

Lecture - 24
X-Ray Crystallography - Phase Problem - Part 2

(Refer Slide Time: 00:43)

X-Ray Crystallography

Phase Problem, Remaining discussion about Heavy atom replacement methods, Phase improvement methods, SIR, MIR, MAD, SIRAS, MIRAS and Molecular Replacement

We are actually at the end of our course document on crystallography. We discussed phase problems in the last class. Today we will again discuss a few parts of the phase problem, the remaining discussions about the heavy atom replacement methods, phase improvement technologies, SIR, MIR, MAD and a little introduction to SIRAS, MIRAS and then we will talk about another technique which is the molecular replacement to solve the phase problem.

(Refer Slide Time: 01:23)

Diffraction Experiments:

Measure the intensities of waves scattered from planes (hkl) within the crystal in order to calculate the electron density at position (xyz) in the unit cell

Perform the summation over all hkl planes

Two separate pieces of information can be found from the diffraction images:
 The first comes from the geometrical arrangement of the reflections, which gives all the information about the crystal lattice and the symmetry of the crystal
 The second comes from the intensity of the reflection that gives part of the information about the content of the lattice
 Unfortunately the second kind, which is the one that we are actually interested in, is only partial - we lack the phases

So, if I go back a bit and start where we discussed a lot, we clone a protein, purify a protein and get the crystal. So, we get the crystal, and we do a diffraction experiment. We measure the intensities of waves scattered from planes within the crystal to calculate the electron density at position xyz in the unit cell.

So, we perform the summation over all hkl planes. What we get from our diffraction experiment is two separate pieces of information. The diffraction image gives us the first information, which is the geometrical arrangement of the reflection gives all the information about the crystal lattice and the symmetry of the crystal. The second information we get from the intensity of the reflection gives part of the information about the content of the lattice.

And as I was discussing throughout the last few lectures in the second one, the intensity information gives us the amplitude, but we are not getting the phase; hence, it is the phase problem.

(Refer Slide Time: 03:10)

Calculation of Electron Density: Fourier Transform

Electron density at (xyz) is the sum of contributions to the point (xyz) of waves scattered from the plane (hkl) whose amplitude depends on the number of electrons in the plane, **added with the correct relative phase relationship**

$$\rho(xyz) = 1/V \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi ihx + ky + lz)$$

Diagram illustrating the Fourier Transform equation for electron density calculation:

- $\rho(xyz)$: Electron Density
- V : Volume of unit cell
- $|F_{hkl}|$: Structure factor amplitude (hkl)
- α_{hkl} : Phase associated with the structure factor
- $\exp(-2\pi ihx + ky + lz)$: Wave scattered from the plane hkl

So, we have to do the calculation of the electron density and we have to perform Fourier transform. So, electron density at xyz is the sum of contribution to the point xyz of waves scattered from the plane hkl whose amplitude depends on the number of electrons in the plane added with the correct relative phase relationship. So, if we again look at the famous formula which we are continually looking at

$$\rho(xyz) = 1/V \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi ihx + ky + lz)$$

$\rho(xyz)$ is the electron density, V is the volume of the unit cell, the absolute value of $|F_{hkl}|$ is the structure factor amplitude, α_{hkl} phase is associated with the structure factor, and the waves scattered from the plain hkl.

(Refer Slide Time: 04:13)

Recovering the Phases:

There is no formal relationship between the amplitudes and phases; only via the molecular structure of electron density

If we can assume some prior knowledge of the electron density or structure, this can lead to values for the phases

Stated another way: phase values **can only be discovered** through some prior knowledge of the structure

So, as we are continuously talking about how to get the phase, the phase problem, and how to recover the phase, there is no formal relationship between the amplitude and the phases but only via the molecular structure of electron density. If we can assume some prior knowledge of the electron density or structure, this can lead to values of the phases. So, phase value can only be discovered through prior structure knowledge.

(Refer Slide Time: 05:10)

Methods for Resolving the Phase Problem:

1. Direct Methods

2. Experimental Phasing

3. Molecular Replacement

- requires some prior knowledge of the crystal structure you want to solve (homologous protein, etc.)

The direct method we discussed details, the experimental phasing using heavy metals we again discussed and molecular replacement which is our topic today. But, we will also talk about some very interesting features related to experimental phasing, especially mathematical ones.

(Refer Slide Time: 05:34)

Direct Method:

Are based on the positivity and atomicity of electron density that leads to phase relationships between the normalized structure factors

Once the phases of some reflections are known, or can be given a variety of starting values, then the phases of other reflections can be bootstrapped

ab initio phase determination are limited to $<1.2\text{\AA}$

Are used in Shake-and-Bake, SHELXD, SHARP to find the heavy atom substructure

Direct method are based on positivity and atomicity of electron density that leads to phase relationships between the normalized structure factor. So, they try to extract data from the intensity, which gives the amplitude the structure factor but from that intensity that diffraction pattern, they want to get more information once the phases of some reflections are known.

Which are some reflections, some reflections where the intensities are very high or can be given a variety of starting values, then the phases of other reflections can be bootstrapped. But this *ab initio* phase determination is limited to less than 1.2 Angstrom resolution. So, lowering the number means you know that the higher the resolution.

But you also remember that the direct method would also help you in protein structure determination. So, they are used in Shake and Bake, SHELXD, SHARP to find the heavy atom substructure. As I told, when the intensity is very high, it is possible to correlate. So, for heavy atoms, we first know the atom and the chemistry, which are very different from the protein (protein contains carbon, nitrogen, and oxygen).

(Refer Slide Time: 07:17)

Introducing a new (heavy) atom into a crystal:

(protein) Crystal

(protein) Crystal
+
(heavy) atom

Get a different structure factor, F_{ph} $F_{ph} = F_p + F_h$

So, we could have got the identity of their information. The protein, we now introduce a heavy atom. So to calculate the F_p , the structure factor of the protein, here with the heavy atom, we get F_{ph} , and we know that

$$F_{ph} = F_p + F_h$$

(Refer Slide Time: 08:07)

Do Heavy atom makes an impact?

How can e.g. a single Cu atom in a 100 kDa protein make any difference to the intensities?
Structure factors add up as a "random walk"

Crick and Magdoff (1956)

$$\frac{\langle \Delta F \rangle}{\langle |F| \rangle} = \frac{Z_H}{Z_{eff}} \sqrt{\frac{2N_H}{N_p}}$$

N_H, N_p are number of heavy, protein atoms
 Z_H, Z_{eff} are atomic numbers of heavy atom and average of protein atoms
 (~6.7)

100 kDa protein with Copper (Z=28) 5.6%

So, before going into details, one thing you should know. Does heavy atom make an impact? So, let us look at how a single copper atom in 100 KDa protein can make any difference to the intensities. So, structure factors add up as a random walk. Crick and Magdoff in 1956 have calculated this,

$$\langle \Delta F \rangle / \langle |F| \rangle = Z_H / Z_{eff} \sqrt{2 N_H / N_p}$$

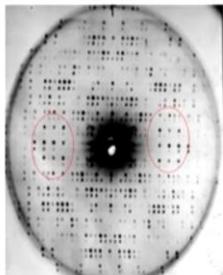
N_H is the number of the heavy atom, and N_p is the number of protein atoms. Z_H is the atomic number of the heavy atom, and Z_{eff} is the average of the protein atom, which is 6.7. This is the change the protein has its vector, and then the copper adds with the structure factor. So, in 100 kDa protein, copper with atomic number 28 adds 5.6%. That is the impact. Now, depending on this, you could choose the heavy metal, understand the concentration, and optimize the concentration of the heavy metal.

