

# Handout 9: Structure solution

Data processing, direct methods, Patterson and  
Fourier maps

Chem 6850/8850

X-ray Crystallography

Department of Chemistry & Biochemistry

[cora.lind@utoledo.edu](mailto:cora.lind@utoledo.edu)



THE UNIVERSITY OF  
TOLEDO

# Before you start...

- **You need to have at least a pretty good idea of your unit cell contents**
  - needed for data scaling
  - also necessary to search for/place atoms
- **Elemental analysis gives stoichiometry**
  - can be “traditional” elemental analysis
  - for organics, could be deduced from spectroscopic data
- **Once we know the stoichiometry, we need to estimate Z**
  - symmetry considerations
  - estimate based on average volume per atom
  - based on density measurement

# The challenge... and the problem

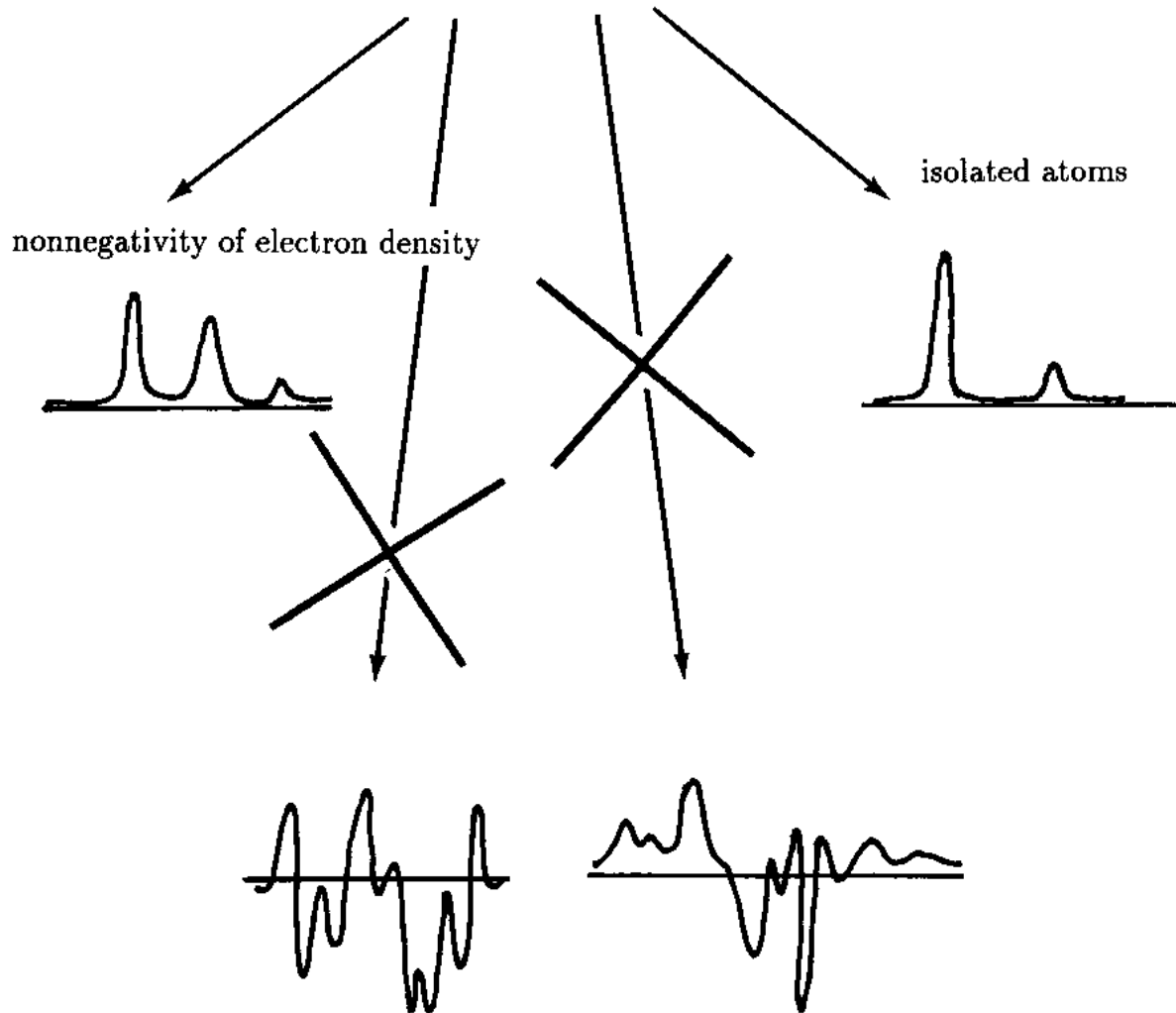
- **We want:**
  - electron density distribution of our crystals
  - can be determined by Fourier synthesis of a large number of  $\mathbf{F}_{hkl}$  values
- **We have:**
  - only  $|F_{hkl}|$  can be obtained from intensity measurements
  - the phases are unknown
- **We need to recover the lost phase information if we want to solve a crystal structure**
  - often referred to as “the phase problem”

# Recovering the phase information

- If there were no constraints on the electron density, trying to estimate what the lost phases are would be a hopeless endeavor
- **Luckily, we know a few conditions that must be fulfilled:**
  - the electron density *must* be positive at all points
  - the electron density should occur in “balls” – the atoms
  - atoms must be placed at least a minimum distance apart
- **With this knowledge, it is often possible to recover the missing phase values**

# Constraints on the electron density

Conditions restricting relative phases



“Crystal Structure Analysis for Chemists and Biologists”, Glusker, Lewis and Rossi, VCH, 1994.

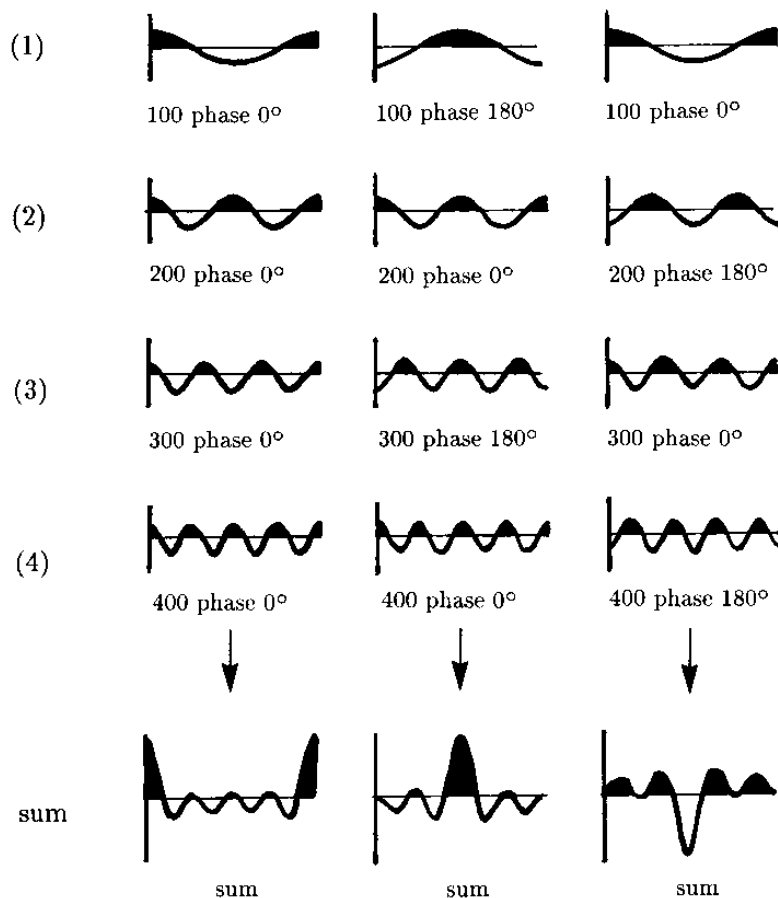


# Possible phase choices

- **For a centrosymmetric structure, the phase angle can only be 0 or 180°**
  - results in  $2^N$  possible sets of phases for N F's
  - only one set gives the right answer!
- **For a non-centrosymmetric structure, the phase angle can take any value between 0 and 360°**
  - in many cases, considering 90° increments works okay
  - still leaves  $4^N$  sets of possible phases for N F's
- **We either need plenty of time and huge computers, or we need to use some more sophisticated methods of guessing!**

# Phase choice in a centrosymmetric crystal

HOW DO I  
COMBINE THE  
ELECTRON-DENSITY  
WAVES?



