CCP4 Fukuoka 2012

SHELXC/D/E

Andrea Thorn

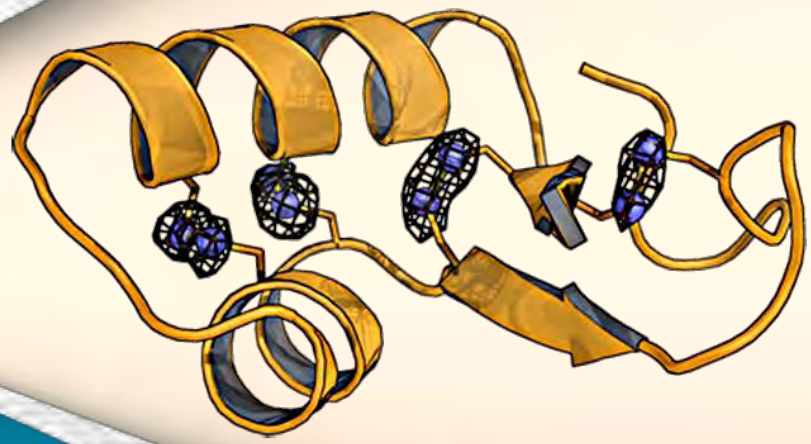


Image property of Madhumati Sevvana

OVERVIEW

Introduction

- Structure factors
- Experimental phasing

Anomalous diffraction

- Substructure search
- Getting α and phasing equations
- Density modification

Practical matters

- Software
- Data Collection

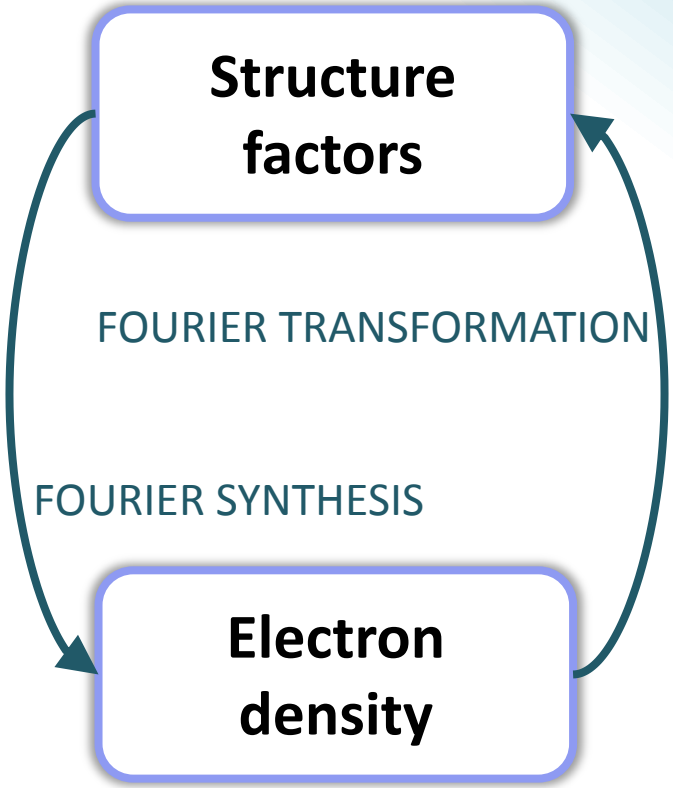
Summary and literature

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.