Why did you need to optimize this? A protein is largely made up of small hydrogen bonds. When a big metal with its high positive charge and high polarizability suddenly comes inside the protein, it is like a bombardment on the weak hydrogen bond network of the protein. A lot of time, they are deleterious. They break the crystal, and sometimes the protein starts changing shapes, and all those problems are there. Also, that is very much concentration dependent.

(Refer Slide Time: 12:03)

Isomorphous Replacement:

Soaking protein crystals in heavy-atom solutions to create isomorphous HA derivatives
Gives measurable intensity changes which could be used to deduce the positions of heavy atoms



2 Protein Diffraction Patterns superimposed and shifted vertically relative to one another

The slide features a central image of two protein diffraction patterns. The patterns consist of a grid of spots (reflections) arranged in a roughly circular pattern around a central dark spot. The two patterns are superimposed on each other, with one pattern shifted vertically relative to the other. Red circles are drawn around specific spots in both patterns to highlight the differences in intensity and position between the native protein and the heavy-atom derivative. The slide has a blue header and footer with some small logos.

We do the soaking up protein, where we have a metal solution, we put the crystal there, or we do co-crystallization, where we add precipitants and the metal at the time of crystallization. Soaking protein crystals in heavy atoms solution to create isomorphous heavy atom derivatives gives measurable intensity changes, which could be used to deduce the position of the heavy atoms.

Two protein diffraction patterns. One is the native protein, and another is the mercury-added protein. Superimposed, we could see the shift of the peaks particularly related to one another, and we could also see the change of intensity very easily. So, this tells that by adding heavy metal there are changes.

(Refer Slide Time: 13:45)

Single Isomorphous Replacement (SIR):

The isomorphous difference, can be used as an estimate of the HA structure factor amplitude to determine the HA positions using Patterson or direct methods:

$$|F_H| \simeq |F_{PH}| - |F_P|$$

HA parameters (xyz positions, occupancies and Debye-Waller thermal factors(B)) can be refined to calculate a more accurate $absF_H$ and its corresponding phase α_H

Using Patterson or direct methods, the isomorphous difference can be used to estimate the heavy atoms structure factor amplitude to determine the heavy atom position.

$$|F_H| = |F_{PH}| - |F_P|$$

Heavy atom parameters, such as xyz positions, occupancies and Debye-Waller thermal factor (B factor), can be defined to calculate a more accurate, absolute F_H the heavy atom structure factor and its corresponding phase α_H .

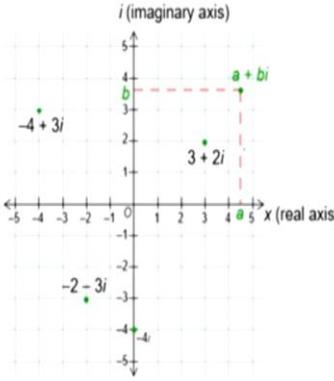
(Refer Slide Time: 14:30)

Argand Diagram:

An Argand Diagram is a plot of complex numbers as points

The complex number $z = x + yi$ is plotted as the point (x, y), where the real part is plotted in the horizontal axis (x axis) and the imaginary part is plotted in the vertical axis (y axis)

If you look at the figure here it would be clear to you how complex numbers can be plotted on an Argand Diagram

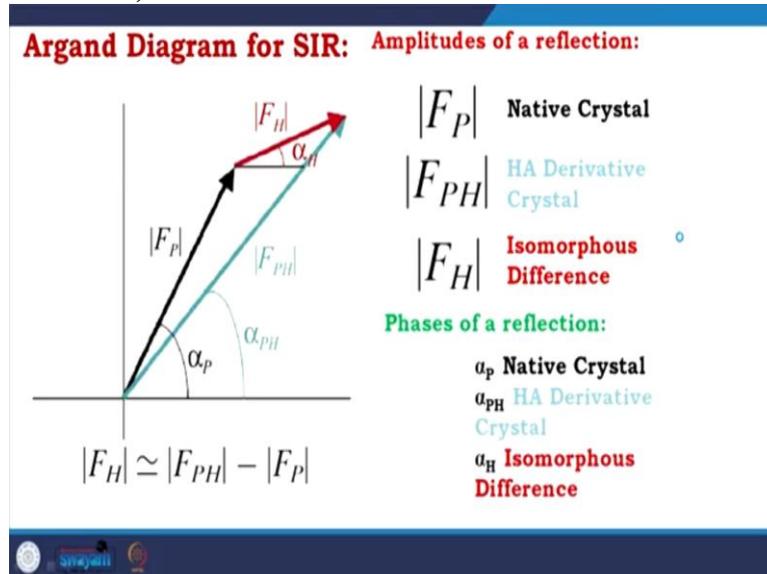


We are going to put an argument diagram and show how it is important in structure solution. The argand diagram in mathematics is a plot of complex numbers as points. The complex number, let us say

$$z = x + yi$$

is plotted as the point x y, where the real part is plotted on the horizontal axis, the x-axis, and the imaginary part is plotted on the vertical axis, the y-axis.

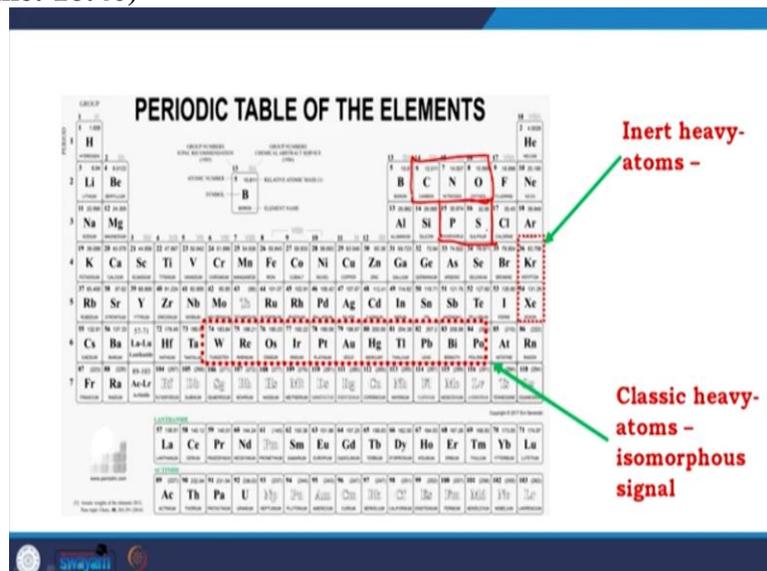
(Refer Slide Time: 15:37)



Now come to the Argand diagram of single isomorphous replacement. You will see many factors here, the amplitude of reflection $|F_P|$, which is the structure factor of the protein, $|F_{PH}|$, which is the absolute value of the structure factor of heavy metal-derived protein. And $|F_H|$ is the difference which is heavy metals contribution.

