Experimental Phasing

Macromolecular Crystallography School

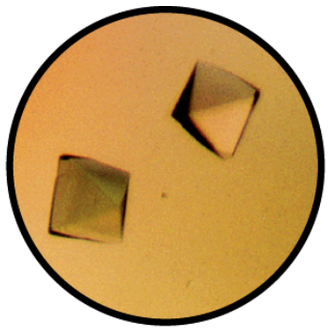
Madrid, May 2017

Paul Adams

Lawrence Berkeley Laboratory and

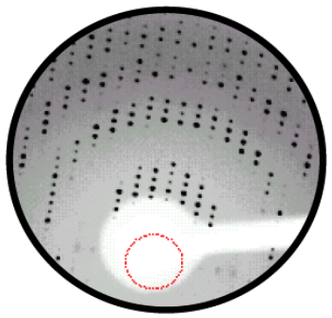
Department of Bioengineering UC Berkeley

The Crystallographic Process

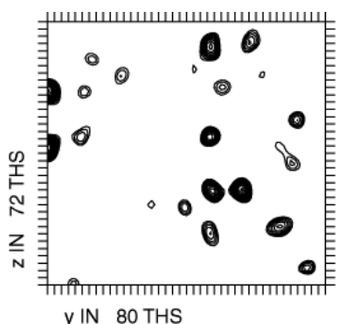


Crystallization

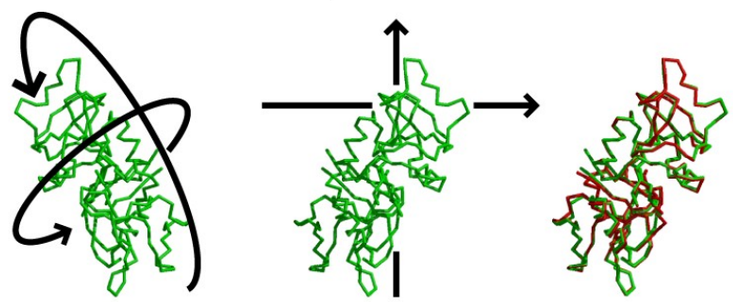
Data collection



Data processing



Anomalous scatterer

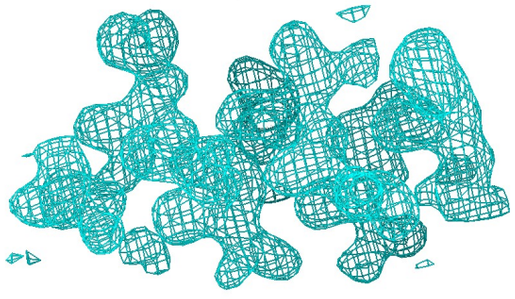


Molecular replacement

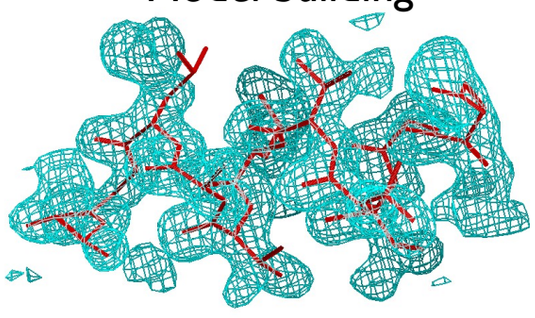
location

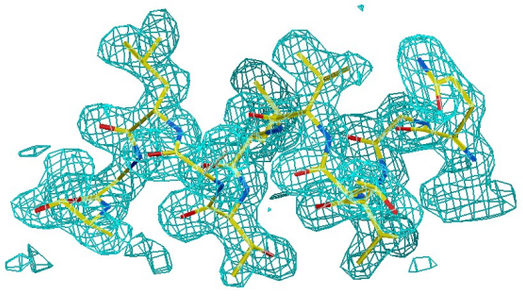
Phase determination

Phase improvement

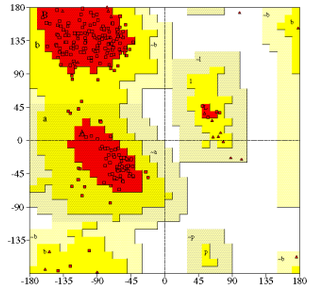


Model building





Model refinement



Validation

Phenix

The Phase Problem

- We want to get an image of the electron density in the crystal
  - Which we can then interpret to generate an atomic model
- The electron density = FT(F_{obs},ϕ)
  - But we can't measure the phase
- Therefore the phases need to be derived using amplitude information alone

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

What can we get from amplitudes?

- The Patterson function (only requires F_{obs}):
  - Gives a map containing all of the vectors between atoms
  - N atoms in the cell gives rise to N^2 peaks
  - For a small structure (~10 atoms)
    - A small number of peaks, atomic positions can be found from the vectors
  - For a macromolecular structure:
    - Many peaks (3000 atoms gives 9 million peaks), interpretation of the vectors in not possible
- Solution:
  - Make the macromolecular case more like a small molecule
  - Locate the positions of a small number of atoms (a substructure)
  - Leads to isomorphous replacement or anomalous diffraction methods

$$\Delta F_H = \left|F_H^{derivative}\right| - \left|F_H^{native}\right| \qquad \Delta F_H = \left|F_H^{+}\right| - \left|F_H^{-}\right|$$

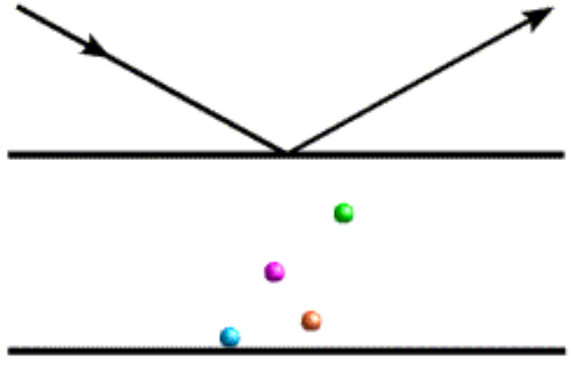
BERKELEY LAB

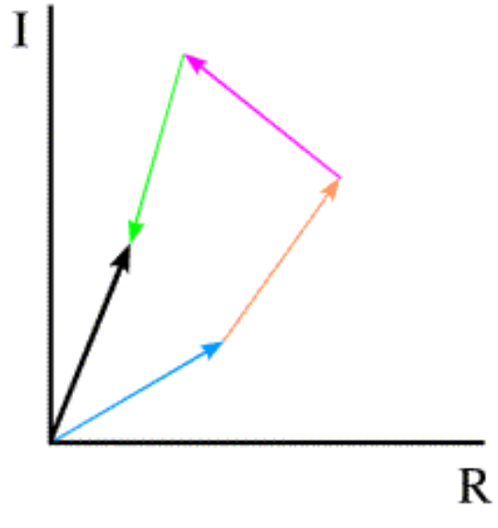
Phenix

Phasing Experiments

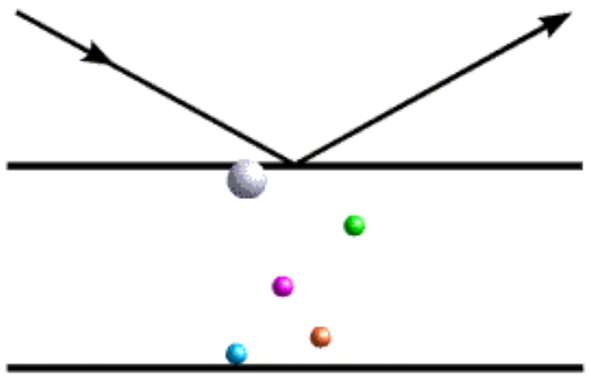
- Multiple isomorphous replacement (MIR)
  - Native data and data from at least 2 crystals soaked with heavy atom solution. Non-isomorphism limits phasing.
- MIR plus anomalous signal (MIRAS/SIRAS)
  - Native data and data from at least 1 crystal soaked with an anomalously scattering heavy atom. Non-isomorphism limits phasing.
- Multi-wavelength anomalous diffraction (MAD)
  - One crystal with an anomalous scatterer, data collected at different wavelengths. Requires a tunable X-ray source. Non-isomorphism is not a major problem (only 1 crystal).*
- Single isomorphous replacement (SIR)
  - Native data and data from 1 derivative soaked with heavy atom solution. Non-isomorphism limits phasing.
- Single wavelength (SAD)
  - One crystal with an anomalous scatterer, data collected at one wavelength with a high anomalous signal.

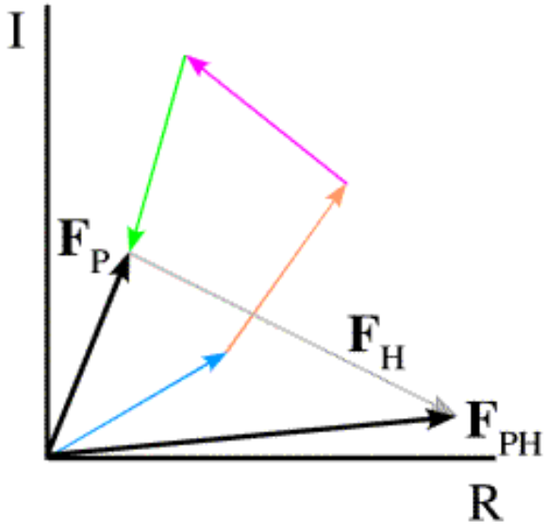
Isomorphous Differences





- Magnitude of differences
  - Can be large (20%+)
- Typically electron dense elements such as mercury, platinum, gold, uranium are used.
- The differences between sulphur and selenium are significant enough to solve a structure.





