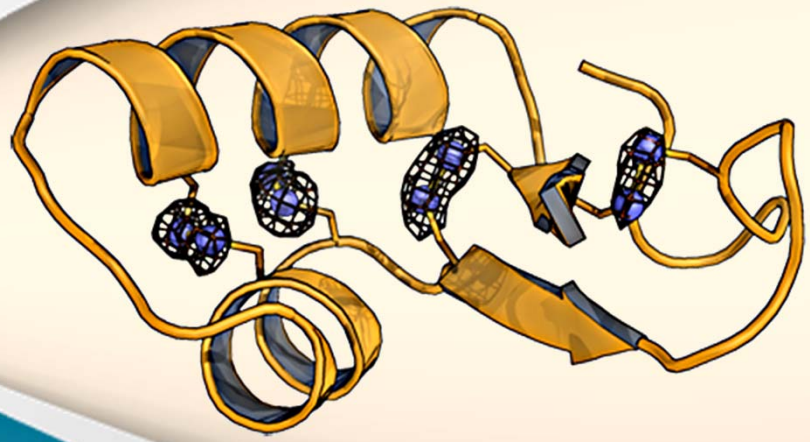
SHELXC/D/E

Andrea Thorn



What is experimental phasing?

Experimental phasing is what you do if MR doesn‘t work.

What is experimental phasing?

Experimental phasing methods depend on intensity differences.

These differences are caused by a marker substructure of certain elements.

MAD and **SAD** exploit the anomalous signal from one or more data sets from the same crystal.

SIR (special case: **RIP**) and **MIR** utilizes several heavy-atom soaked derivative crystals. They have to be isomorphous to be utilized.

Methods

- Single wavelength anomalous diffraction (**SAD**)
  - Native sulfur-based SAD (**S-SAD**)
- Multiple wavelength anomalous diffraction (**MAD**)
- Single isomorphous replacement (**SIR**)
  - Radiation-induced phasing (**RIP**)
- Multiple isomorphous replacement (**MIR**)
- Single isomorphous replacement with anomalous scattering (**SIRAS**)
- Multiple isomorphous replacement with anomalous scattering (**MIRAS**)

If you use SHELX...

SHELXC: α calculation, data analysis, file preparation

SHELXD: Substructure search

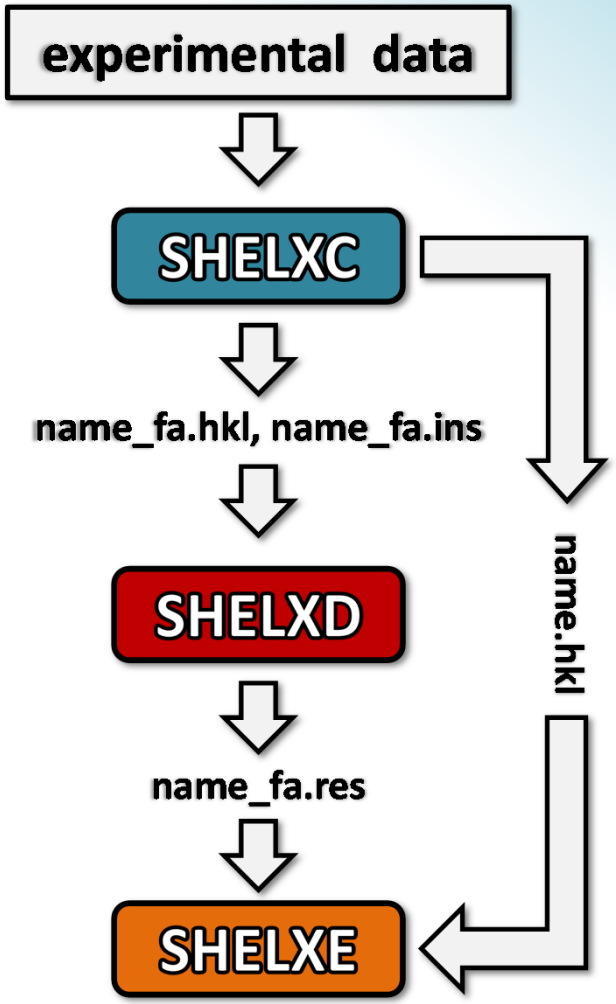
SHELXE: Density modification, tracing*

**** A traced structure is solved; CC (trace against native data) > 25% (for data < 2.5 Å)***

[**ANODE**: Validation]

Pipeline?

Other experimental phasing programs should be considered , in particular for ease of use or problem cases**.



** http://strucbio.biologie.uni-konstanz.de/ccp4wiki/index.php/Experimental_phasing

Theory

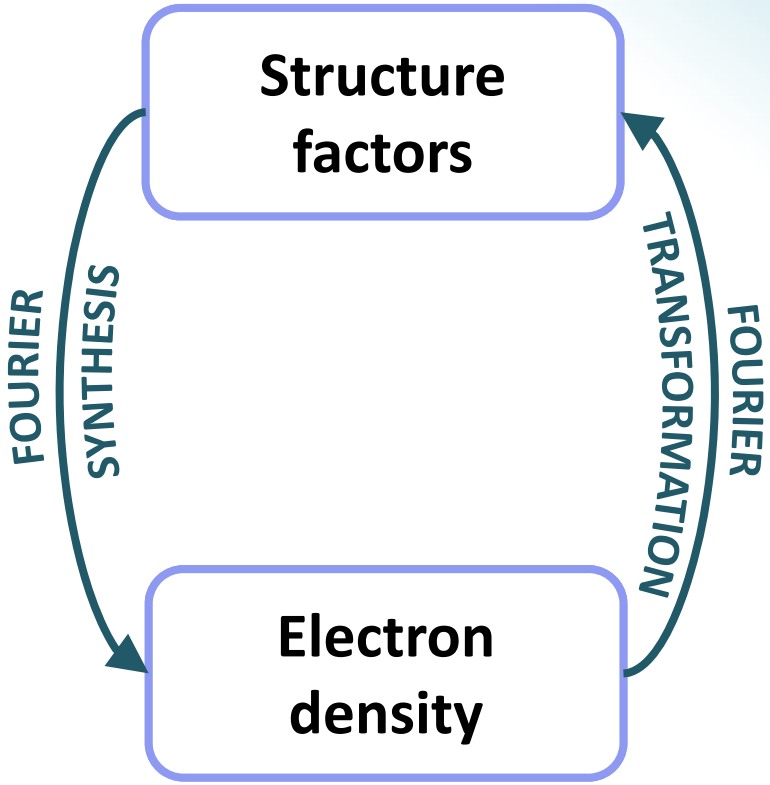
STRUCTURE FACTORS

Structure factors

For each reflection, there is a

structure factor F_{hkl}

If we know the structure factors including their phases for all reflections, we can easily calculate the electron density map, and hence get the structure.

