ICTP School on Synchrotron Radiation and Application

Trieste 10.05.04

Protein Crystallography:
Tutorial on Crystal Structure Determination

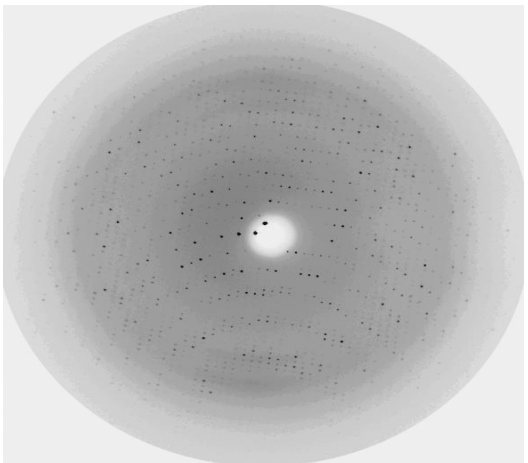
Alberto Cassetta

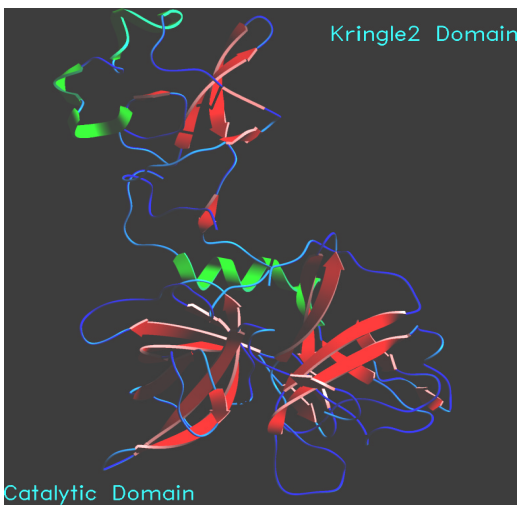
CNR - Istituto di Cristallografia – Trieste

Summary

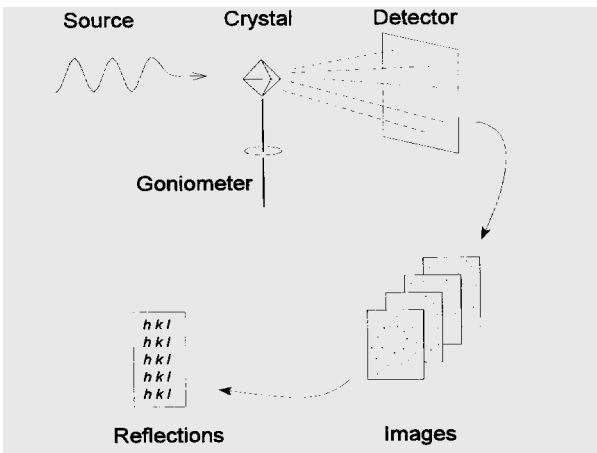
- Diffraction
- Data Collection
- Structure Solution
- Refinement

Diffraction – The method

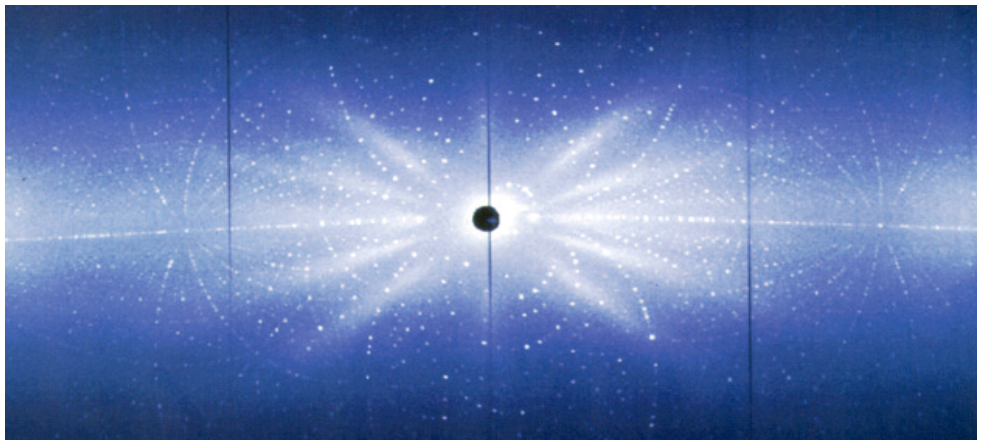


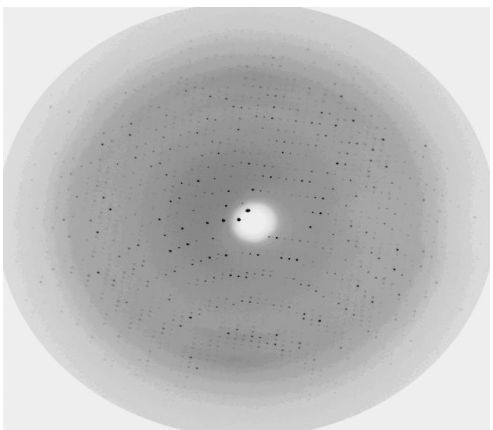


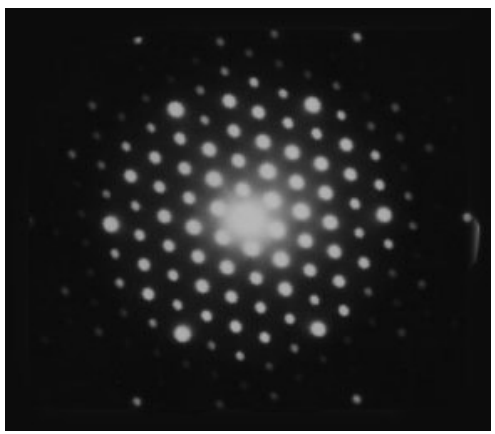
- X-ray diffraction from single crystal is still the main method for proteins structure determination.



Diffraction – Diffraction and Crystals







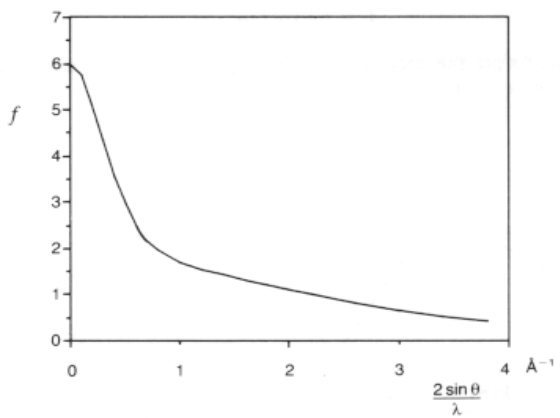
- Crystals behave as a three-dimensional diffraction net.

- The interaction between waves of a given wavelength such as X-rays, but even neutrons and electrons give rise to diffraction phenomena.

Diffraction – X-ray electrons interaction

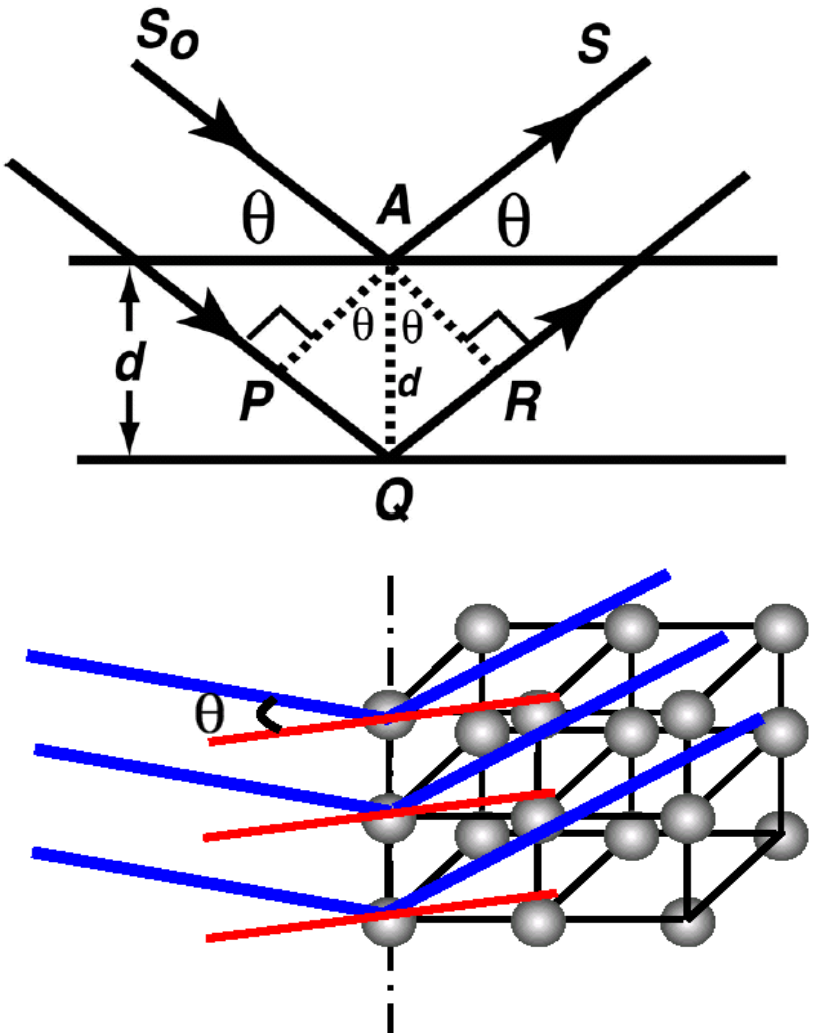
- X-rays interact with electrons, which absorb X-ray photons starting to oscillate in phase with the incident wave. Oscillating electrons emit "secondary" wave which conserve energy and original phase difference (elastic scattering).