“Crystal Structure Analysis for  
Chemists and Biologists”, Glusker,  
Lewis and Rossi, VCH, 1994.

# Electron density equations

- We already established that

$$\rho(x, y, z) = \frac{1}{V} \sum_{\substack{hkl \\ -\infty \\ \infty}} e^{-i2\pi(hx+ky+lz)} \mathbf{F}(hkl)$$

- This can be rewritten as

$$\begin{aligned} \rho(x, y, z) &= \frac{1}{V} \sum_{hkl} |F(hkl)| \cdot e^{-i2\pi(hx+ky+lz)} \cdot e^{i\phi(hkl)} \\ &= \frac{1}{V} \sum_{hkl} |F(hkl)| \cdot \cos[2\pi(hx + ky + lz) - \phi(hkl)] \end{aligned}$$

- Ultimately, the electron density must be real (physical quantity)!



# Friedel's law

- Under normal circumstances, any X-ray diffraction pattern will be centrosymmetric, regardless of crystal class
- This is known as *Friedel's law*, which is often written as

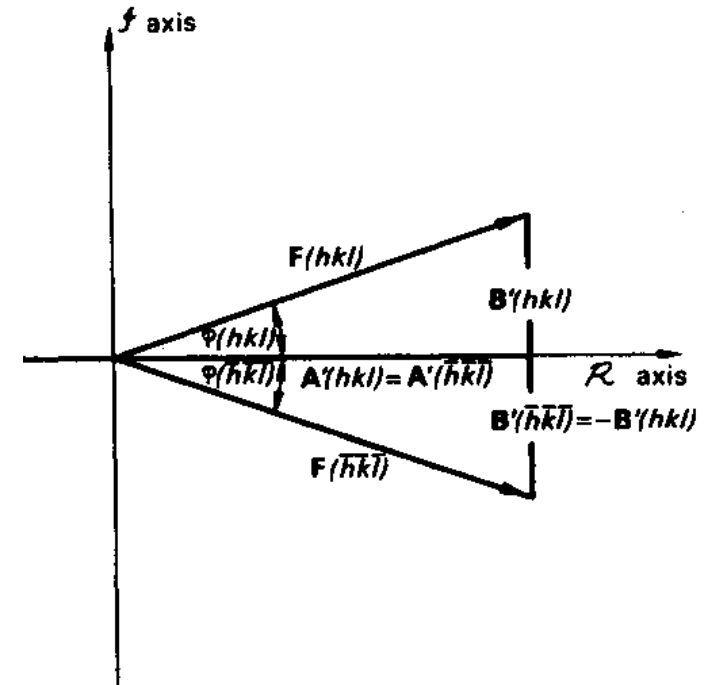
$$I(hkl) = I(\overline{h}\overline{k}\overline{l})$$

- Remember that

$$|F(hkl)|^2 = A(hkl)^2 + B(hkl)^2$$

- This leads to

$$\rho(x, y, z) = \frac{2}{V} \sum_{\substack{hkl \\ 0}}^{\infty} |F(hkl)| \cdot \cos[2\pi(hx + ky + lz) - \phi(hkl)]$$



# Electron density maps

- Electron density maps are most commonly drawn as contour maps where the lines represent traces of equal electron density
  - sometimes drawn as projections
  - often drawn as cross sections so that 3D space can be covered
  - the higher the peak, the more electrons the atom has
  - this is what you will be analyzing during a single crystal analysis... but how do we get it?

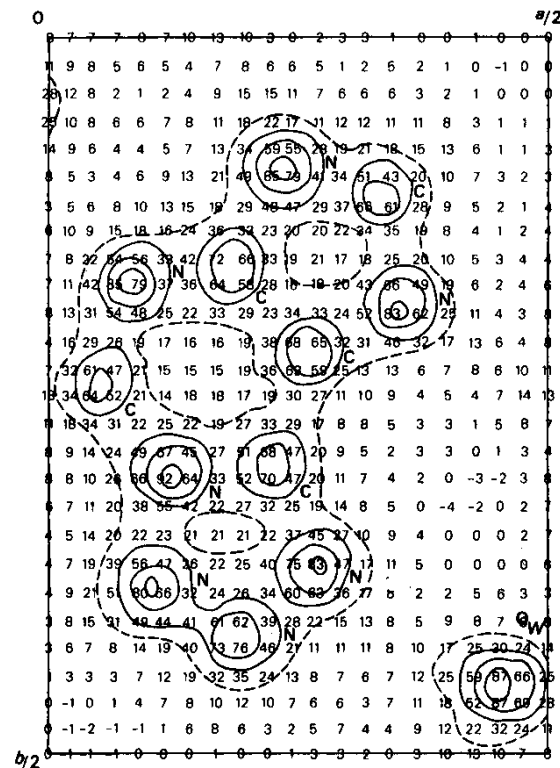


FIGURE 6.11. Two-dimensional electron density projection  $\rho(x, y)$  for azidopurine monohydrate,  $C_5H_3N_7 \cdot H_2O$  [calculated from the data of Glusker *et al.*, *Acta Crystallographica B* **24**, 359 (1968)]. The isolated peak ( $O_w$ ) in the lower right-hand region of the map represents the oxygen atom of the water molecule. Hydrogen atom positions are not obtained in this electron density synthesis (see page 314). The field figures are  $10\rho(x, y)$  in electrons per  $\text{\AA}^2$  contoured at intervals of 20 units.

“Structure Determination by X-ray Crystallography”, Ladd and Palmer, Plenum, 1994.



# The Patterson function

- The Fourier transform relationship between crystal and diffraction space was already recognized in 1915, but the lack of computing facilities combined with the phase problem made extensive Fourier calculations unpopular
- In 1934, Patterson introduced a new Fourier series which could be calculated directly from experimental data
  - lack of phases means that this series does not give atomic positions directly
  - instead, interatomic distances can be seen in the map

# The Patterson function

- The series is a convolution of two electron density functions

$$P(u, v, w) = \frac{2}{V} \sum_{\substack{hkl \\ 0}}^{\infty} |F(hkl)|^2 \cdot \cos[2\pi(hu + kv + lw)]$$

- For one dimension:

$$P(u) = \frac{2}{a} \sum_{\substack{h \\ 0}}^{\infty} |F(h)|^2 \cdot \cos 2\pi hu$$

- Derived from:

$$A(u) = \int_0^1 \rho(x)\rho(x+u)dx$$

# A 1D example

- A Patterson map will *always* be centrosymmetric, regardless of crystal symmetry!
  - this is very different from electron density maps, which will reflect the crystal symmetry precisely
- There will be  $N^2$  peaks in a Patterson map for  $N$  atoms
- Peak heights are proportional to  $Z \cdot Z'$ , so heavy atoms stand out
  - $Z, Z'$  refer to atomic number!
- Highest peak is *always* at origin

