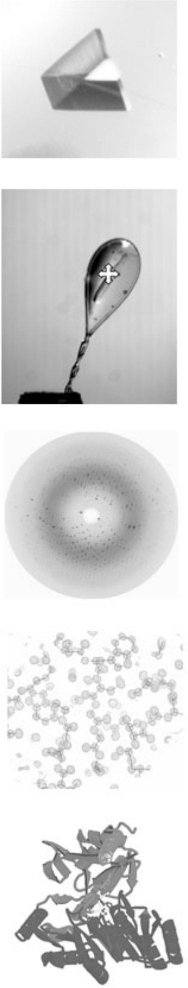
Phase Determination in Macromolecular X-ray Crystallography



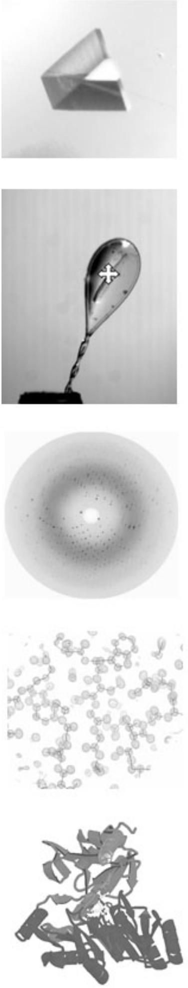
Manfred S. Weiss

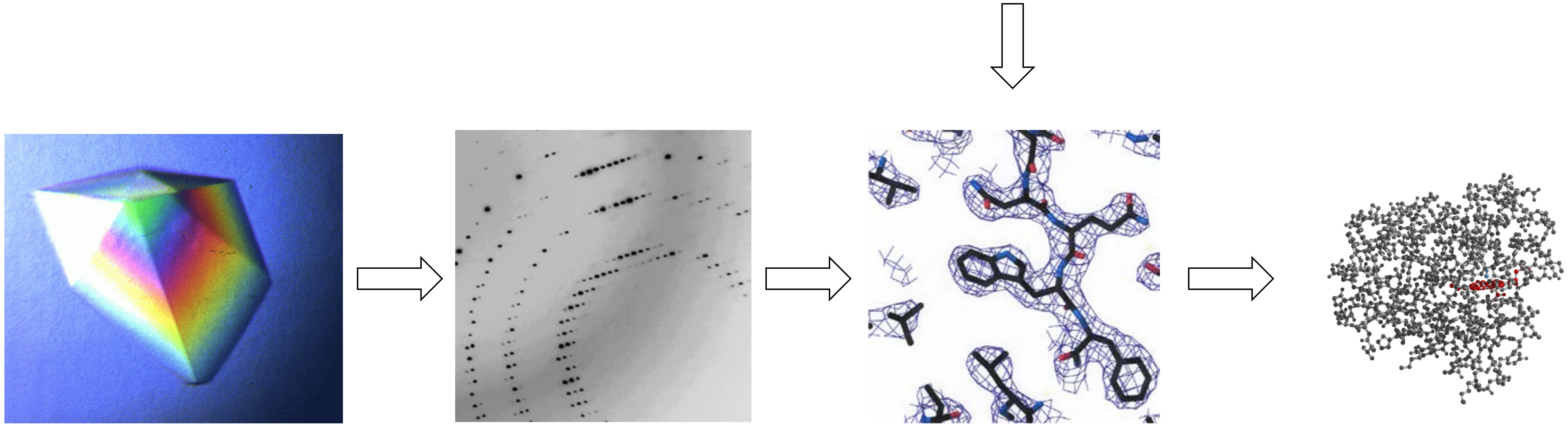
Helmholtz-Zentrum Berlin für Materialien und Energie
Macromolecular Crystallography (HZB-MX)
Albert-Einstein-Str. 15
D-12489 Berlin, Germany

Email: msweiss@helmholtz-berlin.de

@ManfredSWeiss

Structure Determination





Phases $\alpha(hkl)$

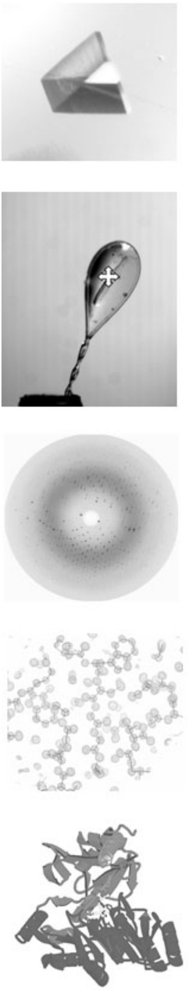
Crystal

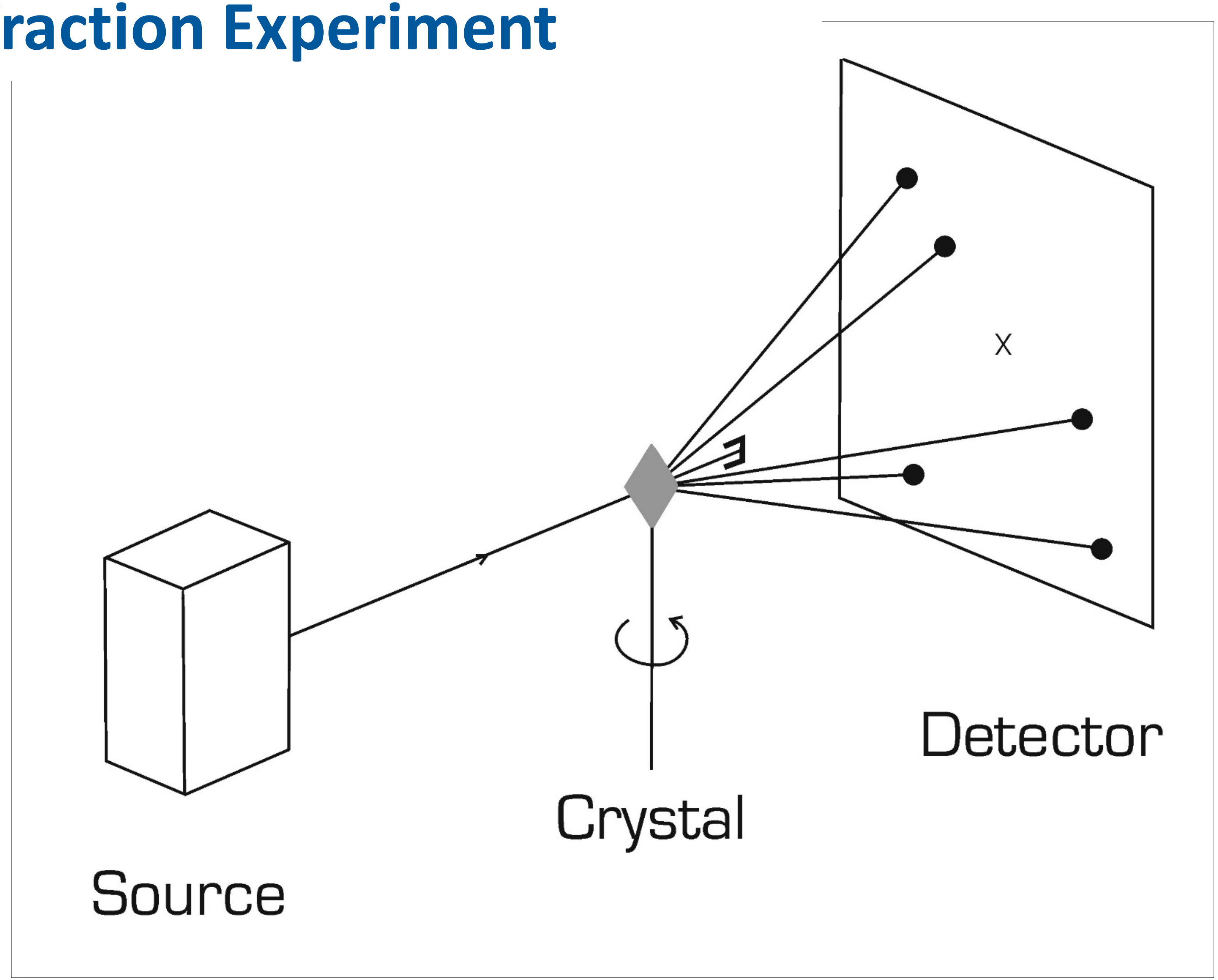
$h, k, l, I, \sigma(I)$

$\rho(x,y,z)$

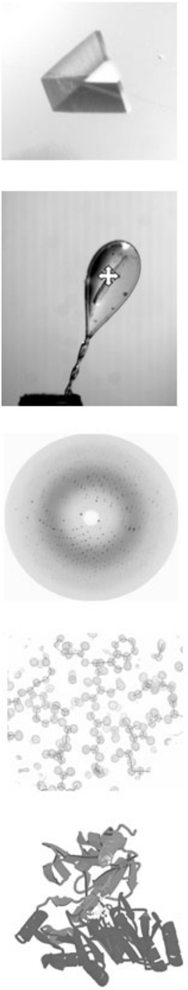
Structure

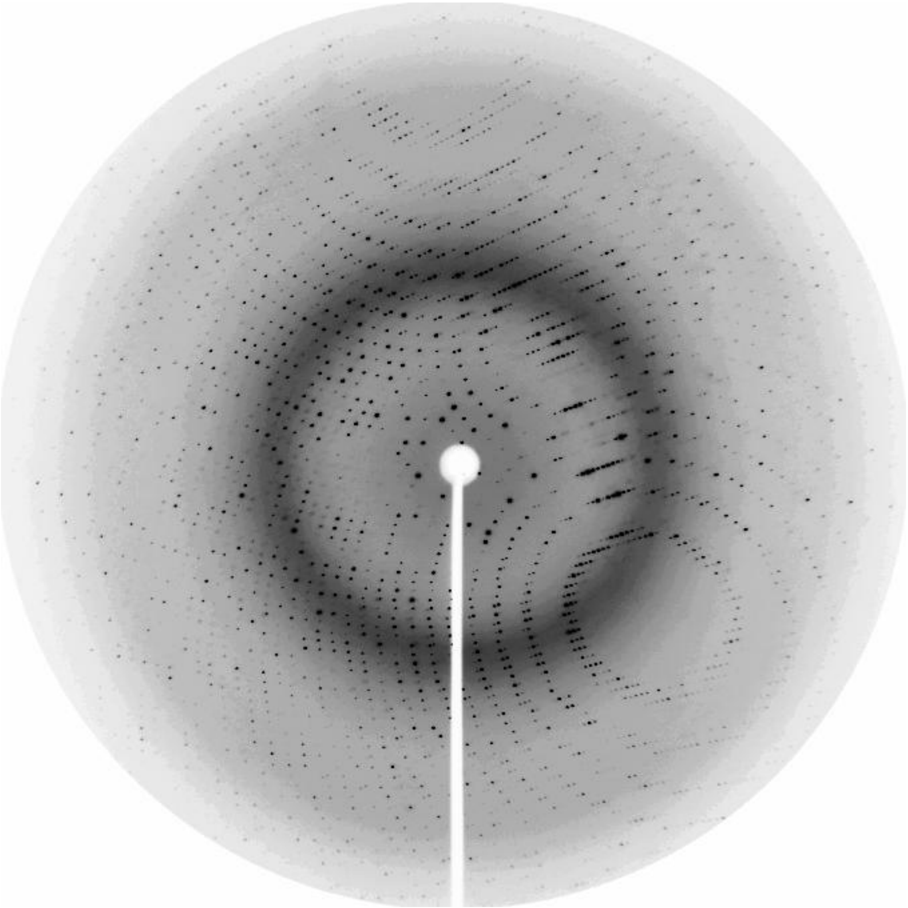
The Diffraction Experiment



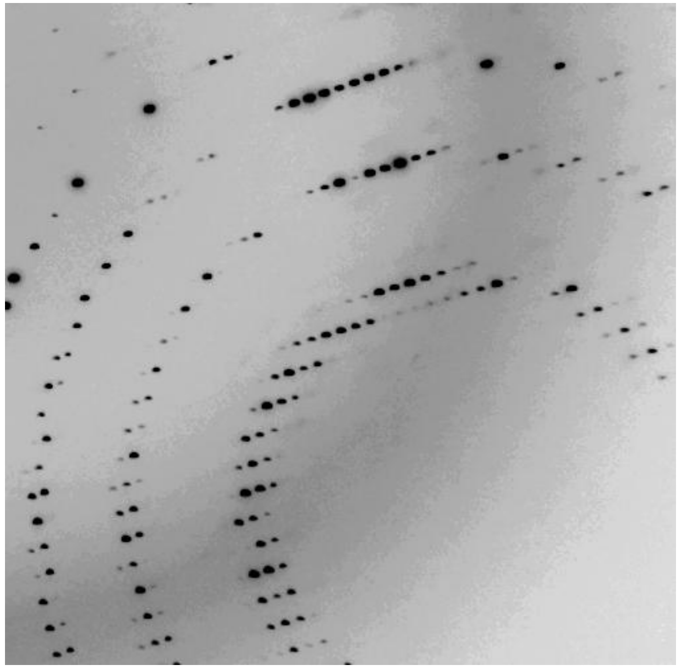


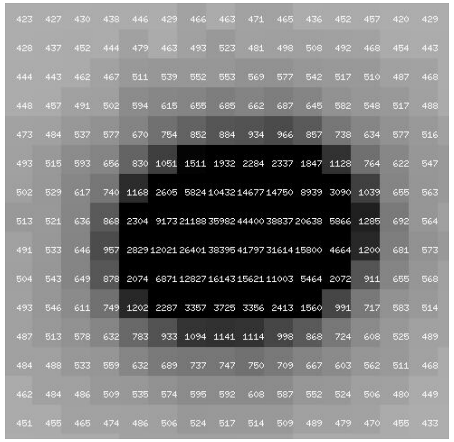
The Diffraction Experiment





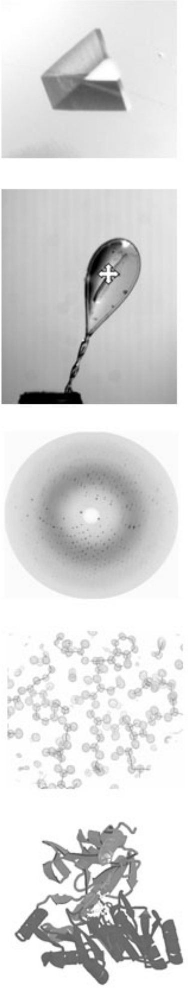
h, k, l	**Miller indices**
I(h,k,l)	**intensity**
σI(hkl)	**error in I**

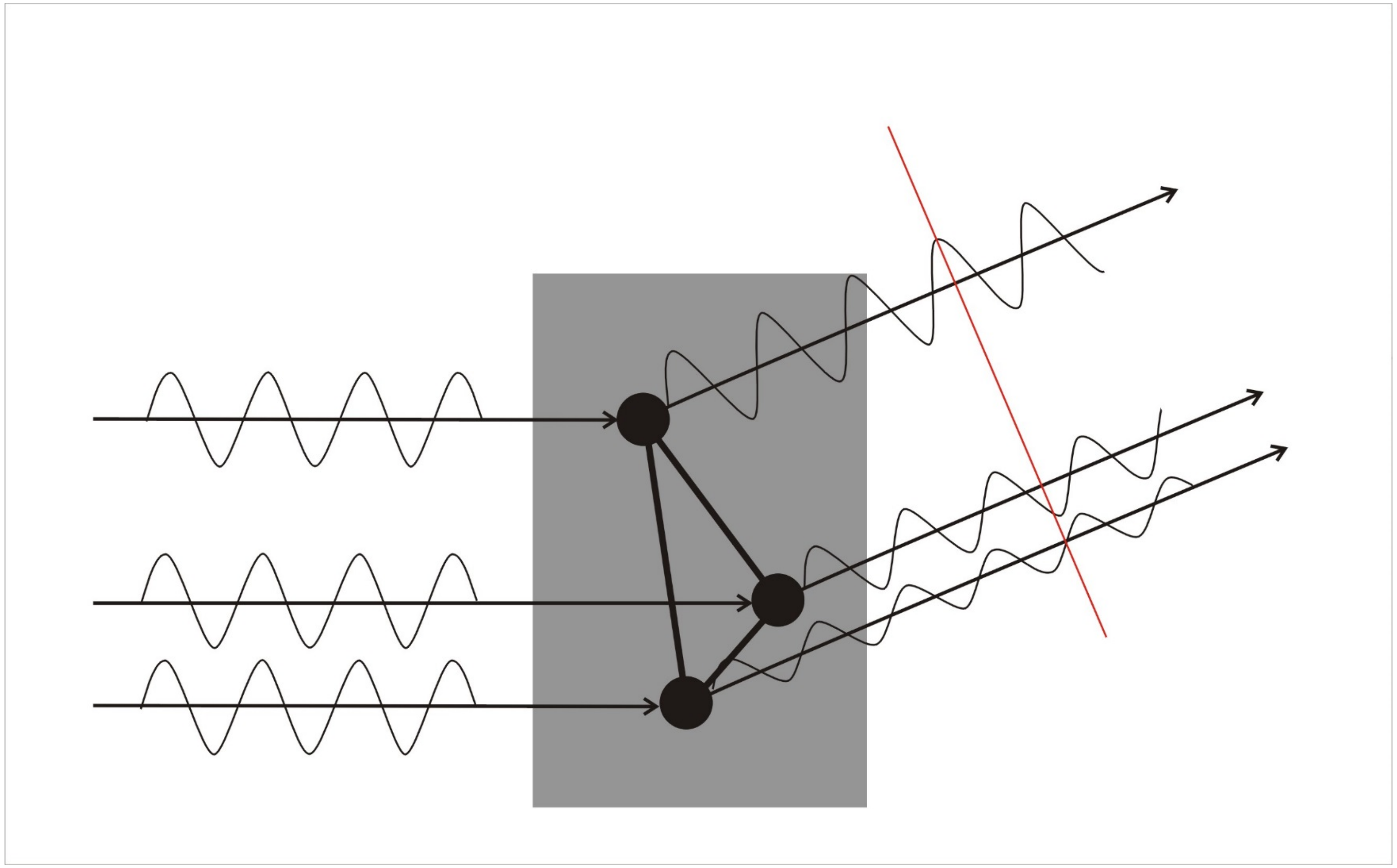




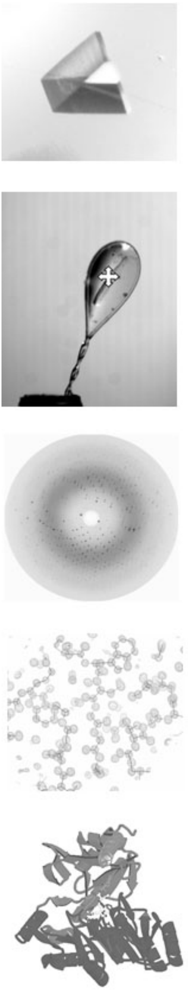
423	427	430	438	446	429	466	463	471	465	436	452	457	420	429
428	437	452	444	479	463	493	523	481	498	508	492	468	454	443
444	443	462	467	511	539	552	553	569	577	542	517	510	487	468
448	457	491	502	594	615	655	685	662	687	645	582	548	517	488
473	484	537	577	670	754	852	884	934	966	857	738	634	577	516
493	515	593	656	830	1051	1511	1932	2284	2337	1847	1128	764	622	547
502	529	617	740	1168	2605	5824	10432	14677	14750	8939	3090	1039	655	563
513	521	636	868	2304	9173	21188	35982	44400	38837	20638	5866	1285	692	564
491	533	646	957	2829	12021	26401	38395	41797	31614	15800	4664	1200	681	573
504	543	649	878	2074	6871	12827	16143	15621	11003	5464	2072	911	655	568
493	546	611	749	1202	2287	3357	3725	3356	2413	1560	991	717	583	514
487	513	578	632	783	933	1094	1141	1114	998	868	724	608	525	489
484	488	533	559	632	689	737	747	750	709	667	603	562	511	468
462	484	486	509	535	574	595	592	608	587	552	524	506	480	449
451	455	465	474	486	506	524	517	514	509	489	479	470	455	433

The Structure Factor Equation





The Structure Factor Equation

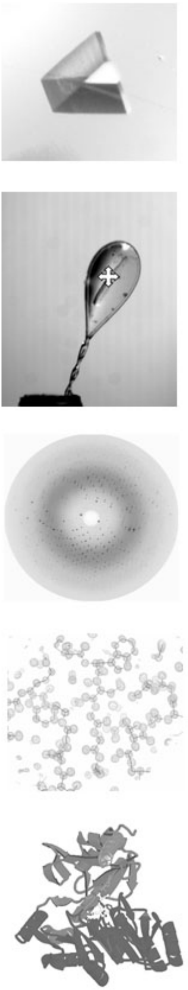


The calculation of F(hkl) from a structure (x_j,y_j,z_j) is nothing but a summation of the waves originating from each atom (j) in the direction defined by (hkl).

$$F(hkl) = \sum_j f_j \, e^{2\pi i(hx_j+ky_j+lz_j)}$$

It is important to note that every atom of the structure contributes to each and every reflection of the diffraction pattern.

