Experimental Phasing

Phenix Workshop
October 14, 2023
80th Pittsburgh Diffraction Conference

Tom Terwilliger
Los Alamos National Laboratory/New Mexico Consortium
Solving a structure with SAD phasing

1. Crystals with Se/Zn/…
2. Collect anomalous SAD data
3. Locate substructure
4. Phasing (calculate density map)
5. Density modification (improve map)
6. Model building
Solving a structure with SAD phasing (Se)

- Planning the experiment
- Automating the analysis
- Improving the map
- Building a model
Will I solve my SAD structure?

Planning the experiment

Automating the analysis

Improving the map

Building a model
Will I find the anomalous substructure?

- How many sites?
- Are sites ordered?
- Anomalous atom?
- Wavelength?
- Accurate data?
- How many reflections?
Key steps in SAD structure determination

1. Find the substructure
   - Anomalous signal $S$

2. Calculate an interpretable map
   - Anomalous correlation $CC^*_\text{ano}$
Anomalous correlation

- Correlation of anomalous differences with ideal
- Accuracy of anomalous data
- Accuracy of phasing
Anomalous signal

- Peak height in anomalous difference Fourier
- "Information per site" (can we find each site)
- Substructure likely to be found if $S > 10$
Will I find the anomalous substructure?

- How many sites?
- Are sites ordered?
- Anomalous atom?
- Wavelength?
- Accurate data?
- How many reflections?
Anomalous signal: information about each site
\textit{(peak height in anomalous difference Fourier)}

\[ \langle S_{\text{ano}} \rangle = CC^*_{\text{ano}} \cdot \frac{\sqrt{N_{\text{refl}}}}{\sqrt{n_{\text{sites}}}} \cdot \frac{1}{f^{1/2}} \]

- **Anomalous correlation**
- **Accuracy of the data**
- **Number of reflections**
- **Number of sites**
- **B-value for anomalous sub-structure**

**Anomalous signal S**

Will I find sites?
Will I solve my SAD structure?

- Planning the experiment
- Automating the analysis
- Improving the map
- Building a model
Why automate structure determination?

- Makes straightforward cases easier
- ... and difficult cases feasible for experts
- Speeds up the process
- Reduces errors
- Allows you to try more possibilities
Decision-making in automation

What does a good electron density map look like?

Using expected features of maps to make decisions and to improve maps
Decision-making in automation

Which map is better?
Histograms of density have positive skew

Typical histogram of electron density

Low density: Points between atoms and in solvent region

High density: Points on top of atoms
Histograms of density have positive skew

Poor map (inverse hand)

Better map
Positive skew in good maps
Estimate map quality from skew

Skew depends on map quality

Estimate map quality from skew

Skew=0.4

CC=0.6-0.7
Density modification

What does a good density map look like?

Use expected features of maps to improve map quality

Key feature of this process: improving density anywhere can improve it everywhere
X-ray density modification: “phase improvement”

Experimental Data

Initial phases

Noisy Map

Improved phases

Clear Map
1. We know a good map when we see it

2. Improvement anywhere improves the phases so there is improvement everywhere
Density modification

Noisy map

Identify local expected density

Clear map

Density everywhere is improved

Find phases consistent with experiment and expected density

Expected density can include...

Flat solvent

NCS

Density matches model

Histogram matching

Connectivity

Expected shapes
Automated model-building

Examples

- Shape-based identification of regular secondary structure
- Extension with short fragments from high-resolution structures
- Probabilistic sequence alignment
Finding regular protein structure
Extending with short fragments from PDB
Assembling best model
Identifying residue type at each position

| #  | G | A | S | V | I | L | M | C | F | Y | K | R | W | H | E | D | Q | N | P | T |
| 1  | 6 | 5 | 4 | 18| 18| 6 | 1 | 1 | 1 | 2 | 6 | 2 | 2 | 1 | 9 | 6 | 1 | 0 | 1 | 4 |
| 2  | 4 | 11| 14| 37| 5 | 2 | 0 | 2 | 0 | 0 | 2 | 3 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 6 |
| 3  | 11| 23| 5 | 12| 5 | 3 | 2 | 0 | 1 | 3 | 7 | 3 | 1 | 0 | 5 | 3 | 2 | 0 | 2 | 2 |
| 4  | 7 | 9 | 6 | 16| 8 | 5 | 2 | 0 | 1 | 3 | 8 | 4 | 1 | 0 | 7 | 6 | 2 | 0 | 3 | 4 |
| 5  | 31| 7 | 3 | 7 | 4 | 2 | 1 | 0 | 1 | 3 | 5 | 4 | 1 | 0 | 6 | 2 | 2 | 0 | 11| 1 |
| 6  | 1 | 3 | 3 | 41| 14| 8 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 4 | 0 | 0 | 1 |
| 7  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15| 63| 1 | 0 | 17| 1 | 0 | 0 | 0 | 0 | 0 |
| 8  | 2 | 3 | 6 | 23| 10| 6 | 2 | 1 | 0 | 1 | 4 | 3 | 0 | 0 | 5 | 16| 1 | 0 | 1 | 6 |
| 9  | 96| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Inserting side chains based on sequence
Automated structure solution

Experimental data, sequence, anomalously-scattering atom, wavelength(s)

Find heavy-atom sites with direct methods or likelihood (HYSS)

Calculate phases (Phaser/Solve)

Improve phases, find NCS, build model (phase_and_build)

Decision to be made:
- Multiple solutions, different derivatives or wavelengths
- Alternative hands of space-group and substructure

phenix.autosol
Iterative map and model improvement

- Experimental data, sequence, phase information or starting model
- Model-building and refinement:
  - Resolve building
  - Secondary-structure only
  - Connect chains
  - Fit loops
  - Build outside model
- Density modification
An NIH/NIGMS funded Program Project

The Phenix Project

Lawrence Berkeley Laboratory
Paul Adams, Pavel Afonine, Dorothee Liebschner, Nigel Moriarty, Billy Poon, Oleg Sobolev, Christopher Schlicksup

University of Cambridge
Randy Read, Airlie McCoy, Rob Oeffner

Los Alamos National Laboratory
New Mexico Consortium
Tom Terwilliger, Li-Wei Hung

UTHealth
Matt Baker

Duke University
Jane Richardson, Vincent Chen, Michael Prisant, Christopher Williams

An NIH/NIGMS funded Program Project