Similarly, the phases, α_p comes from the native crystal, α_{pH} heavy atom derivative crystal, and α_H isomorphous difference.

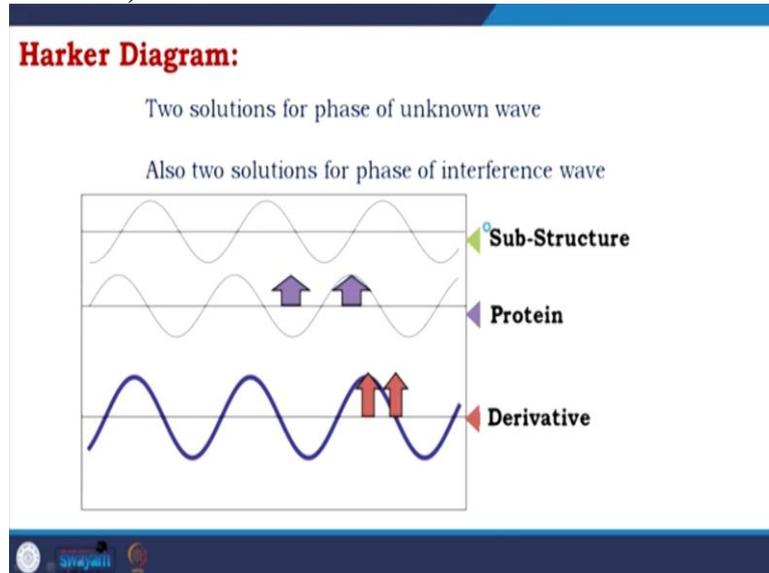
(Refer Slide Time: 16:40)



So, suppose you see the periodic table. In that case, the heavy atoms which you could choose are all in this row, which are tungsten, rhenium, osmium, iridium, platinum, gold, mercury,

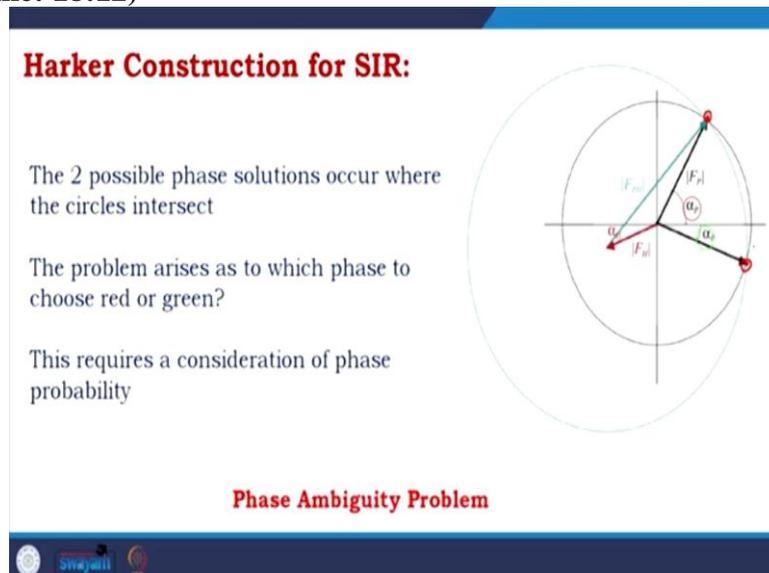
thallium, lead, bismuth, polonium. In addition, we could also use krypton and xenon which are not heavy atoms. So, you can easily see that these are much higher in electron number than carbon, nitrogen, and oxygen, which are present in normal proteins. I also add phosphorus and sulfur because sulfur is present in cysteine, methionine, and phosphorus in phosphor-modified proteins.

(Refer Slide Time: 17:48)



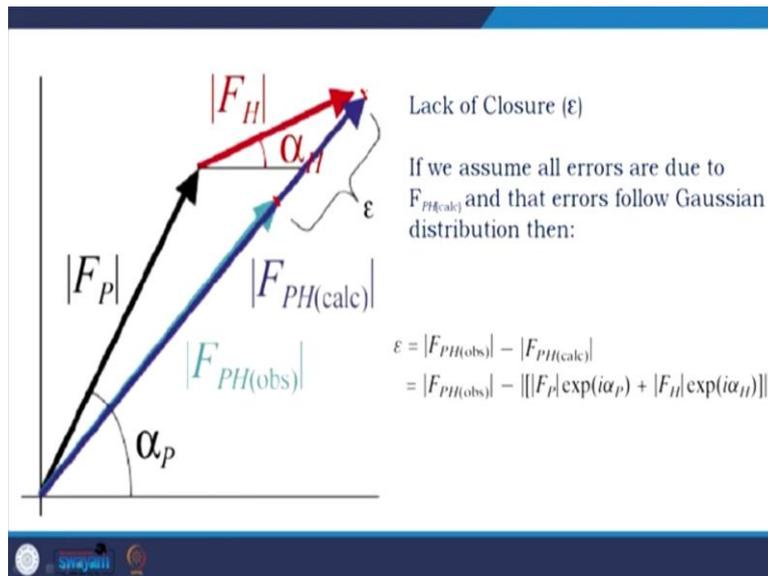
Coming to another geometrical representation called the Harker diagram, the 2 solutions for the phase of an unknown wave where meet.

(Refer Slide Time: 18:12)



Harker construction for a single isomorphous replacement that 2 possible phase solutions occur where the circle intersects. This requires a consideration of phase probability, called the phase ambiguity problem in crystallography.

(Refer Slide Time: 18:54)