- Scattered waves from electrons of all atoms inside the crystal interfere, generating diffracted waves



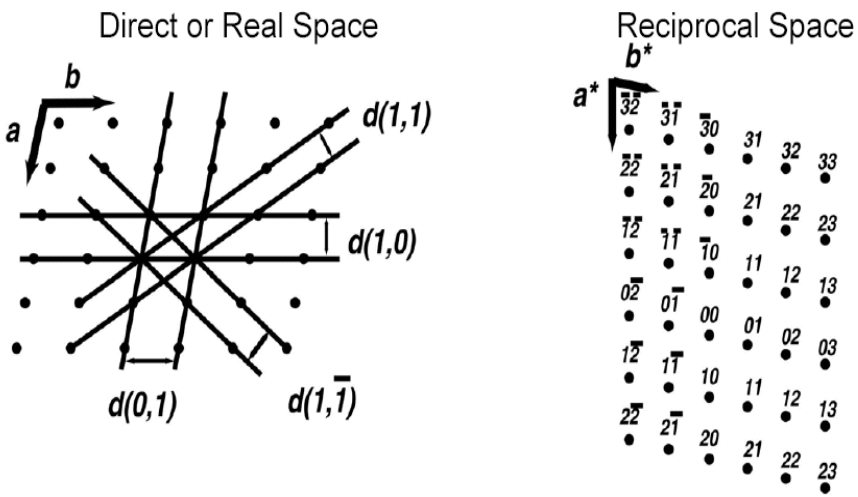
The intensity of X-ray scattered waves from atoms increases with the atomic number and decreases as the scattering angle increase

Diffraction – Bragg Planes



- Diffraction of X-ray by crystals can be viewed as the reflection by families of planes, containing atoms, inside the crystal itself.
- Each family of planes is characterized by 3 numbers (Miller indexes) which identify the Bragg plane orientation relative to the crystal axes.
- The angle between the incident (or scattered) wave and the plane is called Bragg angle.

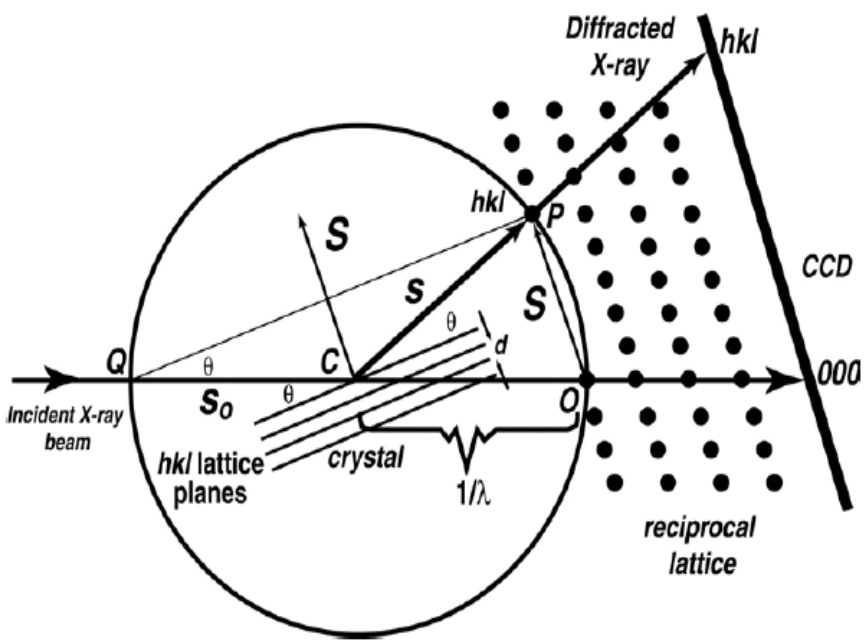
Diffraction – The Reciprocal Lattice



- We can associate at every family of Bragg planes a vector orthogonal to the planes themselves.

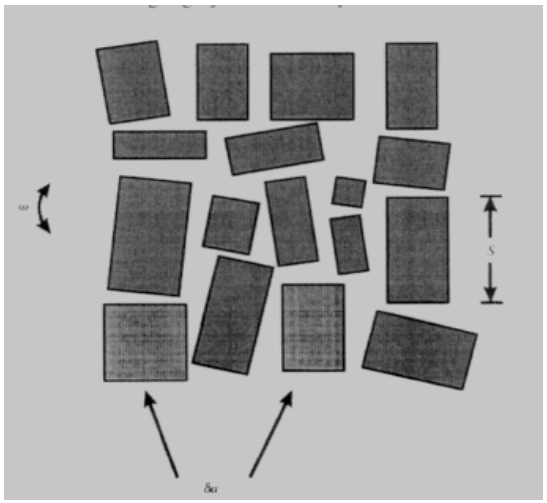
- We can construct in this way a new net called reciprocal lattice, and every discrete point inside it is called reciprocal lattice point (rlp).

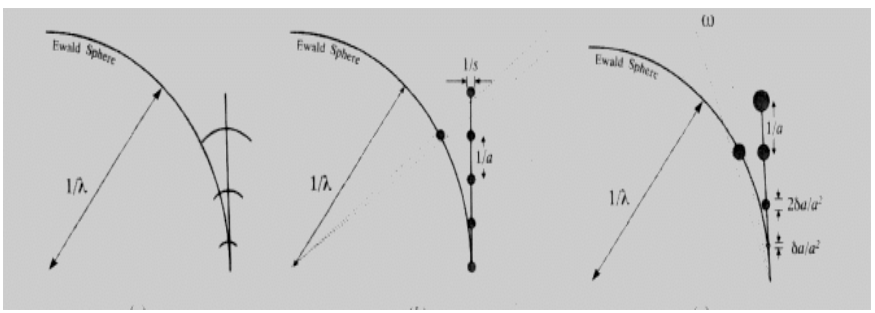
Diffraction – The Ewald Sphere



- A useful tool when considering diffraction by crystals, is the Ewald sphere.
- Every time a reciprocal lattice point lies on the Ewald sphere, that point (or its related Bragg Plane) originates a diffracted ray.

