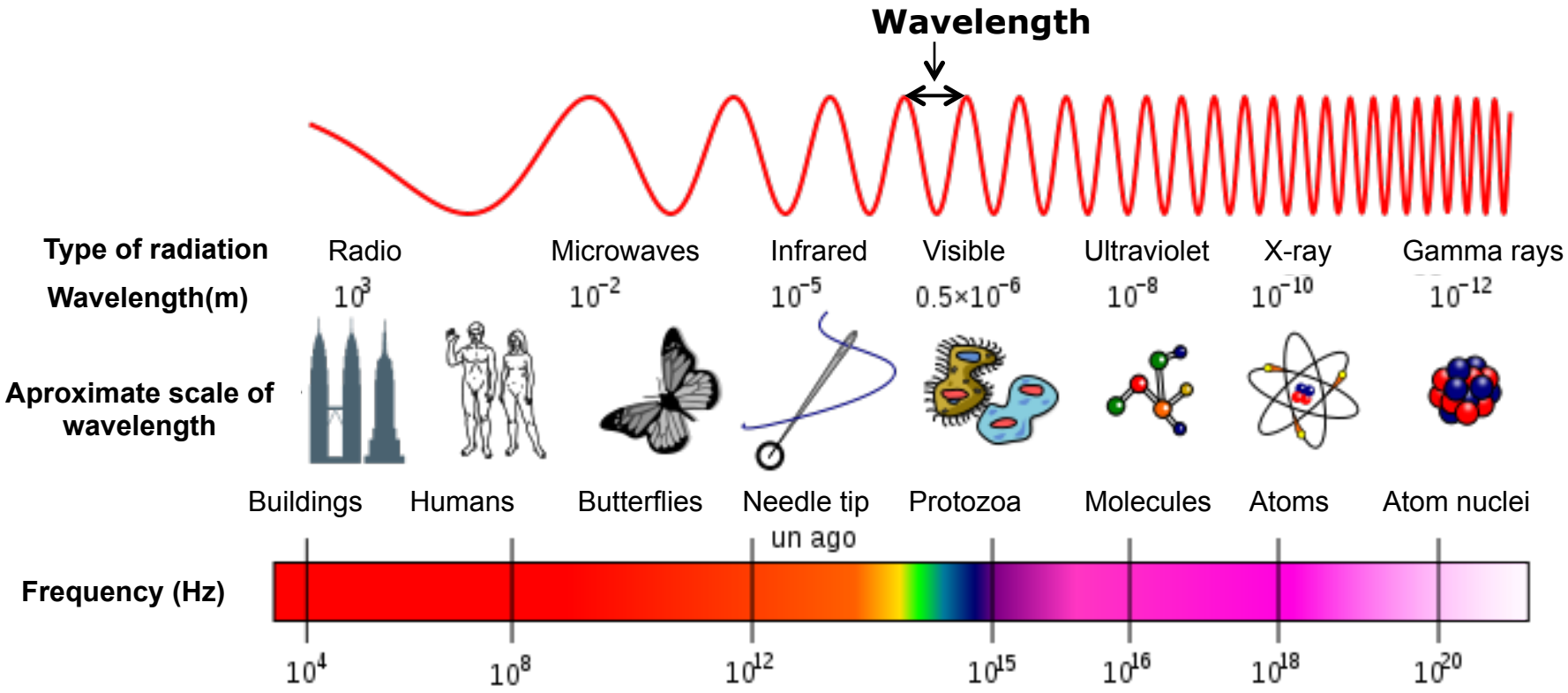
Laue's intuition

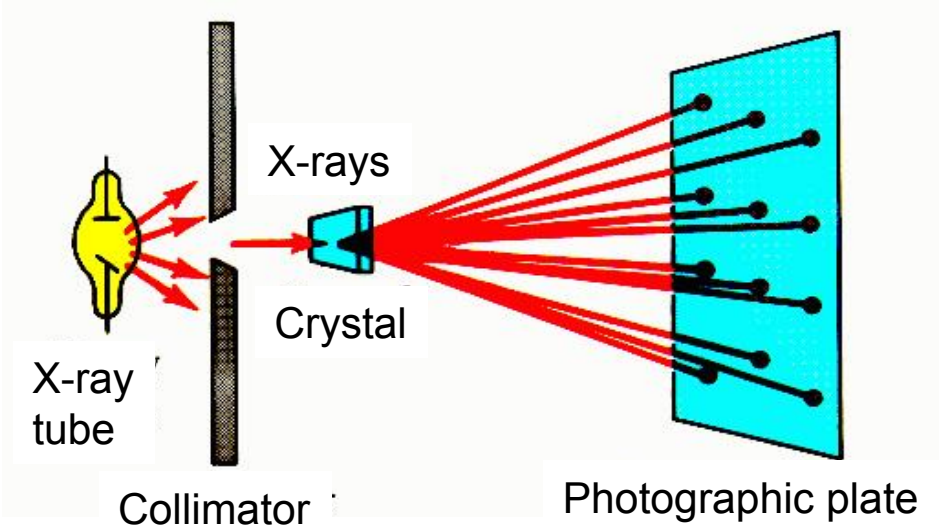
The estimates of the wave length of X-rays gave values of the same order of magnitude as the interatomic distances in the crystals.



Laue's intuition

Laue understood that X-rays could be transmitted through a crystal, producing interference phenomena.

Friedrich and Knipping executed the 1st experiment using the copper sulfate pentahydrate with random orientation of the crystal.

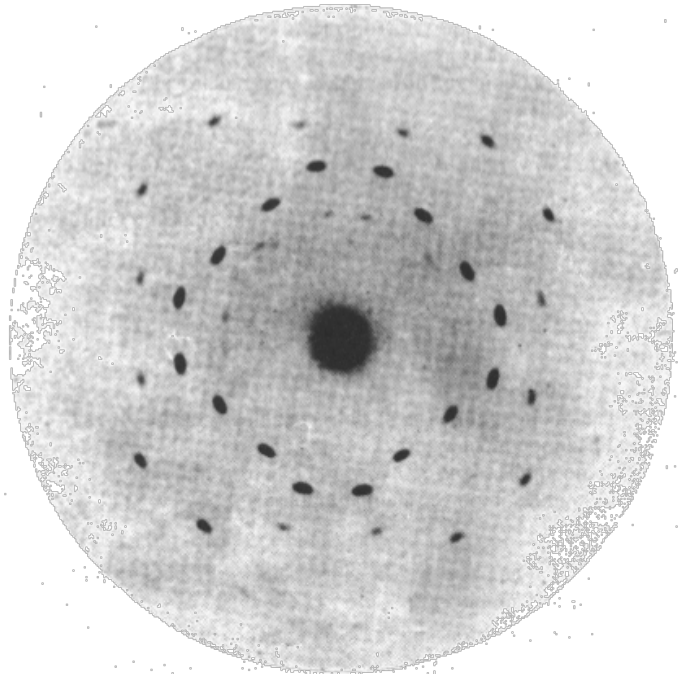


First image of X-ray diffraction

Laue's intuition

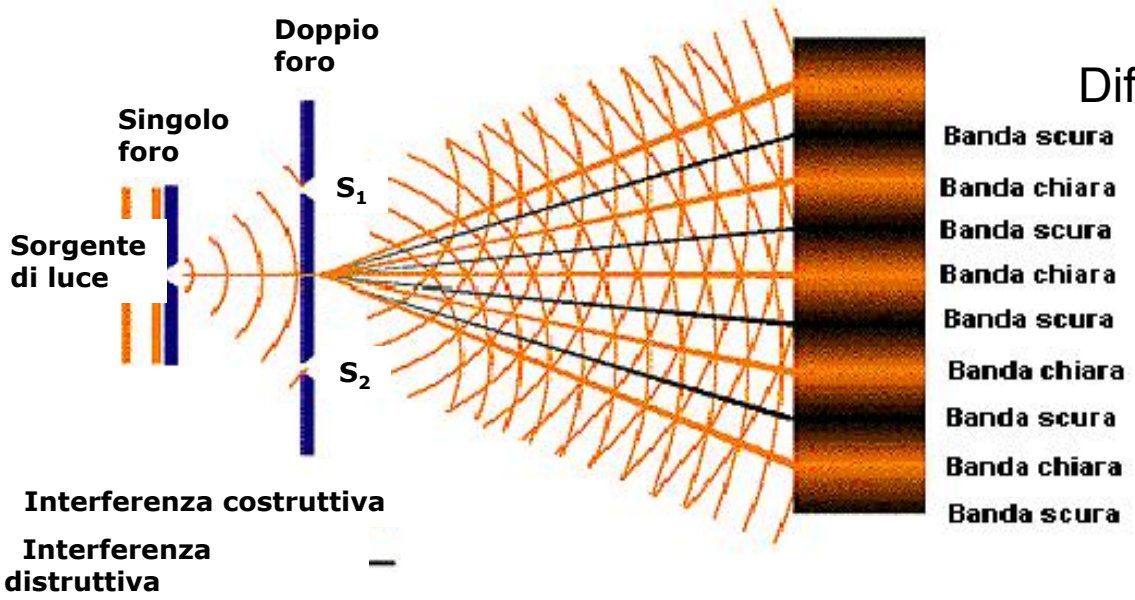
The experiment was repeated with a crystal of zinc blende (ZnS), this time well oriented with respect to the direction of the X-rays.

The black spots show a symmetrical arrangement.
Such regular arrangement can be explained by the properties
of the crystals and the phenomenon of diffraction.

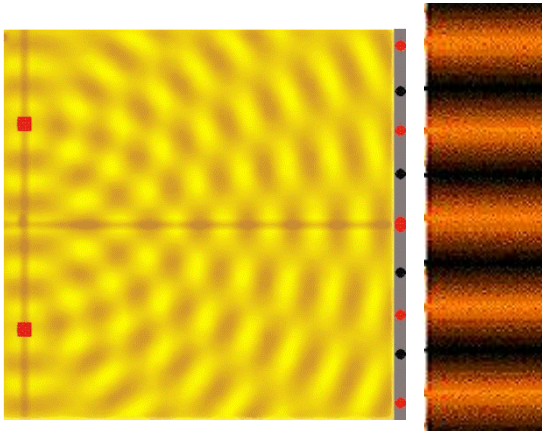


Diffraction

X-rays, as the rest of the visible light, are electromagnetic waves, therefore, subject to phenomena such as interference and diffraction.



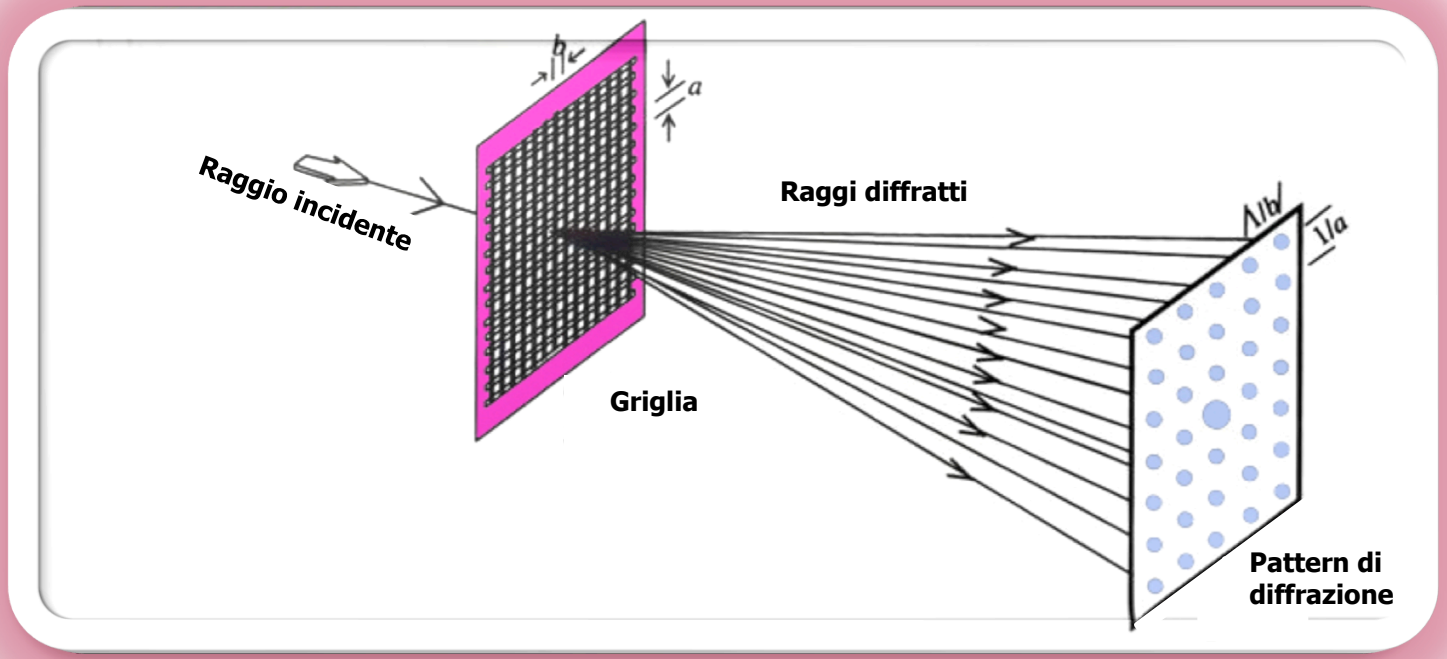
Diffraction occurs when the wave encounters an obstacle of size comparable to its wavelength!