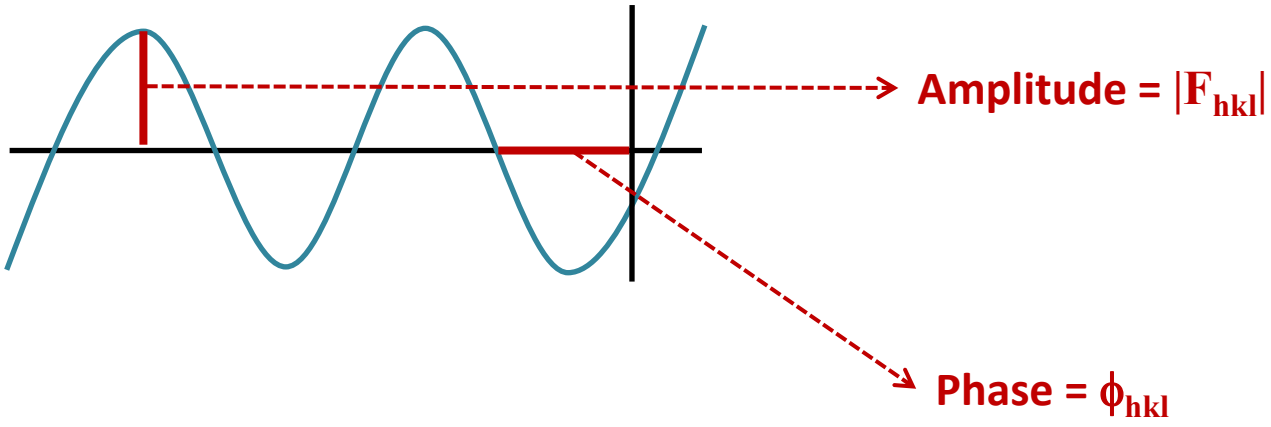


STRUCTURE FACTORS

For each reflection, there is a

structure factor F_{hkl}

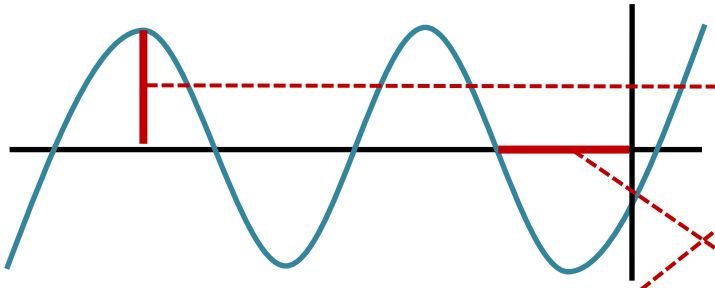
= a wave



STRUCTURE FACTORS

structure factor F_{hkl}

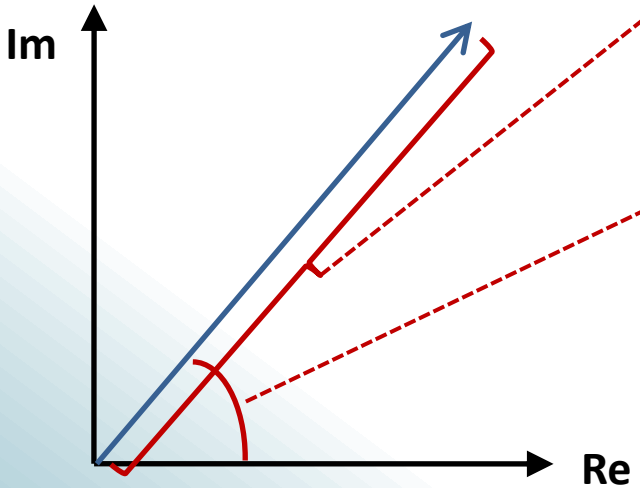
= a wave



Amplitude = $|F_{hkl}|$

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

= a complex number



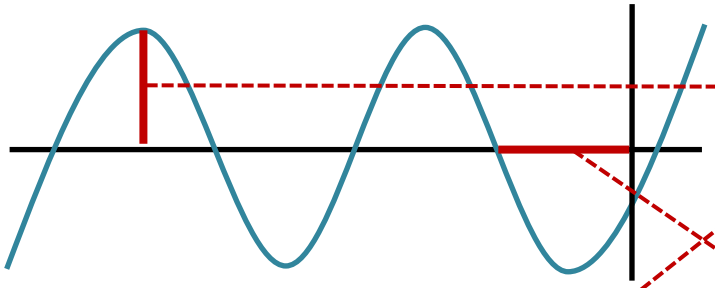
Phase = ϕ_{hkl}

cannot be measured... :-(

STRUCTURE FACTORS

structure factor F_{hkl}

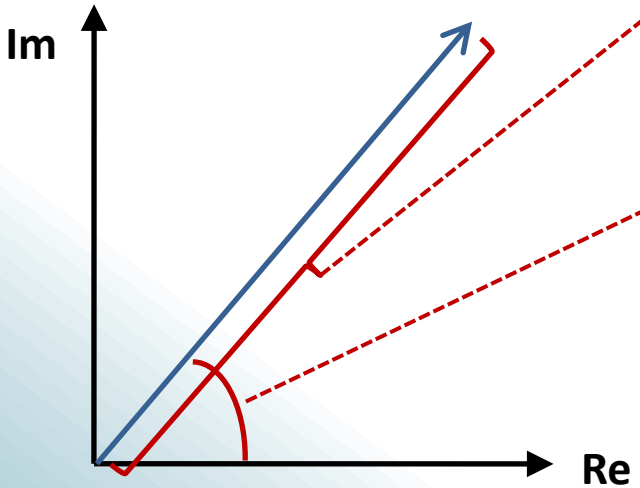
= a wave



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}

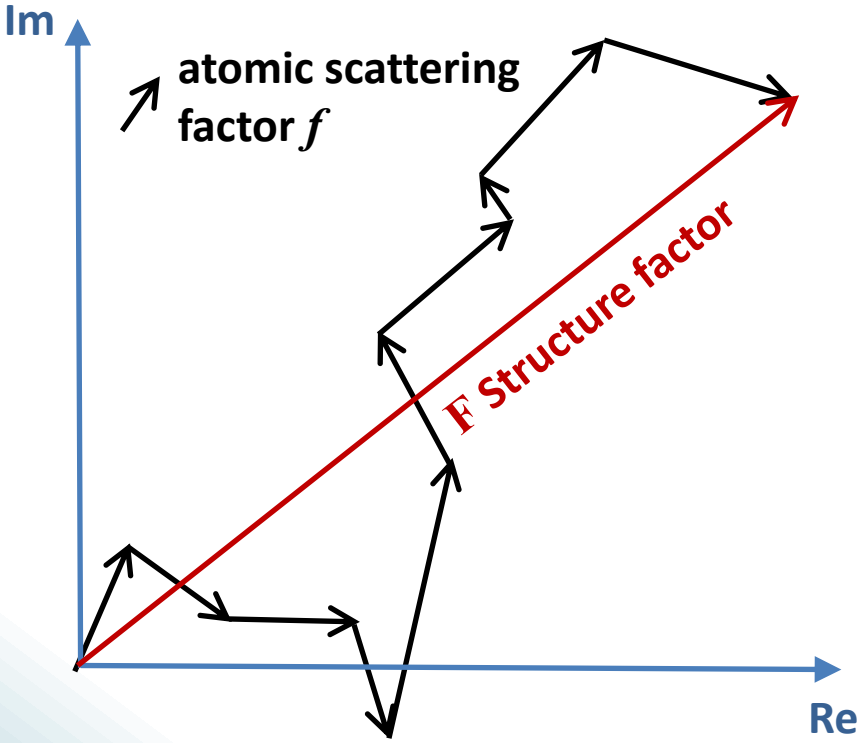
cannot be measured... :-(

PHASE PROBLEM

The central problem of crystallography

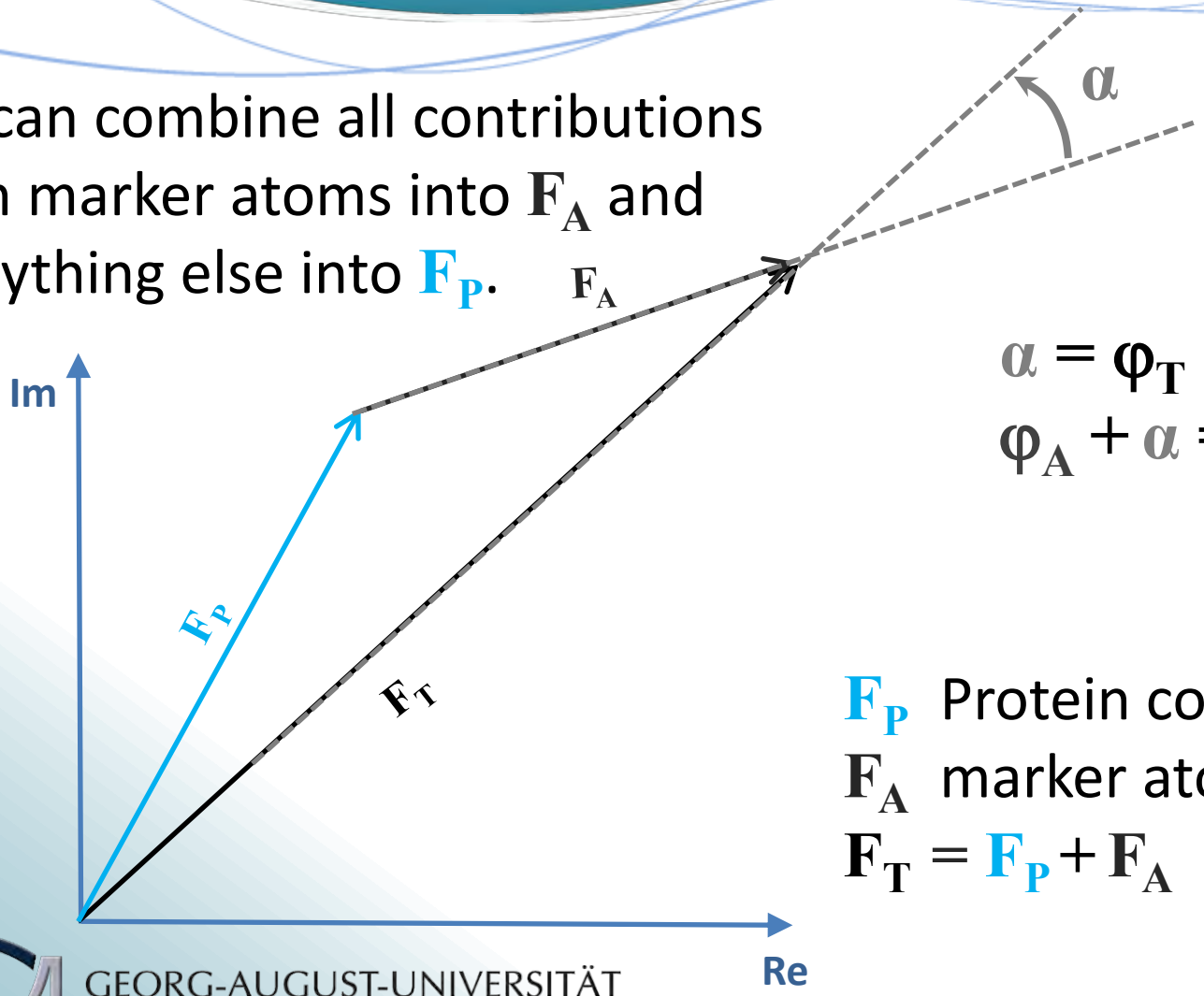
STRUCTURE FACTORS

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



EXPERIMENTAL PHASING

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
$\mathbf{F_T} = \mathbf{F_P} + \mathbf{F_A}$

EXPERIMENTAL PHASING

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 α, we could calculate φ_T and **solve the phase problem**!

EXPERIMENTAL PHASING

With this, both anomalous scatterers as well as heavy elements can be used for

experimental phasing

Whether heavy atoms or anomalous scatterers are used depends on the method.

They will be called **marker atoms** in the following!

EXPERIMENTAL PHASING

- **Single isomorphous replacement (SIR)**
- **Single wavelength anomalous diffraction (SAD)**
- **Native sulfur-based SAD (S-SAD)**

  all require density modification or direct methods support to break the inherent two-phase ambiguity. For SIR, an ismorphous derivative crystal is used.
- **Single isomorphous replacement with anomalous scattering (SIRAS):** Typically, Iodine soaks are used on the crystal.

