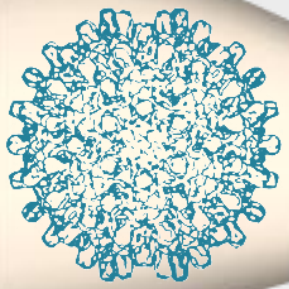
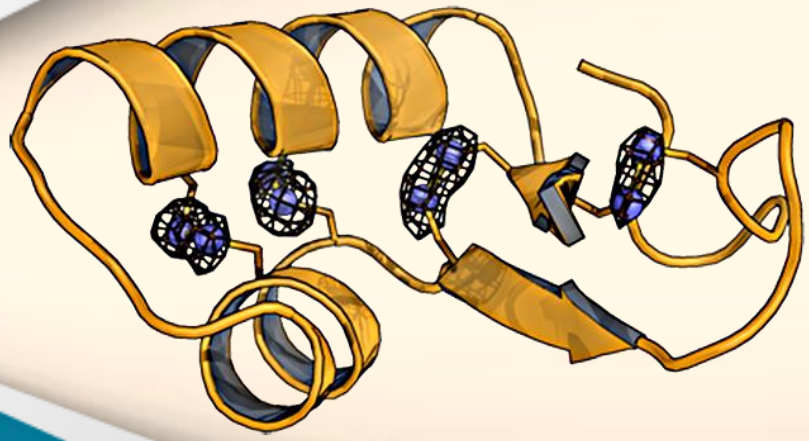
In-House Phasing Using the SHELX Suite





Andrea Thorn

Bruker AXS Webinar
June 25, 2013

BRUKER

Welcome

Sue Byram
Sales & Applications Manager – SC-XRD
Bruker AXS Inc.
Madison, WI, USA

Andrea Thorn
Crystallographic Computing Group
Medical Research Council –
Laboratory of Molecular Biology
Cambridge, UK

What is experimental phasing?

Experimental phasing methods depend on intensity differences. These differences are caused by a „marker atom“ substructure of certain elements.

SAD exploits the anomalous signal.

SIR/MIR utilize derivative crystals.

SHELXC/D/E is a set of programs
for experimental phasing .

Overview of thiw webinar

Theory: **Structure Factors and the Anomalous Signal**

Calculating α: **SHELXC (XC)**

Substructure search: **SHELXD (XM)**

Density modification: **SHELXE (XE)**

Example: **Ga-liquid jet data of Thaumatin**

New feature: **SHELXE & molecular replacement**

Outlook: **ANODE**

Final summary

Theory

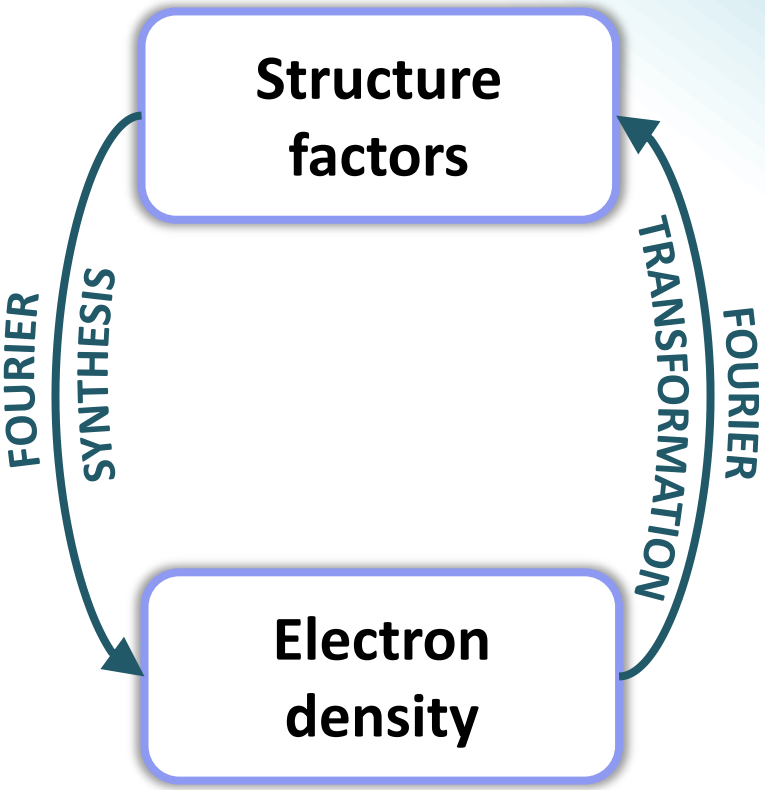
STRUCTURE FACTORS & THE ANOMALOUS SIGNAL

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.