There will be directions along which constructive interference are made of and other conditions for which the interference will be destructive.

Diffraction

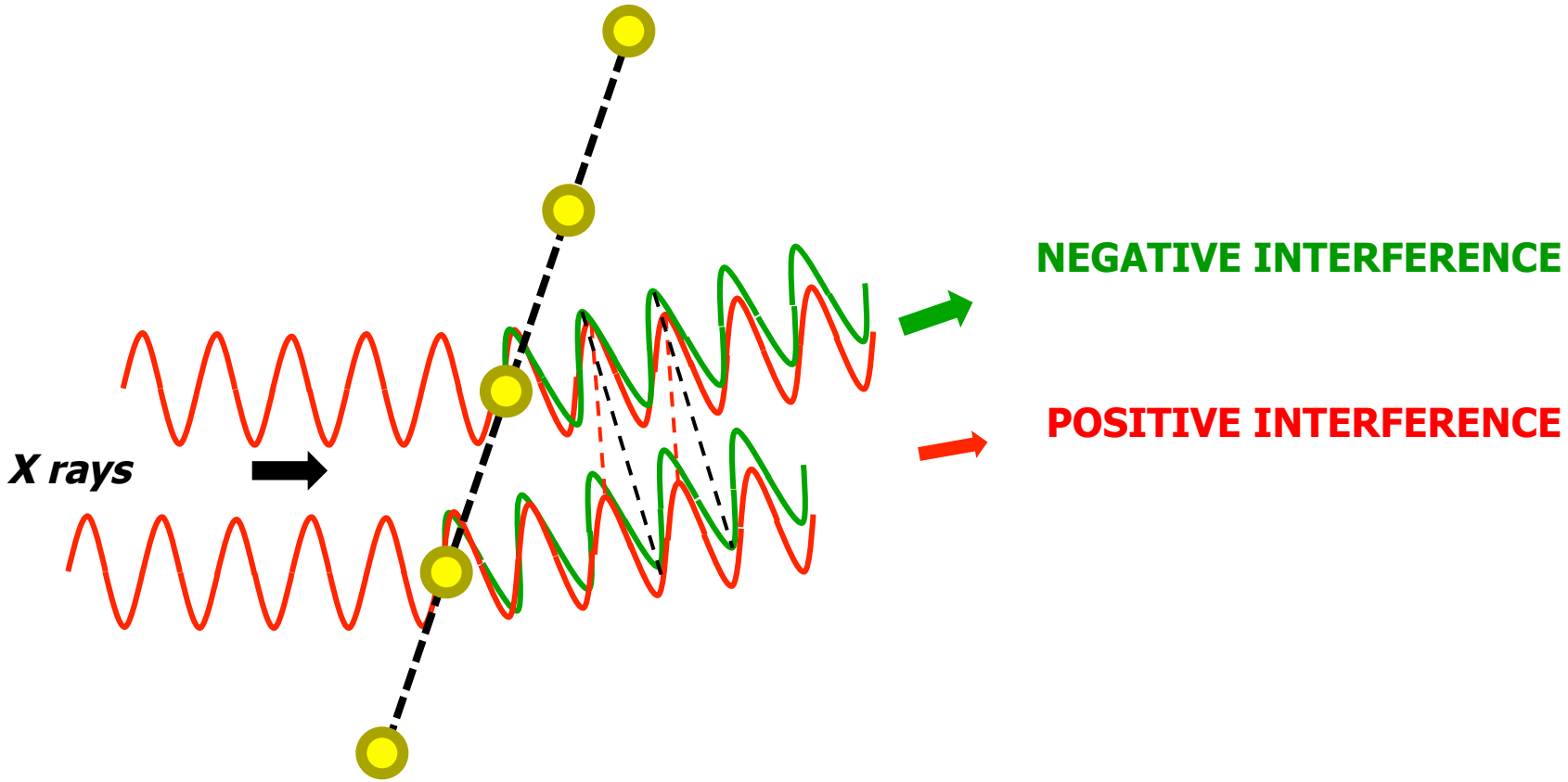
When the "obstacle" has a periodic nature, ie it consists of "objects" arranged at regular intervals according to one, two or three dimensions ....



... the directions along which diffracted beams propagate, produce on a suitable detector (eg. a photographic plate) spots regularly spaced, spaced apart in the reverse way (mutual) than the intervals between the "objects" originating .

Laue diffraction

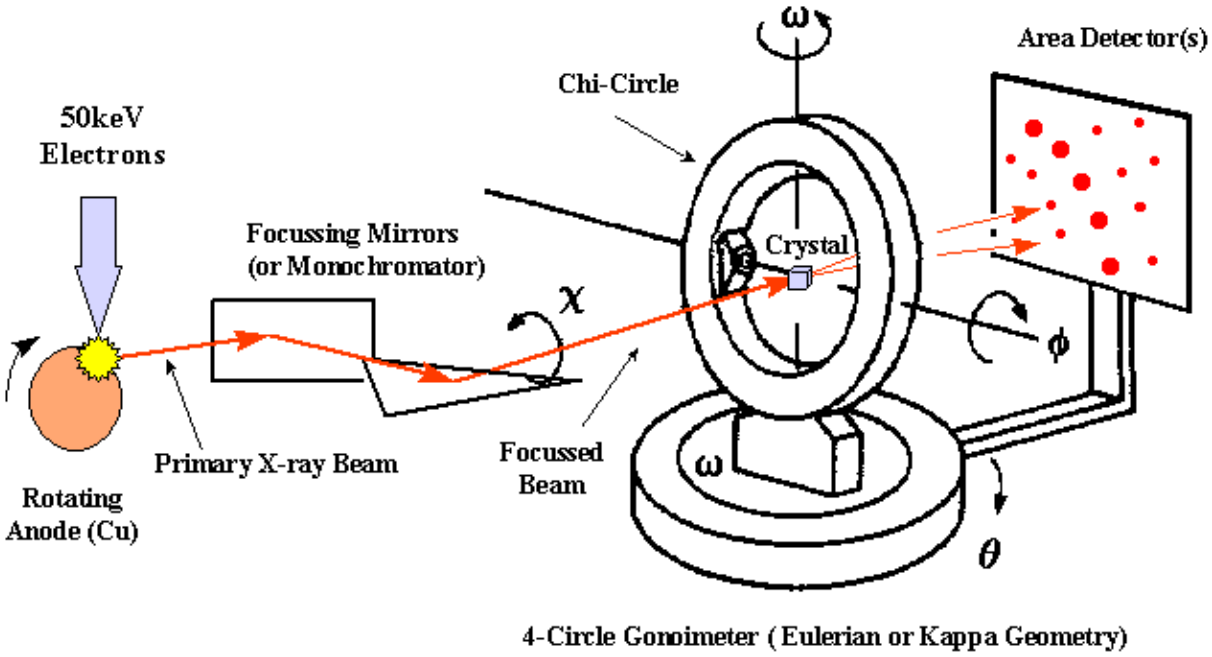
The interaction of X-rays with the atoms, (in reality with their electrons) arranged periodically in the crystals, having interatomic distances similar to their wavelength, causes the diffraction effects observed by Laue, i.e. the crystals behave as a of three-dimensional diffraction grating with respect to X-rays



The first crystal structure of a protein molecule

- 1962: Max Ferdinand Perutz and Sir John Cowdery Kendrew win the Nobel Prize in Chemistry for their studies on the structures of globlular proteins.
- The structure of myoglobin was solved by MIR.

(Max Perutz, 1914-2002)

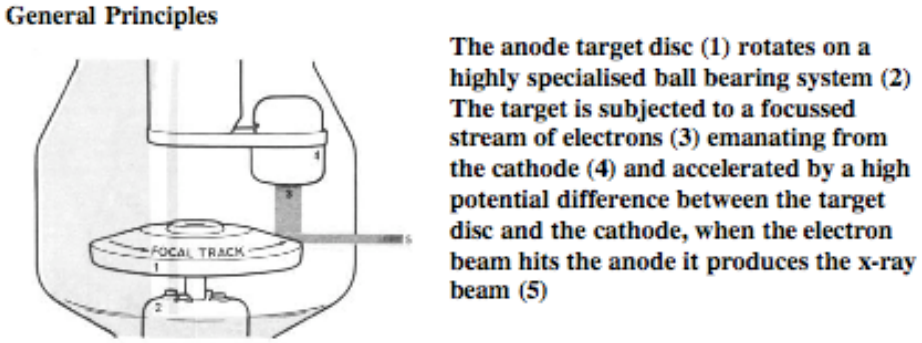


To perform a X-ray diffraction experiment, we need an x-ray source. In many cases a laboratory X-ray source such as a rotating anode generator producing a X-ray beam of a characteristic wavelength is used. Intense, tunable X-ray radiation produced by a Synchrotron provides additional advantages. The primary X-ray beam is monochromated by either crystal monochromators or focusing mirrors. After the beam passes through a helium flushed collimator it passes through the crystal mounted on a pin on a goniometer head. The head is mounted to a goniometer which allows to position the crystal in different orientations in the beam. The diffracted X-rays are recorded using image plates, Multiwire detectors or CCD cameras.

X-ray sources for protein crystallography.

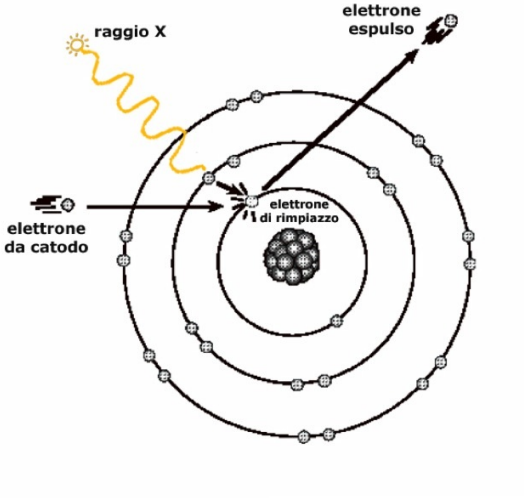
Laboratory X-ray sources.

X-rays of a suitable wavelength range for protein crystallography are generated by two commonly used devices: X-ray tubes and rotating anodes.



X-ray tubes consist of a filament that acts as a cathode. Electrons are emitted by the glowing cathode and accelerated by several tens of kVs across the vacuum towards the anode, which consists of a metal target made of a characteristic material, usually copper or chromium, for protein crystallography.

X-ray sources for protein crystallography.



The Kα component of the X-rays emission are cut out from the bremsstrahlung and other emission lines by filters, monochromators or X-ray mirrors, and the resulting monochromatic X-rays are collimated and focused onto the crystals.

$h\nu = eV$
$hc/\lambda = eV$
$\lambda = hc/eV = 12400/V$

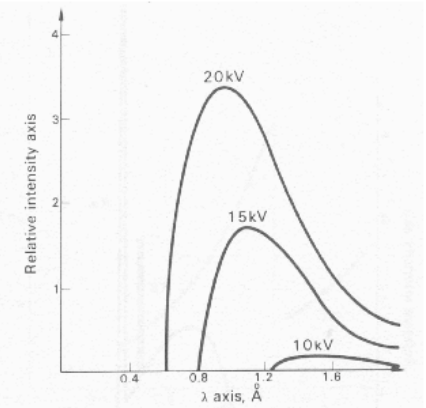


FIGURE A4.2. Variation of X-ray intensity with wavelength λ.

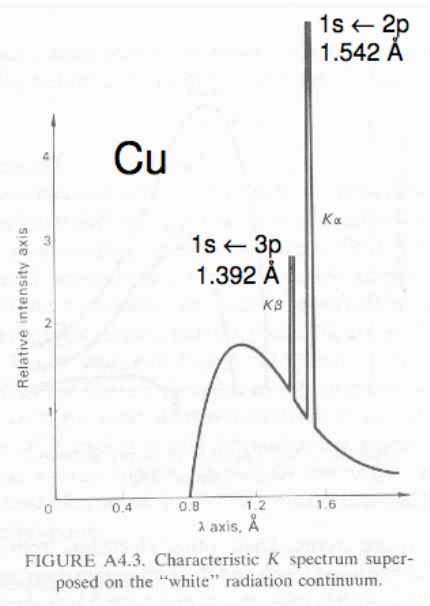


FIGURE A4.3. Characteristic K spectrum superposed on the "white" radiation continuum.