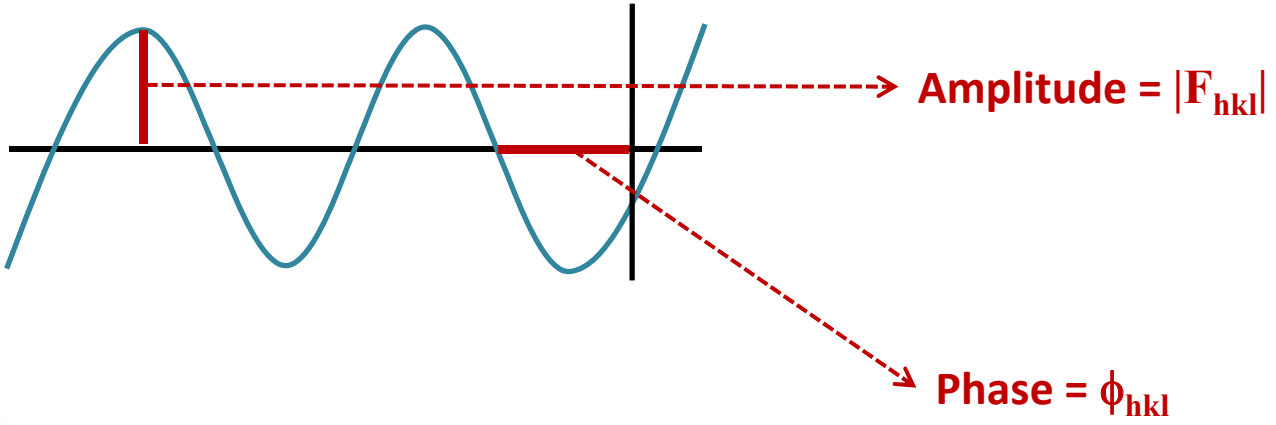


Structure factors

For each reflection, there is a

structure factor F_{hkl}

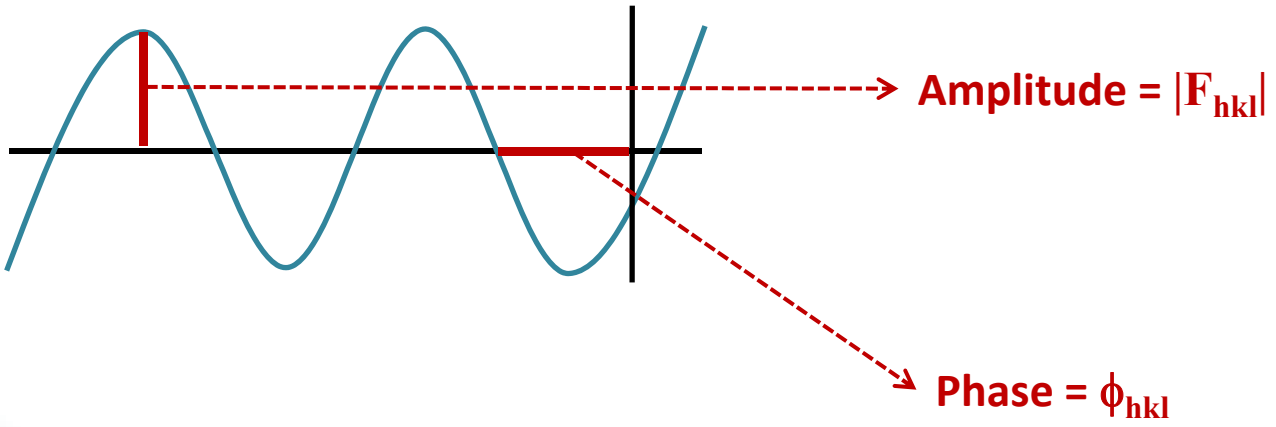
= a wave



Structure factors

structure factor $\mathbf{F_{hkl}}$

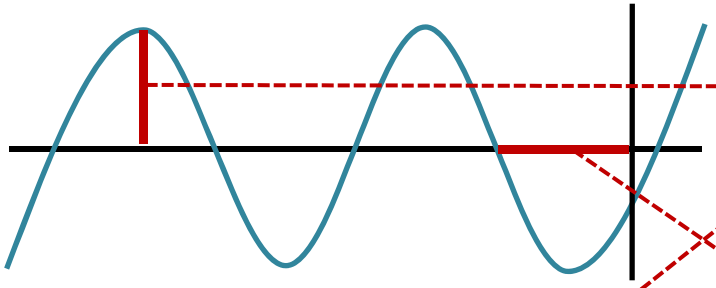
= a wave



Structure factors

structure factor F_{hkl}

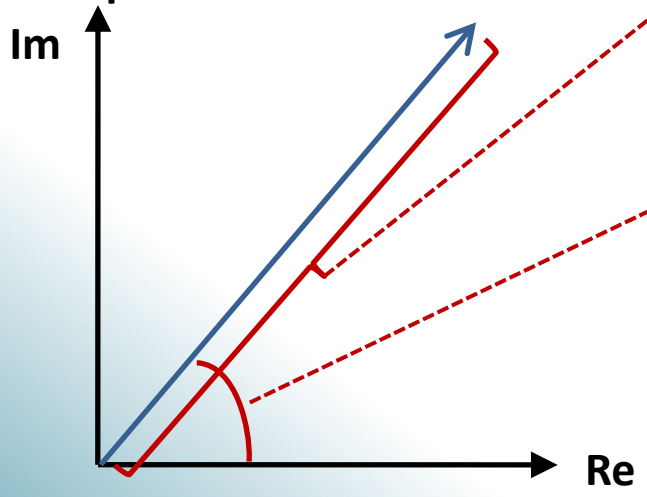
= a wave



Amplitude = $|F_{hkl}|$

$|F_{hkl}|^2 \sim I_{hkl}$ Intensity ☑

= a complex number



Phase = ϕ_{hkl}

cannot be measured... :-(

Structure factors

Amplitude = $|F_{hkl}|$

$|F_{hkl}|^2 \sim I_{hkl}$ **Intensity ☑**

Phase = ϕ_{hkl}

cannot be measured... :-(

PHASE PROBLEM

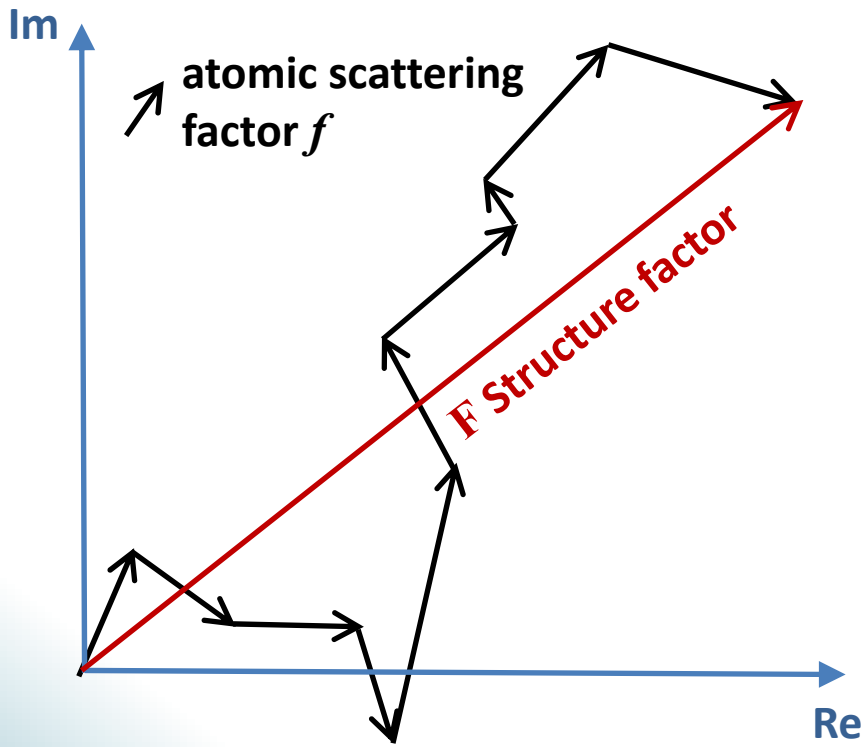
The central problem of crystallography

Theory

ANOMALOUS SCATTERING

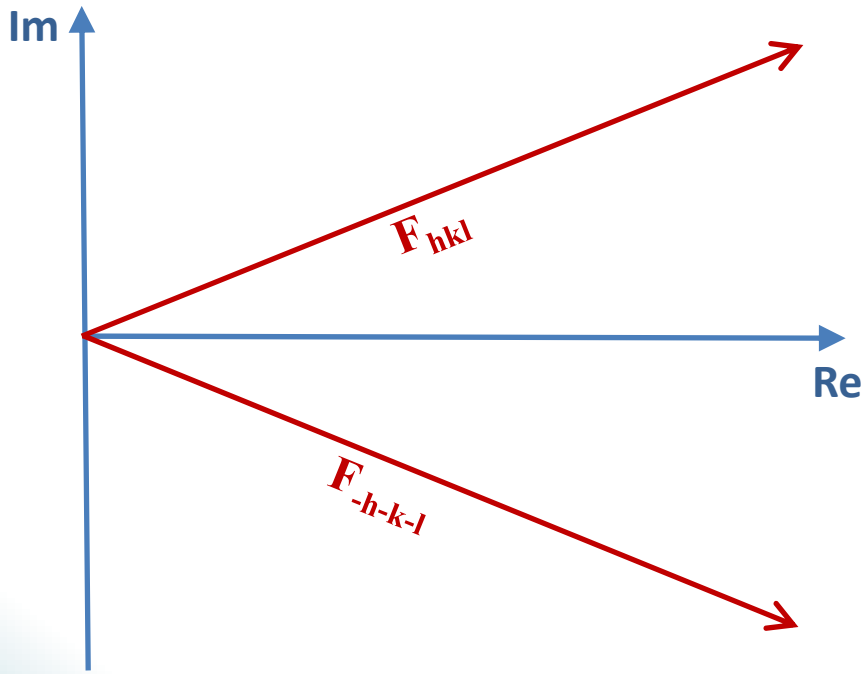
The anomalous signal

Each structure factor is composed of contributions f from each atom:



The anomalous signal

Friedel's law: $|F_{hkl}| = |F_{-h-k-l}|$ $\phi_{hkl} = -\phi_{-h-k-l}$



The anomalous signal

But in reality, there is **anomalous scattering** due to resonance with electronic transitions in the atom:

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

depends solely on resolution and element

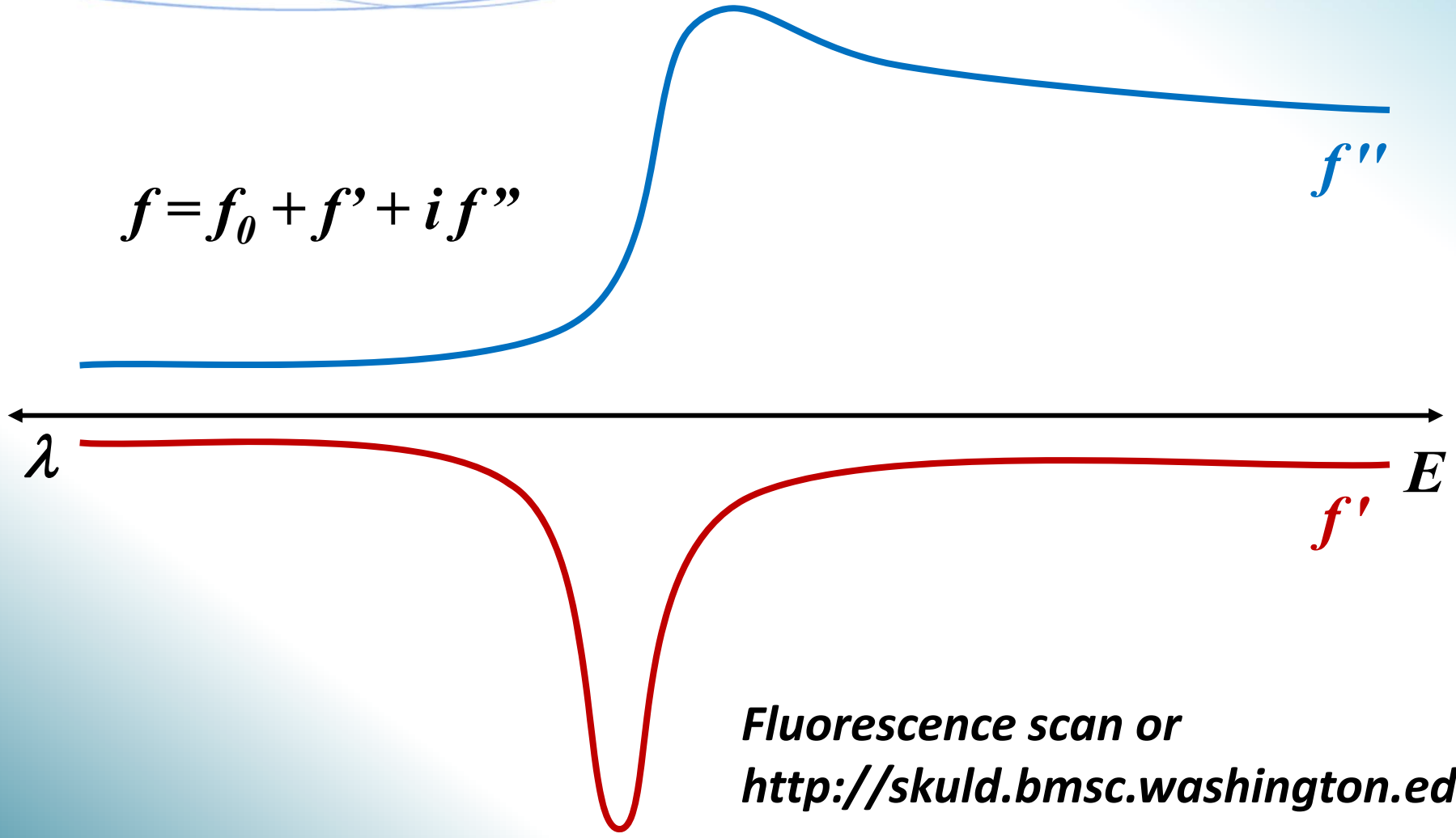
real component

imaginary component

f' and f'' are observed near absorption edges of the atom's element, and are λ-dependent

The anomalous signal

$$f = f_0 + f' + i f''$$



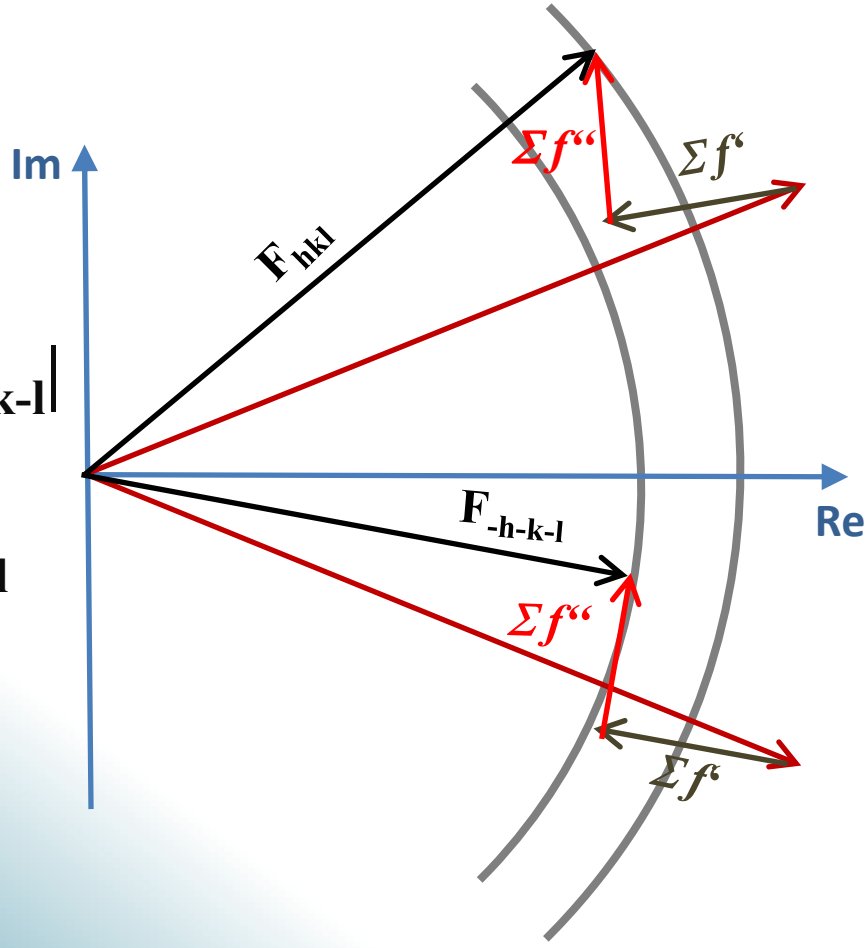
Fluorescence scan or
http://skuld.bmsc.washington.edu