*** Beck et al., Acta Cryst. (2008). D64, 1179.**

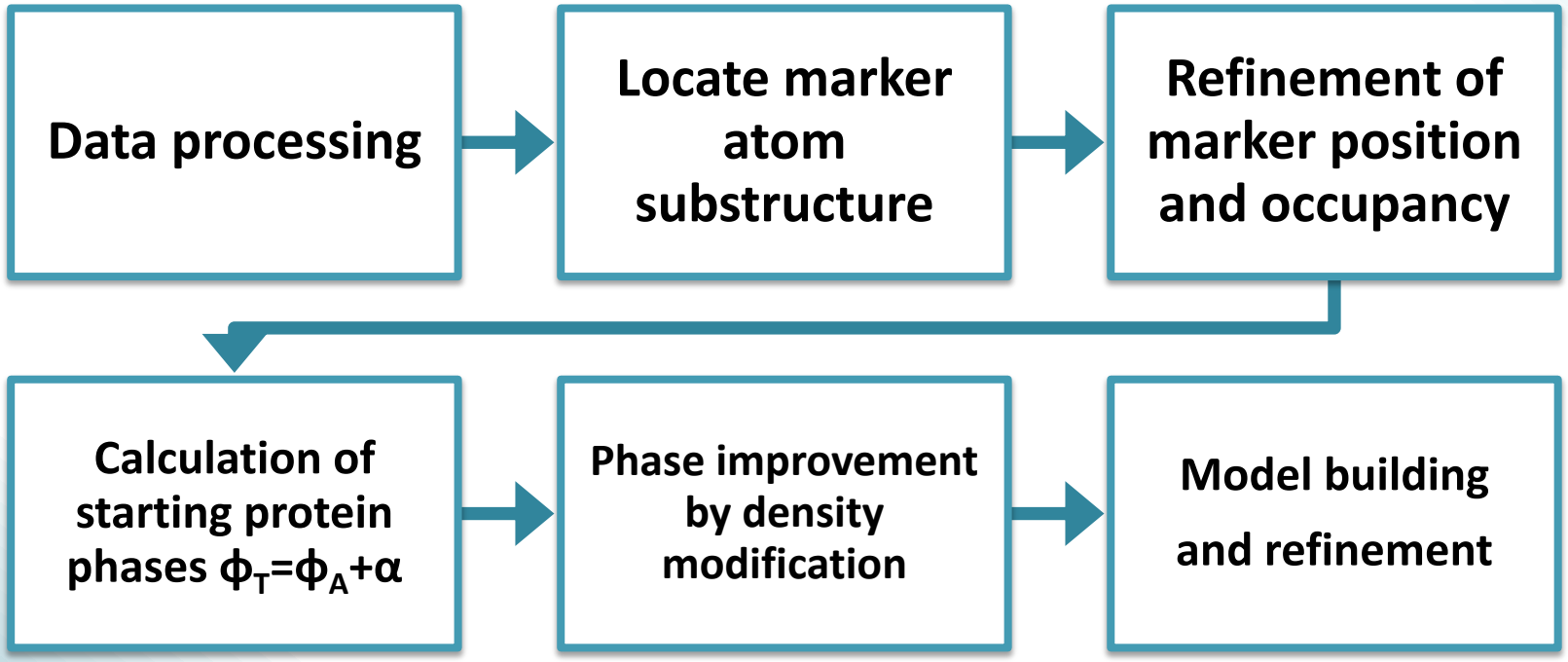
EXPERIMENTAL PHASING

- The two-phase ambiguity can be broken by providing phases from at least a second heavy atom derivative crystal in **multiple isomorphous replacement (MIR)**.
- **Radiation-induced phasing (RIP)** is isomorphous phasing where distinct atoms are lost after high radiation exposure.
- **Multiple wavelength anomalous diffraction (MAD)** provides orthogonal dispersive and anomalous differences from the same crystal.

EXPERIMENTAL PHASING

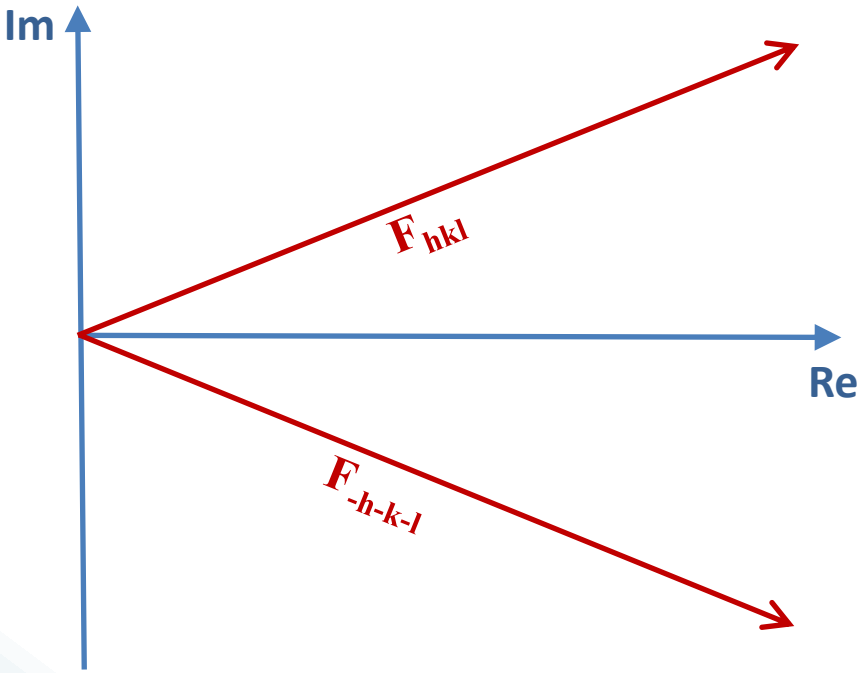
- Single isomorphous replacement (SIR)
- Single wavelength anomalous diffraction (SAD)
- Native sulfur-based SAD (S-SAD)
- Single isomorphous replacement with anomalous scattering (SIRAS)
- Radiation-induced phasing (RIP)
- Multiple isomorphous replacement with anomalous scattering (MIRAS)
- Single isomorphous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

EXPERIMENTAL PHASING



ANOMALOUS DIFFRACTION

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



ANOMALOUS DIFFRACTION

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

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

depends solely on the resolution

real component

imaginary component

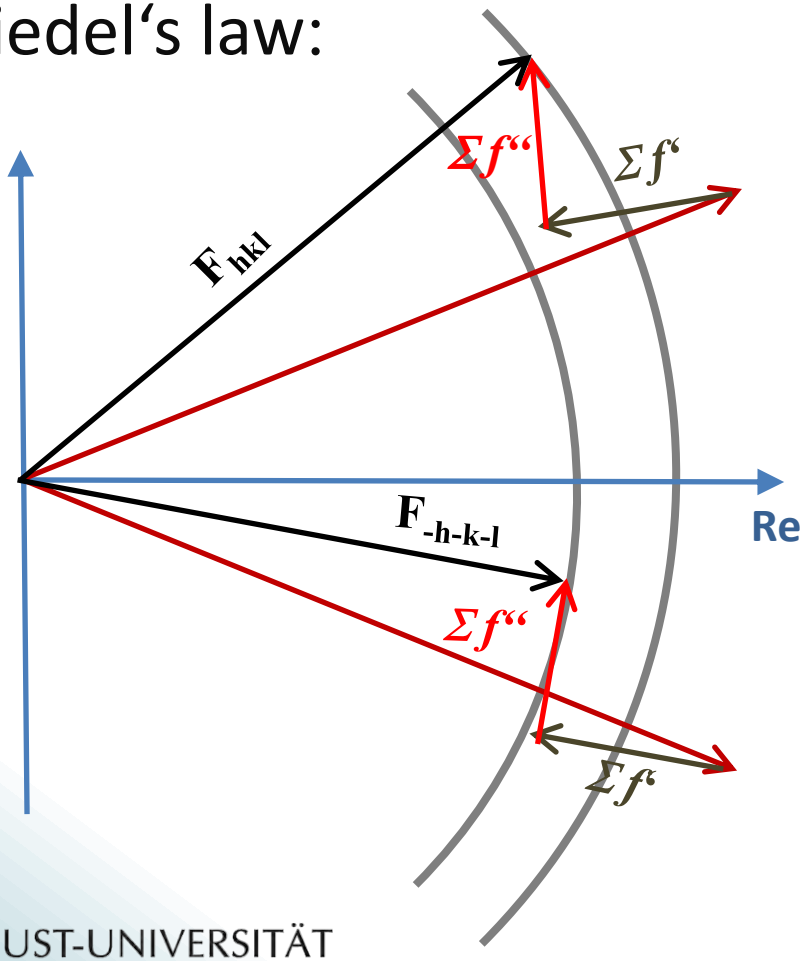
f‘ and *f“* *are observed near absorption edges of the atom‘s element, and are λ-dependent*

ANOMALOUS DIFFRACTION

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!

