Phasing a crystal structure

Osman Mirza

Outline

How to solve a crystal structure

- Defining the phase problem

How to obtain the phases

- Molecular replacement
- Multiple isomorphous replacement
- Multiple wave length anomalous dispersion

Improving phases

- Density modification

Each diffracted beam can be described mathematically by the Structure factor **F**(hkl)

$$F(hkl) = |F_{obs}(hkl)|\, e^{j\alpha_{obs}(hkl)}$$

$$I(hkl) \propto \left|F(hkl)\right|^2$$

α_{obs} cannot be measured during the diffraction experiment!

Why is the complete Structure factor important for us?

$$\rho(r) = \sum_{hkl} \mathbf{F}_{hkl} \exp\left[(-2\pi i(hx + ky + lz)\right]$$

$$= \sum_{hkl} |F_{obs}| e^{j\alpha_{obs}} \exp\left[(-2\pi i(hx + ky + lz)\right]$$

Determining the α_{obs} is essential to determine a macromolecular structure

Different ways to solve the phase problem

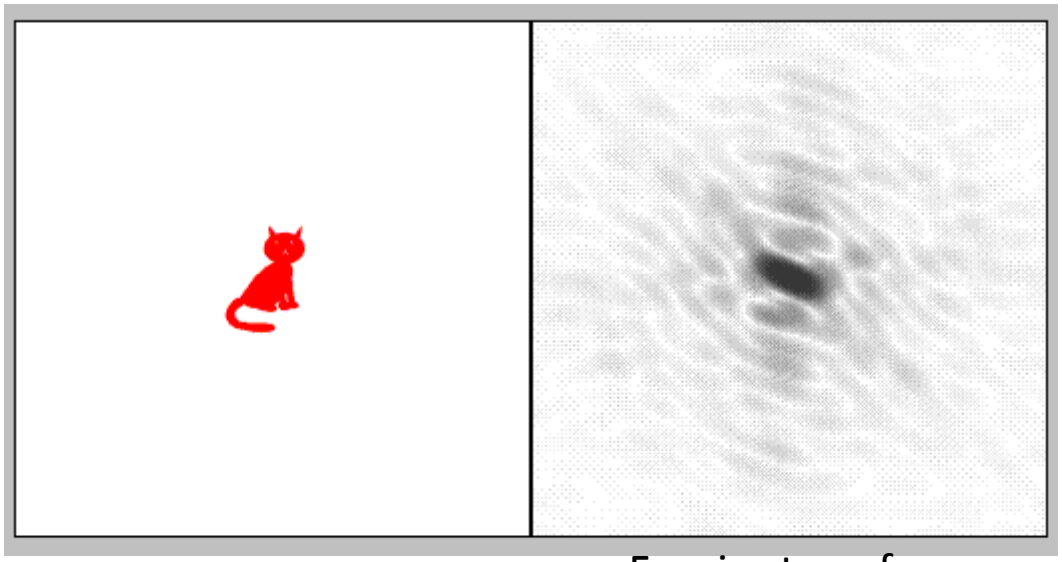
- **MIR** (Multiple Isomorphous Replacement)
  Heavy atom bound in the protein crystal
- **MAD** (Multiple wavelength Anomalous Dispersion)
  Anomalous scatterer in the crystal

- **MR** (Molecular Replacement)
  Homologous model, same fold (~35% sequence identity)

- Direct Methods – small molecules

Principle of Molecular Replacement

In MR the phases come from a known structure

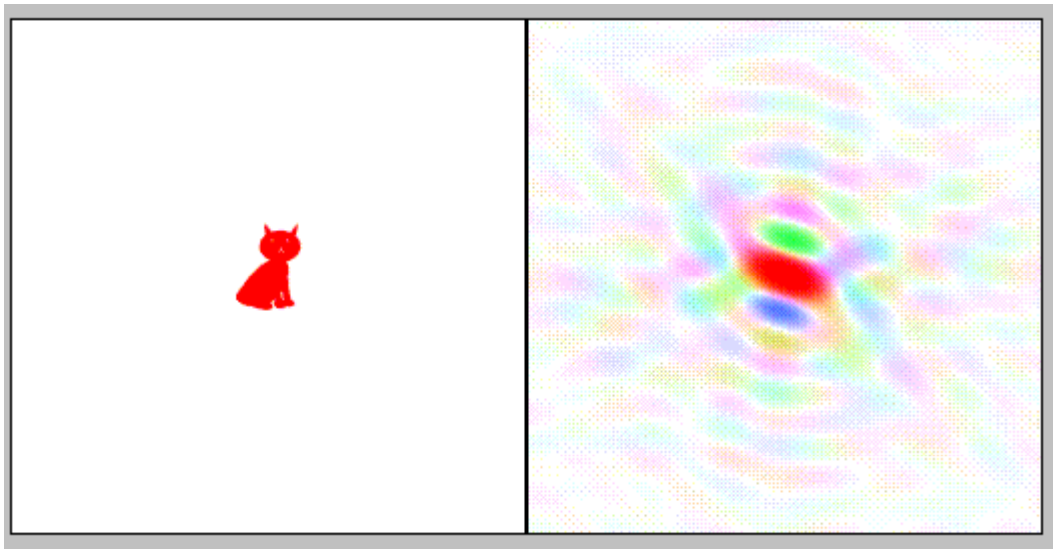


cat with tale

Fourier transform of cat with tale without phases (black/white)

Principle of Molecular Replacement

In MR the phases come from a known structure

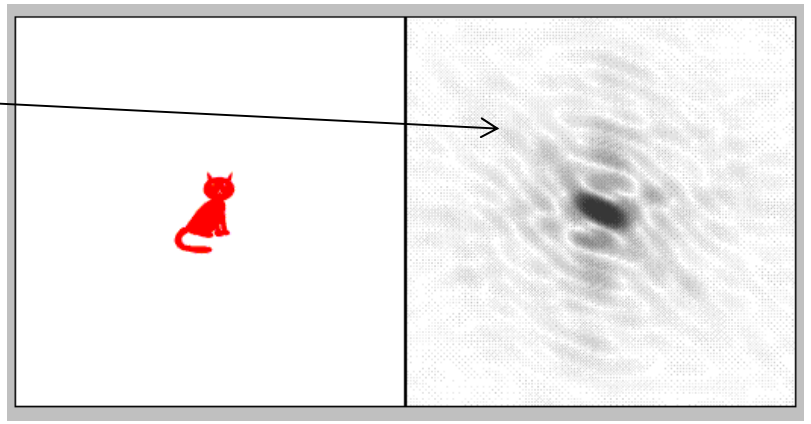


cat without tale

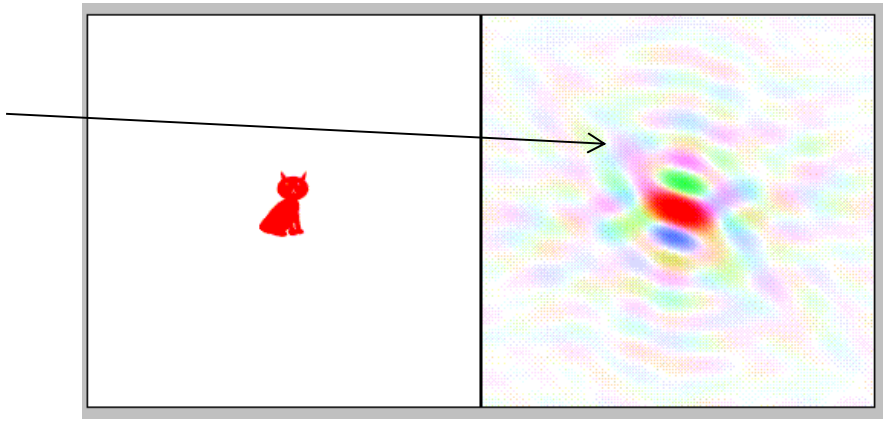
Fourier transform of cat without tale with phases (colour)