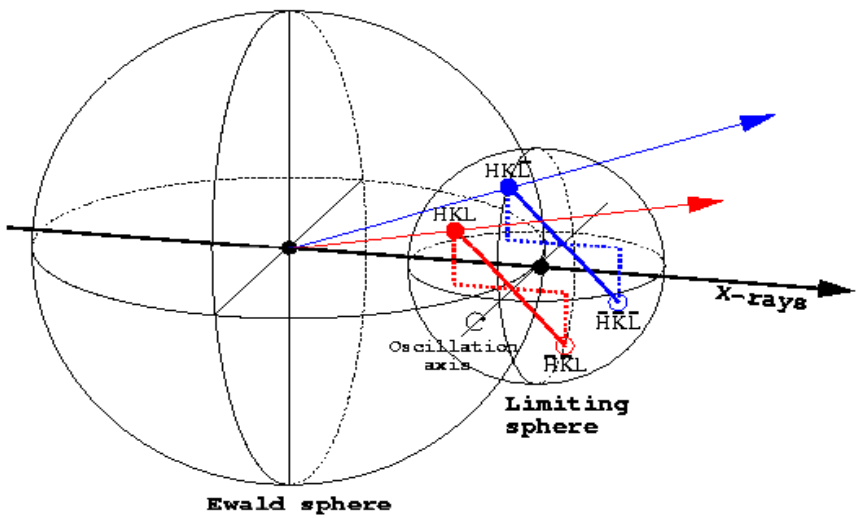
Diffraction – Mosaic Spread





- In principle reciprocal lattice points are zero dimensional, in practice they are not.
- A single crystal can be seen as composed of several blocks, each other misaligned by a more or less small extent.
- The different blocks will be brought in diffraction condition for different but contiguous angles ϕ (while Bragg angle θ does not change), so the diffracted spot will have a distribution (assumed gaussian) around the theoretical position.

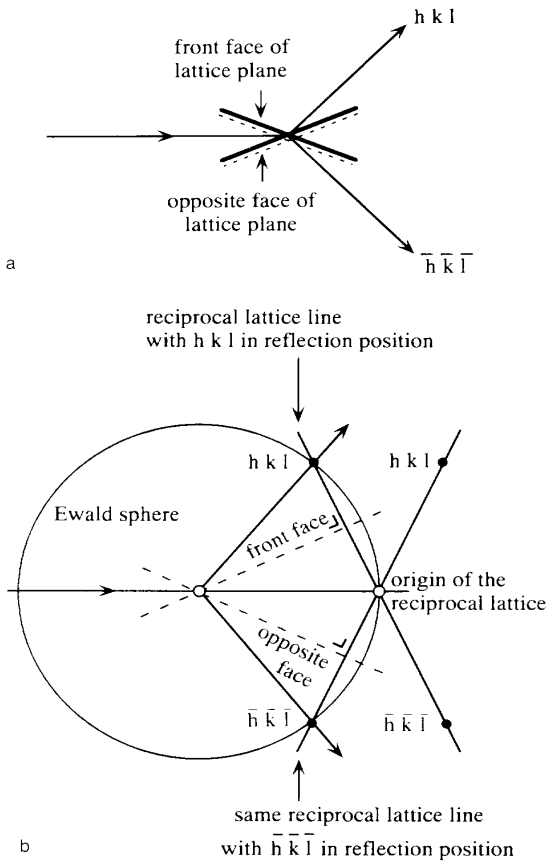
Diffraction – Limiting Sphere



- The intensity of the scattered wave from an atom decreases as the Bragg angle θ increases.
- The internal disorder of the crystal and the thermal motion of the atoms have the same effect on diffraction intensity.
- In practice a maximum angle θ exists above which diffraction is not observed anymore. This angle is related to the level of detail of the final crystal structure (Resolution).

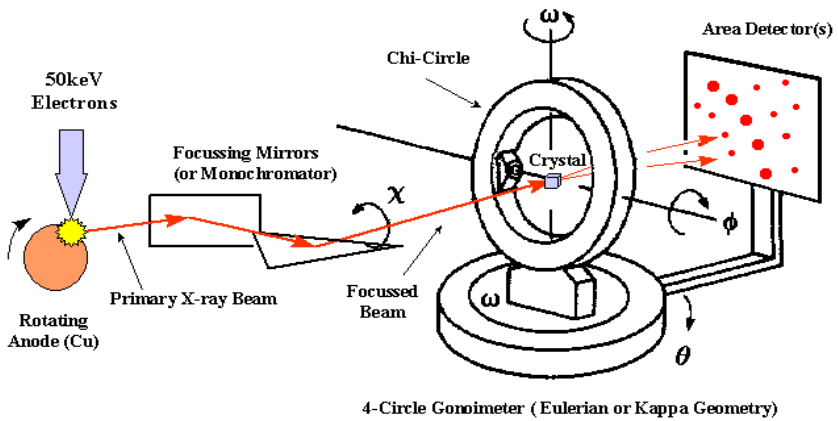
$$\rho(x\ y\ z) = 1/V\ \Sigma_{hkl}\ F(h\ k\ l)\ \exp[-2\pi i(hx+ky+lz) + i\alpha(h\ k\ l)]$$

Diffraction - Friedel Pairs



- The interaction between X-ray and crystals is such that planes related by centrosimmetry, which is planes (h k l) and planes (-h –k –l), have the same diffracted intensity (Friedel law).
- Friedel Law does not hold anymore, if the wavelength of the incident X-ray is near or equal to an absorption edge of an element inside the crystal. This is due to the anomalous scattering contribution.

Data Collection



We collect Intensities of the diffracted waves, but not their phases!

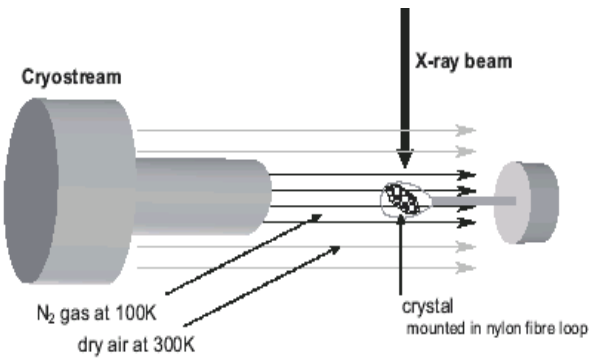
$$I_{hkl}=KI_0\lambda^3LPAV_x|F_{hkl}|^2/V^2_0$$

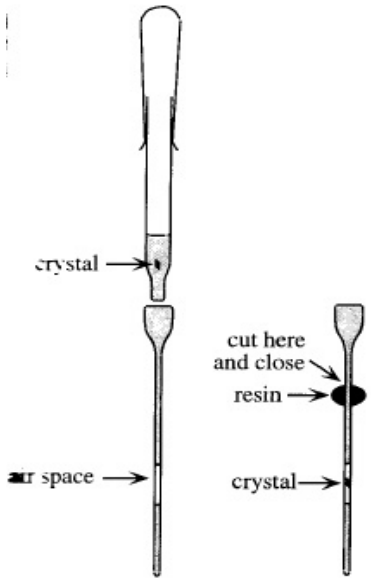
- In a Single Crystal diffraction experiment we have to collect as much as possible diffracted waves, generally indicated as reflections. To do that a general setup consist of:
  - A source of X-rays
  - A goniostat to orientate and rotate the crystal so that a certain number of Bragg planes can be brought in diffraction conditions
  - A detector to acquire the diffracted rays

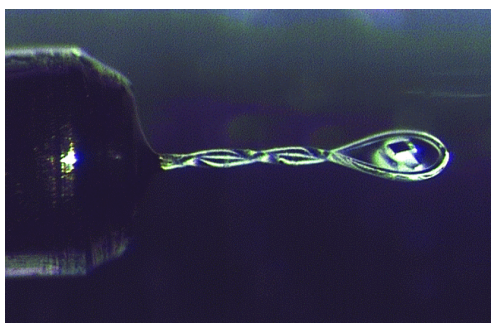
Data Collection – Radiation Damage

- Protein crystals are damaged from the exposure to the X-ray beam
- Radiation damage causes a reduction in lifetime of crystals, which becomes apparent as a reduction of the diffraction pattern maximum resolution, and worsening of the diffracted spot shape.
- At the final electron density map level,disulfide bonds can be broken, and carboxylic groups of exposed aspartic or glutamic acid may be lost.
- The extent of the damage depends on the X-ray dose absorbed, and on the energy of the incoming photons.
- X-rays promote the radiolysis of water molecules, the radicals generated chemically damage the proteins. Migration of radicals can be stopped by freezing protein crystals with an appropriate cryoprotectant.
- X-ray absorption causes the emission of photoelectrons with formation of ions. With an extremely intense X-ray beam (3^{rd} Generation Synchrotron undulators) this effect may become apparent. It is insensitive to low temperature data collection.