SUBSTRUCTURE SEARCH

Finding the „substructure“ of marker atoms

- Direct methods
- Patterson methods

Borrowed from small molecule crystallography

These methods require seperate atomic electron densities to locate atoms.

They work here because the marker atoms have large interatomic distances (disulfides become ‚supersulfurs‘) We assume all marker atoms being of the same element, only with different occupancy.

SUBSTRUCTURE SEARCH

Patterson methods

- The Patterson map contains all interatomic distance vectors between marker atoms. This can be used as a starting point (‚Patterson seeding‘)

Direct methods

- Phases of strong reflections are related (as a result of the non-random distribution of atoms.)
- Relations are relatively easy to resolve for a few atoms.
- **Dual space direct methods** recycle and modify trial structures by peak search in the electron density. Convergence is faster than in reciprocal space alone.

SUBSTRUCTURE SEARCH

Critical factors in substructure search

- Resolution range highly affects the outcome
- The smaller the marker atoms, the better the data quality has to be
- Intensity outliers are problematic
- Scaling (also anisotropic scaling) is needed

BEWARE: Handedness is not resolved at this stage!
(SHELXE allows later to differentiate)

PHASING EQUATIONS

Getting inital phases: Phasing equations

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

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

$$a = \frac{f''^2 + f'^2}{f_0^2} \qquad b = \frac{2f'}{f_0} \qquad c = \frac{2f''}{f_0} \qquad \alpha = \phi_T - \phi_A$$

F_T **Total structure factor**

F_A **Marker substructure structure factor**

PHASING EQUATIONS

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!

PHASING EQUATIONS

Phasing equations

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

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

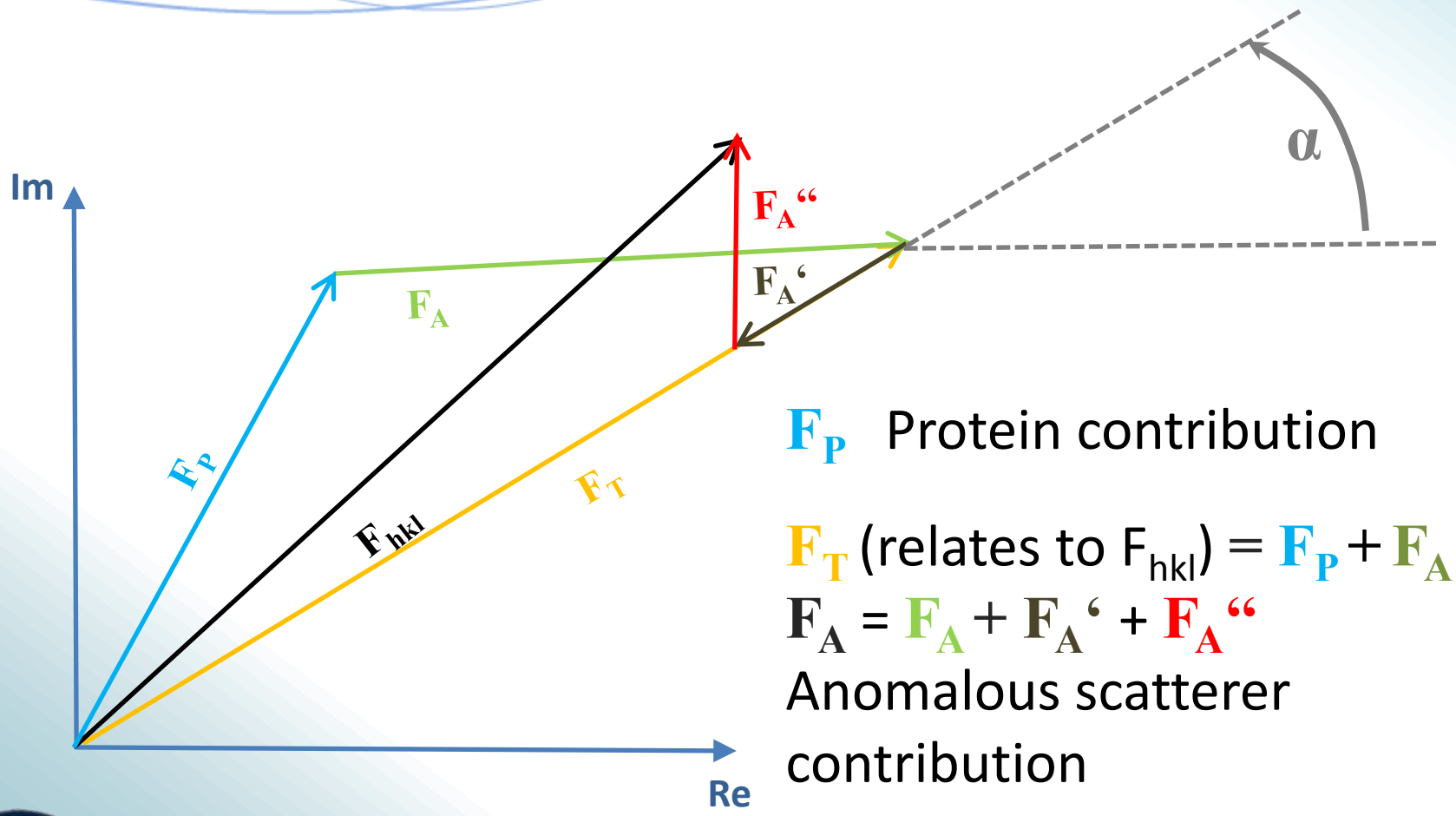
In a **SAD** experiment, we have only two observables, as we measured only one wavelength. So we assume