Principle of Molecular replacement

- Intensities come from measurements

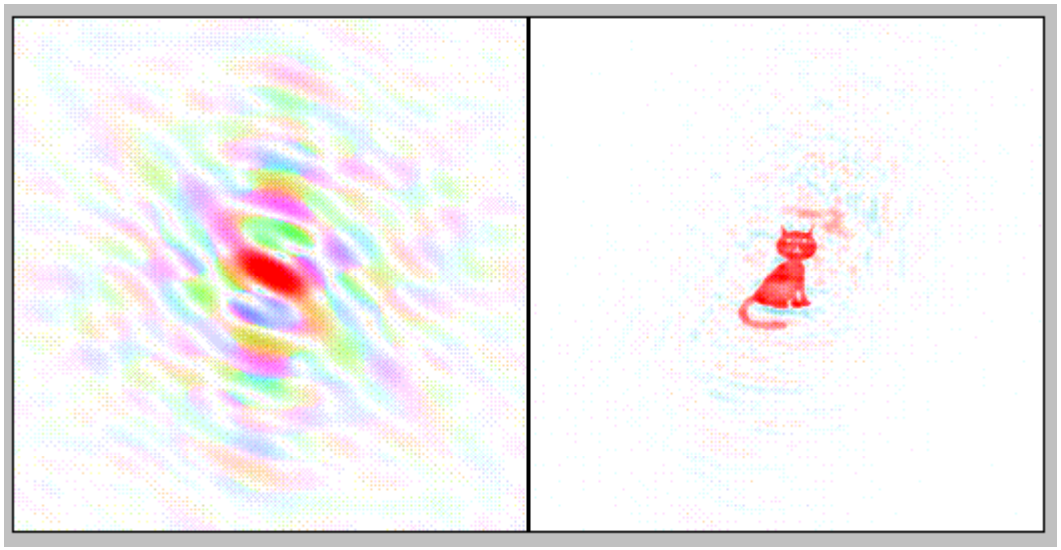


- Phases (colours) come from search model



Molecular Replacement

In MR the phases come from a known structure



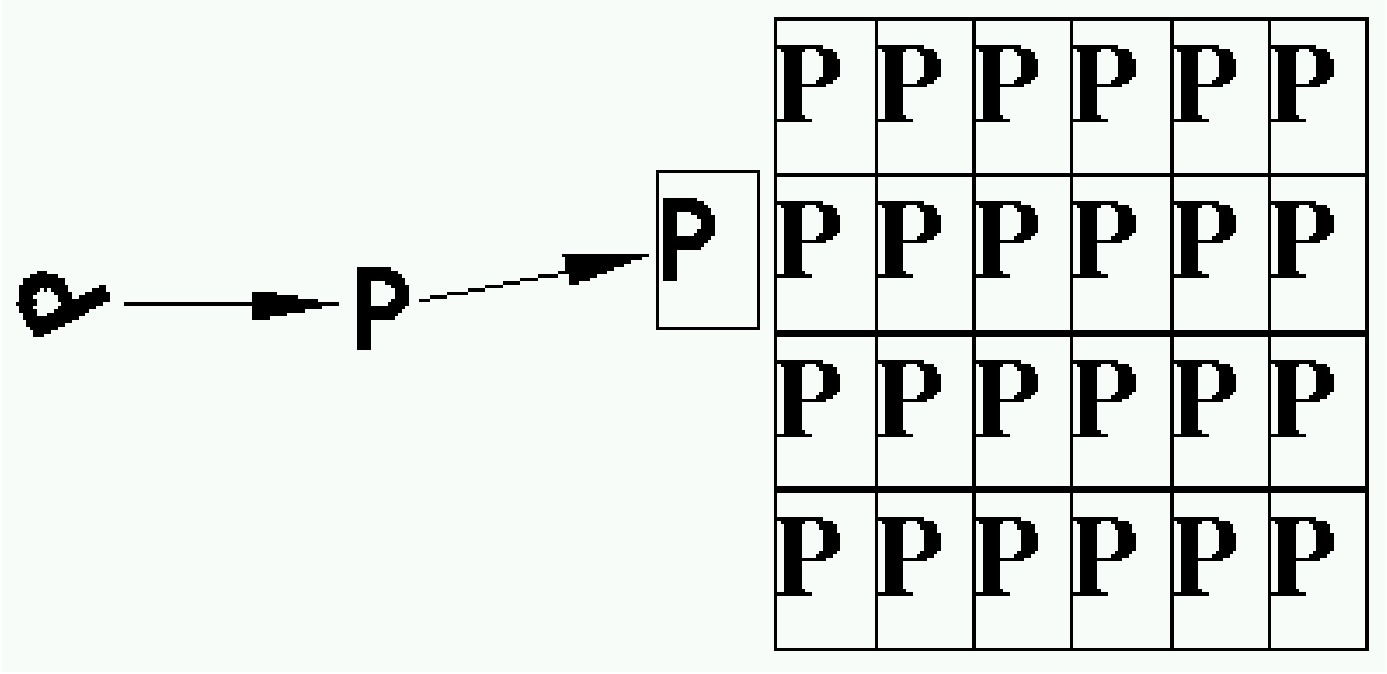
Colours (phases) from FT
of cat without tale are
combined with data
(amplitudes) from cat with tale

prerequisite conditions to make MR

- same protein in another space group
- mutant or complex of the same protein
- homologous model – the closer the better (minimum 35% sequence identity)
- NMR structure/theoretical model
- Fragments (domains) of multi domain protein

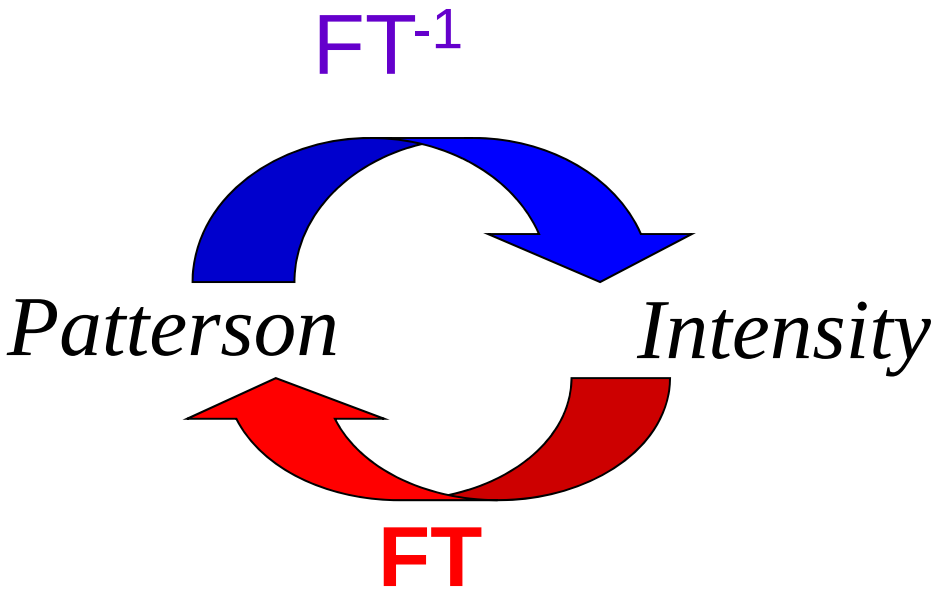
molecular replacement

place a model in the best way in the asymmetric unit



How do we place the search model in the asymmetric unit?

We use the Patterson function



connection between the measured intensity and the Patterson function

Patterson function

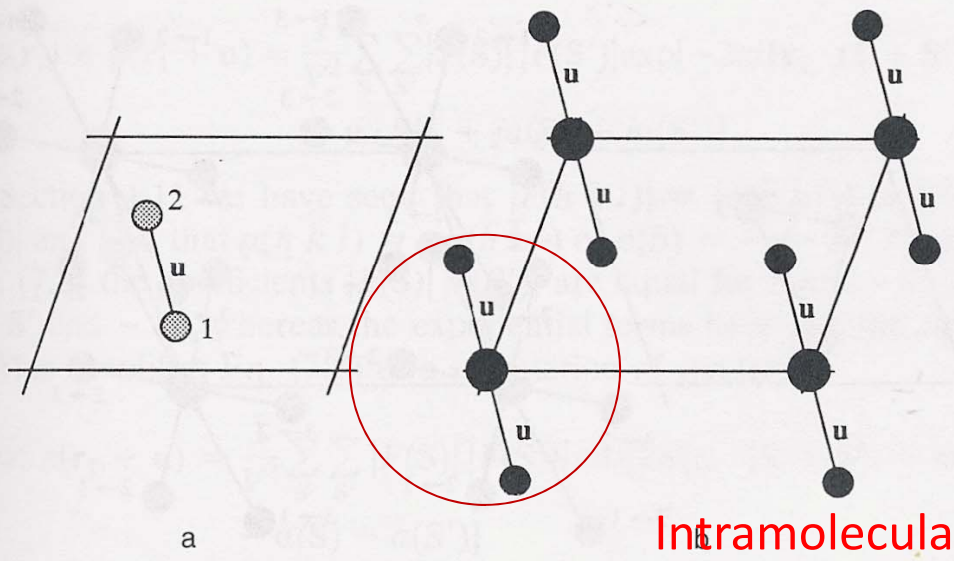
1. peaks corresponding to interatomic distances in the structure
2. N atoms $\rightarrow$ N(N-1)+1 peaks
3. intensity of peaks is proportional to the product of the respective atom numbers (the number of electrons)
4. centrosymmetric (if there is a vector from 1 to 2, there will be a vector from 2 to

- The patterson function may be calculated directly from the pdb-file.
- The patterson function may be calculated directly from the data set

Patterson function from 2 atoms in a 2D cell

Patterson function

Unit cell



Intramolecular vectors

Figure 7.1. (a) A two-dimensional unit cell with only two atoms. (b) The corresponding Patterson cell.

