Using SAD data in *Phaser*

Phasing and extended model completion
Diffraction with anomalous scatterers

- SAD: single-wavelength anomalous diffraction
Harker construction for SAD phasing

\[ H \pm F \]

\[ H - F \]

\[ H + F \]

\[ F^+_O \]

\[ F^-_O \]

\[ F^- \]

\[ F^+_H \]

\[ F^-_H \]

\[ F_H^+ + F_H^- \]

\[ F_P^+ \]

\[ F_P^- \]
Principle of maximum likelihood

• How consistent is the model with the data?

• *What is the probability that the data would be measured if the model were correct?*

\[ L = p(data; model) \]

• Optimise model by adjusting parameters in probability distribution
Illustration of likelihood

- Generate data randomly from Gaussian distribution with $\mu=5$, $\sigma=1$
Illustration of likelihood

- With incorrect mean, some points become improbable
Illustration of likelihood

- With incorrect standard deviation, some points become improbable
Illustration of likelihood

- Plot likelihood as function of mean and sigma
Least squares and likelihood

- Most experiments have multiple sources of error: Gaussian error in observations
  - Central Limit Theorem
- Likelihood for Gaussians = least squares
Least-squares line fitting

\[ p(A_{j,obs}) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left[ -\frac{(A_{j,obs} - A_{j,calc})^2}{2\sigma^2} \right] \]

\[ -\ln \prod_j p(A_{j,obs}) \propto c + \sum_j \frac{(A_{j,obs} - A_{j,calc})^2}{\sigma^2} \]
Why not least squares in crystallography?

- Gaussian error for observations
- Error in predicting observation generally includes difference between structure factors
  - this is Gaussian in *phased* difference
  - *e.g.* $F$ vs. $F_C$ from model, $F_P$ vs. $F_{PH}$
- Phased error usually dominates
  - elimination of unknown phase changes probabilities
Applying likelihood to crystallography

- Find probability distribution for observations
  - start from structure factor probabilities
  - eliminate unknown phase angles
- Adjust parameters to optimise likelihood
- Applications:
  - calculating model phase probabilities
  - structure refinement
  - experimental phasing (isomorphous/anomalous)
  - likelihood-based molecular replacement
Multivariate complex normal distribution

- Complex normal

\[ p(z_1) = \frac{1}{\pi \Sigma} \exp \left[ -\frac{|z_1 - \langle z_1 \rangle|^2}{\Sigma} \right] \]

\[ = \frac{1}{\pi \Sigma} \exp \left[ -(z_1 - \langle z_1 \rangle)^* \Sigma^{-1} (z_1 - \langle z_1 \rangle) \right] \]

- Multivariate complex normal distribution

\[ p(z) = \frac{1}{|\pi \Sigma|} \exp \left[ -(z - \langle z \rangle)^H \Sigma^{-1} (z - \langle z \rangle) \right] \], where

elements of \( \Sigma \) given by \( \sigma_{ij} = \langle (z_i - \langle z_i \rangle)(z_j - \langle z_j \rangle)^* \rangle \)
SAD likelihood function

- Based on probability of $F^+$ and $F^-$ given model
  \[ p\left(F_0^+, F_0^-, H^+, H^-ight) \rightarrow p\left(F_0^+, F_0^-; H^+, H^-ight) \]
- Factor joint probability into two parts
  \[ p\left(F_0^+, F_0^-; H^+, H^-ight) = p\left(F_0^+; F_0^-, H^+, H^-ight) p\left(F_0^-; H^-ight) \]
- Integrate out unknown phases, $\alpha^+$ and $\alpha^-$
Intuitive understanding of SAD phasing

Expected value of $F^{-*}$ ($H^{-*}$)

$P\left(F^{-1}, \alpha^{-1}; H^{-*}\right)$

Expected difference between $F^{+}$ and $F^{-*}$

$P\left(F^{+1}; F^{-1}, H^{+}, H^{-*}\right)$
Intuitive understanding of SAD phasing

Total likelihood is integral of the product of the two distributions under the black circle.