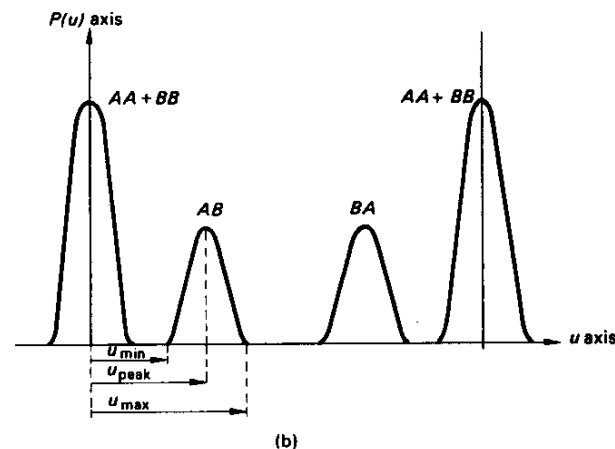
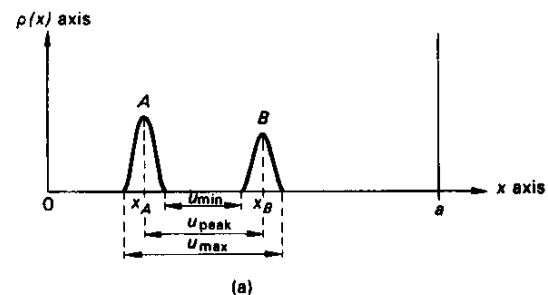
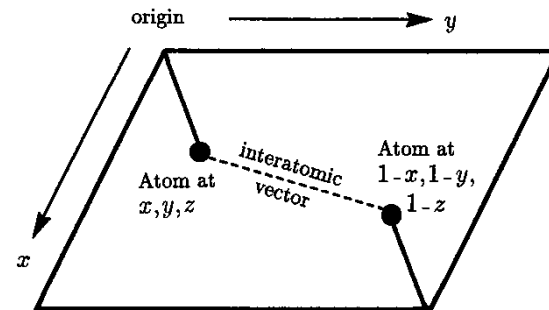
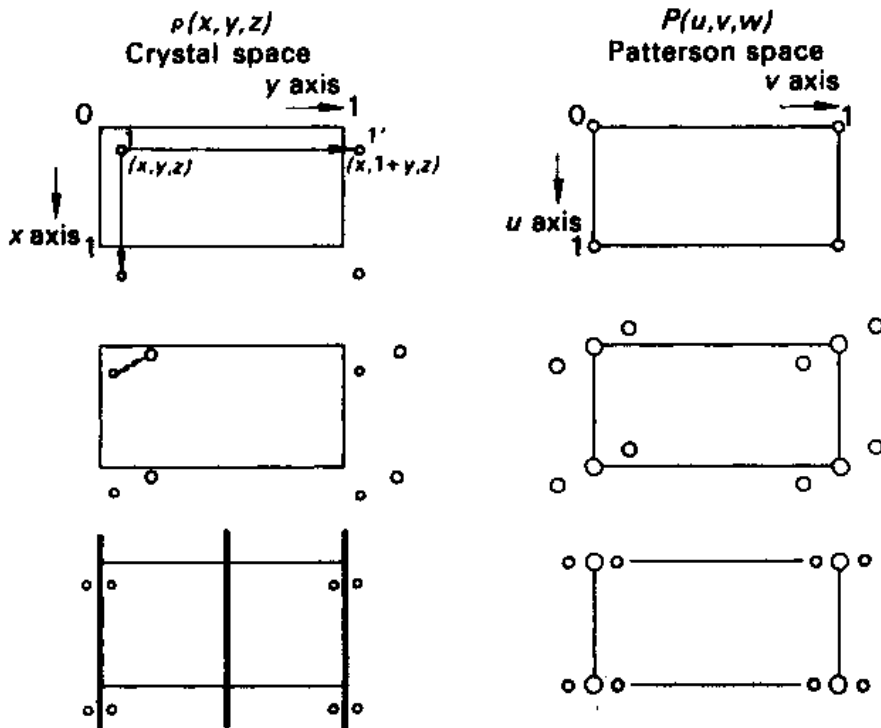


FIGURE 6.13. Development of a one-dimensional Patterson function (b) for a two-atom structure (a). Note the centrosymmetry of the Patterson function that is lacking in the  $\rho(x)$  function.

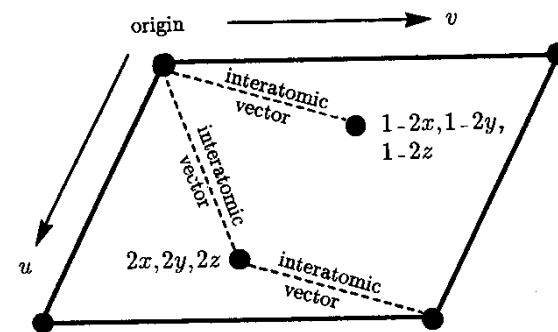
“Structure Determination by X-ray Crystallography”, Ladd and Palmer, Plenum, 1994.



# 2D examples



(a)



(b)

FIGURE 8.13. Interatomic vectors in a centrosymmetric structure. (a) The atomic arrangement and (b) its Patterson map.

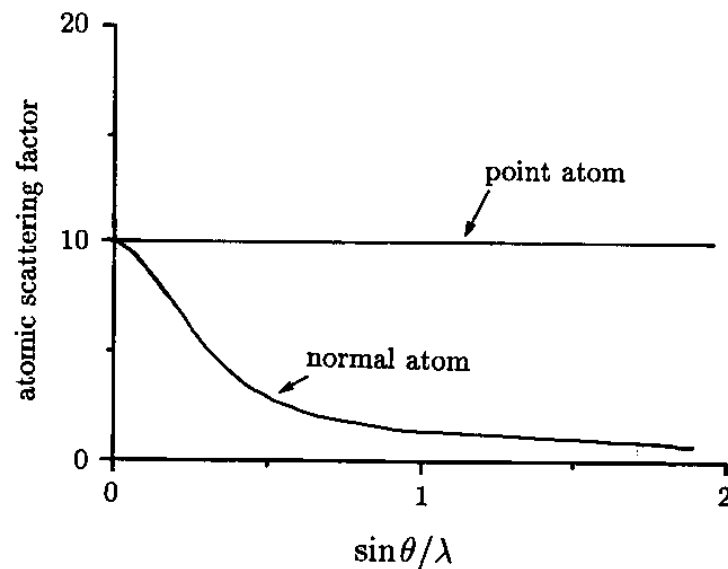
“Structure Determination by X-ray Crystallography”, Ladd and Palmer, Plenum, 1994.

“Crystal Structure Analysis for Chemists and Biologists”, Glusker, Lewis and Rossi, VCH, 1994.



# E-values

- Many calculations during structure solution processes are carried out using normalized structure factors (E-values)
- Removes thermal motion and form factor fall off
  - point atom behavior
- $|E(hkl)| = |F(hkl)| / \epsilon[\sum f_j^2(hkl)]^{1/2}$ 
  - where  $f_j = f_{j0} \exp[-B_j(\sin^2\theta/\lambda^2)]$



“Crystal Structure Analysis for Chemists and Biologists”, Glusker, Lewis and Rossi, VCH, 1994.