The Structure Factor Equation

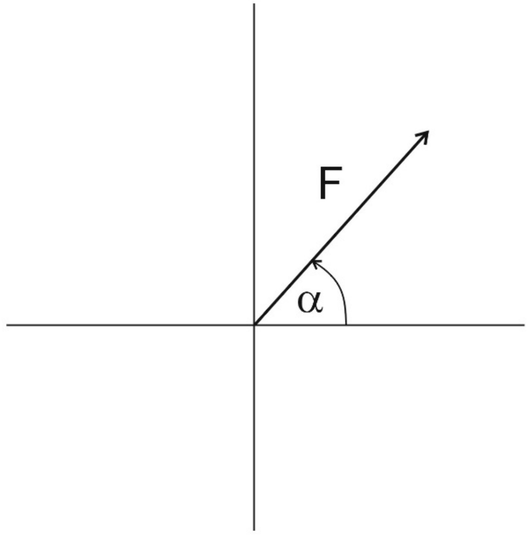


$$F(hkl) = |F(hkl)|\, e^{i\alpha(hkl)} = \sum_j f_j\, e^{2\pi i(hx_j+ky_j+lz_j)}$$

Structure factor amplitude
$|F(hkl)| \propto I(hkl)^{1/2}$

Atomic form factor f_j

Phase $\alpha(hkl)$

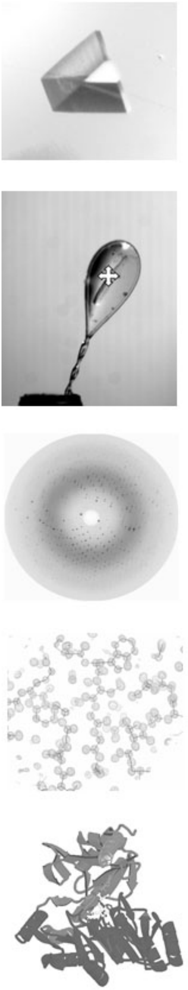


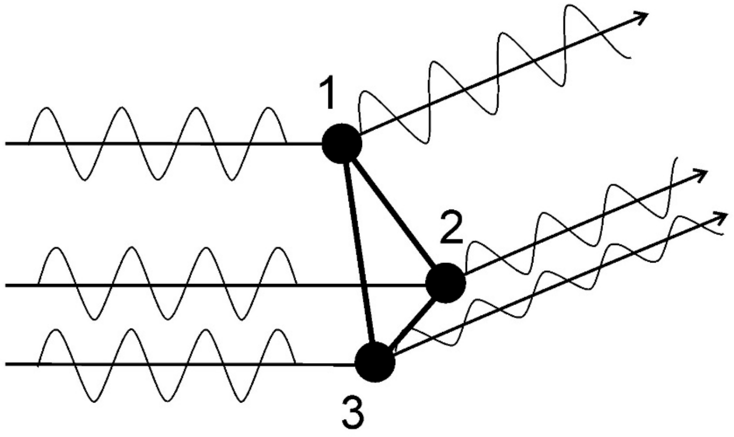
$$F\, e^{i\alpha} = A + i\,B$$

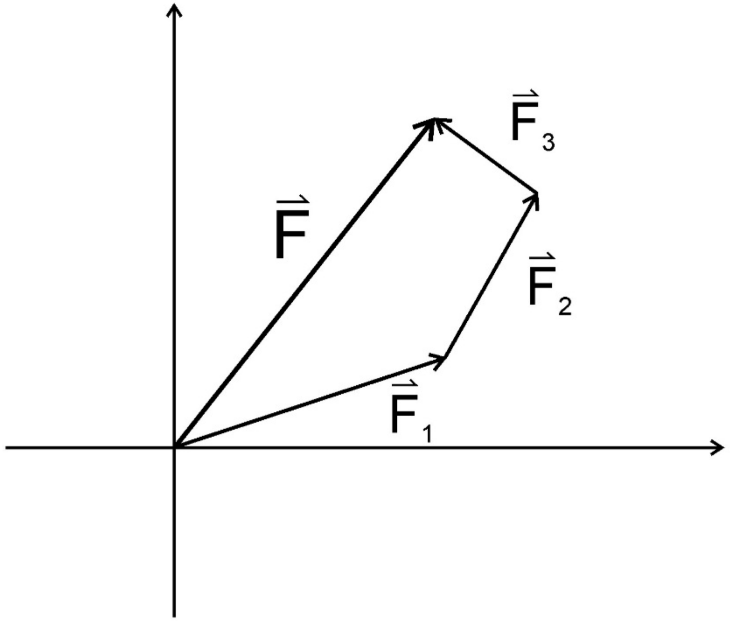
$$A = F \cos \alpha$$

$$B = F \sin \alpha$$

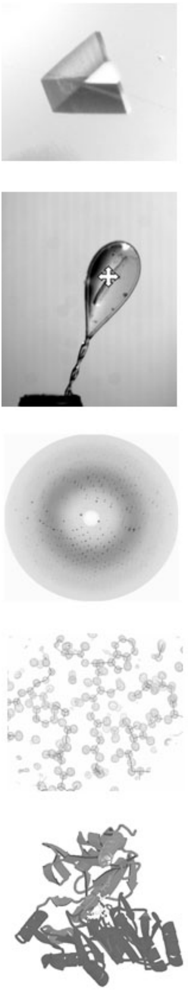
The Structure Factor Equation





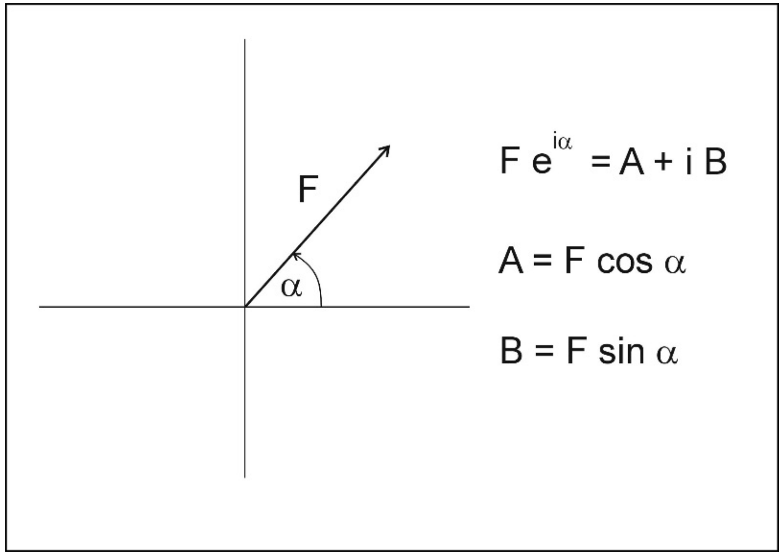


Structure Factor Equation - Summary



- If we know the structure of the protein in a given crystal represented by $(f_j, x_j, y_j, z_j, B_j)$, we can calculate the diffraction pattern exactly.

- The calculated structure factor amplitudes and phases are typically referred to as $|F_{calc}(hkl)|$ and $\alpha_{calc}(hkl)$.

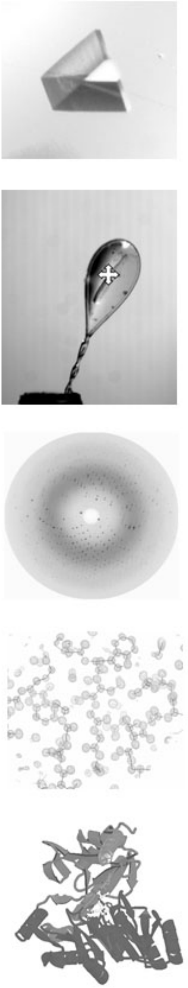


$$F\,e^{i\alpha} = A + i\,B$$

$$A = F \cos \alpha$$

$$B = F \sin \alpha$$

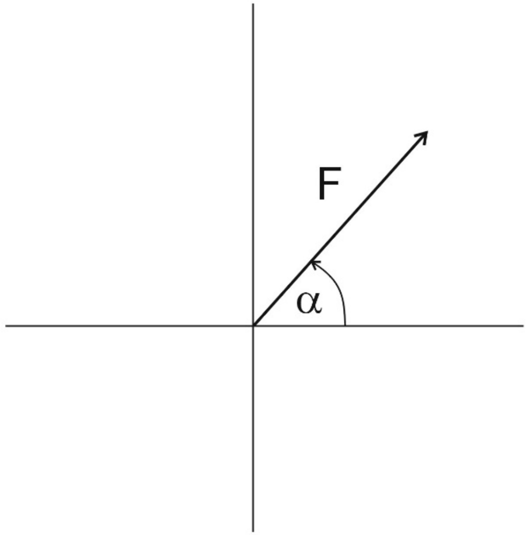
The Electron Density Equation



$$\rho(x,y,z) = 1/V \sum_{hkl} |F(hkl)| \, e^{i\alpha(hkl)} \, e^{-2\pi i(hx+ky+lz)}$$

Structure factor amplitude
$|F(hkl)| \propto I(hkl)^{1/2}$

Phase $\alpha(hkl)$

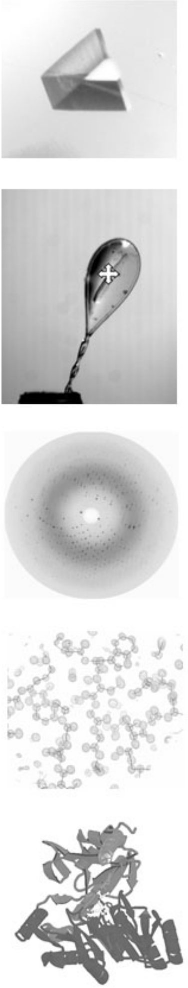


$F e^{i\alpha} = A + i B$

$A = F \cos \alpha$

$B = F \sin \alpha$

The Electron Density Equation

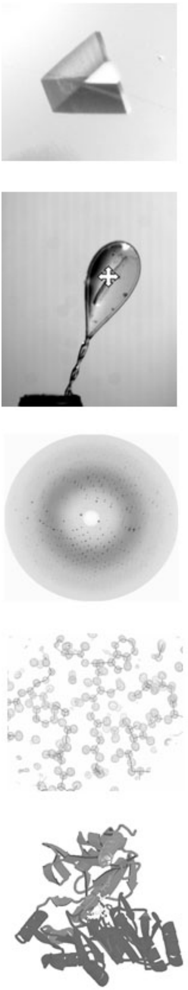


The electron density $\rho(x,y,z)$ is a three-dimensional function (with the unit e/Å^3), which describes where in the unit cell of the crystal the electrons (and therefore the atoms) are. It is basically the image of the structure we want to determine.

$$\rho(x,y,z) = 1/V \sum_{hkl} |F(hkl)| \, e^{i\alpha(hkl)} \, e^{-2\pi i(hx+ky+lz)}$$

It is important to note that every reflection (hkl) of the diffraction pattern contributes to the electron density ρ at each and every position (xyz) in the unit cell of the crystal.

The Phase Problem



From the diffraction pattern, we can only obtain the intensities I(hkl) of the reflections (hkl).

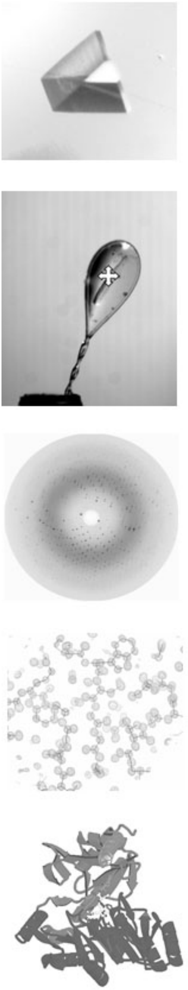
$$I(hkl) \propto F(hkl) \cdot F^*(hkl)$$

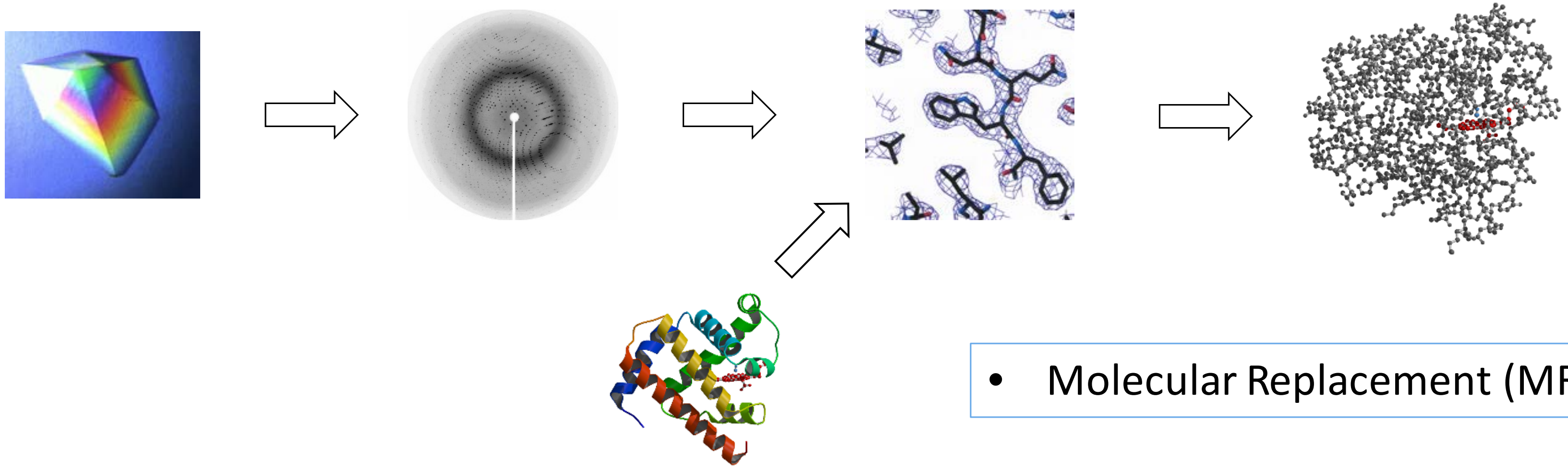
$$= |F(hkl)|\, e^{i\alpha(hkl)} \cdot |F(hkl)|\, e^{-i\alpha(hkl)}$$

$$= |F(hkl)|^2$$

The phases $\alpha(hkl)$ cannot be measured.

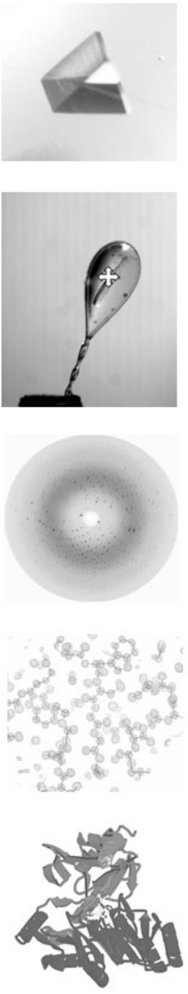
Structure Determination Methods

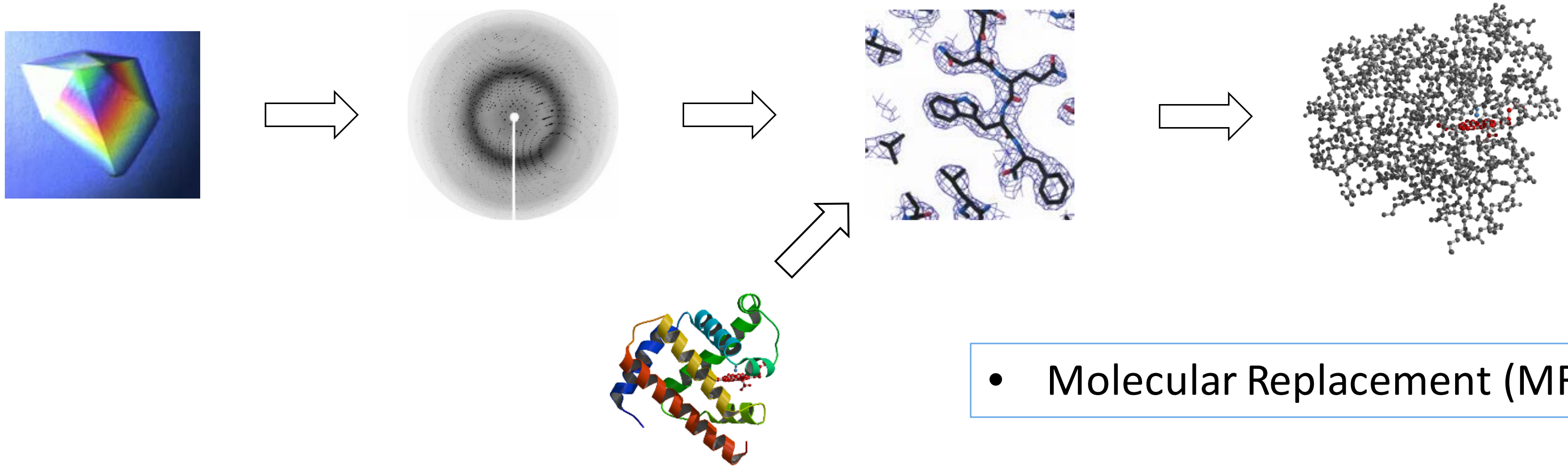




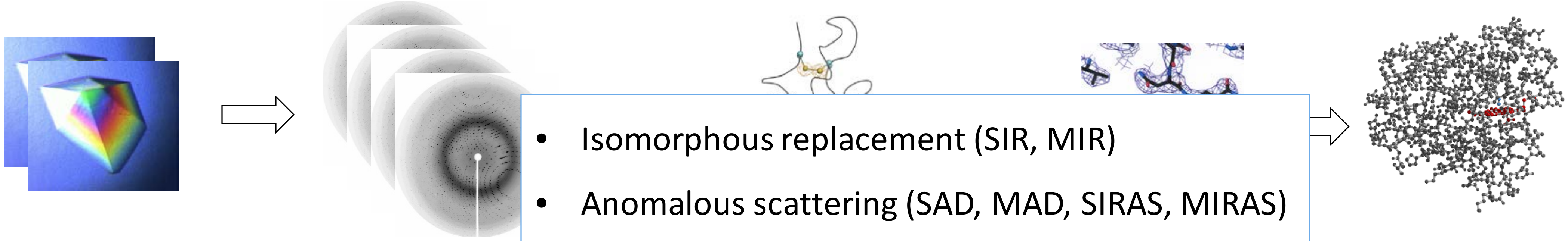
- Molecular Replacement (MR)

Structure Determination Methods



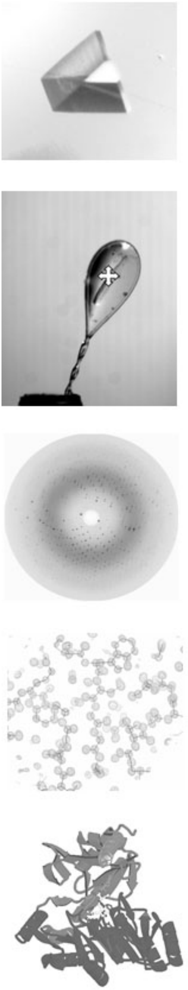


- Molecular Replacement (MR)



- Isomorphous replacement (SIR, MIR)
- Anomalous scattering (SAD, MAD, SIRAS, MIRAS)
- Radiation-damage induced phasing (RIP, RIPAS)

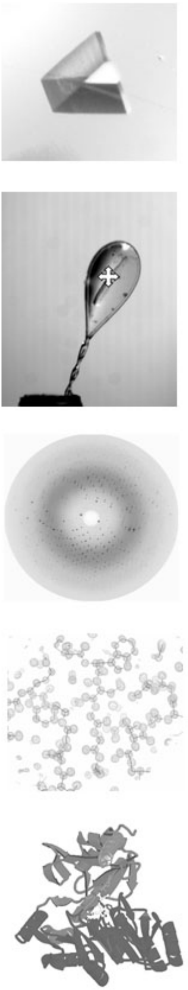
The Molecular Replacement (MR) Method

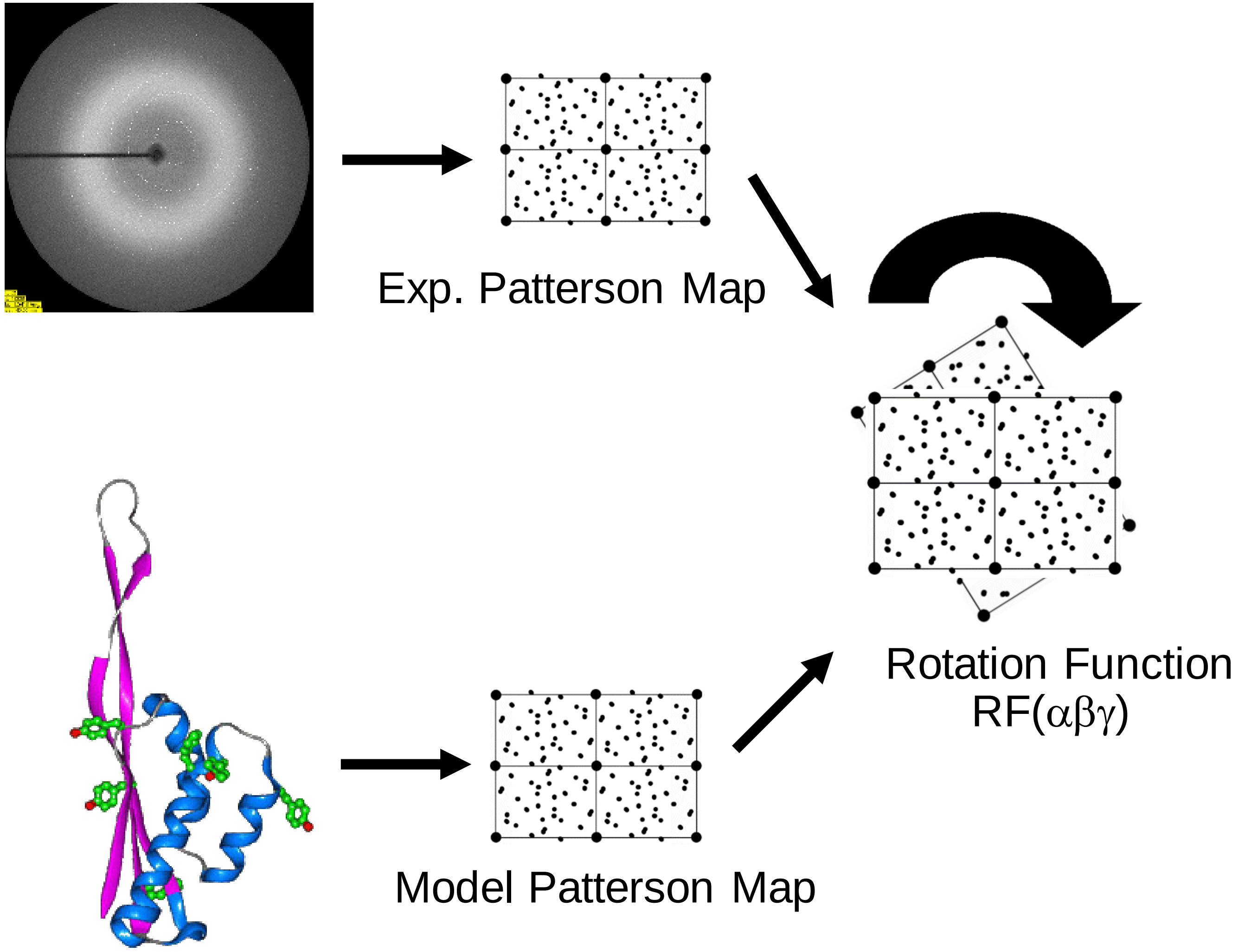


Molecular replacement can be used to solve a structure when a search structure is available.

When the search structure can be correctly oriented and positioned (6 parameters), it can be used as a starting point for phase calculation.