Synchrotron X-ray sources.

At a synchrotron facility, bunches of electrons, several GeV in energy, move in a large, carefully steered, closed electron beam loop containing bending elements and linear segments, collectively called the storage ring.

Properties of synchrotron radiation:

1- Intensity.

- Two order of magnitude stronger than X-ray tubes.

- The intensity of X-rays generated by modern 3rd generation synchrotron sources is so high that radiation damage to crystals has become a major concern.

- Data collection can be reached also for weakly diffracting crystals, such as very small/ tiny crystals or crystals with large unit cells.

- Low divergence of the beam resulting in sharper diffraction spots.

2- Tunability.

- Any suitable wavelength in the spectral range can be selected by monochromators.

Characterization of a crystal.

-What is the quality?

- What are the unit cell dimensions?

- To what space group do they belong?

- How many protein molecules are in the unit cell and in the asymmetric cell?

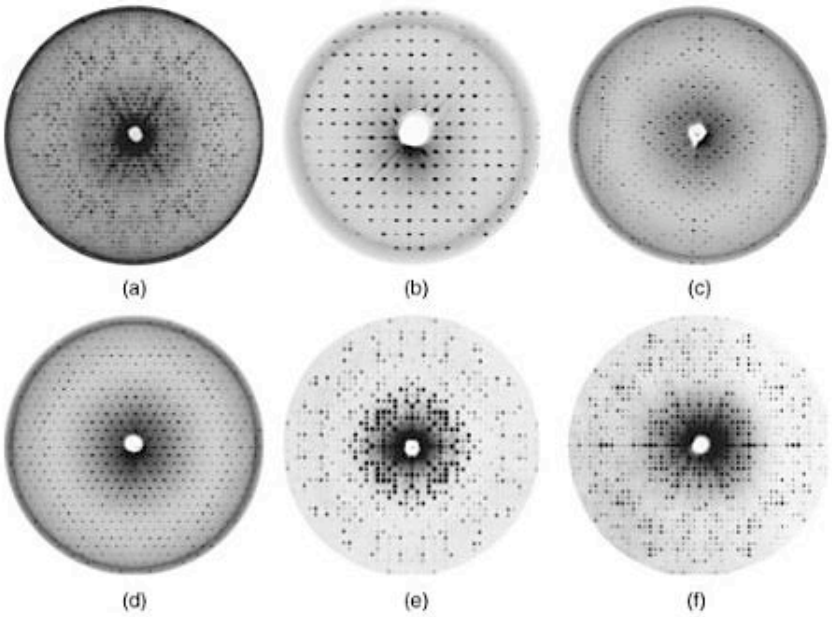


FIGURE 6.2 Diffraction photographs from a variety of protein crystals exhibiting a diversity of reciprocal lattice symmetries. With two images from a crystal, usually orthogonal to one another, the entire symmetry of the three-dimensional lattice can generally be deduced. The protein crystals and the symmetry of the diffraction patterns are (*a*) $\boldsymbol{P2_12_12_1}$ porcine $\boldsymbol{\alpha}$-amylase, the symmetry is ***mm*** (mirror-mirror); (*b*) triclinic (***P1***) glycerol-$3PO_4$-dehydrogenase, the symmetry is $\bar{\mathbf{1}}$; (*c*) centered orthorhombic (***C222***) fructose 1,6 bisphosphatase, the symmetry is ***mm***; (*d*) cubic crystal of phaeseolin viewed along body diagonal, symmetry is ***6mm***; (*e*) tetragonal ***I422*** lactate dehydrogenase, the symmetry is ***4mm***, (*f*) tetragonal crystal of lactate dehydrogenase viewed along a twofold axis, symmetry is ***mm***.

X-ray data collection

-Firstly we need to ascertain the crystal symmetry, the unit cell parameters, the crystal orientation and the resolution limit.

- Armed with this information we derive a data collection strategy which will maximize both the resolution and completeness of the data set.

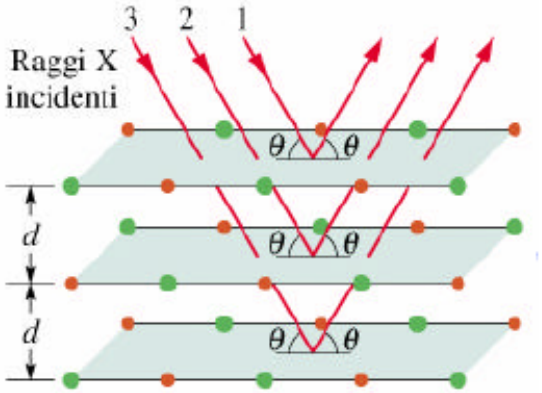
-The method we use is to rotate the crystal through a small angle, typically 1 degree, and record the X-ray diffraction pattern. If the diffraction pattern is very crowded, then the rotation angle should be reduced so that each spot can be resolved on the image.

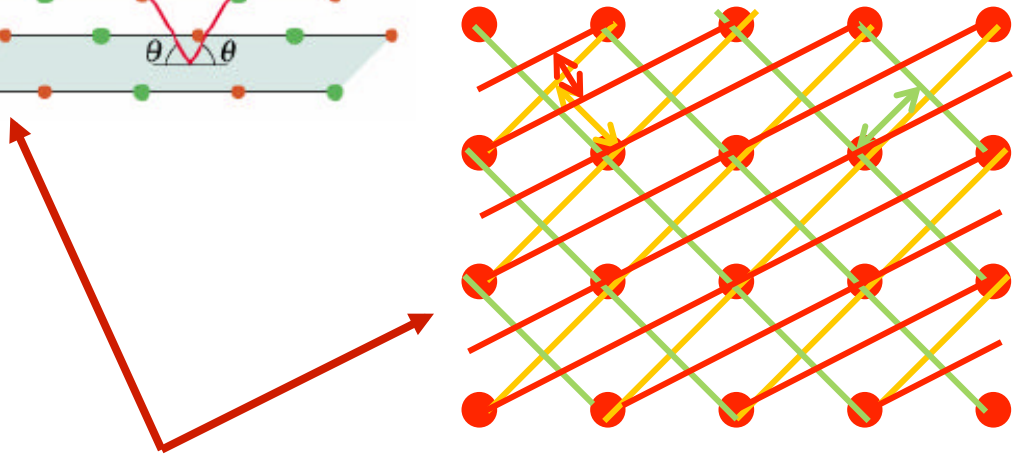
- This is repeated until the crystal has moved through at least 30 degrees and sometimes as much as 180 degrees depending on the crystal symmetry. The lower the symmetry, then more data are required.

- A typical medium resolution data set may take up to 3 days using an 'in house' X-ray source. For high resolution data collection you need to go to a synchrotron where the X-ray intensity is greater and therefore data collection times are shorter - sometimes as fast as 10 minutes! Data sets can contain around 300 images with over 5000 spots per image.

Diffraction and Bragg

In a crystal there are countless families of parallel planes with interplanar distances variables able to give diffraction effects X according to Bragg's law.



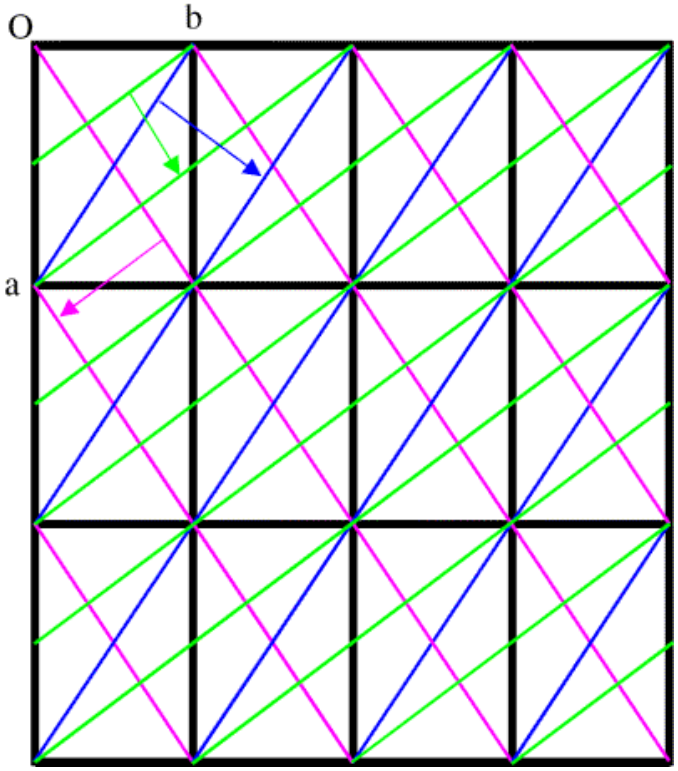


Families of parallel planes

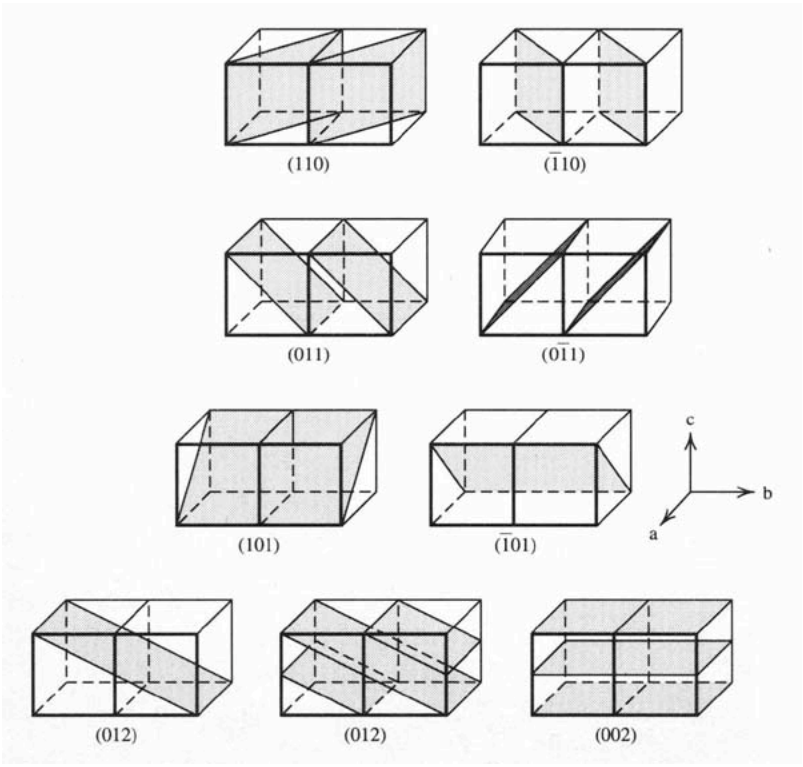
In general, from a small single crystal of dimensions less than a millimeter, it is possible to record from thousands to tens of thousands of diffraction effects also called “reflections"