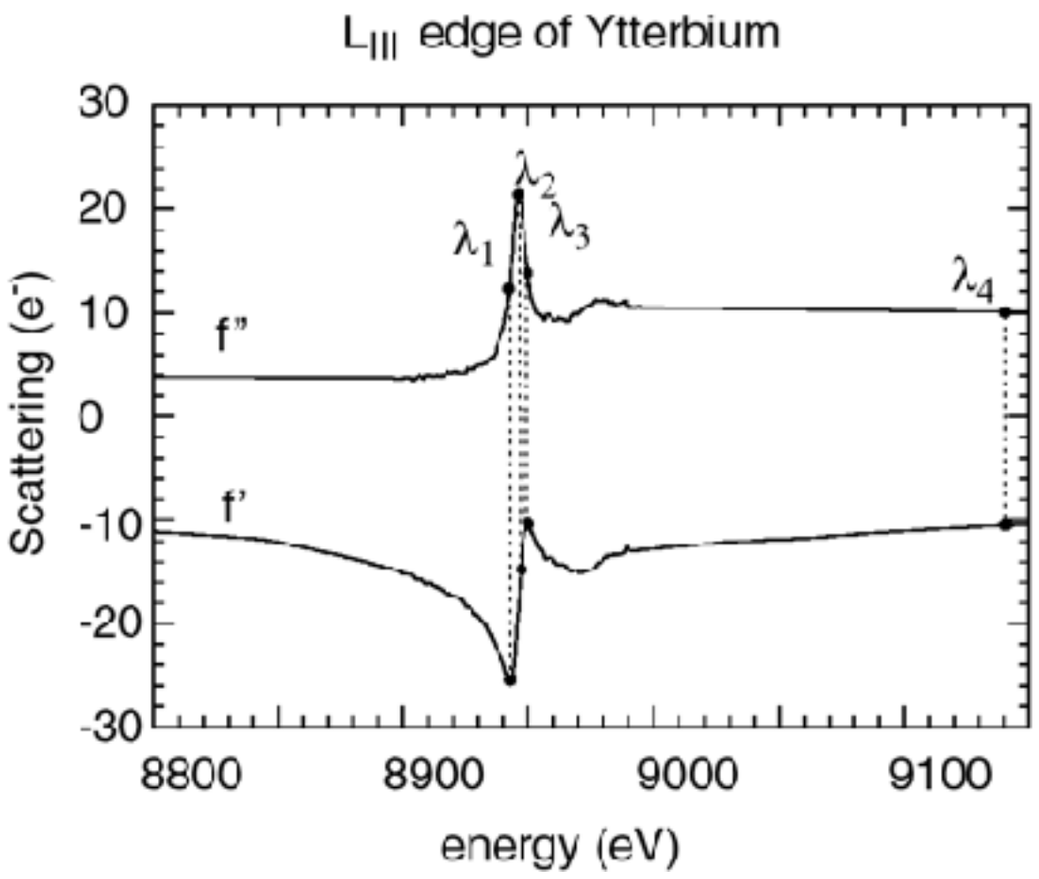
Images from Randy Read, Cambridge University

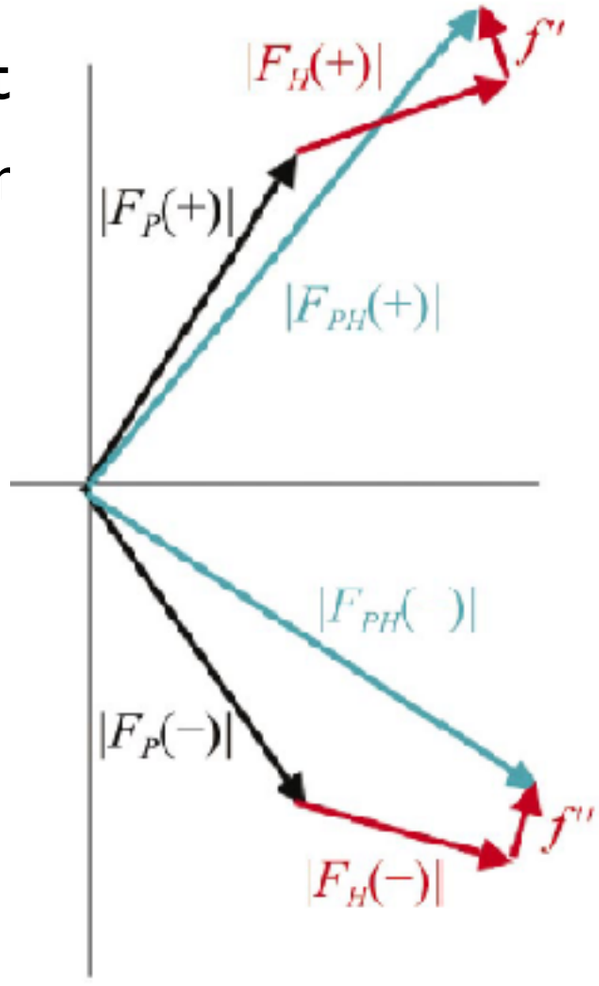
BERKELEY LAB

Phenix

Anomalous Scattering

- Is the result of a resonance effect for elements at
- Usually requires a tunable X-ray source (e.g. a syr





$$f = f^o + f' + i \cdot f''$$

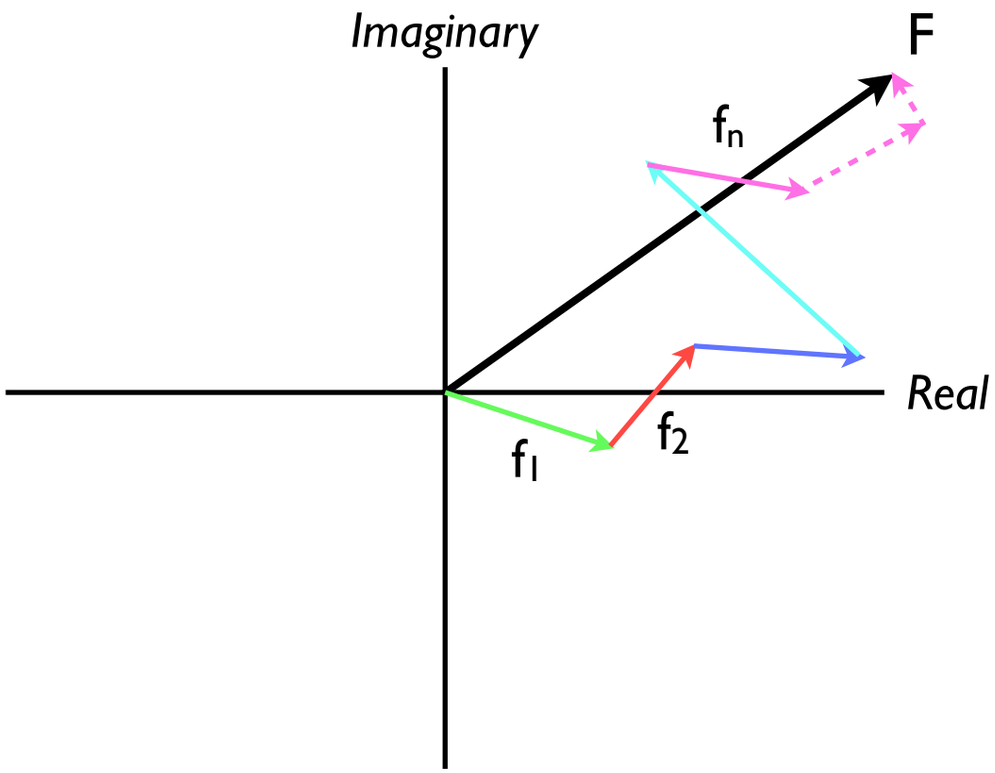
$$f^o(\sin\theta/\lambda) = \sum_{i=1}^{4} a_i \cdot e^{-b_i(\sin\theta/\lambda)^2} + c$$