The anomalous signal

f'' breaks Friedel's law:

$$|F_{hkl}| \neq |F_{-h-k-l}|$$

$$\phi_{hkl} \neq -\phi_{-h-k-l}$$



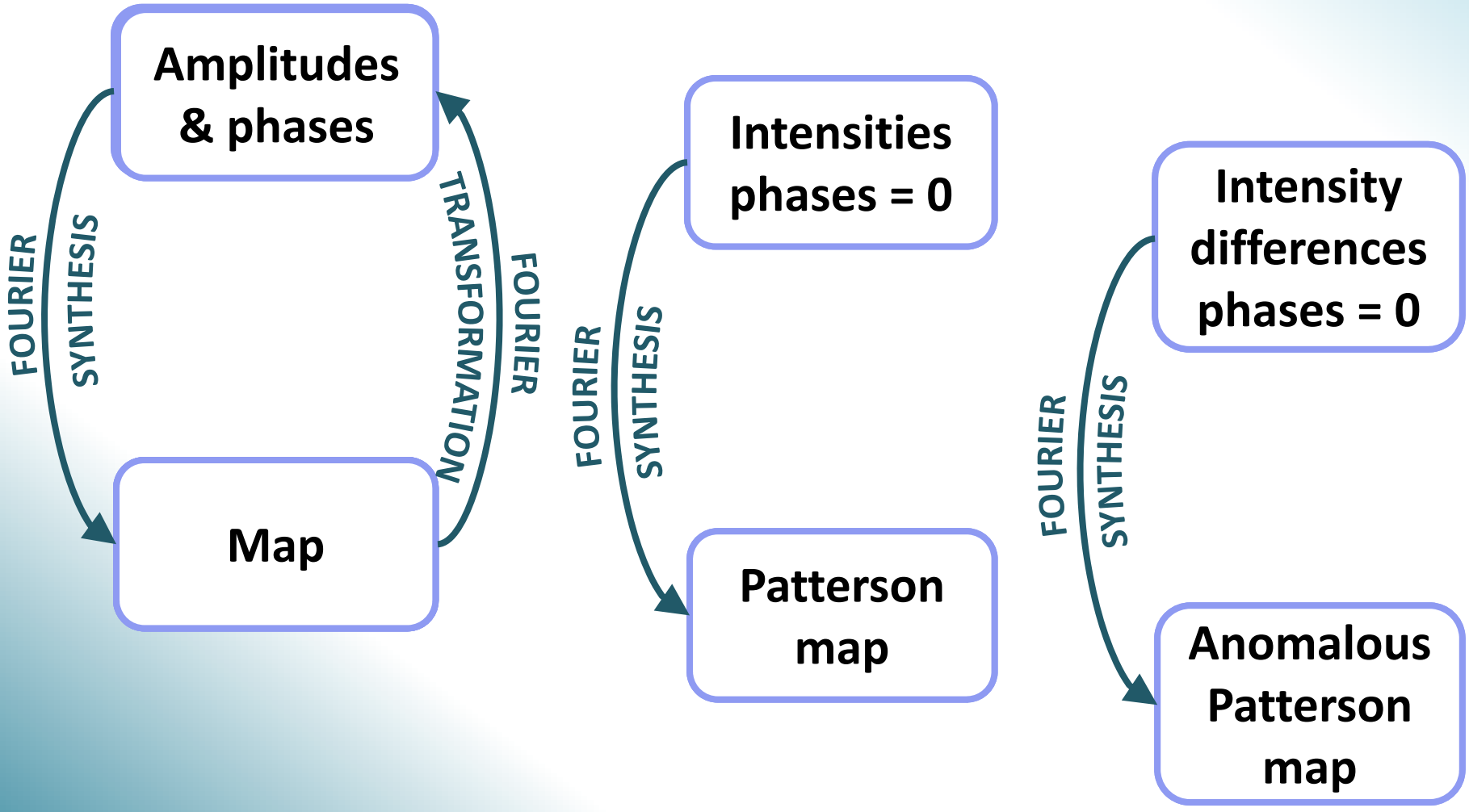
The intensities of Friedel pairs no longer have the same intensity!

This can be used for the absolute structure determination and for experimental phasing!

How to...

SUBSTRUCTURE SEARCH IN SHELXD

Calculating a map - Patterson



Substructure search

Finding the substructure of marker atoms

- Direct methods
- Patterson methods

Borrowed from small molecule crystallography

- These methods require separate atomic electron densities to locate atoms.
- They work here because the marker atoms have large interatomic distances.
- Disulfides become ‚supersulfurs‘.

Direct methods

- Phases of strong reflections are related (as a result of the non-random distribution of atoms.)
  - Triplett equations
  - Sayre equation
- Relations are relatively easy to resolve for few atoms.
- Usage of normalized structure factors (E values):

$$|E_{hkl}|^2 = \frac{|F_{hkl}|^2/\epsilon}{<|F_{hkl}|^2/\epsilon>}$$

$\boldsymbol{\varepsilon}$ scale factor for proper treatment of special position reflections

$<|\boldsymbol{F}_{\boldsymbol{hkl}}|^2/\boldsymbol{\varepsilon}>$ mean per resolution shell

Substructure search

- **Patterson seeding:** The Patterson map contains all interatomic distance vectors between marker atoms. This can be used as a starting point (‚Patterson seeding‘)
- **Dual space direct methods** recycle and modify trial substructures by peak search in the electron density and refining phases in reciprocal space. Convergence is faster than in reciprocal space alone.

Substructure search

An **overdetermined** problem with **noisy** data...

Critical factors in substructure search:

- Resolution range highly affects the outcome
- Good data quality
- Intensity outliers are problematic
- Scaling (also anisotropic scaling) is needed

BEWARE: Handedness is not resolved at this stage!
(Density modification differentiates later.)

How many atoms to search for?

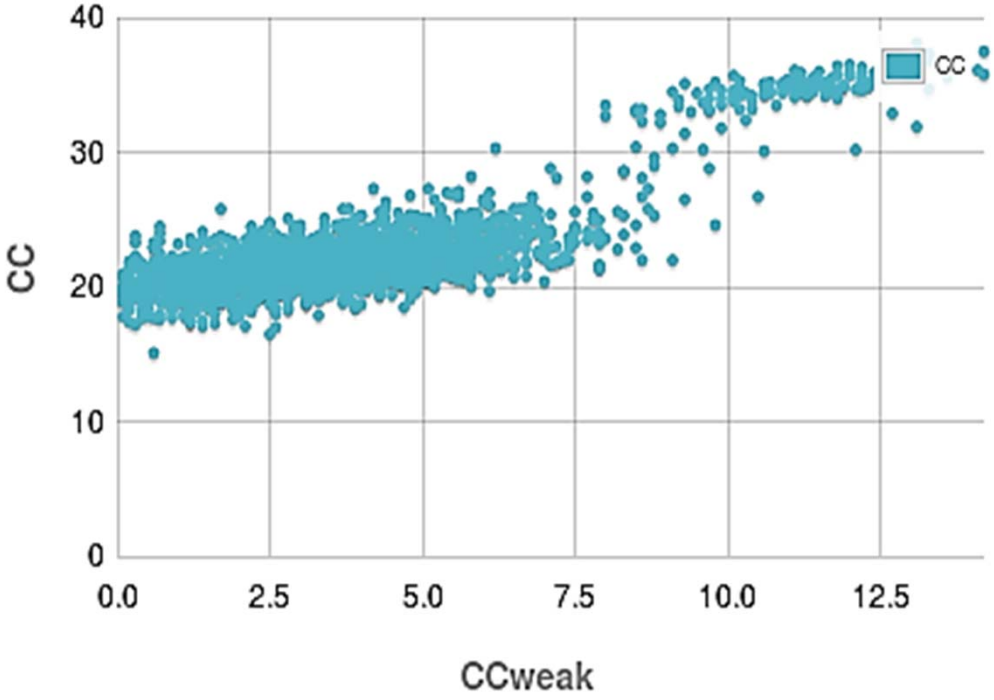
Number of atoms to search for critical.
But how many to search for?

- **Iodine soaks**: (no. of amino acid residues in ASU) / 15
- **RIP:** number of disulphide bridges good starting value.
- **S-SAD**: number of sulphur atoms. Disulphides should be given as one instead of two peaks at resolution > 2.03 Å

How to know if SHELXD worked?

- ***CFOM*** (combined figure of merit) = $\boldsymbol{CC_{All} + CC_{Weak}}$
- ***CFOM* varies for a correct solution**
- clear drop in marker atom occupancies
- substructure with known geometry

The ultimate indication is a successful density modification resulting in an interpretable map!

