes
- presence of posttranslational modifications



“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.



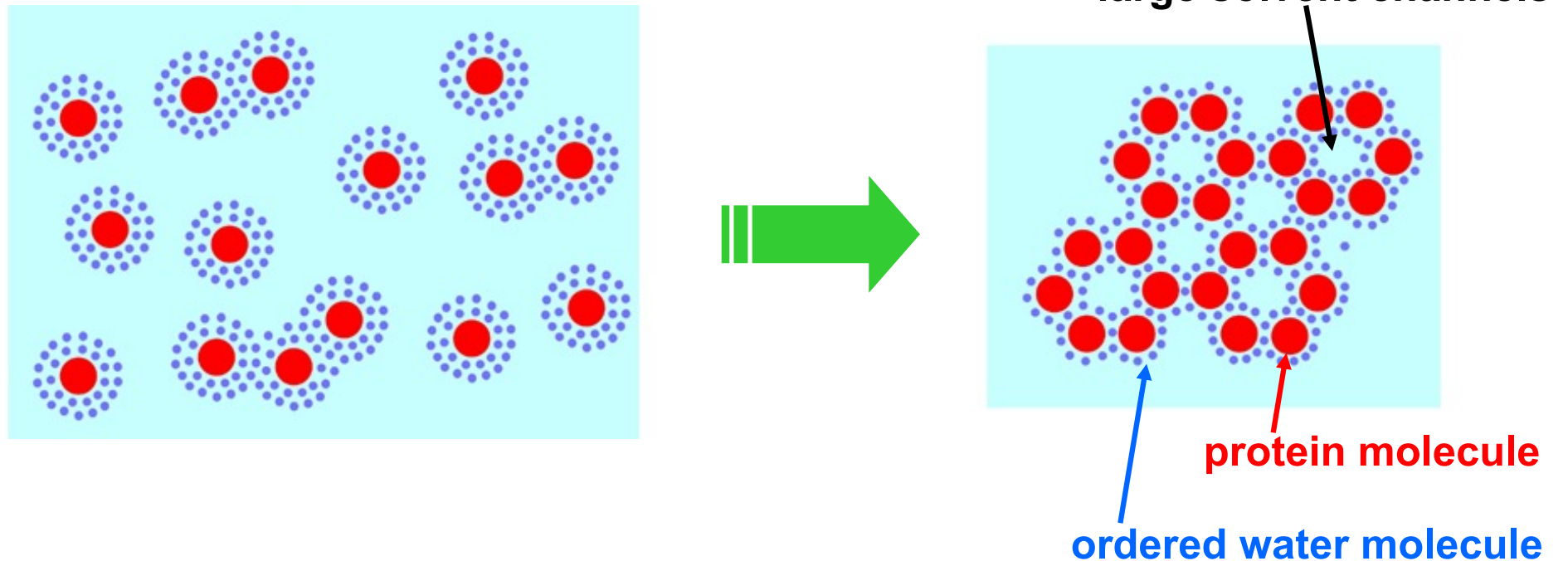
Add ligands/inhibitors/cofactors/metals to stabilise one conformation.



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:

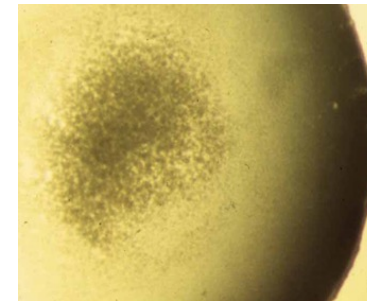
- thermodynamic factors governing the solubility
- kinetic factors governing nucleation and growth