6-dimensional problem

if a molecule is placed in the asymmetric unit 6 parameters are necessary. 3 for the orientation (rotation) and 3 for the position (translation)

reduced to two 3-dimensional problems

first determine the orientation
- afterwards the position

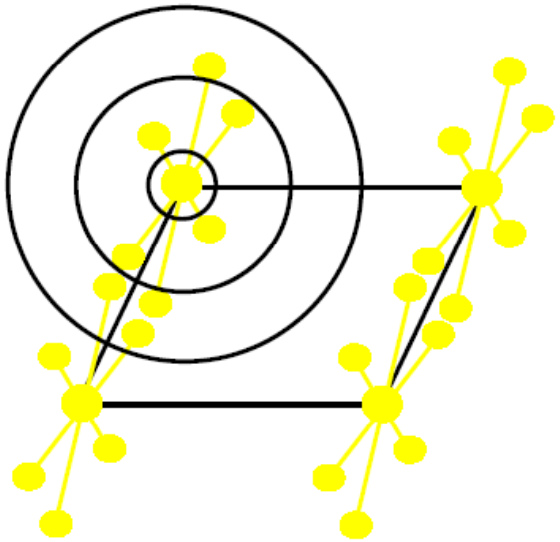
rotation function

the rotation function is determined from the part of the Patterson function **close to the origin**

- the intramolecular vectors are shorther than the intermolecular vectors
- the intramolecular vectors are independent of the position in the unit cell

how **close to the origin** is determined from the size of the molecule

rotation function
(intramolecular vectors)



(to simple)

translation function

- When the rotation function has been fixed, the translation function is determined by using the intermolecular vectors – that is the Patterson peaks further away from the origin

If no template is available the phases have to be determined experimentally

Different ways to solve the phase problem

- **MIR** (Multiple Isomorphous Replacement)
  Heavy atom bound in the protein crystal
- **MAD** (Multiple wavelength Anomalous Dispersion)
  Anomalous scatterer in the crystal

- **MR** (Molecular Replacement)
  Homologous model, same fold (~35% sequence identity)
- Direct Methods – small molecules

Principle of MIR

Native | Heavy atom derivative 1 | Heavy atom derivative 2

The crystals must be isomorphous!

Isomorphous means that the **crystal structures** are exactly the same – the only difference is the heavy atom sites

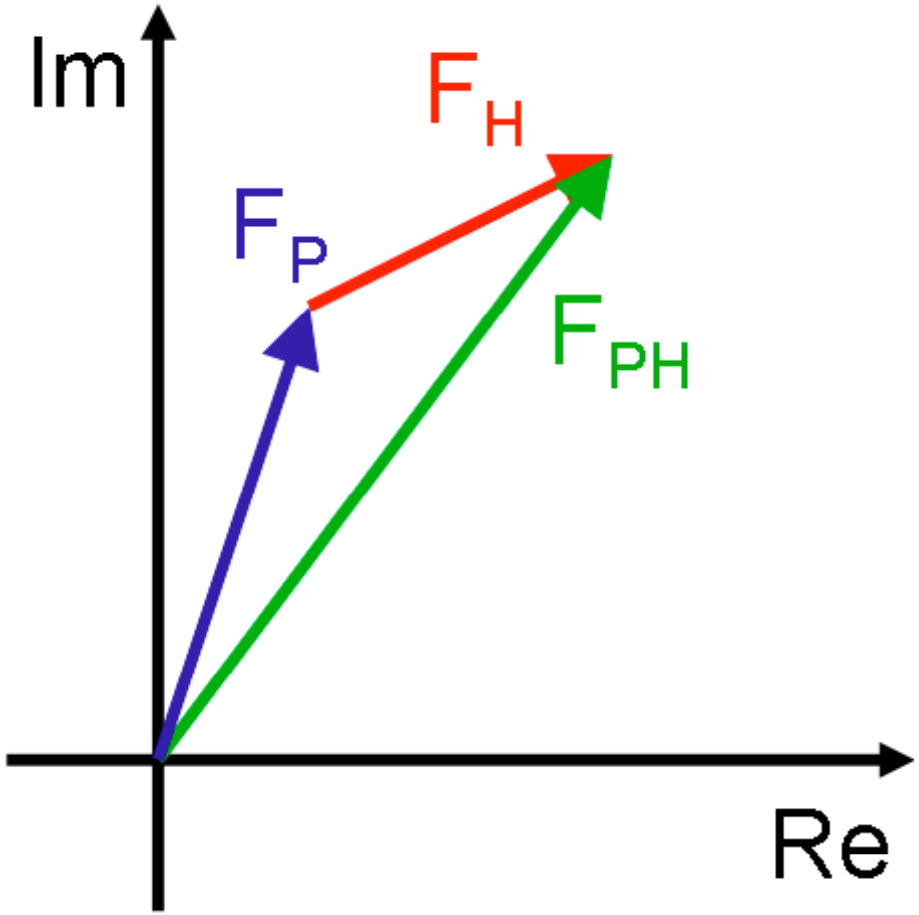
Principle of MIR

Native | Heavy atom derivative 1 | Heavy atom derivative 2

For each reflection, hkl (or **h**), 3 measurements

$F_P(hkl)$	$F_{PH1}(hkl)$	$F_{PH2}(hkl)$
$F_P(\mathbf{h})$	$F_{PH1}(\mathbf{h})$	$F_{PH2}(\mathbf{h})$

isomorphous means: $\mathbf{F_P(h)+F_H(h)=F_{PH}(h)}$

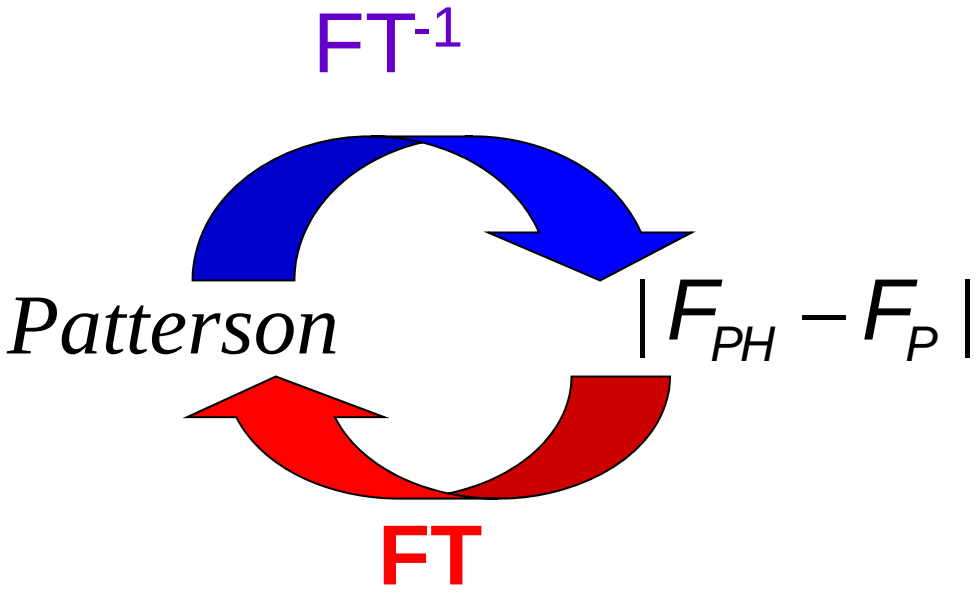


If we can calculate α_H, then we can calculate α_P

How do we calculate α_H?

By determining the positions of the heavy atoms!