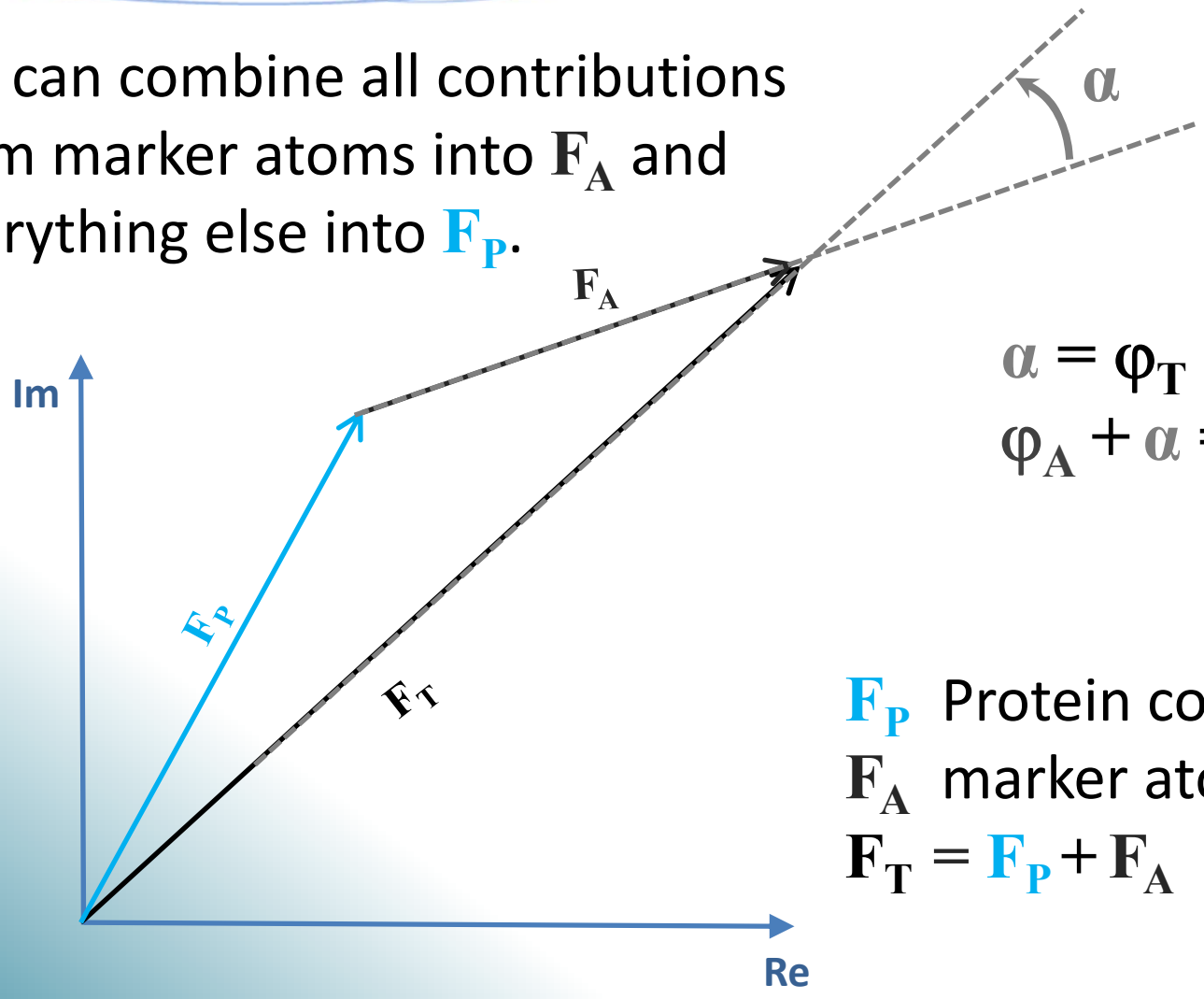


How to...

PHASING THE REST (SHELXC)

From substructure to structure

We can combine all contributions from marker atoms into $\mathbf{F_A}$ and everything else into $\mathbf{F_P}$.



$$\alpha = \varphi_T - \varphi_A$$
$$\varphi_A + \alpha = \varphi_T$$

$\mathbf{F_P}$ Protein contribution
$\mathbf{F_A}$ marker atom contribution
$\mathbf{F_T} = \mathbf{F_P} + \mathbf{F_A}$

From substructure to structure

So, if we would know the anomalous scatterer positions (or heavy atom positions), we could calculate $\mathbf{F_A}$ **:**

$$\alpha = \varphi_T - \varphi_A$$

$$\varphi_A + \alpha = \varphi_T$$

If we could then get $\boldsymbol{\alpha}$, we could calculate φ_T and **solve the phase problem**!

From substructure to structure

Phasing equations If we would have no errors...

$$|F_{hkl}|^2 = |F_T|^2 + a|F_A|^2 + b|F_T||F_A|\cos\alpha + c|F_T||F_A|\sin\alpha$$

$$|F_{-h-k-l}|^2 = |F_T|^2 + a|F_A|^2 + b|F_T||F_A|\cos\alpha - c|F_T||F_A|\sin\alpha$$

$$a = \frac{f''^2 + f'^2}{f_0^2} \quad b = \frac{2f'}{f_0} \quad c = \frac{2f''}{f_0}$$

F_T Total structure factor

F_A Marker substructure structure factor

$$\alpha = \phi_T - \phi_A$$

From substructure to structure

Phasing equations

$$|F_{hkl}|^2 = |F_T|^2 + a|F_A|^2 + b|F_T||F_A|\cos\alpha + c|F_T||F_A|\sin\alpha$$

$$|F_{-h-k-l}|^2 = |F_T|^2 + a|F_A|^2 + b|F_T||F_A|\cos\alpha - c|F_T||F_A|\sin\alpha$$

For each wavelength, we have different **a**, **b**, **c** and two observations. $|\mathbf{F_A}|$, $|\mathbf{F_T}|$ and **α** are unknown. So given good data from at least two wavelengths, the equation can be solved. This would be **MAD** then, and works best if the f' differences and the sum of f'' values would be large!

From substructure to structure

Phasing equations

$$|F_{hkl}|^2 = |F_T|^2 + a|F_A|^2 + b|F_T||F_A|\cos\alpha + c|F_T||F_A|\sin\alpha$$

$$|F_{-h-k-l}|^2 = |F_T|^2 + a|F_A|^2 + b|F_T||F_A|\cos\alpha - c|F_T||F_A|\sin\alpha$$

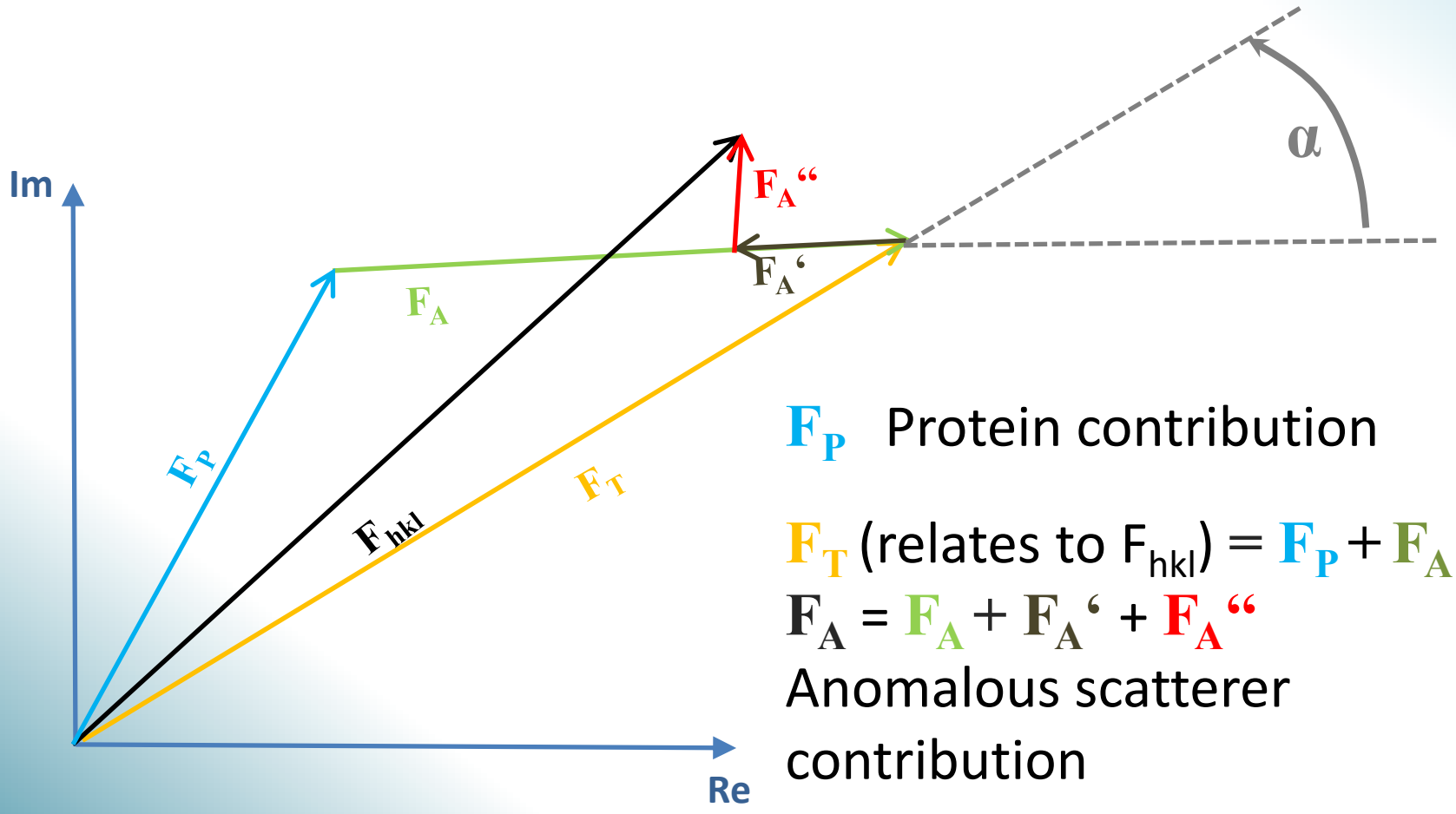
In a **SAD** experiment, we have only two observables, as we measured only one wavelength. So we assume

$$\mathbf{|F_T| = 0.5\,(|F_{hkl}| + |F_{-h-k-l}|)}$$ and get

$$\mathbf{|F_{hkl}| - |F_{-h-k-l}| = c|F_A|\sin\alpha}$$

This is sufficient for the substructure and estimation of $\boldsymbol{\varphi_T}$!