The Molecular Replacement (MR) Method



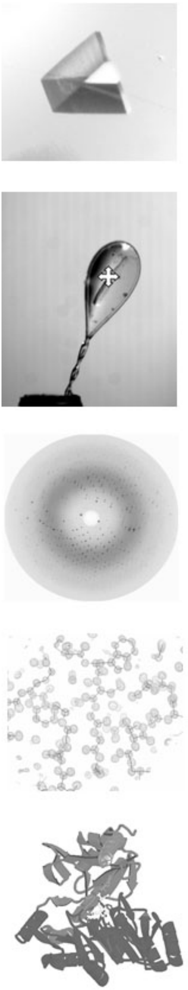


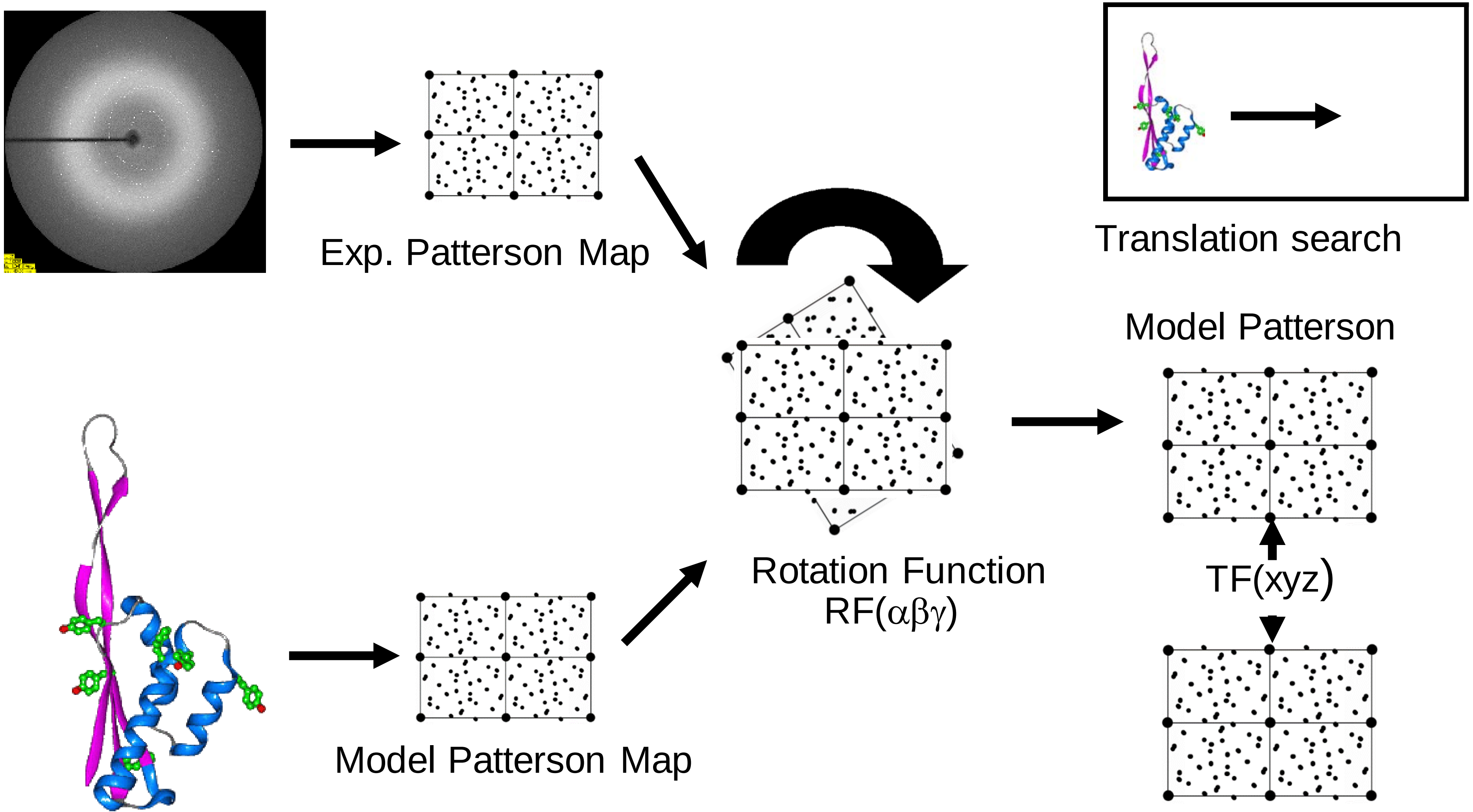
Exp. Patterson Map

Model Patterson Map

Rotation Function
$RF(\alpha\beta\gamma)$

The Molecular Replacement (MR) Method





Exp. Patterson Map

Model Patterson Map

Rotation Function
RF(αβγ)

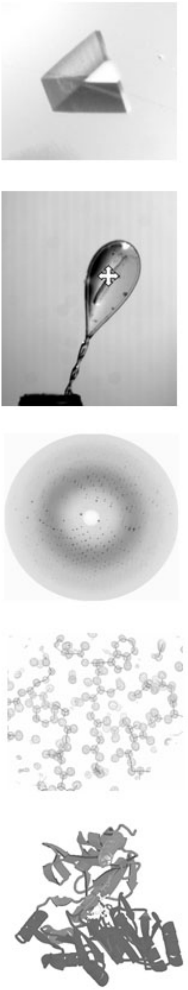
Translation search

Model Patterson

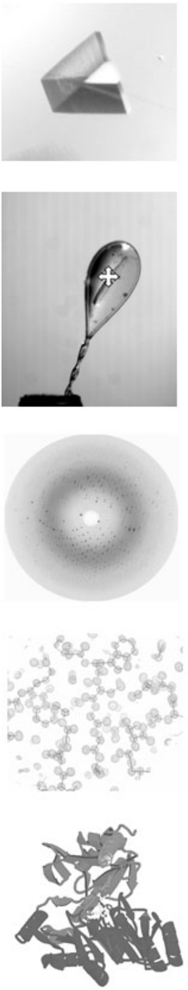
TF(xyz)

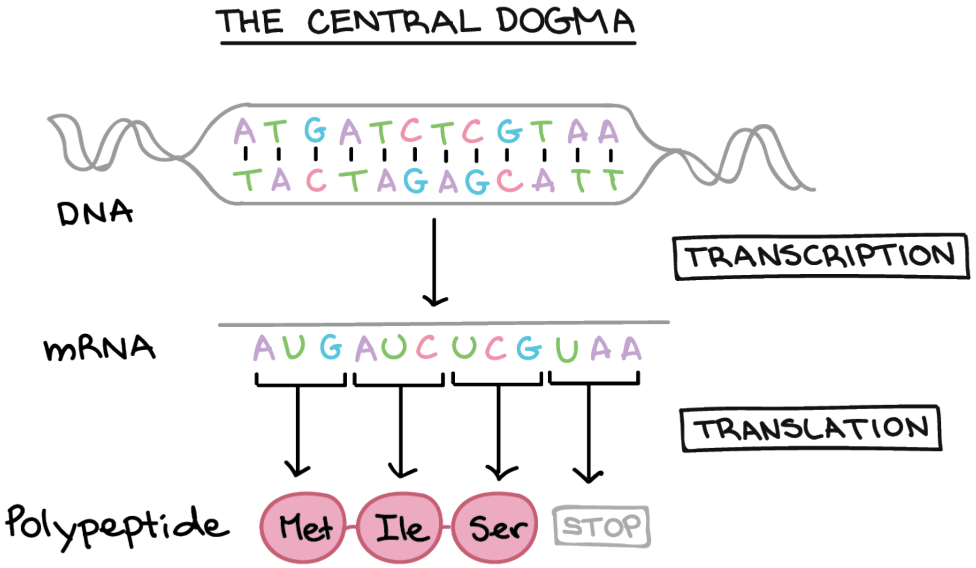
Exp. Patterson

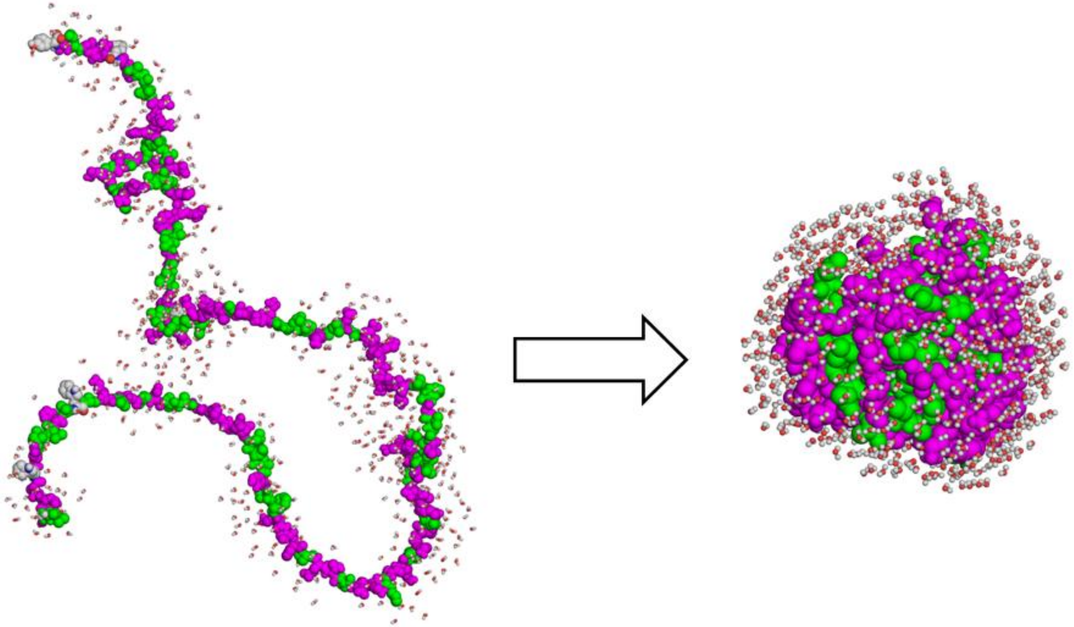
The Search Model



Information and Structure in Biology



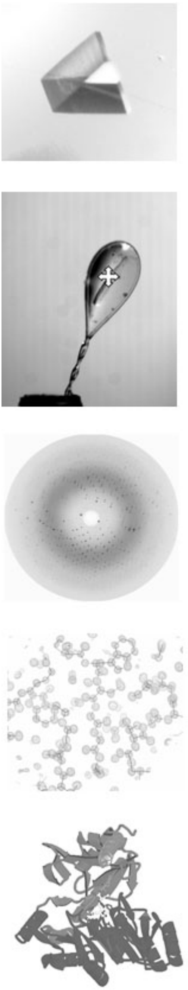




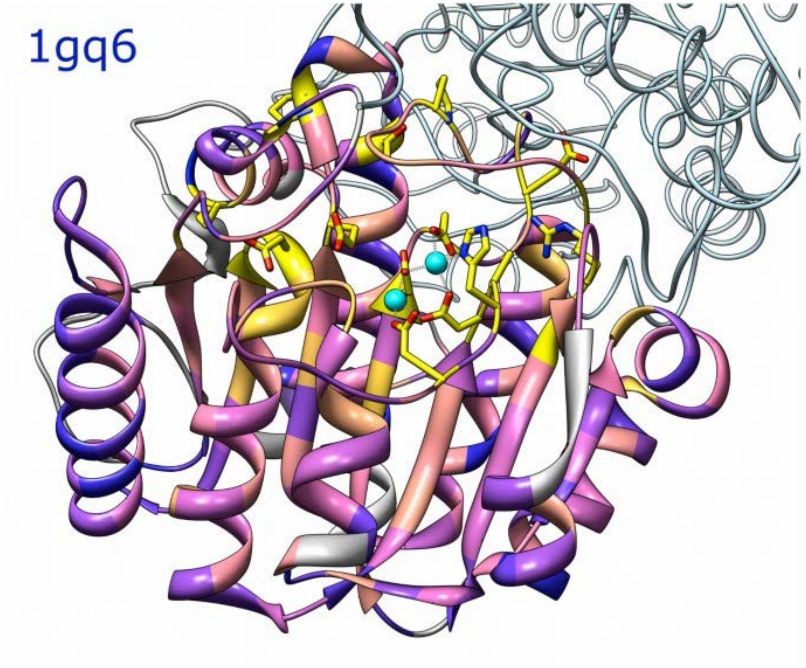
Unfolded

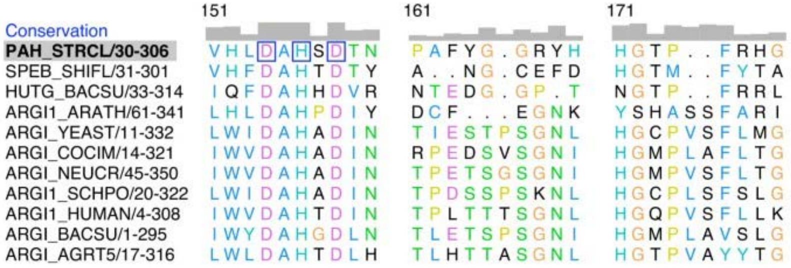
Folded

The Search Model



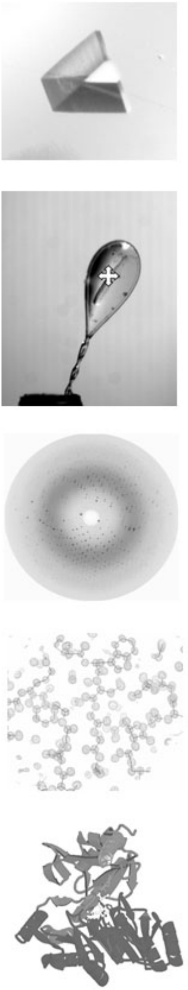
- Proteins, which share a similar amino acid sequence, exhibit a similar three-dimensional structure





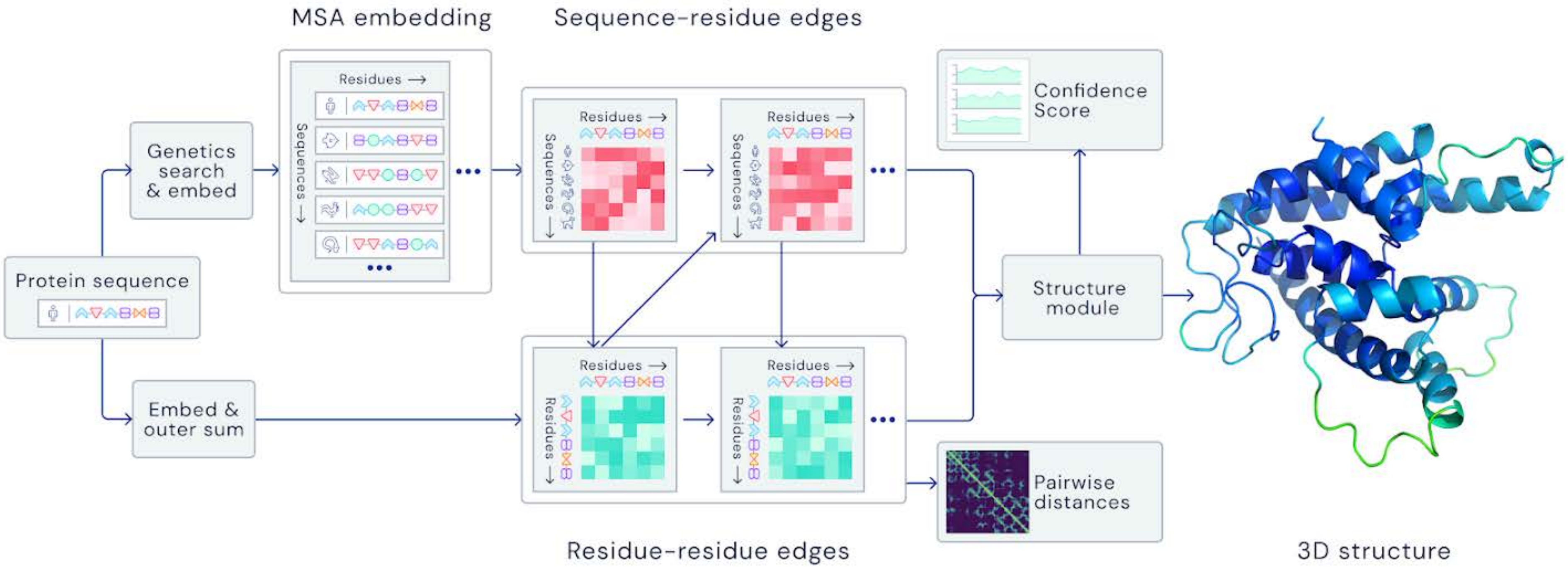
	151	161	171
Conservation			
PAH_STRCL/30-306	VHLDAHSDTN	PAFYG.GRYH	HGTP..FRHG
SPEB_SHIFL/31-301	VHFDAHTDTY	A..NG.CEFD	HGTM..FYTA
HUTG_BACSU/33-314	IQFDAHHDVR	NTEDG.GP.T	NGTP..FRRL
ARGI1_ARATH/61-341	LHLDAHPDIY	DCF...EGNK	YSHASSFARI
ARGI_YEAST/11-332	LWIDAHADIN	TIESTPSGNL	HGCPVSFLMG
ARGI_COCIM/14-321	IWVDAHADIN	RPEDSVSGNI	HGMPLAFLTG
ARGI_NEUCR/45-350	IWVDAHADIN	TPETSGSGNI	HGMPVSFLTG
ARGI1_SCHPO/20-322	LWIDAHADIN	TPDSSPSKNL	HGCPLSFSLG
ARGI1_HUMAN/4-308	IWVDAHTDIN	TPLTTTSGNL	HGQPVSFLLK
ARGI_BACSU/1-295	IWYDAHGDLN	TLETSPSGNI	HGMPLAVSLG
ARGI_AGRT5/17-316	LWLDAHTDLH	TLHTTASGNL	HGTPVAYYTG

The Search Model

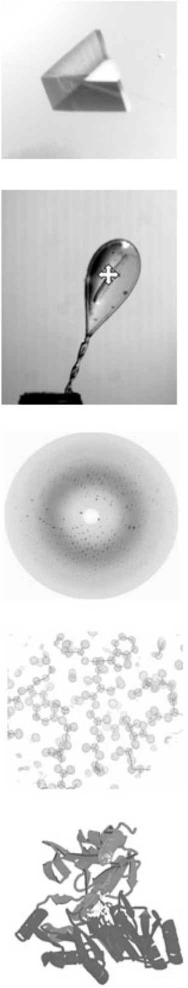


- Proteins, which share a similar amino acid sequence, exhibit a similar three-dimensional structure

- Prediction of the three-dimensional structure of a protein from ist amino acid sequence (AlphaFold2)



The Molecular Replacement (MR) Method

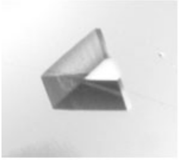


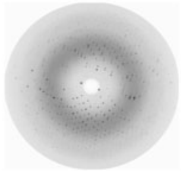
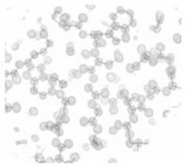
Molecular replacement can be used to solve a structure when a search structure is available.

When the search structure can be correctly oriented and positioned (6 parameters), it can be used as a starting point for phase calculation.

The search structure can either be a homologous structure (from the PDB), identified based on amino acid sequence similarity, or a structure predicted by AlphaFold2.

The “Other” Methods




SIR: **S**ingle **i**somorphous **r**eplacement

MIR: **M**ultiple **i**somorphous **r**eplacement

MAD: **M**ultiple wavelength **a**nomalous **d**ispersion

SAD: **S**ingle wavelength **a**nomalous **d**ispersion

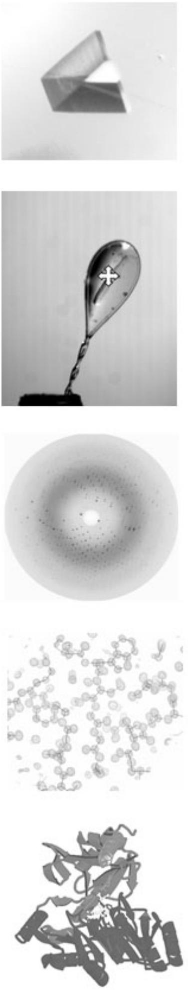
SIRAS: **S**ingle **i**somorphous **r**eplacement with **a**nomalous **s**cattering

MIRAS: **M**ultiple **i**somorphous **r**eplacement with **a**nomalous **s**cattering

RIP: **R**adiation-damage **i**nduced **p**hasing

RIPAS: **R**adiation-damage **i**nduced **p**hasing with **a**nomalous **s**cattering

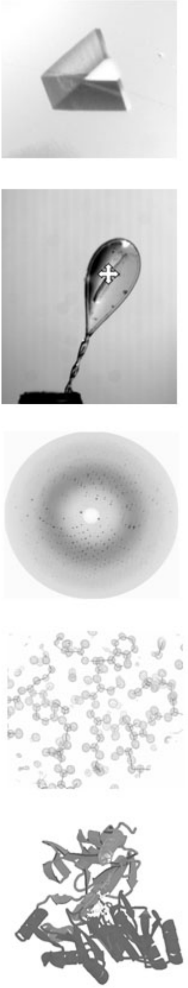
The “Other” Methods



All these methods have in common that data from more than one crystal or from more than one wavelength are compared to one another.

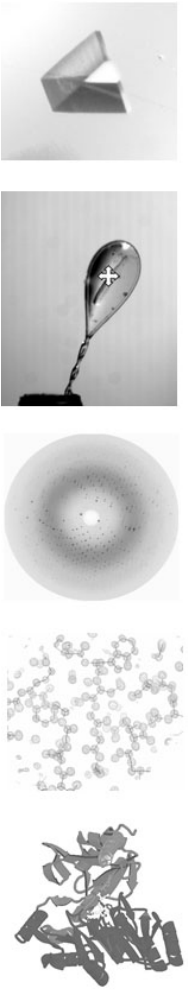
By doing this, the initial step is reduced to the determination of the (heavy atom, anomalously scattering, sulfur) substructure.

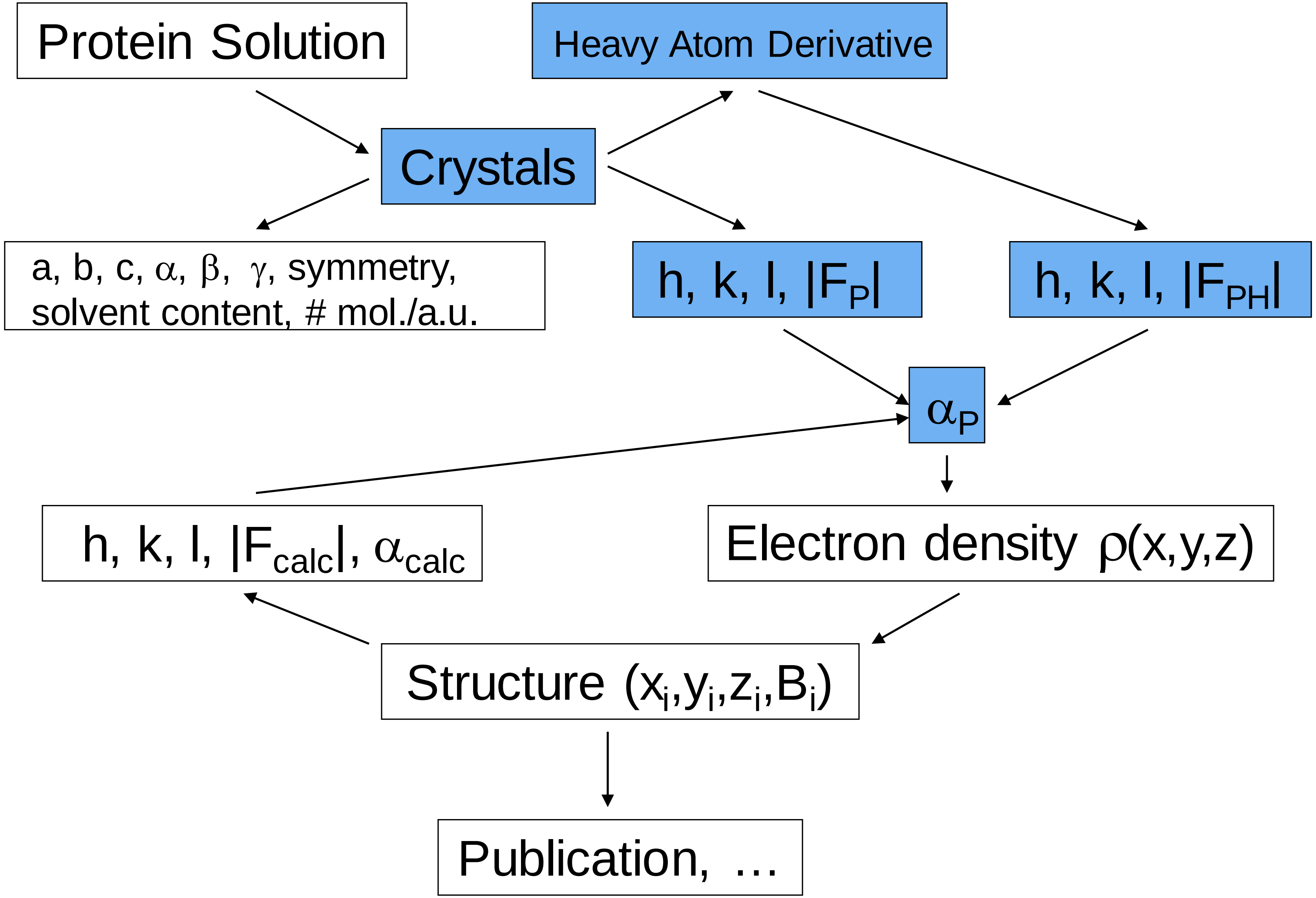
Experimental Approaches



- Soaking or co-crystallization with heavy atom compounds
- Quick-soaking with highly soluble ions (Br$^-$, Rb$^+$, …)
- Pressurization using noble gases (Kr, Xe)
- Covalent modification of proteins (Met → SeMet) or DNA (T → Br-U)
- Utilization of naturally present metals (Fe, Zn, Cu, …)
- Collecting data at different wavelengths
- Collecting data at long wavelengths ($\lambda \approx 2.0$ Å)
- Modifying the crystal by exposure to X-rays or UV

A Flowchart of a Crystal Structure Determination





Protein Solution

Heavy Atom Derivative

Crystals

a, b, c, α, β, γ, symmetry, solvent content, # mol./a.u.