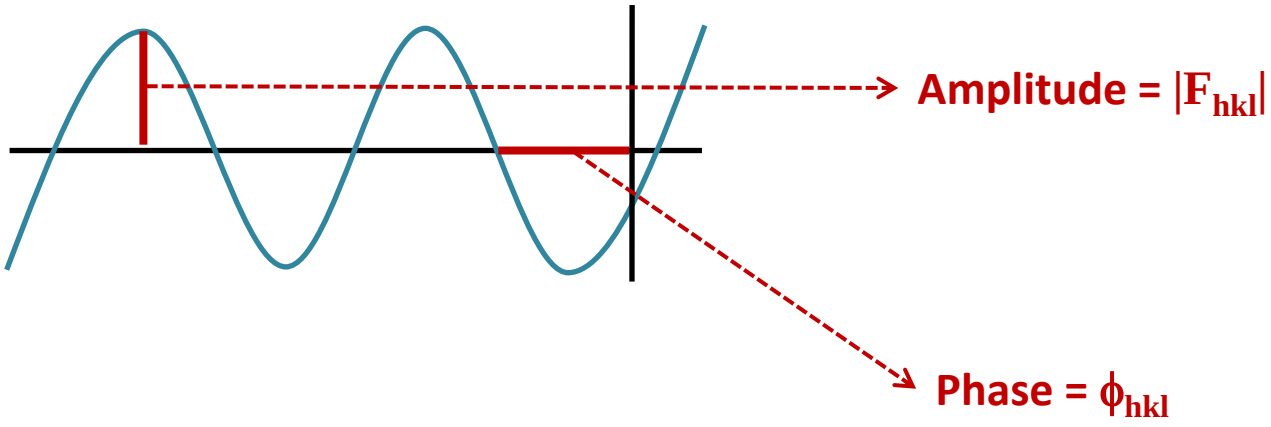


Structure factors

For each reflection, there is a

structure factor F_{hkl}

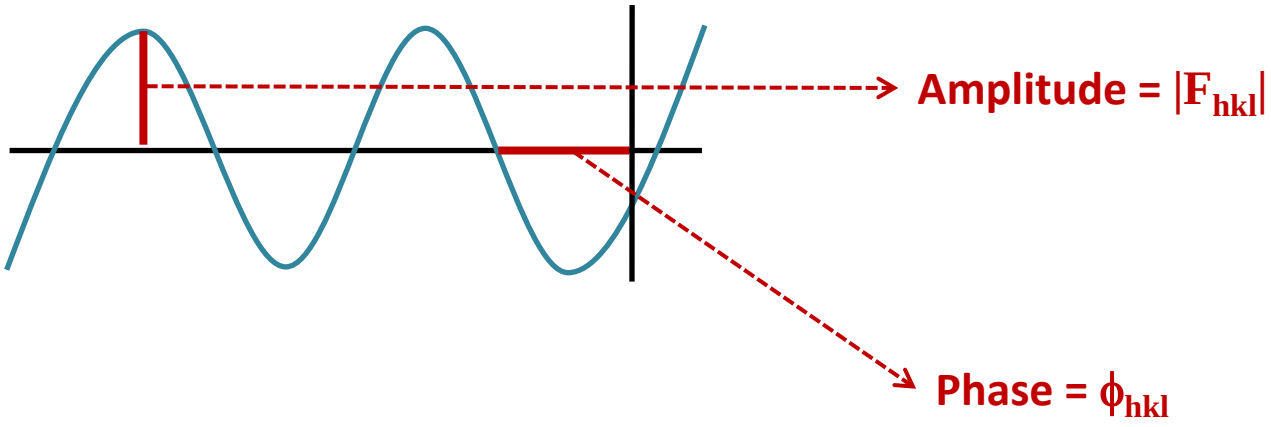
= a wave



Structure factors

structure factor F_{hkl}

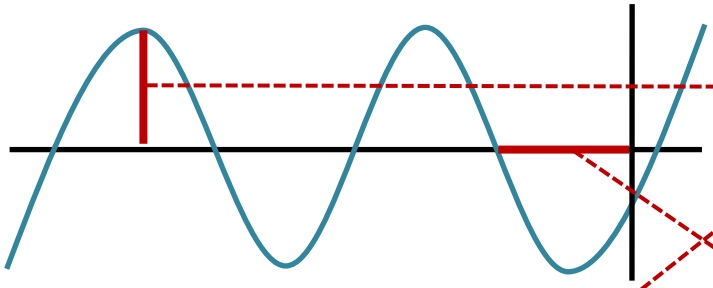
= a wave



Structure factors

structure factor F_{hkl}

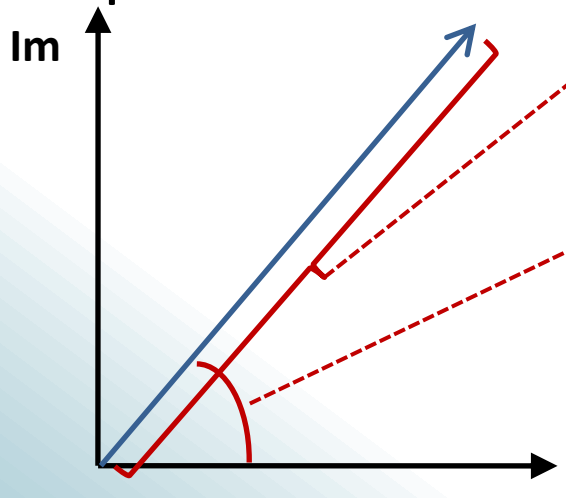
= a wave



Amplitude = $|F_{hkl}|$

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

= a complex number



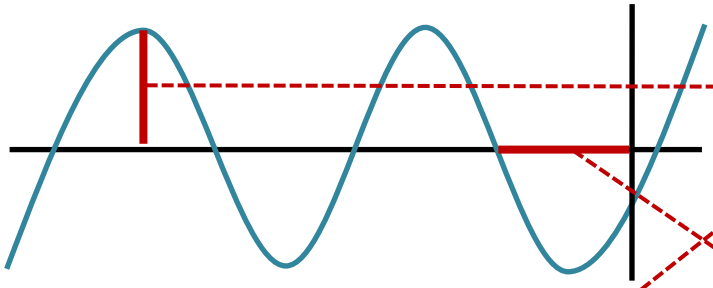
Phase = ϕ_{hkl}

cannot be measured... :-(

Structure factors

structure factor F_{hkl}

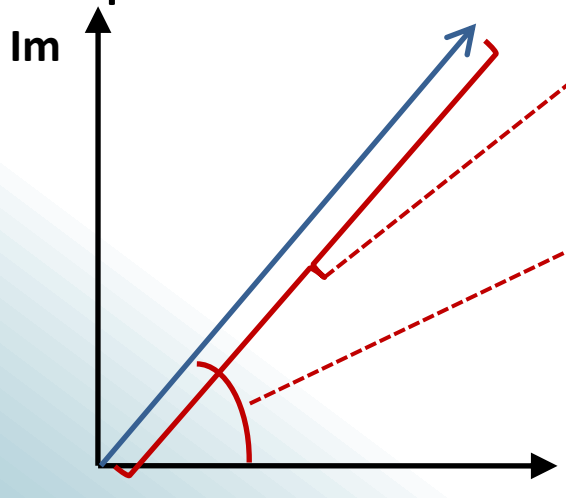
= a wave



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

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

real component

imaginary component

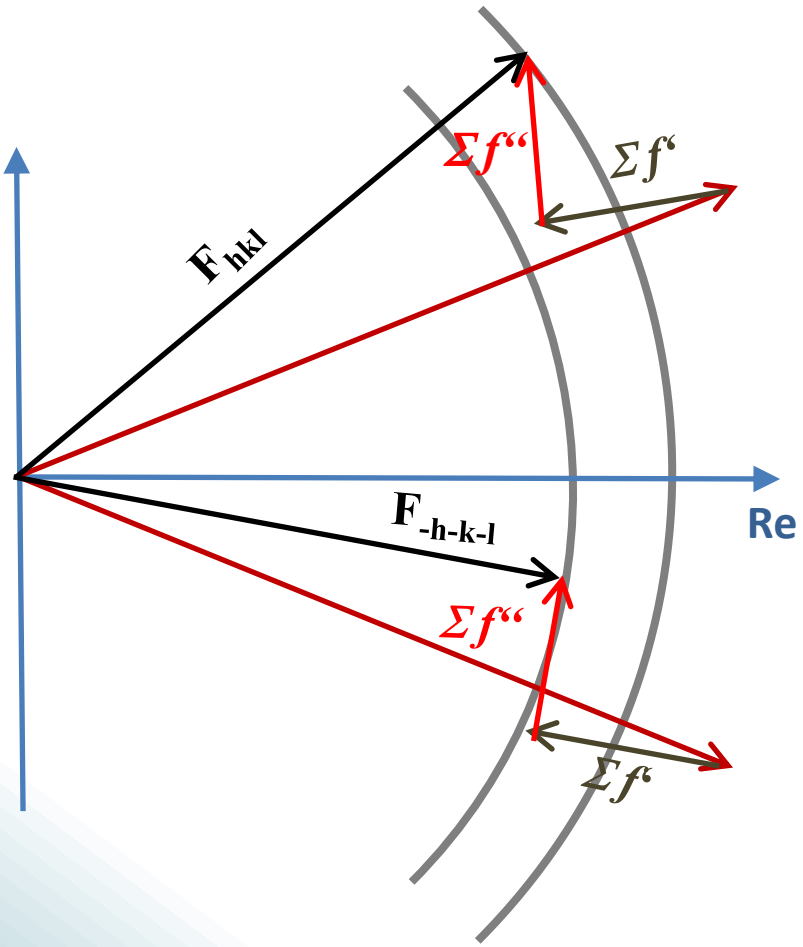
f‘ and *f‘‘* *are observed near absorption edges of the atom‘s element, and are* λ*-dependent*

The anomalous signal

f'' breaks Friedel's law:

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

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

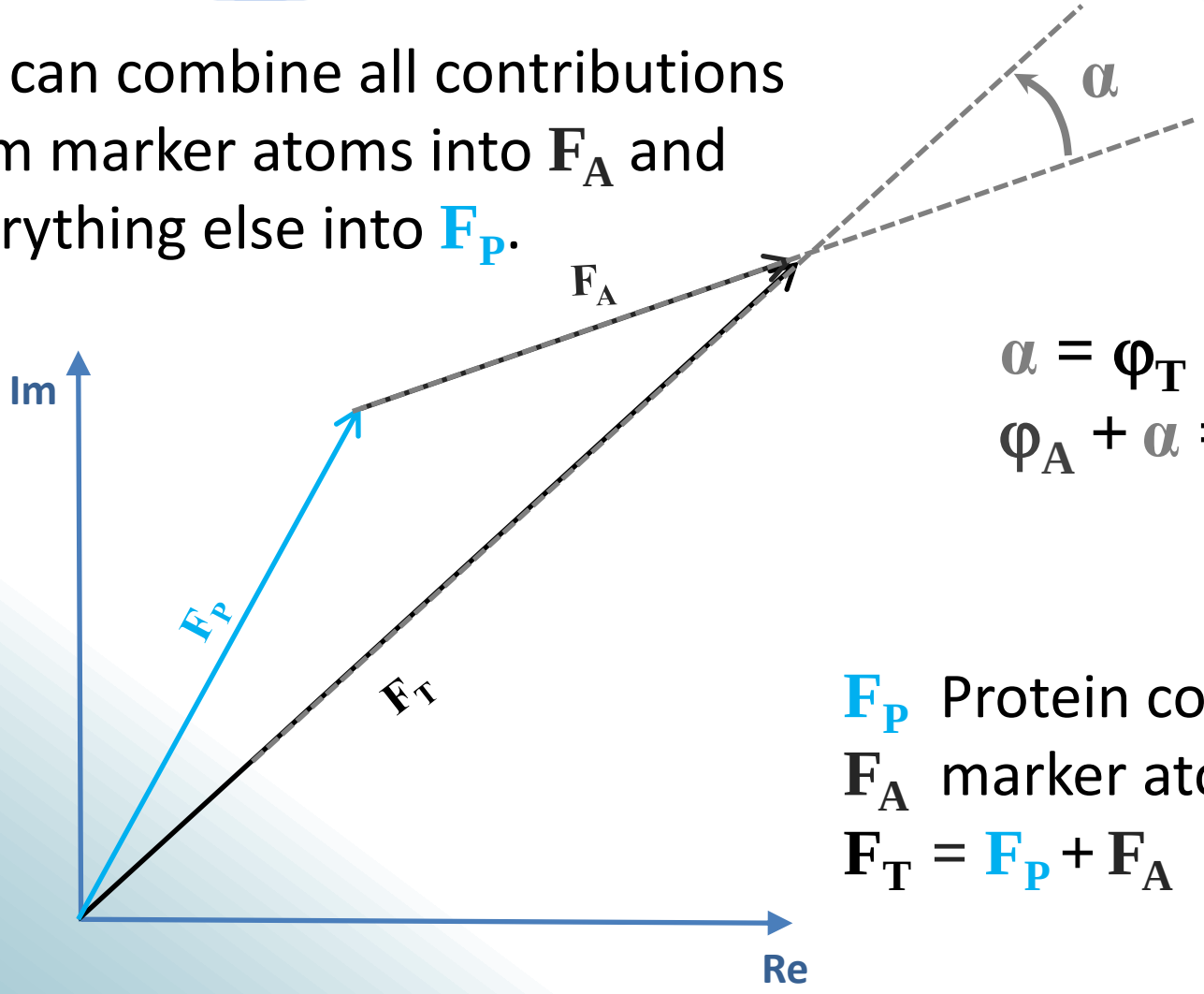


The intensities of Friedel pairs no longer have the same intensity!

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

α angle: A SHELX perspective

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

α angle: A SHELX perspective

So, if we would know the anomalous scatterer positions (or heavy atom positions), we could calculate φ_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**!

SHELXC/D/E

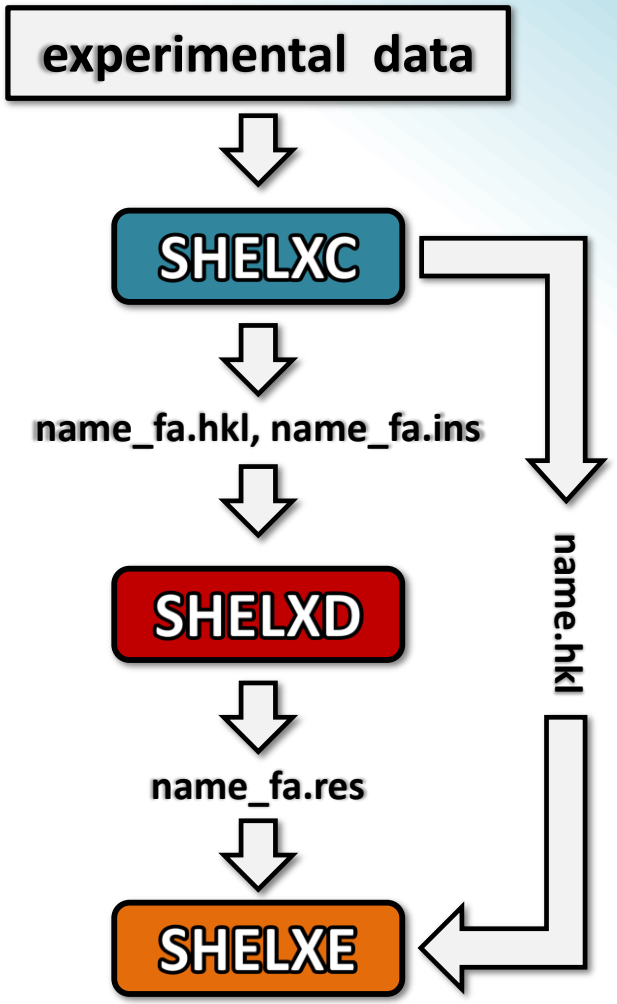
SHELXC: α calculation, data analysis, file preparation