Data Collection – Crystal Mounting





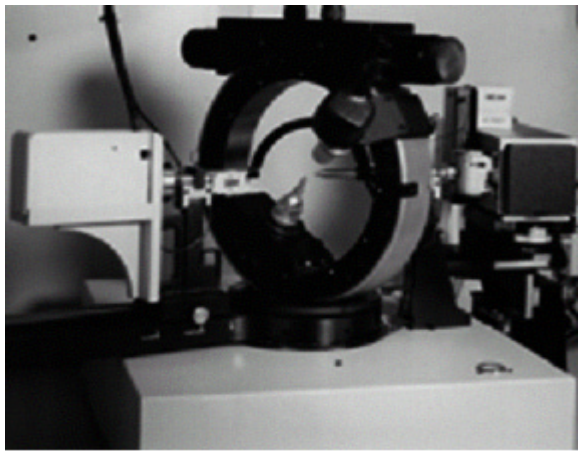


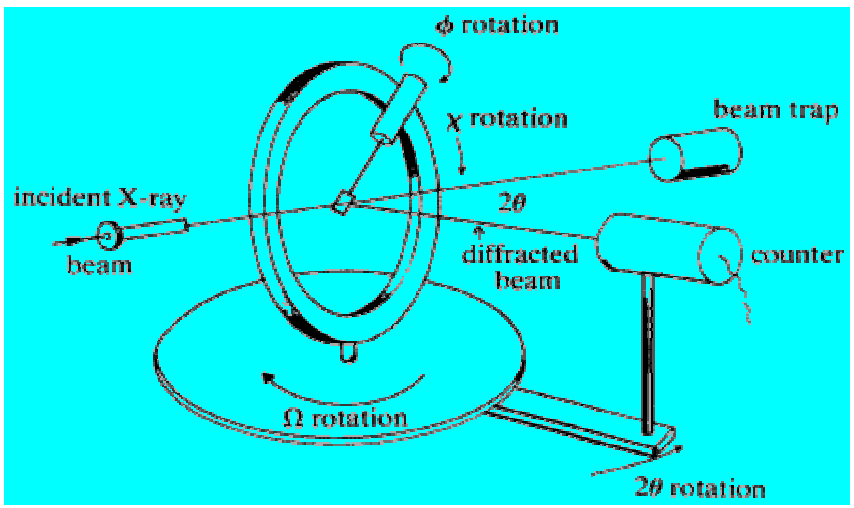
- If crystal decay is not a problem, capillary mounting is an option.
- For Synchrotron data collection, the safest option is to flash freeze crystals soaked with a cryoprotectant.
- Finally crystals are mounted on goniometer head and centered, so that they do not give rise to precession upon spindle axis rotation.

Data Collection - Methods

- Different methods exist for collecting diffraction data, depending on:
  - The incident beam is monochromatic or not
  - The Geometry of diffraction experiment
  - The Detector can be zero, one or bidimensional
- Polychromatic
  - Laue Method
- Monochromatic
  - Single Crystal Diffractometer (with photon counter)
  - Precession Method
  - Rotation Method

Data Collection – Diffractometer

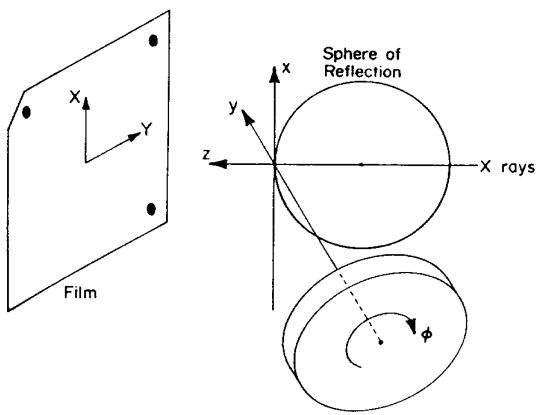




- The Bragg planes are brought into diffraction condition one at time. The diffracted intensity is collected with a zero dimensional detector (photo multiplier)
- Advantage: Data are accurate, semi-empirical absorption correction is possible.
- Disadvantages: Extremely slow! In practice it is not used anymore for Protein crystallography.

Data Collection – Rotation Method

- The crystal is rotated around a generic crystallographic direction so that a certain number of rlps are brought into diffraction conditions.
- Advantages: Fast
- Disadvantages: Distorted Image of rlp planes.
- Rotation Method is the Standard method of data collection for macromolecules.



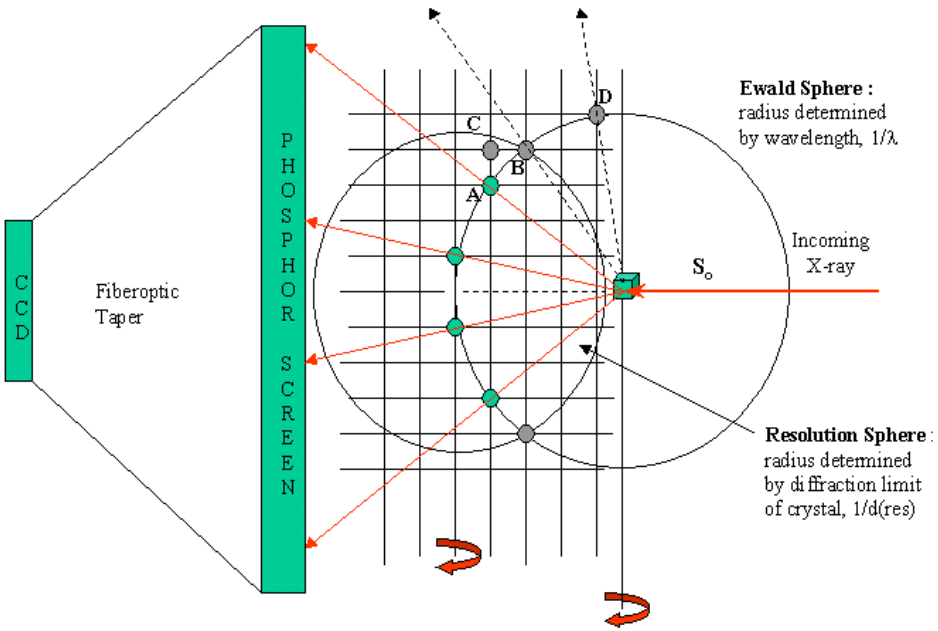
•The simplest setup consist of:

–monochromatic X-ray beam

–single axis goniometer orthogonal to the incident beam

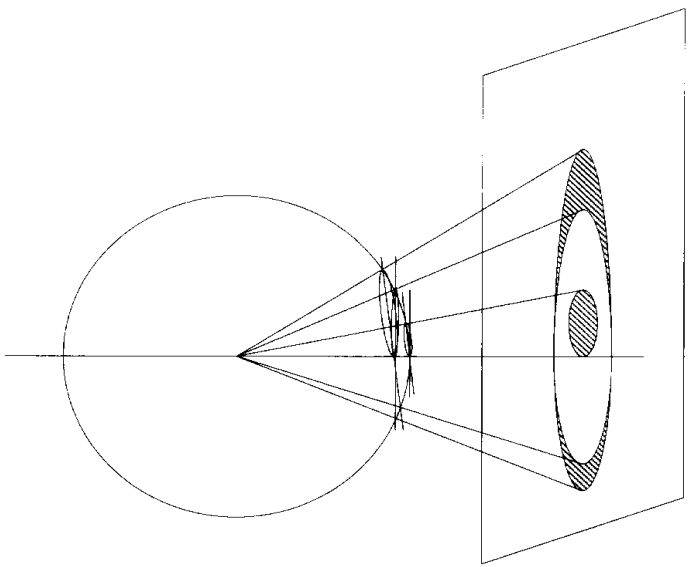
–flat bidimensional detector parallel to the rotation axis. The detector is normally orthogonal to the incident beam, but not necessarily.

Rotation Method - Geometry



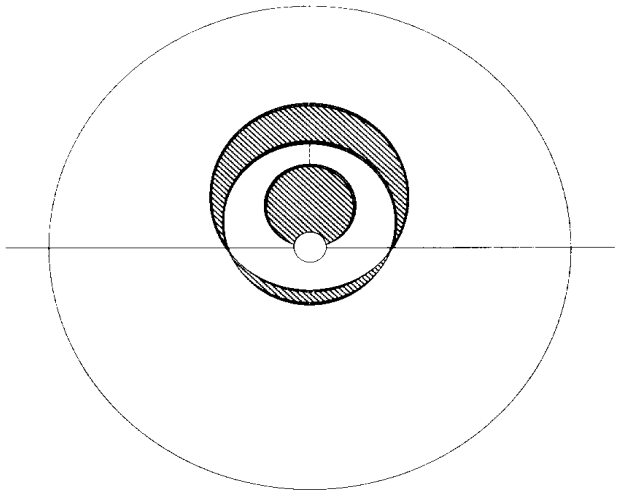
- The crystal is rotated around the goniometer axis of a certain angle so that several rlps will be in diffraction condition during the rotation. The diffracted X-rays are collected from the detector behind the crystal, without any intercepting screen between them.
- The rotation is repeated for contiguous angles until, given the orientation of the crystal, the independent (at least) part of the reciprocal lattice is completely scanned.