Bragg's planes and Miller indeces

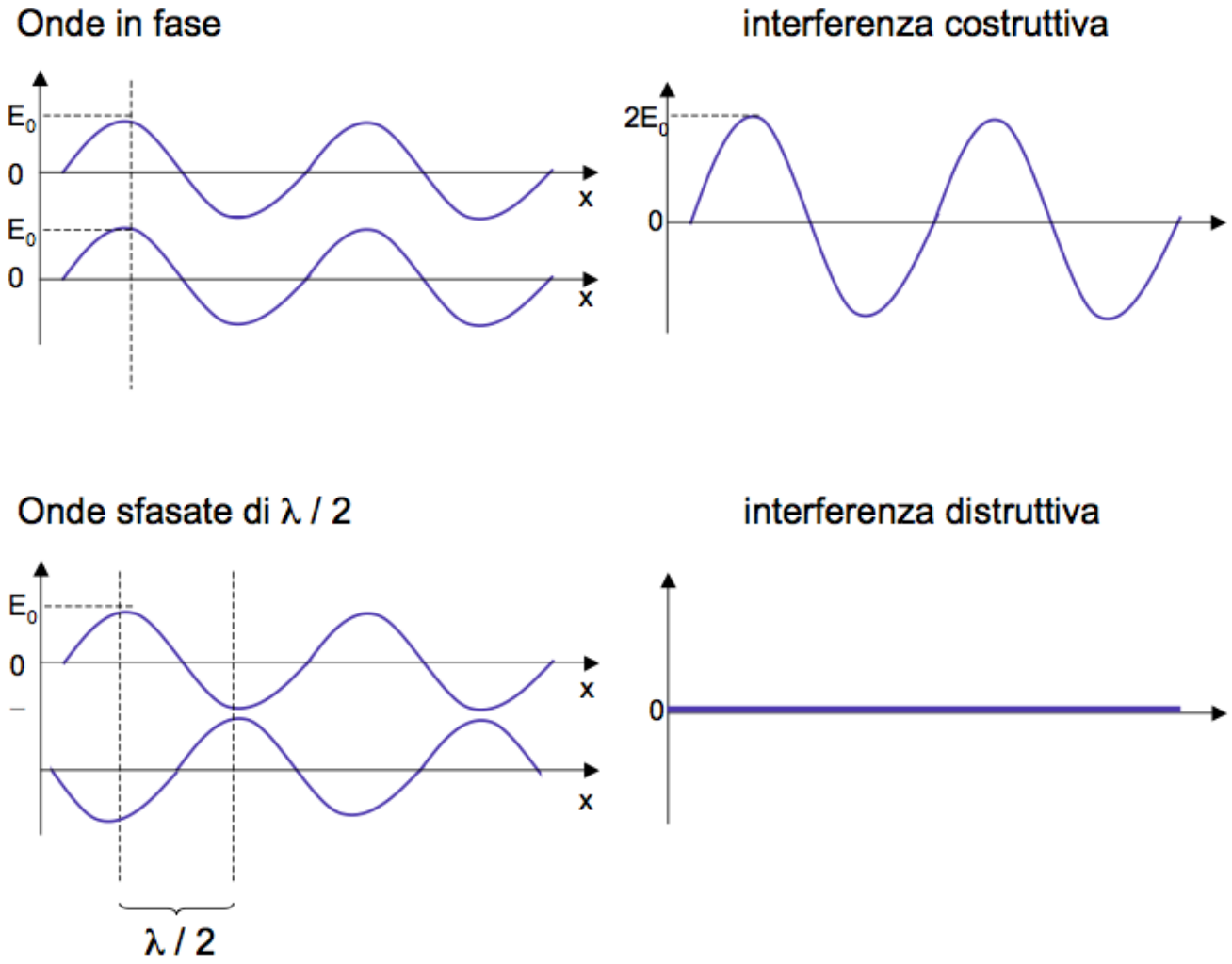
To see diffraction from a crystal, the Bragg planes have to cut all the cell edges an integral number of times. On the other hand, if the planes divide the unit cell edge by a non-integral number, the different unit cells will all diffract out of phase and the waves will cancel out.



blue planes (110)
magenta planes (1-1 0)
green planes (2 1 0)



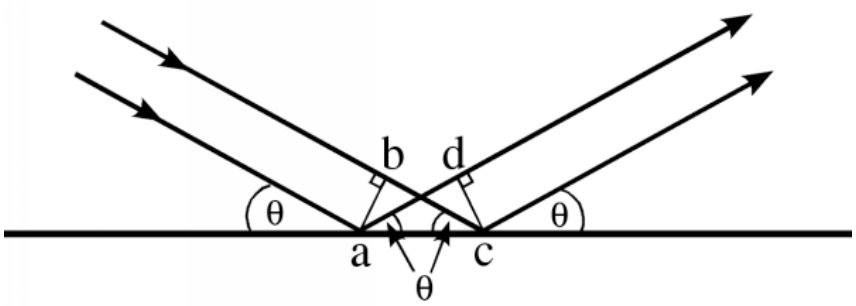
Diffraction: waves in phase



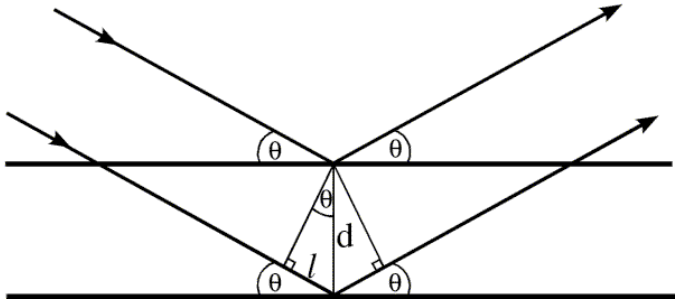
Diffraction: waves in phase

When do waves scatter in phase?

When they have exactly the same pathlength



When their pathlengths differ by a multiple of the wavelength



$l = d \sin\theta$

$n\lambda = 2\, d \sin\theta$

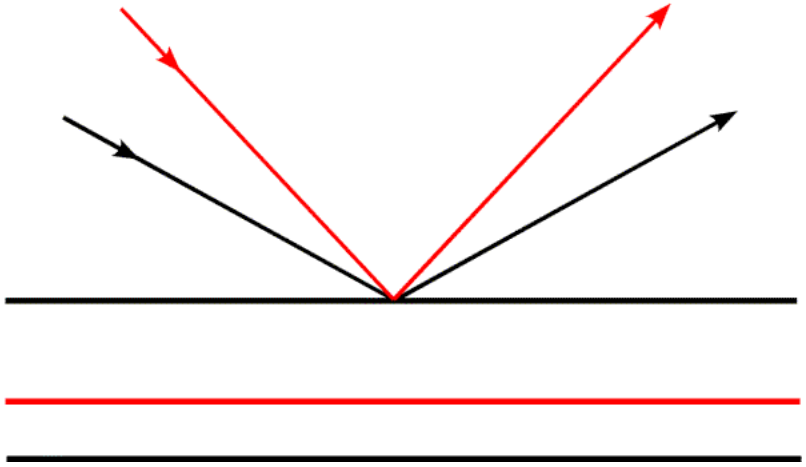
Bragg's low

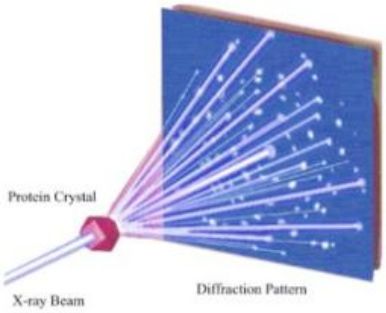
The angle *acb* is the angle of incidence, and the angle *cad* is the angle of reflection. When these angles are the same, then the triangles are indeed congruent and the sides *ad* and *bc* must be the same. So the rays reflected off of two points on the plane have identical pathlengths, and remain in phase with one another.

Bragg's low

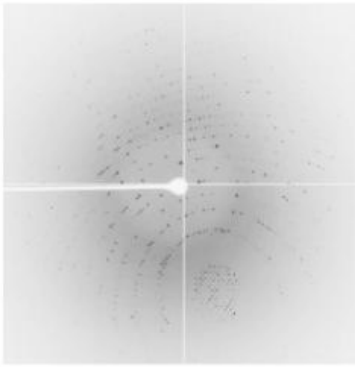
In Bragg's law, as the angle increases, d must become smaller for the pathlength to remain equal to one wavelength.

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





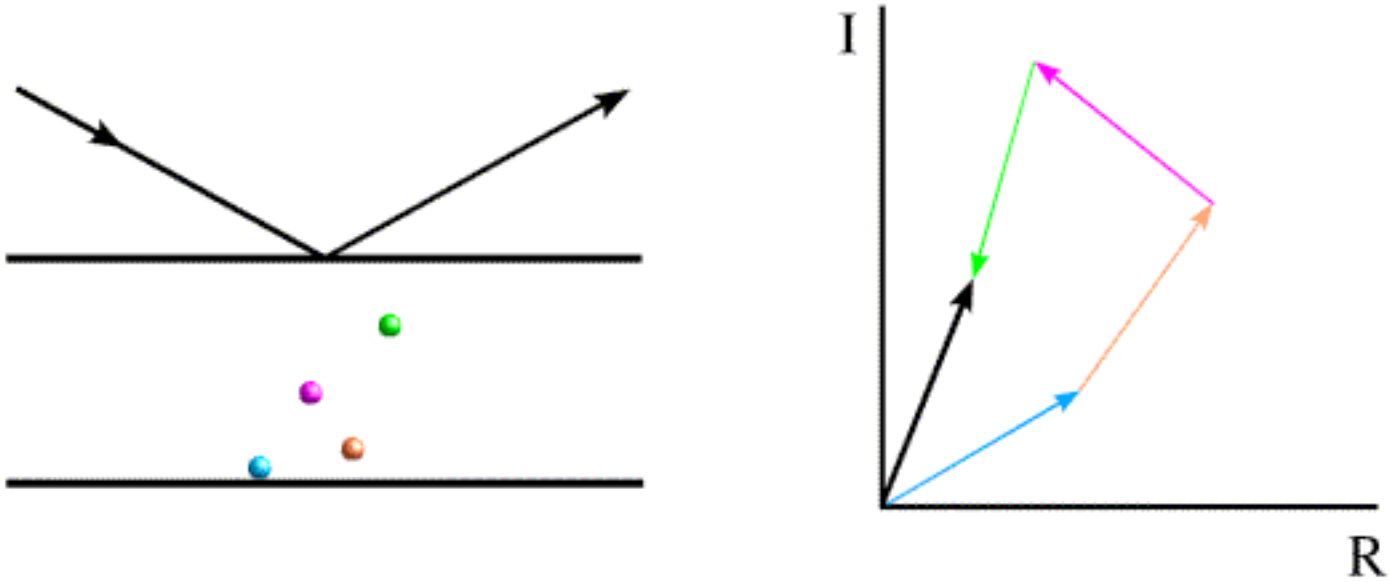
Diffraction Process



Diffraction Pattern from NSLS

The bigger the angle of diffraction, the smaller the spacing to which the diffraction pattern is sensitive.

Structure factor

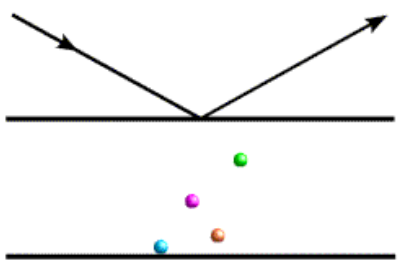


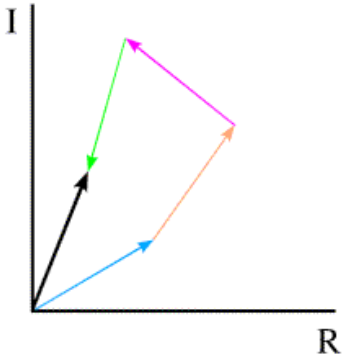
The overall scattered wave is represented by a black vector, which is the sum of the other vectors.

The vector (amplitude and phase or, more properly, the complex number) representing the overall scattering from a particular set of Bragg planes is termed the structure factor, and it is usually denoted **F**.

Structure factors

The structure factor represents the wave that results from diffraction, which means that it is a complex number with amplitude and phase. To put the structure factor on absolute scale, the diffraction from a single electron at a point is defined as having an amplitude of $1e$.





Structure factor from a single electron

$$\mathbf{F}(\mathbf{s}) = 1e \exp(2\pi i \mathbf{s}\cdot\mathbf{r})$$

Structure factor from a number of electrons

$$\mathbf{F}(\mathbf{s}) = \sum_j \exp\left(2\pi i \mathbf{s}\cdot\mathbf{r}_j\right)$$

Structure factor from a crystal

$$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j \exp 2\pi i(hx_j + hy_j + lz_j) = |\mathbf{F}(hkl)| \exp ia(hkl) \qquad (3.63)$$

with $|\mathbf{F}(hkl)|$ – amplitude and a – phase angle. We obtain the intensity of the scattered wave as the structure factor $\mathbf{F}(hkl)$ multiplied by its complex conjugate value according to (Eq. (3.64)):

$$I(hkl) = \mathbf{F}(hkl)\mathbf{F}^*(hkl) = |\mathbf{F}(hkl)|^2 . \qquad (3.64)$$

The electron density equation

In the study of Fourier series, complicated periodic functions are written as the sum of simple waves mathematically represented by sines and cosines. Due to the properties of sine and cosine it is possible to recover the amount of each wave in the sum by an integral. In many cases it is desirable to use Euler's formula, which states that $e^{2\pi i\alpha} = \cos 2\pi\alpha + i \sin 2\pi\alpha$, to write Fourier series in terms of the basic waves $e^{2\pi i\alpha}$.

If we carry out an inverse Fourier transform of the structure factors (amplitudes and phases), we get a picture of the electron density.

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^{N} f_j \exp 2\pi i \mathbf{r}_j \mathbf{S} = \int_{\text{vol. of unit cell}} \rho(\mathbf{r}) \exp 2\pi i \mathbf{r}\mathbf{S} \, dv$$

Each structure factor can be seen to add a single cosine wave to the picture of the electron density. As we know from our study of the structure factor equation, the amplitude of each structure factor tells us the extent to which electrons are concentrated on planes parallel to the Bragg planes. The phase tells us where these concentrations are found.

The temperature factor

- Atoms vibrate around an equilibrium position.