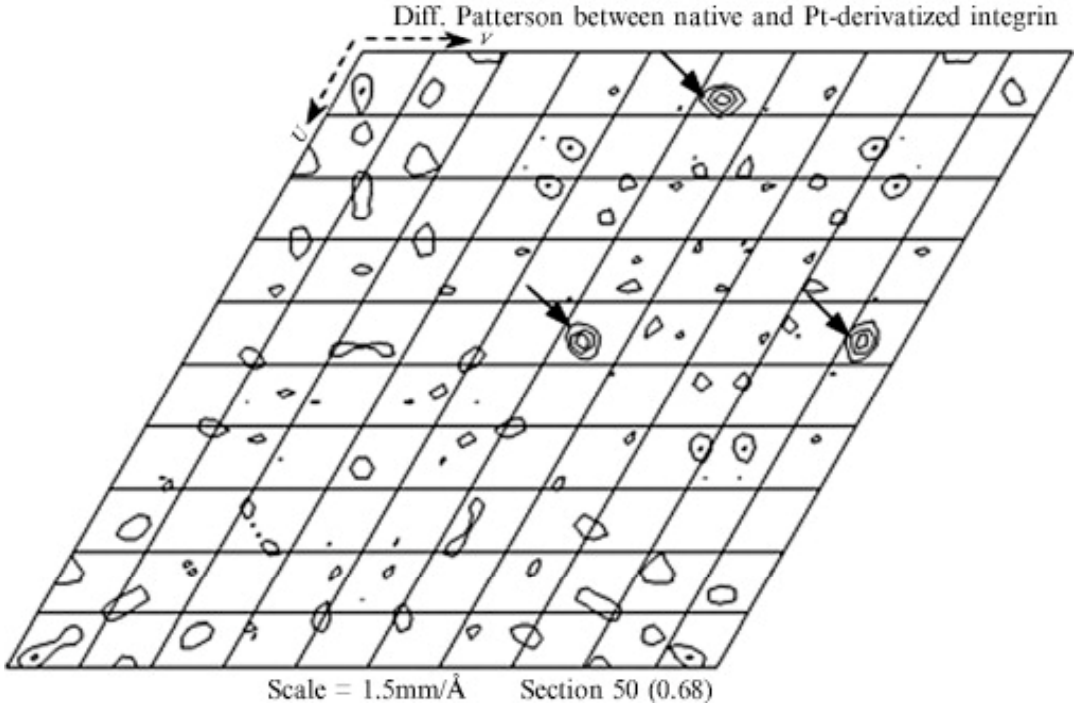
The heavy atom positions can be determined by using a difference patterson map



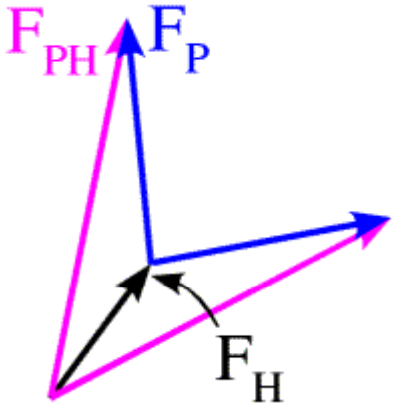
1. peaks corresponding to interatomic distances in the structure

1. intensity of peaks is proportional to the product of the respective atom numbers (the number of electrons)

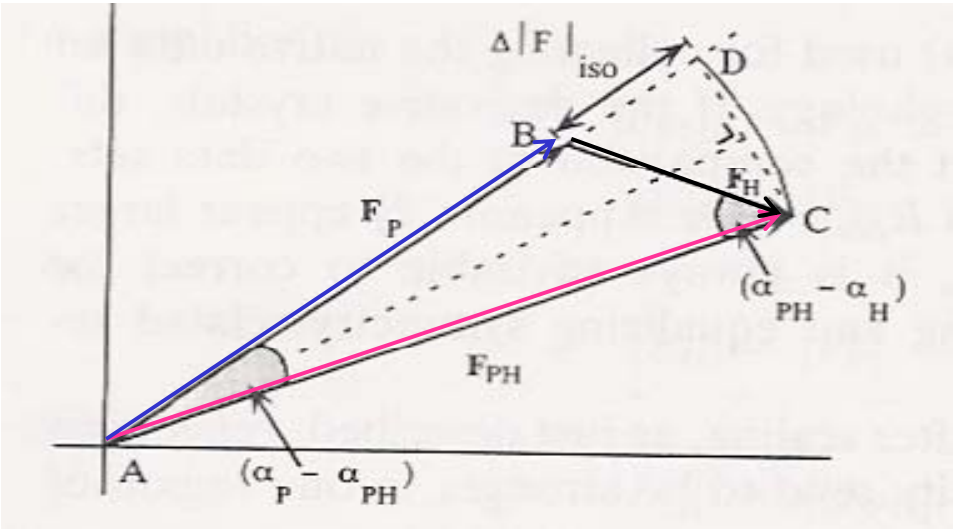
With a few number of heavy atom sites, the Patterson map can be easily interpreted!

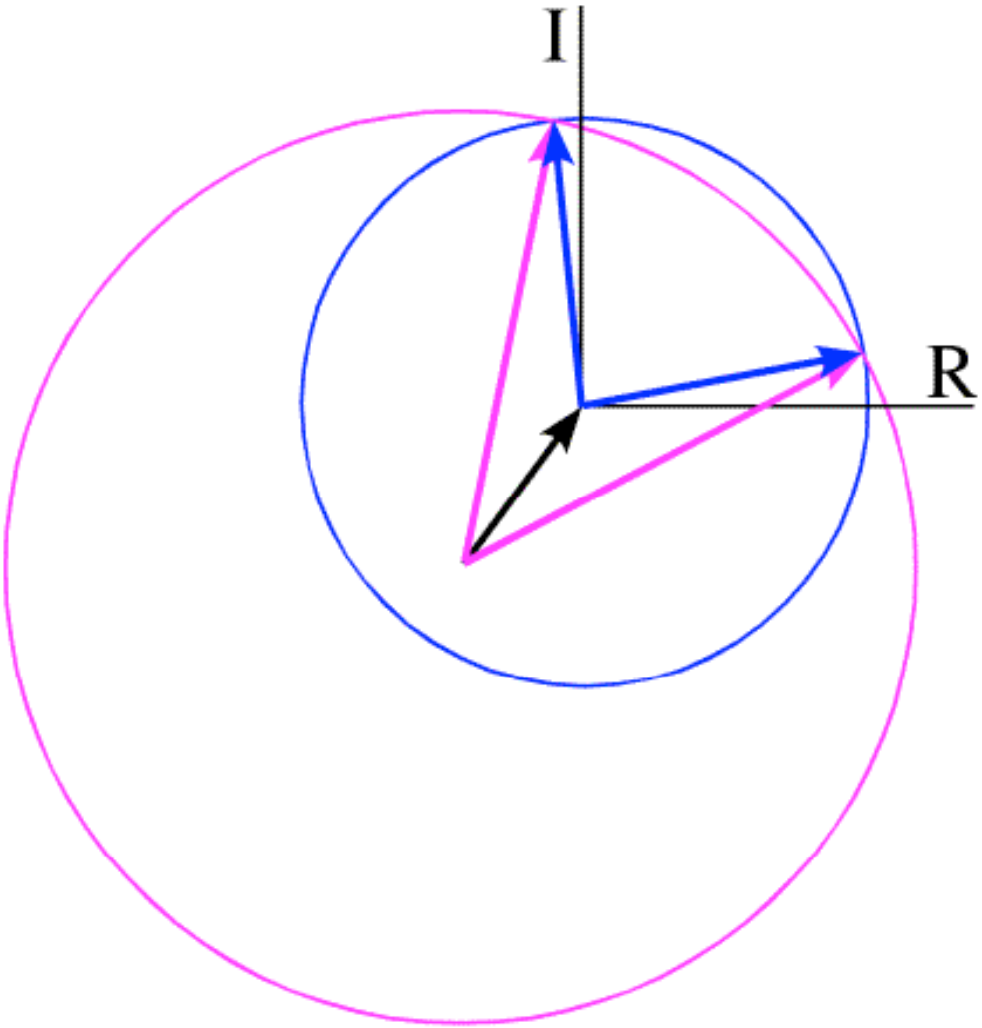


We now have $\mathbf{F}_H$ - both length and phase.
We furthermore have F_{PH} and F_P - only length