SHELXD: Substructure search

SHELXE: Density modification, Poly-Ala backbone tracing*

ANODE: Validation

Maps and coordinate files are compatible with **COOT**



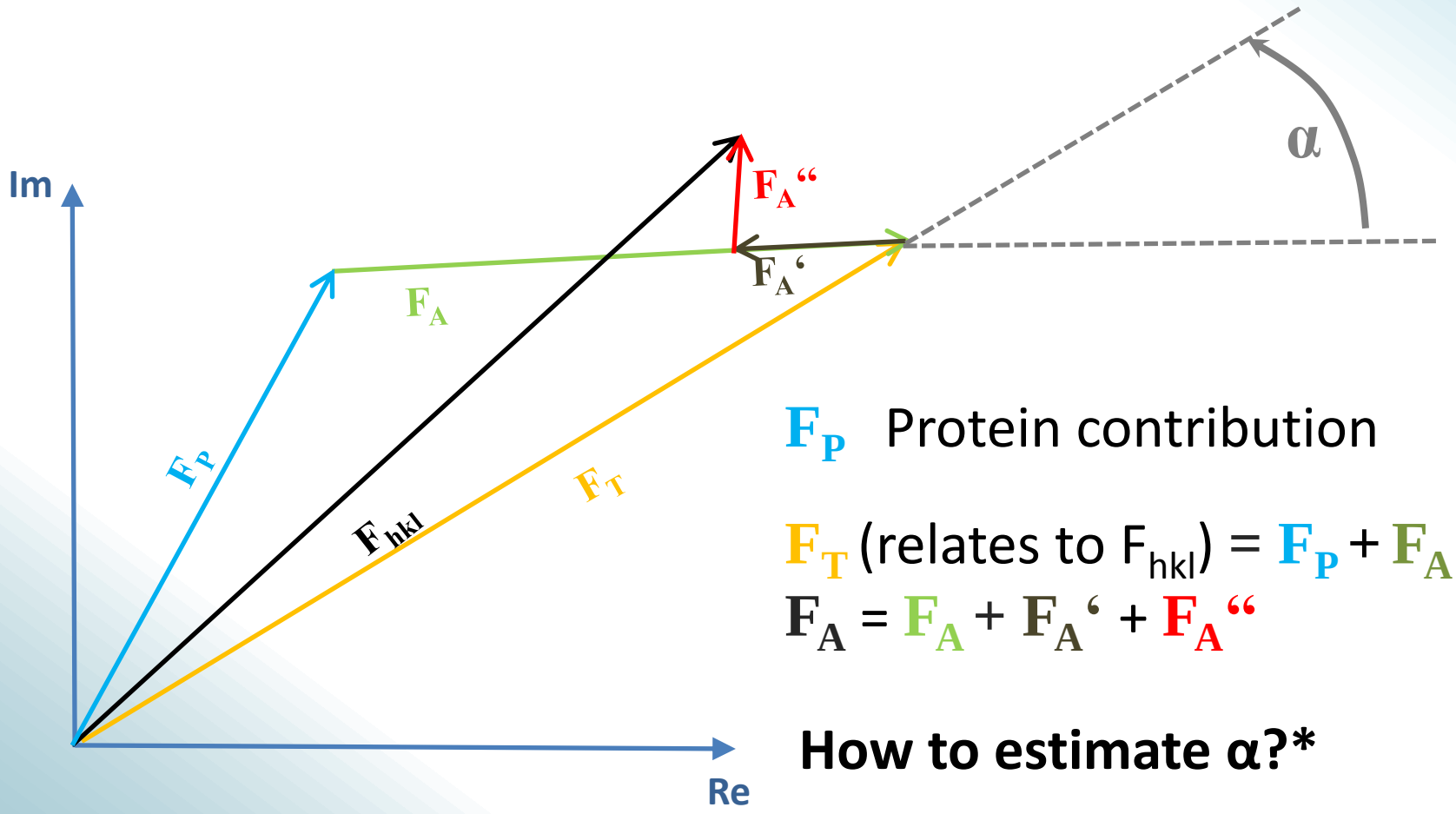
SHELX citation: Sheldrick, Acta Cryst. (2008). A64, 112.

COOT: Emsley et al. (2010) Acta Crystallographica D66, 486.

Calculating α

SHELXC (XC)

From substructure to structure



F_P Protein contribution

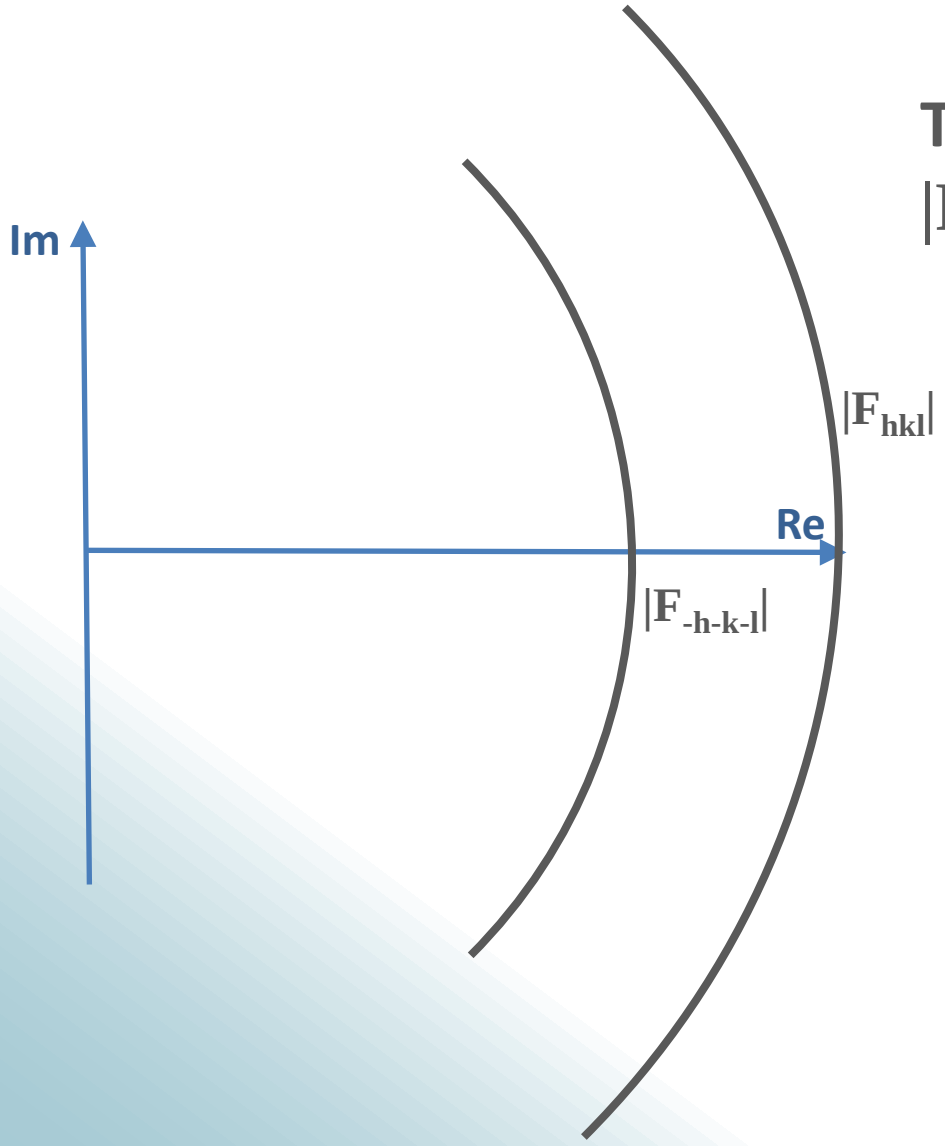
F_T (relates to F_{hkl}) = F_P + F_A

F_A = F_A + F_A‘ + F_A“

How to estimate α?*

* For a mathematical approach: See phasing equations.

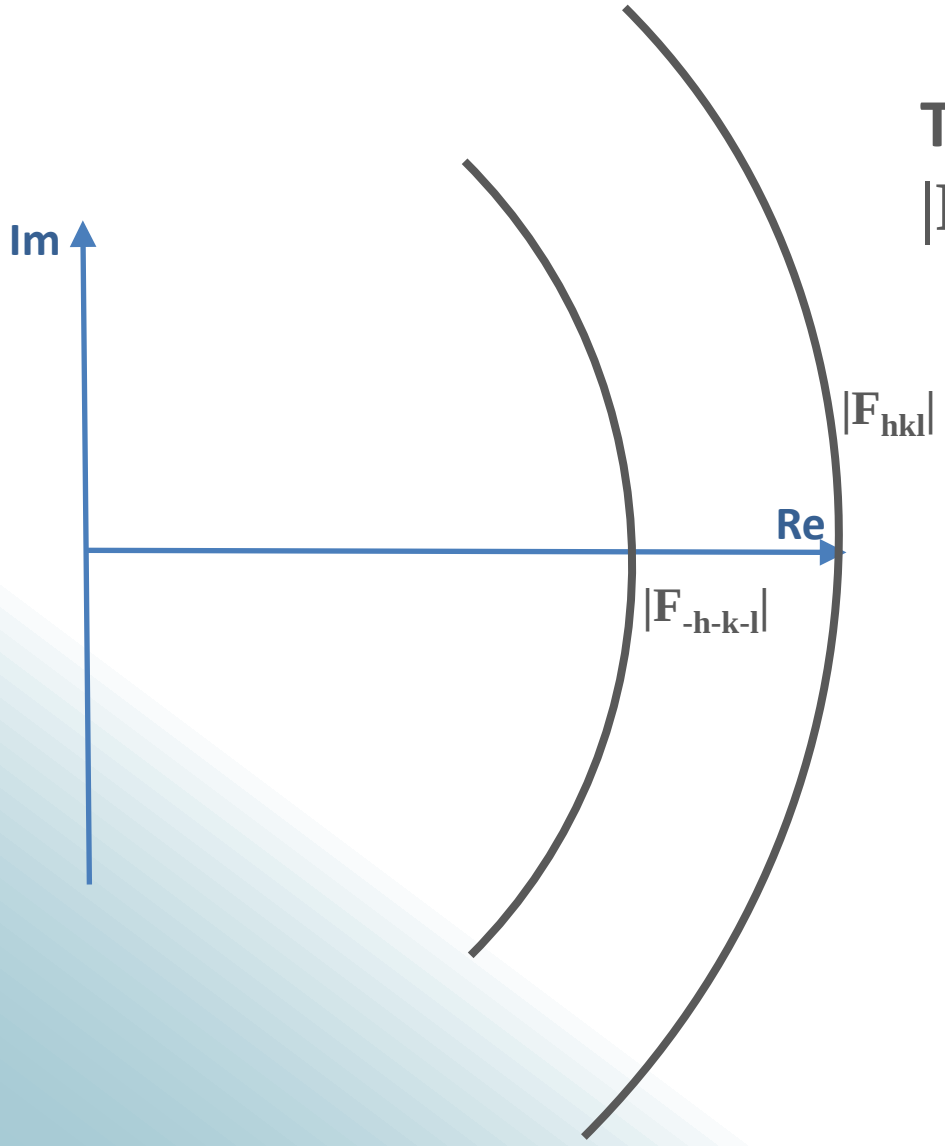
From substructure to structure



This is what we know:

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

From substructure to structure

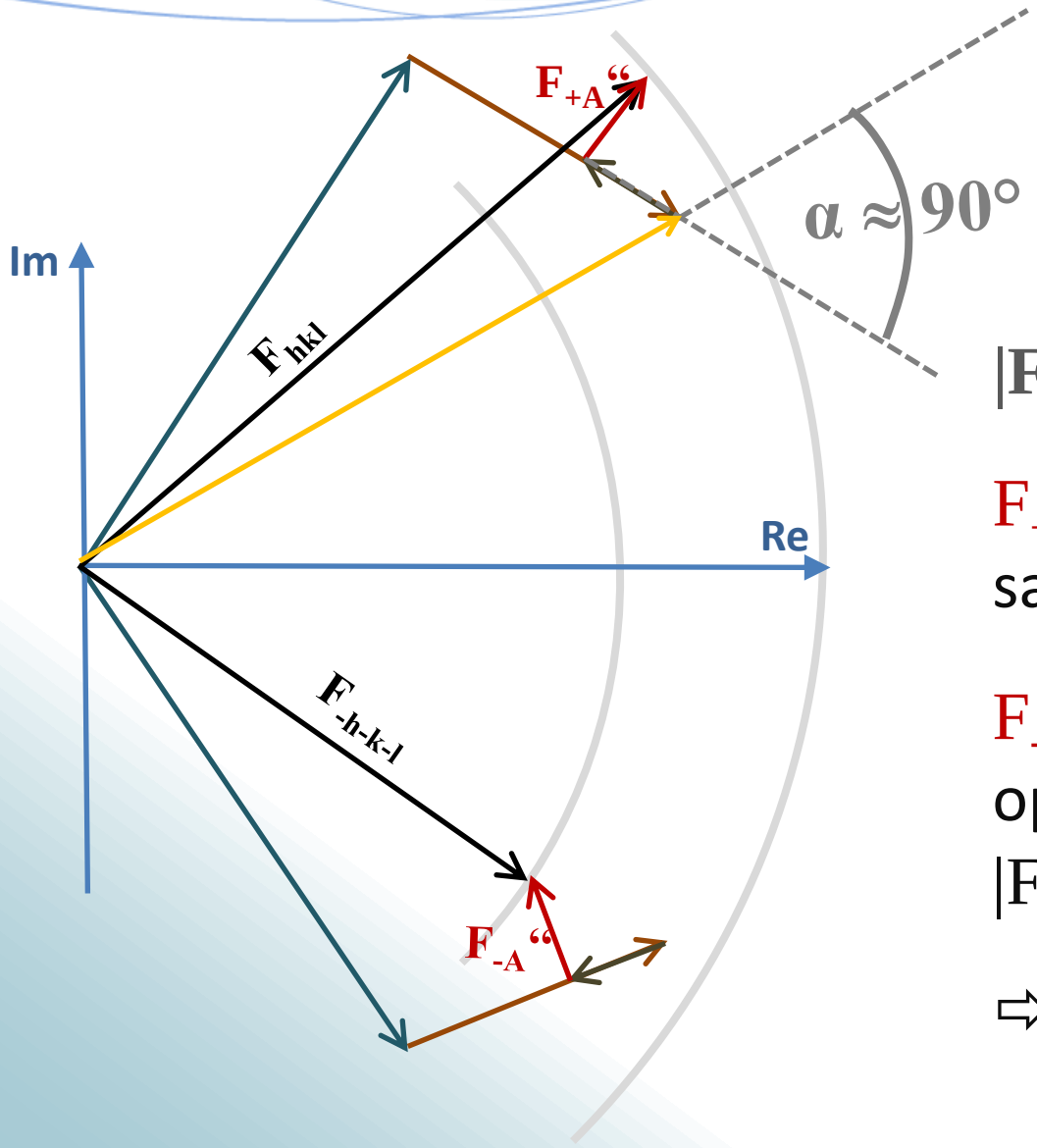


This is what we know:

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

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

From substructure to structure



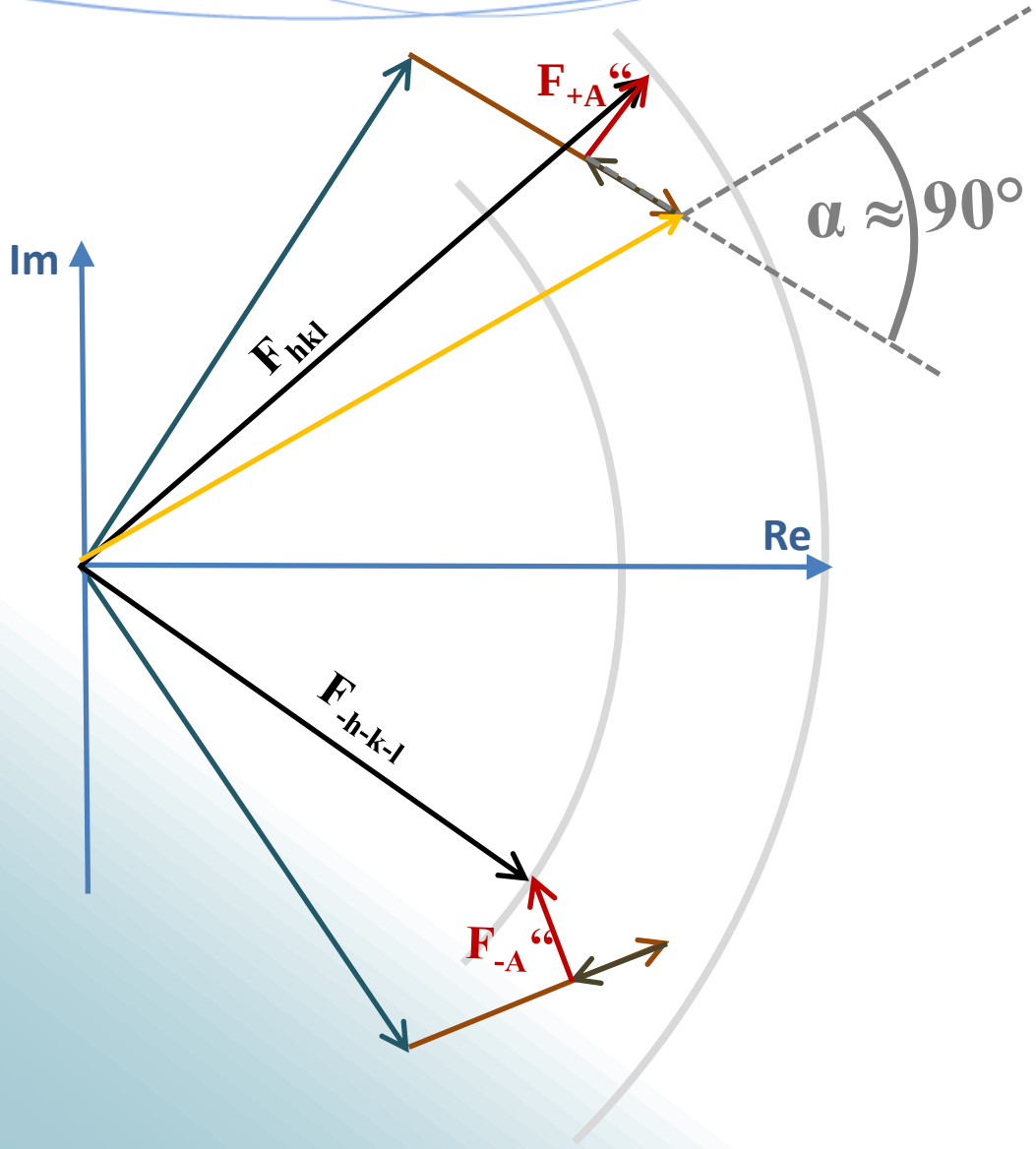
$|F_{hkl}| >> |F_{-h-k-l}|$

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



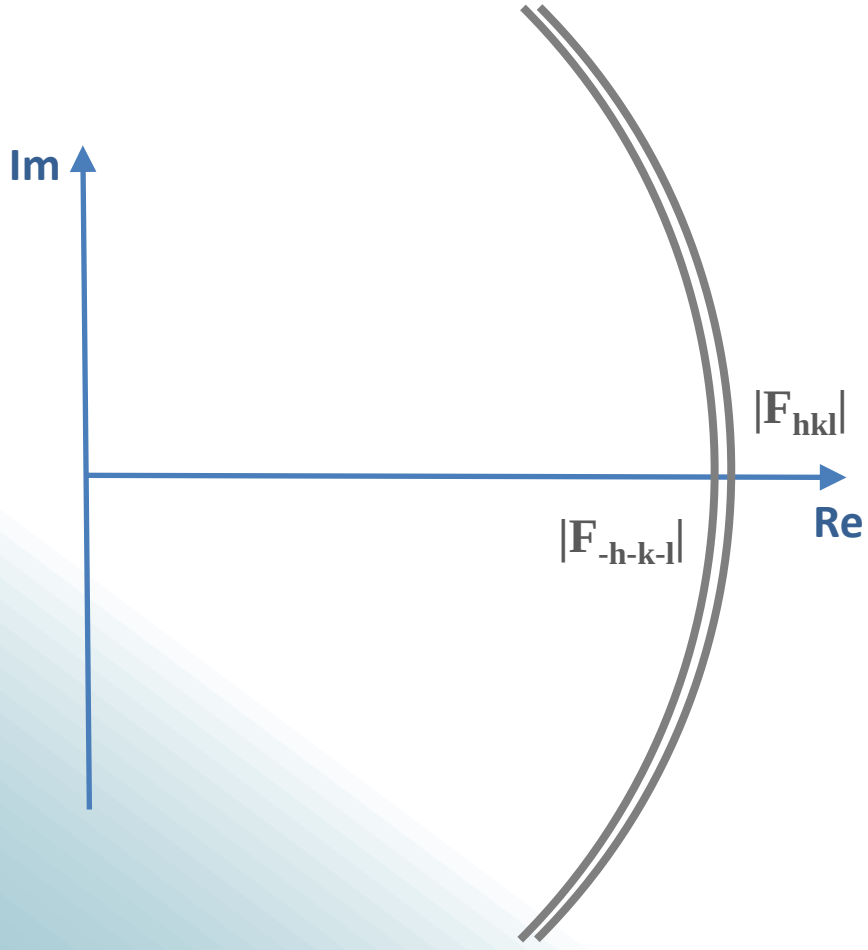
If: $|F_{hkl}| << |F_{-h-k-l}|$

⇨**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°**

Practical matters

- The marker atoms have to have a significant anomalous signal at the wavelength of the X-ray source!
- **Integration**: Check SAINT parameters & active pixel mask
- Check **SADABS** statistics: sad.eps
- Inner shell completeness and outliers?
- 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>
- What could be the best resolution cut-off?

  (SHELXC assumes data resolution + 0.5Å)

New features in SHELXC

New:

- **Input data files in various formats**
  - SHELX/Bruker (*.hkl)
  - SCALA (*.sca)
  - XDS (XDS_ASCII.HKL)
- **Improved statistics** output for low resolution data
- Checks for consistent indexing (several data sets)
  - Re-indexing if necessary with extra output
- $<|E^2-1|>$ quick check on possible twinning

Substructure search

SHELXD (XM)

SHELXD

Originally for *ab initio* solution of big small molecules

Finding the marker atoms is an overdetermined problem with noisy data.