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

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

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

GETTING α



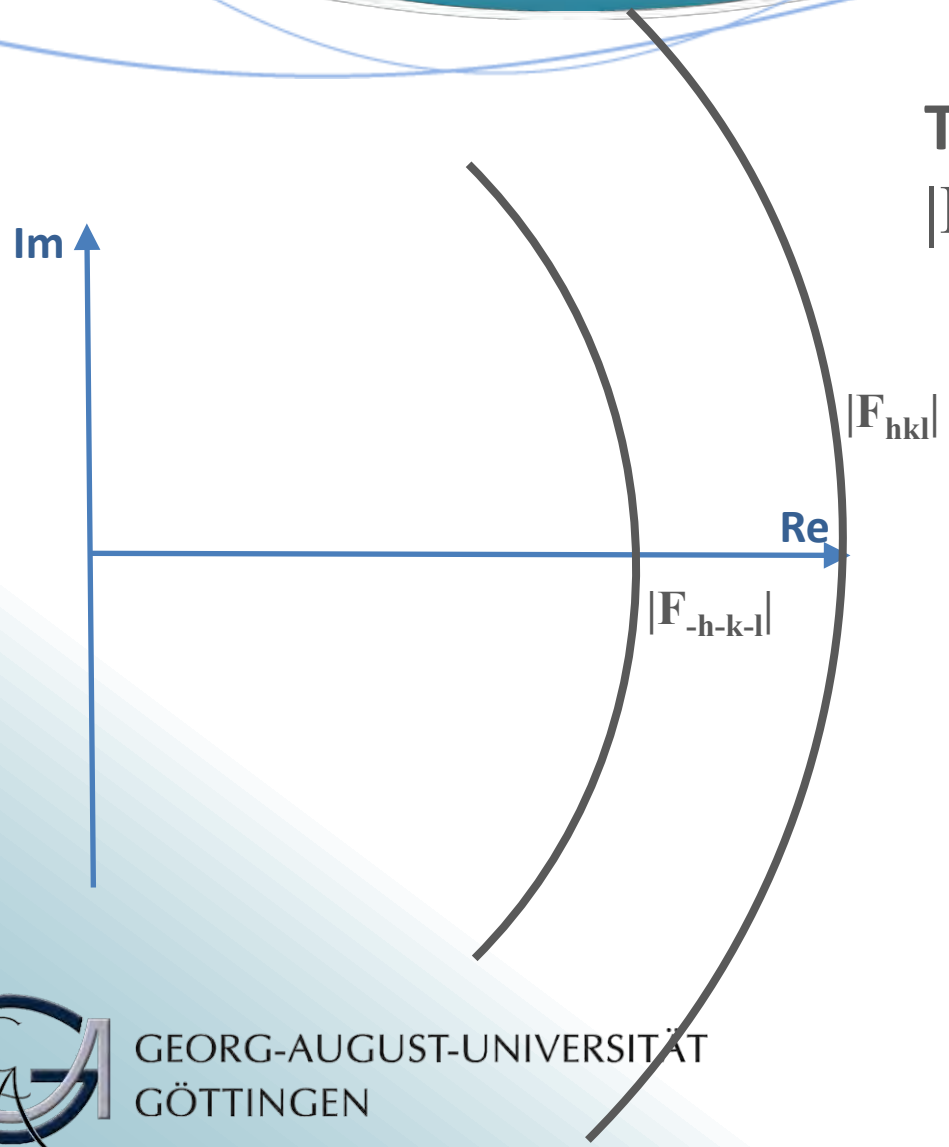
F_P Protein contribution

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

F_A = F_A + F_A‘ + F_A“

Anomalous scatterer contribution

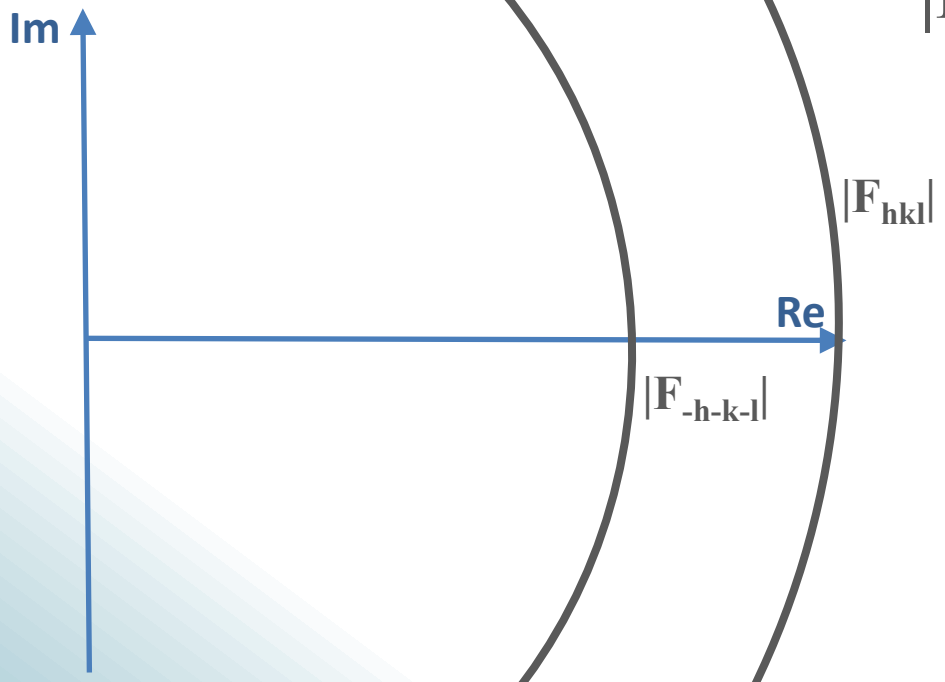
GETTING α



This is what we know:

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

GETTING α



This is what we know:
$|F_{hkl}|$ and $|F_{-h-k-l}|$

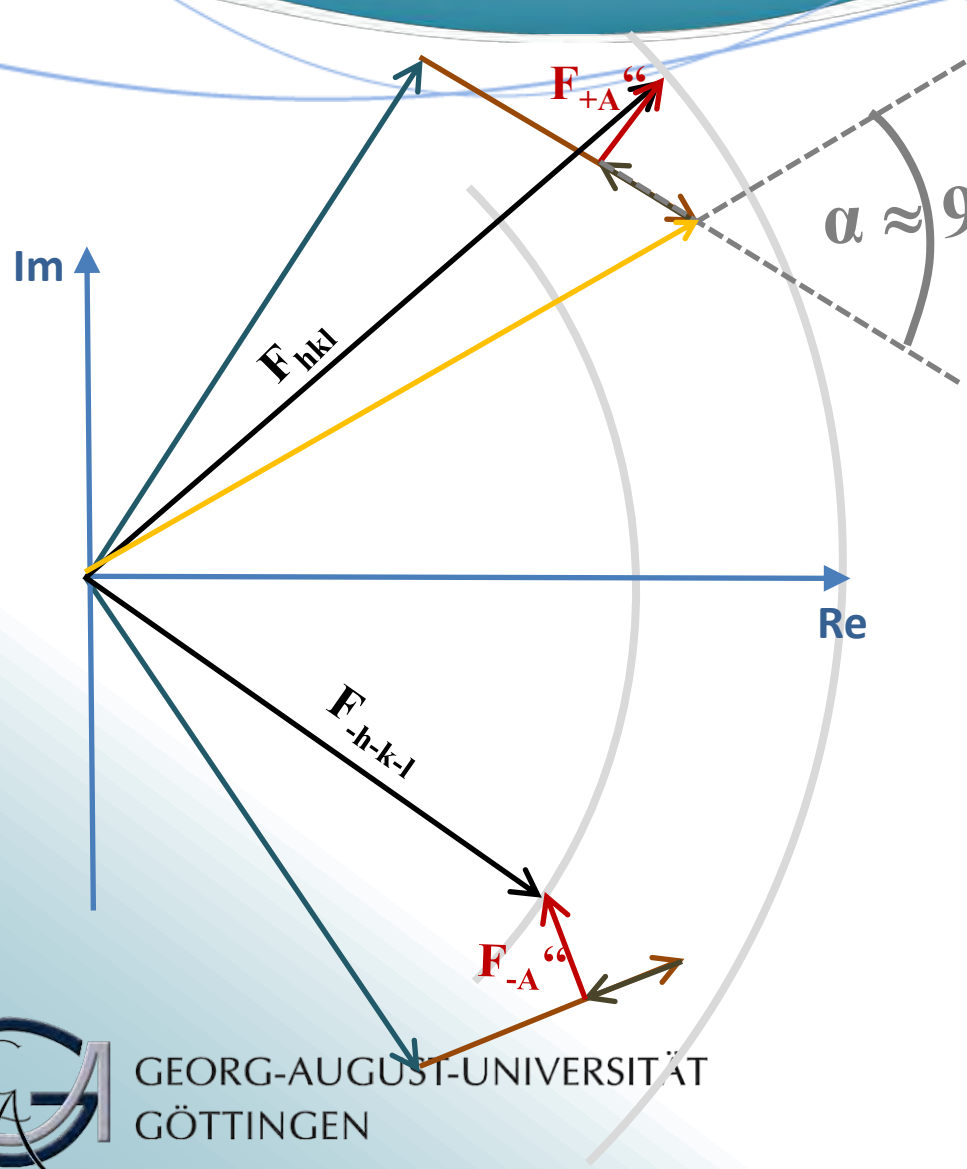
$|F_{hkl}| >> |F_{-h-k-l}|$

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°!**

GETTING α



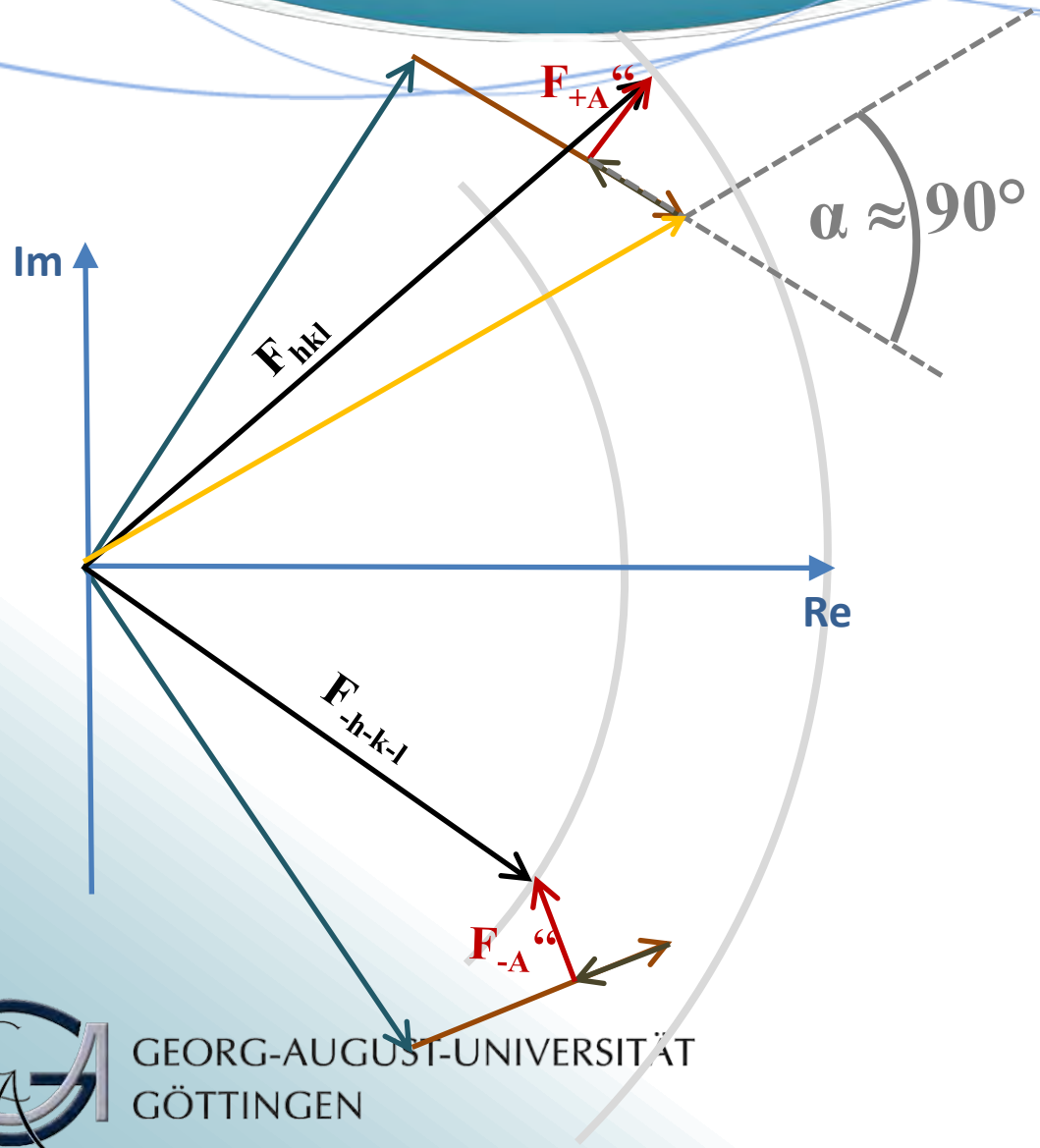
$|F_{hkl}| >> |F_{-h-k-l}|$

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°!**

GETTING α



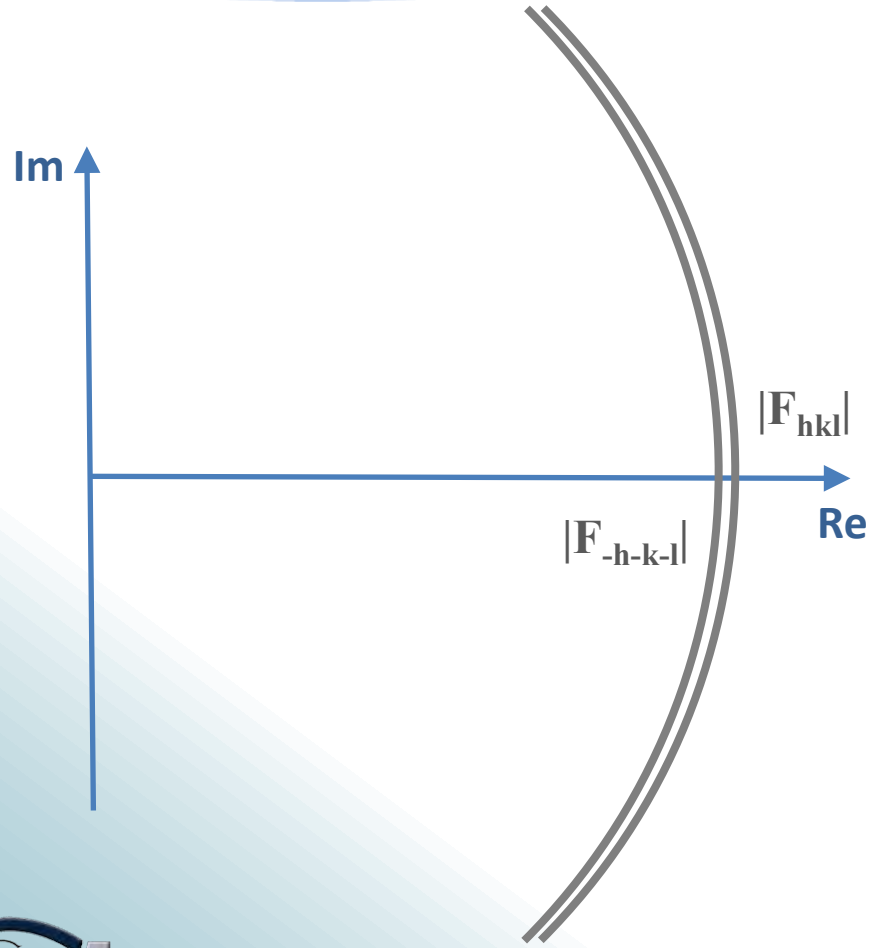
If: $|F_{hkl}| << |F_{-h-k-l}|$

⇨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.

GETTING α



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

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°**

PHASING EQUATIONS

- φ_T can now be computed from the phasing equations!

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

Via Fourier synthesis, an inital map is gained.

- By σ_A coefficients and Sim weights the map is improved.

- But most important: **Density modification** is applied.

DENSITY MODIFICATION

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 mask solvent areas and then use those for flattening. SHELXE uses the sphere of influence algorithm:

- The electron density variance on a spherical surface (r=2.42 Å) is calculated.
- This variance **V** should hint to the center being an actual atomic position.
- For pixels with low **V**, density is flipped.
- For pixels with high **V**, density is sharpened.

DENSITY MODIFICATION

Free lunch algorithm

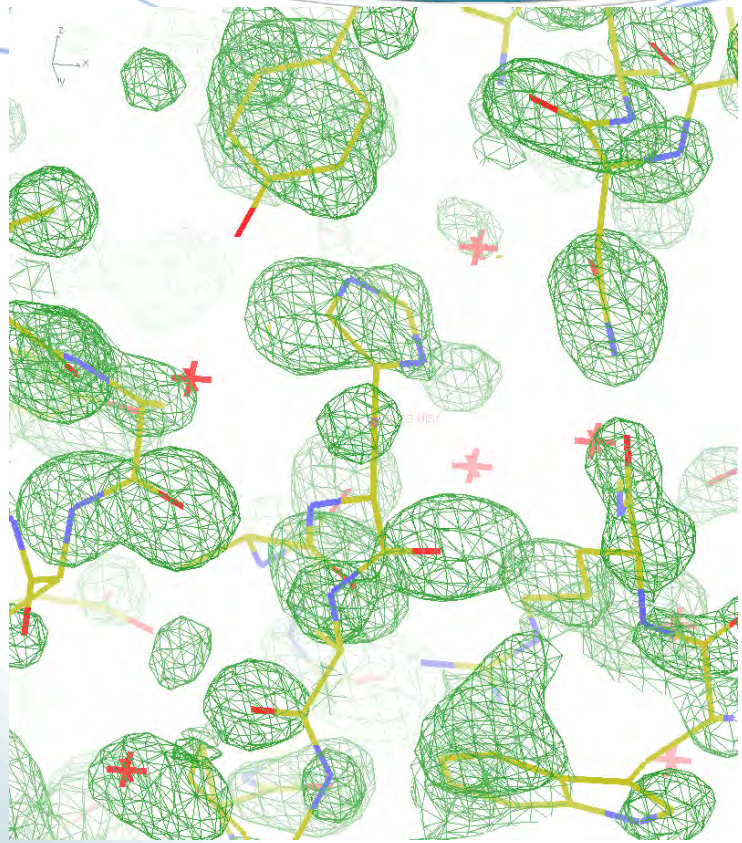
Here, data is expanded with rough guesses for $|F_{hkl}|$ to higher resolution (Typically from 2.0 Å downward).

This, mysteriously, improves the density.