- X-rays do not meet exactly the same position in successive unit cells

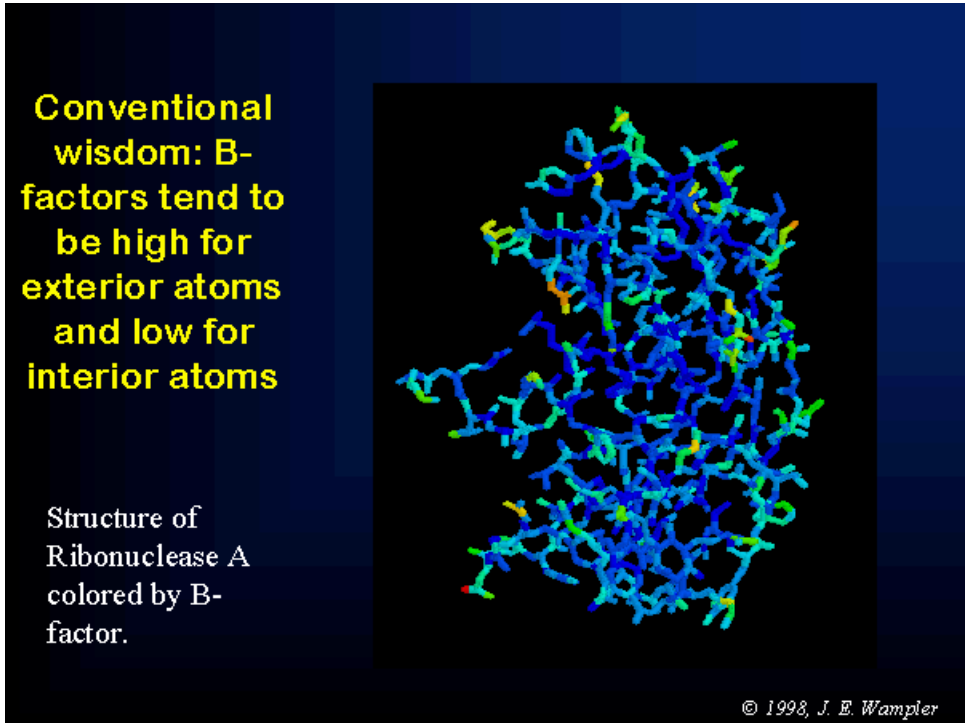
- The thermal vibration has to be considered as it diminishes the scattered X-ray intensities.

- Assuming isotropic and harmonic vibration:

$$B = 8\pi^2 \times \overline{u}^2$$

where $\overline{u}^2$ is the mean square displacement

Average values for B in protein structures ranges from low to 30 Å^2 in well ordered structures. The highest values are found in more or less flexible surface loops



Example.

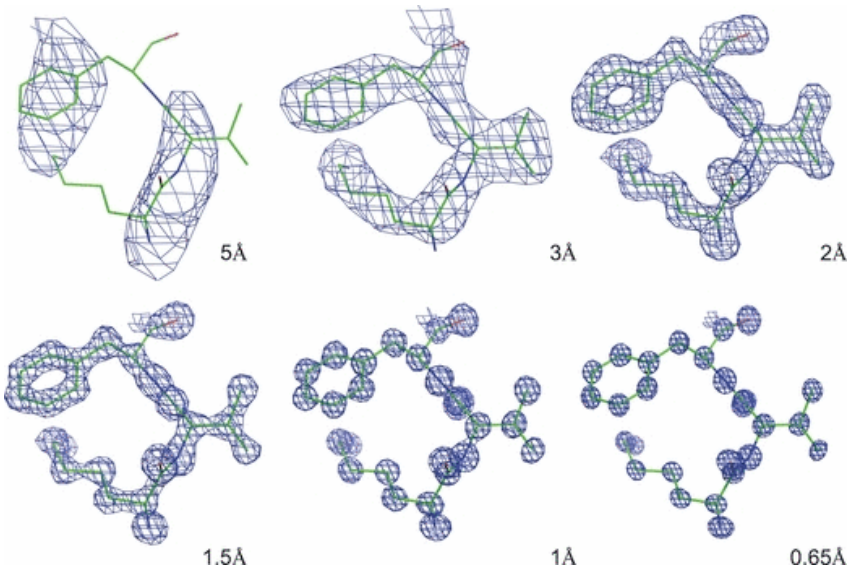
For B = 30 Å^2, the root mean displacement is $\sqrt{30/8\pi^2}$ = 0.62 Å^2

Proteins are not rigid and there is an intrinsic disorder in protein crystal.

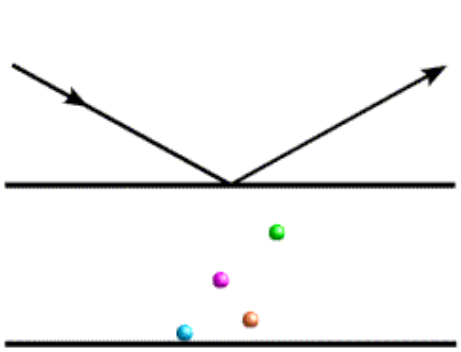
This is the reason why the diffraction pattern fades away at some diffraction angles θ_{max}

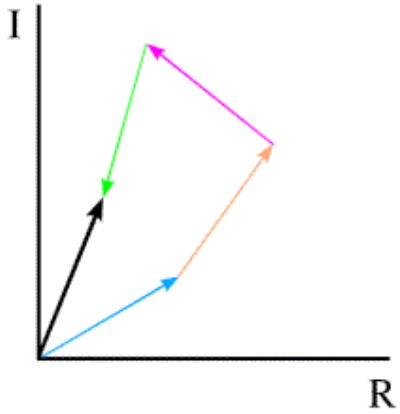
Resolution: $d_{min} = \lambda / 2\sin\theta_{max}$

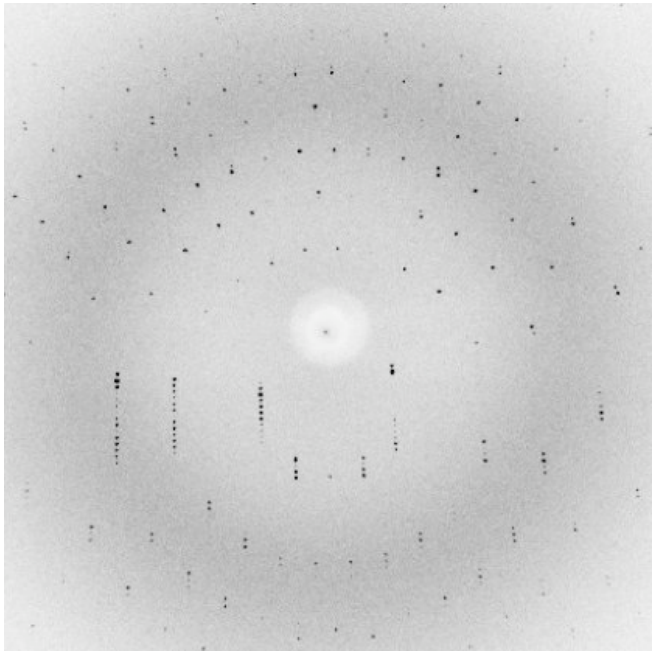
Resolution (Å)	**Visible**
3.5	Polypeptide chain fold
3-2.2	Aa side chains
< 2.0	Aromatic side chains holes



The phase problem







The vector (amplitude and phase or, more properly, the complex number) representing the overall scattering from a particular set of Bragg planes is termed the structure factor, and it is usually denoted **F**. The structure factors for the various points on the reciprocal lattice correspond to the Fourier transform of the electron density distribution within the unit cell of the crystal. A very convenient property of the Fourier transform is that it is reversible; if you apply an inverse Fourier transform to the structure factors, you get back the electron density.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h,k,l} |F(h,k,l)| e^{i\alpha(h,k,l)} \exp\left[-2\pi i(hx+ky+lz)\right]$$

Methods for the solution of the phase problem

- Molecular Replacement (**MR**)
- Heavy metals: Se, Fe, Hg, Pt, Sm, …

MIR

Multiple Isomorphous Replacement (IR)

S/MIRAS

Single/Multiple IR with Anomalous Signal

MAD

Multiple Anomalous Dispersion: derives from the anomalous scattering of “special”atoms present in the protein with energy transition close to the X-ray energy.

The Isomorphous Replacement Method (IR)

Requirements:

- diffraction pattern of the native protein crystal and of at least one heavy atom derivative;

- perfect isomorphism (protein conformation, position and orientation of molecules, unit cell parameters must be the same);

Principle:

The intensity differences between the native and the other patterns are due to the attached heavy atoms. From these differences the position of the heavy atoms can be derived: this is the starting point for the determination of the protein phase angles.

Obstacles:

- Nonisomorphism often occours and gives rise to errors (change in unit cell dimensions, changes in intensities).

- Change in cell dimensions and quality of the diffraction pattern (a change in the cell dimensions of $d_{min}/4$ where d_{min} is the resolution limit, is tolerable).

The Isomorphous Replacement Method: steps.

1. Preparation of at least one, but preferably a few heavy atoms containing derivatives of the protein crystals.
2. Data collection for native and derivative crystals.
3. Application of the Patterson function for the determination of the heavy atom coordinates.
4. Refinement of the heavy atom parameters and calculation of the protein phase angles.
5. Calculation of the electron density of the protein.

The attachment of heavy atoms.

Methods.

-Soaking

- Co-crystallization

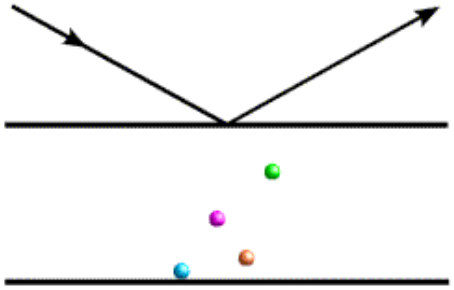
Site of attachment of heavy atoms

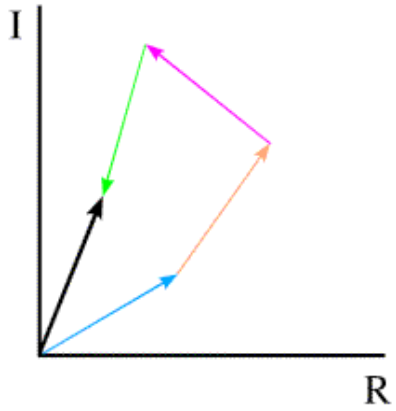
- The choice of heavy atom compounds is mostly an empirical decision
- Several rules (refer to Blundell and Johnson for a detailed discussion):
  - Cys: *Hg, Pt, Au*
  - Met, His: *Pt, Au, Ag*
  - Asp, Glu: *Pb,* UO_2
  - Xe gas can bind to hydrophobic pocket
- Genetically modification
  - Cys mutation
  - Se-Met protein

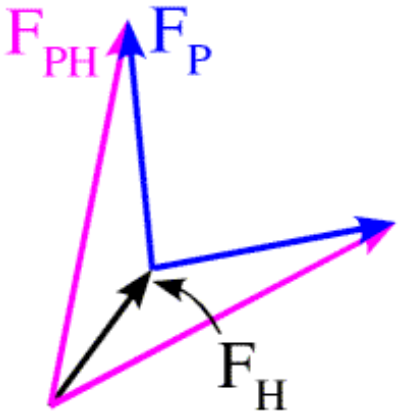
Problems commonly encountered in the search of heavy atom derivatives

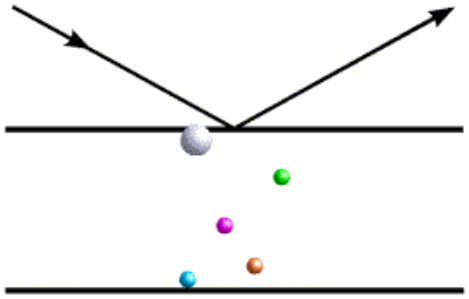
- Phosphate buffer
  - Phosphate buffer is often used in protein purification
  - Phosphates forms insoluble aggregates with many heavy atom compounds, including uranyl ions and rare earth elements
  - Solution: replace phosphate buffer by “good” buffers
- Ammonium sulfate
  - AS is a common precipitant
  - Ammonium can chelate many heavy atoms
  - High ionic strength prevents binding
  - Solution: replace AS by Li_2SO_4 or PEG.

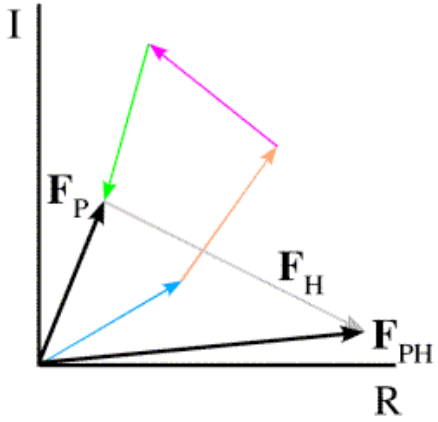
Effect of heavy atoms on X-ray intensity.