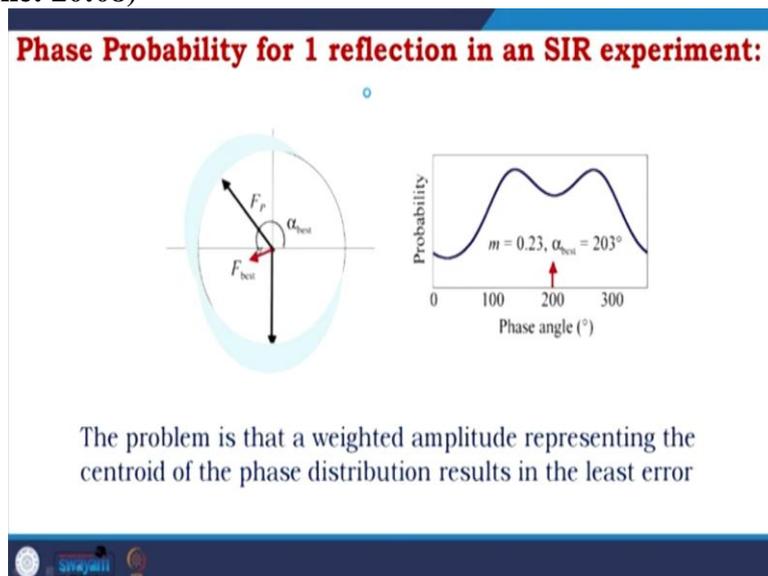


Now, we consider ϵ as the lack of closure, then the lack of closure

$$\epsilon = \text{equal to } |F_{PH(obs)}| - |F_{PHcalc}|,$$

$$= |F_{PH(obs)}| - [|F_P| \exp(i\alpha_p) + |F_H| \exp(i\alpha_H)]$$

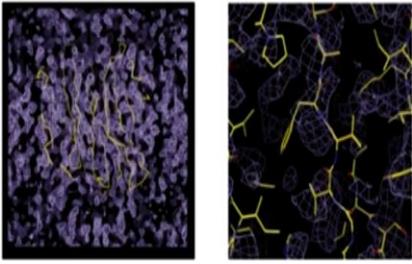
(Refer Slide Time: 20:08)



The phase probability for one reflection in an SIR experiment, the problem is that a weighted amplitude representing the centroid of the phase distribution result in the least error.

(Refer Slide Time: 20:30)

Crystals → diffraction data → Amplitudes
ED map ← (SIR) ← phase



A 2.6 Ang SIR electron density map with the final alpha-carbon trace of the structure are superimposed

$$\rho(\mathbf{x}) = (1/V) \sum m |F_p| \exp(i\alpha_{\text{best}}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

A small section of the map with the final structure superimposed

A 2.6 angstrom single isomorphous replacement electron density map with the final alpha carbon trace of the structure superimposed. So, we had crystals, we got the diffraction data, we tried to solve, it we got amplitude but we did not get the phase. Now we are putting single isomorphous replacement and we get the ED map, this is a 2.6 angstrom resolution density map. But and this is a small part of that where the final structure is superimposed. So, here you see the alpha carbon trace, here in a zoomed one, you see the side chains included here.

(Refer Slide Time: 21:49)

Multiple Isomorphous Replacement (MIR):

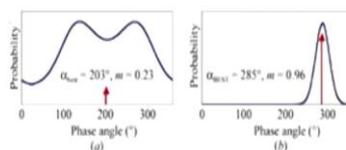
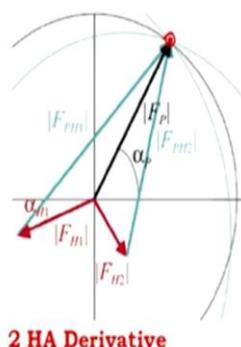
Phase Probability using MIR (more than one HA) to break the phase ambiguity

Multiple isomorphous replacement, in SIR we are using one heavy atom derivative, here we are actually using more than one heavy atom to break the phase ambiguity.

(Refer Slide Time: 22:11)

Multiple Isomorphous Replacement (MIR):

Phase Probability using MIR (more than one HA) to break the phase ambiguity



1 HA Derivative 3 HA Derivative

Instead of one, we use two, there is no ambiguity. If you see that, if we compare two heavy atom derivative, this is a one heavy item derivative, you get that confused two peak, and the 3 heavy atom (3 HA) derivative you get a very sharp peak. So, the problem of phase ambiguity is solved.

(Refer Slide Time: 22:48)

Phase Improvement:

Experimental Phases are the starting point

Improvement based on some prior knowledge

Methods Include:

- Solvent Flattening
- Histogram Matching
- Non-Crystallographic Symmetry (NCS) averaging

We could improve the phase: Experimental phases are the starting point. So, we did the heavy atom replacement and get the experimental phase. The improvement based on some prior knowledge. The methods are solvent flattening, Histogram matching and non crystallographic symmetry(NCS), and averaging.

(Refer Slide Time: 23:31)

Solvent Flattening:

Sets values of electron density in the solvent regions to a typical value of 0.33 e\AA^{-3} , in contrast to typical protein electron density of 0.43 e\AA^{-3} .

Sets the values of electron density in the solvent region to a typical value which is given here in contrast to the typical protein electron density. So actually we are differentiating between a protein and a solvent, so that the solvent could develop its phase better.

(Refer Slide Time: 23:31)

Histogram Matching: histogram matching is the transformation of an image so that its histogram matches a specified histogram

electron density distribution from refined atomic coordinates

electron density distribution from isomorphous replacement phases

Alters the values of electron density points to concur with an expected distribution of values

Histogram matching is the transformation of an image, so that its histogram matches a specified histogram. So, if you look at, we are talking about two peaks, which are plotted against electron density value (in the x axis) and probability of observing electron density value. We get two peaks, electron density distribution from refined atomic coordinates and electron density distribution from isomorphous replacement phases. It alters the values of electron density points to concur with an expected distribution of values.

(Refer Slide Time: 25:35)

Non Crystallographic Symmetry (NCS):

A symmetry operation that is not compatible with the periodicity of a crystal pattern (in two or three dimensions) is called a **non crystallographic symmetry**

In protein crystallography, the term 'non crystallographic symmetry' is often, but improperly, used to indicate a symmetry relationship between similar subunits within the crystallographic asymmetric unit

This use comes from the fact that the operation required to superimpose one subunit on another is similar to a space group operation, but it operates only over a local volume, and the superposition may be inexact because the subunits are in different environments

The 'subunit' can be a molecular aggregate, a single molecule, a monomer unit of an oligomeric molecule, or a fragment of a molecule.

A symmetry operation that is not compatible with the periodicity of a crystal pattern (in 2 or 3 dimension) is called a non crystallographic symmetry.

In protein crystallography, the term non crystallographic symmetries is often, but importantly used to indicate a symmetry relationship between similar sub units within that crystallographic asymmetric unit.

This use comes from the fact that, the operation required to super impose one sub unit on another is similar to a space group operation, but it operates only over a local volume. And the superposition may be inexact because the sub units are in different environments. The subunit can be a molecular aggregate, a single molecule, a monomer unit of an oligomeric molecule or a fragment of a molecule.

(Refer Slide Time: 27:38)

Non Crystallographic Symmetry (NCS):

The superposition is inexact because protein subunits in different environments are never identical

At the very least, surface side chains are differently ordered, and solvation is different because of different interactions with adjacent subunits.

This use of the term 'noncrystallographic symmetry' is improper for two reasons:

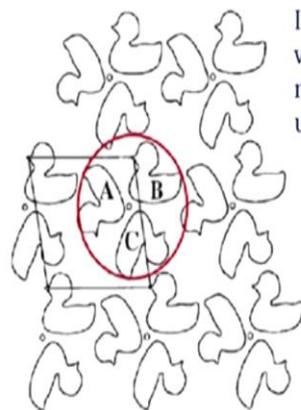
1. A symmetry operation acting on a subspace of the crystal space is called a local or partial symmetry operation; it is a space groupoid operation;
2. An operation that superposes two objects only approximately is called a pseudo symmetry operation.