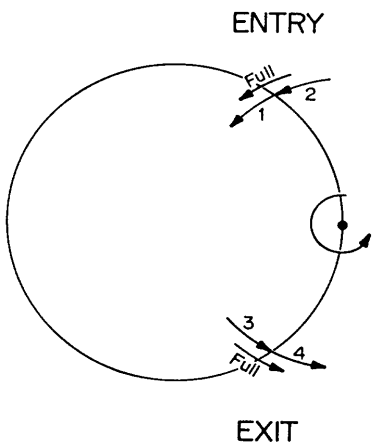
Rotation Method – Diffraction Pattern



- The diffraction pattern shows typical diffraction figures called “lunes”, due to the sweep through the Ewald sphere of the reciprocal lattice, during the crystal rotation
- The rotation angle must be small enough in order to avoid spots superposition. In practice the lunes must be well resolved.

Rotation Method – Partial Reflections





- Due to the finite size of the rlps, it may happens that a diffraction spot is not completely recorded on a single image (partial reflection).
- The number of partial and full reflection on a single image depends on the mosaicity of the crystal, divergence of the X-ray beam, $\delta\lambda/\lambda$ of radiation and the rotation angle for a single image $\Delta\phi$.
- The overall intensity of a reflection distributed among several images is the sum of the partial intensities.

Rotation Method – $\Delta\phi$ Choice

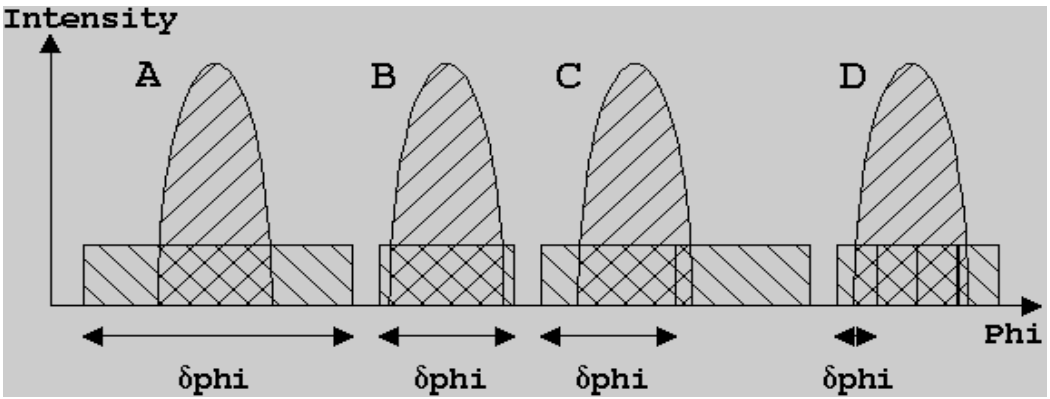
$$\Delta\phi = d/a * 180/\pi - \eta$$

d = maximum resolution (Å)

a = longest real unit cell axis (Å)

- The $\Delta\phi$ must be small enough to avoid the superposition of the spots.
- The $\Delta\phi$ can be roughly estimated on the basis of the unit cell dimension, maximum resolution and mosaicity.
- in order to minimize the total number of exposures, a better estimation of $\Delta\phi$, for different zones of reciprocal lattice, can be determined by appropriate software.

Rotation Method – Fine Slicing



- Using small $\Delta\phi$ (0.1 – 0.2 deg.) may be advantageous because less background is acquired on a single image, and a better guess of the spot profile can be achieved.

On the other hand a huge amount of images is acquired (lot of disk space required), and the acquisition time may be longer due to the detector readout time (problematic with X-ray sensitive

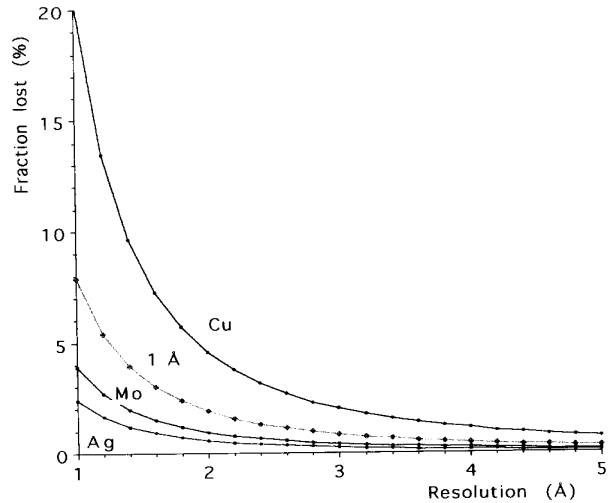
Rotation Method – Total rotation angle

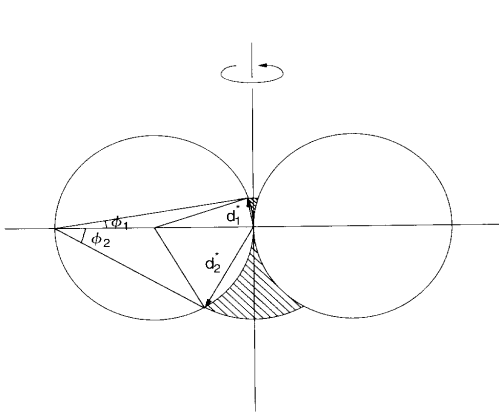
TOTAL ROTATION REQUIRED FOR COMPLETE DATA IN CASE OF SYMMETRIC DETECTOR POSITION

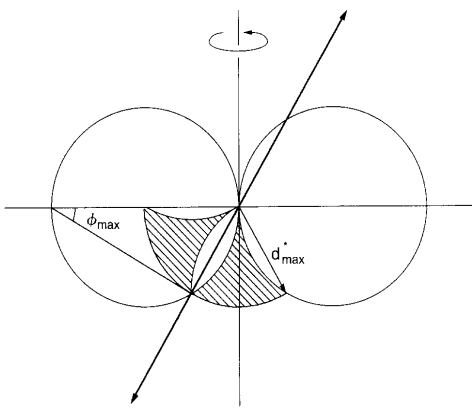
Crystal class	Point group	Rotation required for	
		Standard data	Anomalous data
Triclinic	*1*	180°	360°
Monoclinic	*2*	180° (b^*), 90° (a^*, c^*)	180° (a^*, b^*, c^*)
Orthorhombic	*222*	90° (a^*, b^*, c^*)	90° (a^*, b^*, c^*)
Tetragonal	*4*	90° (a^*, b^*, c^*)	90° (c^*), 180° (a^*, b^*)
	422	45° (c^*), 90° (a^*, b^*)	45° (c^*), 90° (a^*, b^*)
Trigonal	*3*	60° (c^*), 90° (a^*, b^*)	120° (c^*), 180° (a^*, b^*)
	321	30° (c^*), 90° (a^*, b^*)	60° (c^*), 180° (a^*, b^*)
	312	30° (c^*), 90° (a^*, b^*)	60° (c^*), 180° (a^*, b^*)
Hexagonal	*6*	60° (c^*), 90° (a^*, b^*)	60° (c^*), 180° (a^*, b^*)
	622	30° (c^*), 90° (a^*, b^*)	30° (c^*), 90° (a^*, b^*)
Cubic	*23*	About 60°	About 70°
	432	About 35°	About 45°

- The total rotation angle must cover at least the independent part of the reciprocal lattice.
- If Friedel pairs have to be collected in order to exploit the anomalous signal, the total rotation angle could be greater, depending on the crystal symmetry and crystal orientation
- Acquiring redundant data set is better, if the crystal does not suffer for radiation damage. We have a better estimate of the diffracted intensities, this is especially important for a good estimation of the small anomalous differences.

Rotation Method – Blind Region







- Geometry of the data collection clearly shows that exist a set or rlps that will never meet the diffraction condition.

- This is not problematic in general because the set of rlps that will never meet diffraction conditions are collected as symmetry related. If the symmetry axis lies along or near the rotation axis, this situation is not true, in that cases k-goniometer can help.

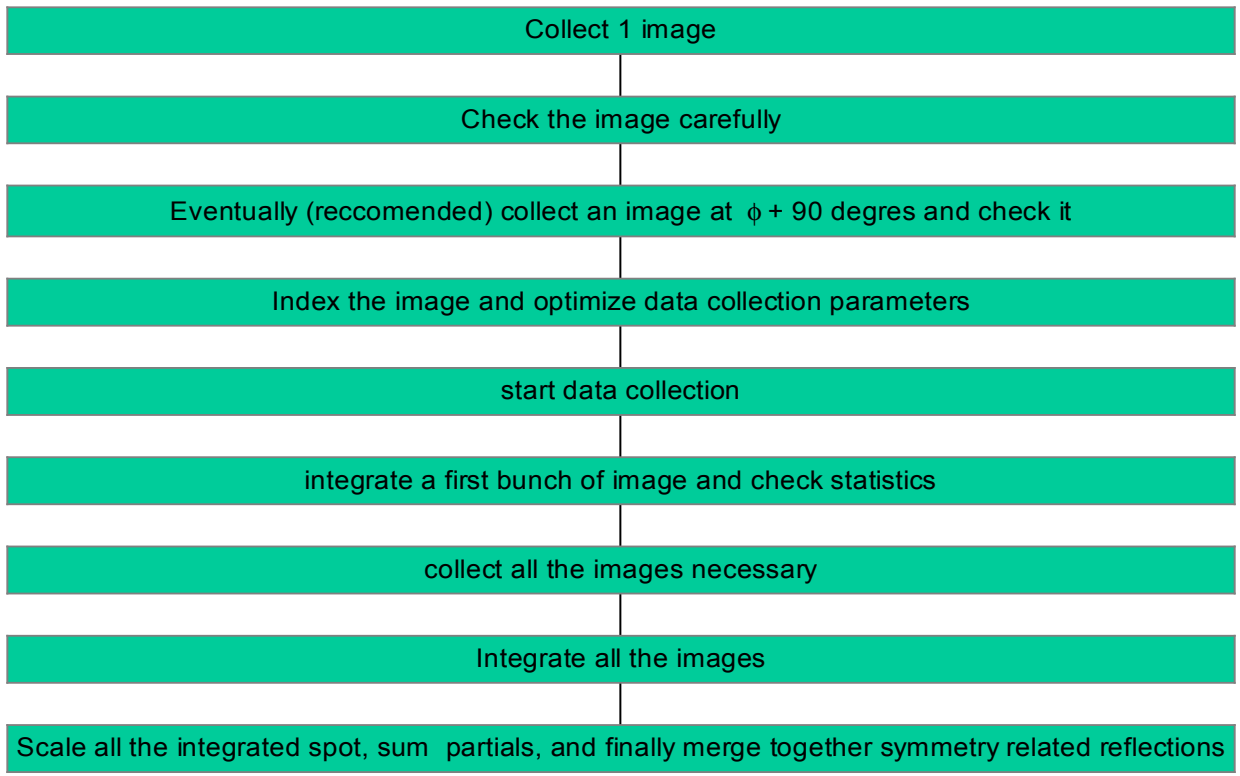
Rotation Method –Exposure Time

- In order to achieve an higher resolution data set, the exposure time could be increased, taking care that a complete dataset is still possible. In any case consider that:
- Long exposure times can result in many spots being saturated
- To double $\langle I/\sigma \rangle$ of the spots the exposure time must be increased by a factor of four
- Long exposure times can be critical for radiation sensitive crystals
- Sometimes it is better to have a shorter exposure time but collect more images. In this case the disadvantages are: 1) many images, 2) more readout noise, 3) worse duty cycle (important with Imaging Plate).

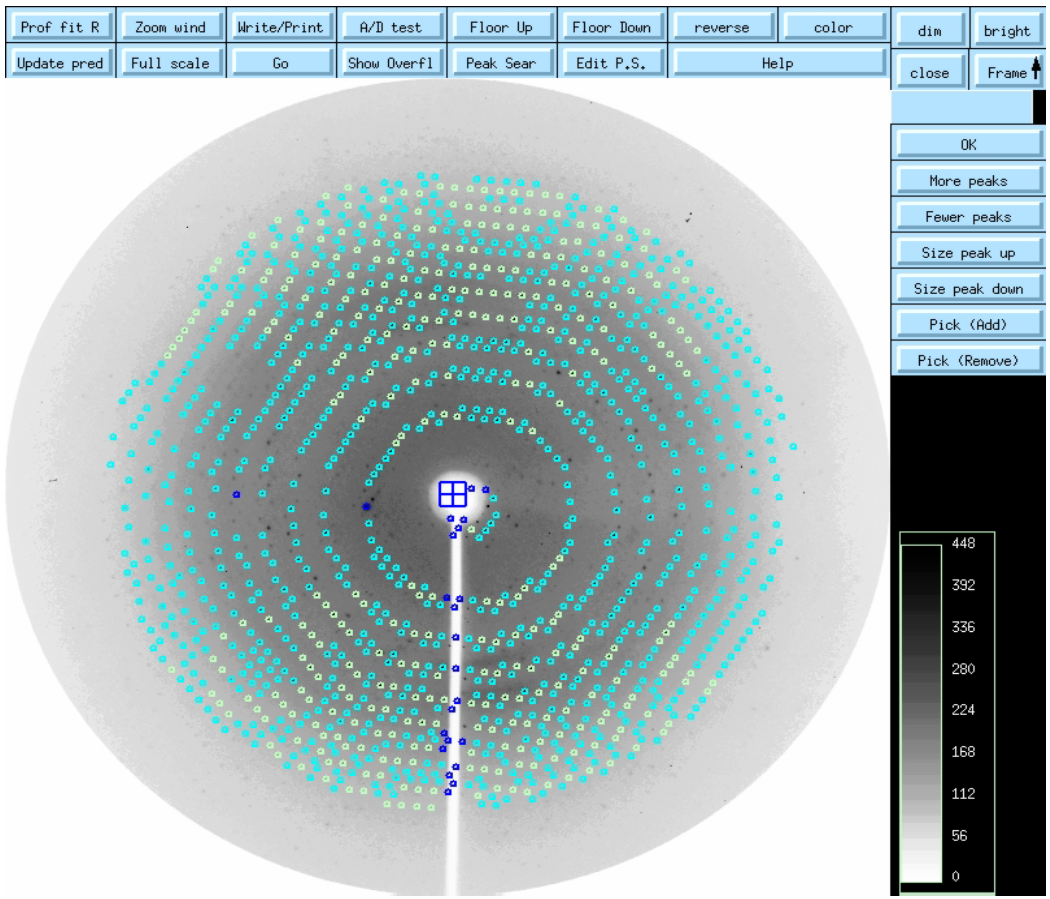
Rotation Method – General Checks

- Spots not well resolved: check $\Delta\phi$ and crystal to film distance
- Spots are split or streaked: check cryo conditions, crystallization conditions or protein homogeneity
- Ice rings: Check cryo conditions
- No spots in the outer region of the detector: Increase (not a large amount) the exposure time if you are confident in your resolution limits, otherwise adjust the crystal to film distance to fill the detector (better $\langle I/\sigma \rangle$ at longer distance)
- High mosaic spread: Try annealing or check cryo conditions or crystallization conditions
- Always check images at 0 and 90 degrees for anisotropic behavior
- Strong background: Try smaller collimation (be careful of alignment of the crystal and/or experimental setup). If possible use an Helium purging path.