Expected value of $F^- \cdot (H^-)$

Expected difference between $F^+$ and $F^-$
Breakdown of Friedel’s law

• Friedel’s law breaks down for mixture of scatterers differing in real:anomalous ratio
SAD log-likelihood gradient (LLG) map

• Compute derivative of log-likelihood with respect to heavy atom structure factor
• Fourier transform gives map of where likelihood target would like to see changes in anomalous scatterer model
• Very sensitive to minor sites
  • picks up sites identified as water molecules in refined structures determined by halide soaks
• [http://www-structmed.cimr.cam.ac.uk/phaser/tutorial](http://www-structmed.cimr.cam.ac.uk/phaser/tutorial)
  • tutorial with data for lysozyme iodide soak
SAD with partial model

Expected value of $F^- (DF_C^*)$

Expected difference between $F^+$ and $F^-^*$

$P \left( F_O^-, \alpha_O^- \mid F_C^{-*} \right)$

$P \left( F_O^+ \mid F_O^-, F_C^+, F_C^{-*} \right)$
SAD with partial model

Expected difference between $F^+$ and $F^{-*}$

Expected value of $F^{-*} (DF_{C}^{-*})$

$|F_{O}^{-}|$
Combining MR and SAD

- CuKα data to 1.9Å on hen egg-white lysozyme
  - can’t find sulfurs with HySS or SHELXD
- Solve by MR with goat alpha-lactalbumin (40% identical)
- Use MR model as “substructure” for SAD
  - look for S atoms in LLG map (finds all 10 S, 5-9 Cl⁻)
  - phases automatically combine MR and SAD
- Automated fitting with density-modified map
- [http://www-structmed.cimr.cam.ac.uk/phaser/tutorial](http://www-structmed.cimr.cam.ac.uk/phaser/tutorial)
  - tutorial with these data
Iterative model-building with SAD

- Nitrate reductase structure (Natalie Strynadka)
  - integral membrane protein, 1976 residues
  - contains 21 Fe atoms, 1 Mo, 118 S, 5 P (146 total)
  - solved using combination of Fe-MAD, MIRAS

- Fe peak SAD data only
  - find 11 “Fe” sites with phenix.hyss
    - several are super-sites of Fe₄S₄ clusters
  - phase and complete adding Fe, Mo, S with Phaser
    - total of 57 sites: 20 Fe, 6 Mo, 31 S
    - superatoms are resolved, 51 of 57 are identified correctly
    - correct hand indicated by number of sites, LLG score
Iterative model-building and phasing

- Improve phases by density modification
- Build with ARP/wARP (or Resolve)
  - 1607 residues, 1368 docked in sequence
- LLG completion from ARP/wARP model
  - 105 sites, 92 correctly identified
- Repeat DM and ARP/wARP
  - 1813 residues, 1775 docked in sequence
Automation of SAD phasing

- Functions are all available from Python
  - used for SAD in AutoSolve wizard
  - can run from HAPPy (CCP4)
- Log-likelihood-gradient completion
  - look for one or several types of scatterer
    - start from MR model or partial substructure
  - analyse map to add sites, make atoms anisotropic
  - delete atoms that fade away
  - change atom type if occupancy far from one
  - repeat to convergence
Absolute scaling

• SAD target uses real (partial structure) scattering and anomalous scattering
  • best results if f” known precisely
  • helps to have data on absolute scale

• use BEST data from Sasha Popov
  • average intensities as function of resolution

• get Wilson B-factor, absolute scale
  • have to define composition of crystal
Practical aspects of SAD phasing in *Phaser*

- Provide information about cell content
  - sequence, molecular weight, percent solvent…
  - used to put data on absolute scale
    - occupancies are reasonably accurate
- Provide information about $f''$ values
  - wavelength (table lookup) or measured
  - refined by default if only one atom type
- Try both hands if uncertain
  - separate completion if mixture of atom types
SAD phasing in CCP4

- ccp4i interface has *Phaser SAD* phasing module
- Two modes:
  - “Single-wavelength anomalous dispersion (SAD)”
    - start from substructure of anomalous scatterers
    - can test both hands, complete with multiple scatterers
  - “SAD with molecular replacement partial structure”
    - start from substructure of non-anomalous scatterers
    - optionally include known anomalous scatterers
SAD phasing in Phenix

- Use AutoSolve wizard
  - GUI version prompts for necessary information
    - command-line version is faster
  - finds sites with Hyss
  - automatically uses *Phaser* for phasing if SAD data
  - tests both hands, chooses best hand
  - carries out Resolve density modification and model-building
Background information

  • plus papers cited here


• [http://www-structmed.cimr.cam.ac.uk/phaser](http://www-structmed.cimr.cam.ac.uk/phaser)

• [http://www-structmed.cimr.cam.ac.uk/Course](http://www-structmed.cimr.cam.ac.uk/Course)
Acknowledgments

• Molecular replacement
  • Airlie McCoy, Laurent Storoni, Gabor Bunkoczi, Rob Oeffner

• Experimental phasing
  • Raj Pannu, Airlie McCoy, Laurent Storoni

• PHENIX collaboration
  • Ralf Grosse-Kunstleve, Nigel Moriarty, Paul Adams
  • Tom Terwilliger