The superposition is inexact because proteins subunit in different environments are never identical. At the very least, surface side chains are differently ordered and solvation is different because of different interaction with adjacent subunits.

The use of the term non crystallographic symmetry is improper for two reasons. One is, a symmetry operation acting on a subspace of a crystal space is called local or partial symmetry operation, it is a space groupoid operation. Second, an operation that superposes 2 objects only approximately is called a pseudo symmetry operation.

(Refer Slide Time: 29:21)

Non-Crystallographic Symmetry (NCS) Averaging:



Imposes equivalence on electron density values when more than one copy of a molecule is present in the asymmetric unit

$$\rho(x_B) = R_B \rho(x_A) + t_B$$

$$\rho(x_C) = R_C \rho(x_A) + t_C$$

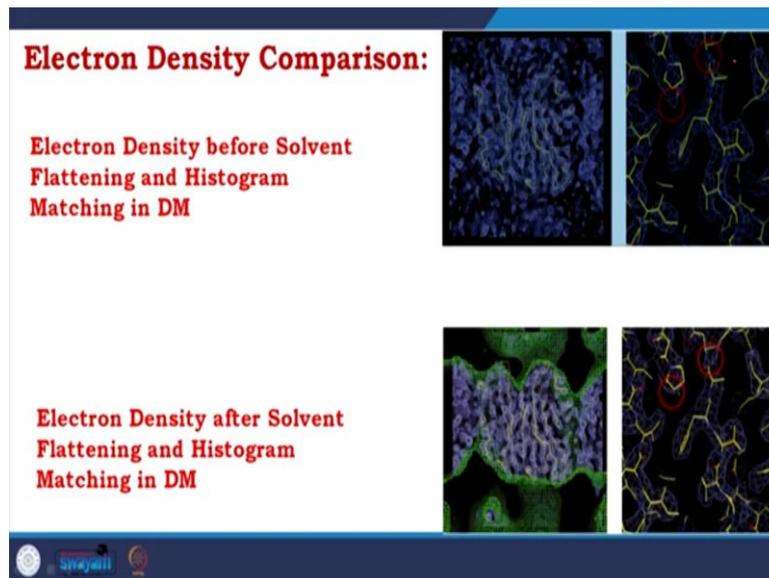
NCS Symmetry Operators

You see A, B and C and they imposes equivalence on electron density values when more than one copy of a molecule is present in the symmetric unit.

$$\rho(X_B) = R_B \rho(X_A) + t_B$$

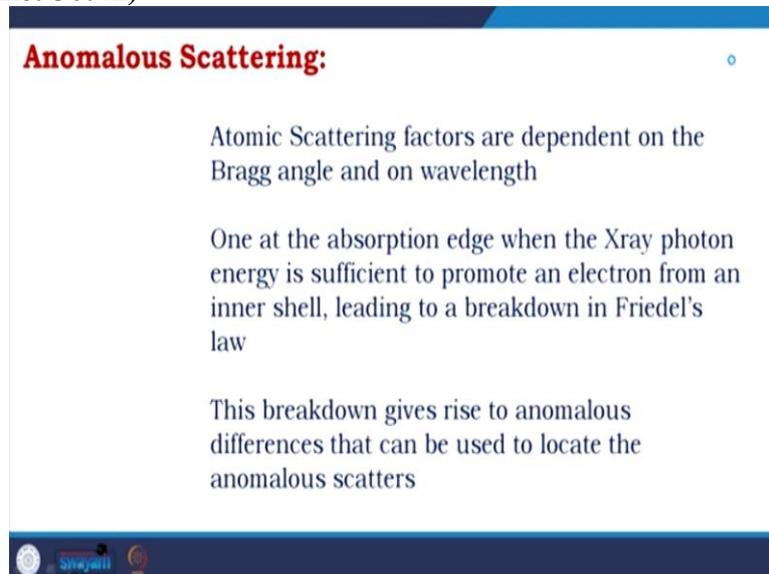
$$\rho(X_C) = R_C \rho(X_A) + t_C$$

(Refer Slide Time: 29:51)



We have done all those solvent flattening, histogram matching and all you see there is a beautiful change, the maps are sharper the phase is definitely improved and that is where the critical role of these methods solvent flattening, histogram matching and averaging of non crystallographic symmetry.

(Refer Slide Time: 30:41)



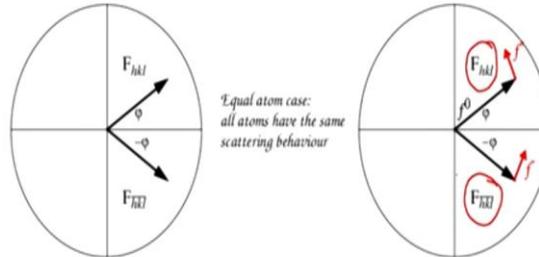
Atomic scattering factors are dependent on the Bragg angle and on wavelength. One at the absorption edge when the X ray photon energy is sufficient to promote an electron from an inner cell, leading to a break down in Friedels law. This breakdown gives rise to anomalous differences that can be used to locate the anomalous scatterers.

(Refer Slide Time: 31:17)

Friedel's Law:

Members of a Friedel pair have equal amplitude and opposite phase:

If all atoms scatter equally, then the amplitudes remain equal but the phase relationship no longer holds. This is because the term is always positive.



The members of a Friedel pair have equal amplitude and opposite phase. If all atoms scatter equally then the amplitudes remain equal but the phase relationship no longer holds. This is because the term is always positive. So, if you compare an equal atom case, all atoms have the same scattering behavior. So, when you compare between F_{hkl} and F_{hkl}^- you would not get, the difference.

(Refer Slide Time: 32:02)

When an anomalous scatterer is present:
And the Bijvoet Difference:
 $f(\theta, \lambda) = f_0(\theta) + f'(\lambda) + if''(\lambda)$
 $F_{hkl} \neq F_{hkl}^-$

Breakdown of Friedel's Law
 $\Delta F^\pm = |F_{PH}(+)| - |F_{PH}(-)|$

When an anomalous scatterer is present there is by void differences,

$$f(\theta, \lambda) = f_0(\theta) + f'(\lambda) + if''(\lambda)$$

$$F_{hkl} \neq F_{hkl}^-$$

And the breakdown of Friedels law is

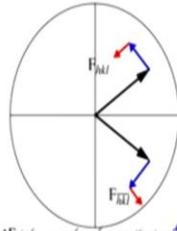
$$\Delta F^\pm = |F_{PH}(+)| - |F_{PH}(-)|$$

(Refer Slide Time: 32:39)

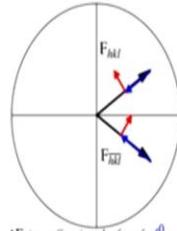
Breaking Friedel's Law:

The contribution of f_0 will be greatest at $\pi/2$

and smallest when f_0 is zero



ΔF is largest when the contribution f_0^0 from the anomalous scattering atoms is $\pi/2$ out of phase with that of the non-anomalous atoms



ΔF is smallest (zero) when the f_0^0 contribution is in phase with that of the non-anomalous atoms

The contribution of f_0 will be greatest at $\pi / 2$ and smallest when f_0 is 0. So, if you see that ΔF is largest when the contribution f_0 from the anomalous scattering atoms is $\pi / 2$ out of phase with that of the non anomalous atoms and ΔF is smallest when the f_0 contribution is in phase with that of the non anomalous atoms.