h, k, l, $|F_P|$

h, k, l, $|F_{PH}|$

α_P

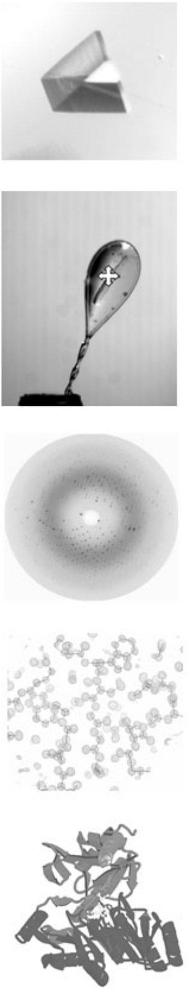
h, k, l, $|F_{calc}|$, α_{calc}

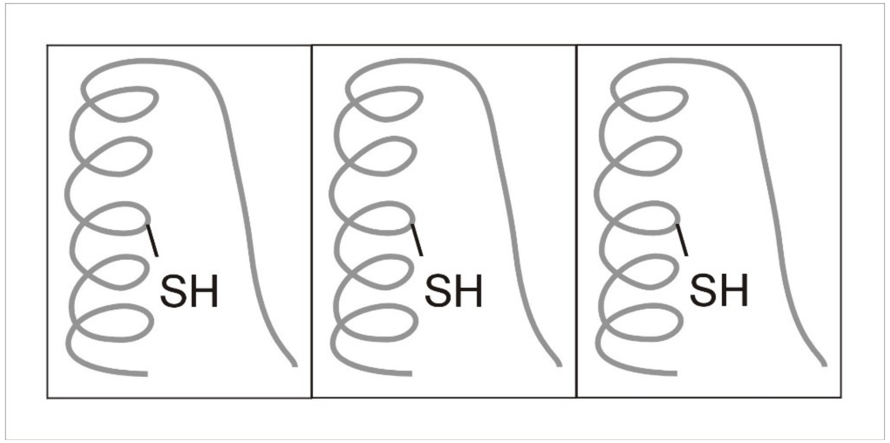
Electron density $\rho(x,y,z)$

Structure (x_i,y_i,z_i,B_i)

Publication, …

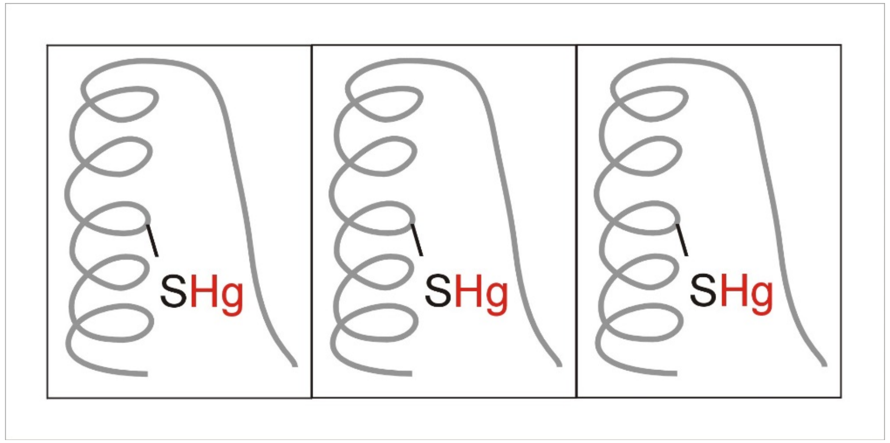
Isomorphous Replacement





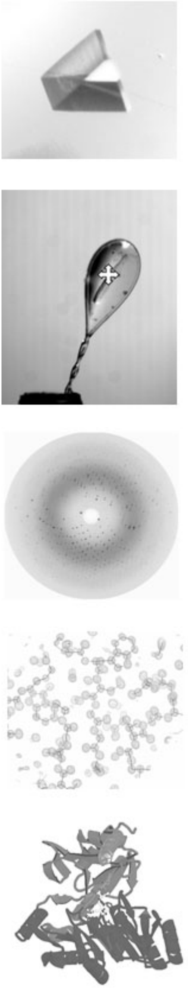
$\Rightarrow$ h, k, l, $|F_P(hkl)|$

$\alpha_P(hkl)$



$\Rightarrow$ h, k, l, $|F_{PH}(hkl)|$

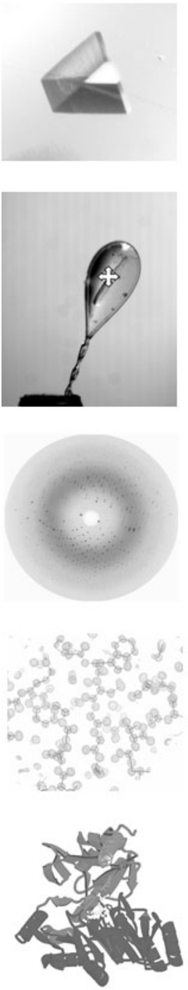
Isomorphous Replacement



<u>Step 1</u>: put $|F_P(hkl)|$ and $|F_{PH}(hkl)|$ on the same scale

$$(R = 100 \bullet \Sigma_{hkl} \,|\, |F_{PH}| - |F_P| \,|\, / \Sigma_{hkl} |F_P|)$$

Isomorphous Replacement



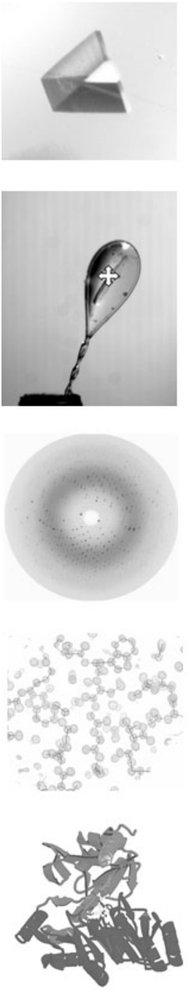
Step 1: put $|F_P(hkl)|$ and $|F_{PH}(hkl)|$ on the same scale

$$(R = 100 \bullet \Sigma_{hkl} \, | \, |F_{PH}| - |F_P| \, | \, / \, \Sigma_{hkl} \, |F_P| \,)$$

Step 2 : determine the positions of the heavy atoms (F_H)

from the differences ($|F_{PH}(hkl)| - |F_P(hkl)|$)

Isomorphous Replacement



Step 1: put $|F_P(hkl)|$ and $|F_{PH}(hkl)|$ on the same scale

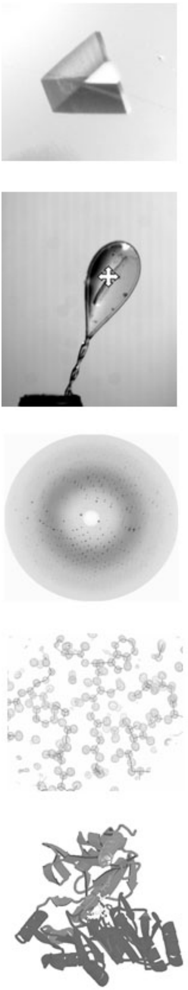
$(R = 100 \bullet \Sigma_{hkl} \,|\, |F_{PH}| - |F_P| \,|\, / \Sigma_{hkl} |F_P|)$

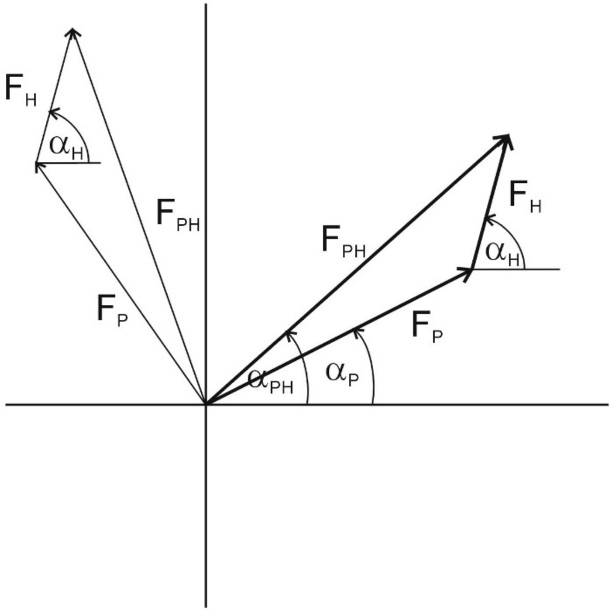
Step 2 : determine the positions of the heavy atoms (F_H)

from the differences ($|F_{PH}(hkl)| - |F_P(hkl)|$)

Step 3 : calculate $\alpha_P(hkl)$ from $|F_P(hkl)|$, $|F_{PH}(hkl)|$ and F_H

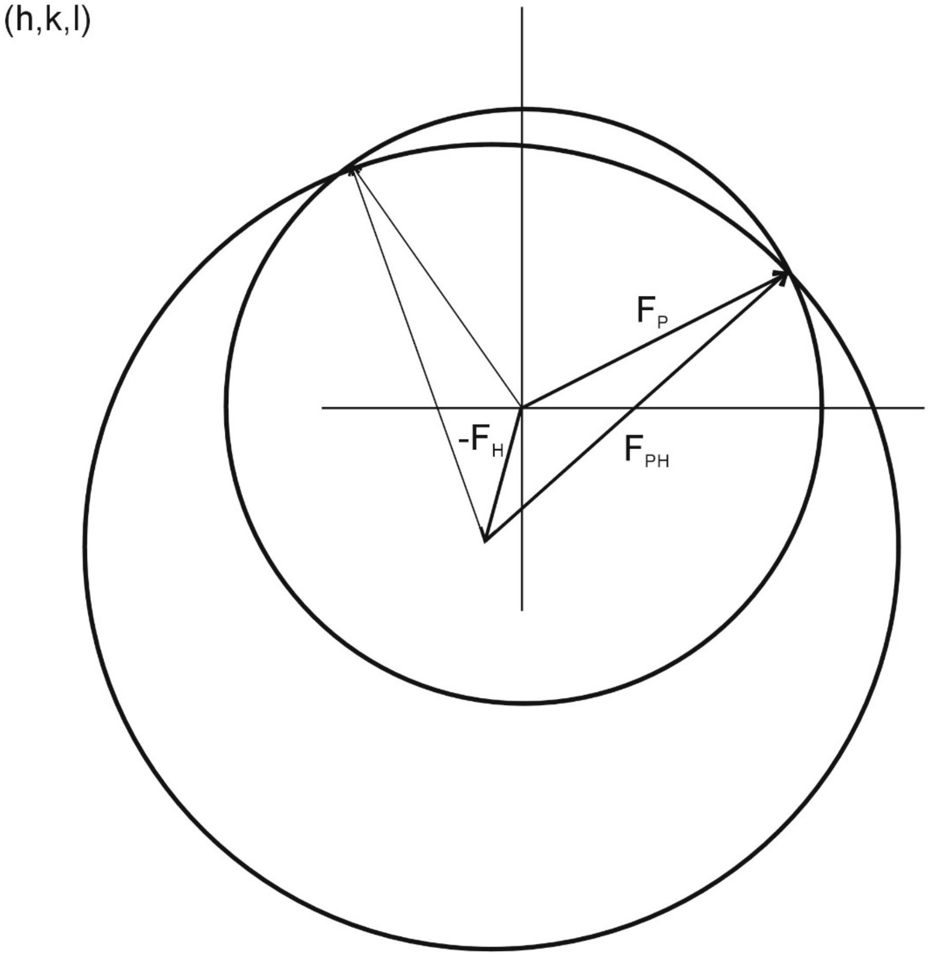
Harker Construction - SIR



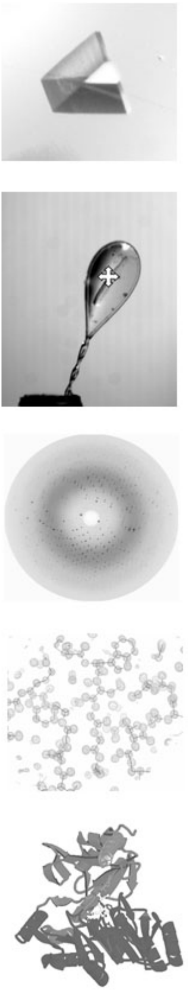


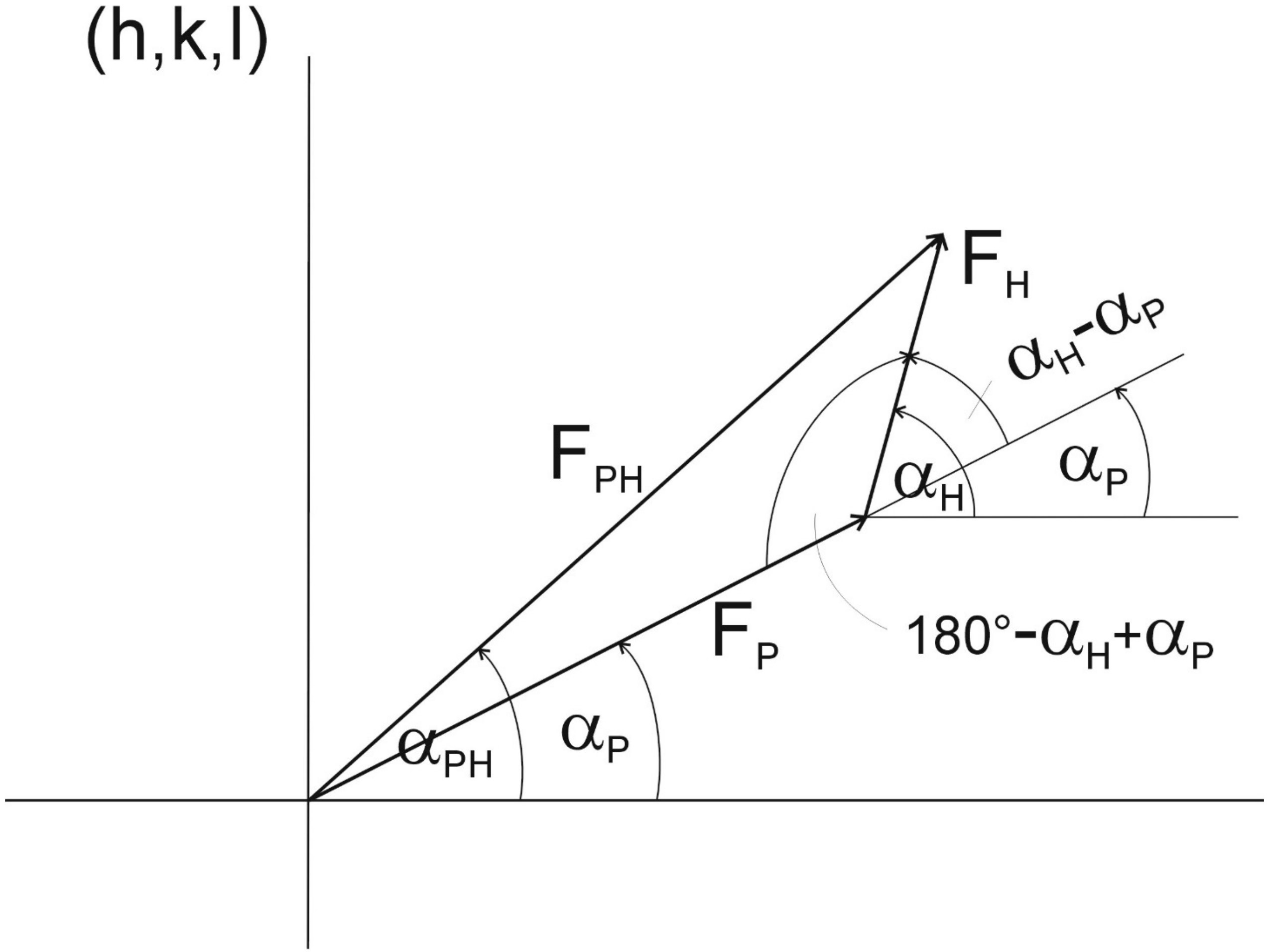
$$F_P + F_H = F_{PH}$$

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



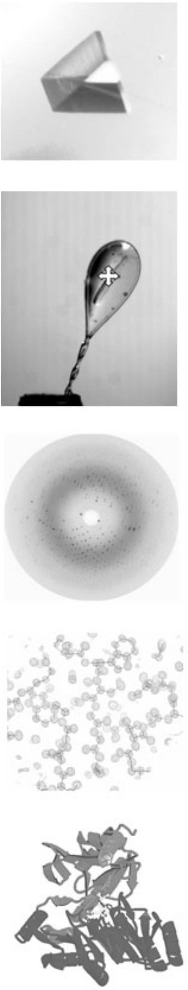
Harker Construction - SIR





$$\alpha_P = \alpha_H + \arccos \frac{|F_{PH}|^2 - |F_P|^2 - |F_H|^2}{2\,|F_P|\,|F_H|}$$

Harker Construction - SIR



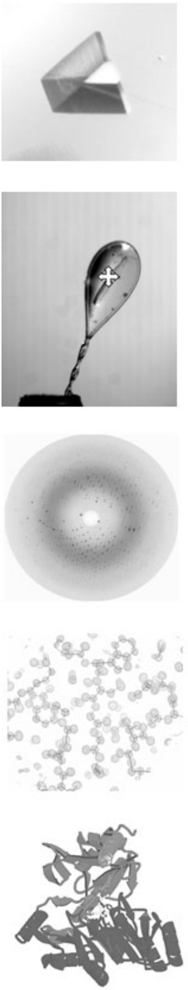
Instead of one value for α_P(hkl) we obtain two possibilities

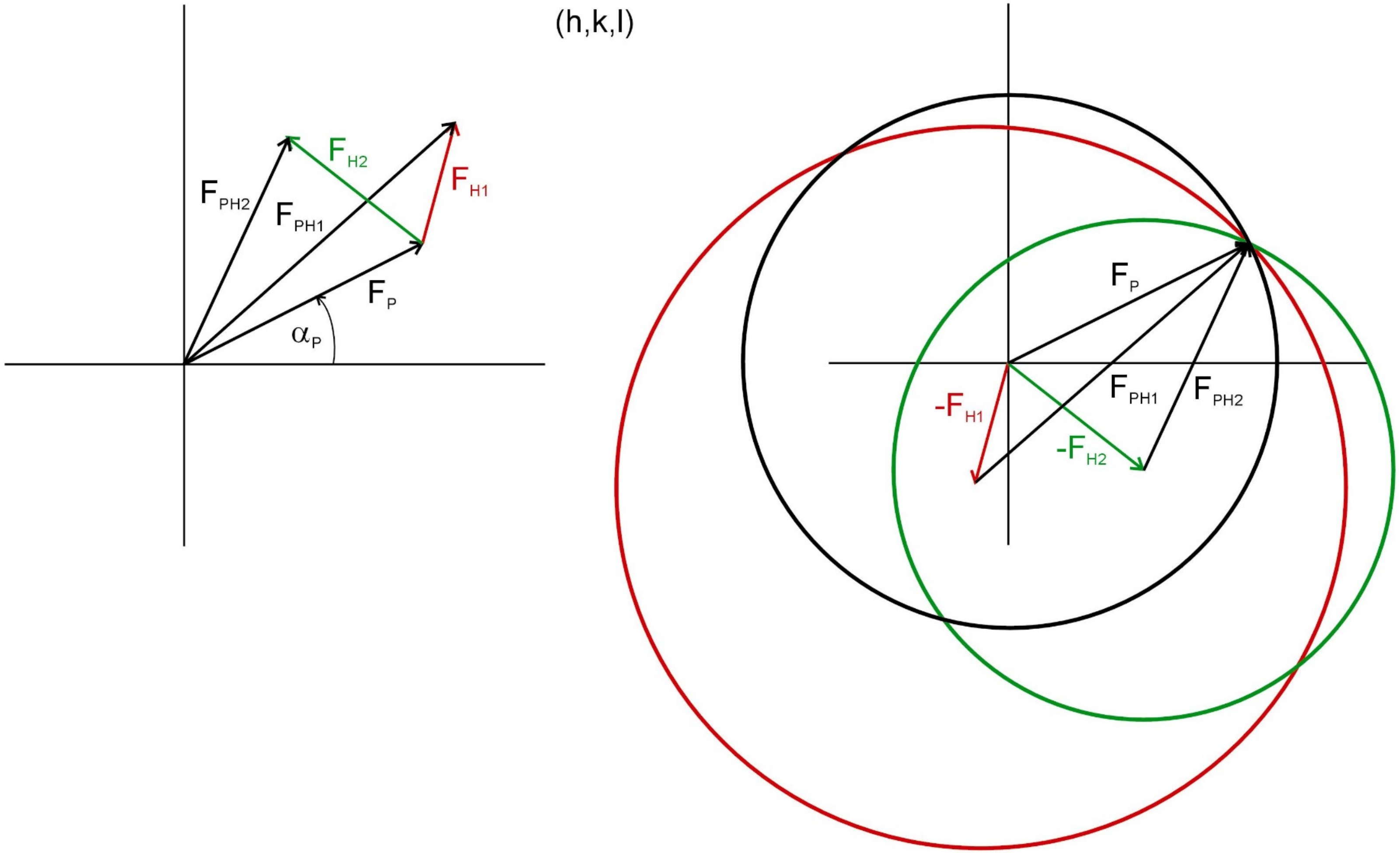
→ *phase ambiguity*

How can this be made unambiguous ?

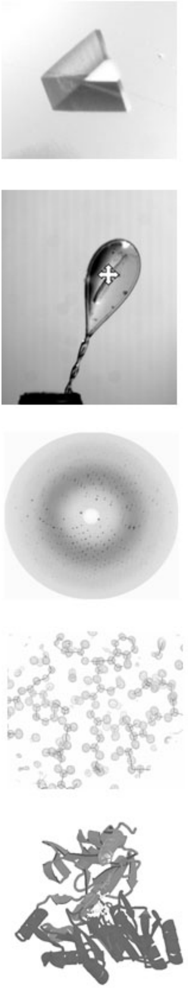
→ a second heavy atom derivative (MIR)

Harker Construction - MIR





Harker Construction - SIR



Instead of one value for α_P(hkl) we obtain two possibilities

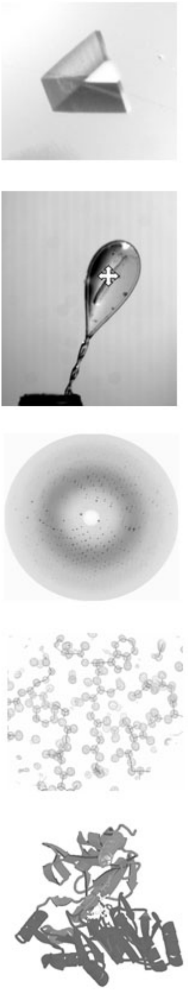
→ *phase ambiguity*

How can this be made unambiguous ?