- **Patterson seeding**
- **Dual space direct methods**

Borrowed from small molecule crystallography

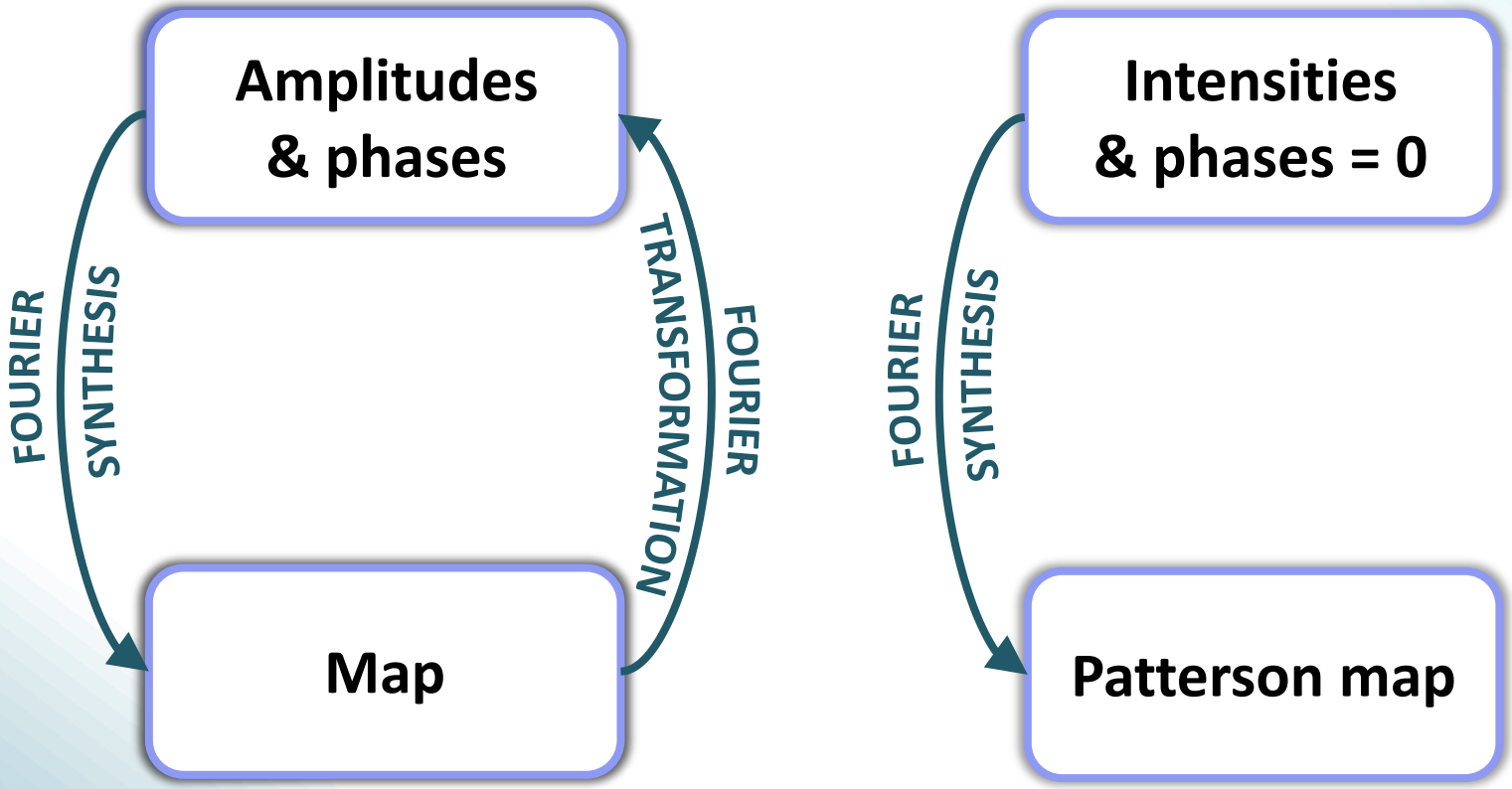
require separate densities to locate marker atoms.

They work here because the marker atoms have large interatomic distances.

Disulfides become ‚supersulfurs‘.

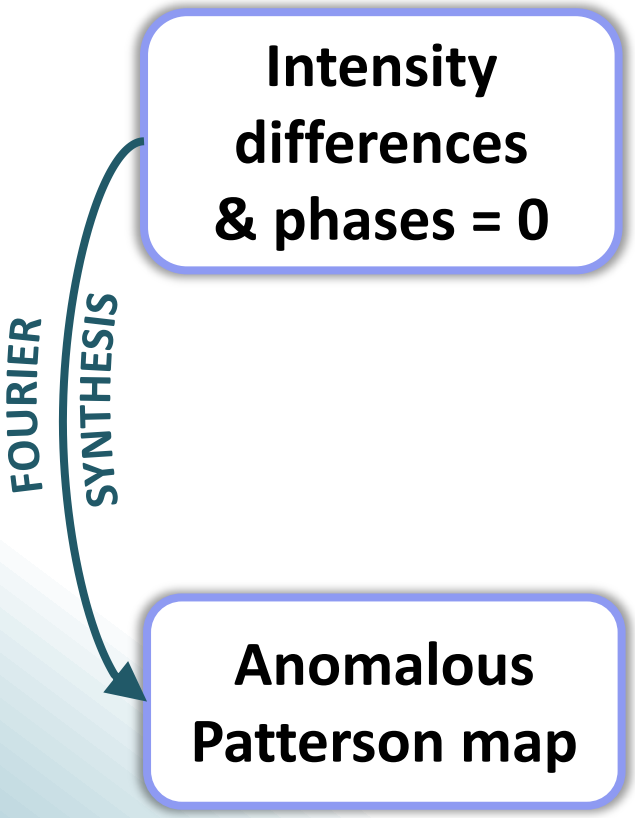
- **E-Values**

Calculating a map - Patterson



- The Patterson map gives information about interatomic vectors. The handedness is not resolved.
- Macromolecules: Too many vectors and resolution issues.

Anomalous Patterson map



Anomalous Patterson map

- Works even at low resolution
- Exploited in **SHELXD** for **Patterson seeding.**

- Can be calculated with **XPREP** (using the SAD option) or by reading in **name_fa.hkl** from SHELXC.

Direct methods & E-values

- Phases of strong reflections are related (as a result of the non-random distribution of atoms.)
- Relations are relatively easy to resolve for few atoms.
- **Dual space direct methods** recycle and modify trial substructures by peak search in the density map and refining phases in reciprocal space.
- E-values are used instead of intensities:

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

ε	scale factor for special position reflections
$<\|F_{hkl}\|^2/\varepsilon>$	mean per resolution shell

Practicalities

Required information:

- *Substructure type:* Which elements/molecules?
- *Number*: Could any marker atoms ‚fuse' into bigger blobs of density because of resolution cut-off?
- *Disulfides*? Use the DSUL command!
- *Resolution cut-off* (see SHELXC)

Indication of correct solution:

- Best CFOM
- Sharp drop in occupancy
- High CC_{all} and CC_{weak}, distinct from non-solutions

SHELXD

New features:

- Multiple CPUs are automatically used if available, with dynamical distribution of tries
- Improved Patterson seeding – usable in all holoedries

Density modification

SHELXE (XE)

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. **The handedness is not yet resolved**.

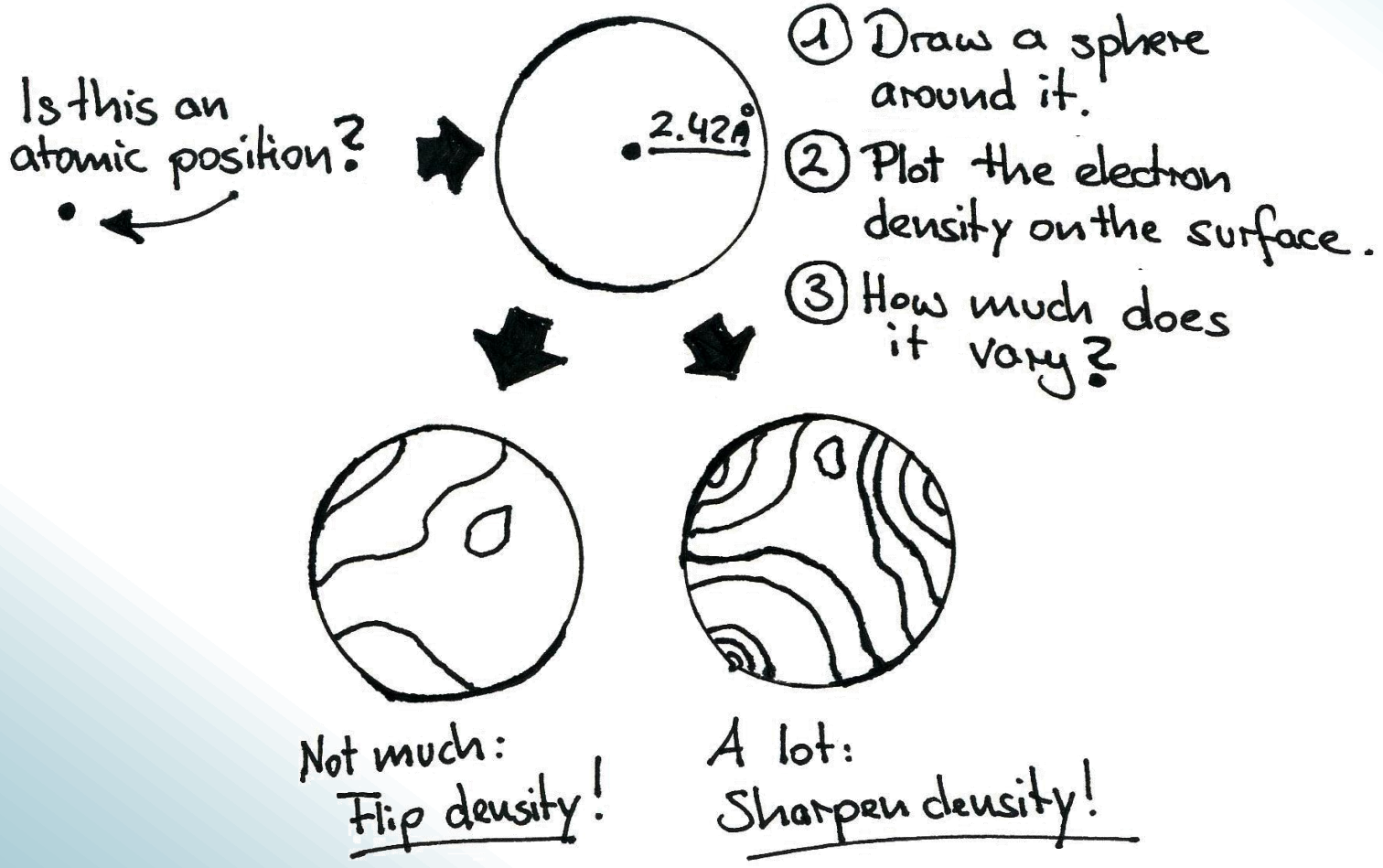
SHELXE applies **density modification**.

- Based on areas filled by disordered solvent
- Solvent area is flattened or flipped
- High solvent content gives often better improvement
- Only one of both substructure hands will give a map will look like protein.

Solvent content has to be estimated before!

DENSITY MODIFICATION

SHELXE uses the sphere-of-influence method:



SHELXE

New:

- **Marker atom substructure optimization (-z option)**
- MR and MR-SAD facilities
- **Poly-Alanine backbone auto-tracing**
  - **main purpose**: Best "experimentally phased map"
  - **byproduct**: poly-Ala trace that can be used as input to ARP/wARP, Buccaneer, Resolve or Main
- **A structure traced is a structure solved**:

  CC of partial structure against native data > 25%

  (2.5Å / 25% rule)

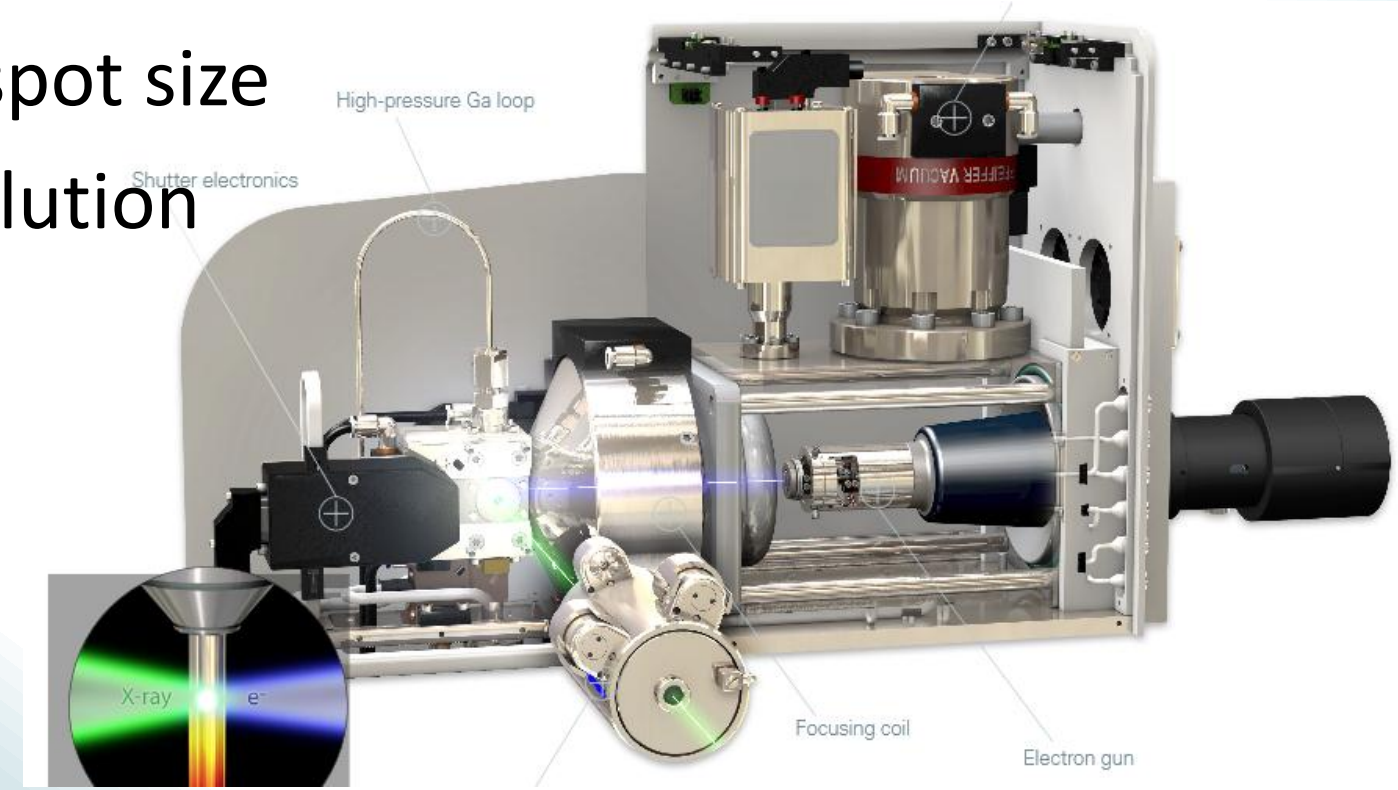
Auto-tracing: Sheldrick, Acta Cryst. (2010). D66, 479-485

Example

GA-LIQUID JET DATA OF THAUMATIN

Ga metal jet

- λ = 1.3414Å
- Liquid Ga Anode
- Variable spot size
- High resolution
- Bright



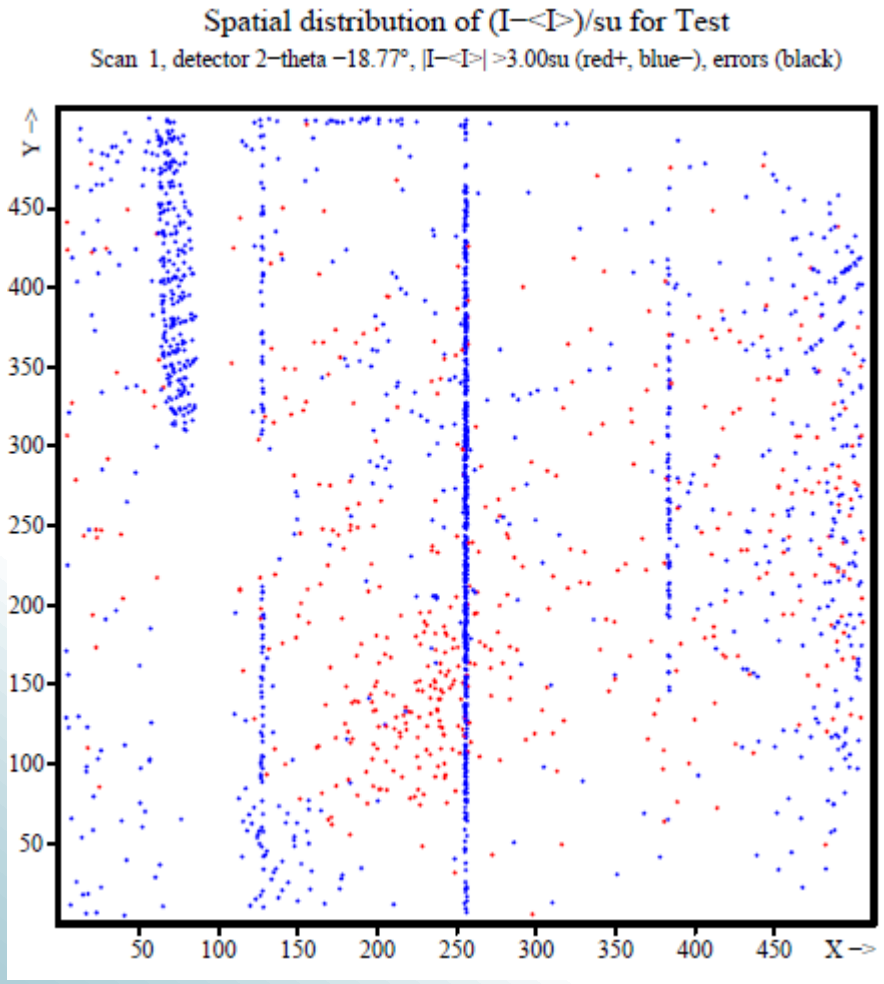
Picture courtesy of BRUKER AXS.

SADABS

293319 total and 34104 unique reflections (…)

292361 total and 34104 unique reflections left after |I-<I>|/su test

SADABS



Data collection parameters

Detector CMOS PHOTON 100	
Exposure time	20 sec
Collection time	3 hours
Rotation angle	0.5°
Max. resolution	1.65 Å
R_{merge}	0.043 (0.33)
$\langle I/\sigma I \rangle$	30.73 (3.88)
Multiplicity	9 (12)

Courtesy of BRUKER AXS.

XPREP (instead of SHELXC)

```
Resl.   Inf.  3.67  2.91  2.54  2.31  2.15  2.02  1.92  1.83  1.76  1.70  1.65
N(data)   3101  2943  2906  2833  2731  2907  2768  3065  2789  2760  2633
Chi-sq    1.13  0.88  0.75  0.90  1.13  1.03  1.15  1.03  1.07  1.11  1.15
<I/sig>   92.3  72.6  45.3  33.3  25.3  19.1  14.4  10.0   7.3   5.5   3.8
%Complete 98.5  99.6  99.8  99.8  99.9  99.8  99.9  99.9  99.9  99.9  99.9
<d"/sig>  1.42  1.03  0.90  0.80  0.83  0.80  0.81  0.78  0.78  0.78  0.75
CC(1/2)   50.7  17.9  12.6   2.9  -2.3  -0.6   0.5  -4.2  -2.0  -5.8  -7.5

For zero signal <d'/sig> and <d"/sig> should be about 0.80

   20045 Reflections written to file thau_fa.hkl for input to SHELXD/E

   31436 Reflections written to file thau.hkl for input to SHELXE
```

Good data quality vs. weak anomalous signal: The cut-off will play an important role!

```
2    4    1 2935.01 2316.70  90
3    4    1 1573.05 1089.56 270
1    5    1 3457.25 2688.83 270
2    5    1 2592.79 2161.85 270
3    5    1 3172.40 1087.23 270
4    5    1 3247.10 1482.29  90
1    6    1 2093.32 1144.36 270
2    6    1 9075.51 1014.62  90
3    6    1 2302.90  875.88  90
4    6    1 4769.11  893.33  90
5    6    1  482.08 1800.12  90
1    7    1 3570.51 1041.23 270
```

Ga-data: SHELXD

name_fa.ins:

```
(...)
 SFAC S                  Only relevant for output
 SHEL 999  3.2           Resolution cut-off
 PATS                    Use Patterson seeding (not random)
 FIND 6                  Find 6 peaks
 DSUL 6                  6 of them super-sulfurs
 MIND -3.5 0.1           Do not search on special positions.
 NTRY 1000               Do a thousand trials.
(...)
```

Courtesy of BRUKER AXS.

Ga-data: SHELXD

name_fa.res:

```
S001    1   0.984724   0.302571   0.093172    1.0000   0.2
S002    1   1.098938   0.484413   0.151469    0.9965   0.2
S003    1   1.078568   0.215660   0.055399    0.9600   0.2
S004    1   1.112525   0.148108   0.226177    0.8433   0.2
S005    1   1.163547   0.042923   0.097446    0.7522   0.2
S006    1   1.086253  -0.030050   0.064236    0.7452   0.2
S007    1   1.044640   0.457344   0.228629    0.7281   0.2
S008    1   1.108905   0.650631   0.242303    0.6752   0.2
S009    1   1.107883  -0.054044   0.069621    0.6059   0.2
S010    1   1.087523   0.125345   0.222393    0.5949   0.2
S011    1   1.143530   0.049860   0.086654    0.5139   0.2
S012    1   1.091531   0.663787   0.231690    0.4448   0.2
S013    1   1.042419   0.448683   0.222095    0.3051   0.2
```

Courtesy of BRUKER AXS.