$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

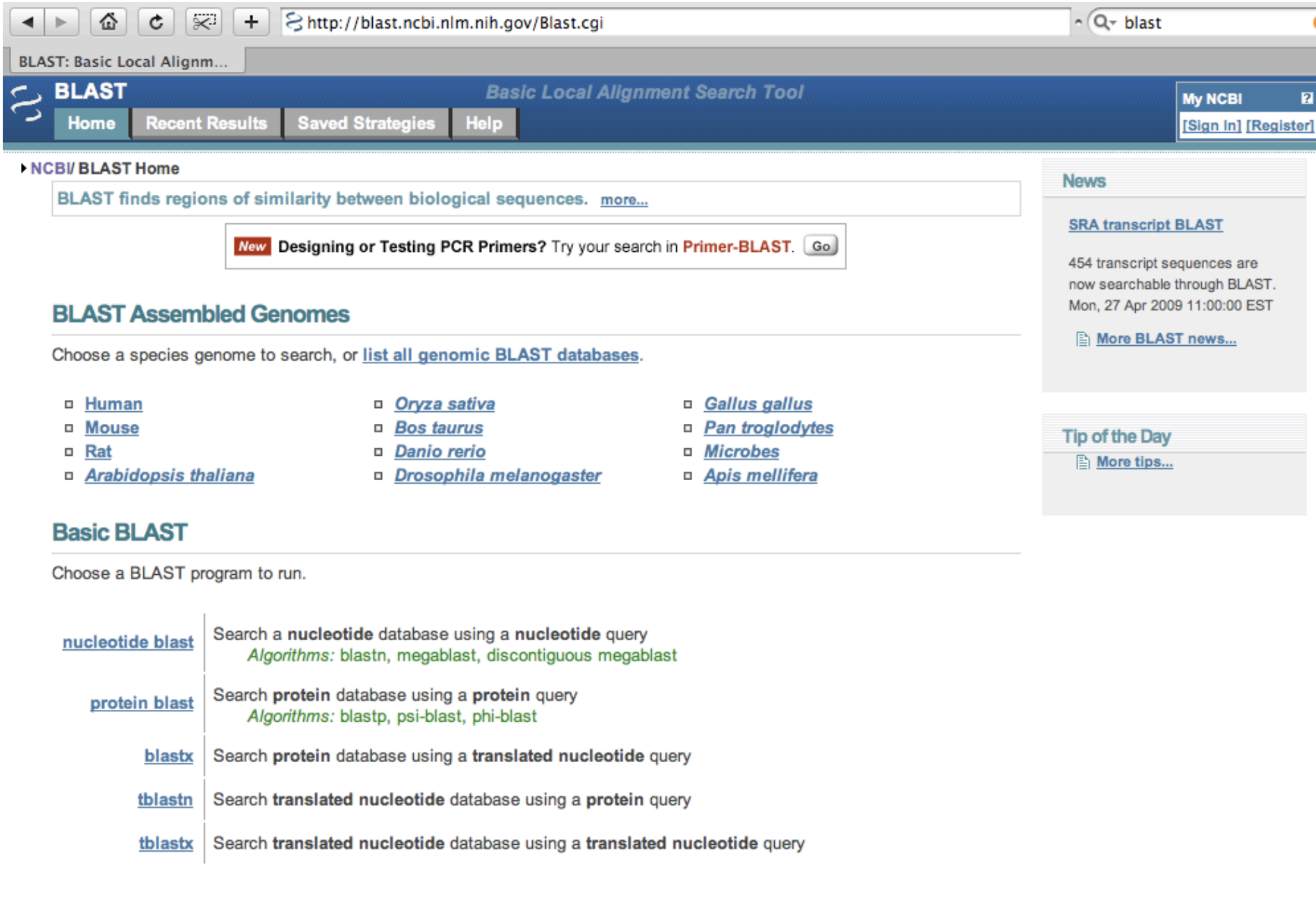
Molecular replacement

- Molecular replacement can be used to solve a structure when you have a good model for a reasonably large fraction of the structure in the crystal.

-The level of resemblance of two protein structures correlates well with the level of sequence identity.

- As a rule of thumb, molecular replacement with modern programs will probably be fairly straightforward if the model is fairly complete and shares at least 30% sequence identity with the unknown structure.

Searching homologous sequences: blast



PDB: the Protein Data Bank (http://www.rcsb.org/)

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

RCSB Protein Data Bank

MyPDB Login

A MEMBER OF THE PDB

RCSB PDB PROTEIN DATA BANK

An Information Portal to Biological Macromolecular Structures

As of Tuesday Jun 09, 2009 there are 58083 Structures | PDB Statistics

WHAT'S NEW | CONTACT | FEEDBACK | HELP | PRINT | PDB ID or keyword | Search | Adv. Search

Home | Search

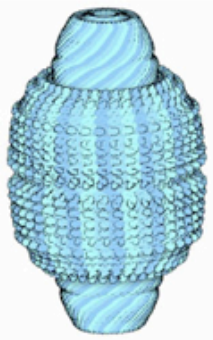
- Home
- Getting Started
- Structural Genomics
- BioSync
- Electron Microscopy
- Download Files
- Deposit and Validate
- Dictionaries & File Formats
- Software Tools
- General Education
- General Information
- Acknowledgements
- Frequently Asked Questions

A Resource for Studying Biological Macromolecules

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

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

Molecule of the Month: Vaults



Our cells are filled with compartments, each performing a specific function. Some of these compartments, such as mitochondria and lysozomes, are very large and enclose many different molecular machines. Other intracellular compartments are smaller, such as the transport vesicles that shuttle proteins from site to site inside the cell. Most of these compartments, including mitochondria, lysozomes and transport vesicles, are surrounded by membranes. However, in special cases, cells build smaller compartments surrounded by a protein shell. In our own cells, vaults are a spectacular example of these protein-enclosed compartments.

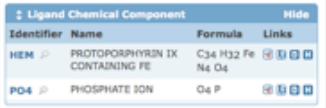
Read more ... | **Previous Features**

News

- **Complete News**
- **Newsletter**
- **Discussion Forum**
- **Job Listings**

09-June-2009

Customizable Structure Summary Pages



Tabbed Structure Summary pages let users explore information about any given structure. Located above the structure ID and title, each tab offers different types of information -- a **general overview**, **derived data**, **sequence details**, **sequence similarity**, **literature view**, **biology & chemistry**, **methods**, **geometry**, and **links to external resources**.

More >>

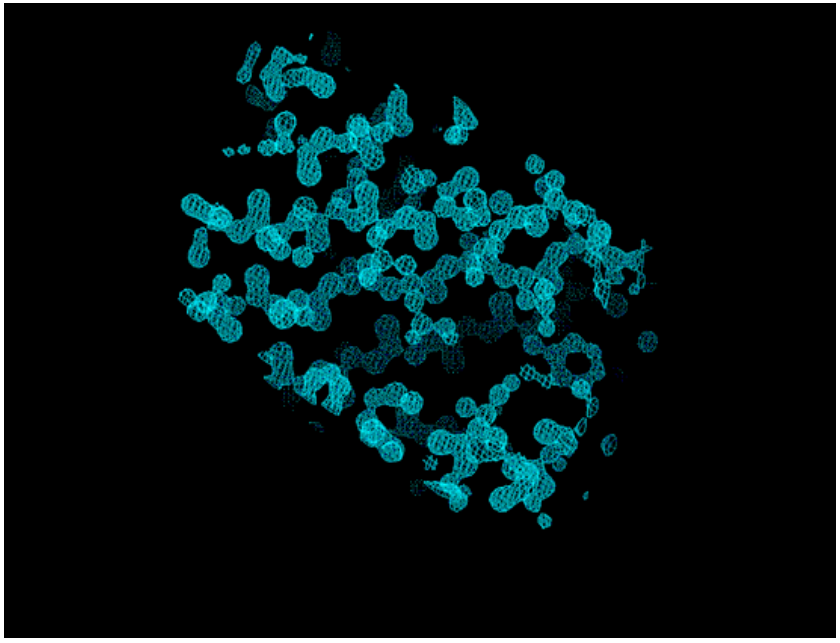
Fitting, refinement and validation

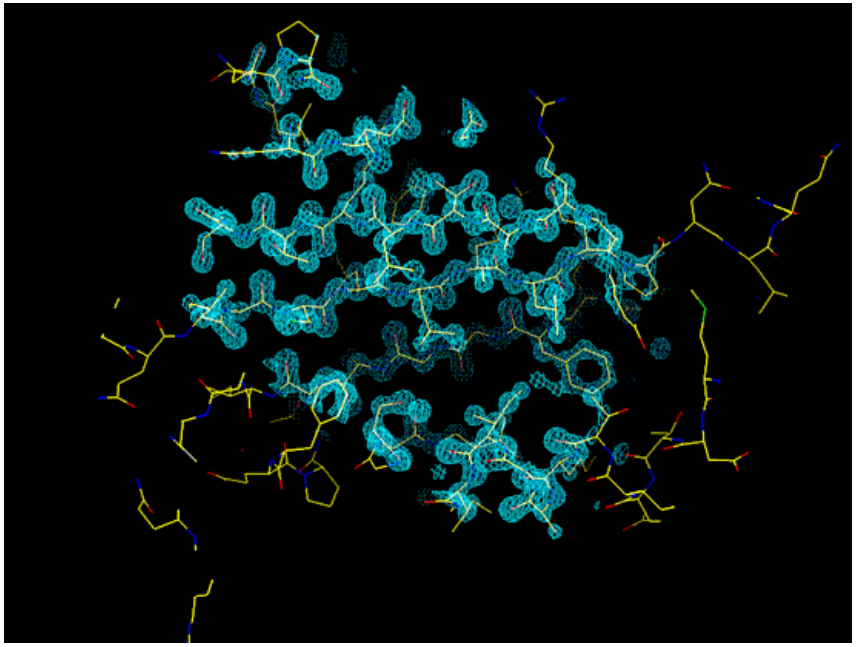
Electron density maps, whether from an initial structure determination (eg by MIR, MAD) or during refinement, are not self-explanatory. They need human or machine intelligence to interpret.

The task of electron-density fitting is to interpret the electron-density maps in the light of chemical knowledge, of basic stereochemistry, the chemical sequence, and the nature of any bound ligands

Resolution

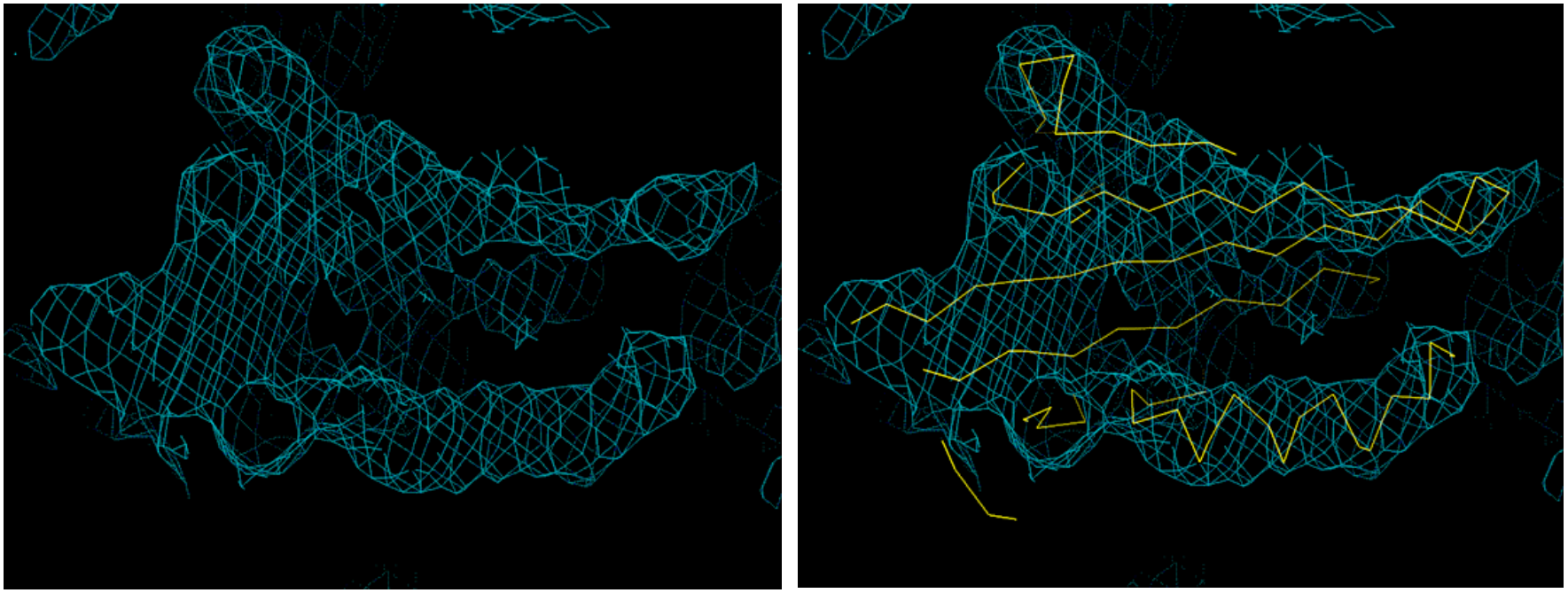
At very high resolution, individual atoms can be fitted, and the problem is join-the-dots. Labelling remains a problem, but a relatively easy one.