→ a second heavy atom derivative (MIR)

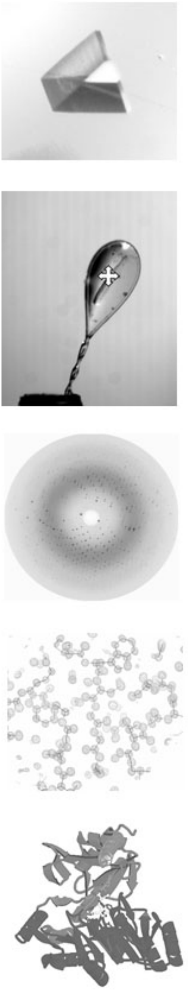
→ density modification

Desnity Modification



Density Modification is a technique that uses additional information to improve an electron density

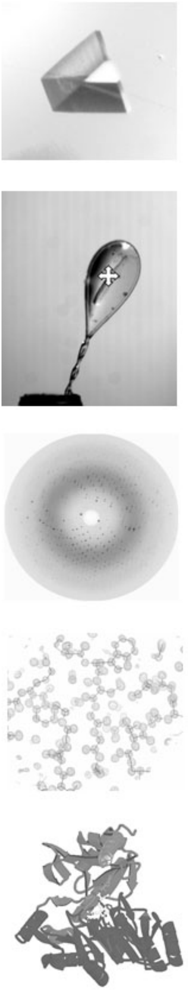
Desnity Modification

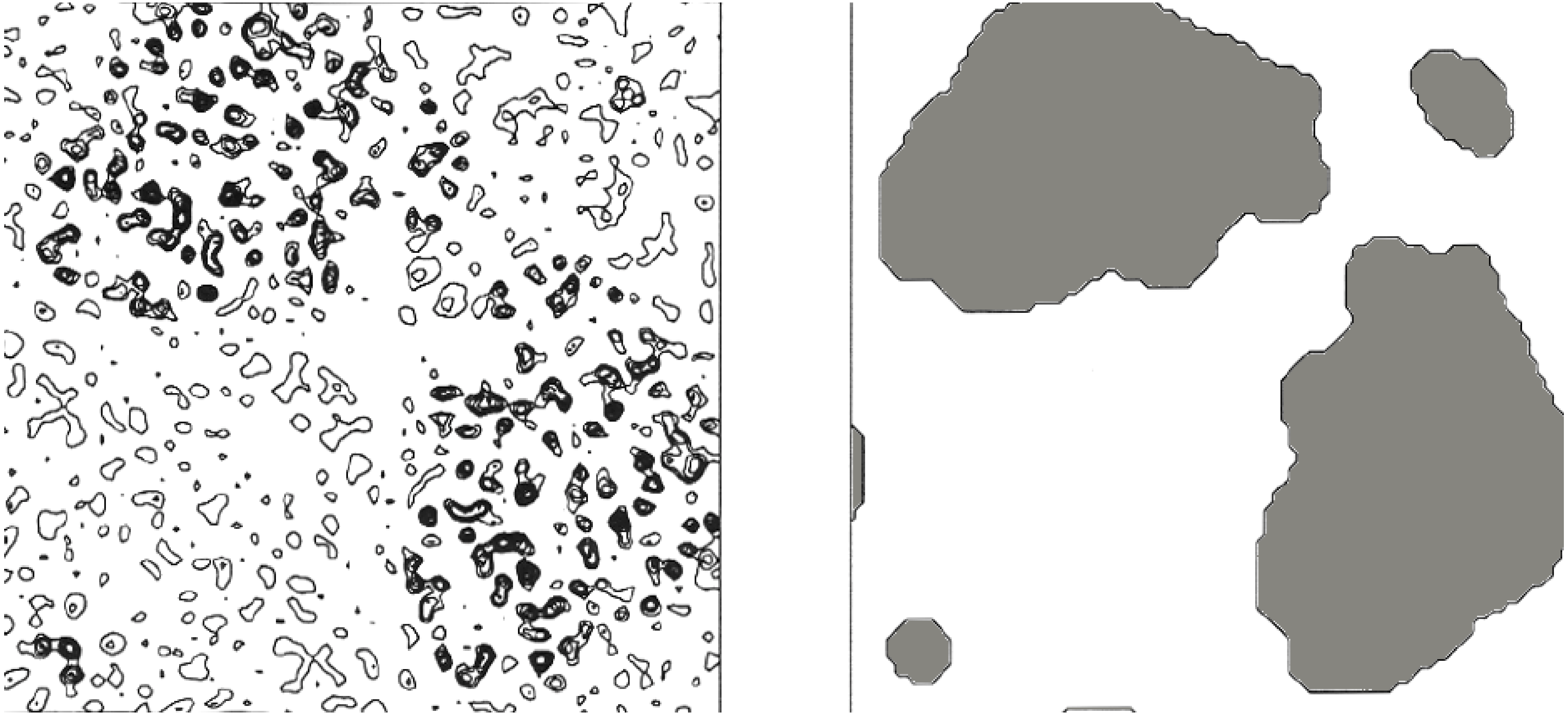


Density Modification is a technique that uses additional information to improve an electron density

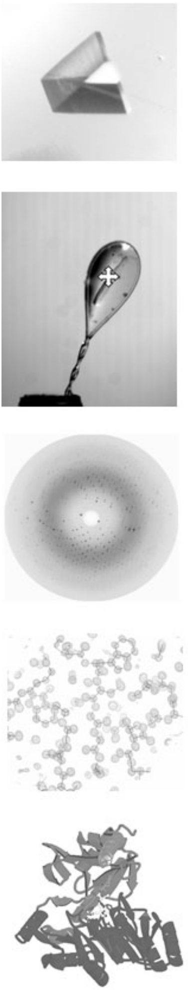
- ❑ solvent flattening

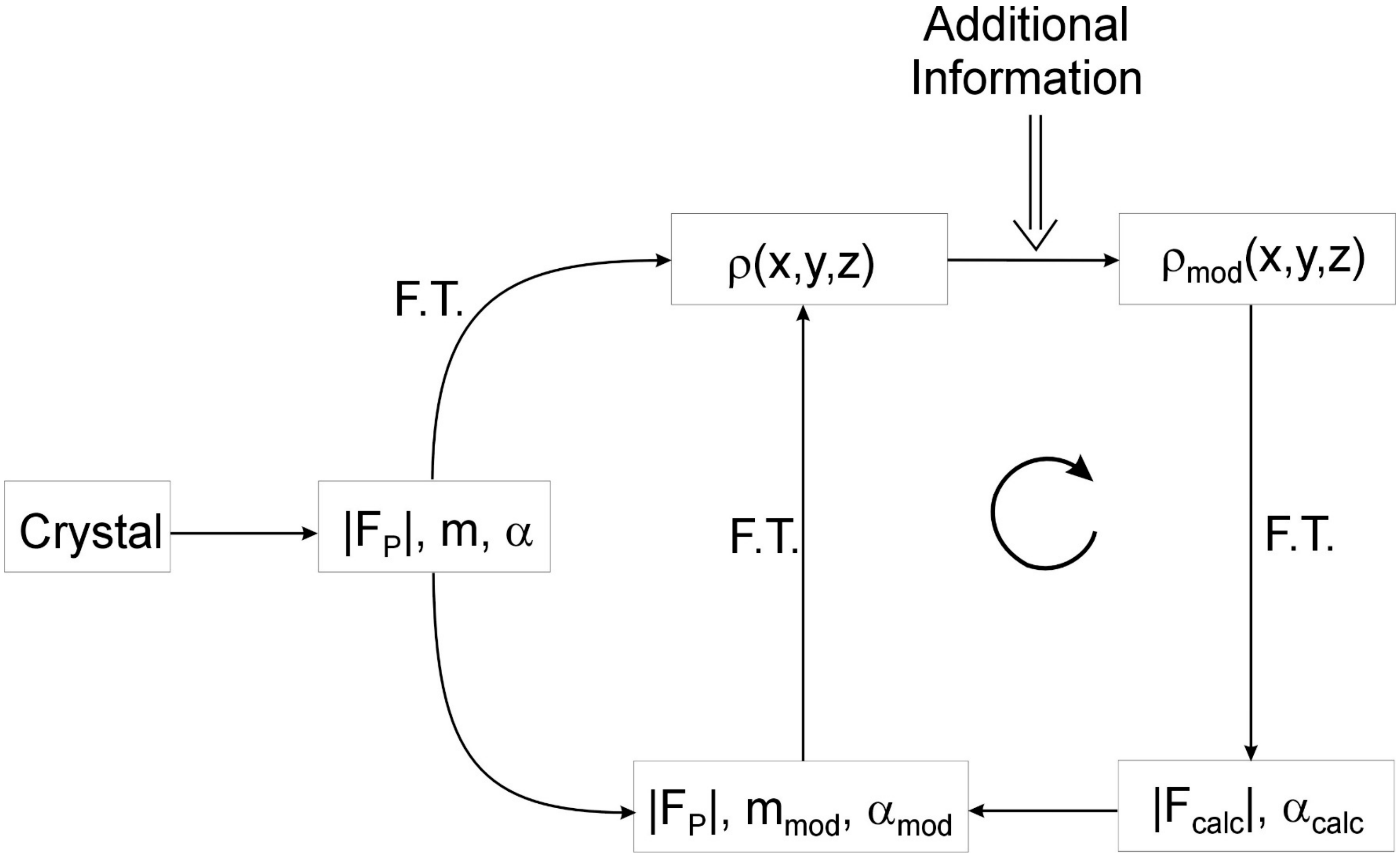
Solvent Flattening





Density Modification





Additional Information

Crystal

$|F_P|, m, \alpha$

F.T.

$\rho(x,y,z)$

$\rho_{mod}(x,y,z)$

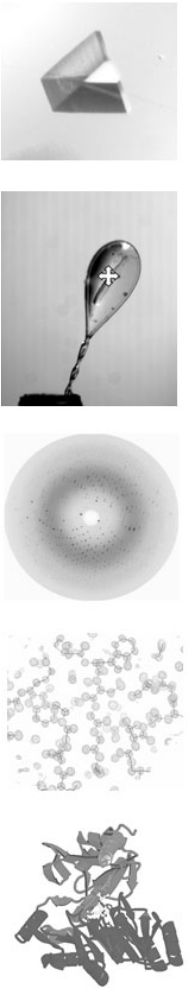
F.T.

F.T.

$|F_P|, m_{mod}, \alpha_{mod}$

$|F_{calc}|, \alpha_{calc}$

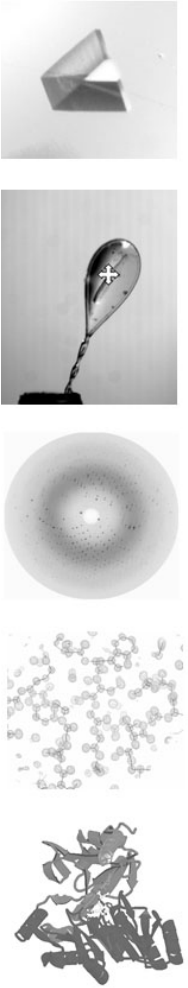
Desnity Modification



Density Modification is a technique that uses additional information to improve an electron density

- ❑ solvent flattening
- ❑ non-crystallographic symmetry averaging
- ❑ histogram matching
- ❑ map interpretation and refinement

Harker Construction - SIR



Instead of one value for α_P(hkl) we obtain two possibilities

→ *phase ambiguity*

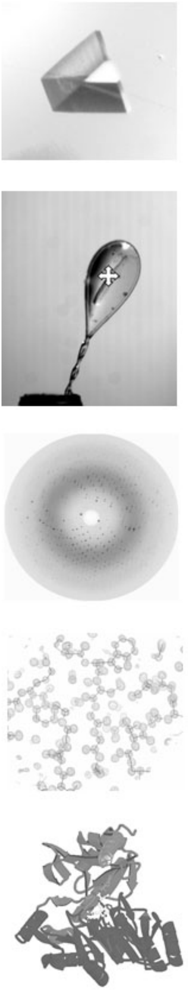
How can this be made unambiguous ?

→ a second heavy atom derivative (MIR)

→ density modification

→ incorporation of anomalous scattering (SIRAS)

Harker Construction - SIR



Instead of one value for α_P(hkl) we obtain two possibilities

→ *phase ambiguity*

How can this be made unambiguous ?

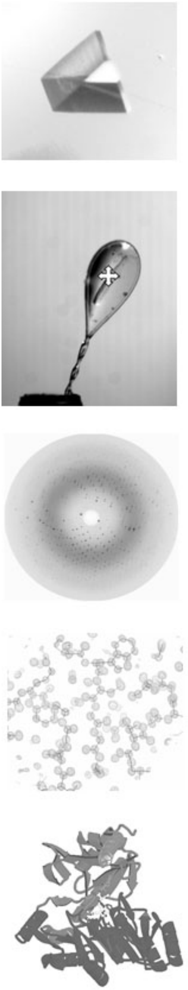
→ a second heavy atom derivative (MIR)

→ density modification

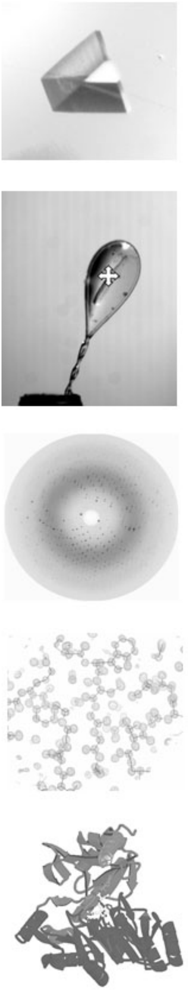
→ incorporation of anomalous scattering (SIRAS)

→ second derivative plus anomalous scattering (MIRAS)

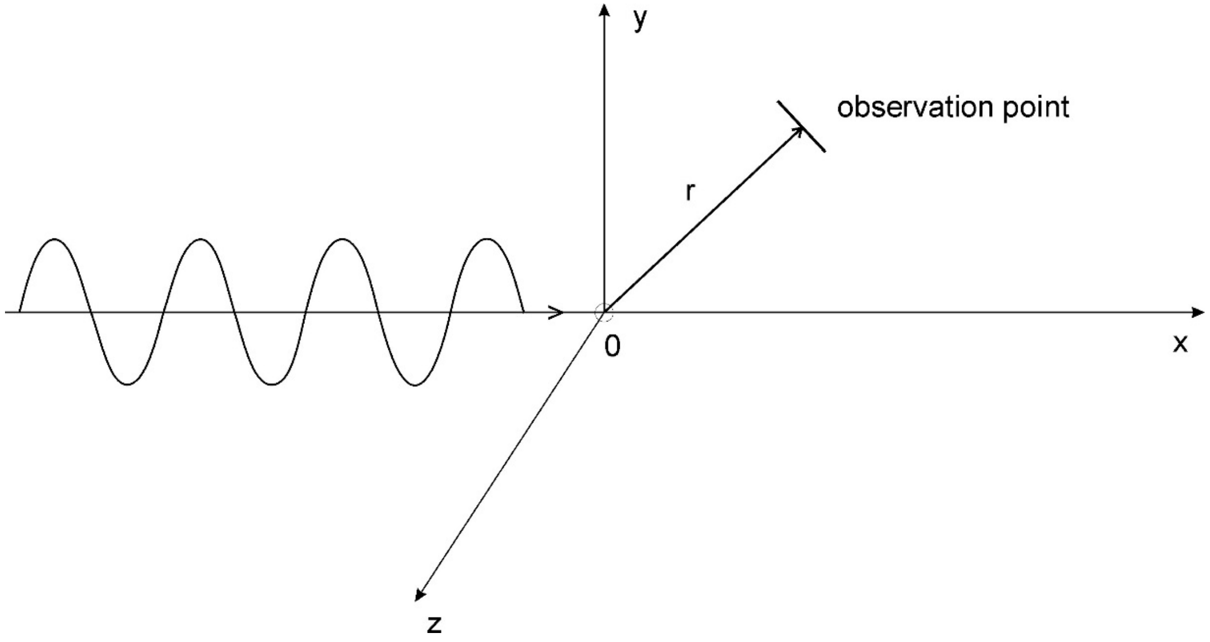
Anomalous Scattering



Normal Scattering - 1

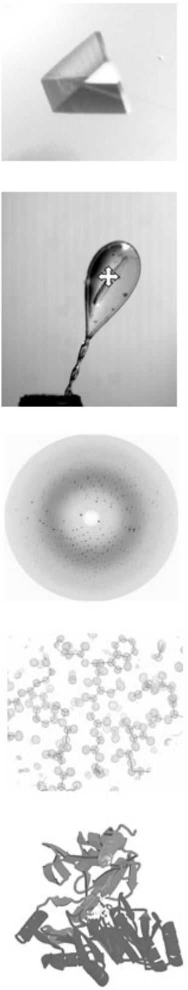


aka Elastic Scattering or Thomson Scattering or Scattering without Loss of Energy

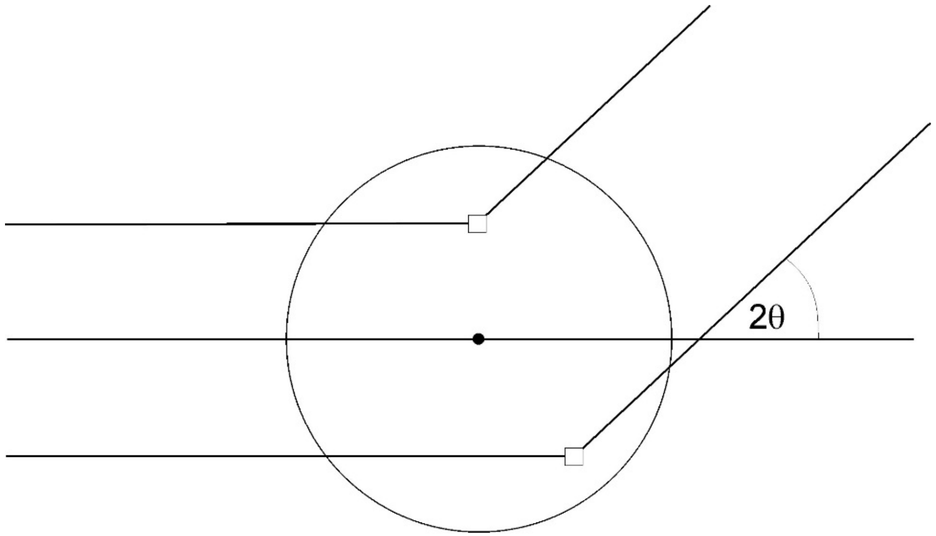


An X-ray wave impinges on an atom and causes the electrons to oscillate in the same frequency as the incoming X-ray wave.

Normal Scattering -2



The moving electrons emanate X-rays in all directions of space



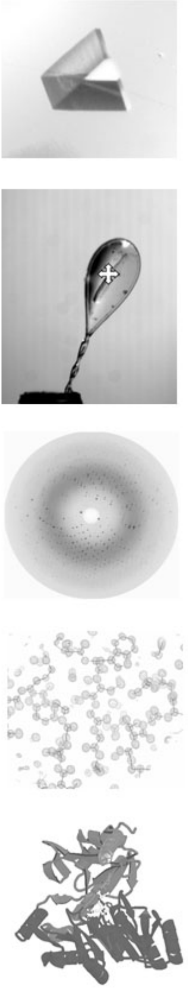
The amplitude of the scattered wave depends on the number of electrons and the scattering angle

$$f(\vec{s}) = \int_V \rho(\vec{r}) \cdot e^{2\pi i \vec{r}\vec{s}}$$

f(s) is the atomic form factor

f(s) is a scalar function, dependent only on the atom type and scattering angle 2θ

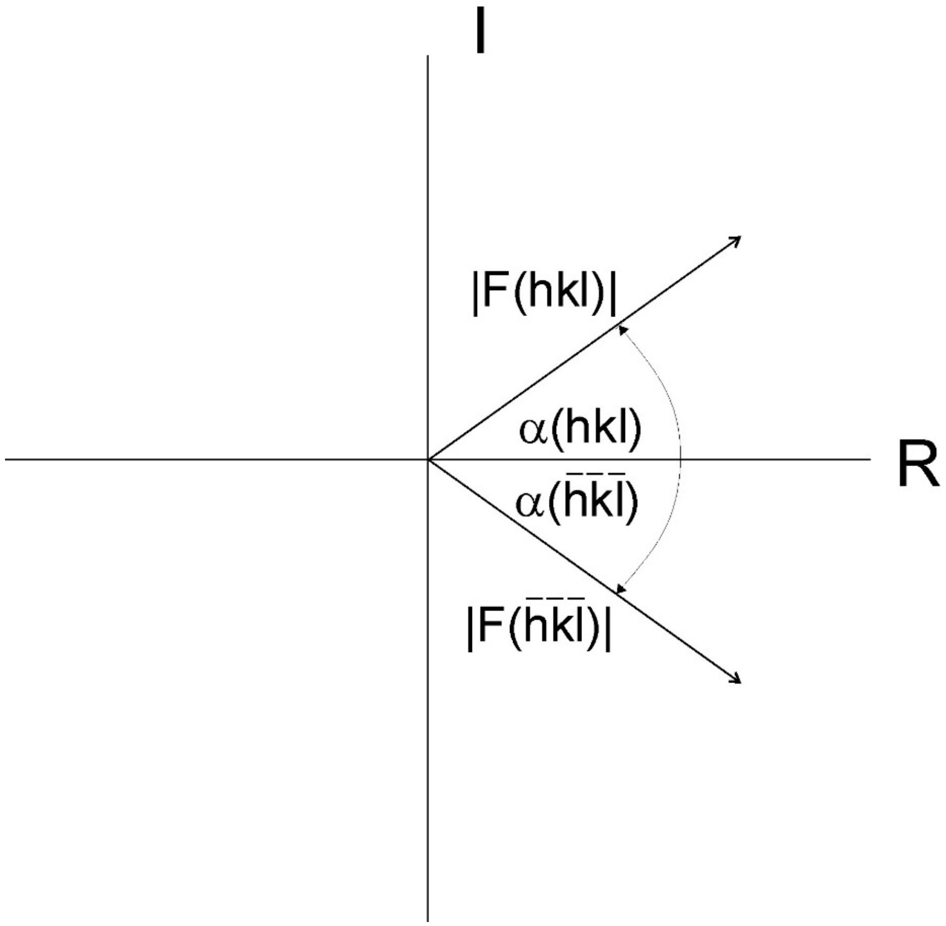
Normal Scattering -3



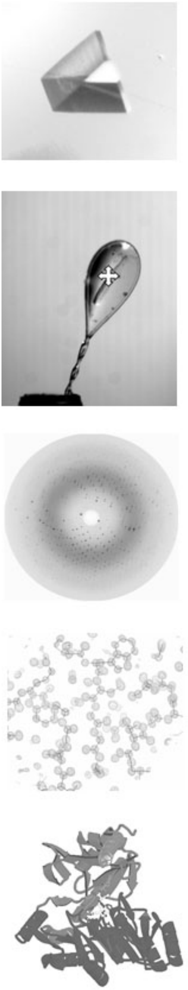
since ρ(r) = ρ(-r) → f(s) = f(-s), this results eventually in Friedel's Law

|F(hkl)| = |F(-h-k-l)|

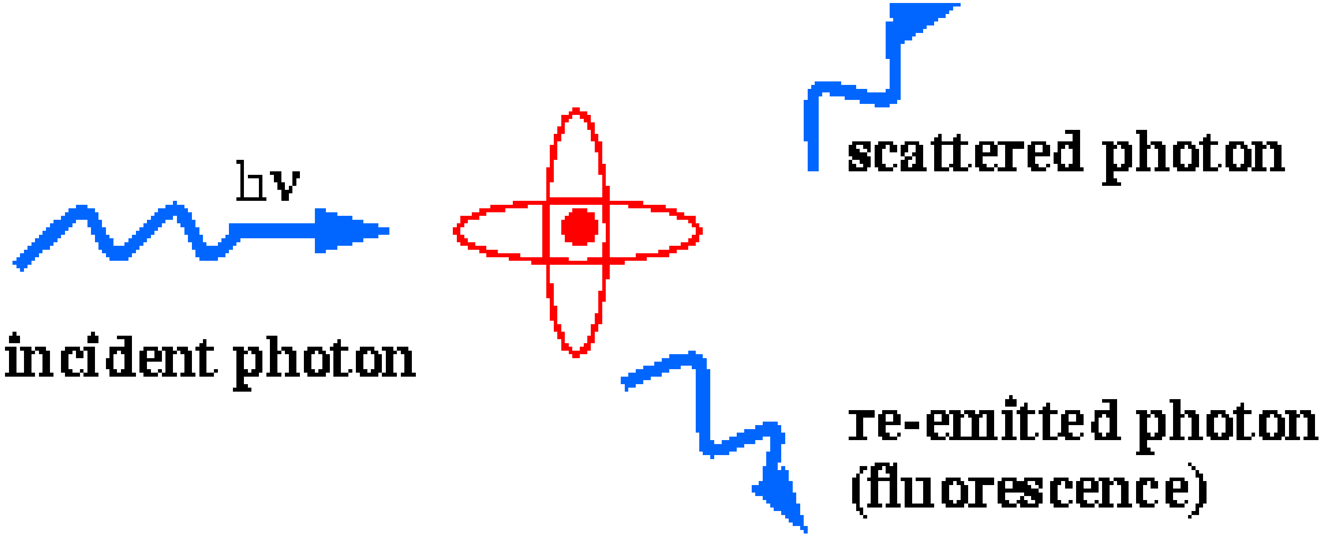
α(hkl) = - α(-h-k-l)



Anomalous Scattering - 1

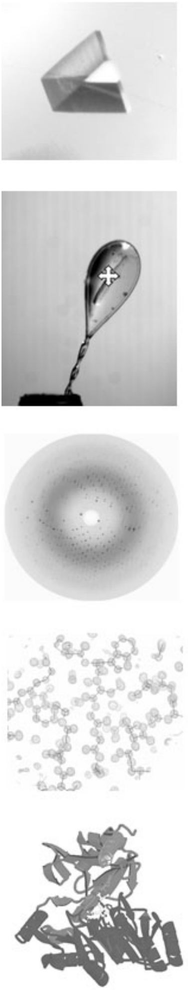


When the energy of the incident X-ray wave is close to an elemental absorption edge, something else happens in addition.