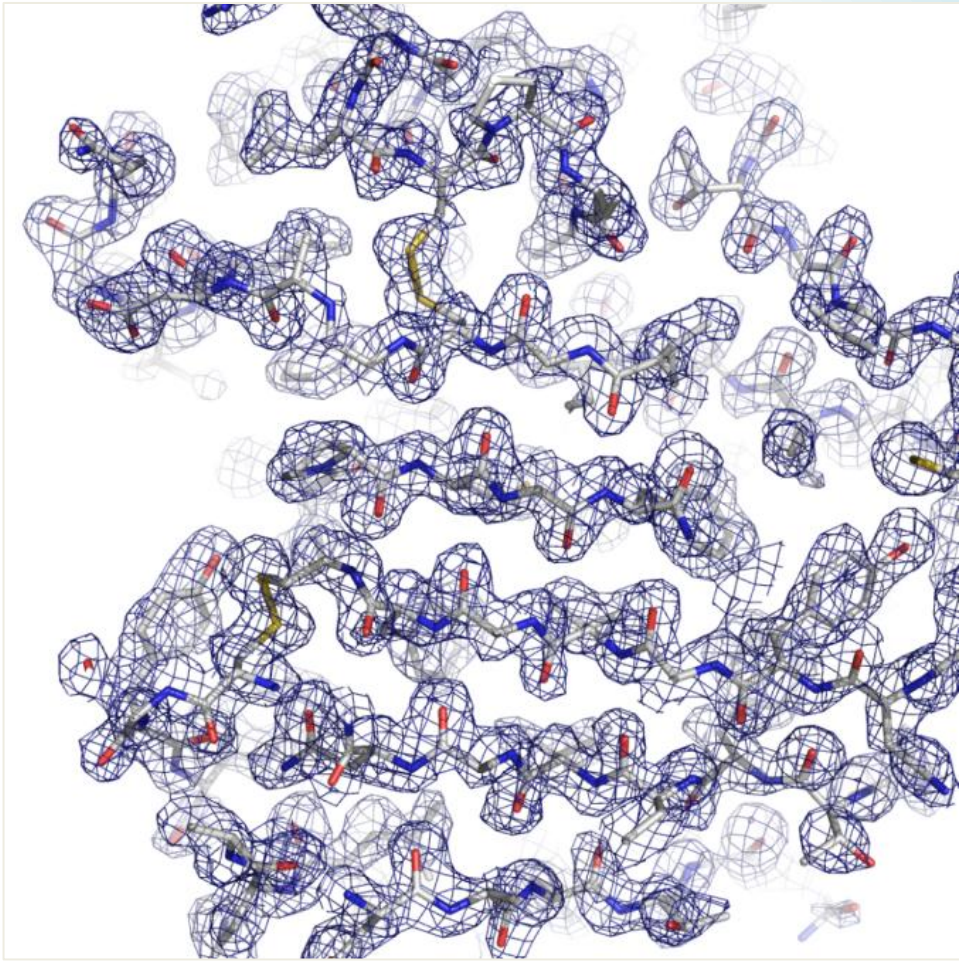
Ga-data: SHELXE

```
shelxe name name_fa -s0.50 -q -a -z (-i)
```

CC = 46.50%

(wrong hand 11.07%)

199 residues in 5 chains



Courtesy of BRUKER AXS.

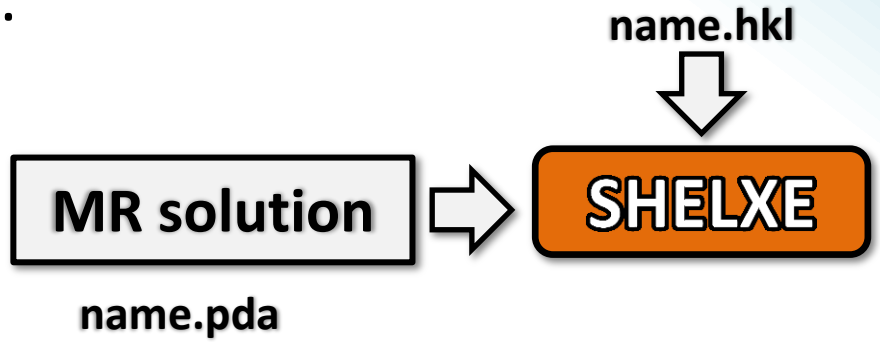
New feature

SHELXE & MOLECULAR REPLACEMENT

How to...

Input phases for SHELXE auto-tracing can be from a molecular replacement model.

- Improve phases
- Expand the map
- Remove bias



Data < 2.1 Å are required.*

* 2.5Å for auto-tracing in combination with experimental phase information.

Thorn & Sheldrick: Extending Molecular Replacement solutions with SHELXE, Acta Cryst. D, submitted

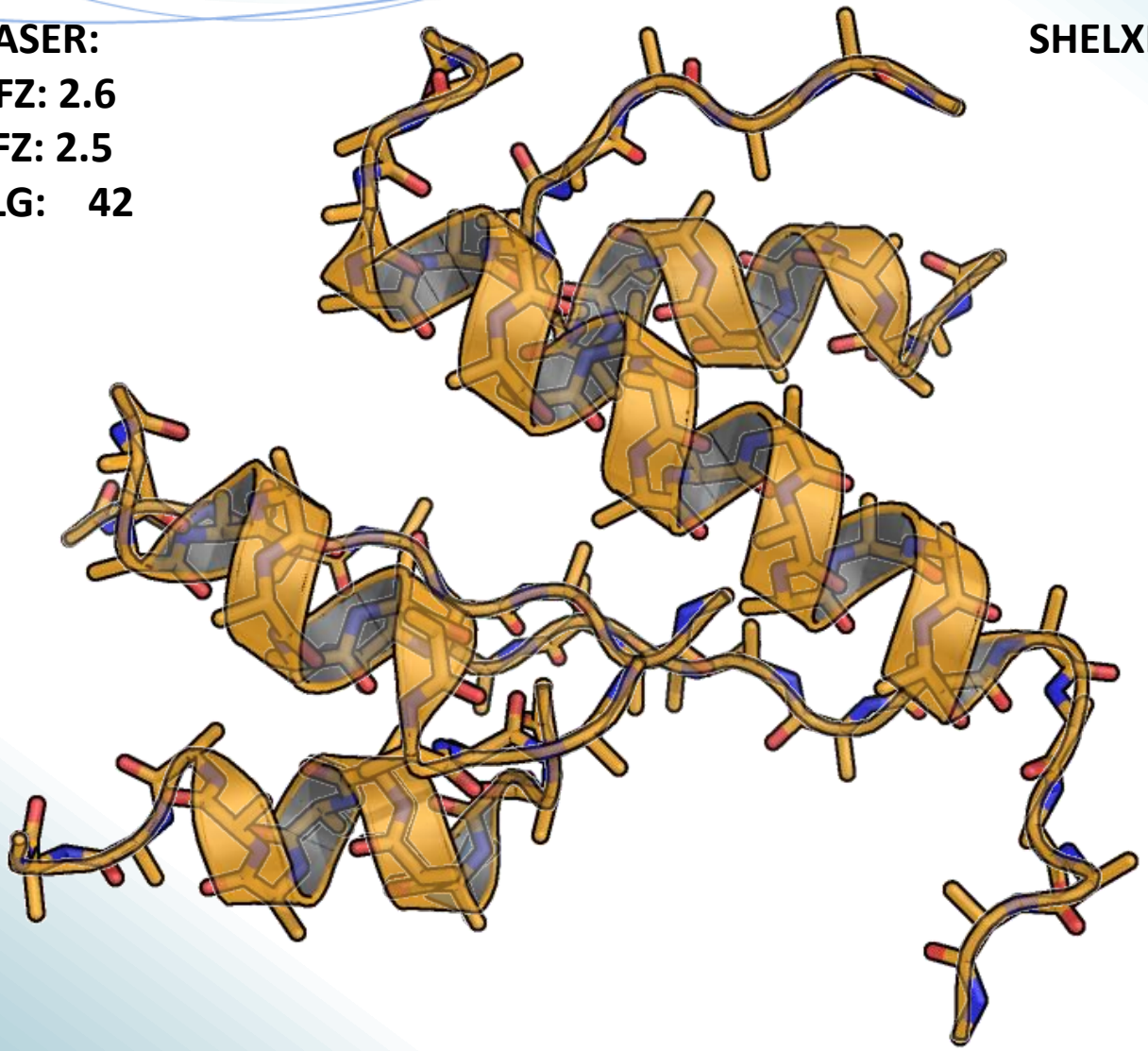
Example: 1ZZK

AMPLE/PHASER:

RFZ: 2.6

TFZ: 2.5

LLG: 42



SHELXE CC : 32.17%

86 residues

AMPLE: Bibby et al., *CCP4 Newsletter* (2012) 48.

Outlook

ANODE

Wait a moment...

If I can use the anomalous signal for a Patterson, can I also calculate a map from the anomalous signal?

Yes! And it can be used after phasing for atom type identification, radiation damage assessment *et cetera*.

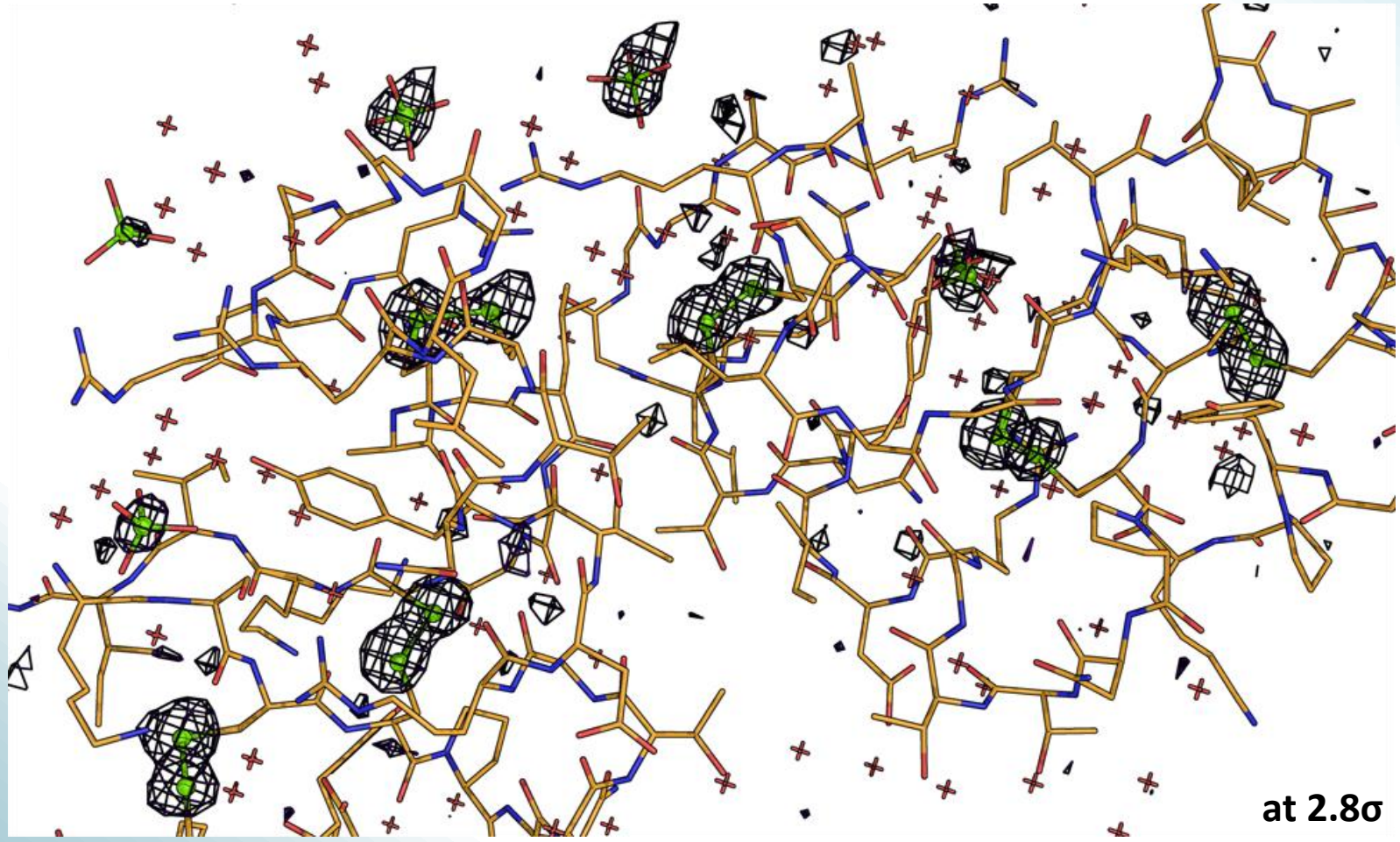
One possibility to get such a map quickly from SHELX type input is the program **ANODE**.

Upcoming Bruker webinar:

11/25/13

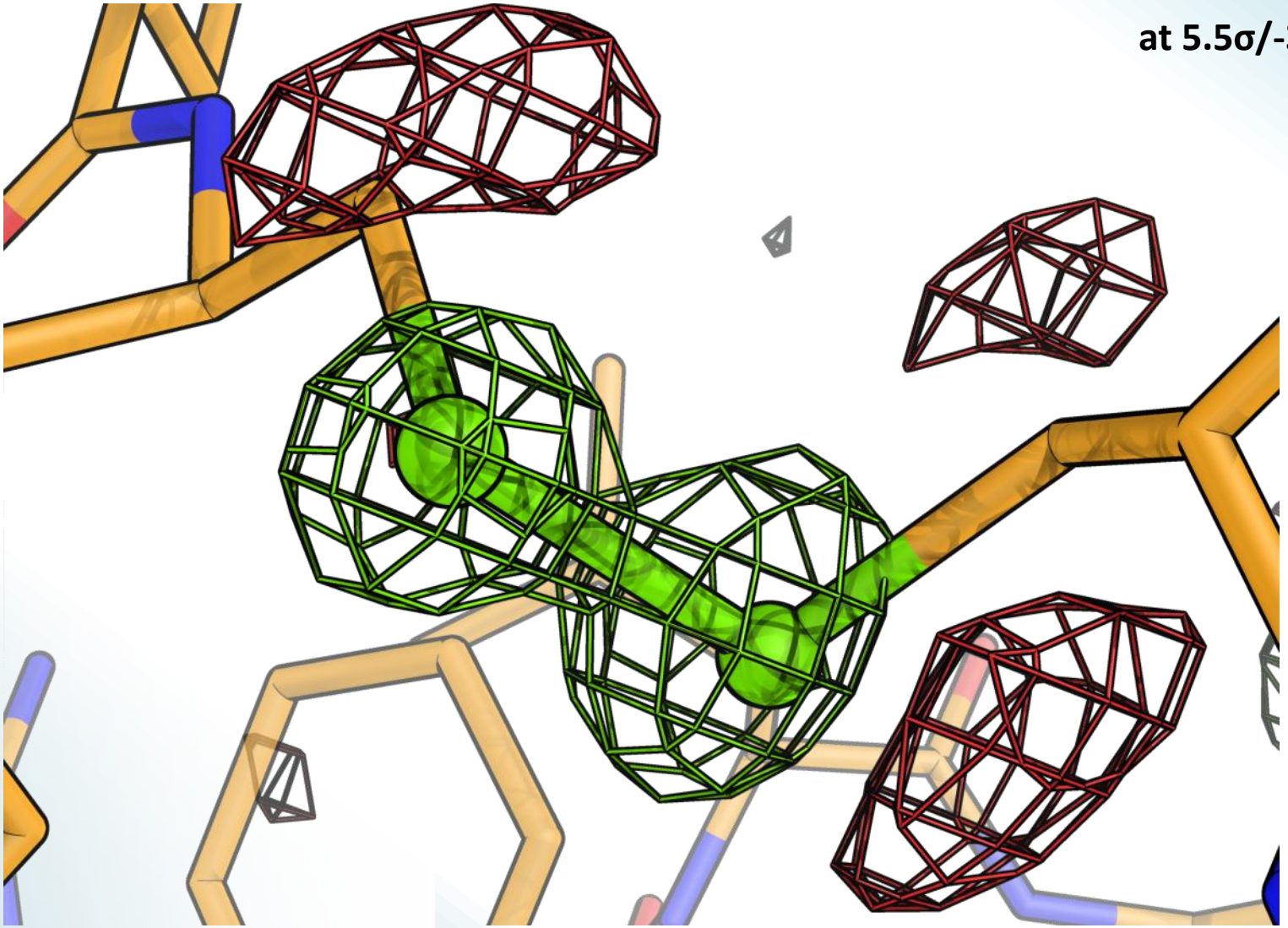
ANODE: Anomalous and Heavy-atom Density Calculation

Anomalous density maps



PDB 2V9B, Pal et al. (2008). *Acta Cryst. D*64, 985-992.

RIP density maps



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

Final

SUMMARY

SUMMARY

- SHELXC/D/E or XC/XM/XE:
  Set of programs for experimental phasing
- All kinds of experimental phase information are treated similar in SHELXC/D/E.
- New options allow for better indication of data quality and successful phasing and for the combination of phase information from different methods.
- Data quality and processing are key steps.

ACKNOWLEDGEMENTS

George Sheldrick
Ronan Keegan, Max Nanao, Isabel Usòn

New 2013 release available via **PROTEUM update** or **http://shelx.uni-ac.gwdg.de/SHELX/**

SHELX_2012 replaces **all** previous versions.

More information and tutorials:
http://shelx.uni-ac.gwdg.de/~athorn/

MRC | Laboratory of Molecular Biology

Any questions?

Please type any questions you may have for our speakers in the Q&A panel and click Send.

How did we do?

When you exit the webinar, please fill out our evaluation survey to let us know. We appreciate your feedback.

Thank you!

Globally interacting with experts

Related links

- News
- Events

Contact

Home • Service • Education & Training • Webinars •

SC-XRD Webinars

Live and On-Demand

More SC-XRD webinars available at
www.bruker.com/service/education-training/webinars/sc-xrd.html

Webinar	Content
Apr 02, 2013 **Fast, Intuitive Structure Determination II: Crystal Indexing and Data Collection Strategies**	Attend this one-hour webinar to learn skills for crystal indexing and data collection. Dr. Bruce Noll will cover the APEX2 tools that can be used to select reflections and sort them into arrays to facilitate indexing routine, twin and modulated structures. He will also examine the diffraction behavior of the sample to determine the best data collection parameters. Register now
Feb 26, 2013 **Advanced Crystallography: Data Collection and Data Reduction Techniques for Modulated Structures**	Drs. Michael Ruf and Charles Campana give a one-hour presentation demonstrating the steps required to collect and process data for modulated structures. Specific topics include: how to recognize modulated structures; introduction to q-vectors; APEX2 software tools; preparation of files for JANA2006; and examples of the solution of modulated structures from single-crystal diffraction data. View recording Download slides
Dec 05, 2012 **PROTEUM2 and Foreign Frame Formats**	View this 45-minute webinar to learn about Bruker's PROTEUM software, a suite of crystallographic data integration and reduction tools for macromolecules. Dr. Matt Benning and Dr. Michael Ruf will discuss how to reduce data from various detectors and show results from synchrotron beamlines. View recording Download slides
Nov 13, 2012 **Advanced Crystallography: Refinement of Disordered Structures**	Dr. Charles Campana gives a one-hour presentation demonstrating the steps required to refine disordered crystal structures. Along with specific examples, topics covered include SHELXL, SHELXTL, free variables, PART n and –n instructions, constraints vs restraints, and advanced refinement instructions. View recording Download slides
Dec 06, 2011 **High-Pressure Crystallography**	Dr. Francesca Fabbiani of Georg-August-Universität Göttingen takes you through the steps for performing a high-pressure single-crystal X-ray diffraction experiment with a diamond-anvil cell (DAC). You will learn DAC loading techniques and useful strategies for data collection, data reduction, and structure refinement. View recording Download slides
Oct 13, 2011 **Advanced Crystallography – Publication of Crystal Structures**	Publication of crystal structures is often a researcher's least favorite task. This 90-min webinar will help. It describes the steps required to prepare crystal structures for publication in chemical and crystallographic journals, including collection of high-quality intensity data, common refinement problems, data evaluation, error analysis, preparation of CIF files, and the use of CIF-checking programs. View recording Download slides
Jun 22, 2011 **CMOS: The Next Generation in X-ray Detector Technology**	CCDs have long been one of the most important detector technologies for X-ray crystallography. Recently, however, new detectors based on CMOS technology have begun to supplant CCDs. View this webinar to learn about Bruker AXS' new X-ray detector based on CMOS technology, the PHOTON 100. View recording Download slides
May 03, 2011 **Using APEX2 to Analyze Twinned Crystals**	Twinned crystals are common occurrences in many laboratories and pose special challenges in structure solution. View this 76-minute webinar to learn how to visualize, index, integrate, scale, solve and refine structure data from twinned crystals with the APEX2 software suite. (Note: playback begins at 14:00 minutes into the 1:30:20 hour presentation.) View recording Download slides
	Accurate 3D structures determined by X-ray diffraction require

Like what you learned in this webinar?

Subscribe to Bruker's ***FIRST Newsletter*** to get webinar announcements, technical articles, and X-ray crystallography news delivered right to your inbox

Subscribe at:

https://www.bruker.com/about-us/register.html

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

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

Additional material

Phasing equations

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

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

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

$\mathbf{F_T}$ **Total structure factor**

$\mathbf{F_A}$ **Marker substructure structure factor**

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

Additional material

Phasing equations

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

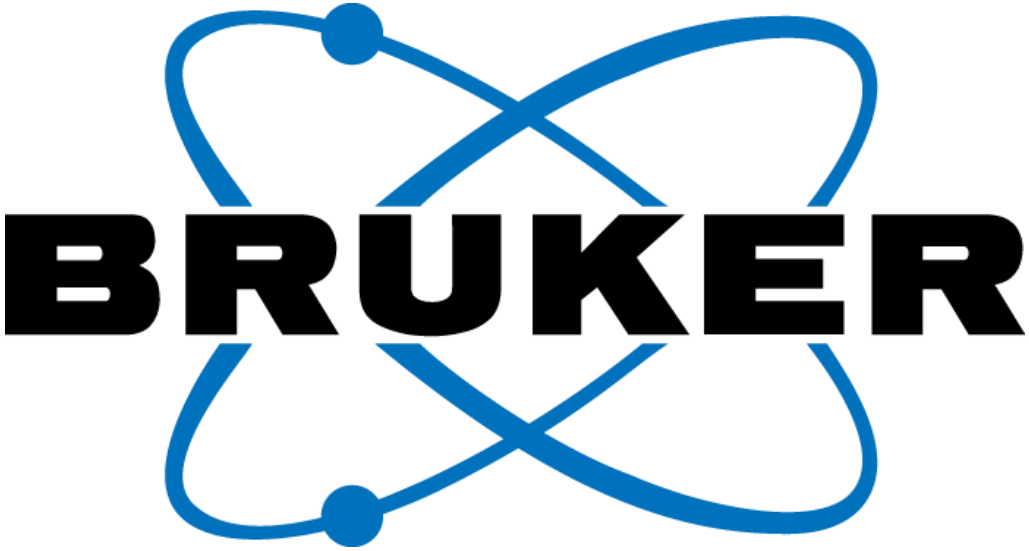
$$|F_{-h-k-l}|^2 = |F_T|^2 + a|F_A|^2 + b|F_T||F_A|\cos\alpha - c|F_T||F_A|\sin\alpha$$

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 $\boldsymbol{\varphi_T}$!



Innovation with Integrity