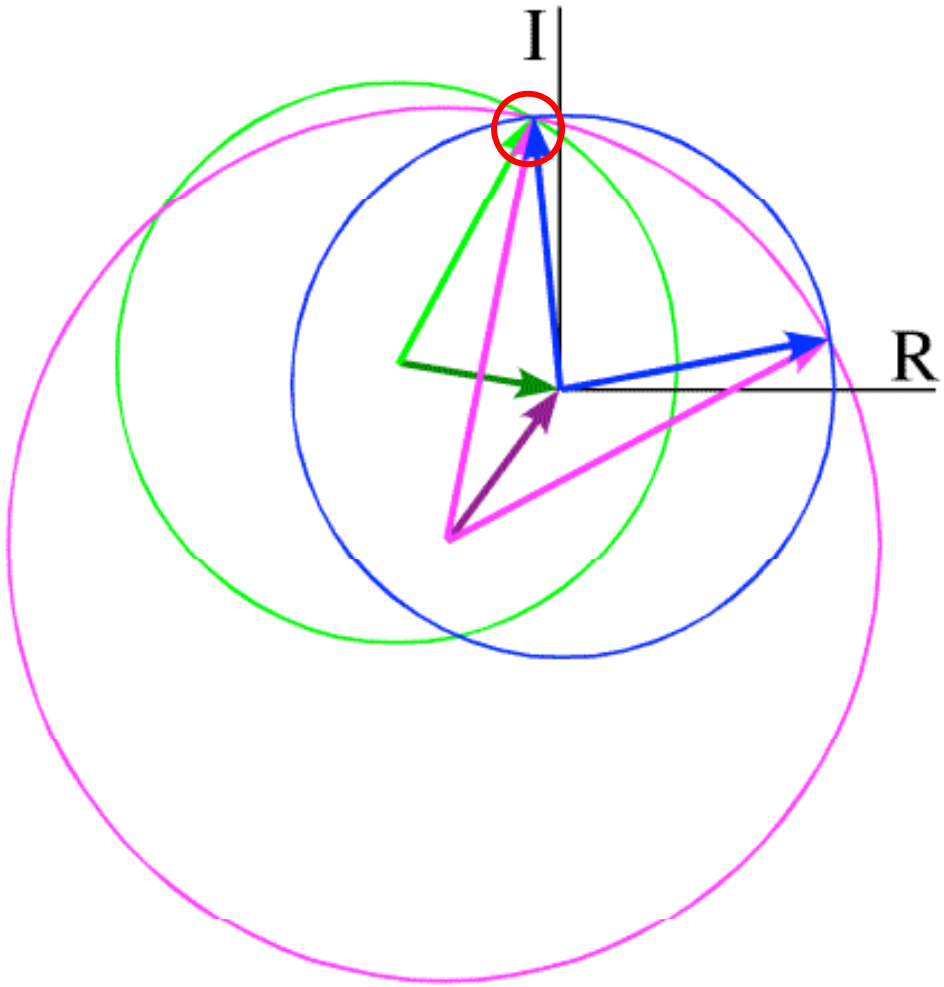


- phase information can be obtained, but it is ambigous
  - two possible phases for each reflection F(**h**)




I
R

Including a second derivative solves the ambiguity!



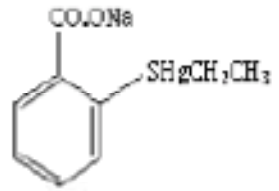
As the crystals are not perfectly isomorphous the two possible solutions are not necessarily coinciding

1. For each derivative there is a phase probability function for each reflection
2. These are multiplied
3. the best phase is the weighted average (the centroid of the joint phase probability function)

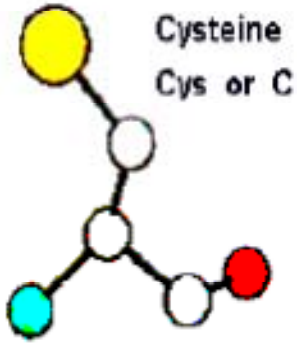
Frequently used heavy atoms

platinum	-	78 electrons
gold	-	79
mercury	-	80
uranium	-	92
iridium	-	77
lead	-	82

HAZ

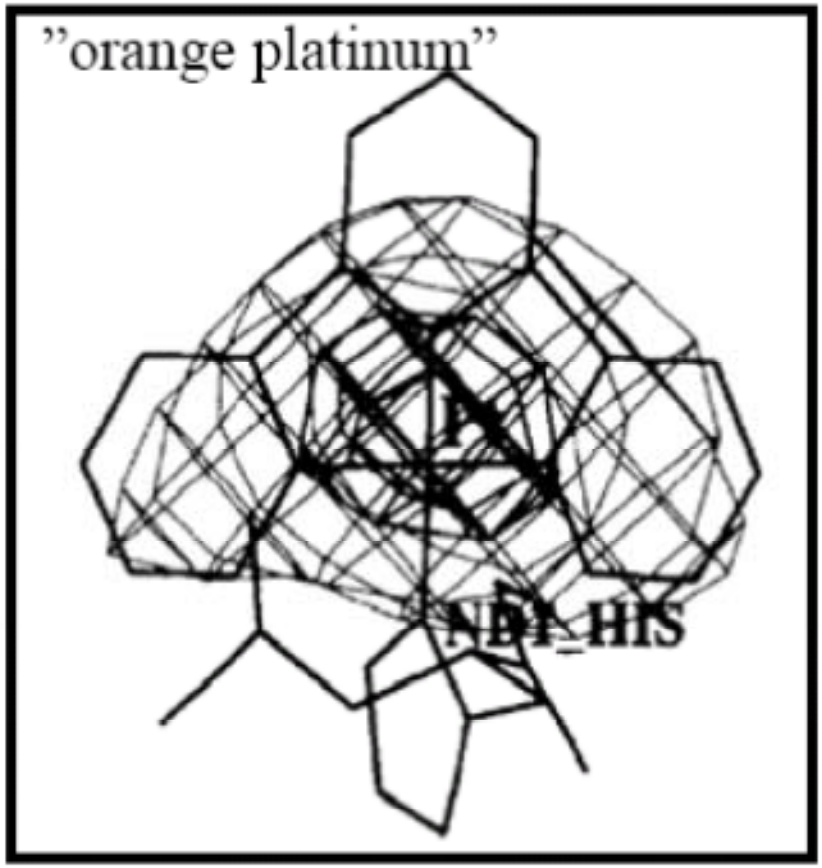


Thiomersal or Ethyl mercury thiosalicylate

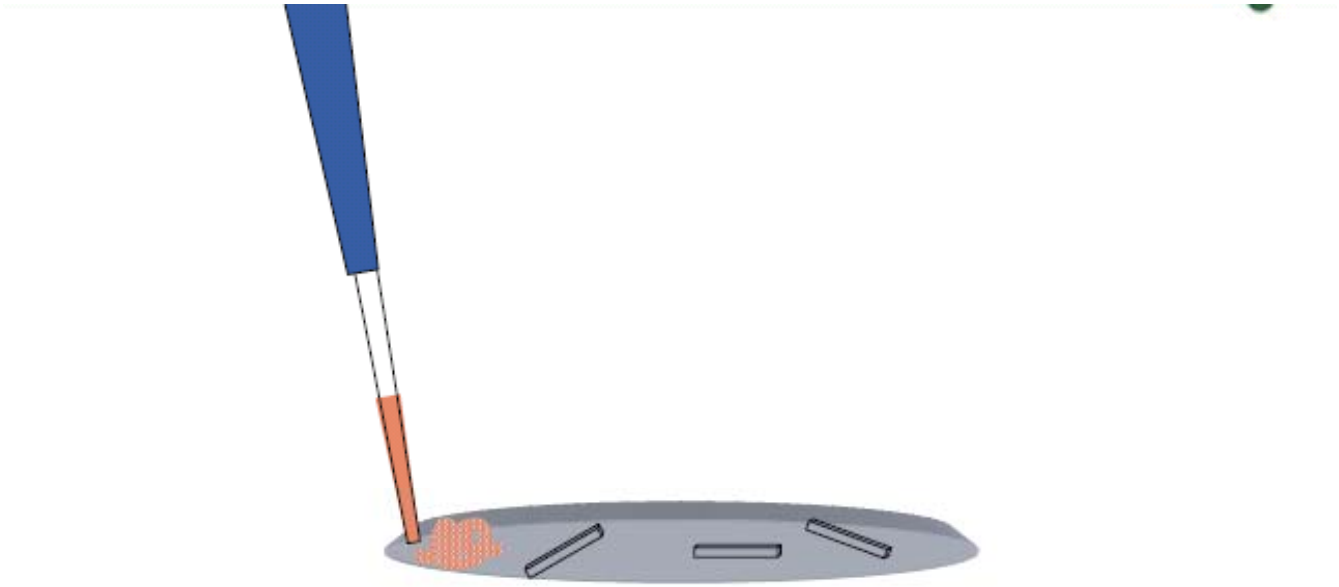


Cysteine

Cys or C



soak the crystals



1 hour – several days

In some cases non-isomorphism may be serious issue that cannot be overcome

Different ways to solve the phase problem

- **MIR** (Multiple Isomorphous Replacement)
  Heavy atom bound in the protein crystal

- **MAD** (Multiple wavelength Anomalous Dispersion)
  Anomalous scatterer in the crystal

- **MR** (Molecular Replacement)
  Homologous model, same fold (~35% sequence identity)