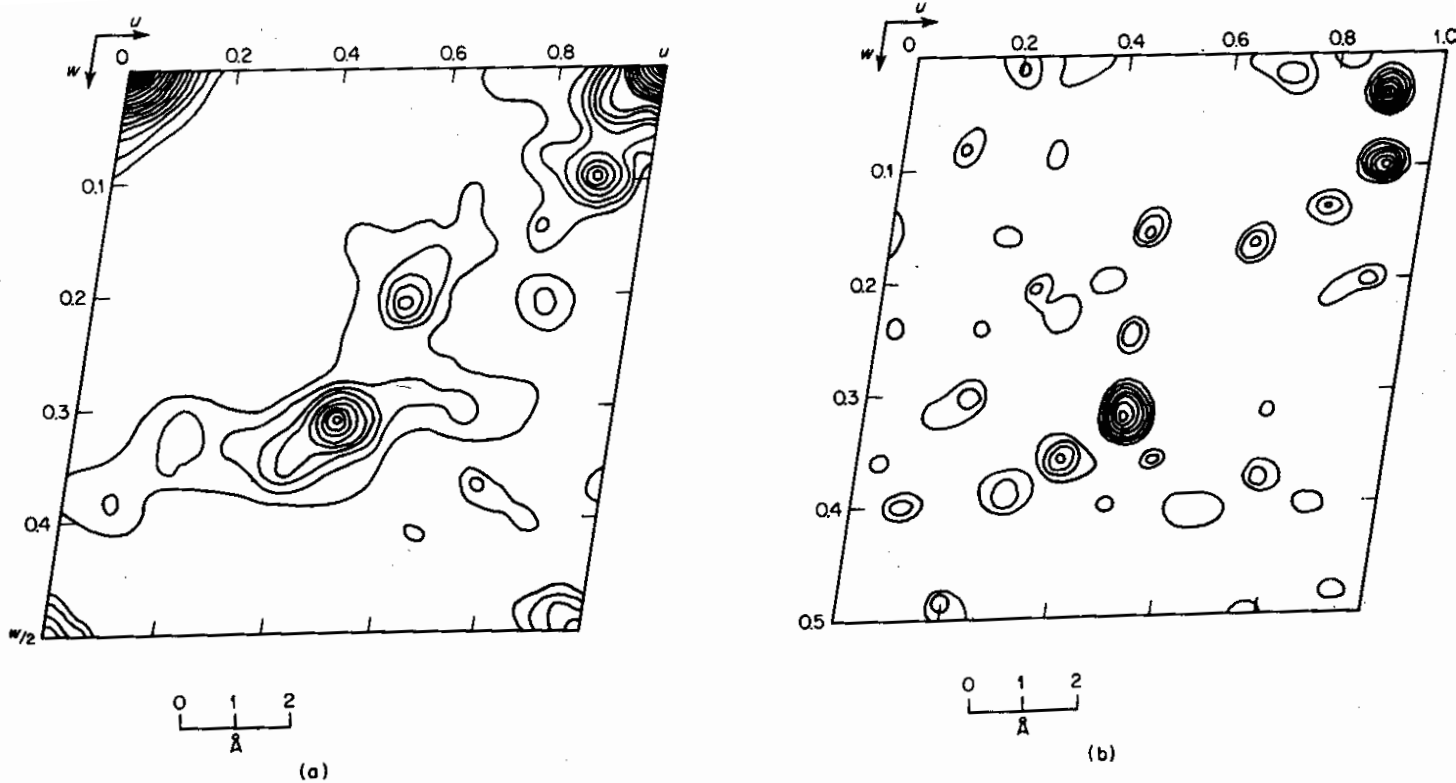


# Different types of Patterson maps

- E-values are frequently used to emphasize features in the map
- $|E|^2$  and  $|E^2-1|$  maps are common
- $|E|^2$  gives super sharpened maps
- $|E^2-1|$  gives maps without origin peak
- Over-sharpening can result in spurious peaks from truncation errors



# Effects of sharpening and origin removal



**Figure 12.3.** (a) Asymmetric part of the  $u_0w$  section of the Patterson map of a natural product. (b) Same section after sharpening and removal of the origin peak.

# Harker lines and planes

- Peaks in Patterson maps often concentrate along lines or on planes
- These Harker lines and planes can be used to determine the space group symmetry
  - lines and planes arise from symmetry equivalent atoms

# Harker lines

- **Example: Space group P m**

- $x, y, z$  and  $x, -y, z$  are symmetry equivalent
  - there will be peaks at  $0, 2y, 0$  in the Patterson
  - along the Harker line  $0, v, 0$

- **Example: Space group P c**

- $x, y, z$  and  $x, -y, z+1/2$  are symmetry equivalent
  - Patterson peaks will occur at  $0, 2y, 1/2$
  - along the Harker line  $0, v, 1/2$

# Harker planes

- **Example: Space group  $P 2_1$**

- $x, y, z$  and  $-x, y+1/2, -z$  are symmetry equivalent
  - peaks at  $2x, 1/2, 2z$  in the Patterson map
  - on the Harker plane  $u, 1/2, w$

- **Example: Space group  $P 2$**

- $x, y, z$  and  $-x, y, -z$  are symmetry equivalent
  - Patterson peaks will occur at  $2x, 0, 2z$
  - on the Harker plane  $u, 0, w$

# Using Patterson maps

- **Patterson maps are usually used for heavy atom location**
  - heavy atoms stick out
- **After a heavy atom is located, phases can be estimated for each reflection from its  $x, y, z$  coordinates**
  - assumption: Heavy atom dominates scattering
  - use phases to calculate a density map and look for residual peaks to locate lighter atoms
- **Does not work well for complicated structures, but is a common approach for small molecules (especially metal oxides etc.)**

# Locating more atoms

- **After locating some initial atoms, Fourier difference maps can be used to find missing atoms**
  - this is almost always required for structure completion, regardless of method used
  - all that differs between methods is how the initial atoms are located
- **Often referred to as  $|F_o| - |F_c|$  map**

# $F_{\text{obs}}$ and $F_{\text{calc}}$ maps

- $F_{\text{obs}}$  maps are calculated using experimental  $|F|$  together with phases calculated from some structural model
  - can show peaks that do not correspond to atoms in the current model
  - allows location of missing atoms
- $F_{\text{calc}}$  maps are constructed from phases and amplitudes that are calculated from a structural model
  - map only has peaks corresponding to atom positions in the model used for calculation
  - not very informative

# Difference maps

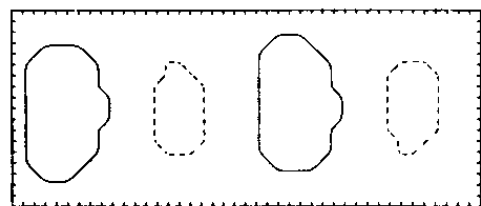
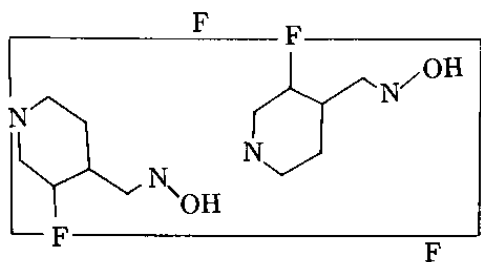
- $|F_{\text{obs}}| - |F_{\text{calc}}|$  (or  $|F_{\text{o}}| - |F_{\text{c}}|$ )
- Measured and calculated  $F$ 's are combined with calculated phases
- The difference map emphasizes incorrect features of the model used
  - peaks indicate missing atoms
  - valleys correspond to incorrectly placed atoms



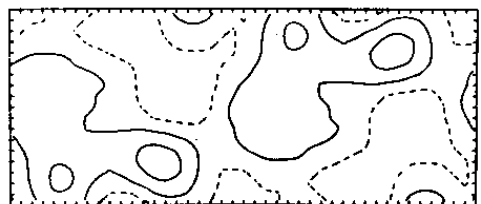
# Map resolution

- A complete Fourier synthesis would involve all  $F(hkl)$  values (summation from  $-\infty$  to  $\infty$ )
- In reality, only a finite amount of data is available
- Neglecting high resolution data (e.g.,  $F(hkl)$  with small d-spacings) has **two effects**:
  - structural detail cannot be seen
  - unwanted ripples in the map can arise due to termination errors