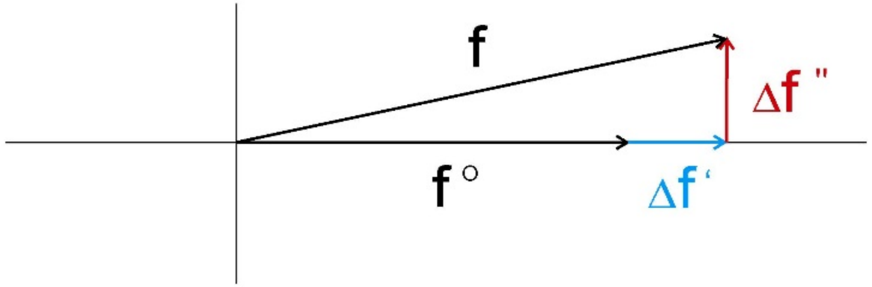


In addition to being scattered elastically, the incident X-ray wave is absorbed. It can then be re-emitted or it can eject a core electron (photoelectric effect). Subsequently, a higher-shell electron then falls back to a lower shell, emitting X-radiation (X-ray fluorescence).

Anomalous Scattering - 2



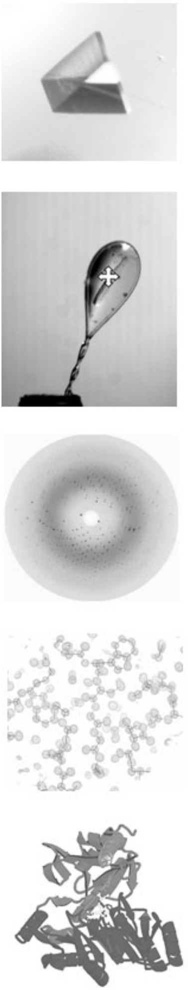
As a net result of absorption, the scattered wave is retarded and the atomic form factor becomes a vectorial quantity.



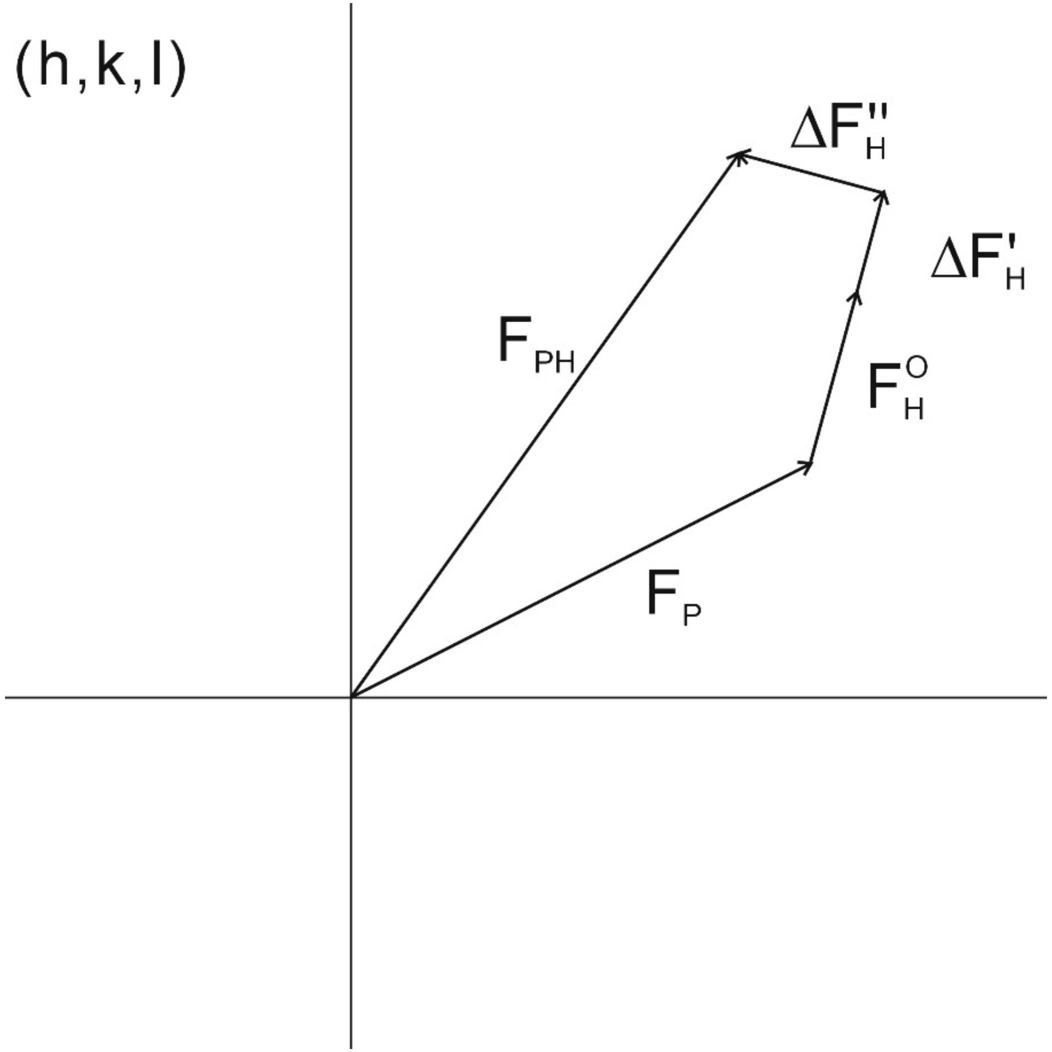
$$f = f^{\circ} + \Delta f' + i\Delta f''$$

f ° is the normal atomic form factor, Δf ’ is called the ***dispersive component*** and Δf ’’ the ***anomalous component***.

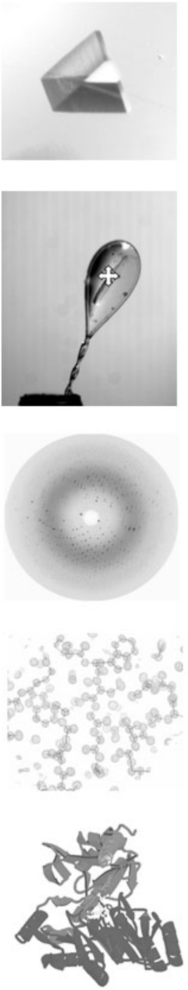
Anomalous Scattering - 3

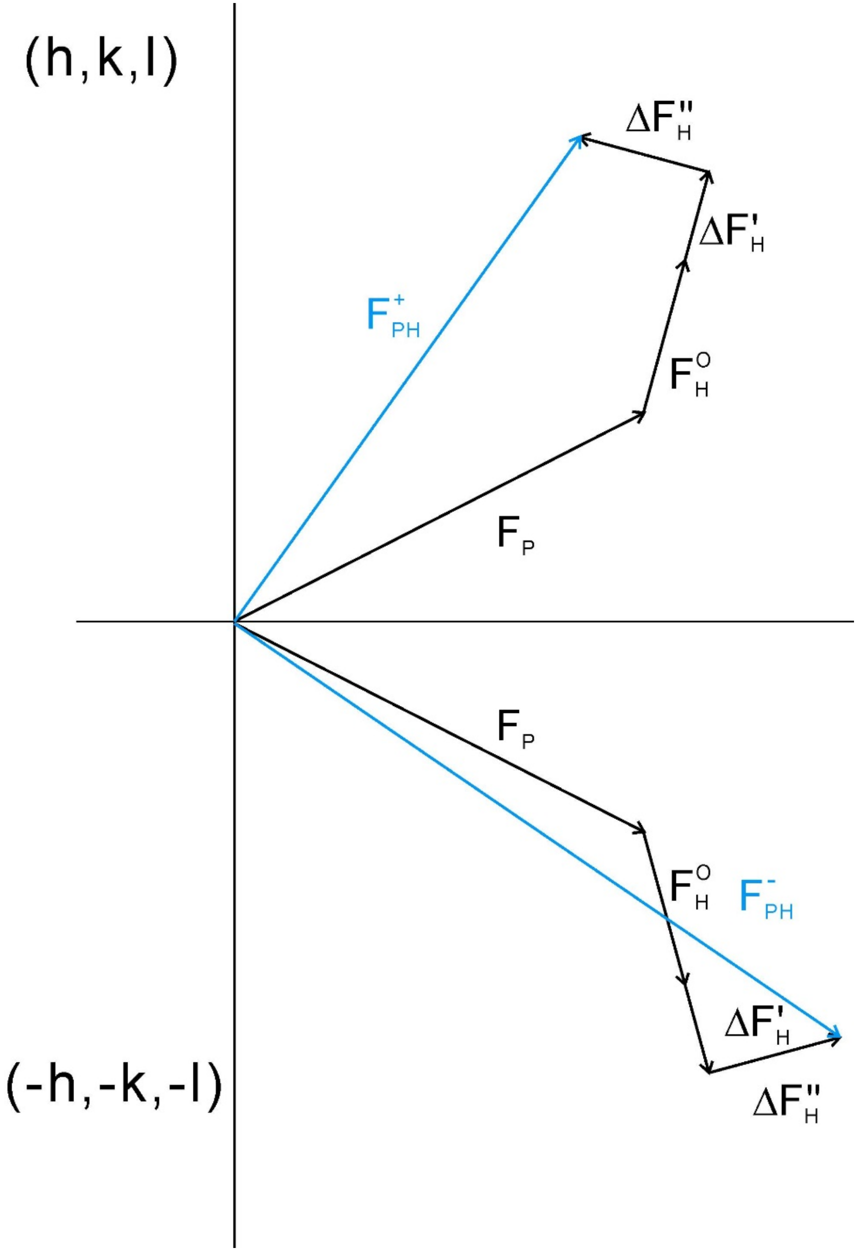


Thus, the total structure factor F_{PH} contains the ***normal scattering*** of all atoms $F_P+F_H^0$, the ***dispersive component*** of the anomalously scattering atoms $\Delta F_H'$ and the ***anomalous component*** of the anomalously scattering atoms $\Delta F_H''$.



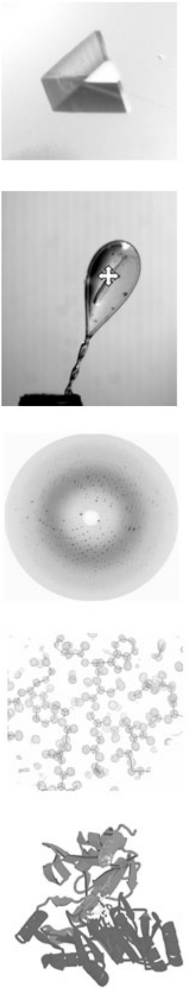
Anomalous Scattering - 4

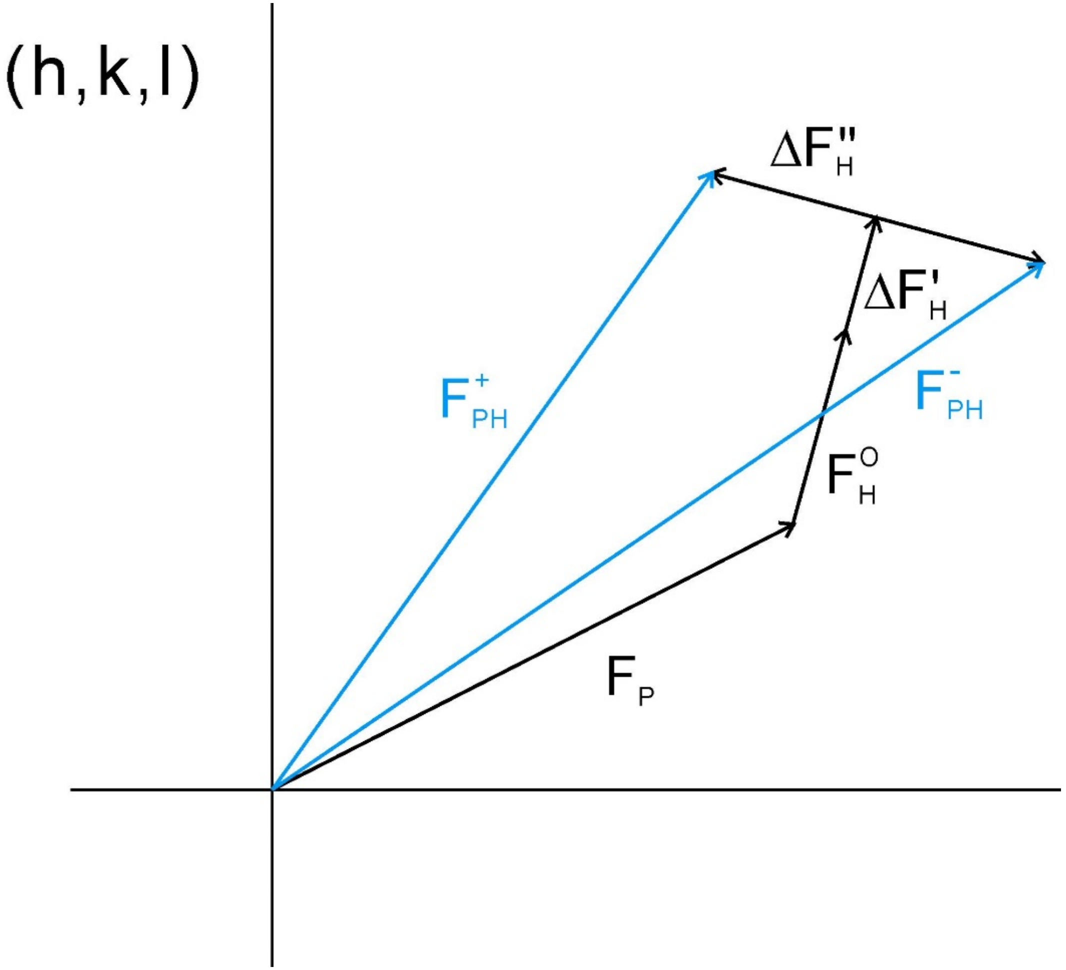




As consequence of anomalous scattering, Friedel’s Law is not valid anymore.

Anomalous Scattering - 5



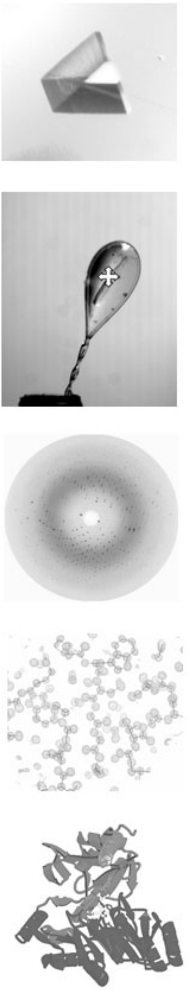


As consequence of anomalous scattering, Friedel’s Law is not valid anymore.

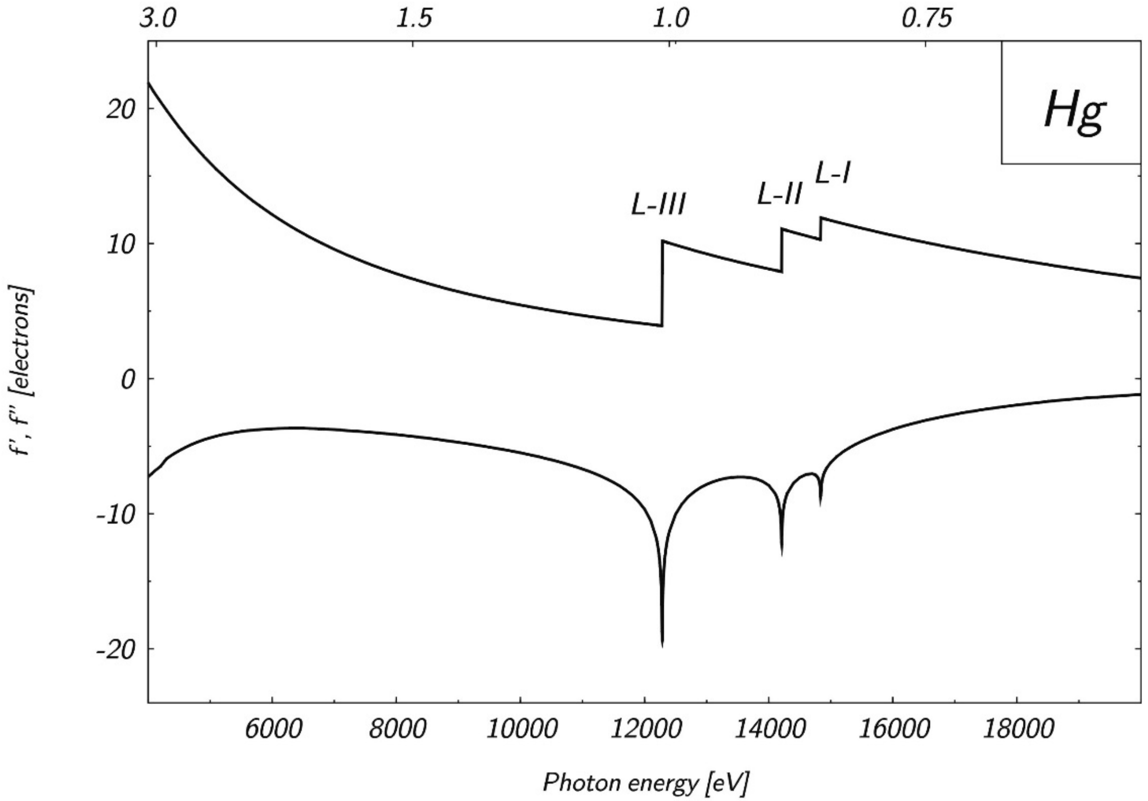
$|F(hkl)| \neq |F(-h-k-l)|$

$\alpha(hkl) \neq -\alpha(-h-k-l)$

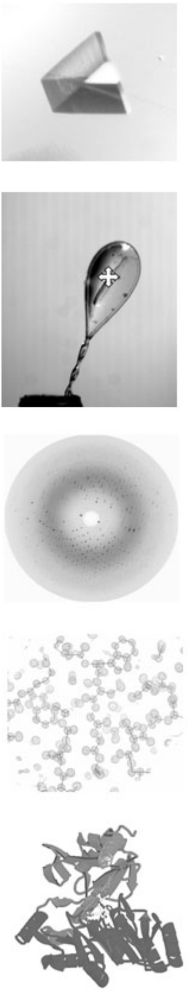
Anomalous scattering - 6

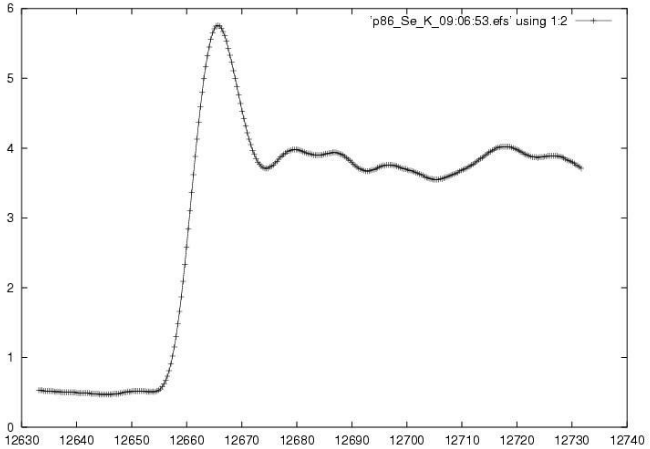


- The extent of anomalous scattering depends on the X-ray energy used
- Tunable synchrotron radiation can be used to maximize the effect of anomalous scattering
- Energies at which anomalous scattering changes abruptly are called absorption edges
- Absorption edges are element-specific; they also depend a bit on the environment.

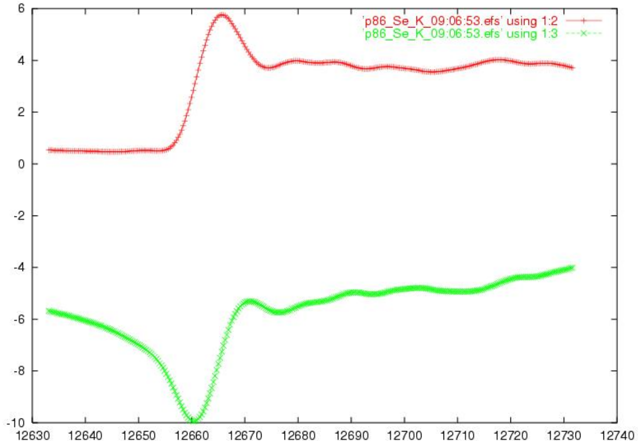


Anomalous scattering - 7





Measured fluorescence



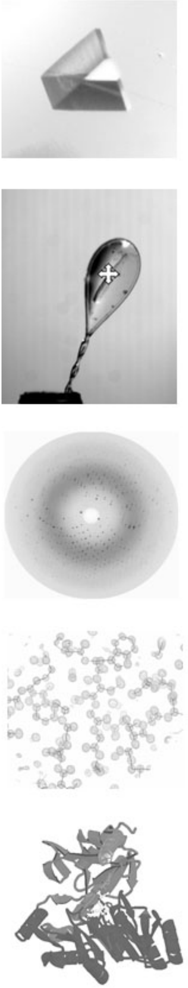
Derived anomalous scattering factors

$$f''(\omega) = \left(m\omega c / 4\pi N e^2 \right) \mu(\omega)$$

Kramers-Kronig equations

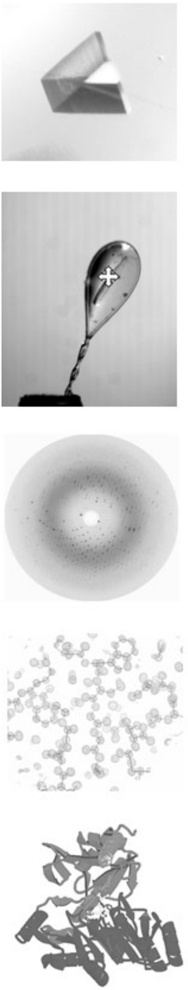
$$f'(\omega) = (2/\pi) \int_0^\infty \left[\omega' f''(\omega') / \left(\omega^2 - \omega'^2 \right) \right] d\omega'$$

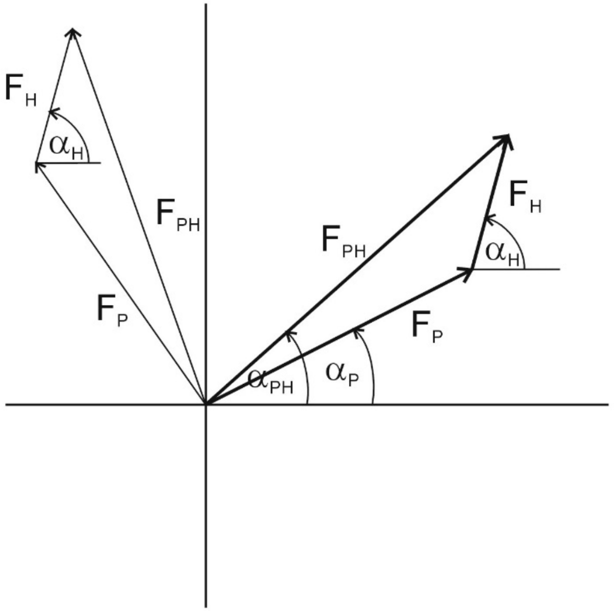
What is this good for?