From substructure to structure



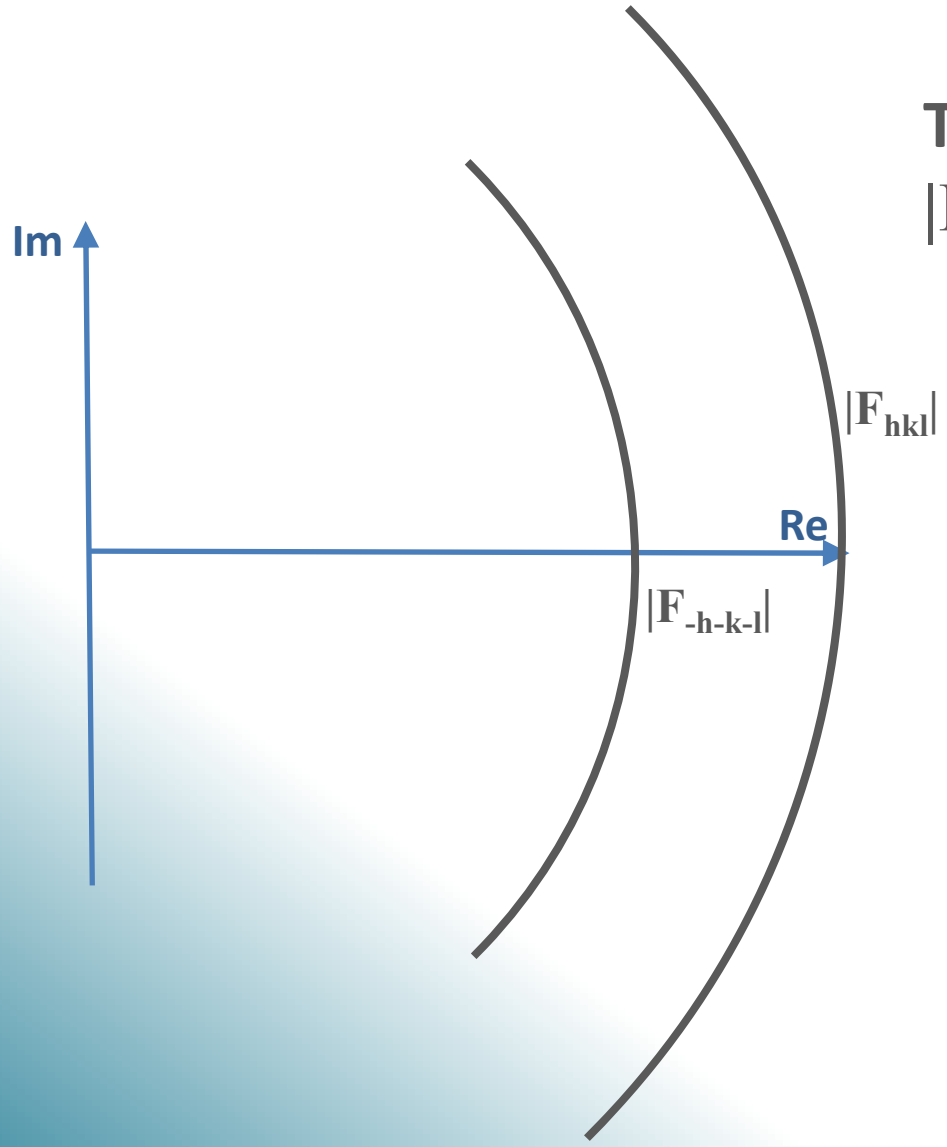
F_P Protein contribution

F_T (relates to F_{hkl}) = F_P + F_A

F_A = F_A + F_A‘ + F_A“

Anomalous scatterer contribution

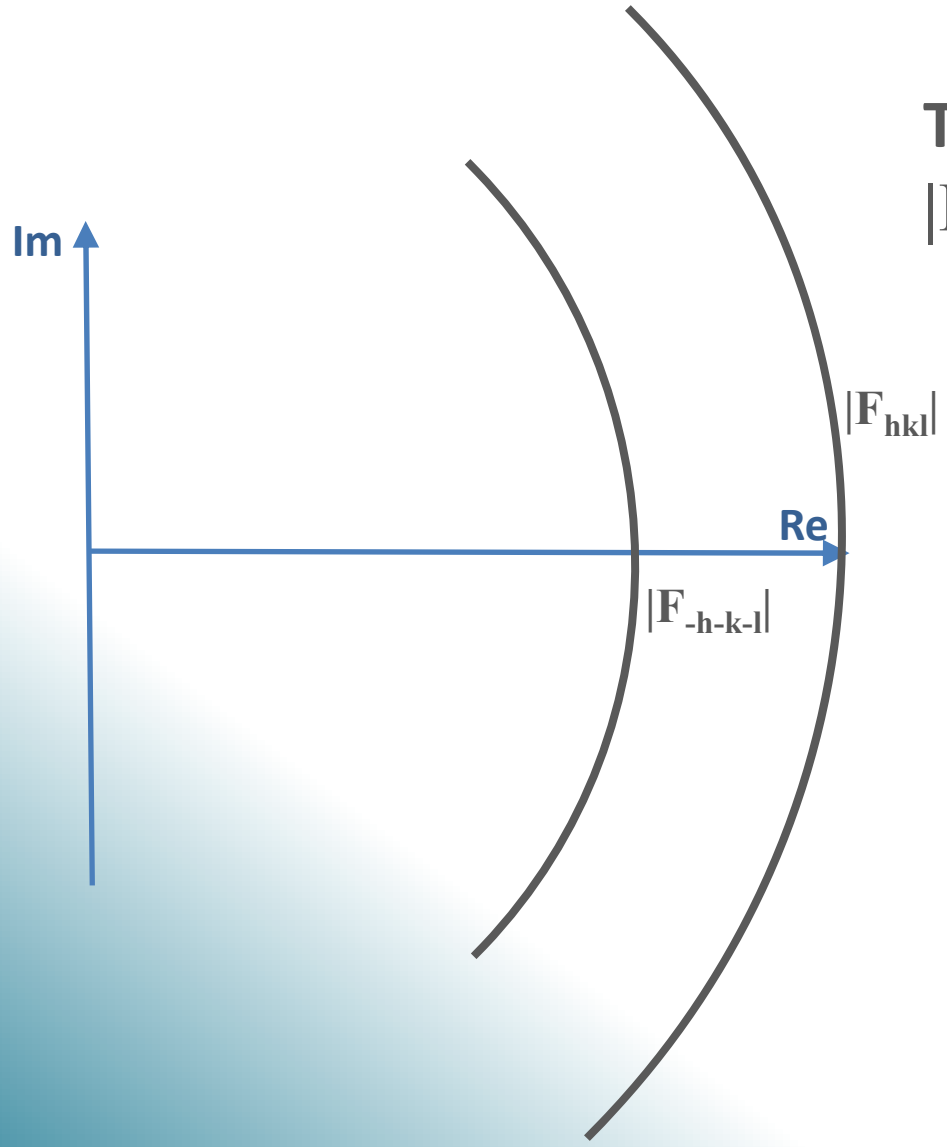
From substructure to structure



This is what we know:

$|F_{hkl}|$ and $|F_{-h-k-l}|$

From substructure to structure

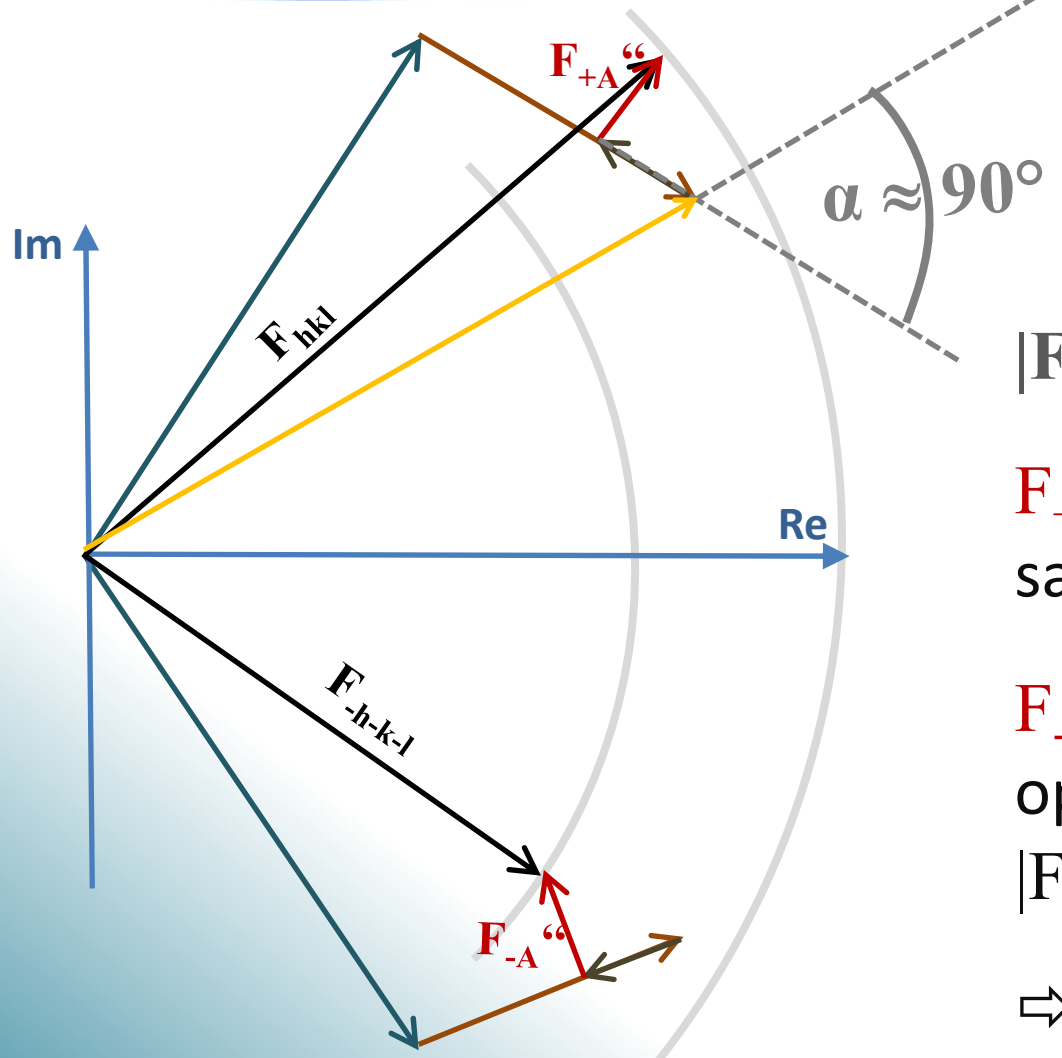


This is what we know:

$|F_{hkl}|$ **and** $|F_{-h-k-l}|$

$|F_{hkl}| >> |F_{-h-k-l}|$

From substructure to structure



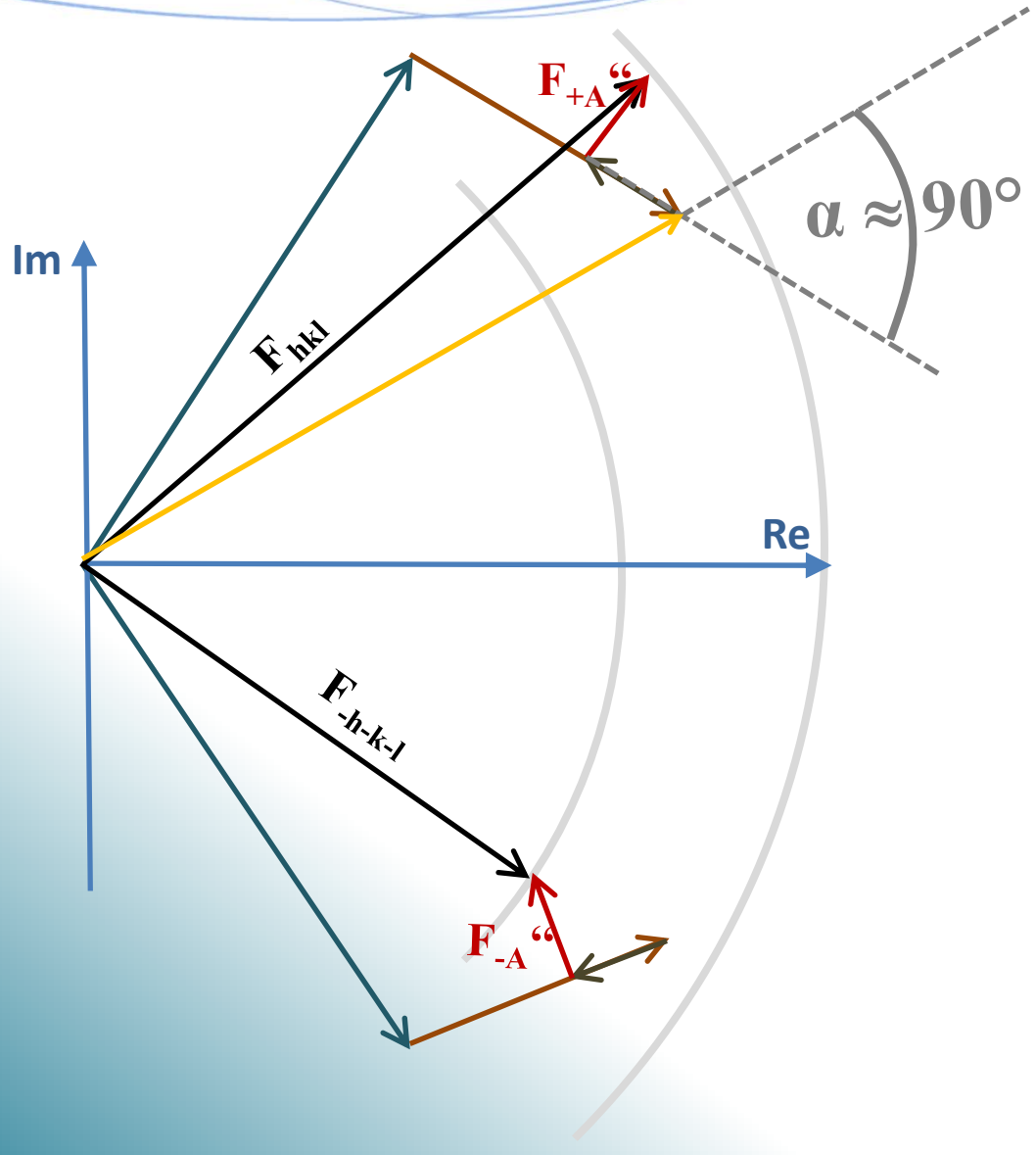
$|F_{hkl}| >> |F_{-h-k-l}|$

F_{+A}“ has to point in the same direction as $|F_{hkl}|$

F_{-A}“ has to point in the opposite direction as $|F_{-h-k-l}|$

⇨**α must be close to 90°!**

From substructure to structure



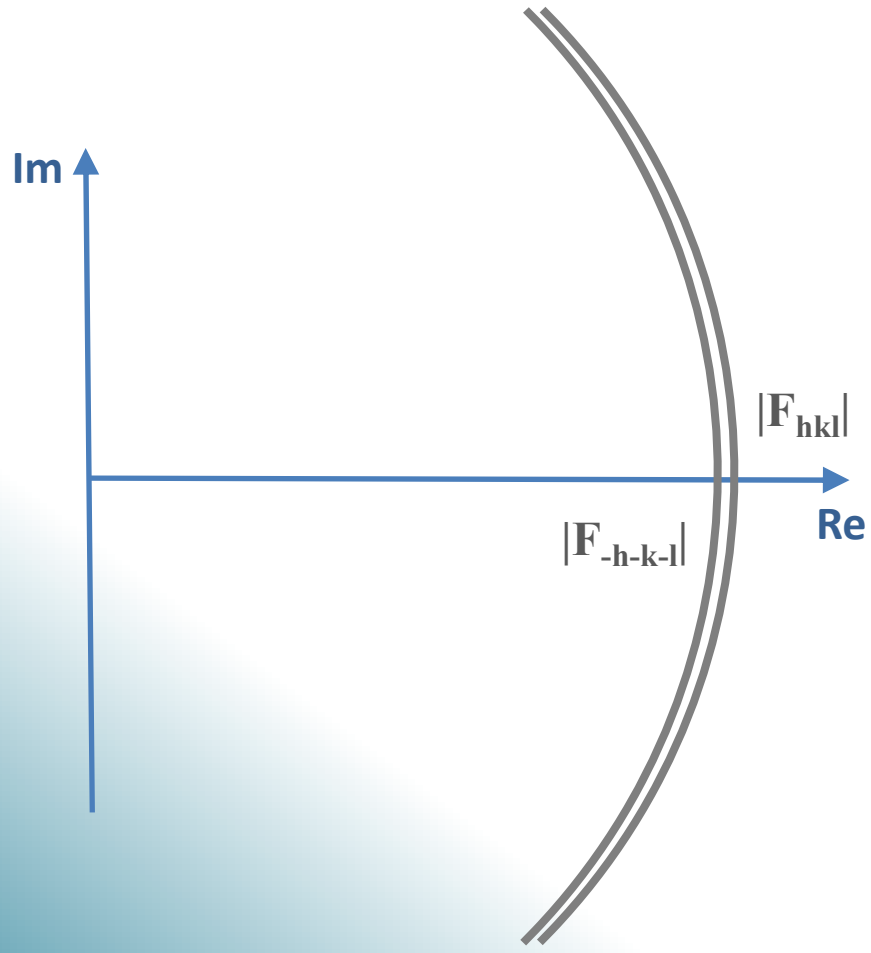
If: $|F_{hkl}| << |F_{-h-k-l}|$

⇨**α must be close to 270°!**

Reflections with the largest anomalous differences must be closest to $\alpha = 90°$ or $\alpha=270°$.

As you can easily see, estimation is rough.

From substructure to structure



$|F_{hkl}| \approx |F_{-h-k-l}|$

F_{+A}“ and F_{-A}“ must be very small or almost perpendicular to F_{hkl} or F_{-h-k-l}, respectively.

⇨**α must be close to 0° or 180°**

Density modification

- φ_T can now be computed from the phasing equations!

$$\varphi_A + \alpha = \varphi_T$$

Via Fourier synthesis, an initial map is gained.

- By σ_A coefficients and Sim weights the map is improved.
- But most important: **Density modification** is applied.

How to...

DENSITY MODIFICATION IN SHELXE

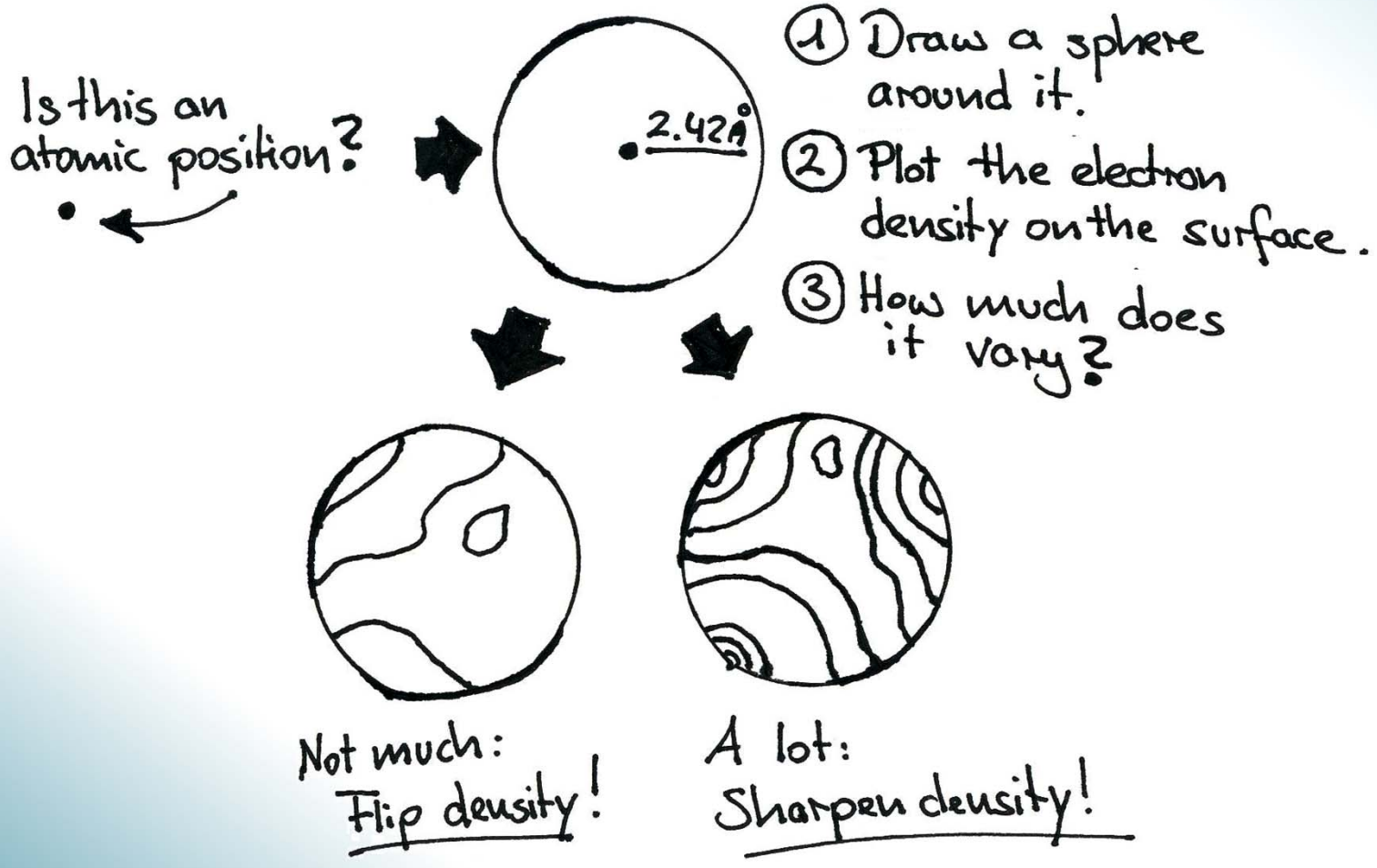
Density modification

Especially SAD phases are still ambiguous as well as inaccurate. **Density modification** dramatically improves initial phases, electron density and **resolves handedness**!

- Based on areas filled by disordered solvent
- Solvent area is flattened or flipped
- NCS averaging can improve map quality
- High solvent content gives often better improvement

Density modification

Most programs use a mask. SHELXE uses the sphere-of-influence method for density modification:



Density modification

After several cycles, one of the two maps (one for each substructure enantiomer) looks ‚like protein‘.

The other has less connectivity and looks ‚ragged‘.

After density modification, the structure is solved! Experimental phasing has led to initial phases.

DENSITY MODIFICATION

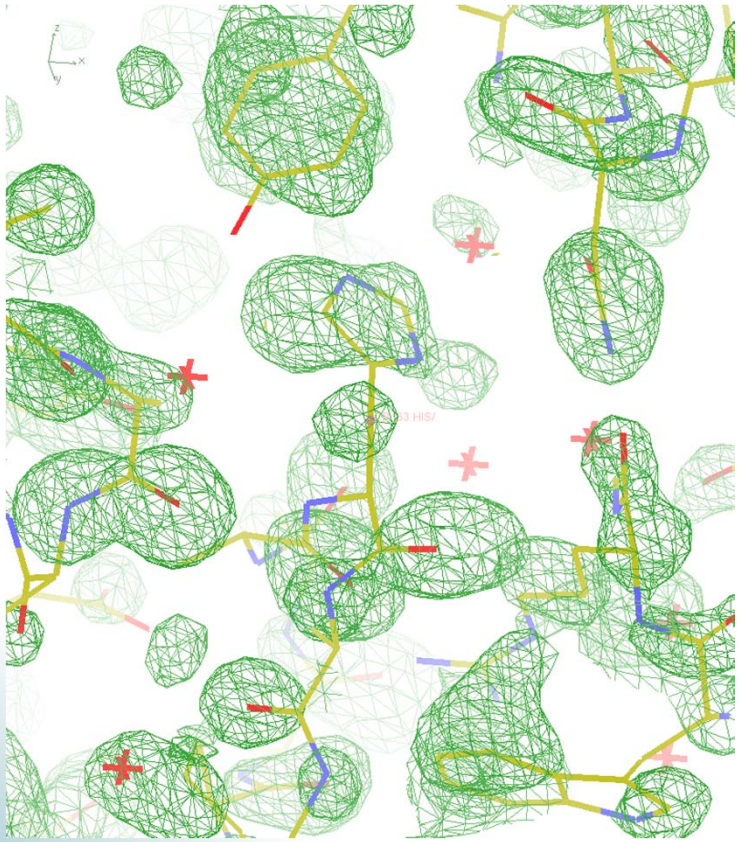
Free lunch algorithm

Here, data is expanded with rough guesses for $|F_{hkl}|$ to higher resolution (Typically from 2.0 Å downward).