BERKELEY LAB
Lawrence Berkeley National Laboratory

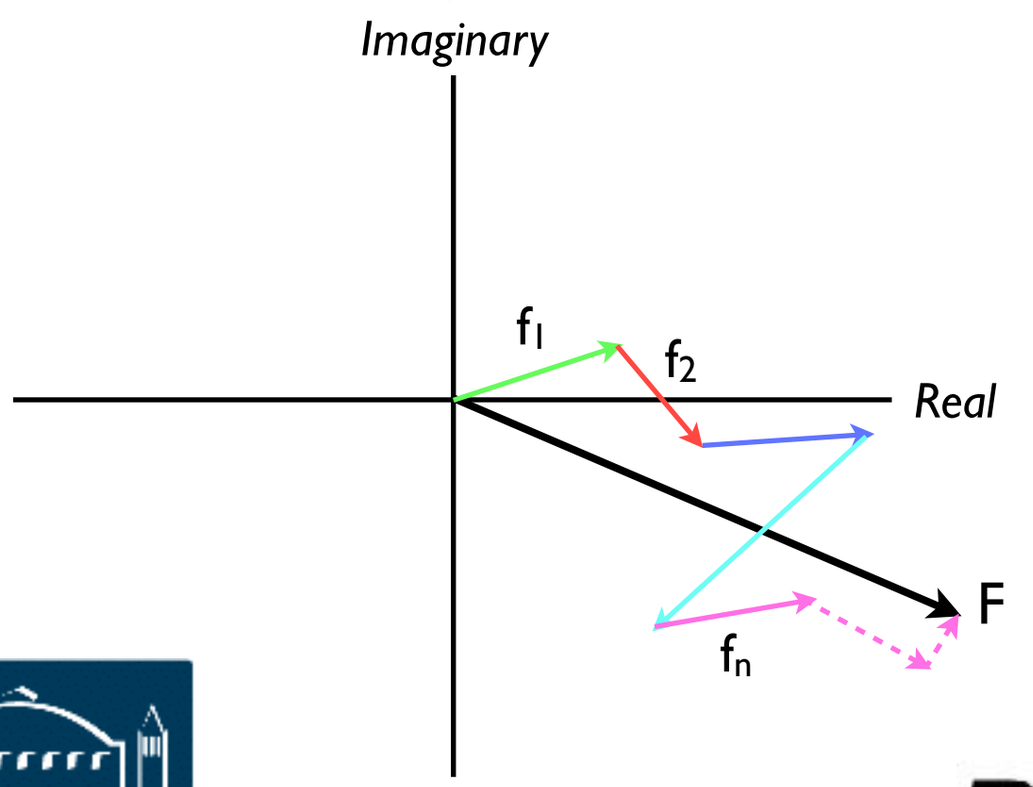
Phenix

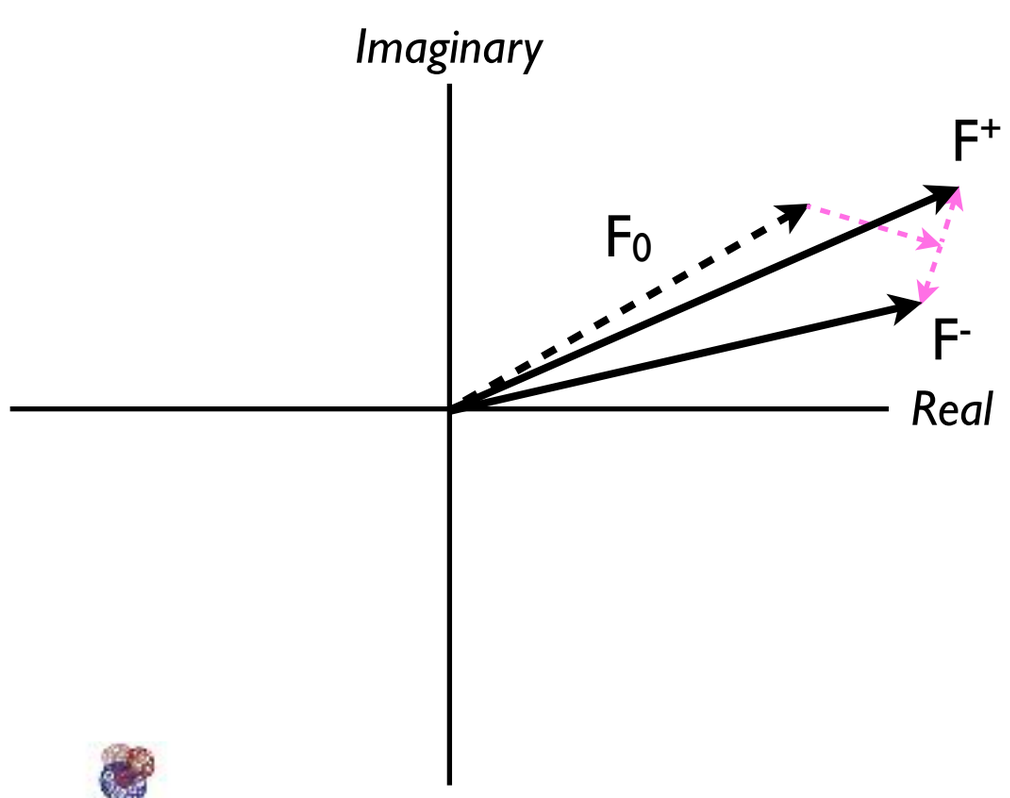
THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Anomalous Scattering



- The phase shift is always +90 (i.e. it does not obey Friedel's law)
- Therefore Friedel's law breaks down in the presence of anomalous scattering
- If we measure F_{hkl} and $F_{\bar{h}\bar{k}\bar{l}}$ we will find that they have different magnitudes (and phases)
- The two measurements are called F^+ and F^- and have an anomalous difference





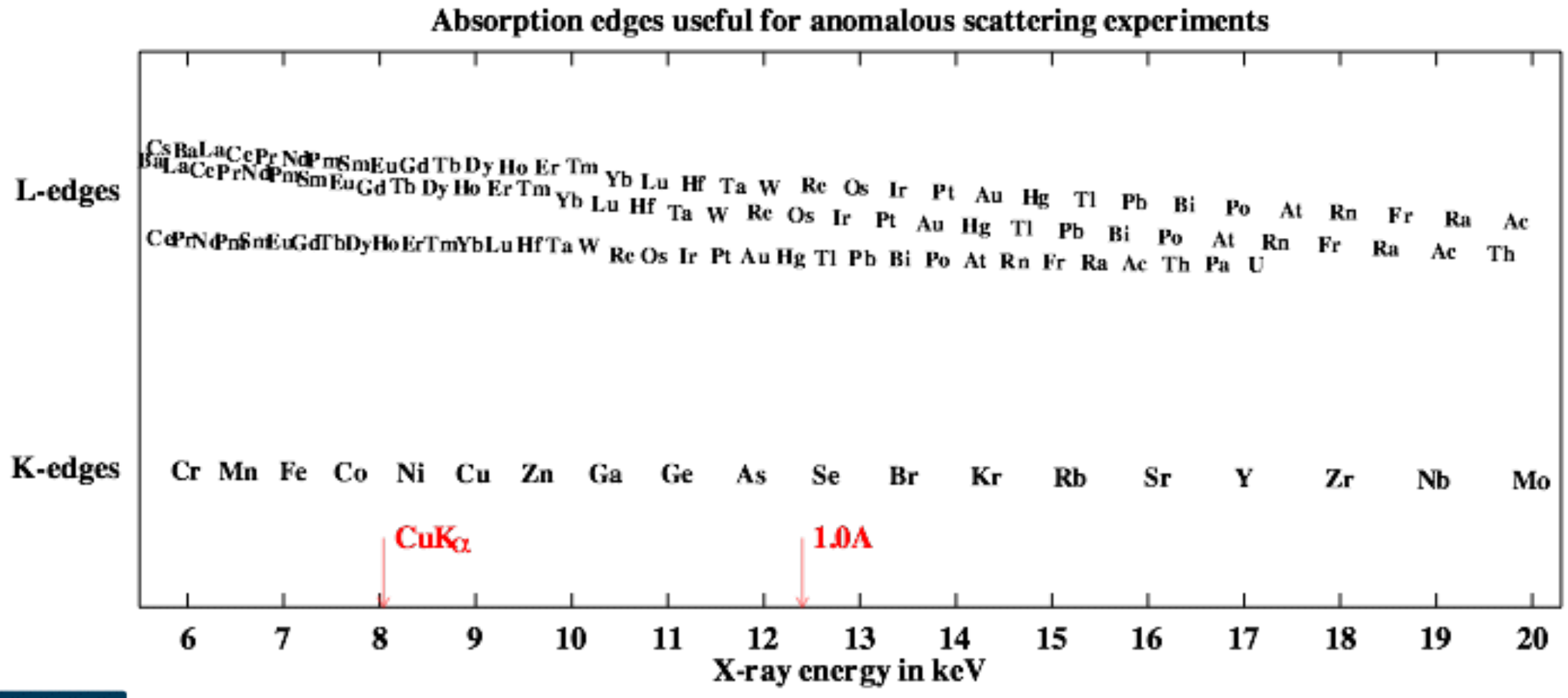
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Anomalously Scattering Atoms

- All atoms exhibit anomalous scattering
- Practically, not all edges are accessible at wavelengths routinely available and useful for crystallography (6keV to 17keV)
- Experiments do not have to be performed at the edge



skuld.bmsc.washington.edu/ scatter/AS_chart.html

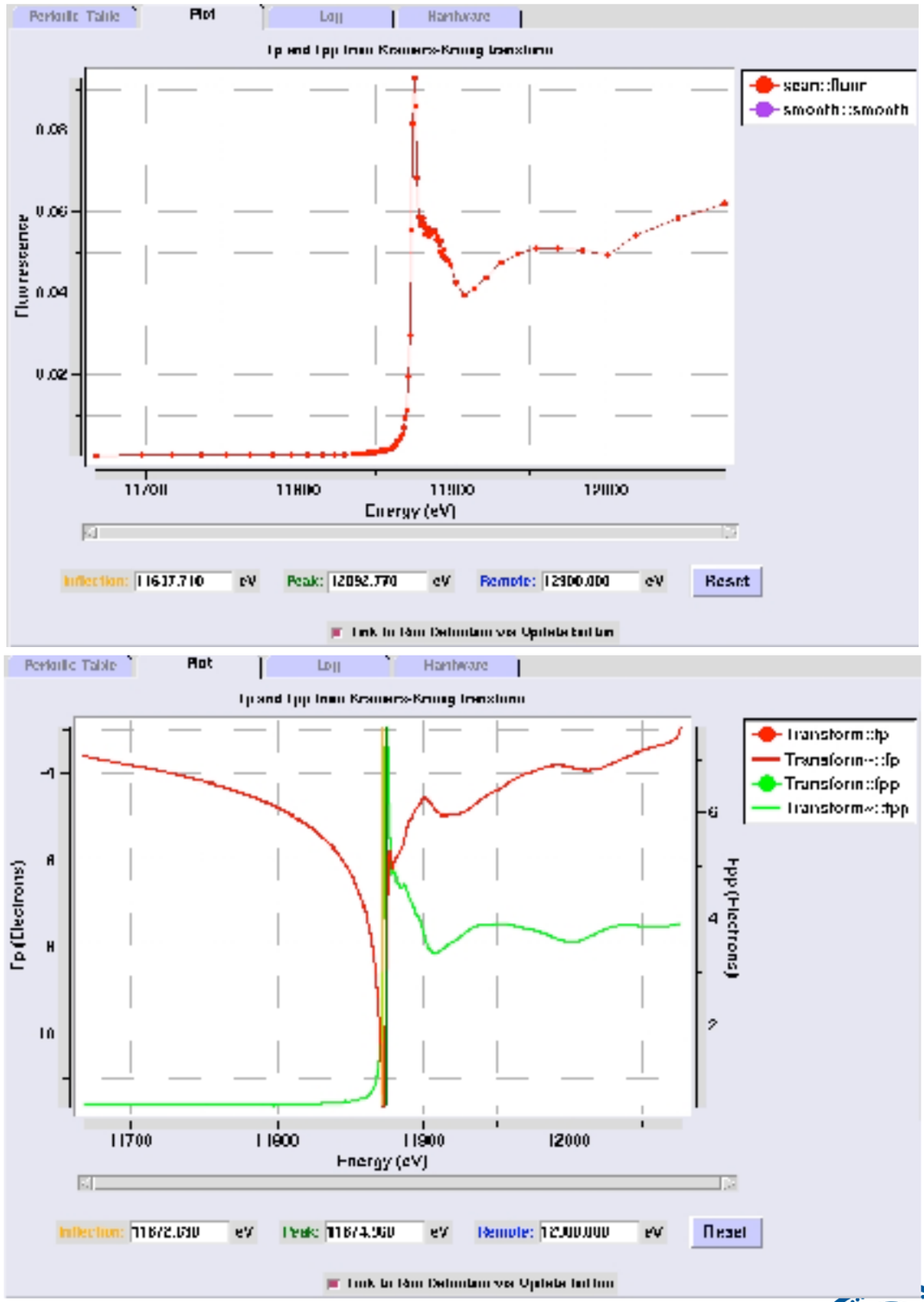
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Measuring Anomalous Scattering

- The fluorescence scattering increases rapidly near the absorption edge
- This can be measured using an X-ray fluorescence detector and varying the wavelength
- The anomalous scattering parameters, f' and f", can be obtained by calculating the first and second derivatives of the fluorescence curve (the Kramers-Kronig transform)
- The sharpness and features of the fluorescence scan will vary between elements



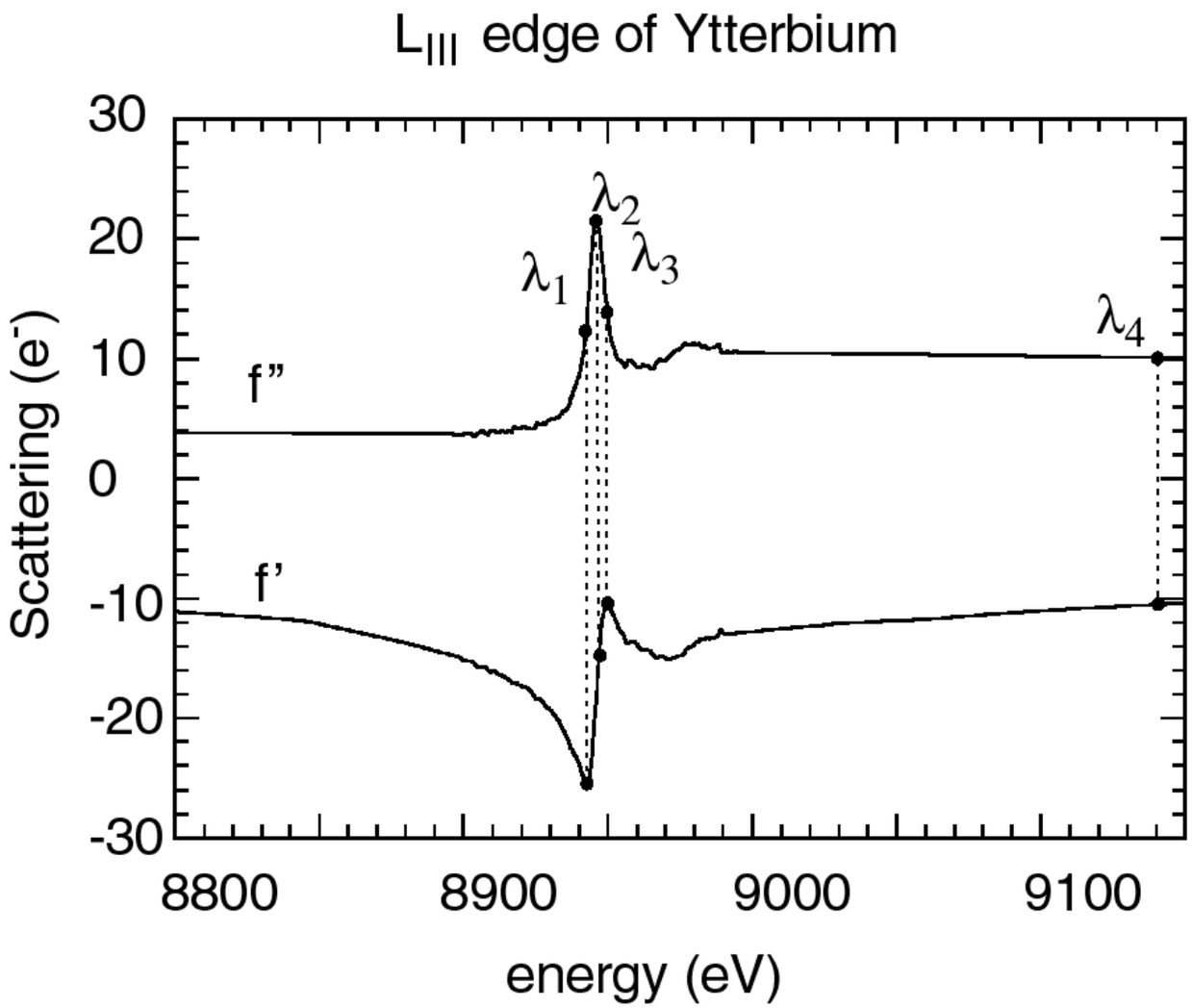
Macromolecular Crystallography Group, SSRL

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

MAD Data Collection



- Inflection point: maximizes f', has moderate f" contribution
- Peak: maximizes f", has low f' contribution
- High energy remote: has modest contribution for f' and f"
- Low energy remote: minimizes f' and f"

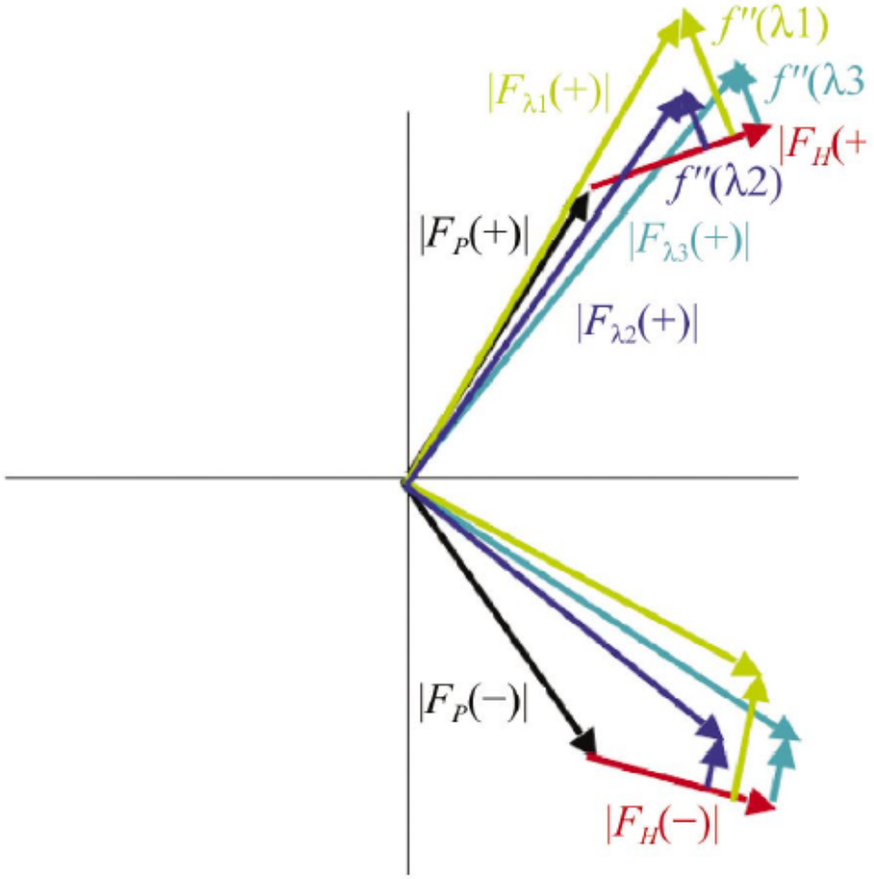
- There are multiple approaches to data collection
- How many wavelengths, which order, wedges?
- The maximal anomalous contribution (the peak) is also likely to be the wavelength with maximal radiation damage for the anomalous scatterer

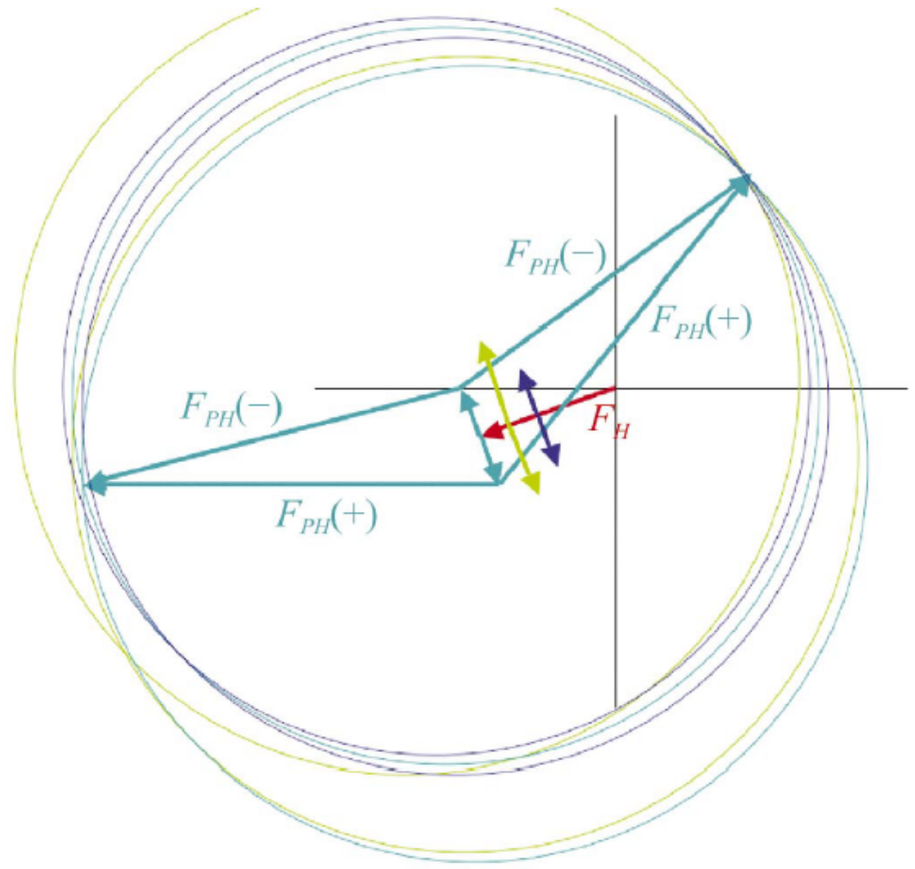
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Multi-wavelength Anomalous Diffraction





- Non-isomorphism not a significant problem
  - Except for radiation damage
- Correlated errors between wavelengths are a problem
  - c.f. several derivatives with the same substructure sites

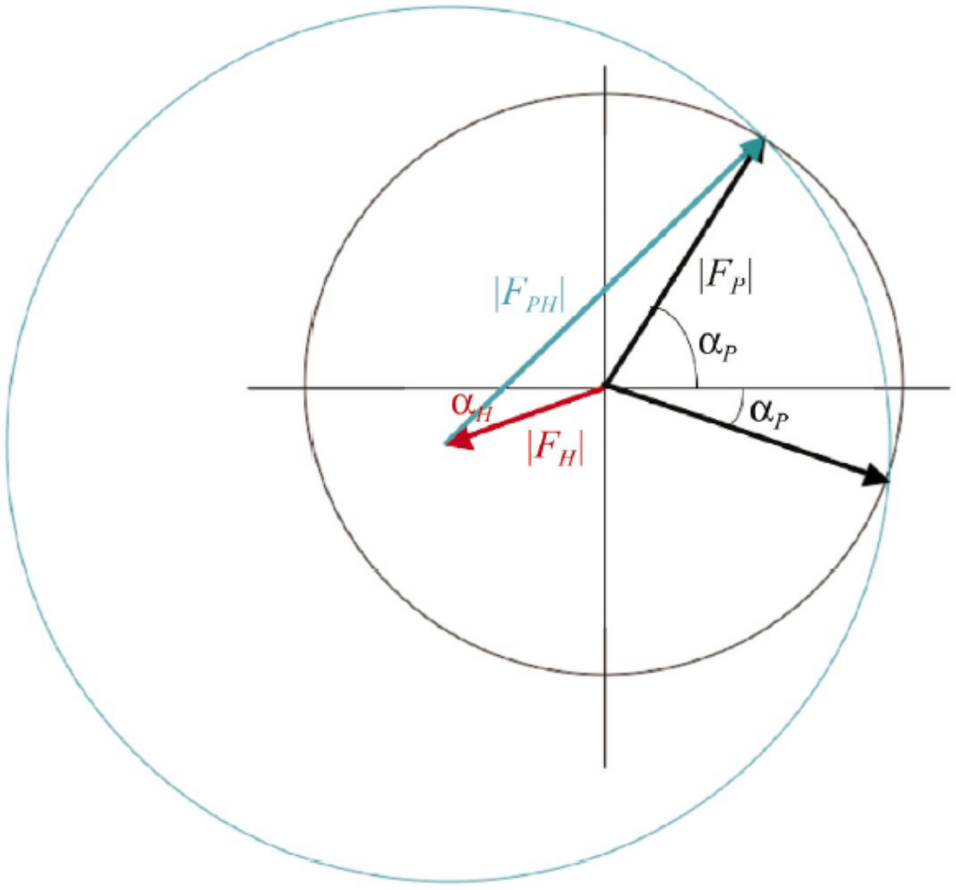
Images from G. Taylor, Acta Cryst. D, 59, 1881-1890 (2003)

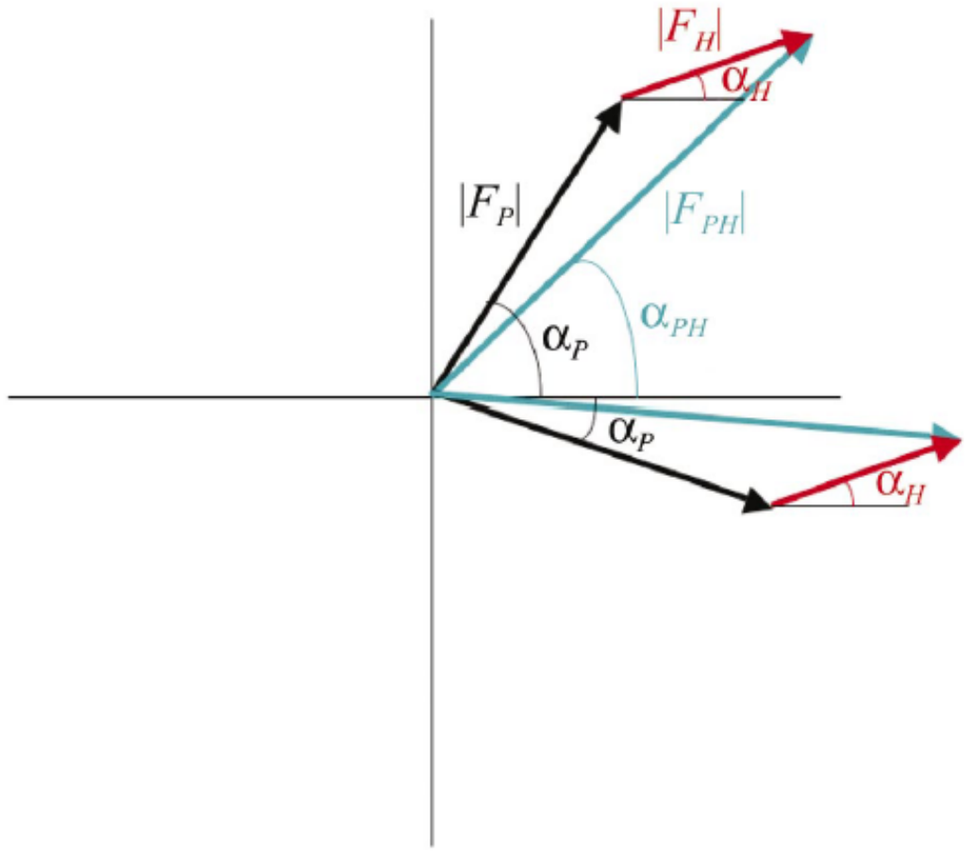
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Solutions to the Unknown Phase





- The agreement between the measurements and calculated information is greatest when the amplitude circles intersect.
- Note that if there are only two measurements there are two solutions.
- This assumes that there are no errors and that the amplitudes are such that the circles do intersect

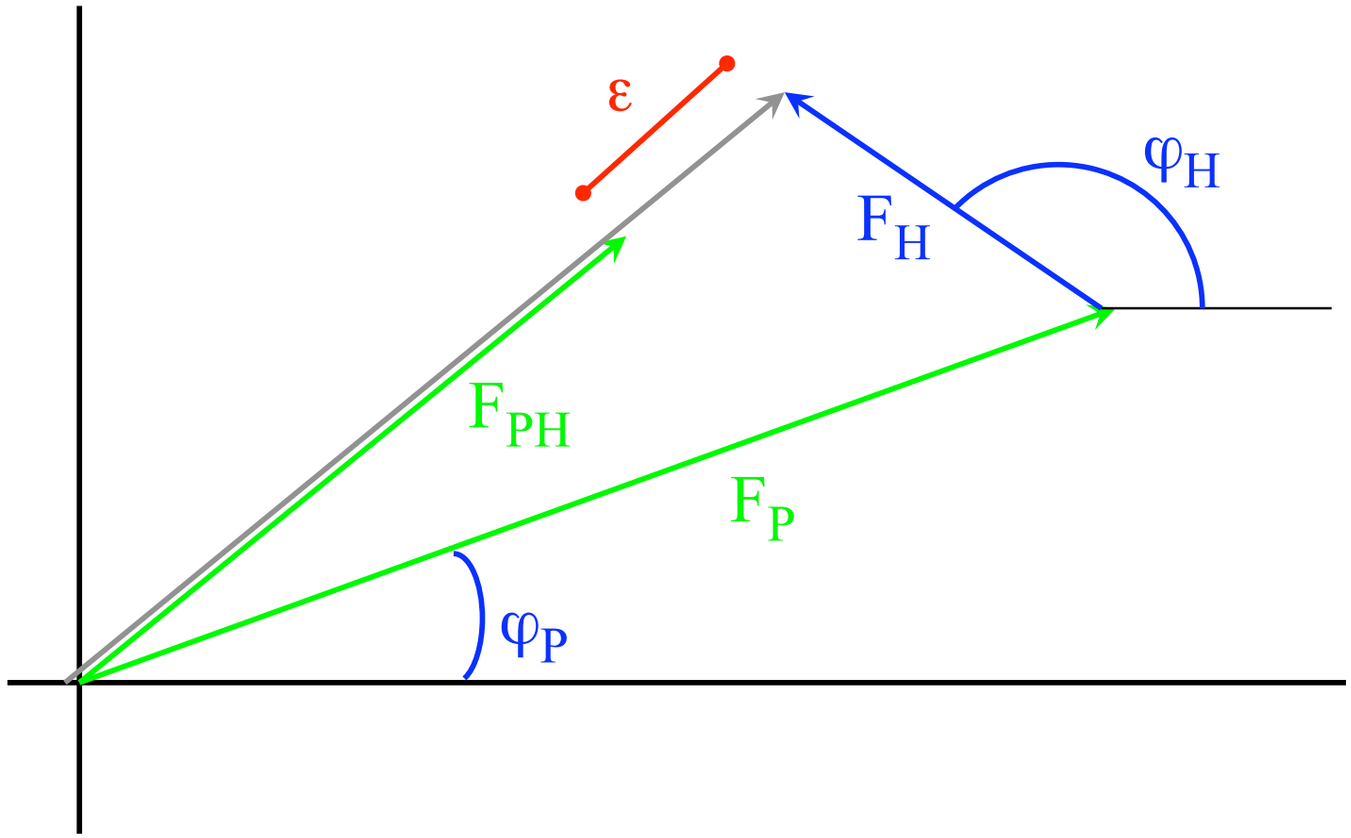
Images from G. Taylor, Acta Cryst. D, 59, 1881-1890 (2003)

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Goal of Phasing



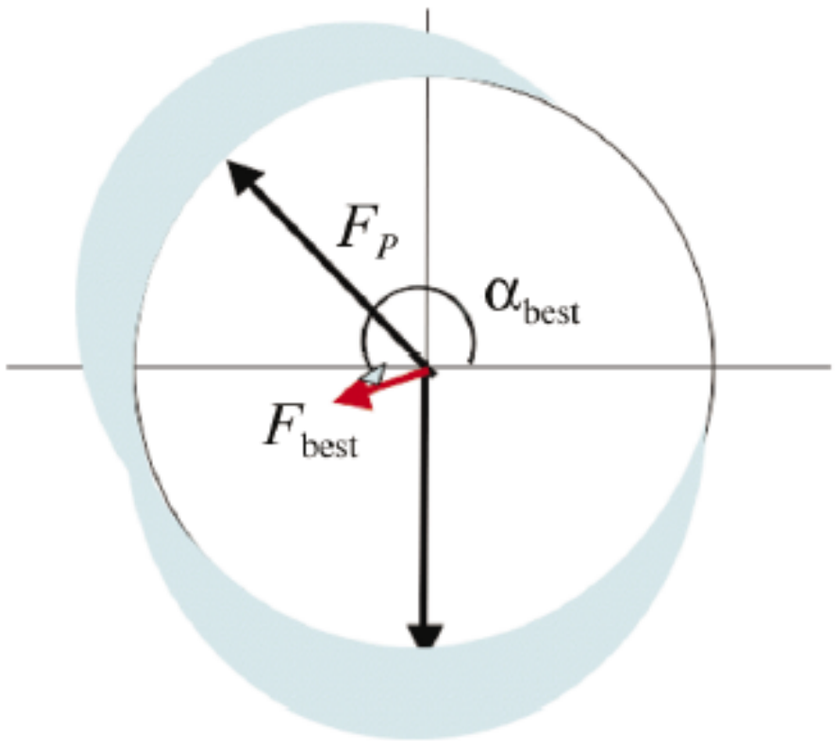
- The goal in phasing is to generate a set of phases that are consistent with the observed data and the heavy atom model
- The phases should minimize the lack-of-closure
- There are many observations and only a few model parameters
- However, there are many unknowns (phases)

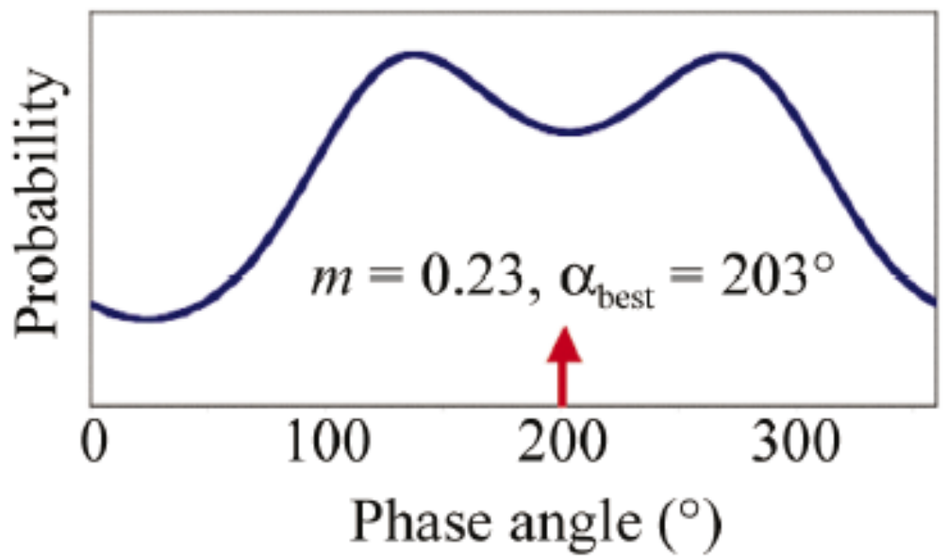
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Phase Probability Distributions





$$P(\alpha_P) \propto \exp(-\varepsilon^2/2E^2)$$

$$E = \langle [F_{PH(\text{obs})} - F_{PH(\text{calc})}]^2 \rangle$$

- The phase information is described by a phase probability distribution.
- This is calculated from the lack-of-closure at each phase angle.
- The best phase is defined as the centroid of the distribution.
- The figure-of-merit (FOM) describes the width of the distribution

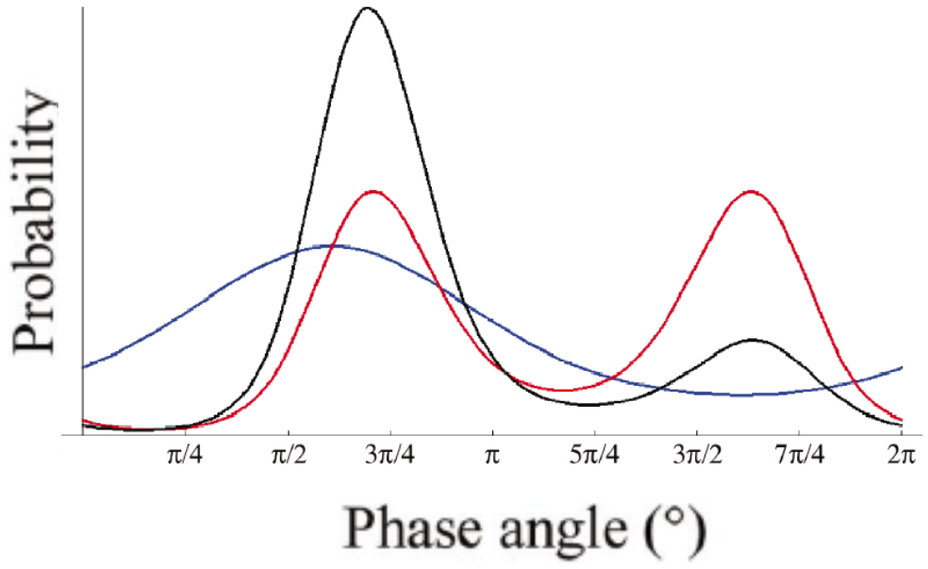
Images from G. Taylor, Acta Cryst. D, 59, 1881-1890 (2003)

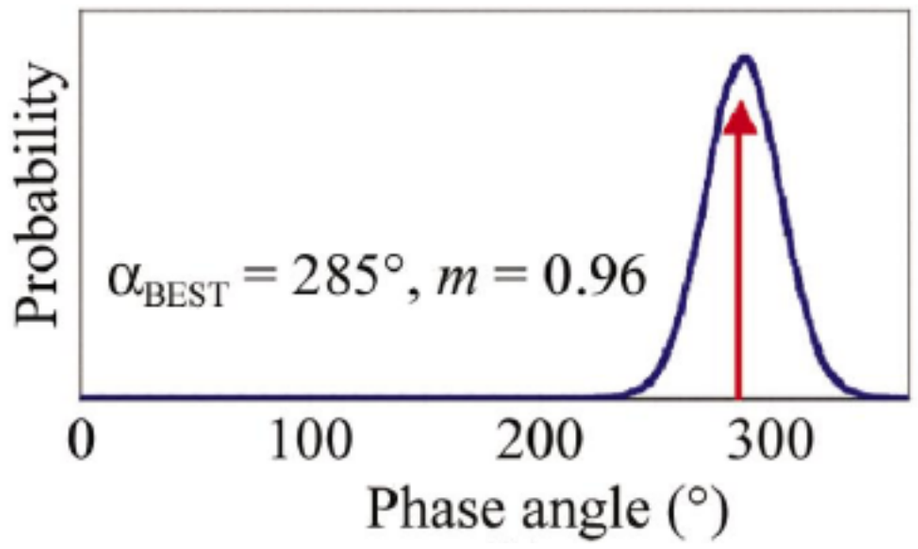
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Phase Probability Distributions





$$P(\alpha_P) \propto \prod_{i=1}^{\text{No. of derivatives}} \exp(-\varepsilon_i^2 / 2E_i^2)$$

- Phase probability distributions are typically represented with Hendrickson-Lattman coefficients (an approximation to Gaussians using sine/cosine terms – 4 in total).
- The probability distributions can be easily multiplied by simple mathematical operations on the HL coefficients.
- HL coefficients contain more information than a centroid phase and figure-of-merit.
- The contribution from the heavy atom model can be included

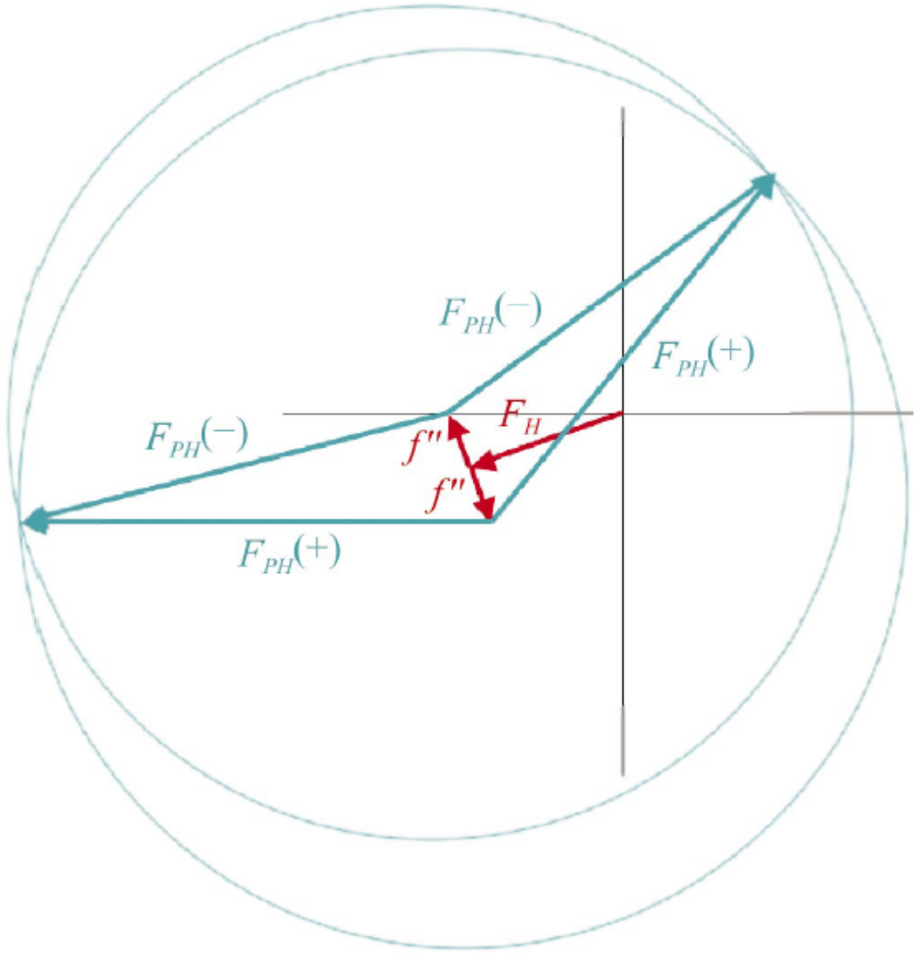
Images from G. Taylor, Acta Cryst. D, 59, 1881-1890 (2003)

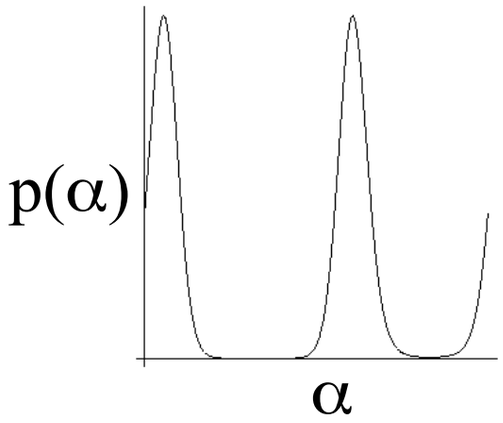
BERKELEY LAB
Lawrence Berkeley National Laboratory

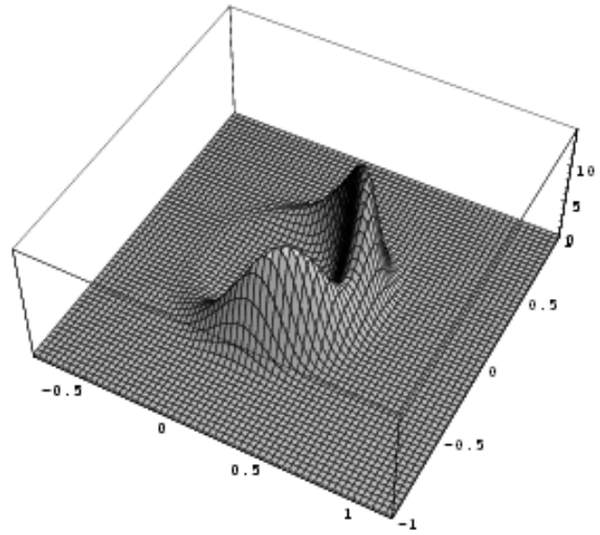
Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Single-wavelength Anomalous Data







- Single-anomalous diffraction is a special case of MAD
- Requires less wavelengths, but higher redundancy
- Has an implicit phase ambiguity, which needs to be resolved
- Is used to solve more than 50% of experimentally phased structures annually

Images from G. Taylor, Acta Cryst. D, 59, 1881-1890 (2003)