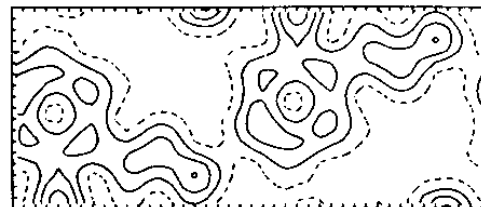
# Map resolution – an example



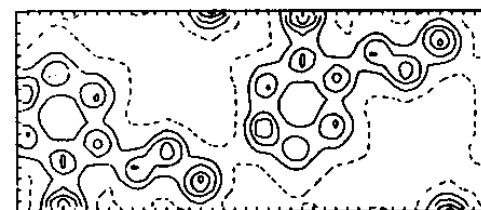
6 Å data



2.5 Å data



1.5 Å data



0.8 Å data

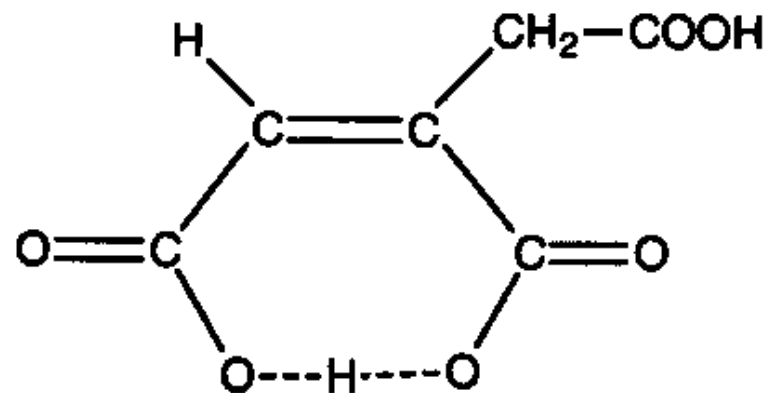
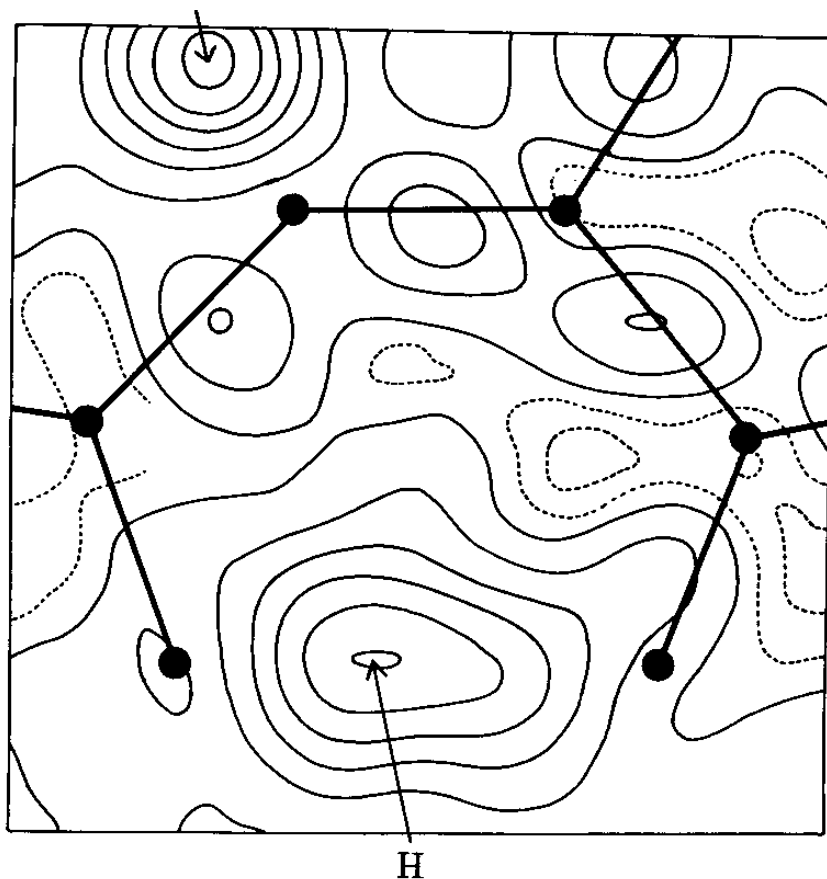
**FIGURE 9.10.** Different resolutions of the same structure. These diagrams were obtained by only including terms that diffract to the quoted resolution.

“Crystal Structure Analysis for Chemists and Biologists”, Glusker, Lewis and Rossi, VCH, 1994.

# Location of light atoms

- **Even in difference maps, light atoms can be difficult to locate**
  - depends also on how much lighter than the rest of the structure they are
- **In many cases, it is not possible to find hydrogen atoms in electron density maps created from X-ray diffraction data**
- **Neutrons can help to overcome this problem**
  - remember that the neutron form factors do not depend linearly on  $Z$ !

# An example of a difference map



"Crystal Structure Analysis for Chemists and Biologists", Glusker, Lewis and Rossi, VCH, 1994.

# Isomorphous replacement

- Consider a structure that contains only one heavy atom
  - $\mathbf{F}_1 = \mathbf{F}_{M1} + \mathbf{F}_R$
- If we can prepare two structures that are identical except for the identity of the heavy atom
  - $\mathbf{F}_2 = \mathbf{F}_{M2} + \mathbf{F}_R$
- We can calculate a difference map:
  - $\mathbf{F}_1 - \mathbf{F}_2 = \mathbf{F}_{M1} - \mathbf{F}_{M2}$

# Isomorphous replacement

- We can locate the heavy atoms using a Patterson map
  - this gives us phase and amplitude of  $\mathbf{F}_{M1}$  and  $\mathbf{F}_{M2}$
- We know  $|F_1|$  and  $|F_2|$  from our diffraction experiment
- For a centrosymmetric structure ( $\phi$  can only be 0 or  $180^\circ$ ), we can uniquely determine the phases of  $F_1$  and  $F_2$ 
  - the phase problem reduces to a sign problem (+ or -) in centrosymmetric structures

# Example calculation for a centrosymmetric structure

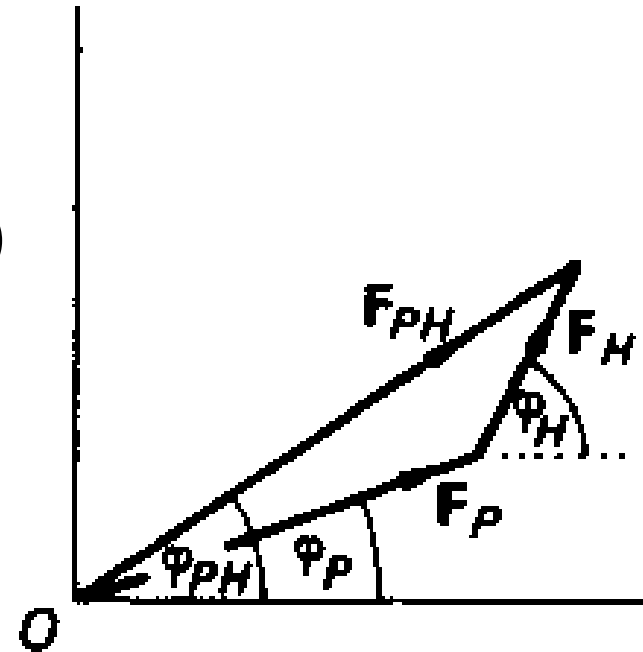
	$F_{M1}$ $-F_{M2}$	$ F_1 $	$ F_2 $	$ F_2  >$ $ F_1  ?$	Possibilities	Deductions on sign of $F_1$ and $F_2$
(a)	+3	9	12	yes	$(+3) + (+9) = (+12)$	+ +
(b)	-3	9	6	no	$(-3) + (+9) = (+6)$	+ +
(c)	-3	9	12	yes	$(-3) + (-9) = (-12)$	- -
(d)	+3	9	6	no	$(+3) + (-9) = (-6)$	- -

"Crystal Structure Analysis for Chemists and Biologists", Glusker, Lewis and Rossi, VCH, 1994.

# Use of isomorphous replacement for macromolecules

- Isomorphous replacement is commonly used for solving protein structures
- Native proteins contain a lot of water
- Replace some water with a heavy atom (e.g.,  $K^+$ )
- If things go well, you now have a native protein and a heavy atom derivative that are isomorphous

○  $F_{PH} = F_P + F_H$



“Structure Determination by X-ray Crystallography”,  
Ladd and Palmer, Plenum, 1994.