Data Collection – General layout

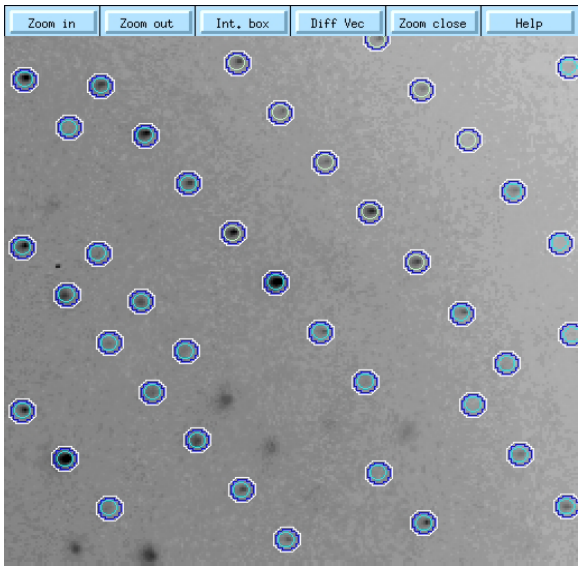


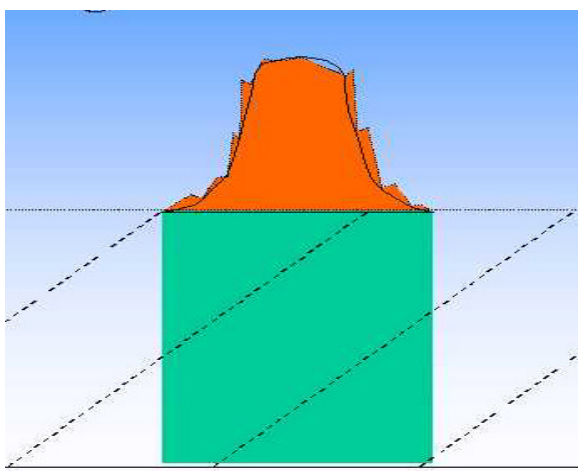
Data Processing - Indexing



- At the indexing stage we determine the lattice symmetry of the crystal and the orientation of the lattice with respect to the laboratory orthogonal coordinate system
- In general this is done finding the highest order lattice symmetry that well predicts as much spots as possible on the diffraction pattern

Data Processing - Integration





- Once the first image has been indexed, on the basis of refined indexing parameters such as unit cell, detector and beam geometry, mosaicity and integration box, the intensities for all the predicted spots is determined; this process is then automatically repeated for all the subsequent images.
- The spot intensity is in general estimated by using a profile function obtained by fitting the intensity profile of strong reflections in the same detector zone.
- Correction factors are applied to intensities (Lorentz and Polarization)

Data Collection – Scaling and Merging

- All the integrated spots from different images are put on a common scale.
- The partial reflections are summed together.
- Better estimate of the unit cell parameters are determined (Postrefinement)
- Symmetry related reflections are merged together giving a final dataset useful for the subsequent structure solution and refinement.

- Several figures of merit are issued at the end of the scaling and merging stage, giving an idea of the goodness of the dataset.

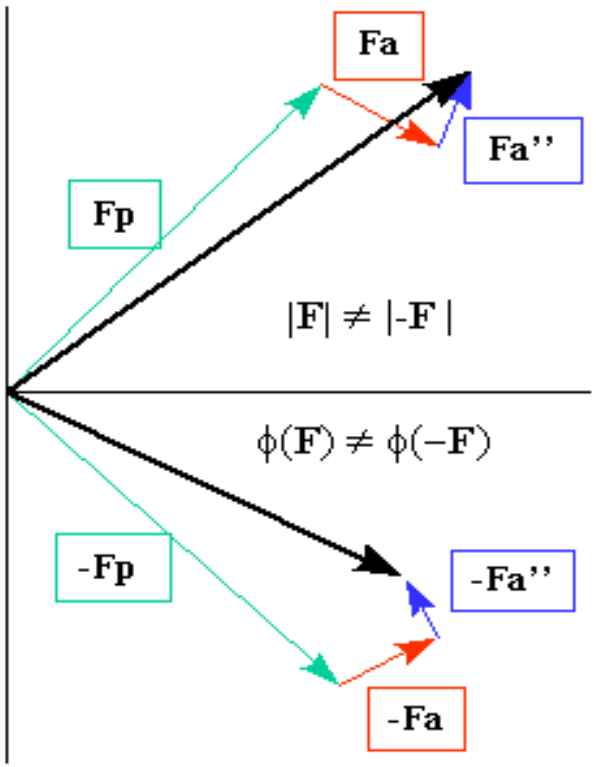
Rotation Method – Quality Indicator

- Completeness: indicates the percentage of reflections experimentally determined with respect to the theoretical ones for that resolution (should be as near as possible to 100 %, check the low resolution completeness.
- $\langle I/\sigma \rangle$: It gives an indication of the signal to noise ratio, and then how strong the diffraction from the crystal is compared to the backround. It is very important in deciding the high resolution limit. A suggested limit for the high resolution shell is $\langle I/\sigma \rangle = 2$.
- Redundancy: Indicates the number of times that the reflections are measured as such or as symmetry related. In general the higher the redundancy is, the better the reflections intensity estimation.
- R_{merge}: One definition is

  $R_{merge} = \Sigma_{hkl} \Sigma_i | I_i(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} | \langle I(hkl) \rangle |$

  In general the lower the better, but strongly depends on the redundancy (other indicators exist that take redundancy into account)

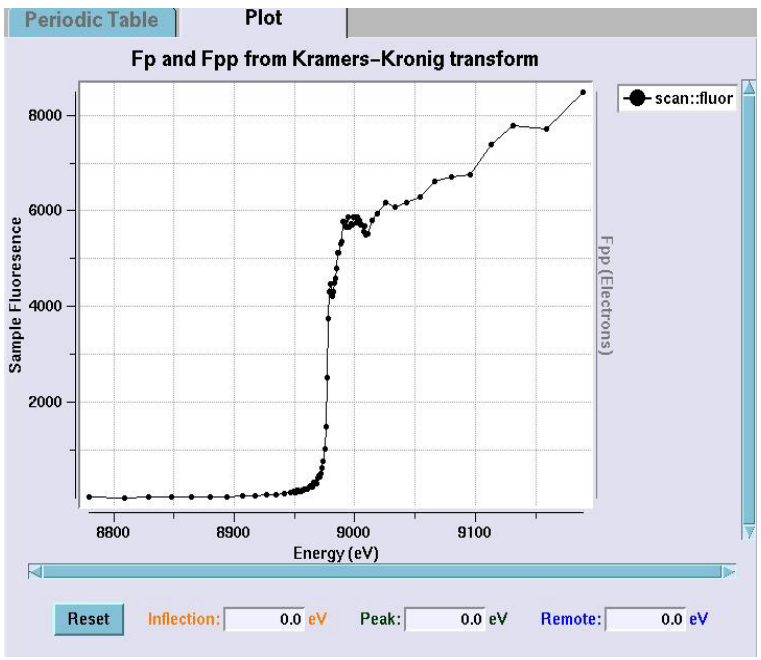
Multiwavelenght Anomalous Dispersion

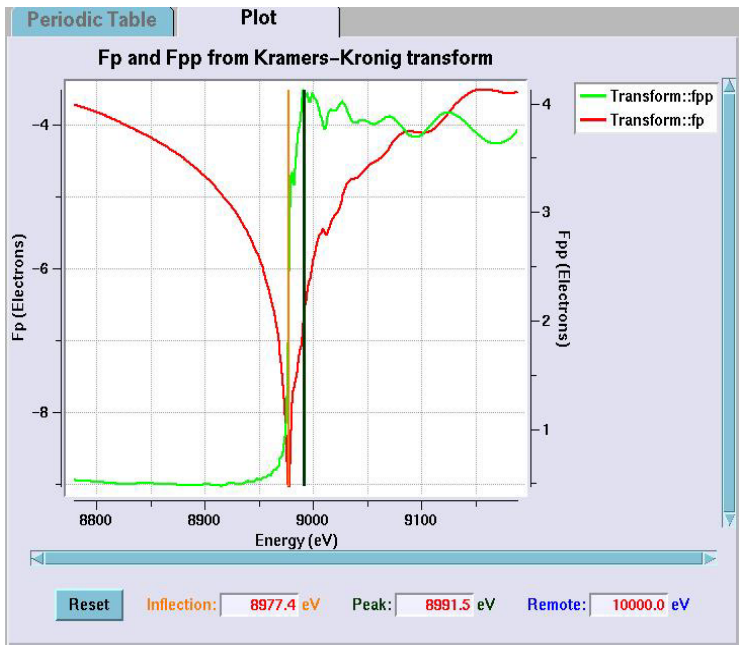


- The MAD (Multiwavelength Anomalous Dispersion) phasing method exploits the abrupt changes in scattering power of heavy atoms (such as transition metals, lanthanides, Se or Br) in the vicinity of absorption edges. The changes in scattering power result in differences between the diffracted intensities measured at different wavelengths as well as differences between reflections of the type (hkl) and (-h-k-l) (anomalous pairs) measured at the same wavelength. These differences can be usefully exploited in order to gain phase informations.