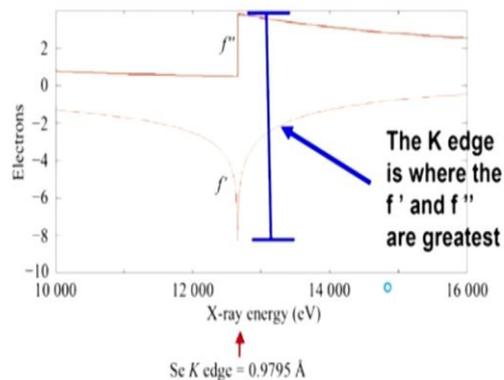
(Refer Slide Time: 33:22)

Variations in Anomalous Scattering at the K (peak) edge of Selenium:

$$f(\theta, \lambda) = f_0(\theta) + f'(\lambda) + if''(\lambda)$$

Dispersive term

Absorption term



Variation in anomalous scattering at the K (peak) edge of selenium,

$$f(\theta, \lambda) = f_0(\theta) + f'(\lambda) + if''(\lambda)$$

where a prime lambda is the dispersive term and double prime lambda is the absorption term.

So, if you see here, the K edge is where the f' and f'' are greatest for selenium, the K edge is 0.9795 angstrom.

(Refer Slide Time: 34:05)

SIRAS and MIRAS:

Single Isomorphous Replacement with Anomalous Scattering (SIRAS) uses anomalous scattering to break the phase ambiguity in a single isomorphous replacement experiment

Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS) uses multiple anomalous scatterers to break the phase ambiguity in a multiple isomorphous replacement experiment

Both of these use the anomalous or Bijvoet difference in the same way as the isomorphous difference in Patterson or direct methods to locate the anomalous scatterers

SIRAS is single isomorphous replacement with anomalous scattering uses anomalous scattering to break the phase ambiguity in a single isomorphous replacement experiment, you could consider it as a hybrid of both the isomorphous replacement along with the anomalous dispersion or anomalous scattering. Multiple isomorphous replacement with anomalous scattering is MIRAS, it uses multiple anomalous scatterers to break the phase ambiguity in a multiple isomorphous replacement experiment. Both of these use the anomalous or Bijvoet difference in the same way as the isomorphous difference in Patterson or direct methods to locate the anomalous scatterers.

(Refer Slide Time: 35:31)

Molecular Replacement Method (MR):

Molecular replacement (or MR) is a method of solving the phase problem in X-ray crystallography

MR relies upon the existence of a previously solved protein structure which is similar to our unknown structure from which the diffraction data is derived

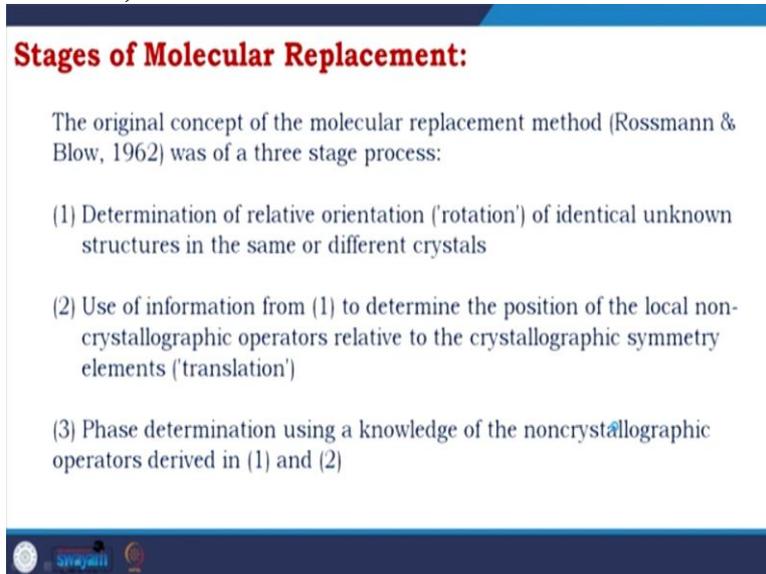
This could come from a homologous protein, or from the lower-resolution protein NMR structure of the same protein



Molecular replacement (MR) is a method of solving the phase problem in X ray crystallography. This process was developed by Michael Rossmann. MR relies upon the existence of a previously solved protein structure which is similar to our unknown structure

from which a diffraction data is derived. This could come from a homologous protein which is having similar property to the protein you are working or from a lower resolution protein NMR structure of the same protein.

(Refer Slide Time: 41:47)



Stages of Molecular Replacement:

The original concept of the molecular replacement method (Rossmann & Blow, 1962) was of a three stage process:

- (1) Determination of relative orientation ('rotation') of identical unknown structures in the same or different crystals
- (2) Use of information from (1) to determine the position of the local non-crystallographic operators relative to the crystallographic symmetry elements ('translation')
- (3) Phase determination using a knowledge of the noncrystallographic operators derived in (1) and (2)

The original concept of the molecular replacement method was of a three stage process.

1. Determination of relative orientation (rotation) of identical unknown structure in the same or different crystal
2. use of information from one, to determine the position of the local non crystallographic operators relative to the crystallographic symmetry element (translation).
3. Phase determination using knowledge of the non crystallographic operators derived in 1 and 2. So, you do the rotation, you do the translation and you utilize them to get a phase.

(Refer Slide Time: 42:44)

Molecular Replacement in Solving Phase Problem:

Molecular replacement is the process of solving the phase problem for an unknown structure by placing the atomic model for a related, known structure in the unit cell of the unknown structure in such a way as to best reproduce the observed structure factors

The known model, once placed, may be used to calculate phases which, in combination with the observed structure factors for the unknown structure, allow the model to be rebuilt and refined

Molecular replacement is the process of solving the phase problem for an unknown structure by placing the atomic model for a related known structure in the unit cell of the unknown structure in such a way as to best reproduce the observed structure factors. The known model, once placed, may be used to calculate phases which in combination with the observed structure factors for the unknown structure, allow the model to be rebuilt and refined. So, from the experiment, you get the structure factor, you get the amplitude. Somehow by doing the theoretical operation of rotation and translation, you find that with the previous result structure and get the phase and solve the new structure.