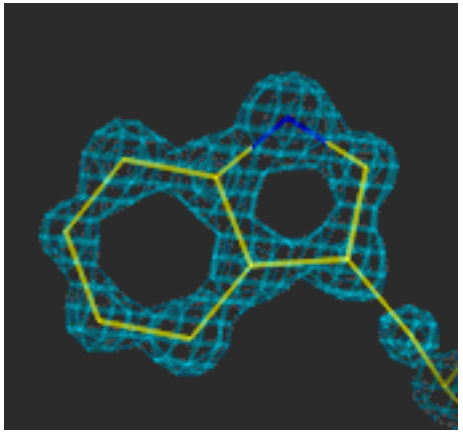
1 Å resolution

At the other extreme, large chunks must be fitted at one time. Alpha-helices are clear at 6 Å resolution, but beta-sheets are not. At lower resolutions than about 8 Å, only whole molecules can be placed.

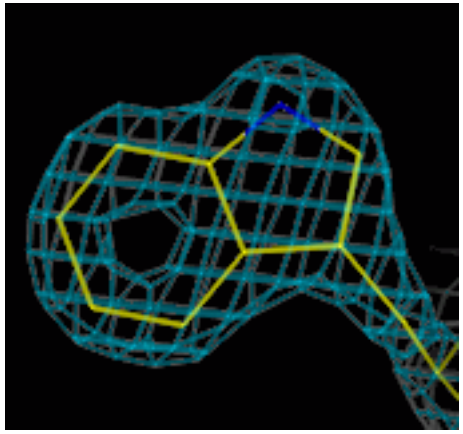


6 Å resolution

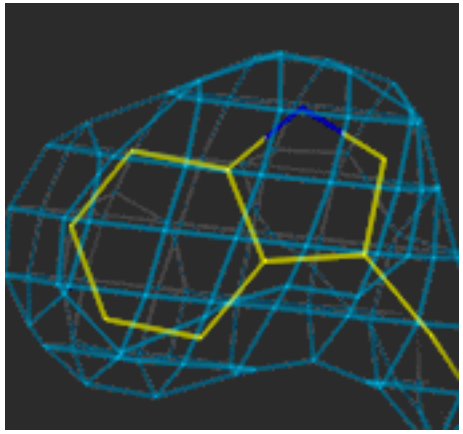
At 1.0 Å, there is no problem fitting individual atoms (and the N atom is bigger than the carbons). At 2.5 Å the ring is easily fitted, at 3.0 Å less easily, and at 4 Å the fit is very uncertain.



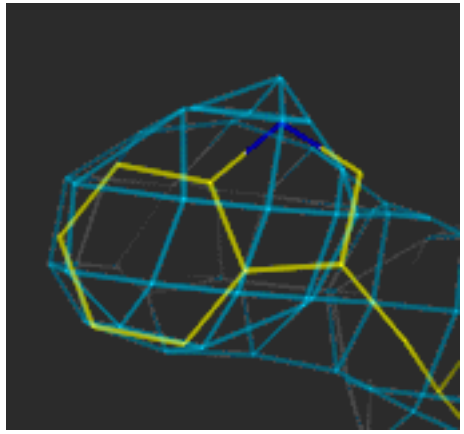
1.0 Å



2.5 Å



3.0 Å



4.0 Å

Fitting

Automatic fitting programs.

There are now a number of programs which automate some or all of the process: these and others are under active development, and will undoubtedly improve in the future.

Manual fitting

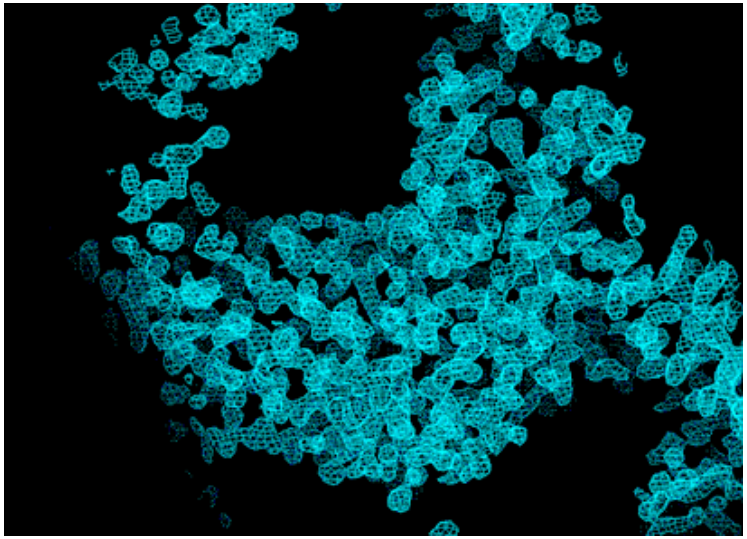
Model building is too complicated to do in one step, but it may be broken down into stages.

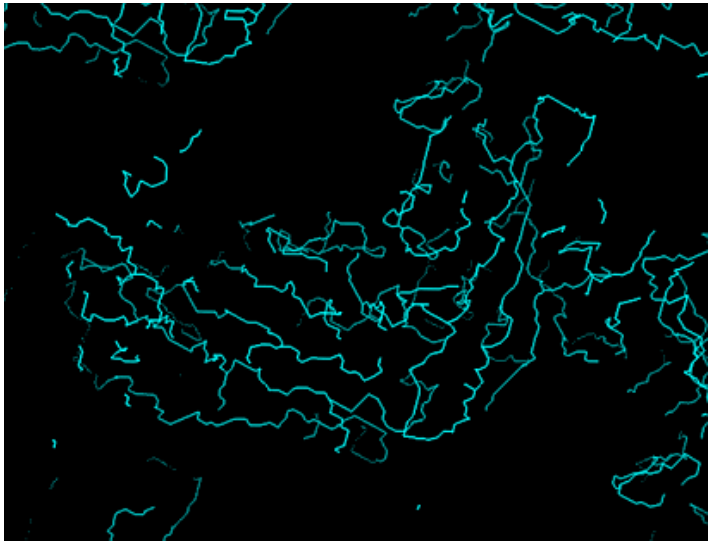
Remember that what you are constructing is a **model**, a **hypothesis**, and that it may make sense to build multiple models for parts (or all) of the structure. Anything can be changed.

- Polynucleotides can be fitted in a similar way in principle, but the programs are less well developed.

Tracing the chain with a skeleton (bones).

The skeleton is a simpified representation which allows the whole molecule and major secondary structure features to be seen at once, to get an overall view of the structure.

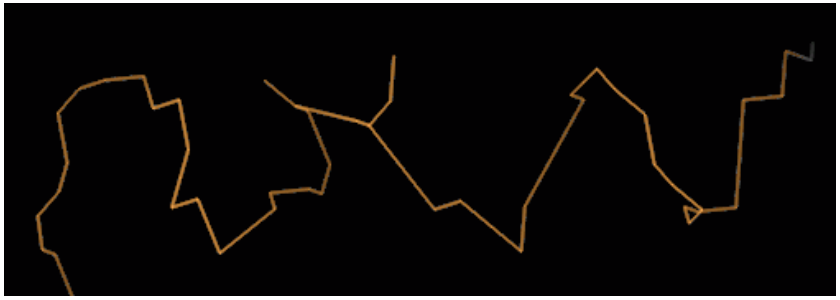


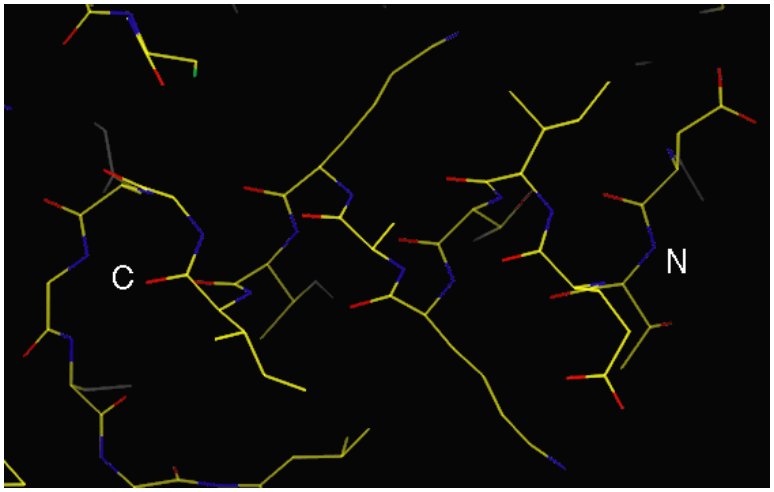


Positioning the sequence on the trace.

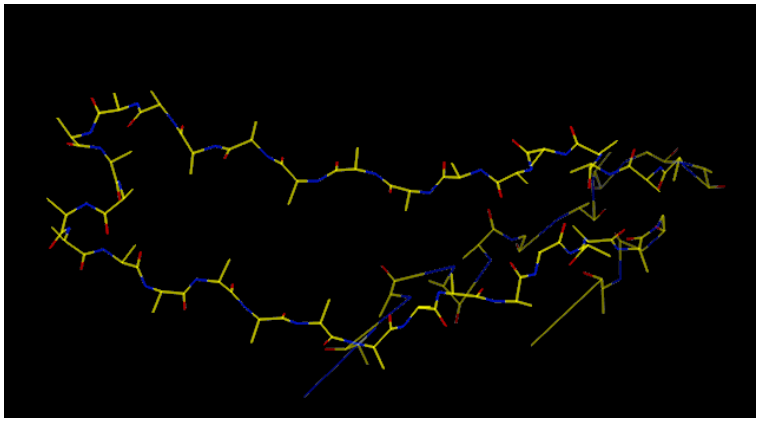
A. Markers: Recognisable aminoacids (SeMet, Hg - Cys), active site, prosthetic groups etc

B. Direction

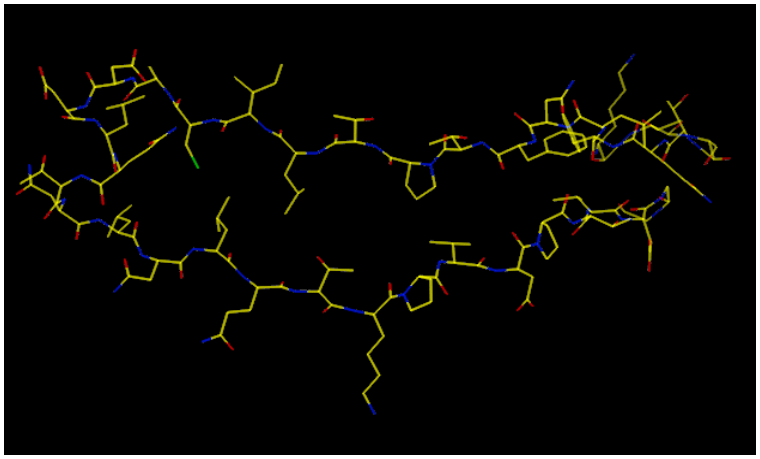




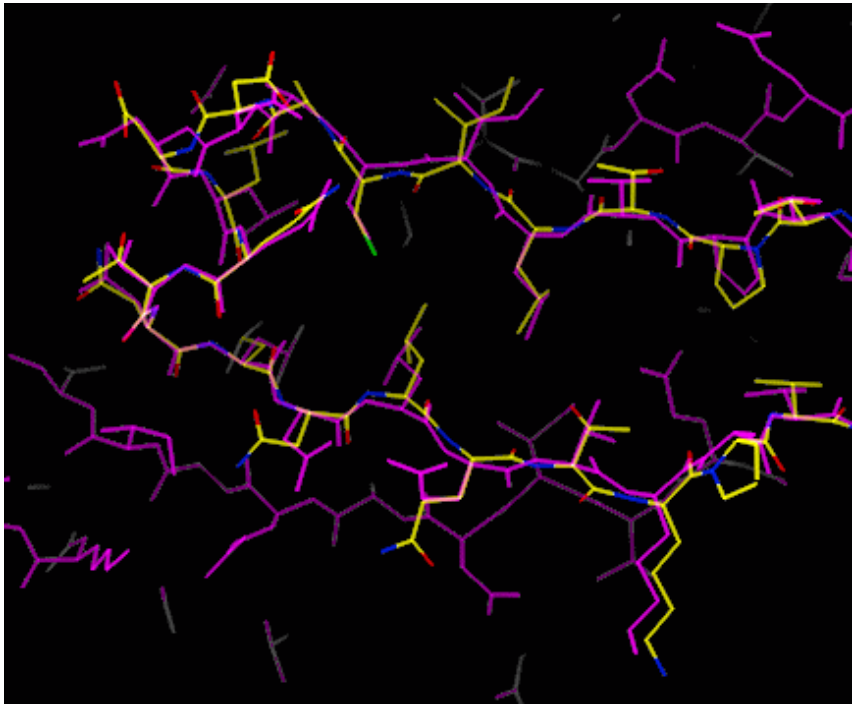
First CA trace along skeleton



Mainchain fitted to CA guides



Sidechains built in most common rotamer



The model may be quite good at this stage, but will have some mistakes .