1. Phase Determination
    - e.g. SIRAS (single isomorphous replacement with anomalous scattering)

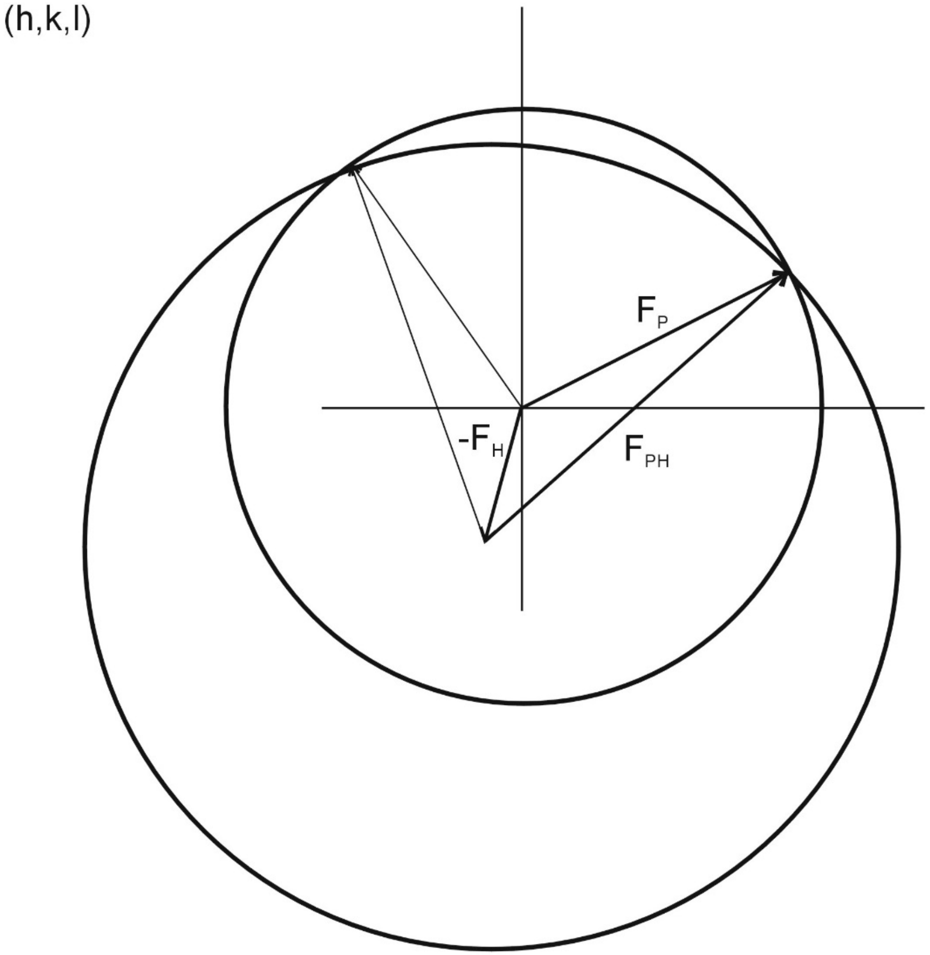
Harker Construction - SIR



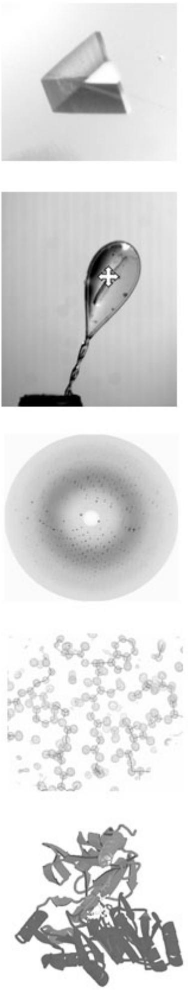


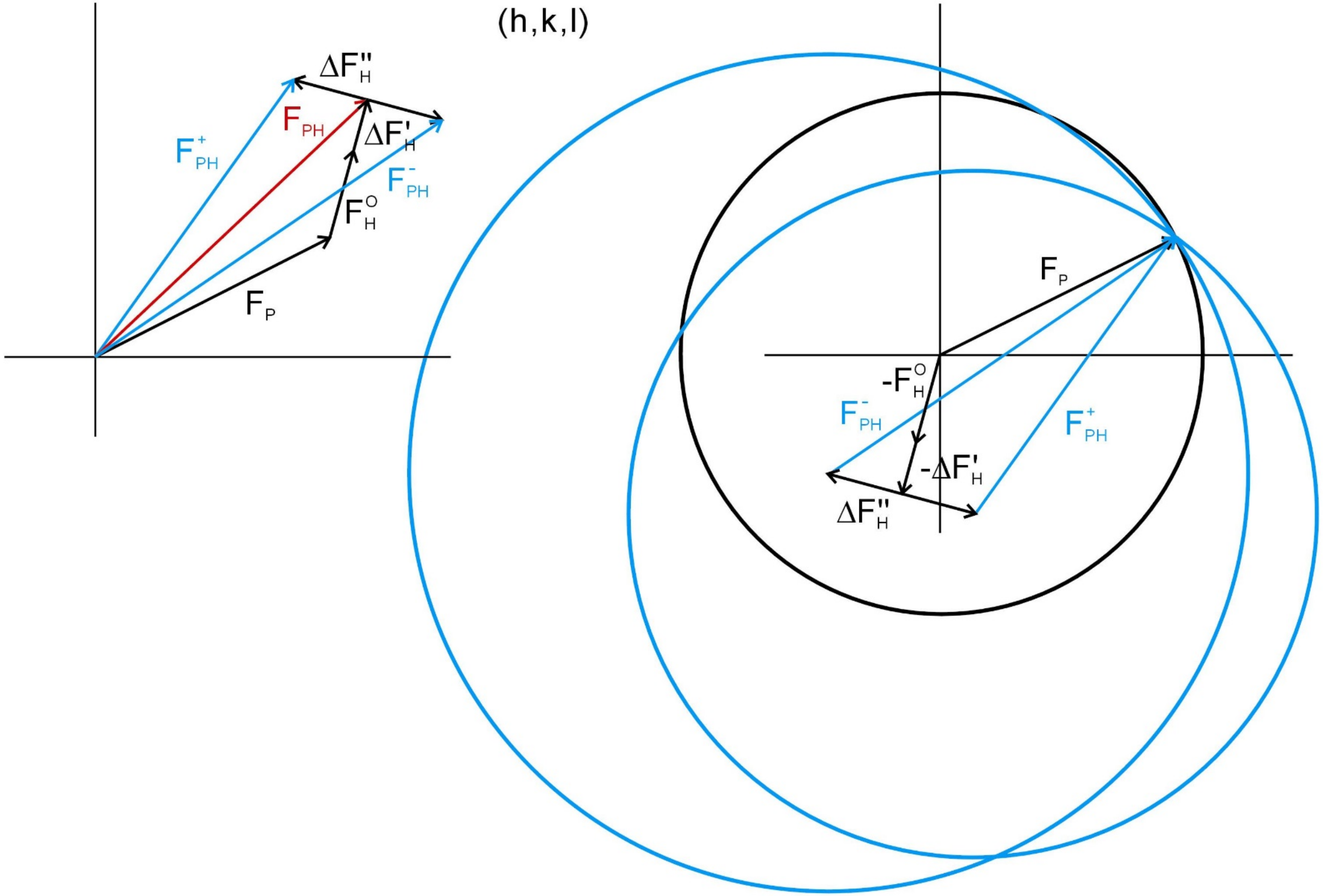
$$F_P + F_H = F_{PH}$$

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

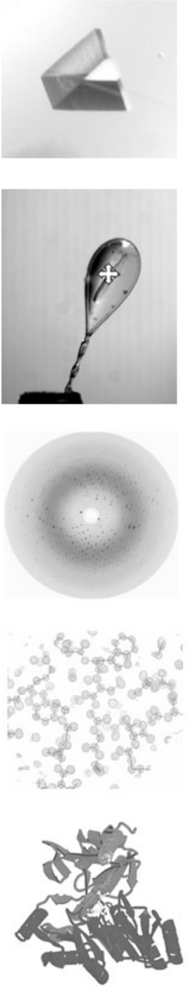


Harker Construction - SIRAS





What is this good for?



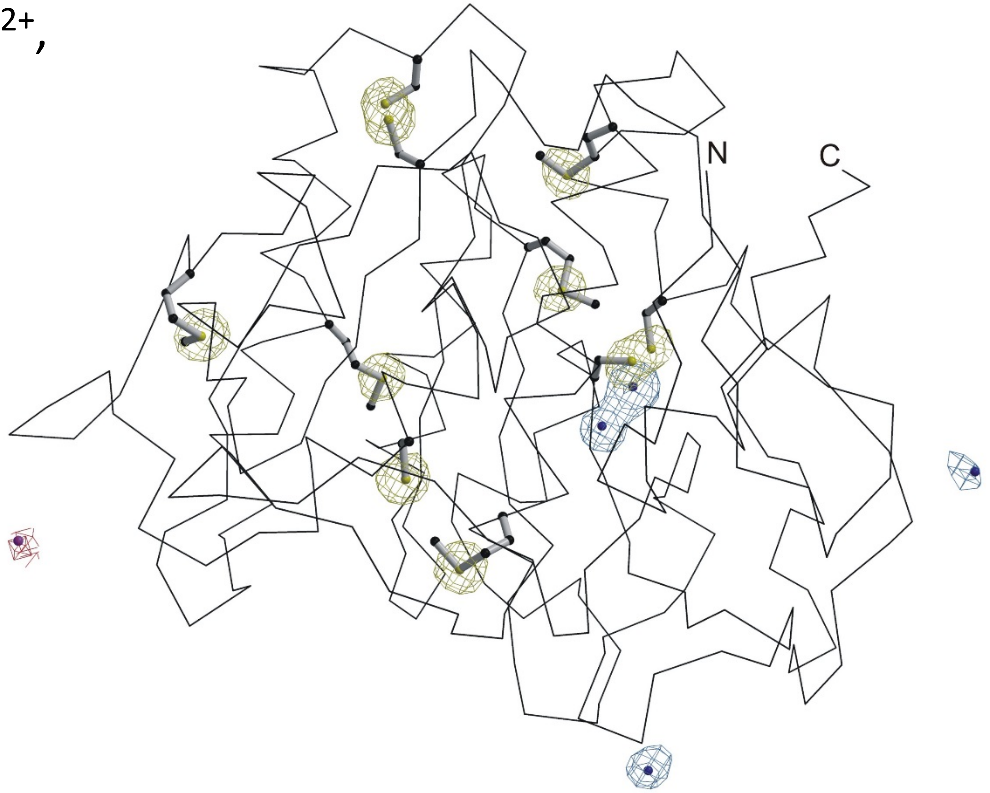
1. Phase Determination
2. Structure Validation

Proteinase K:

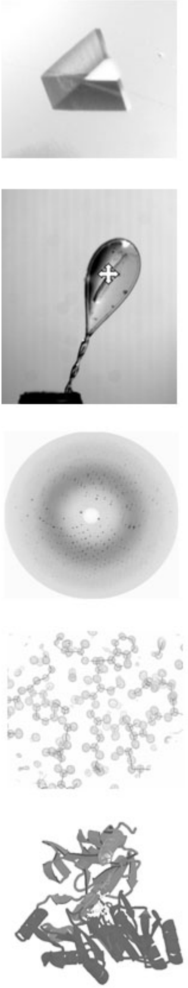
2 Ca^{2+},

2 K^{+},

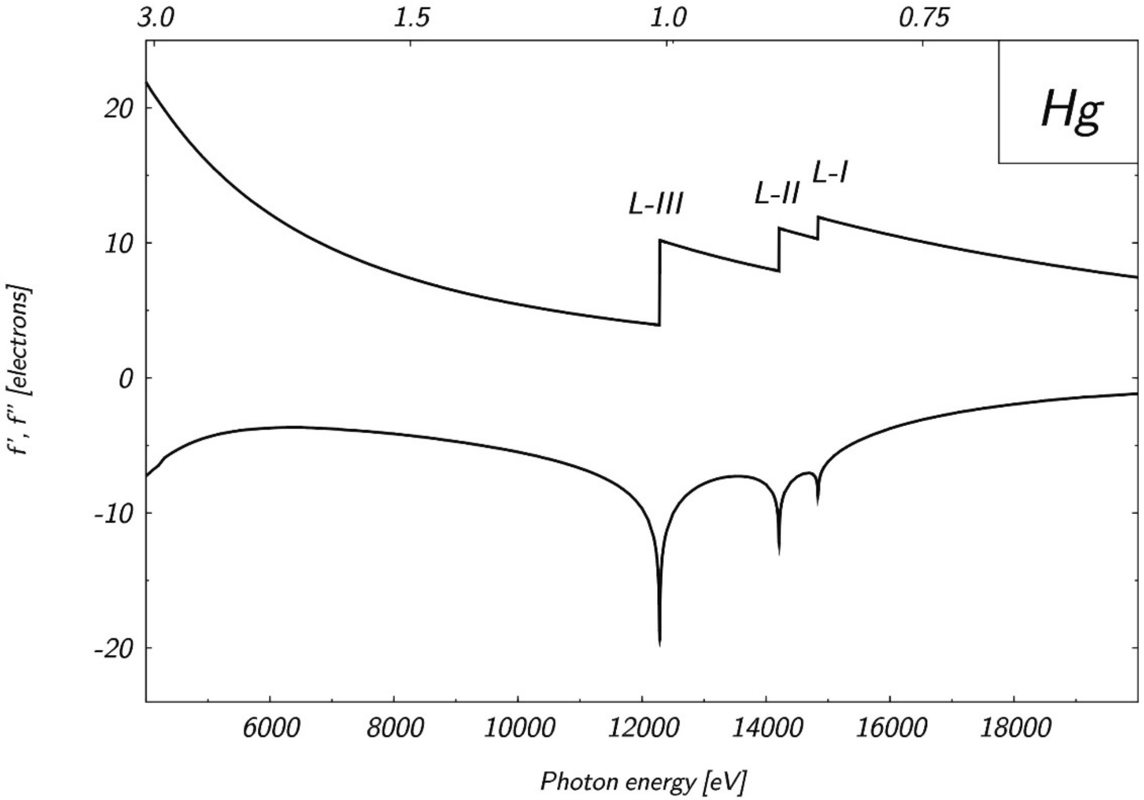
1 Cl^{-}



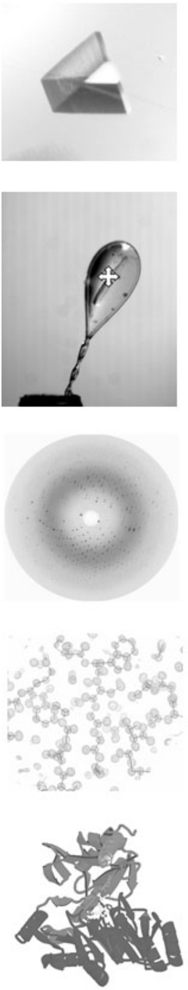
What is this good for?



1. Phase Determination
2. Structure Validation
3. Element Identification

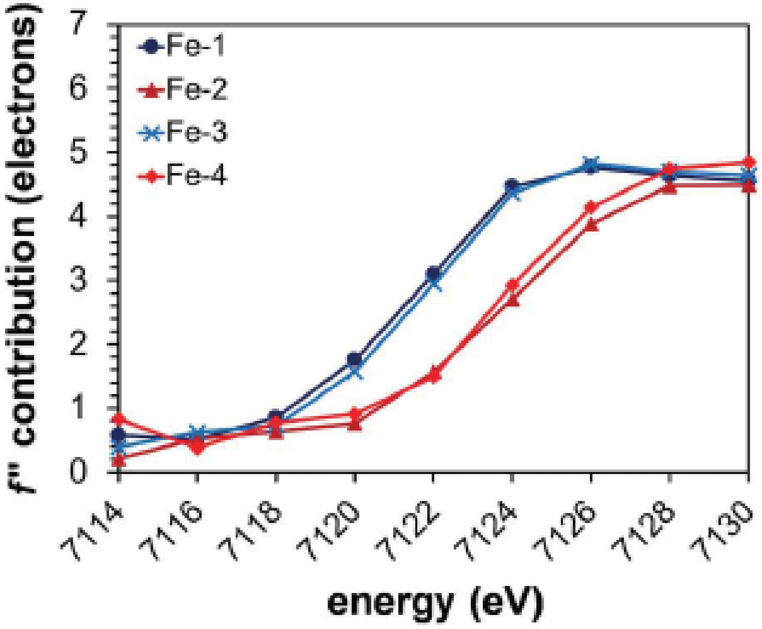


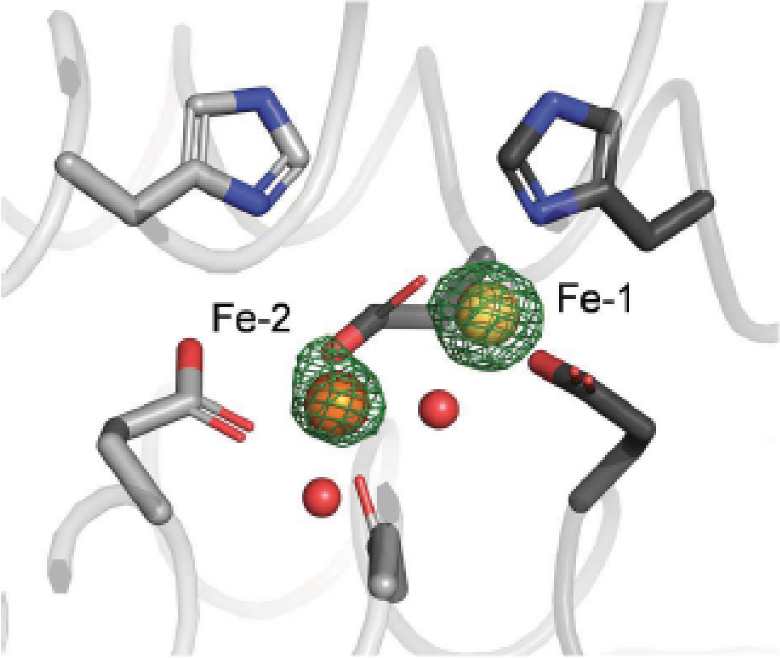
What is this good for?



1. Phase Determination
2. Structure Validation
3. Element Identification
4. Determination of the

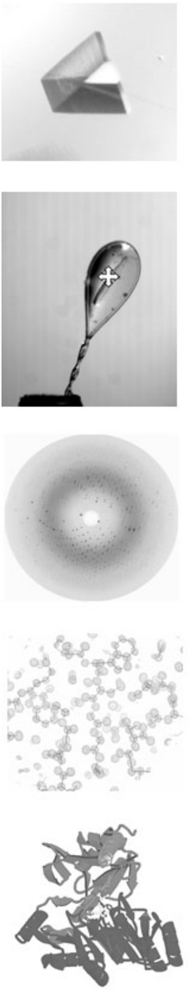
   Oxidation State of Elements

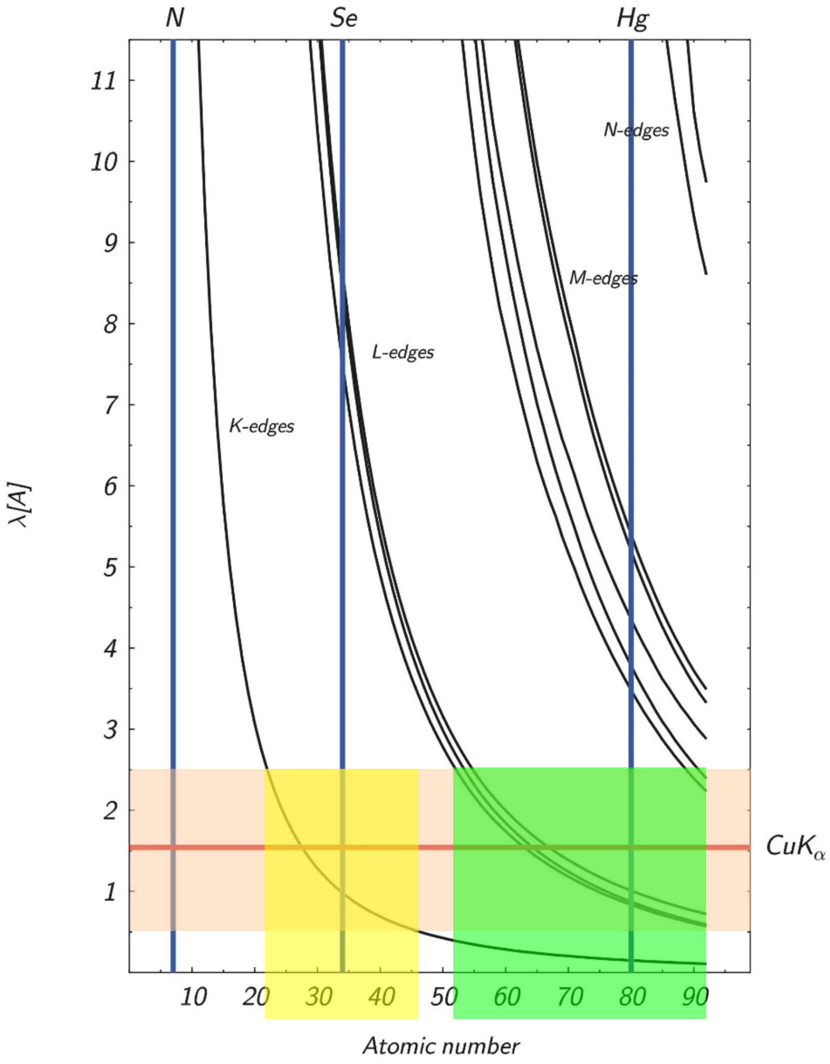


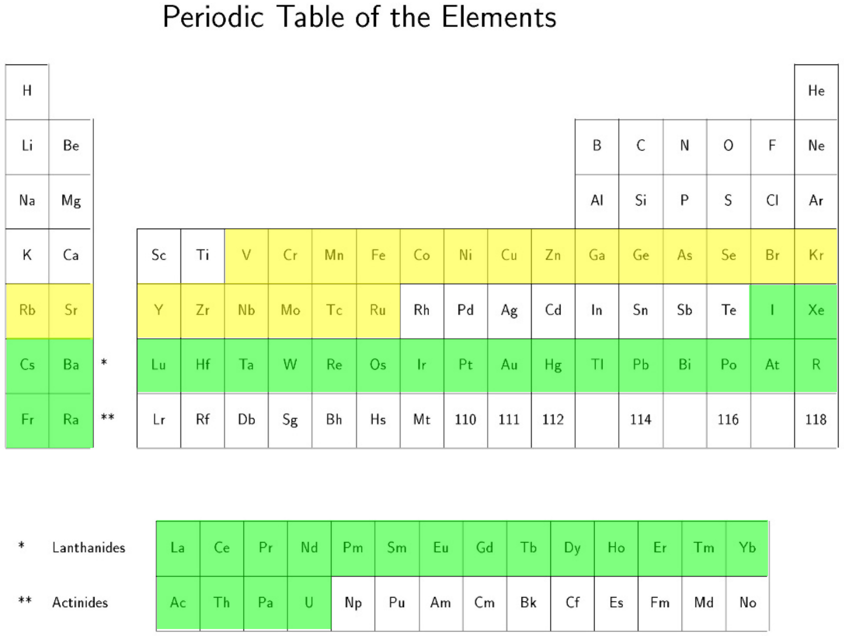


Lennartz *et al*. (2022). *Acta Cryst.* **D78**, 238-247.

Which Elemental Edges Are Accessible?