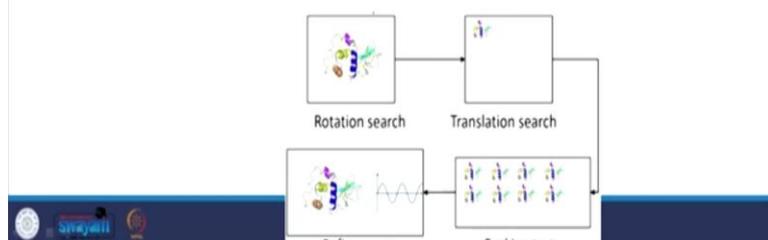
(Refer Slide Time: 43:49)

The calculation involves a 6 dimensional search over all possible orientations and translations of the known model in the unit cell of the unknown structure

This calculation is generally too time consuming to perform in full, so it is usually split into two parts:

A 3 dimensional search over all possible orientations to determine the orientation of the model

A 3 dimensional search over all possible translations to determine the position of the orientated model.



The calculation involves a 6 dimensional search over all possible orientation and translation of the known model in the unit cell of the unknown structure. This calculation is generally too time consuming to perform in full, so usually it is divided into 2 parts, the first one, a 3 dimensional search over all possible orientation to determine the orientation of the model, and second, a 3 dimensional search over all possible translations to determine the position of the oriented model.

(Refer Slide Time: 44:54)

Things to check to start the process:

Number of molecules to search for:

Given molecular weight of the target and the dimensions of the target cell it will return a set of probabilities for the number of molecules in the asymmetric unit

Matthews Coefficient

Check for twinning:

C_{truncate} or S_{check} can be used to examine the data to assess the likelihood that the data is twinned

Is there NCS?

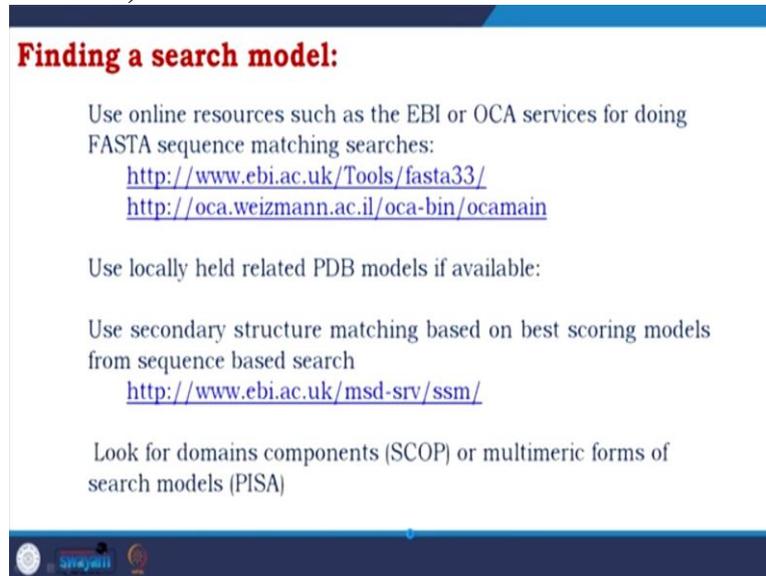
- helps with refinement
- use self rotation function

There are things to check to start the process of molecular replacement, one the number of molecules to search for. Given the molecular weight of the target protein and the dimension of the target cell, it will return a set of probabilities for the number of molecules in the asymmetric unit which is Matthews coefficient. We also have to check for twinning, there are

programs C_{truncate} or Sf_{check} which could be used to examine the data to assess the likelihood that the data is twinned or not.

Is there non crystallographic symmetry? You also have to check, this information would be helpful with refinement and you could use self rotation function.

(Refer Slide Time: 45:56)



Finding a search model:

Use online resources such as the EBI or OCA services for doing FASTA sequence matching searches:
<http://www.ebi.ac.uk/Tools/fasta33/>
<http://oca.weizmann.ac.il/oca-bin/ocamain>

Use locally held related PDB models if available:

Use secondary structure matching based on best scoring models from sequence based search
<http://www.ebi.ac.uk/msd-srv/ssm/>

Look for domains components (SCOP) or multimeric forms of search models (PISA)

So, first we have to do a finding a search model. There are online resources such as the EBI or OCA services for doing, fasta sequence matching searches. You could also do locally held related PDB models if available. Suppose, you already solved one of your structure and you are making mutants, you could use the local PDB to do that. You could also perform secondary structure matching based on best scoring models from sequence based search. You could look for domain components from SCOP or multimeric forms of the search models from PISA.

(Refer Slide Time: 46:50)

Good Search Model: (MGR) → MR → Same Protein
Extremely high identity

General rule-of-thumb is that the sequence identity of homologue to the target must be > 30% for the process to work

Where sequence identity is low it is important to get as good a sequence alignment as possible:

- Use multiple alignment pulling in many related sequences rather than pair-wise
- Profile fitting alignment e.g. Blast

General rule of thumb is that the sequence identity of homologue to the target must be greater than 30% for the process to work. Even when the sequence identity is lower than 30%, it is important to get a good sequence alignment as possible. You could use multiple sequences alignment pulling in many related sequences rather than pair wise, also you could go for global profile fitting alignment like BLAST.

(Refer Slide Time: 48:15)

Preparing your search model:

Signal-to-noise problem:
Anything that is in your model that is not likely to be in your target structure needs to be removed as it will only contribute to the background noise

- Prune back side-chains that aren't aligned
- Cut out flexible loops
- Cut out waters

How to prepare your search model? Signal to noise problem, anything that is in your model which is not likely to be in your target structure needs to be removed as it will only contribute to the background noise. They have been very different in terms of binding ligands, solvent, different loops, very different flexibility in different part of the loops, side chains. You have to take care initially, cut out the flexible loops, crystallographic solvent or precipitant etc.

(Refer Slide Time: 50:16)

CCP4:

Collaborative Computational Project No. 4:
Developing Software for Macromolecular X-Ray
Crystallography



CCP4 exists to produce and support a world-leading, integrated suite of programs that allows researchers to determine macromolecular structures by X-ray crystallography, and other biophysical techniques.

CCP4 aims to develop and support the development of cutting edge approaches to experimental determination and analysis of protein structure, and integrate these approaches into the suite.

CCP4 is collaborative computational project 4, it is a developing software for macromolecules X ray crystallography. CCP4 exist to produce and support a world leading integrated suite of programs that allows researchers to determine macromolecules structures by X ray crystallography and other biophysical techniques. CCP4 aims to develop and support the development of cutting edge approaches to experimental determination and analysis of protein structure and integrate these approaches into the suite.

(Refer Slide Time: 51:14)

CCP4 is a community based resource that supports the widest possible researcher community, embracing academic, not for profit, and for profit research.

CCP4 aims to play a key role in the education and training of scientists in experimental structural biology. It encourages the wide dissemination of new ideas, techniques and practice.

Set up in 1979 to support collaboration between researchers working on such software in the UK

CCP4 was originally supported by the UK Science and Engineering Research Council (SERC), and is now supported by the Biotechnology and Biological Sciences Research Council (BBSRC).