Magenta: final model;

Yellow: model from fitting

Adjustment of model (rebuilding)

Tools for rebuilding

Tools based on stereochemical libraries offer choice of:

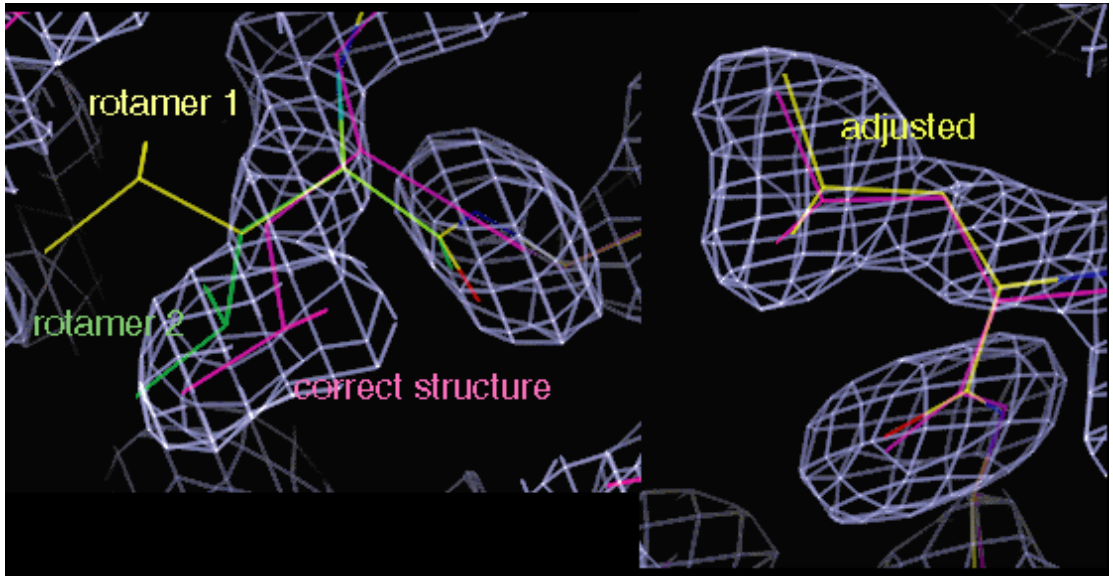
- loop which best fits CA atoms

- sidechain libraries: most sidechains are in rotamer conformations, so these should always be preferred over alternatives

- torsion angle rotation.

Free movement tools

Move around part of structure, possibly without concern for preserving stereochemistry. This is often useful when you can see where to go, but not how to get there.



A leucine sidechain changed to rotamer 2, and adjusted into density

Refinement

A major role of refinement is to **produce new electron density maps** for examination and for correcting errors in the model (manually or automatically). Maps from refinement are typically better than those from experimental phases, and should improve as the model improves.

Two main sorts of maps are useful:

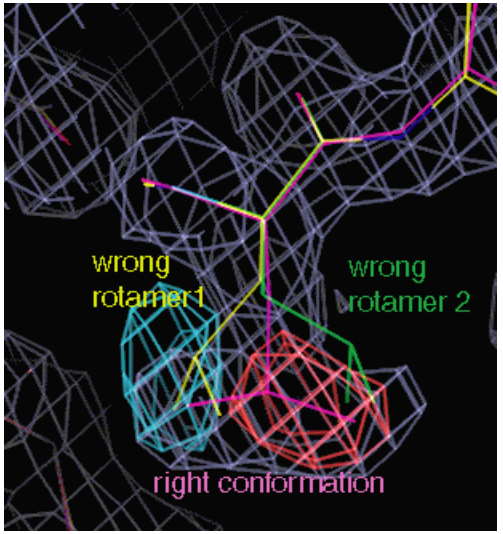
• "2Fobs - Fcalc" type: these show the current best estimate of the electron density for the structure. This is the map which the model should fit.

• Difference map, mFobs - DFcalc: this shows the best estimate of the difference between the true structure and the current model. Ideally, positive density indicates atoms should be added, negative density that they should be removed (or moved elsewhere), and a positive/negative pair indicates that atoms should move torwards the positive density.

During rebuilding, it is useful to display both the 2Fo-Fc map, and positive and negative contours of the difference map, coloured differently (note that at least two colour conventions are in use

(1) red is positive

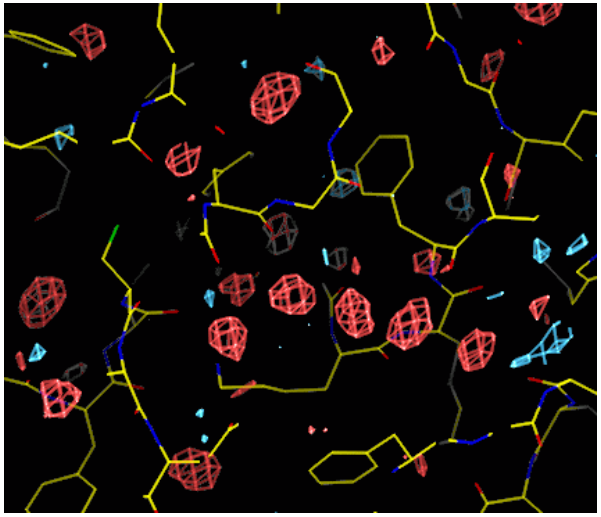
(2) blue is negative



Red: positive difference density; blue: negative: this is a leucine, unusually not in a standard rotamer conformation, but clear

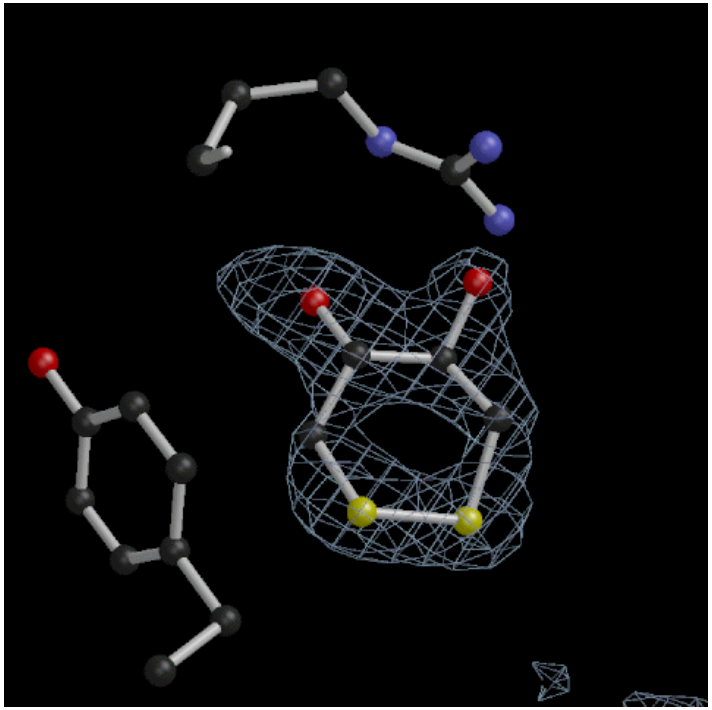
Water molecules and other UFOs

Waters are an important part of the structure: a well-ordered water molecule contributes more to the X-ray scattering than a poorly ordered part of the protein. They are clearly visible in expermental maps and particularly in difference maps, at least at medium to high resolution. At resolutions worse than 2.8 - 3 Å, waters cannot generally be placed reliably: the free R-factor is a good guide to whether adding waters improves the model.

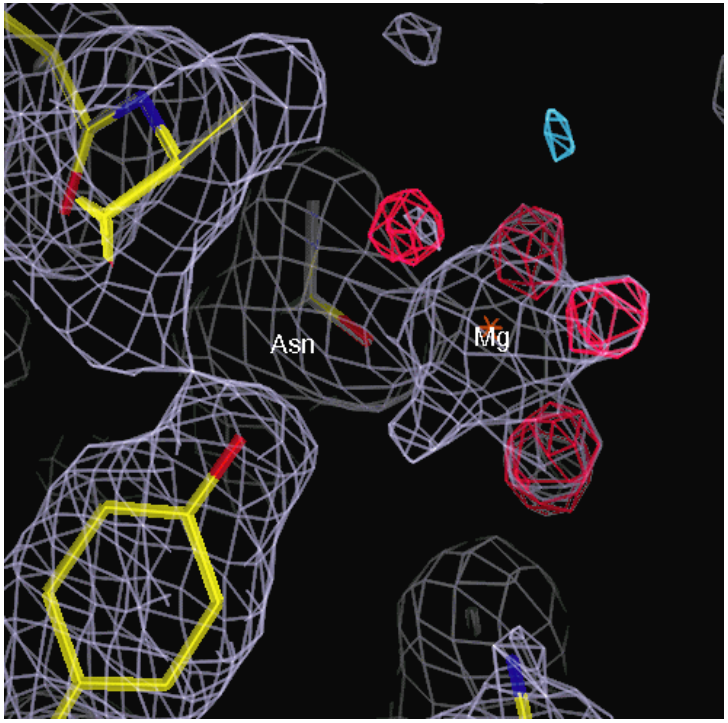


Waters appear as spherical positive features in the difference map

Other features such as expected or unexpected ligands and bound ions may appear in the maps and should be added to the model when they can be interepreted (what did you put into your crystallisation mix?)



Oxidised DTT, 1.7 Å resolution - probably more common than is recognised



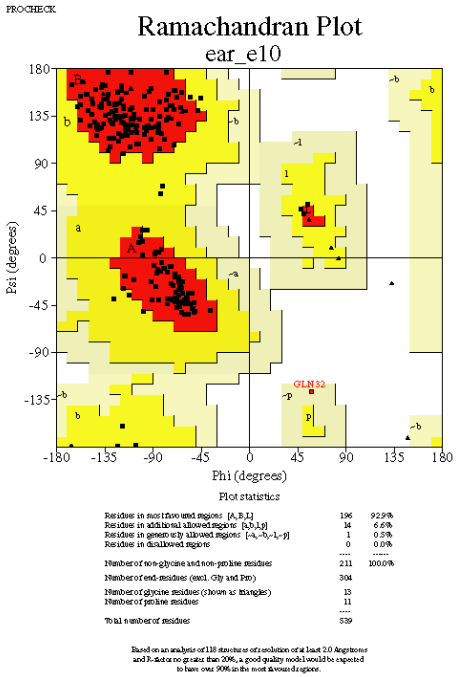
Hydrated magnesium bound to sidechain, ocatahedral coordination.

Validation: how do you know that you are right?

Most of these checks compare the model to common properties of other similar macromolecules or small molecules. These properties reflect the energetics of molecular conformation, so a region of the model which deviates significantly from normal is either a mistake, or is an unusual high-energy conformation and thus may be important.

Mainchain torsions: the Ramachandran plot

The mainchain Phi (N-CA) & Psi (CA-C) torsion angles are highly constrained by steric hindrance, and any residues falling outside low energy regions are suspect. Glycine residues are more tolerant of unusual conformations.



Useful references.

Tutorials about X-Ray Crystallography:

A Protein Crystallography Course http://www-structmed.cimr.cam.ac.uk/course.html

Protein Crystallography Tutorial Site (Crystallography 101)
http://www.ruppweb.org/Xray/101index.html

Interactive Tutorial about Diffraction
http://www.mineralogie.uni-wuerzburg.de/crystal/teaching/teaching.html

Webmineral - The different crystal froms http://webmineral.com/crystall.shtml

Book:

Principles of protein x-ray crystallography. By Jan Drenth