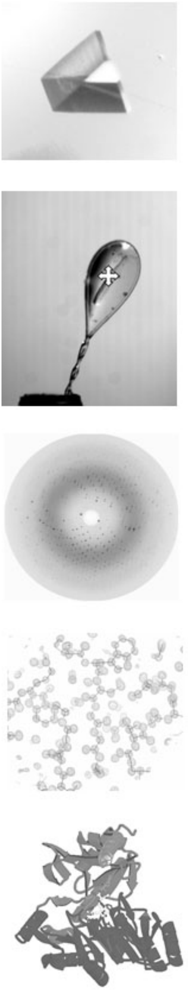


Periodic Table of the Elements

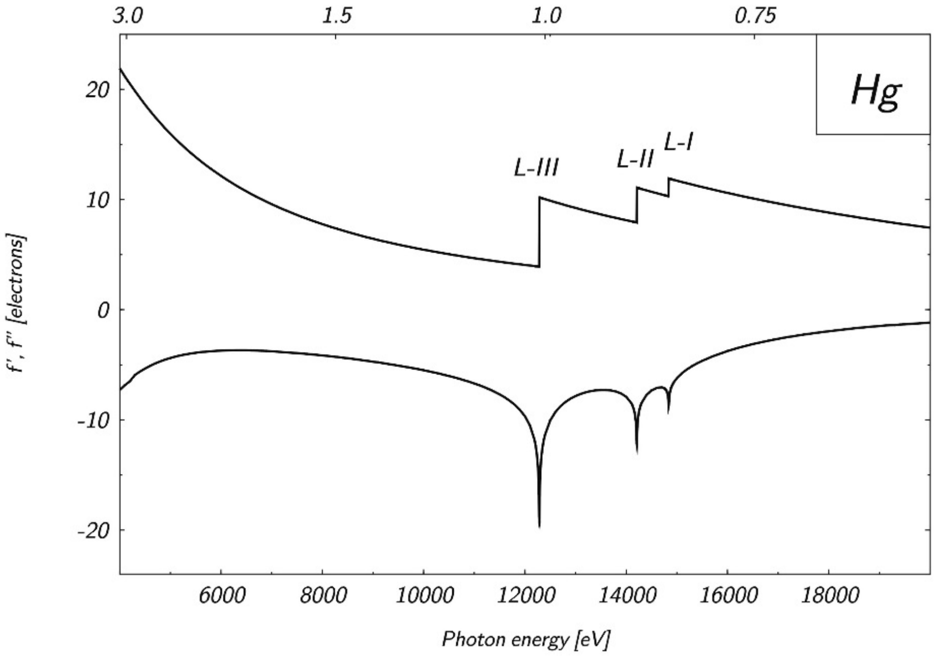
H																		He
Li	Be													B	C	N	O	F	Ne
Na	Mg													Al	Si	P	S	Cl	Ar
K	Ca		Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr		Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	*	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	R
Fr	Ra	**	Lr	Rf	Db	Sg	Bh	Hs	Mt	110	111	112		114		116		118

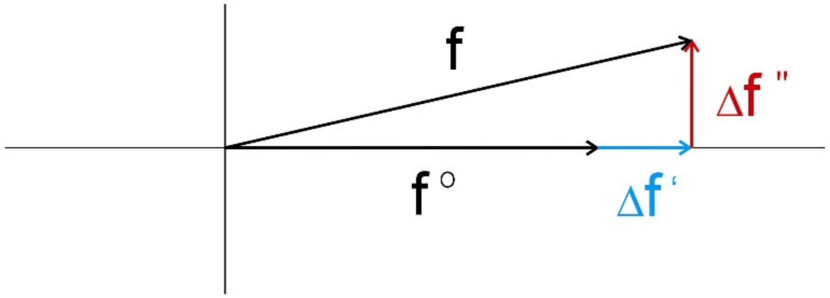
* Lanthanides	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb
** Actinides	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No

The MAD Method



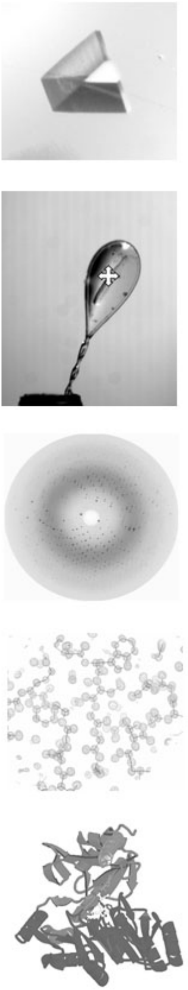
Instead of using two different crystals, we can also collect data from one crystal at two (or more) different wavelengths.

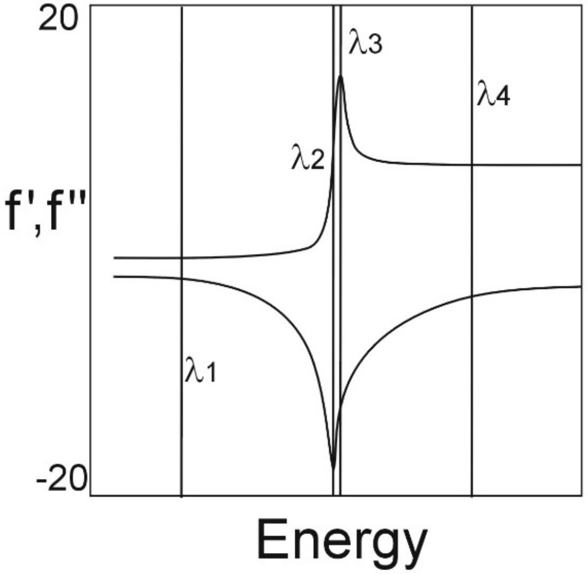


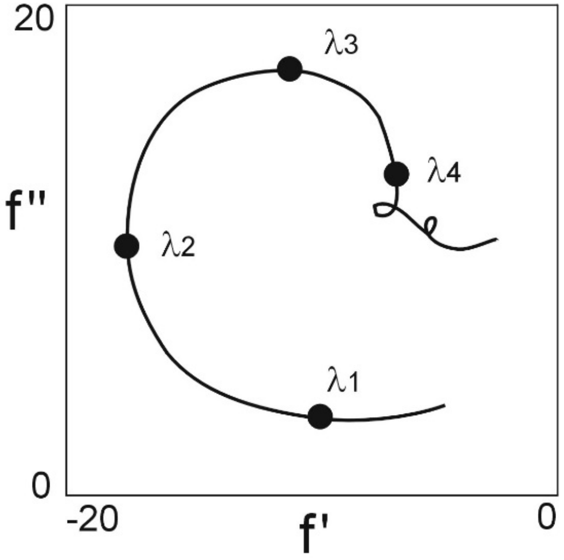


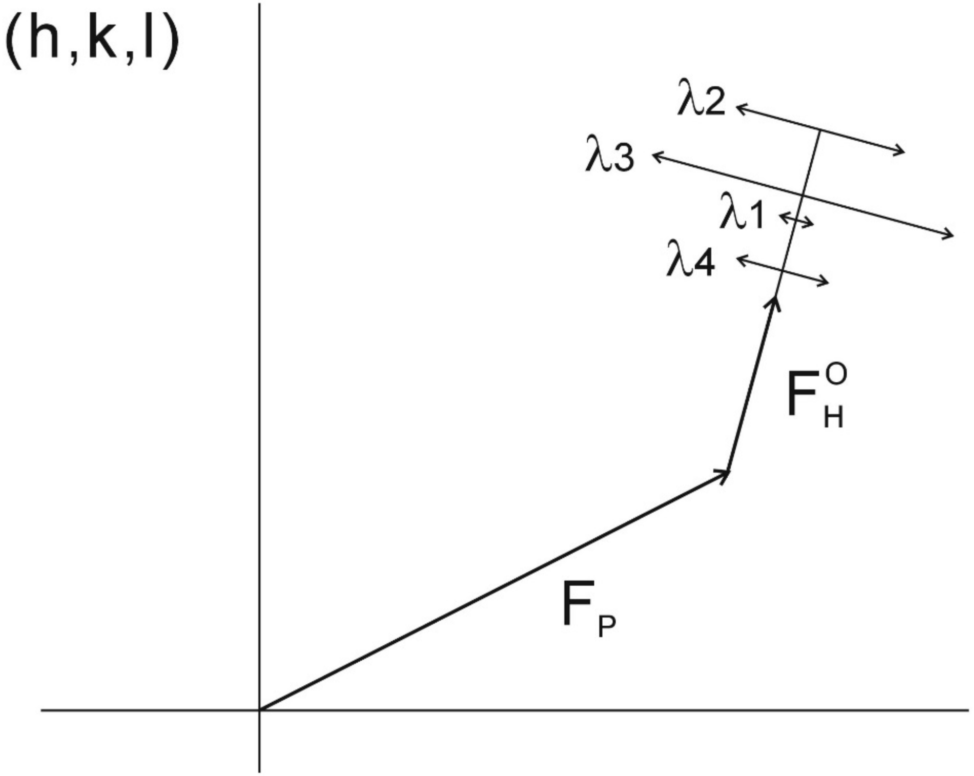
$$f = f^\circ + \Delta f' + i\Delta f''$$

The MAD Method

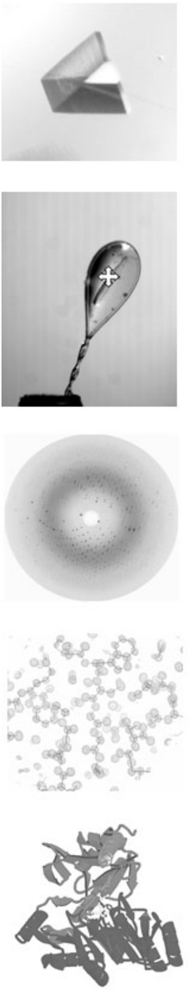






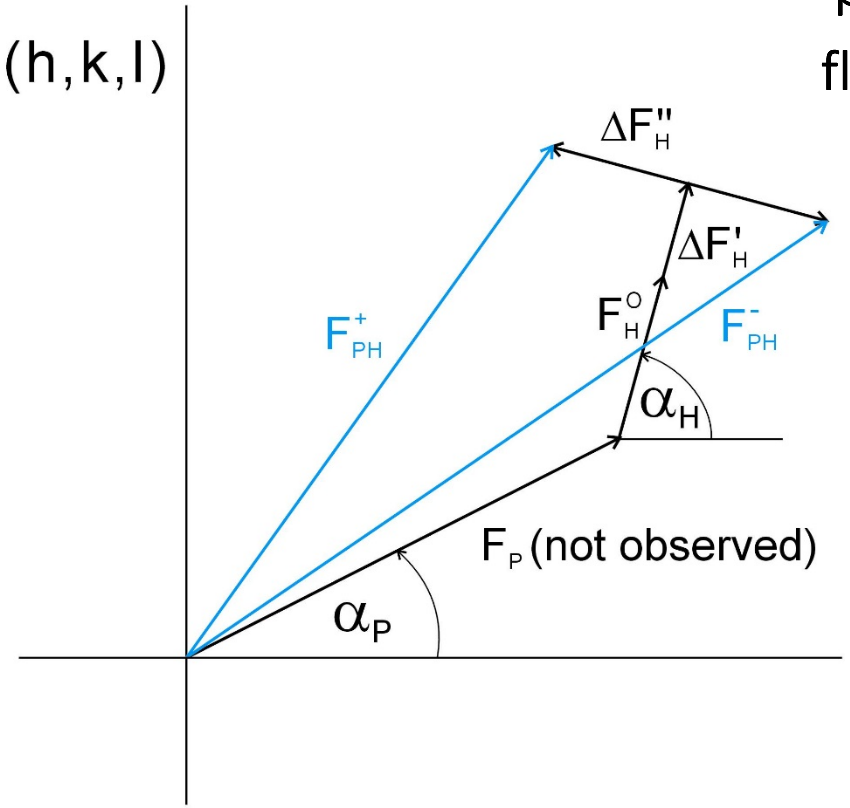


The SAD Method



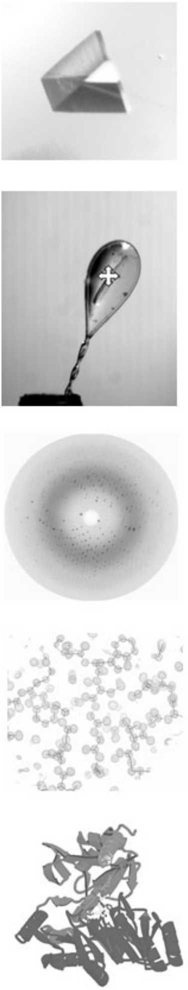
In principle, two, three or four wavelengths are not needed.

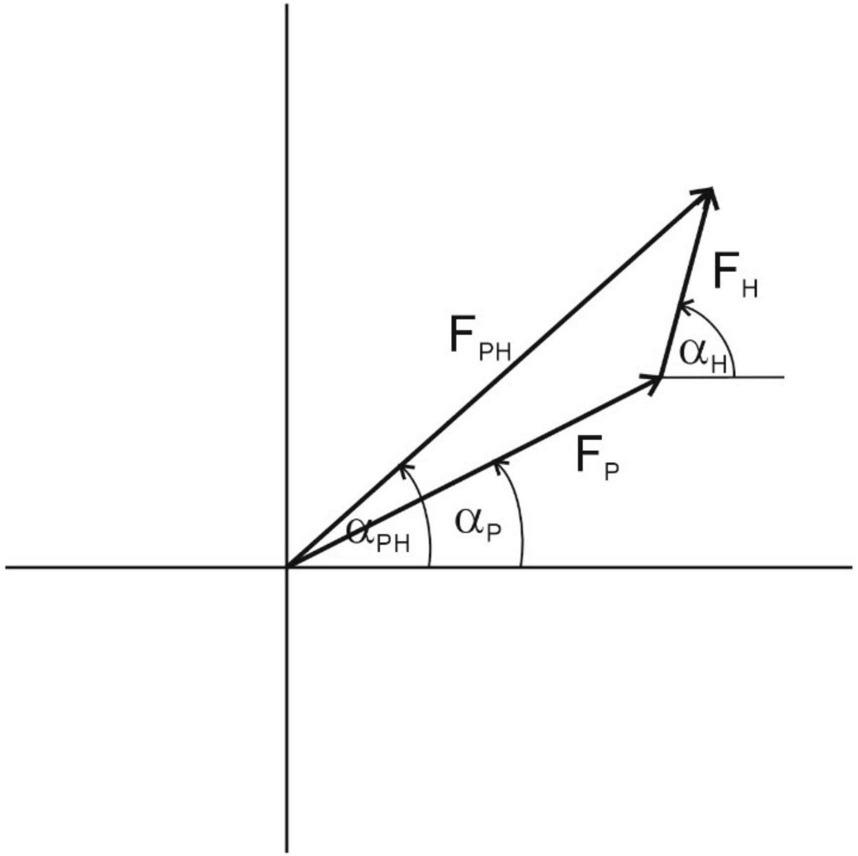
One should be enough if other information is available, e.g. possibility for solvent flattening.

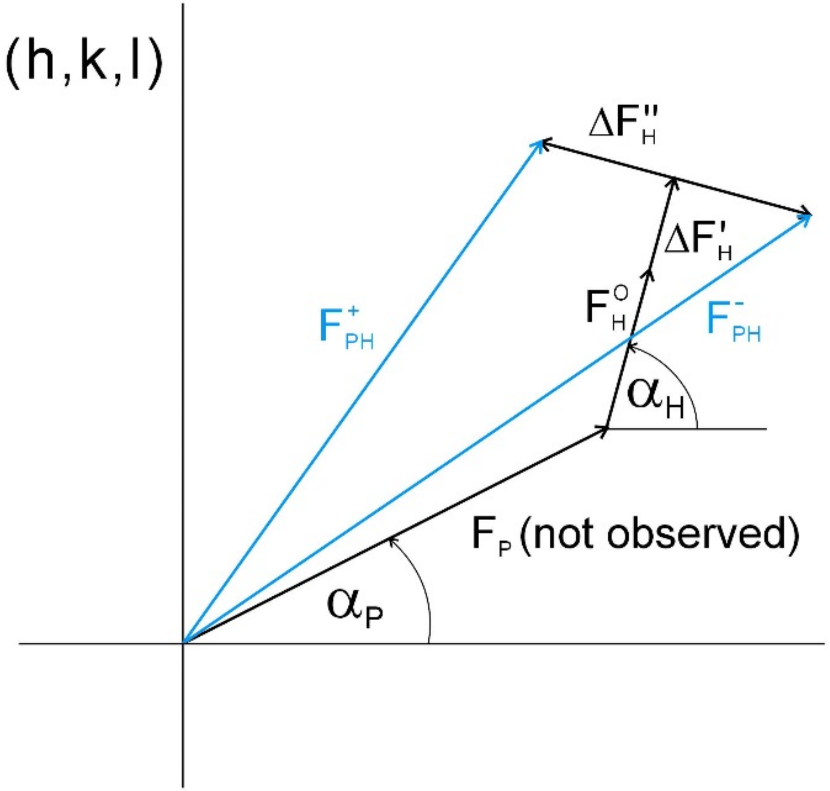


The situation is exactly analogous to the SIR case.

SIR and SAD

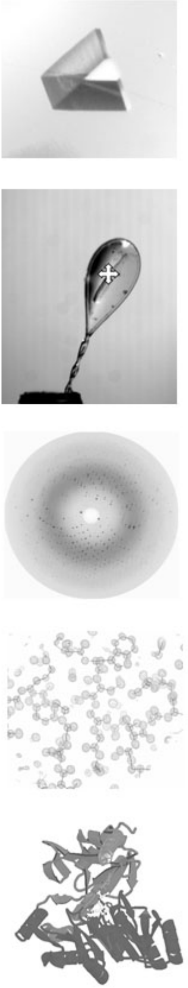


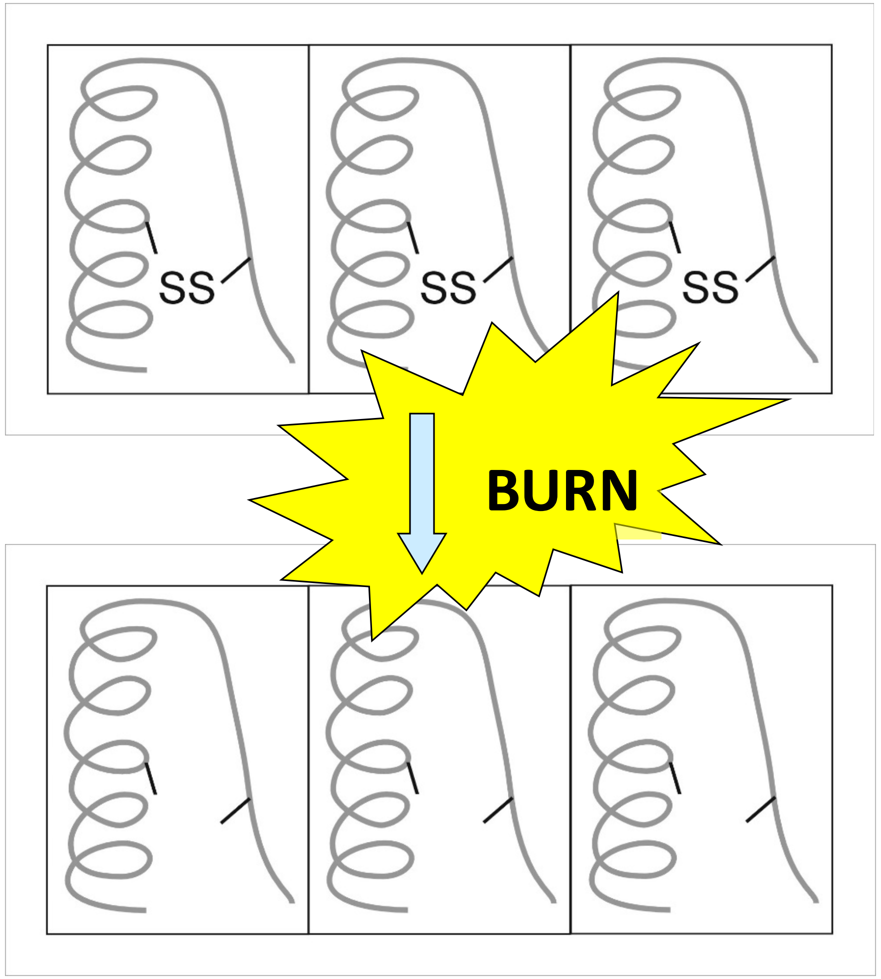

F_{PH}
F_H
α_H
F_P
α_{PH}
α_P


(h,k,l)
$\Delta F''_H$
$\Delta F'_H$
F^O_H
F^+_{PH}
F^-_{PH}
α_H
F_P (not observed)
α_P

The RIP Method

Radiation damage induced phasing is a relatively new method.





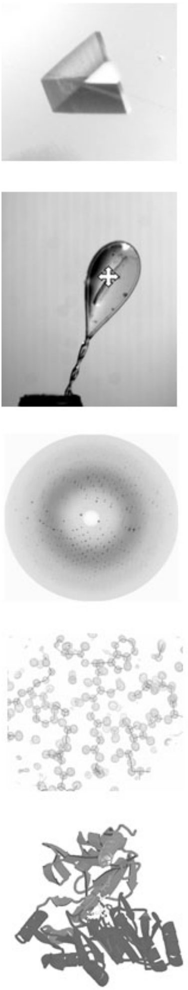
$\rightarrow$ $h, k, l, |F_{before}(hkl)|$

$\alpha(hkl)$

$\rightarrow$ $h, k, l, |F_{after}(hkl)|$

The RIP method is in principle applicable to all crystals, which are damaged by X-rays (or UV).

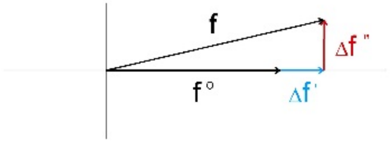
Summary



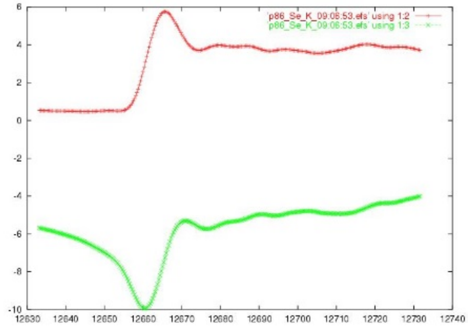
- Two Fundamental Equations in Crystallography,

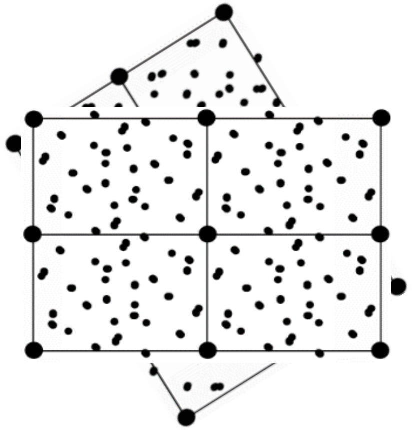
F(hkl) and ρ(x,y,z)

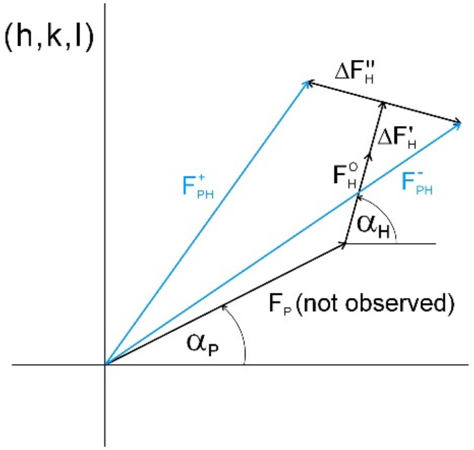
- The Phase Problem
- Normal and Anomalous Scattering
- Determination of Δf ’ and Δf ’’
- Phase Determination Methods (MR, SIR(AS), MIR(AS), MAD, SAD, RIP(AS))

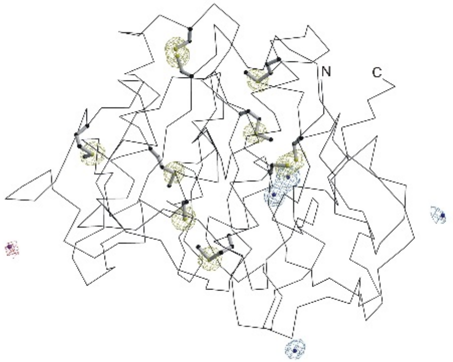


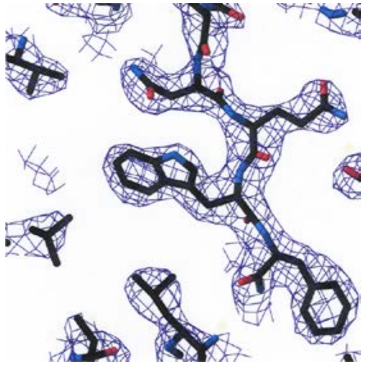
$f = f^{\circ} + \Delta f' + i\Delta f''$

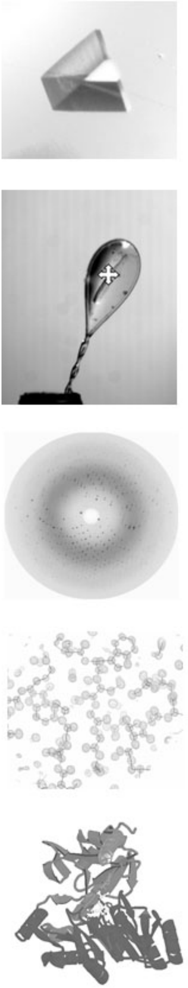













Thank you!