# Phase ambiguity with one derivative

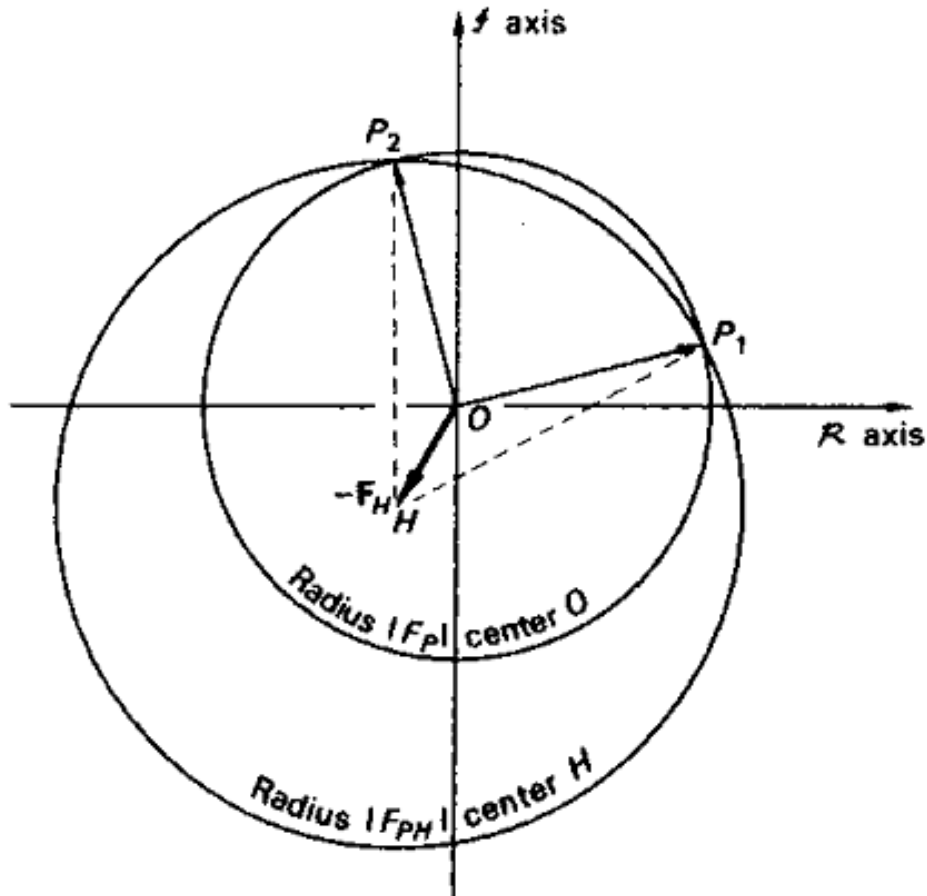


FIGURE 6.33. Single isomorphous replacement (SIR) phase-amplitude diagram in the single isomorphous replacement method.  $OH(-\mathbf{F}_H)$  is the known reversed heavy-atom vector. The triangles  $OHP_1$  and  $OHP_2$  both satisfy (6.81), giving a twofold ambiguity with either  $OP_1$  or  $OP_2$  as the solution for  $\mathbf{F}_p$ .

“Structure Determination by X-ray Crystallography”, Ladd and Palmer, Plenum, 1994.



# Using multiple derivatives

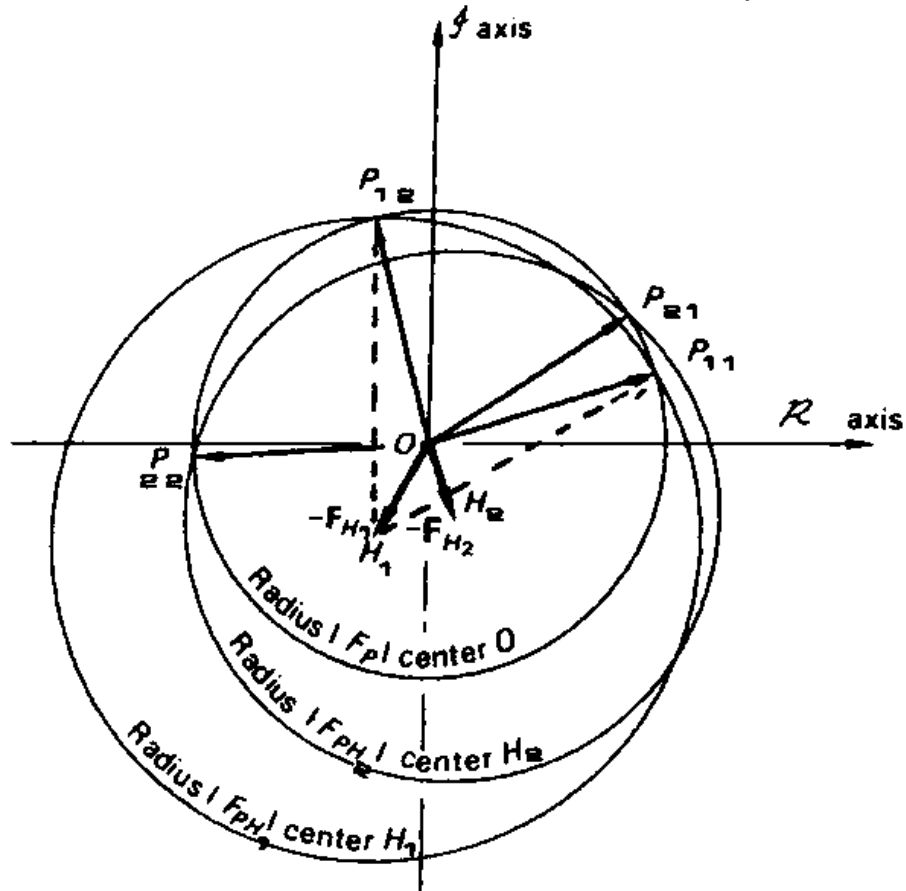


FIGURE 6.34. Multiple isomorphous replacement (MIR) resolution of the phase ambiguity with a second isomorphous heavy-atom derivative  $PH_2$ . The determined direction of  $F_P$  is near  $OP_{11}$  and  $OP_{21}$ . In practice,  $P_{11}$  and  $P_{21}$  rarely coincide, due to inaccuracies in the heavy-atom parameters and lack of complete nisomorphism.

“Structure Determination by X-ray Crystallography”, Ladd and Palmer, Plenum, 1994.

# What if...?

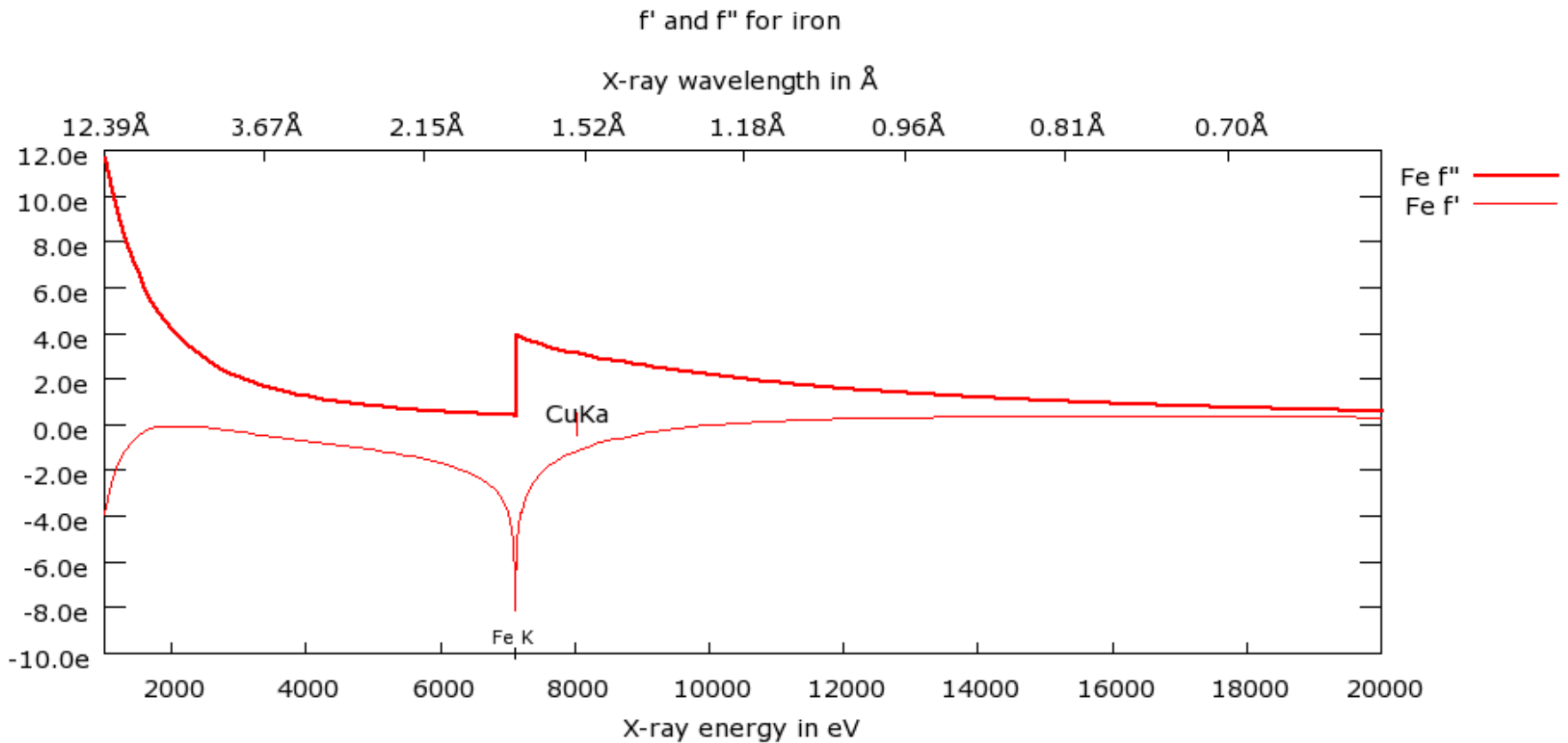
- **What if we cannot prepare several derivatives of a protein?**
  - maybe only certain metals will substitute into the structure
- **What if our “isomorphous” materials refuse to be isomorphous?**
  - changing the metal/cation could significantly alter the structure (even the crystal system)
- **What if we can put two different metals in, but they have almost the same number of electrons?**
  - or for that matter, we have a structure with neighboring elements to start with?