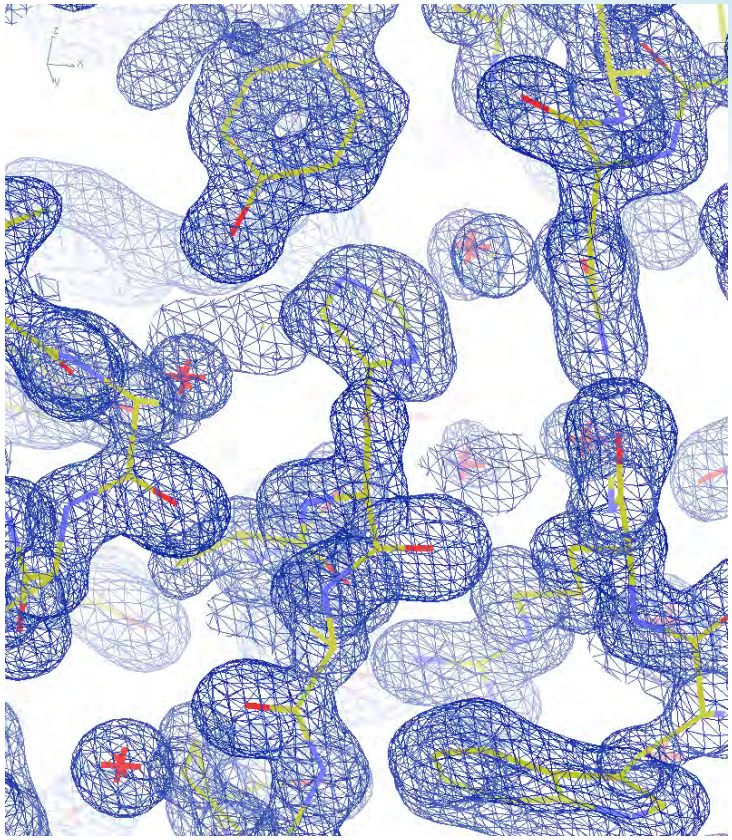
Possibly because of:

- 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

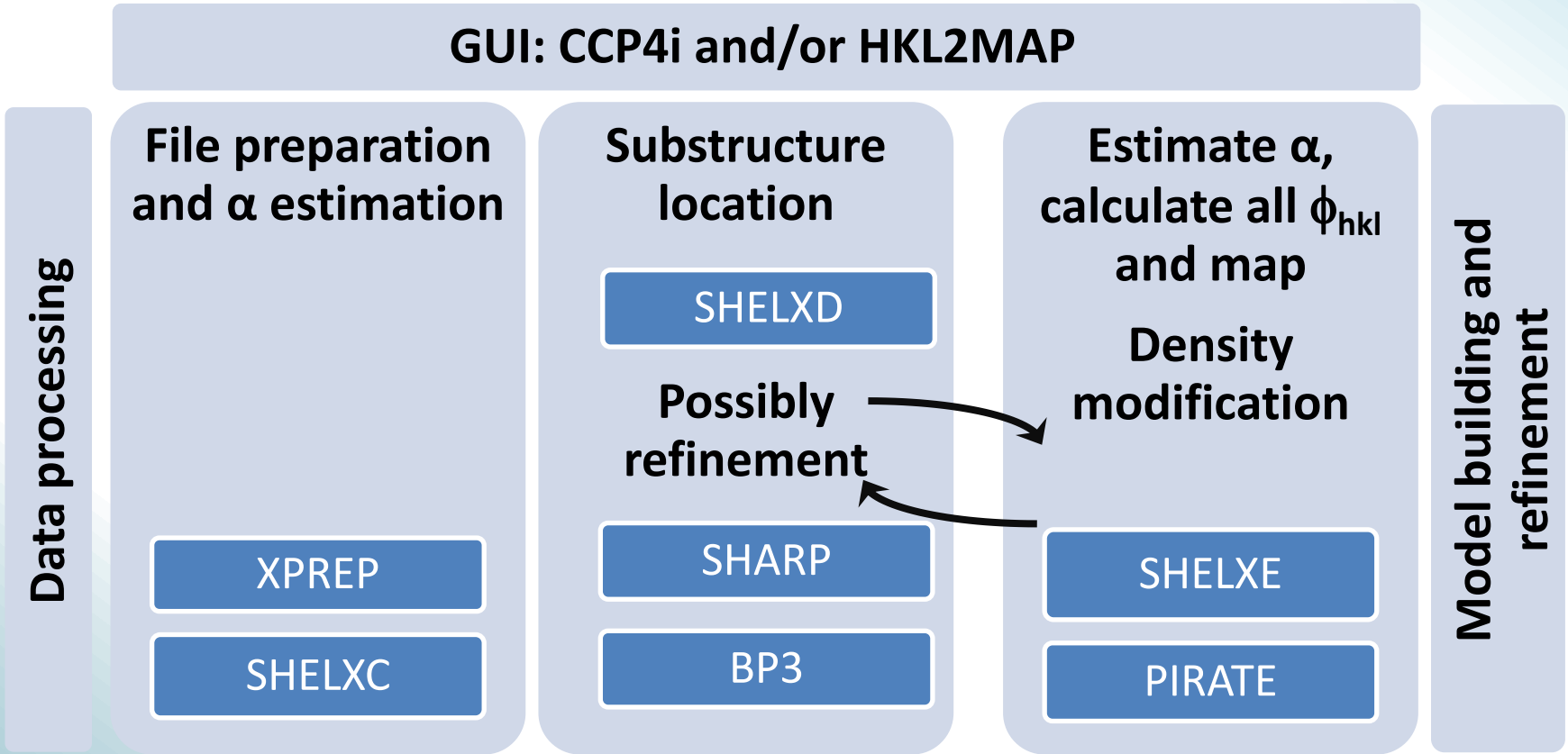


After expansion to 1.0 (MapCC 0.94)

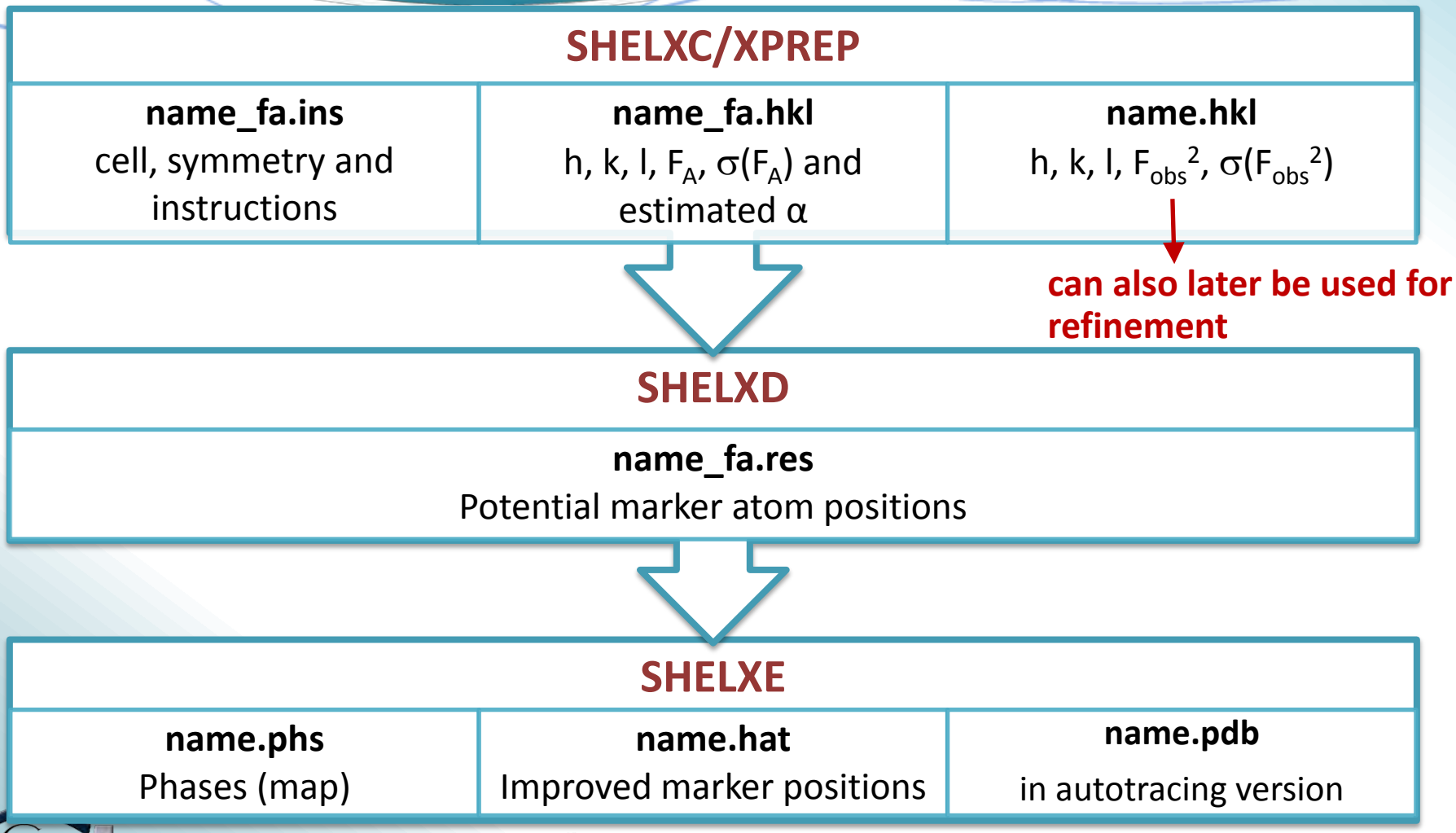
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