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



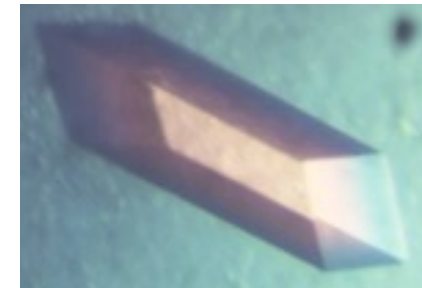
crystalline state

micro-crystals

needles

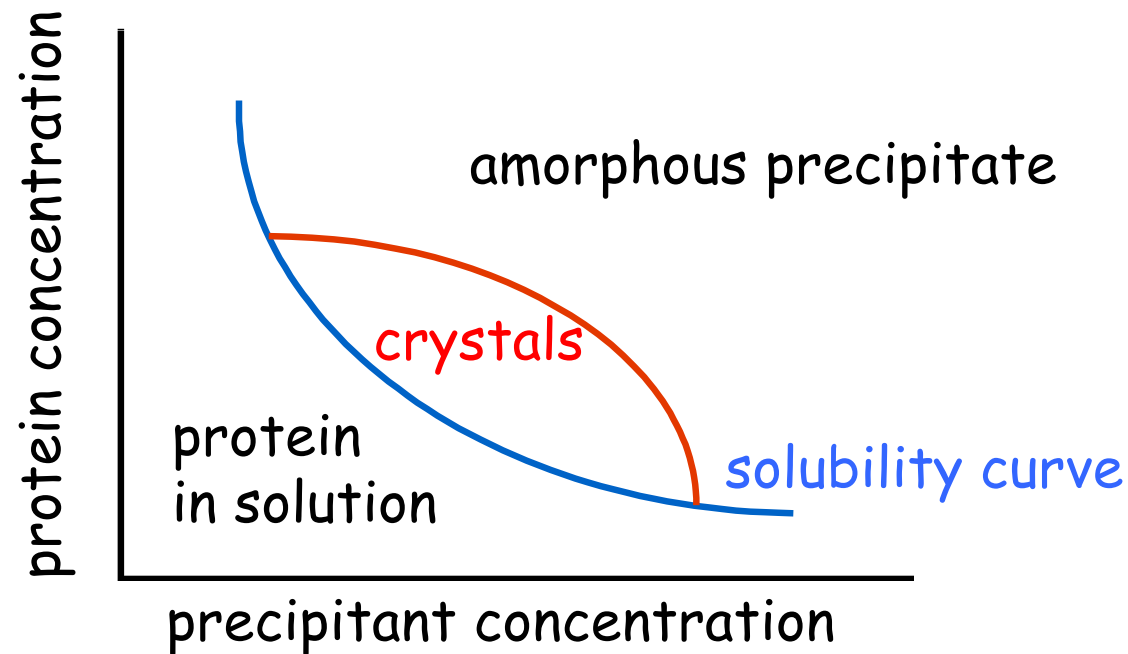
plates

large single crystals



Phase diagram

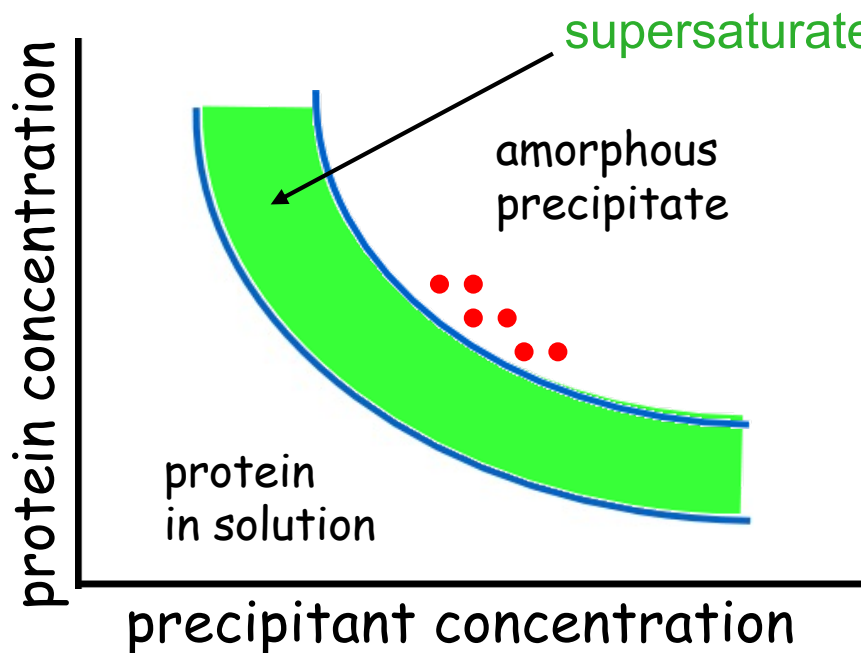
The process of crystal formation can be understood by using a phase diagram. Because several factors influence protein solubility, the phase diagram has many dimensions. A very important factor is the precipitant concentration, and a two-dimensional diagram is shown here:



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.

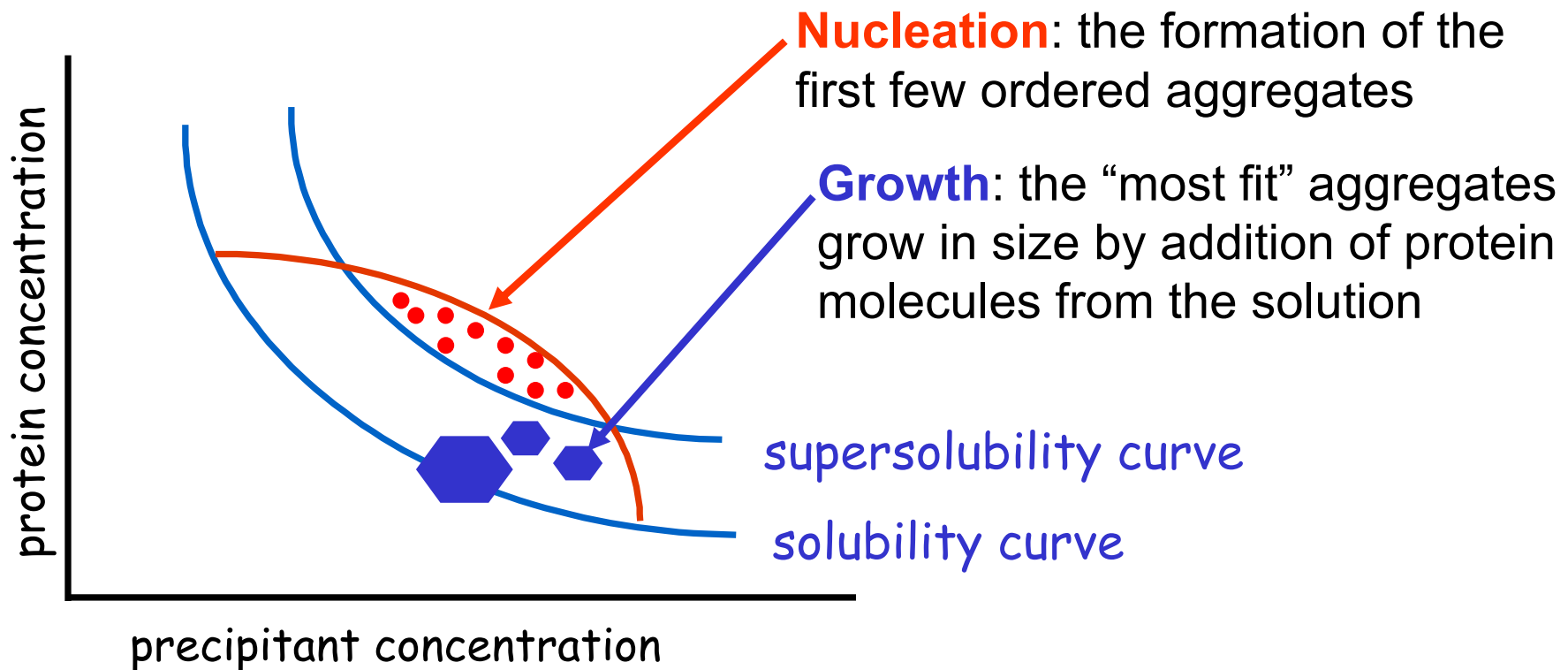
supersolubility curve

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:



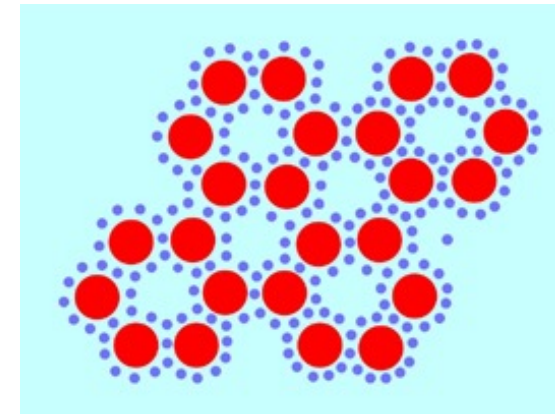
Crystallising proteins

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.

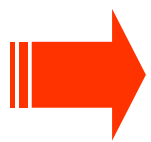
The protein solution has to be homogeneous not only in terms of composition, but also in terms of structure and conformation.



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.

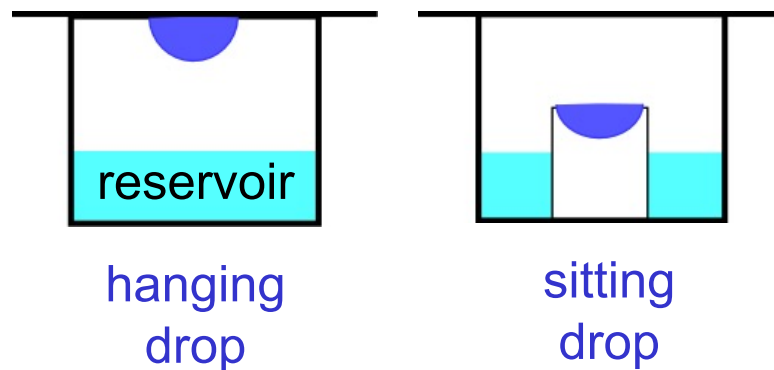


One needs large amounts of very pure protein that can be concentrated to several mg/ml without aggregation.

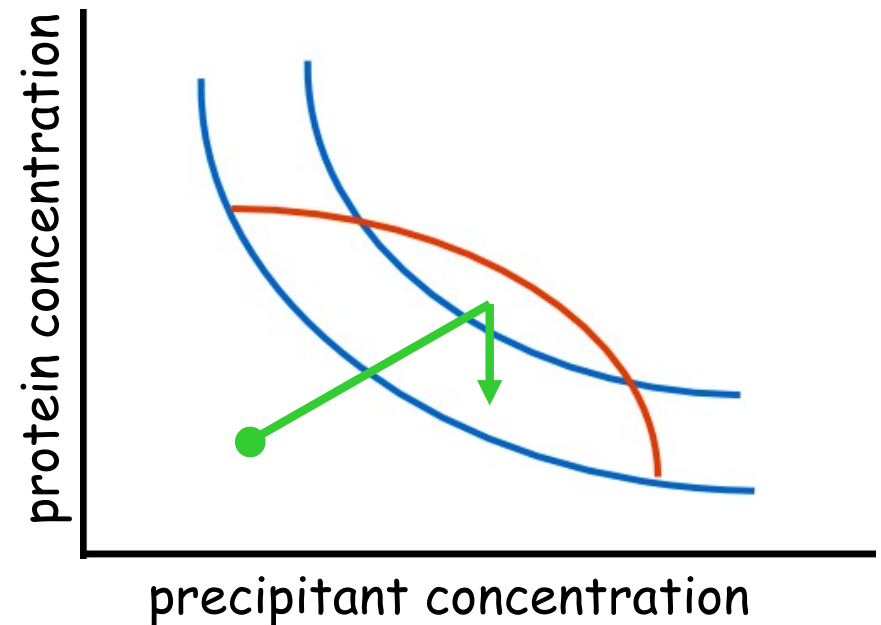
Robotic screening allows testing of ~1000 conditions/mg of protein.