CCP4 is a community based resource that supports the widest possible researcher community, embracing academic, not for profit and for profit research. CCP4 aims to play a key role in the education and training of scientists in experimental structural biology. It encouraged the wide dissemination of new ideas, techniques and practice. Set up in 1979 to support collaboration between researchers working on such software's in United Kingdom,

CCP4 was originally supported by the UK Science and Engineering Research Council which is a SERC and is now supported by the Biotechnology and Biological Sciences Research Council which is BBSRC.

(Refer Slide Time: 52:36)

Softwares: For Preparing your search model

Various programs in CCP4 to help do this:

- Chainsaw:** prunes side chains based on a given alignment
- Molrep:** creates its own alignment and prunes side chains accordingly
- Cutting loops:** look at B-factors, if above an acceptable threshold cut out those residues using PDBCur
- Removing waters and other small molecules:** use PDBset

We need software's for preparing the search model. There are various programs in CCP4 to help doing this. Few I am naming, if you go and play with the software you will know more. Chainsaw: prunes side chains based on given alignment. Molrep: creates its own alignment and prunes side chain accordingly. Cutting loops: look at B factors, if above an acceptable threshold cut out those residues using PDBCur. Removing water and other small molecules: use PDBset.

(Refer Slide Time: 53:19)

CCP4 & Molecular Replacement:

- Molecular Replacement programs:**
 - Molrep
 - Phaser
 - Amore
- Automated MR:**
 - Balbes
 - MrBUMP
- Helper Applications:**
 - Matthes_coef, Chainsaw, Pdbcur, Pdbset, Coordformat, Superpose, PISA

The screenshot shows the CCP4 Program Suite 6.1.13 CCP4Interface 2.0.6 running. The 'Molecular Replacement' menu is expanded, showing options like 'Analysis', 'Model Generation', 'Run Phaser', 'Run Molrep - auto MR', 'Run MrBUMP', 'AMoRe Suite', and 'Utilities'.

CCP4 and molecular replacement. There are molecular replacement programs Molrep, Phaser, Amore. Automated MR, Balbes, MrBUMP. Helper applications: Matthes coef,

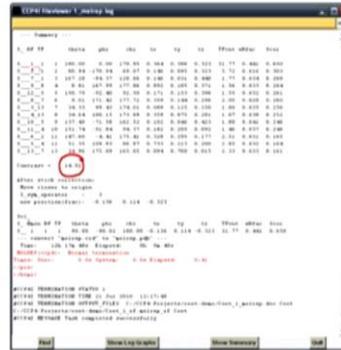
Molrep Output:

Look at the output log file

- Examine RF scores
- Examine TF scores
- Look at the Contrast score

Check to see if the number of molecules asked for have been found

Output PDB file will contain the best positioned model



If you look at the Molrep output, you could look at the output log file, examine the RF (the rotational function score), examine the translational function score and also you look at the contrast score. Check to see, if the number of molecules asked for have been found. Output PDB file will contain the best positioned model.

(Refer Slide Time: 56:45)

Cross Rotation Function: Euler angles (CCP4) R factor

polar angles

List of top RF peaks

Number of molecules	chi	alpha	beta	gamma	RF	RF/alpha
1	85.13	99.14	178.82	100.00	0.8097E+06	4.70
2	85.09	98.46	176.38	78.18	0.7883E+06	4.09
3	84.79	96.06	172.19	59.13	0.7161E+06	3.72
4	71.34	125.45	160.96	97.79	0.6883E+06	3.57
5	84.66	94.93	169.32	49.83	0.6660E+06	3.56
6	167.01	-166.17	34.35	87.07	0.6700E+06	3.48
7	154.83	177.46	139.15	19.82	0.6633E+06	3.44
8	88.33	76.16	173.20	12.31	0.6475E+06	3.36
9	167.89	176.71	158.02	7.94	0.6474E+06	3.36
10	68.18	120.10	156.08	90.42	0.6457E+06	3.35
11	13.90	173.29	34.00	99.82	0.6268E+06	3.25
12	165.41	175.04	155.76	7.55	0.6137E+06	3.18
13	83.59	48.65	174.48	25.29	0.6068E+06	3.15
14	162.10	173.19	155.84	5.87	0.6049E+06	3.14
15	159.88	172.36	153.90	6.22	0.5924E+06	3.07
16	86.11	43.33	179.79	41.80	0.5905E+06	3.06
17	85.01	47.56	173.60	14.81	0.5831E+06	3.03
18	119.85	-116.97	161.87	80.80	0.5782E+06	3.00
19	82.06	105.27	172.41	79.63	0.5535E+06	2.87
20	162.97	-174.47	89.34	52.14	0.5486E+06	2.85
21	93.82	-136.05	177.46	62.32	0.5479E+06	2.84
22	86.40	87.79	163.26	20.92	0.5473E+06	2.84
23	164.30	-175.87	90.92	49.76	0.5457E+06	2.83
24	88.95	88.05	163.67	5.30	0.5434E+06	2.82
25	85.51	89.56	163.91	28.56	0.5414E+06	2.81
26	156.61	171.11	108.83	29.05	0.5392E+06	2.80
27	84.04	46.50	178.72	40.35	0.5375E+06	2.78
28	23.14	141.60	29.65	65.28	0.5328E+06	2.77
29	29.68	123.51	23.36	43.70	0.5253E+06	2.73
30	65.12	134.50	150.58	82.94	0.5252E+06	2.73

number of peaks : 30

INFO: Relations between peaks see in molrep.doc

So, this is details about the cross rotation function, you will see the list of top rotational function peaks and you see how polar angles, Euler angles which is calculated specifically in CCP4 and you get the R factor. R factor, it will talk about the organizer the structure, we will talk in details in the next class.

(Refer Slide Time: 57:29)

Translation Function:

polar angles **fractional translation** **R factor Score**

List of top solutions:

contrast of solution

In the translational function, you see the top solution, polar angles, fractional translations like translation, R factor, the score. More refinement you do, less the R factor would be, and that is one of the very good estimation of how your refinement is going.

(Refer Slide Time: 58:15)

Summary (MR):

- Molecular replacement is the process of retrieving the phase information for a target structure using a related, known structure
- A revolution in the field of Protein Crystallography
- CCP4 provides several programs and helper tools to perform MR
- If you are having difficulty try them all!
- If no search model use the program Balbes or MrBUMP to do the work for you
- Before experiment use MrBUMP or Balbes "Model Search" modes to check to see if good models are available
- This also open the possibility of theoretical modelling for getting 3D structures.

In summary, molecular replacement is the process of retrieving the phase information for a target structure using a related known structure. It is a revolution in the field of protein crystallography. CCP4 provides several programs and helper tools to perform molecular replacement. If you are having difficulty, you should try them all. If no search model, use the program Balbes or MrBUMP to do the work for you. Before experiment, you could use MrBUMP or Balbes, model search modes to check to see if good models are available or not.

Thank you very much. We have kind of finishing up the X ray crystallography module. We have only one class left. And I will discuss about how to submit and refine the structure. And then after refinement, we will deposit it.