The MAD experiment

- The first step in a MAD experiment, is the acquisition of a fluorescence spectrum from the protein sample, in order to exactly determine the wavelengths of the absorption peak (where *f''* is maximized) and the inflection point (*f'* is minimized). Once the the peak and inflection point are determined, at least three data collection must be carried out (the third one is the so called 'remote', at a wavelength far away from the absorption edge (usually at lower wavelength).





Structure Solution

In order to calculate the electron density inside the crystal, we need the complex structure factors *F(hkl)* but from a diffraction experiment, we know only their magnitude |*Fhkl*|

$$\rho(x\ y\ z) = 1/V\ \Sigma_{hkl}\ F(h\ k\ l)\ \exp[-2\pi i(hx+ky+lz) + i\alpha(h\ k\ l)]$$

Phases can be obtained by means of different alternative techniques:

Experimental Techniques:

Multiple Isomorphous Replacement (MIR, Requires the introduction of heavy atoms in the crystal)

Multiwavelength Anomalous Dispersion (MAD, Requires a tunable source of X-ray and an anomalous scatterer

Computational Techniques

Molecular Replacement (MR, Requires a suitable starting model)

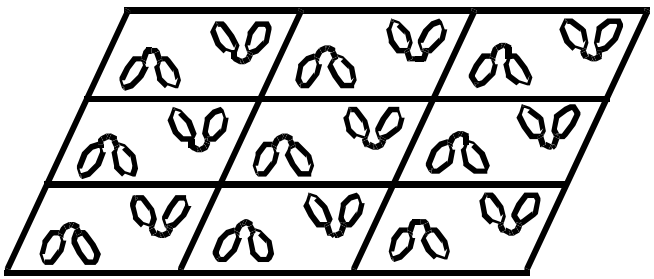
Ab-initio (Requires very high resolution data, few cases solved up to now)

Molecular replacement

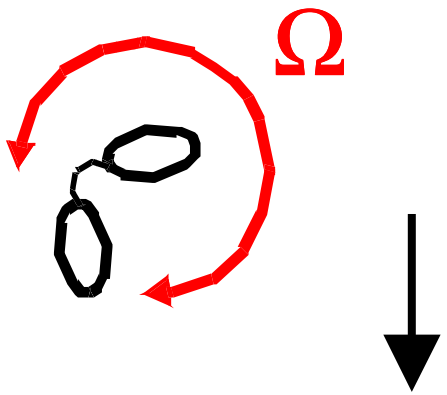
- **Molecular replacement can be used when you have a good model for a reasonably large fraction of the structure in the crystal.**
- To carry out molecular replacement, you need to place the model structure in the correct orientation and position in the unknown unit cell.
- To orient a molecule you need to specify three rotation angles and to place it in the unit cell you need to specify three translational parameters. So if there is one molecule in the asymmetric unit of the crystal, the molecular replacement problem is a 6-dimensional problem.
- It turns out that it is usually possible to separate this into two 3-dimensional problems. A **rotation function** can be computed to find the three rotation angles, and then the oriented model can be placed in the cell with a 3D **translation function**.

Molecular Replacement Stages

Crystal

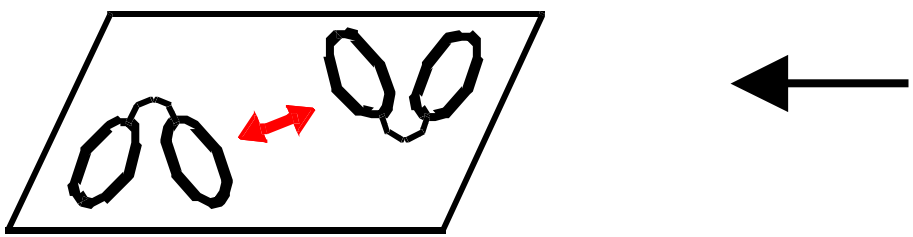


1. Rotation Search



2. PC Refinement

3. Translation Search



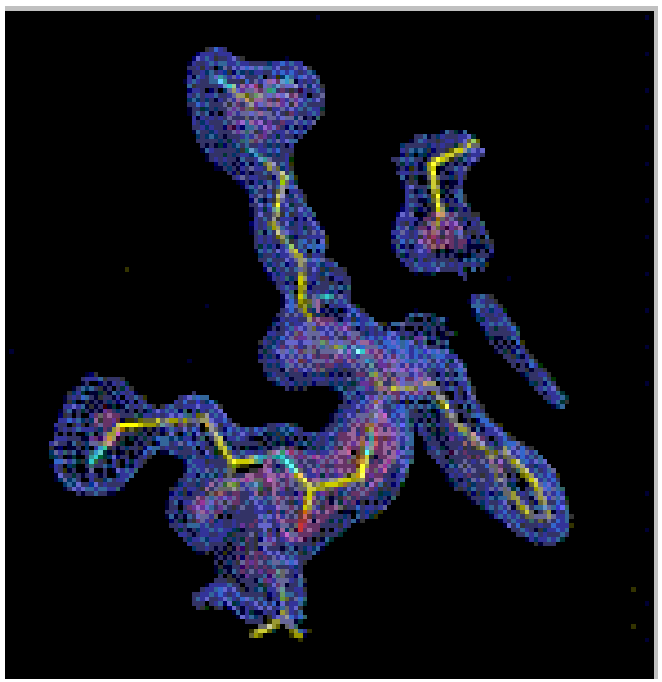
Refinement

- Once the phase problem is solved, it is possible to calculate the electron density inside the crystal and to fit an initial molecular model into it.

- **Refinement** is the process of adjusting the model to find a closer agreement between the calculated and the observed structure factors.

- Refinement technique are based on the principle of least-squares, or maximum likelihood. Molecular model coordinates x, y, z and thermal parameters are changed in order to improve the agreement between the calculated $|F_c|$ and the observed $|F_o|$
- Given the unfavorable ratio between the number of observations and the refined parameters, it is necessary to introduce some geometrical restrains (expected bond length, bond angle and torsional angle).

Refinement –Quality indicator



- The quality of the final model can be established on the base of several quality indicator:
  - Connectivity of the resulting electron density.
  - R-factor/Free R-factor (least square residual for the working and test dataset).
  - Final model geometry.

$$\text{R-factor} = \Sigma_{hkl} \left| |Fo| - k|Fc| \right| / \Sigma_{hkl} |Fo|$$

Bibliography (1)

- Protein Crystallography:
  - Blundell, T. L. ,and Johnson, L. N. (1976) "Protein Crystallography", New York: Academic Press
  - Drenth J. (1999) "Principles of Protein X-Ray Crystallography", Springer: New York.
  - Giacovazzo C, et al. (1992)"Fundamental of Crystallography", Oxford University Press

Bibliography (2)

- Rotation Method:
  - Arndt, U. W. and Wonacott, A. J. (1977) "The Rotation Method in Crystallography", North-Holland
  - Wycoff, H H et al (Editors) - Methods in Enzymology "Diffraction Methods for Protein Crystallography" Vol. 114, Academic Press.
  - Carter, C. W. and Sweet R. M (Editors) - Methods in Enzymology "Macromolecular Crystallography" Vol 276, Academic Press
  - Acta Cryst (1999) D55 "Data Collection And Processing – CCp4 Study Weekend". Many interesting articles.
- Mosaicity:
  - Dobrianov, I. et al: 196 (1999), 511
  - Nave, C: Acta Cryst D54, (1998), 848
  - Helliwell, J. R.: J. Crystal Growth 90 (1988), 259

Bibliography (3)

- Radiation Damage:
  - Henderson, R. : Proc. R. Soc. London (1990), B241, 6
  - Teng, T and Moffatt, K.: J. Synchrotron Rad. (2000), 7, 313
  - Ravelli, R. B. G. and McSweeney, S. M: Structure (2000), 8(3), 315
  - Burmeister, W. P.: Acta Cryst D56, 328

- Synchrotron Radiation and Protein Crystallography:
  - Helliwell, J. R. (1992) "Macromolecular Crystallography with Synchrotron Radiation", Cambridge University Press.

Bibliography (4)

- Detectors for Protein crystallography:
  - See Methods of Enzymolgy 276
  - Lewis R: J Synchrotron Rad. (1994), 1, 43
  - Amemiya, Y: J. Synchrotron Rad. (1995), 2, 13.
  - Gruner S. M. Transactions ACA: vol 34 (1999), 11