# Anomalous scattering

- The scattering contribution of each atom in the unit cell is represented by its form factor  $f$ 
  - so far assumed to be a function of scattering angle only
- Form factors consist of several contributions:
  - $f = f_0 + \Delta f' + i\Delta f''$
- $f'$  and  $f''$  show little dependence on  $\theta$
- $f''$  is approximately zero - *unless the wavelength is close to the absorption edge of the atom!*
  - changes the scattering and thus all phases without disturbing the structure

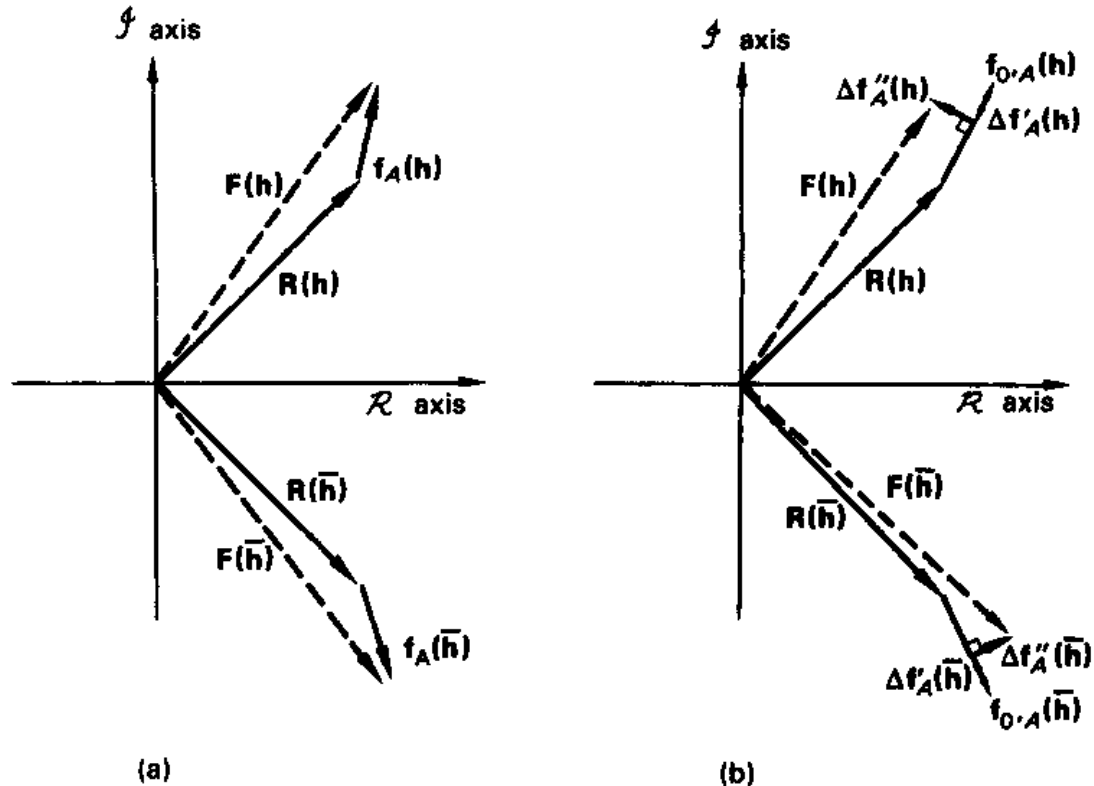
# f' and f''

Edgeplots web tool <http://www.bmsc.washington.edu/scatter/>



[http://skuld.bmsc.washington.edu/scatter/AS\\_form.html](http://skuld.bmsc.washington.edu/scatter/AS_form.html)

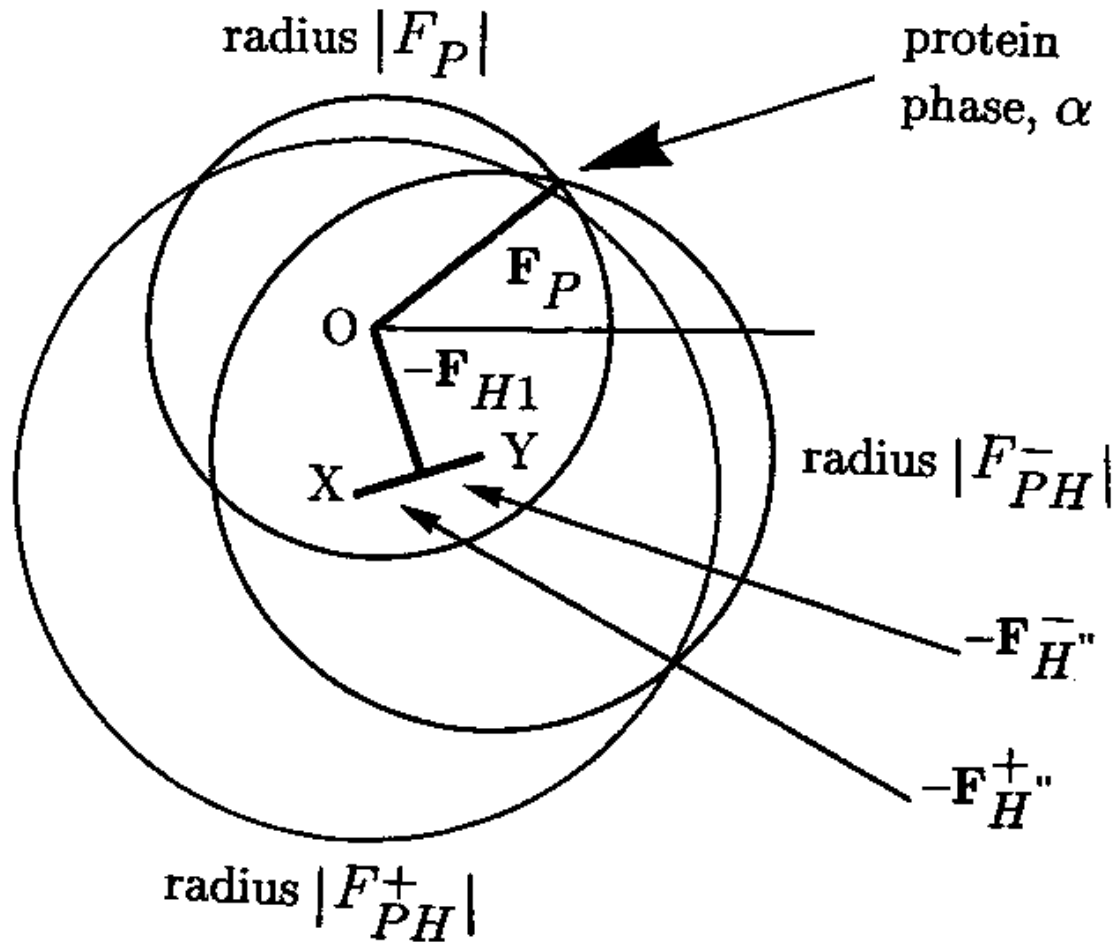
# The effect of anomalous scattering



**FIGURE 6.45.** Anomalous scattering of atom  $A$  with respect to the rest of the structure  $R$ :  
 (a) normal case— $|F(\mathbf{h})| = |F(\bar{\mathbf{h}})|$ ; (b) anomalous case— $|F(\mathbf{h})| \neq |F(\bar{\mathbf{h}})|$ .

“Structure Determination by X-ray Crystallography”,  
 Ladd and Palmer, Plenum, 1994.

# Phases from anomalous scattering



“Crystal Structure Analysis for Chemists and Biologists”, Glusker, Lewis and Rossi, VCH, 1994.