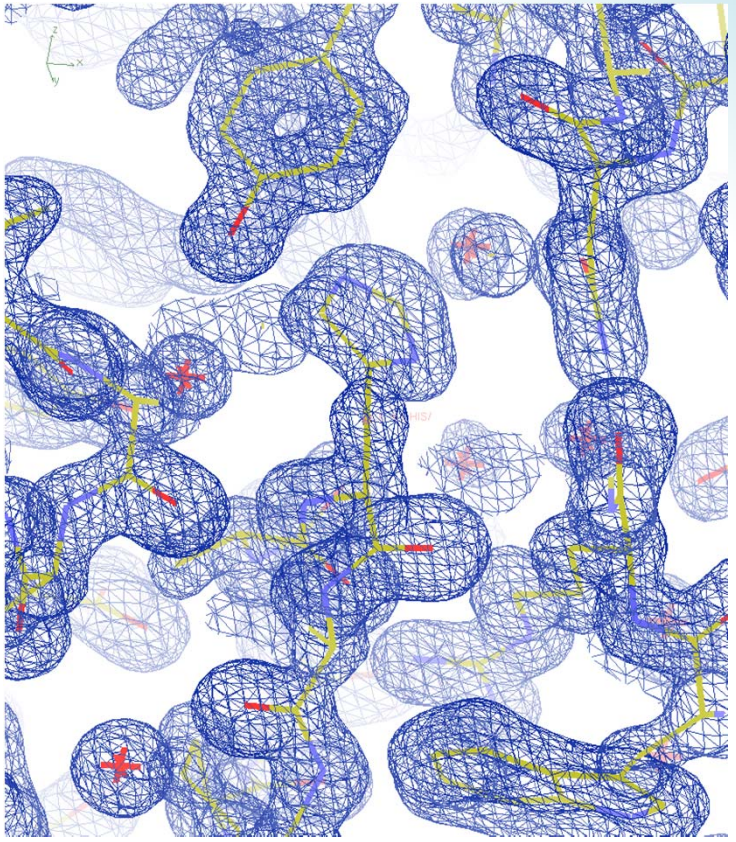
This improves the density, because of:

- Correction of Fourier truncation errors
- Phases are the major influence
- **0** is a bad estimation for a not measured intensity.

DENSITY MODIFICATION



After density modification (MapCC 0.57), incomplete reflections to 1.3Å



After expansion to 1.0 (MapCC 0.94)

Usón *et al*, *Acta Cryst.* D63 (2007) 1069.

Has SHELXE worked?

- Connectivity (continuous stretches of electron density representing the connectivity of the macromolecule in question) high
- Map contrast (clear delineation between ordered and disordered regions of the crystal) high.
- **The correct hand of the substructure** is the one with the higher map contrast.

The best indication of successful density modification is a map that looks like protein and in which the macromolecule can be built easily.

Experimental phasing, for real

PRACTICALITIES

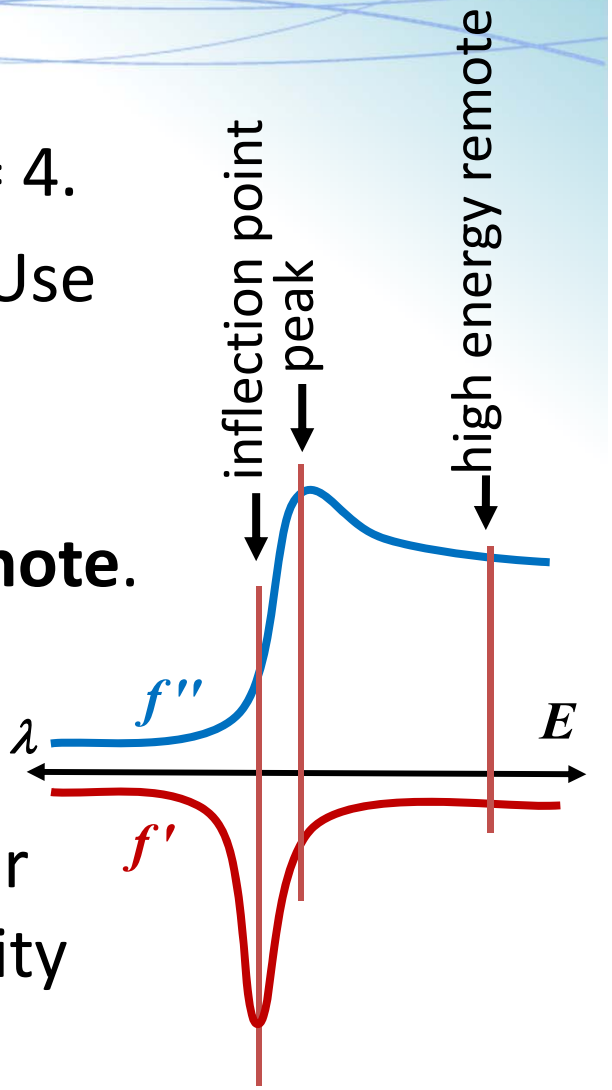
Data collection

- High multiplicity is good.
- Radiation damage is often bad.
- Precise intensity measurements are good.
- Near to the absorption edge, the crystal absorbs most energy, therefore radiation damage is high.
- A **fluorescence scan** can prove the presence of anomalous scatterers in the crystal.
- Good low resolution completeness

Pictures courtesy of Airlie McCoy

Data collection: MAD

- Collect **peak** with at least multiplicity = 4.
- Radiation damage? Stop and try SAD! Use a second crystal to collect high energy remote.
- No damage? Measure **high energy remote**.
- Last data set should be **inflection –** so *f*‘ is maximized.
- A **higher resolution data set** with lower redundancy may prove useful for density modification and for refinement.



$$f = f_0 + f' + i f''$$

Data evaluation

- The general data quality should be good – multiplicity, completeness, R_{PIM} etc.
- If scaling was applied, check statistics.
- Check the mask; inner shell completeness?
- Data set files well distinguishable?
- If you have made a fluorescence scan, keep it.
- Is there an anomalous signal in the collected data?
  - Anomalous correlation within a data set: $CC_{anom(1/2)}$
  - <d”/σ> and/or <d’/σ>
  - Anomalous correlation of data sets: CC_{anom}

Things you want to have an idea about

- Space group? (Twinning?)
- How many marker atoms do you expect?
- Substructure: Which elements/molecules?
- What could be the best resolution cut-off?

  (SHELXC assumes data resolution + 0.5Å)
- Could any marker atoms ‚fuse' into bigger blobs of density because of resolution cut-off? Disulfides?
- Merging of data from different crystals/runs?
- Expected solvent content and residue numbers?

If you use SHELX...

SHELXC: α calculation, data analysis, file preparation

SHELXD: Substructure search

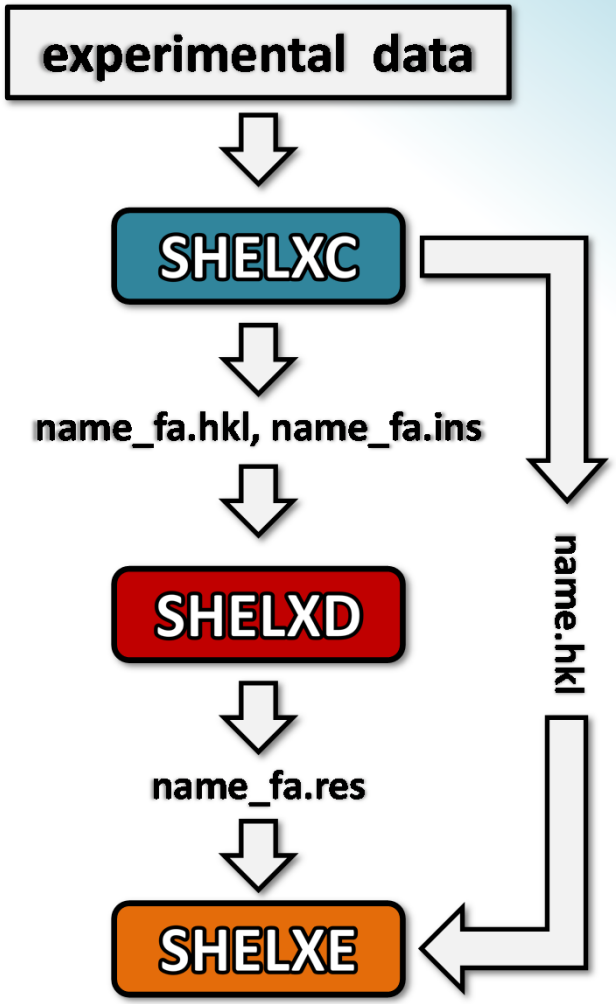
SHELXE: Density modification, tracing*

**** A traced structure is solved; CC (trace against native data) > 25% (for data < 2.5 Å)***

[**ANODE**: Validation]

Pipeline?

Other experimental phasing programs should be considered , in particular for ease of use or problem cases**.



** http://strucbio.biologie.uni-konstanz.de/ccp4wiki/index.php/Experimental_phasing

SOFTWARE

The new SHELXE version (2013) also can do autotracing of the backbone.

A structure that can be traced is a structure solved*.

This proves particularly useful

- in borderline cases
- in pipelines, like ARCIMBOLDO or AUTORICKSHAW

*** Solved: CC (trace against native data) > 25% for data < 2.5 Å!**

SOFTWARE

MR-SAD

- Not enough phase information from SAD alone

  or only partial Molecular Replacement solution

  or severe model bias

- Use Molecular Replacement to bootstrap SAD phases
- (Partial) Molecular Replacement solution can be put into SHELXE, even without experimental differences

Schuermann & Tanner (2003). Acta Cryst. D59, 1731.