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.



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

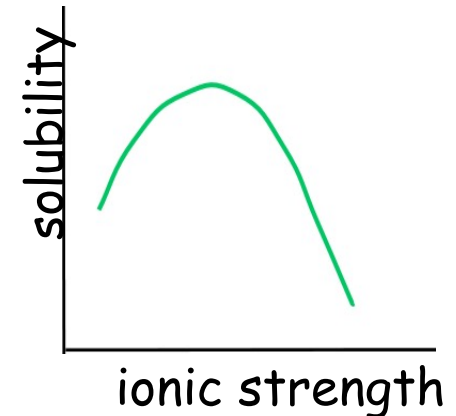


Vapour diffusion is best suited for robotic nanolitre crystallisation.

Precipitants

SALTS

Solubility changes with the ionic strength of the solution, causing salting in/salting out effects. Common ionic precipitants include $(\text{NH}_4)_2\text{SO}_4$, 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:

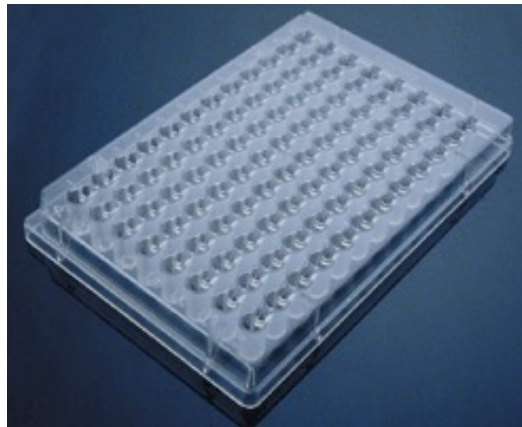
- slow and time consuming
- error prone and not always reproducible
- expensive in terms of amount of purified protein (drops: 1-2 μ l)

Robotic crystallisation:

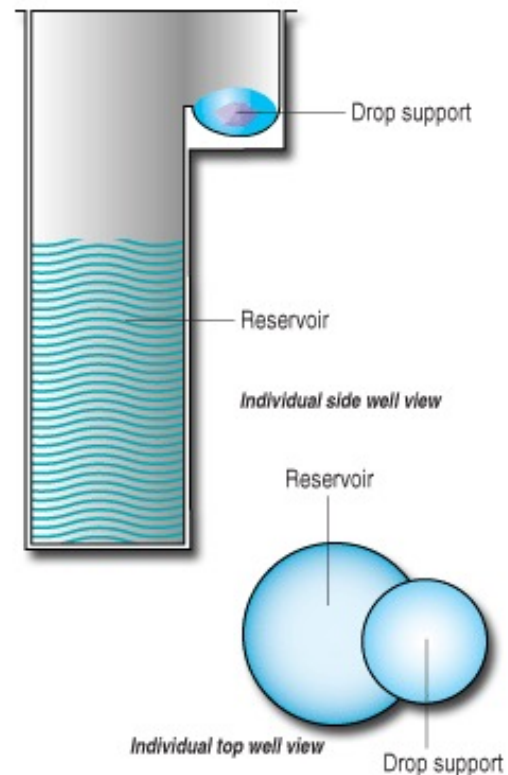
- faster and more efficient
- more accurate and reproducible
- smaller sample sizes (down to 100 nl drops) cut down on expenditure of purified protein

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