# Direct methods

- **Common for small organics or other small molecules with little scattering contrast between atoms**
  - e.g.,  $\text{AlF}_3$ :  $\text{Al}^{3+}$  and  $\text{F}^-$  have identical electron configurations
- **Phase probability methods**
  - makes use of non-negativity constraint on electron density
  - actively uses the constraint that the electron density should be pooled in atoms
- **Uses normalized structure factors**



# Symmetry information from $|E|$ statistics

- The statistical distribution of  $E$  values can yield information about crystal symmetry

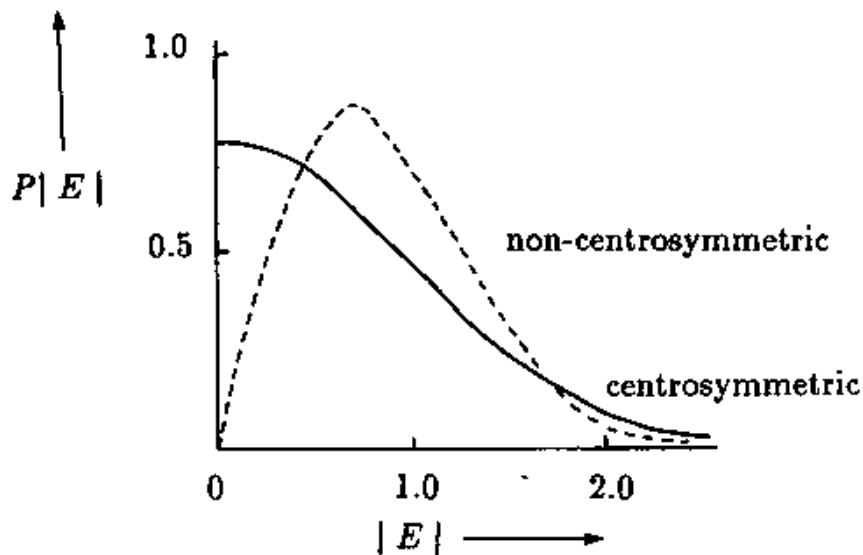


TABLE 7.2. Some Theoretical and Experimental Values Related to  $|E|$  Statistics

Mean values	Theoretical values		Experimental values and conclusions	
	$P(\bar{C})$	$P(A)$	Crystal 1	Crystal 2
$ E ^2$	1.00	1.00	0.99	0.98
$\frac{ E }{ E ^2 - 1}$	0.80	0.89	0.85 <i>A/C</i>	0.84 <i>A/C</i>
$  E ^2 - 1 $	0.97	0.74	0.91 <i>C</i>	0.82 <i>A</i>
Distribution	%	%	%	%
$ E  > 3.0$	0.30	0.01	0.20 <i>C</i>	0.05 <i>A</i>
$ E  > 2.5$	1.24	0.19	0.90 <i>C</i>	0.98 <i>C</i>
$ E  > 2.0$	4.60	1.80	2.70 <i>A/C</i>	2.84 <i>A/C</i>
$ E  > 1.75$	8.00	4.71	7.14 <i>C</i>	6.21 <i>A/C</i>
$ E  > 1.5$	13.4	10.5	12.9 <i>C</i>	10.5 <i>A</i>
$ E  > 1.0$	32.0	36.8	33.7 <i>C</i>	37.1 <i>A</i>

"Crystal Structure Analysis for Chemists and Biologists", Glusker, Lewis and Rossi, VCH, 1994.

"Structure Determination by X-ray Crystallography", Ladd and Palmer, Plenum, 1994.



# Choice of origin

- **All phase angles are defined relative to an origin, which can be chosen arbitrarily**
  - changing the origin will only change the phase of  $F_{hkl}$ , but not the value of  $|F_{hkl}|$
- **Assigning phases to three reflections will fix the origin**
  - for non-centrosymmetric structures, a fourth reflection is used to fix the handedness
- **The reflections that were initially assigned phases can be used as a basis set for the phase probability methods**
  - use strong reflections for better/more reliable statistics

# Triplets

- In 1952, Sayre showed that for centrosymmetric structures

$$s(hkl) \cdot s(h'k'l') \cdot s(h-h', k-k', l-l') \approx +1$$

- $s(hkl)$  corresponds to the sign (phase angle 0 or 180°) of each hkl-reflection
- $\approx$  represents the underlying probabilistic approach and should be read as “is likely to be”
- **Three reflections with hkl values related by this relationship are called a triplet**
  - the corresponding  $\mathbf{d}^*$  vectors form a triangle
- **If two signs are known, the third can be deduced**

# $\Sigma_2$ formula

- Hauptman and Karle introduced a more general form of the triplet formula in 1953:

$$s(hkl) \approx \sum_{h'k'l'} s(h'k'l') \cdot s(h-h', k-k', l-l')$$

- this formula uses all vector pairs with known signs that fulfill the triplet relationship
  - can be used for probability calculations
- For non-centrosymmetric structures, estimation of relative phases is also possible, but the math gets much more complicated

# An example of using the $\Sigma_2$ formula

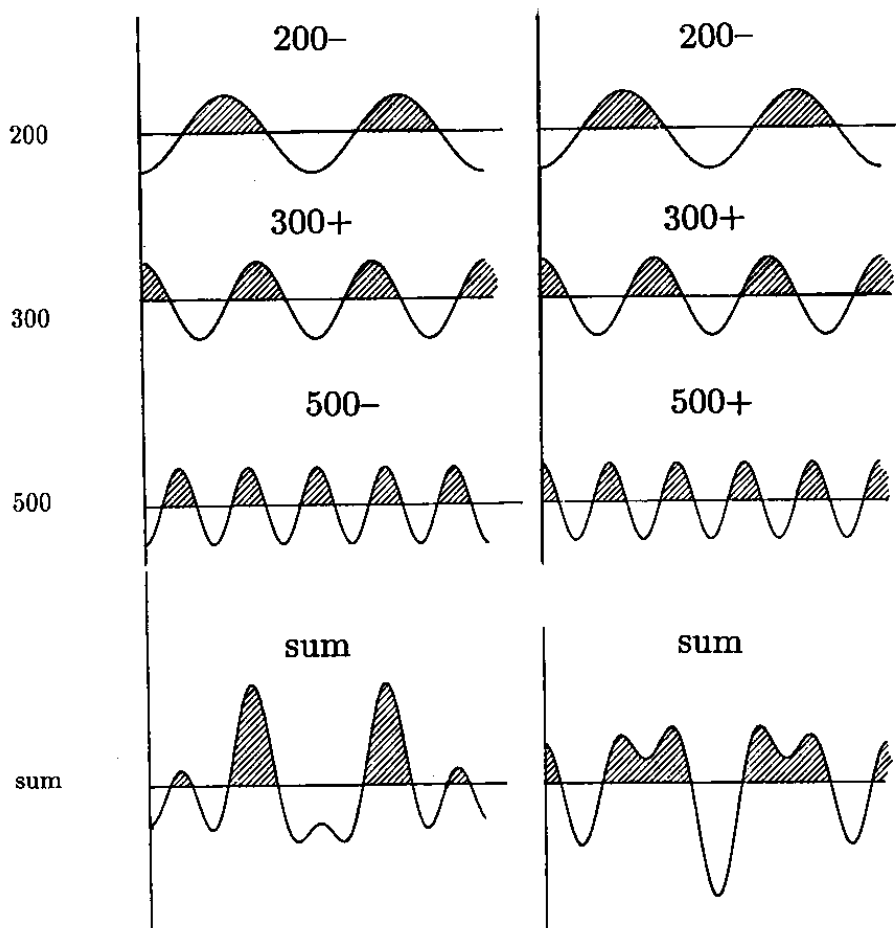


FIGURE 8.8. Direct phase determination for a centrosymmetric crystal structure. Two possible summations of waves are shown (read vertically). If  $|F(200)|$ ,  $|F(300)|$ , and  $|F(500)|$  are all large, they will make a significant contribution to the electron-density map. If it is known that  $|F(200)|$  has a negative sign (relative phase  $180^\circ$ ) and  $|F(300)|$  has a positive sign (relative phase  $0^\circ$ ), then it is most likely, by summing terms, that  $|F(500)|$  is negative (relative phase  $180^\circ$ ), since this gives a sum with the less negative electron-density map (compare the summations on the left and on the right). The sum (electron-density map) on the left is better because its background is less negative.

“Crystal Structure Analysis for Chemists and Biologists”, Glusker, Lewis and Rossi, VCH, 1994.