ANODE

ANOMALOUS MAPS

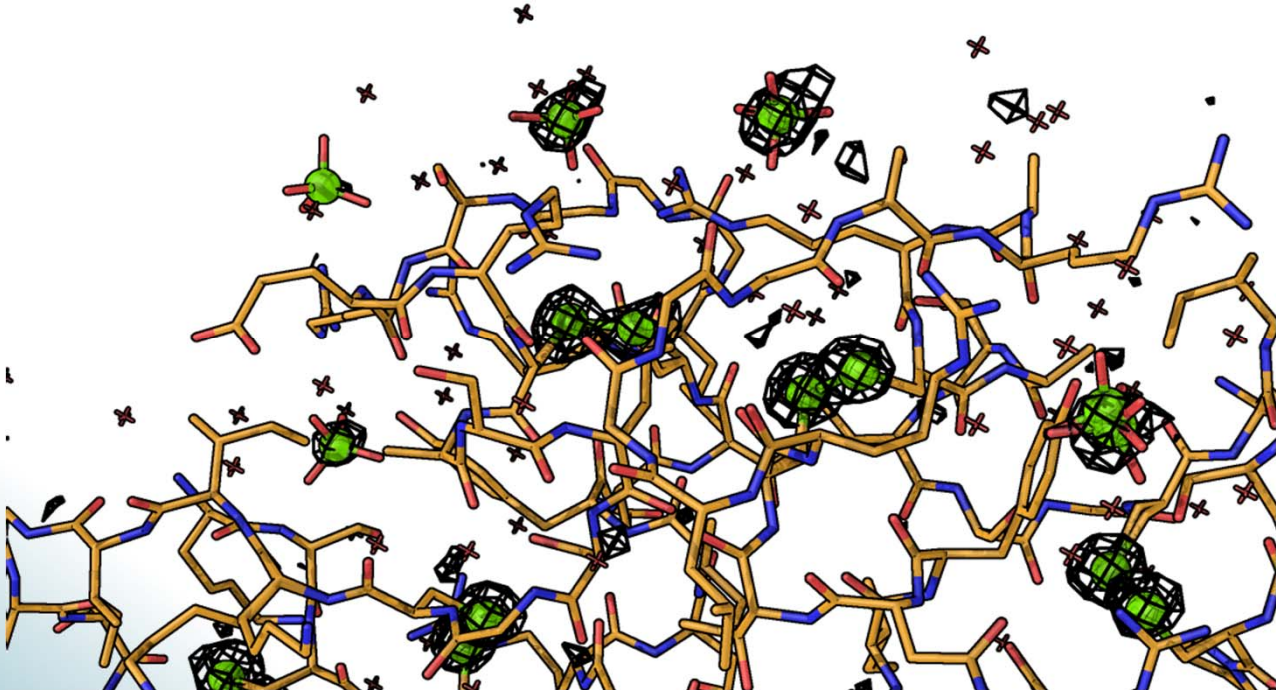
ANODE

Intensity differences & phases

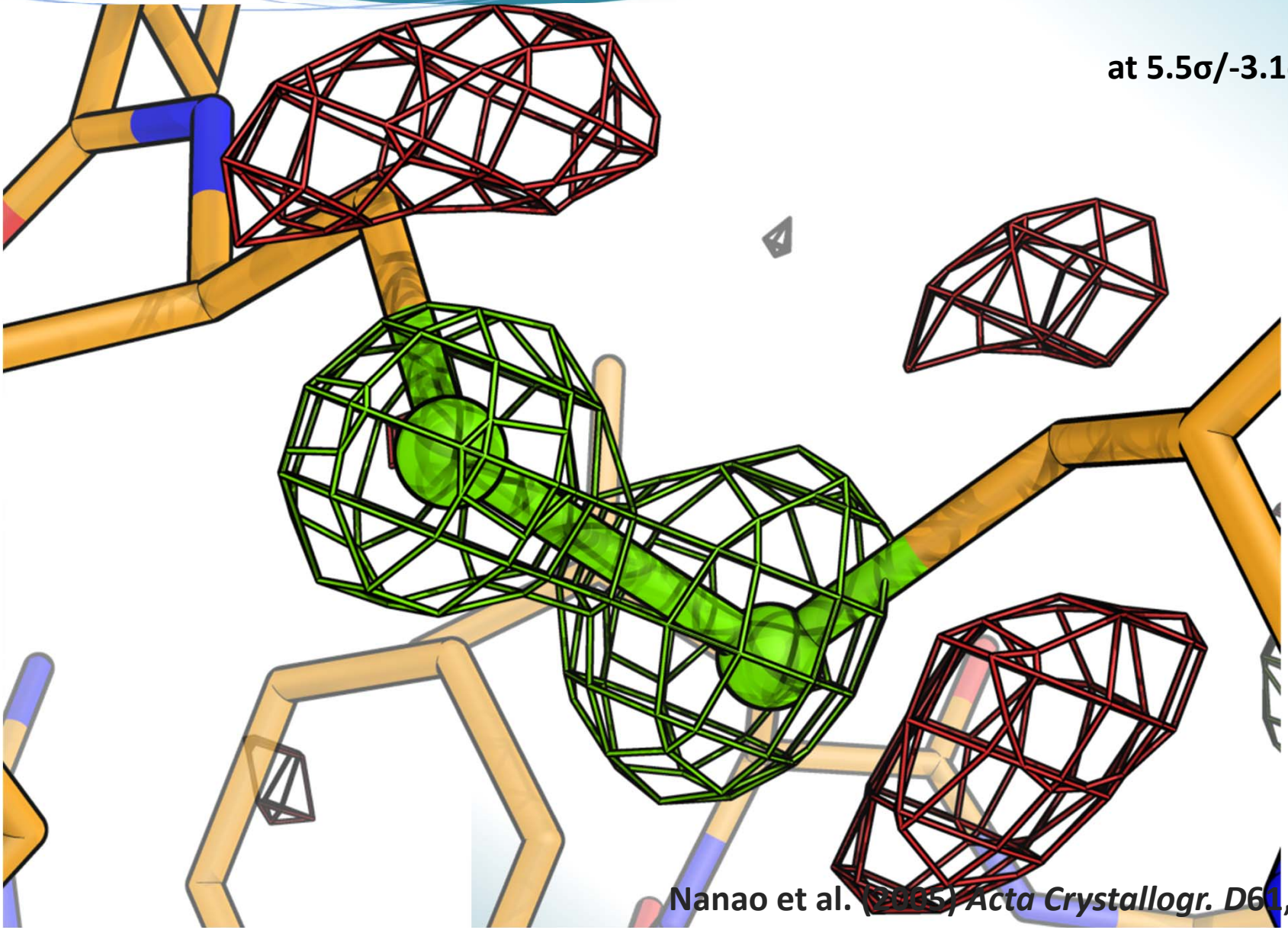
FOURIER SYNTHESIS

Anomalous map

- Structure validation
- Analysis of the anomalous signal



RIP density maps



at 5.5σ/-3.1σ

Nanao et al. (2005) *Acta Crystallogr. D*61, 1227

Final

SUMMARY

SUMMARY

- Experimental phasing methods use marker substructures of certain elements to solve the phase problem via the phasing equations. Patterson maps can help.
- **MAD** and **SAD** exploit the anomalous signal from one or more data sets from the same crystal.
- **SIR** and **MIR** utilizes several heavy-atom soaked derivative crystals. They have to be isomorphous to be utilized.
- Experimental phase solutions **do not define the enantiomorph**; after solution, the map that looks like protein has to be chosen!

LITERATURE

- Bernhard Rupp, **Biomolecular Crystallography**: Principles, Practice, and Application to Structural Biology, 2004

- Kai Diederichs, P. Andrew Karplus, **Improved R-factors for diffraction data analysis in macromolecular crystallography.** Nat Struct Biol. (1997). 4, 269-75.

- Manfred S. Weiss, **Global indicators of X-ray data quality**, J. Appl. Cryst. (2001). 34, 130-135

LITERATURE

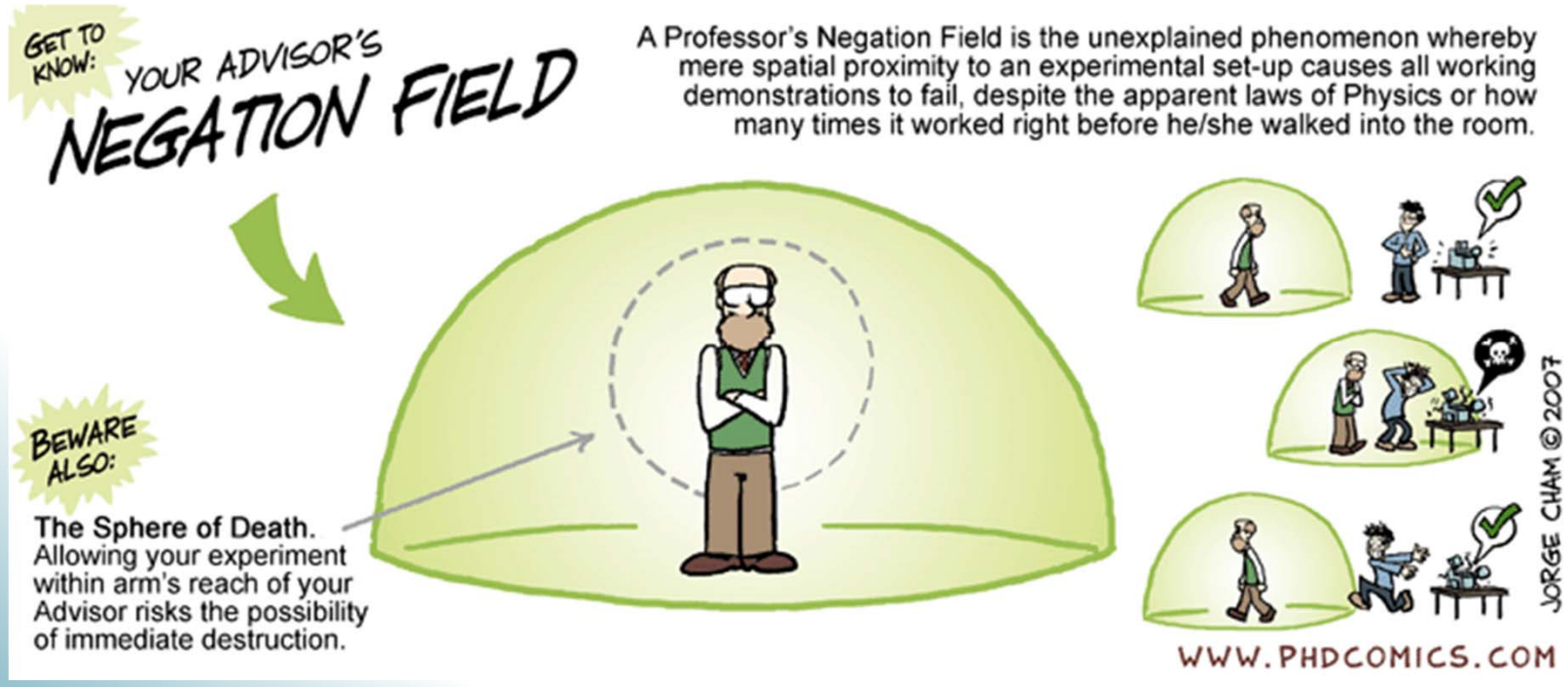
- George M. Sheldrick, **A short history of SHELX**, Acta Cryst. (2008). A64, 112-122
- George M. Sheldrick (2002). **Macromolecular phasing with SHELXE**, *Z. Kristallogr*. 217:644-650.
- George M. Sheldrick, **Experimental phasing with SHELXC/D/E: combining chain tracing with density modification**, Acta Cryst. (2010). D66, 479-485
- A. Thorn & G.M. Sheldrick: "**ANODE: ANOmalous and heavy-atom DEnsity calculation**" *J. Appl. Cryst.* 44 (2011), 1285-1287

More material: shelx.uni-ac.gwdg.de/~athorn/

http://shelx.uni-ac.gwdg.de/SHELX/

ACKNOWLEDGEMENTS

I am grateful to **George Sheldrick** who freely shared with me hisknowledge, ideas and material. I also want to thank the Murshudov lab for their support!



MRC Laboratory of Molecular Biology

SEVENTH FRAMEWORK PROGRAMME

MARIE CURIE ACTIONS