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

SOFTWARE


GUI: CCP4i and/or HKL2MAP
Data processing
File preparation and α estimation
XPREP
SHELXC
Substructure location
SHELXD
Possibly refinement
SHARP
BP3
Estimate α, calculate all ϕ_{hkl} and map
Density modification
SHELXE
PIRATE
Model building and refinement

SOFTWARE


SHELXC/XPREP
name_fa.ins
cell, symmetry and instructions
name_fa.hkl
h, k, l, F_A, $\sigma(F_A)$ and estimated α
name.hkl
h, k, l, F_{obs}^2, $\sigma(F_{obs}^2)$
can also later be used for refinement
SHELXD
name_fa.res
Potential marker atom positions
SHELXE
name.phs
Phases (map)
name.hat
Improved marker positions
name.pdb
in autotracing version

SOFTWARE

The new SHELXE beta version 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 < 2.3 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

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

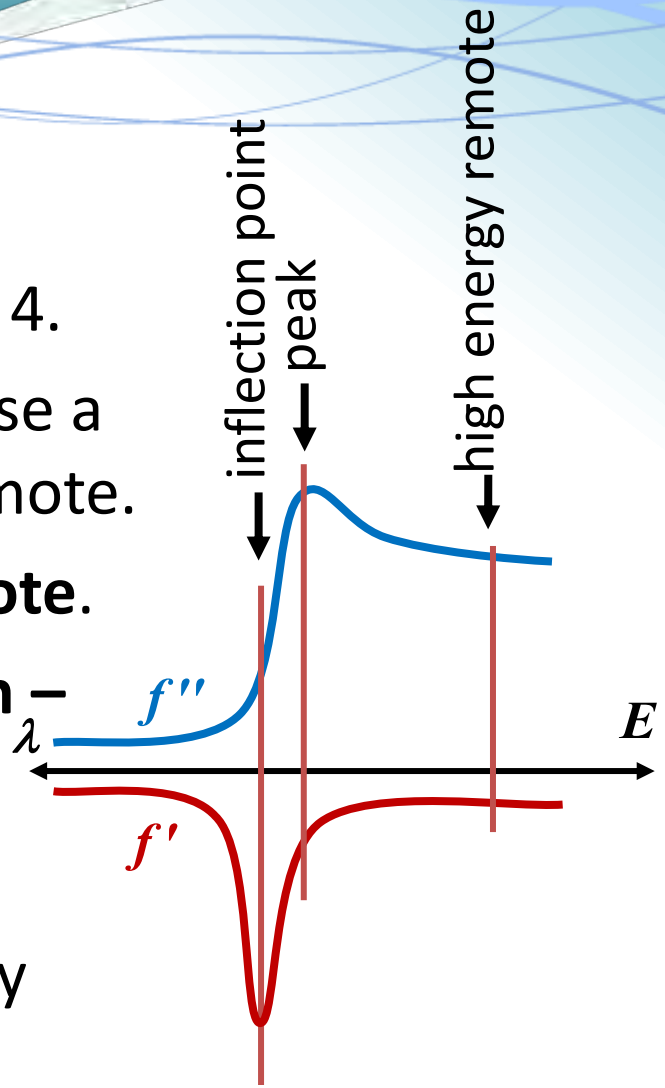
DATA COLLECTION

- Preliminary criterion: $R_{anom}/R_{pim} > 1.5$
- 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 proof the presence of anomalous scatterers in the crystal.

DATA COLLECTION

For MAD

- Collect **peak** with at least multiplicity of 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 always 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 COLLECTION

For SAD

- Best wavelength right and not too close from peak
- Beware not to hit the „white line“ near the peak
- A bit away from peak, radiation damage will be less
- SAD data measured at Peak are often the result of a MAD experiment attempt!

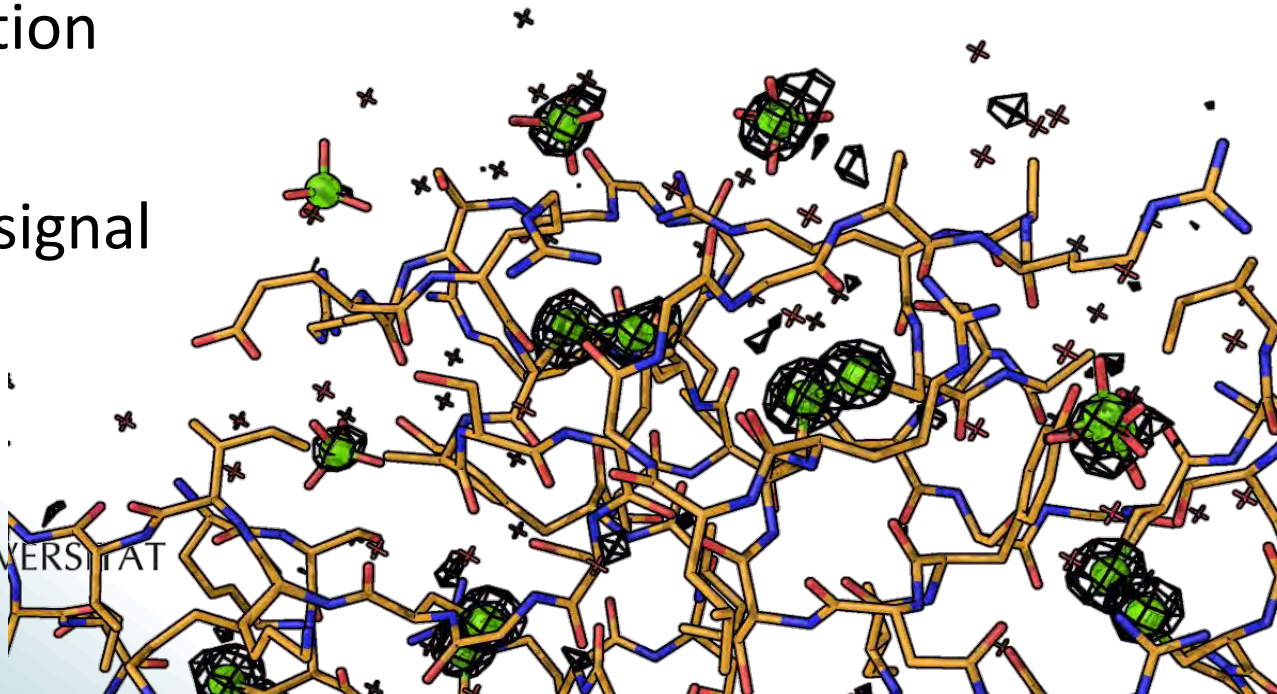
ANODE

New software tool: ANODE

Input: SHELXC file with $|F_A|$ and α

Output: Map for COOT and screen text

- Structure validation
- Analysis of the anomalous signal



SUMMARY

- Experimental methods use marker substructures of certain elements to solve the phase problem via the phasing equations.
- **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

ACKNOWLEDGEMENTS

I am grateful to **George Sheldrick** and **Tim Grüne**, who freely shared with me their knowledge, ideas and material.

I also want to thank the Sheldrick lab for all the support!

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN