Experimental phasing basics

Andrea Thorn
**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**)

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.
Theory

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

Amplitude = $|F_{hkl}|$

Phase = $\phi_{hkl}$
Structure factors

$F_{hkl}$ = a wave

Amplitude = $|F_{hkl}|$

Phase = $\phi_{hkl}$
Structure factors

structure factor $F_{hkl}$

= a wave

= a complex number

Amplitude = $|F_{hkl}|$

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

Phase = $\phi_{hkl}$

cannot be measured... :-(

$\text{Im}$

$\text{Re}$
Structure factors

structure factor $F_{hkl}$

= a wave

= a complex number

Amplitude $= |F_{hkl}|$

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

Phase $= \phi_{hkl}$

cannot be measured... :-(

Im

Re
Structure factors

Amplitude = $|F_{hkl}|$

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

Phase = $\phi_{hkl}$

cannot be measured... :-(

PHASE PROBLEM

The central problem of crystallography
ANOMALOUS SCATTERING
The anomalous signal

Each structure factor is composed of contributions $f$ from each atom:
Friedel’s law: $|F_{hkl}| = |F_{-h-k-l}| \quad \phi_{hkl} = -\phi_{-h-k-l}$
The anomalous signal

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

\[ f = f_0 + f' + i f'' \]

- \( f' \) and \( f'' \) are observed near absorption edges of the atom’s element, and are \( \lambda \)-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
Direct methods

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

\[ |E_{hkl}|^2 = \frac{|F_{hkl}|^2 / \varepsilon}{\langle |F_{hkl}|^2 / \varepsilon \rangle} \]

\( \varepsilon \) scale factor for proper treatment of special position reflections

\( \langle |F_{hkl}|^2 / \varepsilon \rangle \) mean per resolution shell
Finding the substructure of marker atoms

- Direct methods
- Patterson methods

These methods require separate atomic electron densities to locate atoms.

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

Disulfides become 'supersulfurs'.
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.
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 to...

PHASING THE REST (SHELXC)
We can combine all contributions from marker atoms into $F_A$ and everything else into $F_P$.

\[ \alpha = \varphi_T - \varphi_A \]
\[ \varphi_A + \alpha = \varphi_T \]

$F_P$  Protein contribution
$F_A$  marker atom contribution
$F_T = F_P + F_A$
So, if we would know the anomalous scatterer positions (or heavy atom positions), we could calculate $F_A$:

\[
\alpha = \phi_T - \phi_A
\]

$\phi_A + \alpha = \phi_T$

If we could then get $\alpha$, we could calculate $\phi_T$ 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. \(|F_A|, |F_T|\) and \(\alpha\) 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

\[ |F_T| = 0.5 (|F_{hkl}| + |F_{-h-k-l}|) \]

and get

\[ |F_{hkl}| - |F_{-h-k-l}| = c |F_A| \sin \alpha \]

This is sufficient for the substructure and estimation of \( \varphi_T \)!
From substructure to structure

Protein contribution

\[ F_T \text{ (relates to } F_{hkl} \text{)} = F_P + F_A \]

\[ F_A = F_A + F_A' + F_A'' \]

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:

\[ |F_{hk\ell}| \text{ and } |F_{-h-k-l}| \]

\[ |F_{hk\ell}| \gg |F_{-h-k-l}| \]
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 \( |F_{-h-k-l}| \)

\( \Rightarrow \alpha \text{ must be close to } 90^\circ! \)
From substructure to structure

If: \( |F_{hkl}| \ll |F_{-h-k-l}| \)

\[ \Rightarrow \alpha \text{ must be close to } 270^\circ ! \]

Reflections with the largest anomalous differences must be closest to \( \alpha = 90^\circ \) or \( \alpha = 270^\circ \).

As you can easily see, estimation is rough.
$|F_{hkl}| \approx |F_{-h-k-l}|$

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

$\Rightarrow \alpha$ must be close to $0^\circ$ or $180^\circ$
Density modification

• $\varphi_T$ can now be computed from the phasing equations!

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

Via Fourier synthesis, an initial map is gained.

• By $\sigma_A$ 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:

1. Draw a sphere around it.
2. Plot the electron density on the surface.
3. How much does it vary?

- Not much: Flip density!
- A lot: Sharpen density!
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.
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.

$$f = f_0 + f' + if''$$
Data evaluation

• The general data quality should be good – multiplicity, completeness, \( R_{\text{PIM}} \) 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: \( CC_{\text{anom}}(1/2) \)
  – \( <d''/\sigma> \) and/or \( <d'/\sigma> \)
  – Anomalous correlation of data sets: \( CC_{\text{anom}} \)
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**: $\alpha$ 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**.

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


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

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