- Direct Methods – small molecules

$$\mathbf{F}_{hkl} = \sum_j f_j \exp\left[(-2\pi i(hx + ky + lz)\right]$$

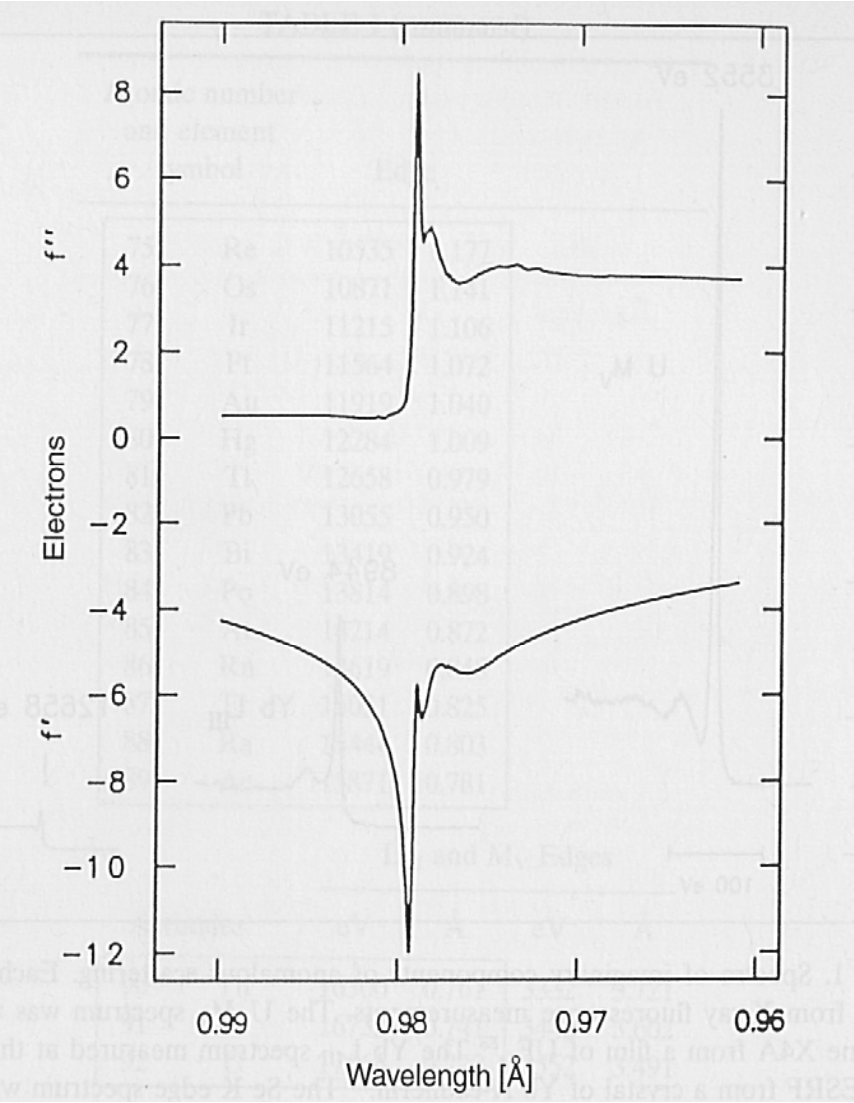
$$f_j = f_0 + f' + if''$$

The anomalous part of the scattering factor becomes significant close to the Absorption edge of the atom.

Selenium anomalous signal

f''

f'



consequences?

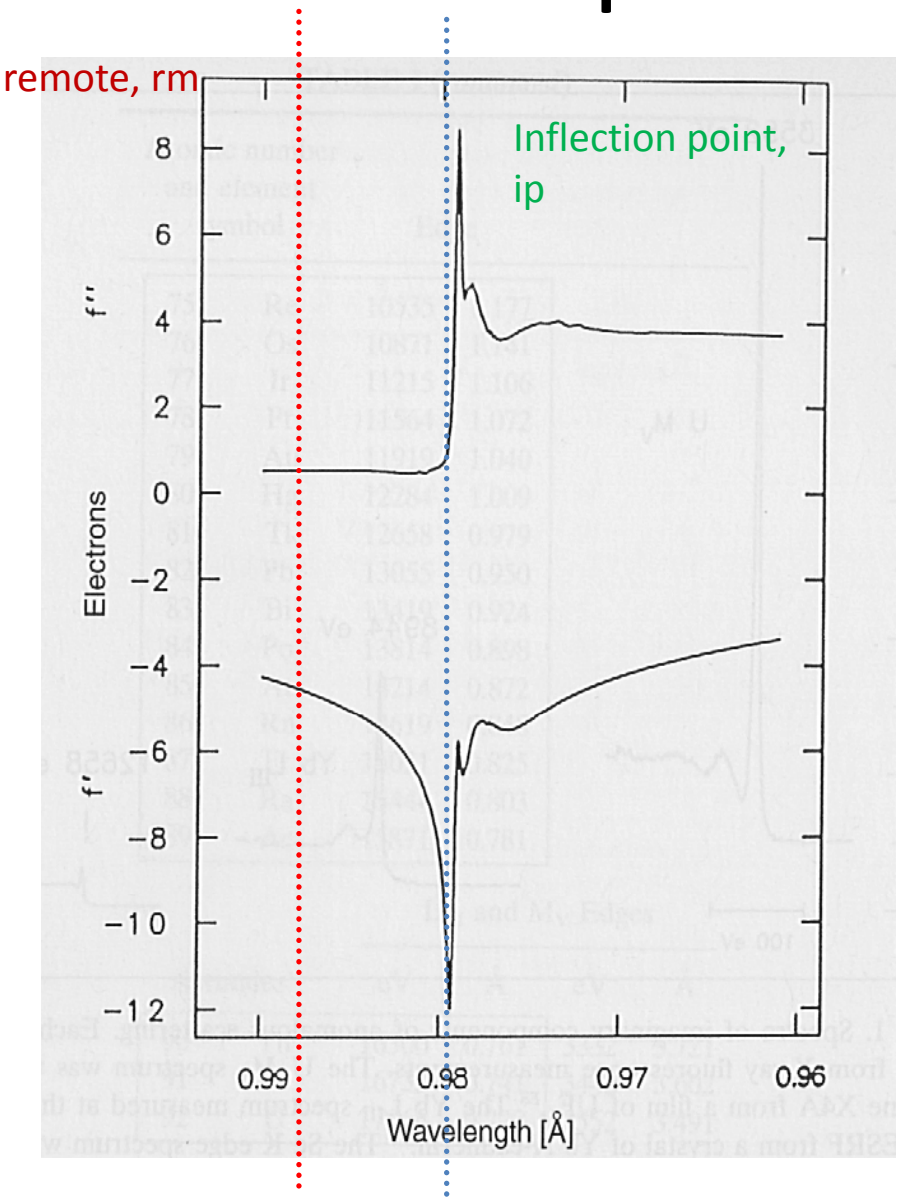
- Change λ and you change the form factor from one particular type of atoms

- Friedel's law breaks down (because f'' is imaginary)

The use of these consequences to solve protein crystal structures is called MAD

- Dispersive difference
- Bijvoet difference

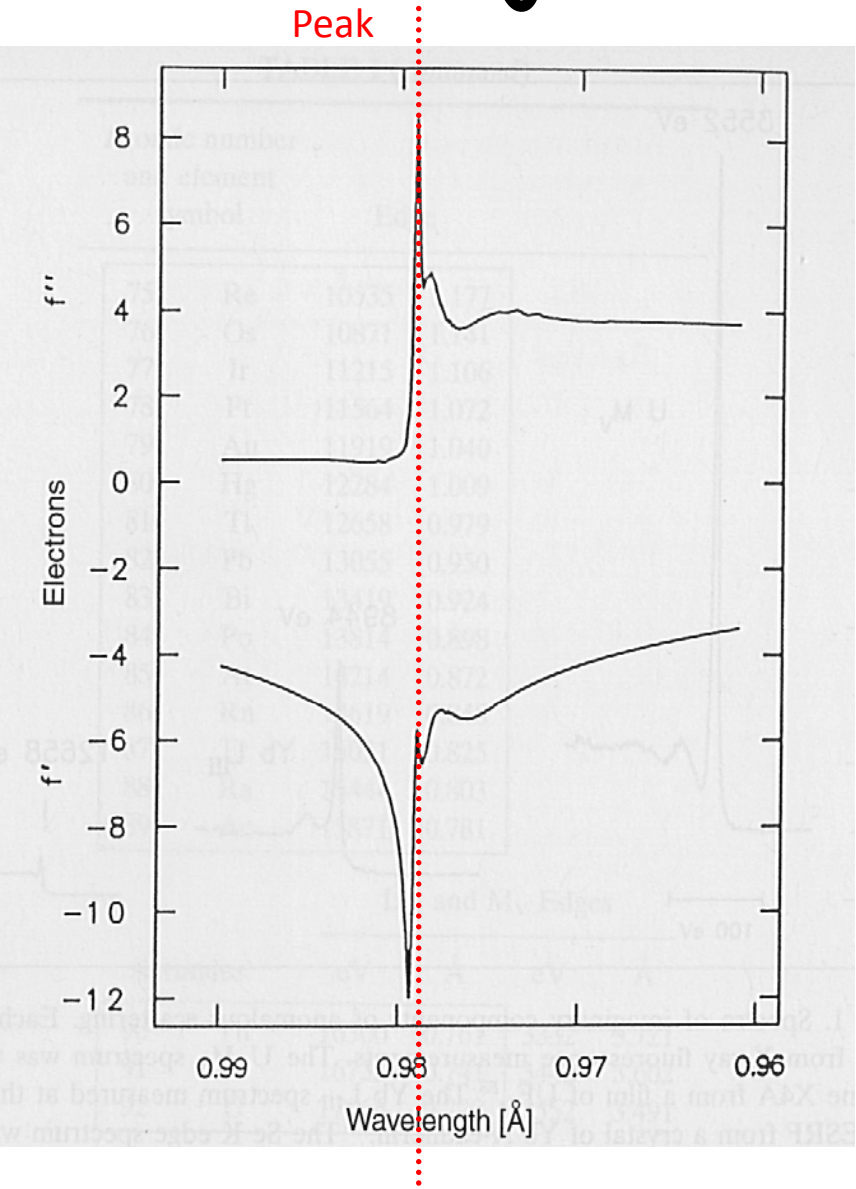
Dispersive differense



- two data sets where *f'* differs

  (perfectly isomorphous MIR case)
- $|\Delta F'(\lambda)|$ is used to find the positions of the anomalous scatterers

Bijvoet differense



- A single data set at the wave length where f'' is max

  $\Delta|F|_{ano} = |F(hkl)| - |F(-h-k-l)|$ is used to find the positions of the anomalous scatterers

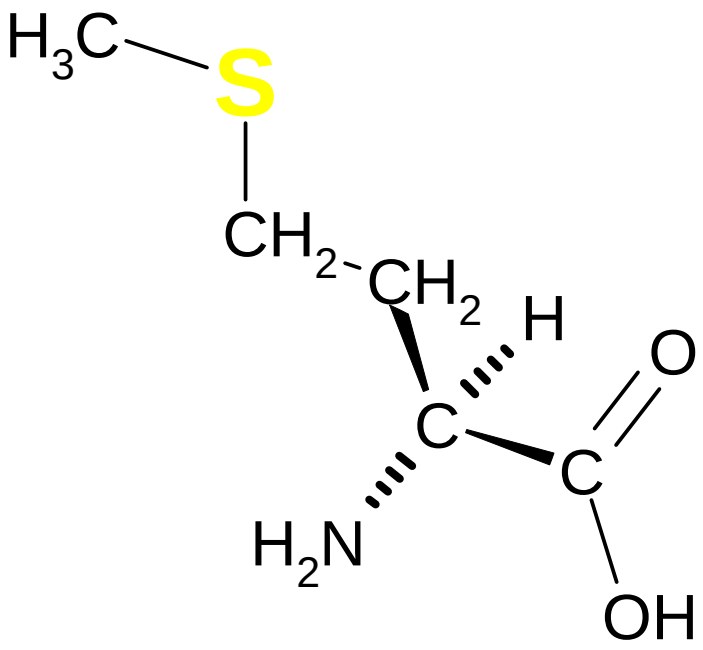
success full MAD-experiment:

- beam time on MAD beam-line
- cryo cooling
- Automatic solution (from computer)

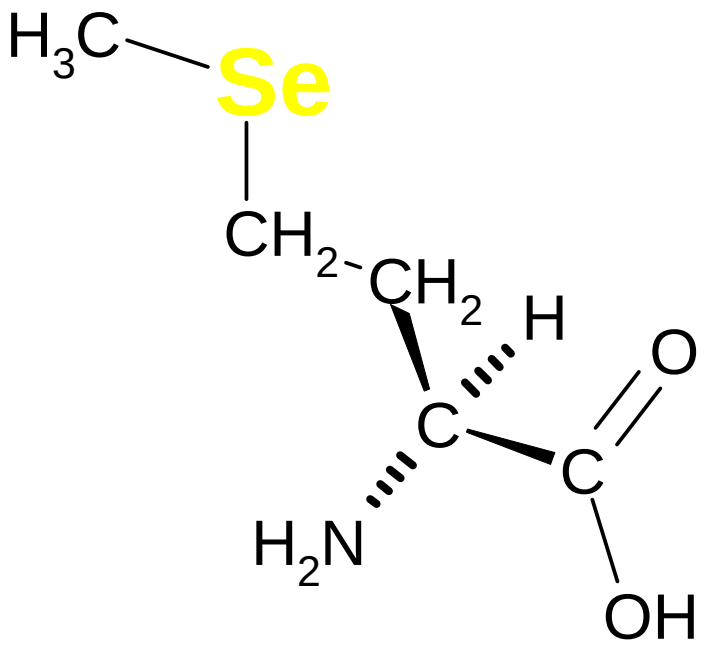
  – and AN ANOMALOUS SCATTERER IN THE CRYSTAL

Likely targets for MAD measurements

Category	*Anemalous scatterer*
Metalloproteins	
Transition metals	*Fe, Cu, Zu, Mn*
Other metals	*Ca, Mo*
Metal replacements	
Lanthanides for Ca^{2+}, Mg^{2+}	*Tb, Ho, Yb*
Mercury for Zn	*Hg*
Heavy atom complexes	
Common protein derivatives	*Pt, Au, Hg, Pb, W, U*
Cluster compounds	*Ta, W*
Building - unit analogs	
Selenomethionine or selenocysteine	*Se*
Telluromethionine	*Te*
Brominated or iodinated nucleotides	*Br, I*



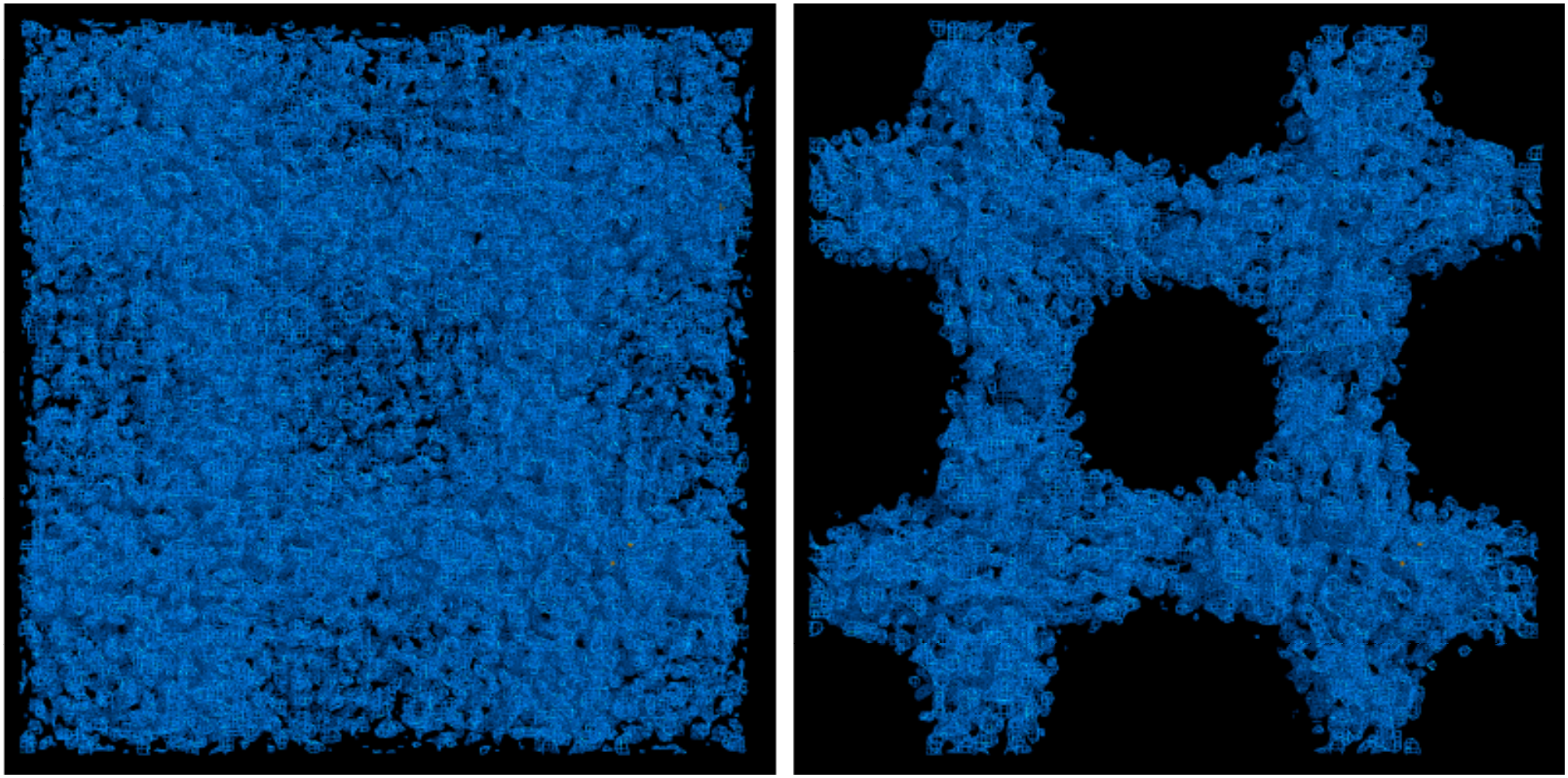
L-Methionine



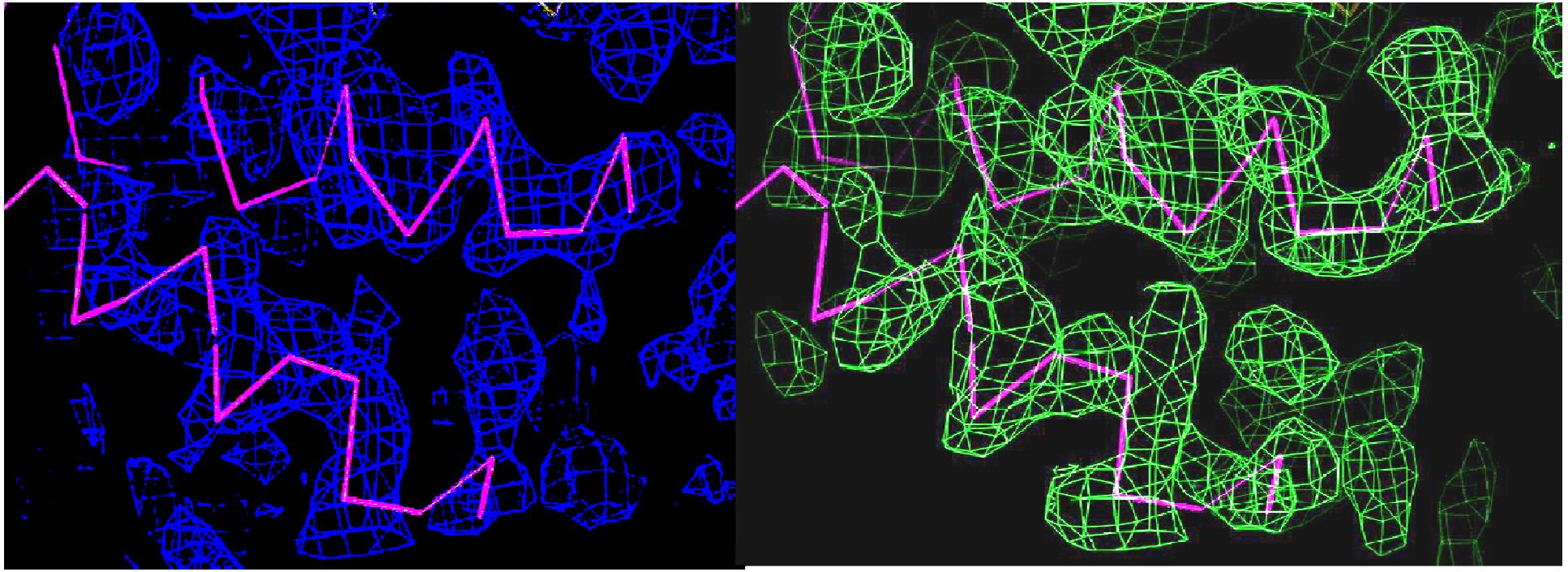
Seleno-L-Methionine

density modification

- improvement of the electron density, so that it is easier to interprete

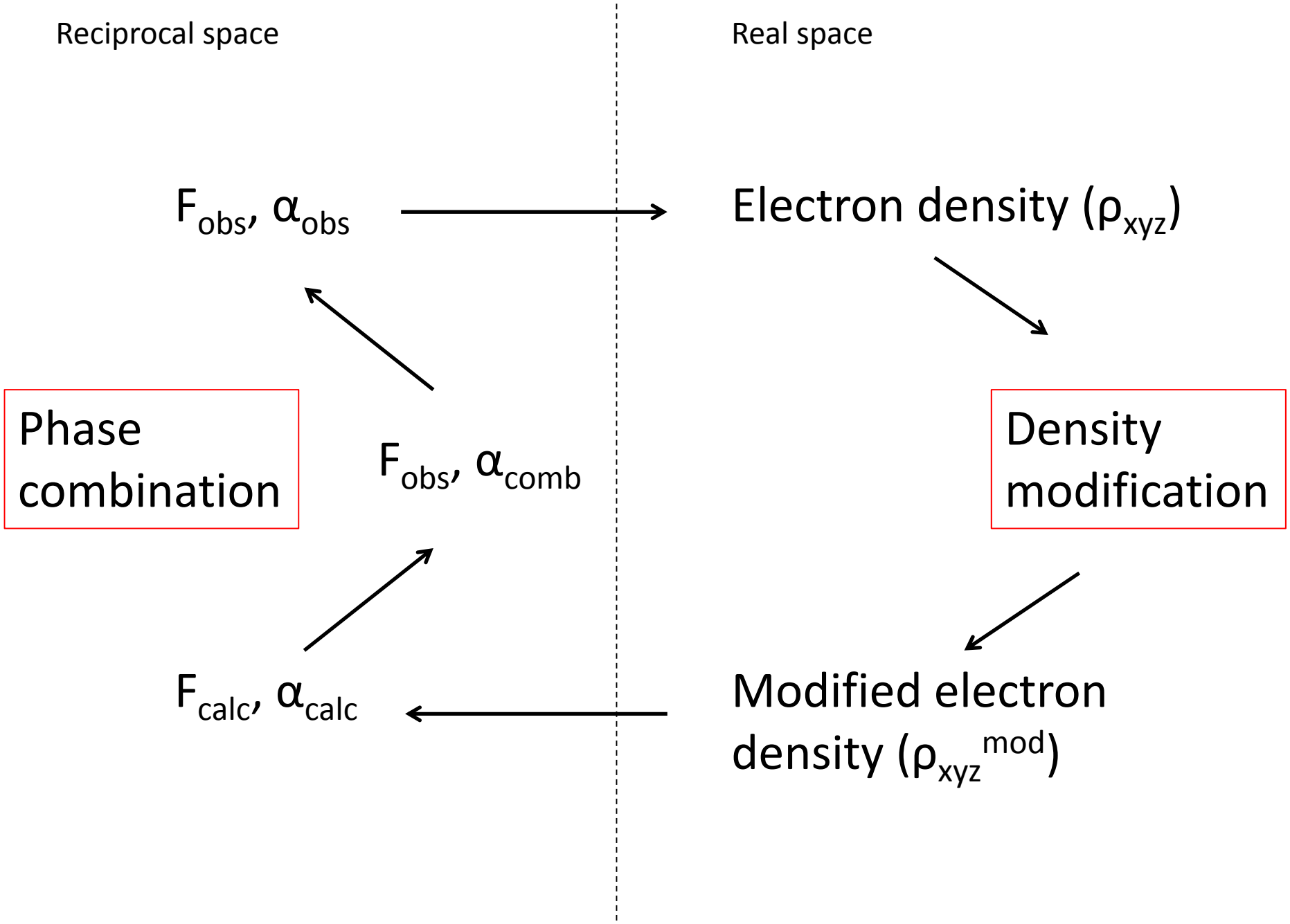


small changes can be important for building the molecule



- phases come from MIR, MAD or MR – how to improve them? – get help from our knowledge on the protein and the crystals
- improvements are going on in real space (not in reciprocal space).

  Fourier transform back to F_{calc} and α_{calc} and calculate a new map from F_{obs} and α_{calc}.


Reciprocal space
Real space
F_{obs}, α_{obs}
Electron density (ρ_{xyz})
Phase combination
F_{obs}, α_{comb}
Density modification
F_{calc}, α_{calc}
Modified electron density (ρ_{xyz}^{mod})

1. solvent channels contain disordered solvent – that is a flat electron density solvent flattening

2. more molecules in the asymmetric unit – they must be (almost) identical ncs-averaging (non-crystallographic symmetry)

1. empirically it has been observed that the electron density of a protein must have a certain profile, try to fit it to that profile by histogram matching