BERKELEY LAB
Lawrence Berkeley National Laboratory

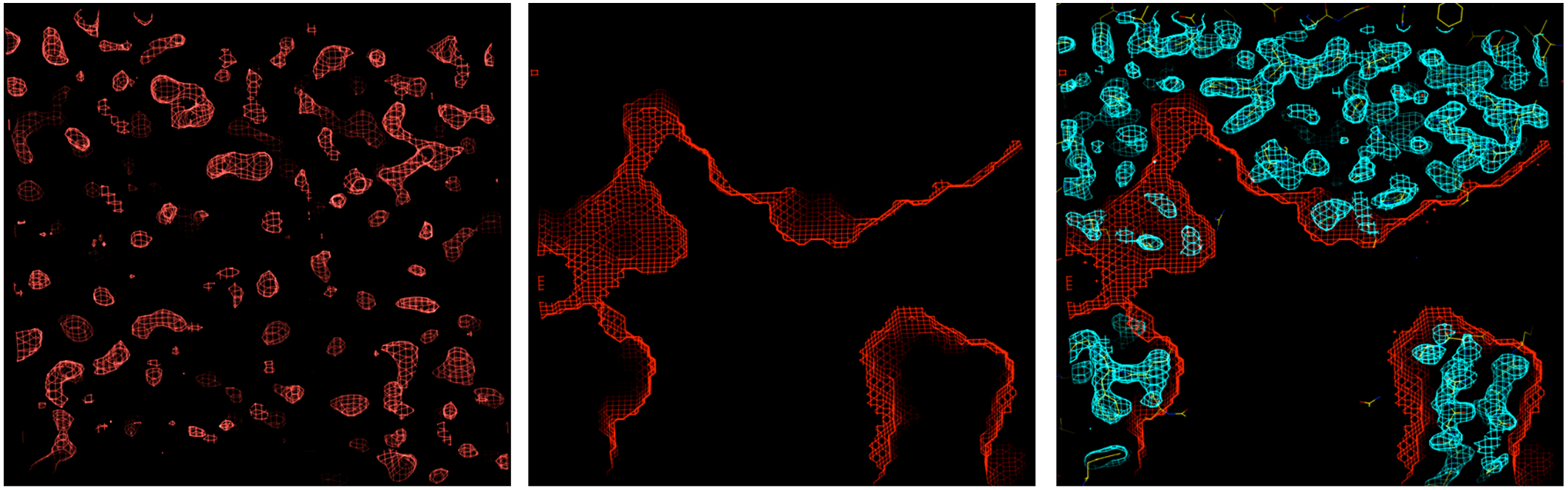
Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Resolving the Phase Ambiguity

- SAD phases are bimodal
- Centroid phases can be calculated
- The map produced is the superposition of the "correct" structure and noise
- The noise is removed by iterative filtering (density modification)

SAD Phases *Mask Generation* *Density Modification*



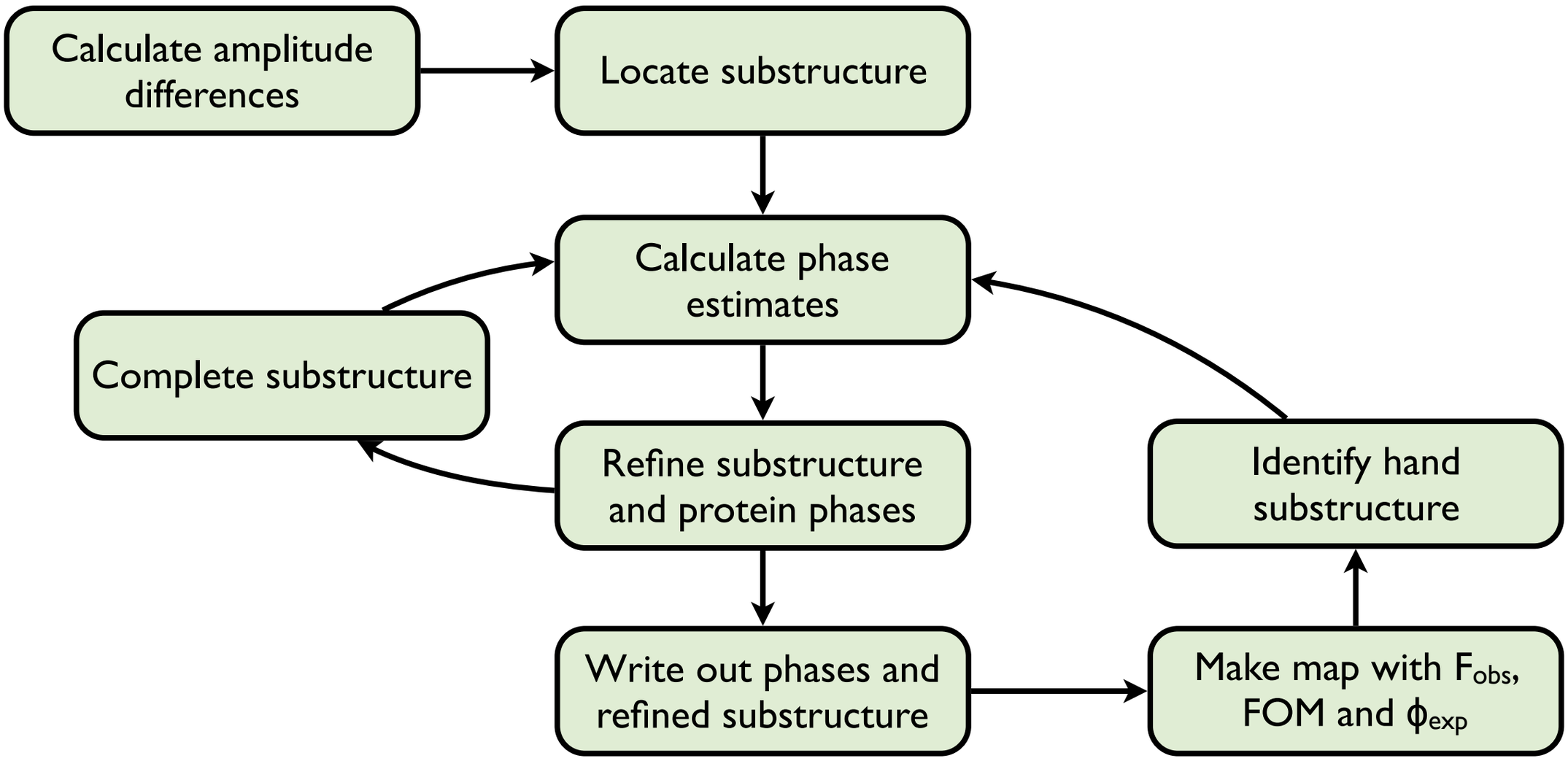
- *ISAS procedure: B.C.Wang, Methods in Enzymology, 1985*

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA · BERKELEY · 1868

Overview of Experimental Phasing



- Phasing typically relies on small differences between measured amplitudes

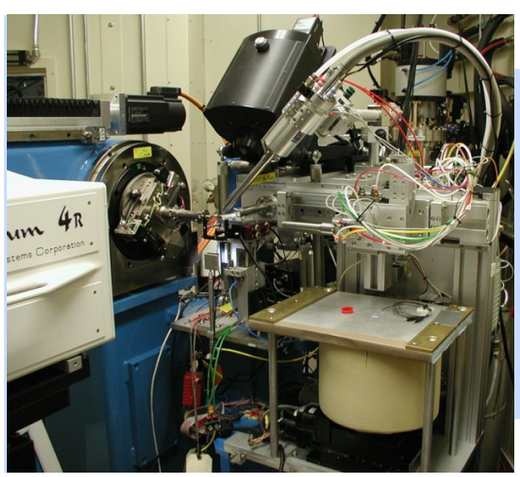
BERKELEY LAB
Lawrence Berkeley National Laboratory

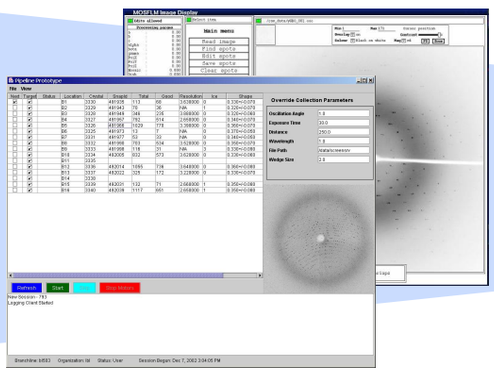
Phenix

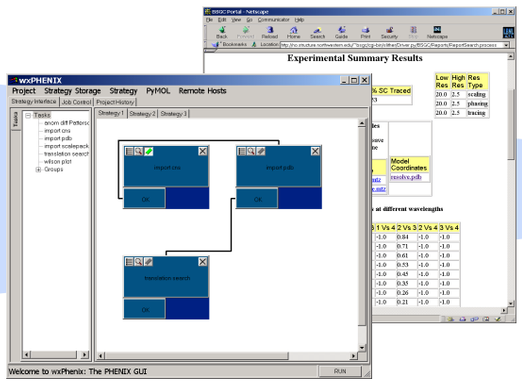
THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

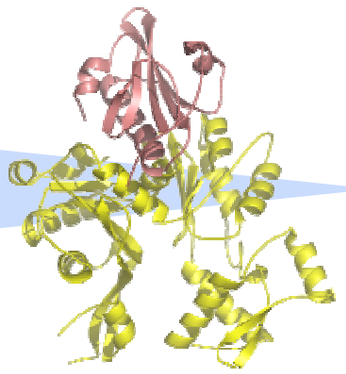
Automation

Data collection **Screening** **Data processing** **Data analysis** **Structure Solution**









- Automation can increase efficiency, and reduce human error
- Education becomes even more important

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

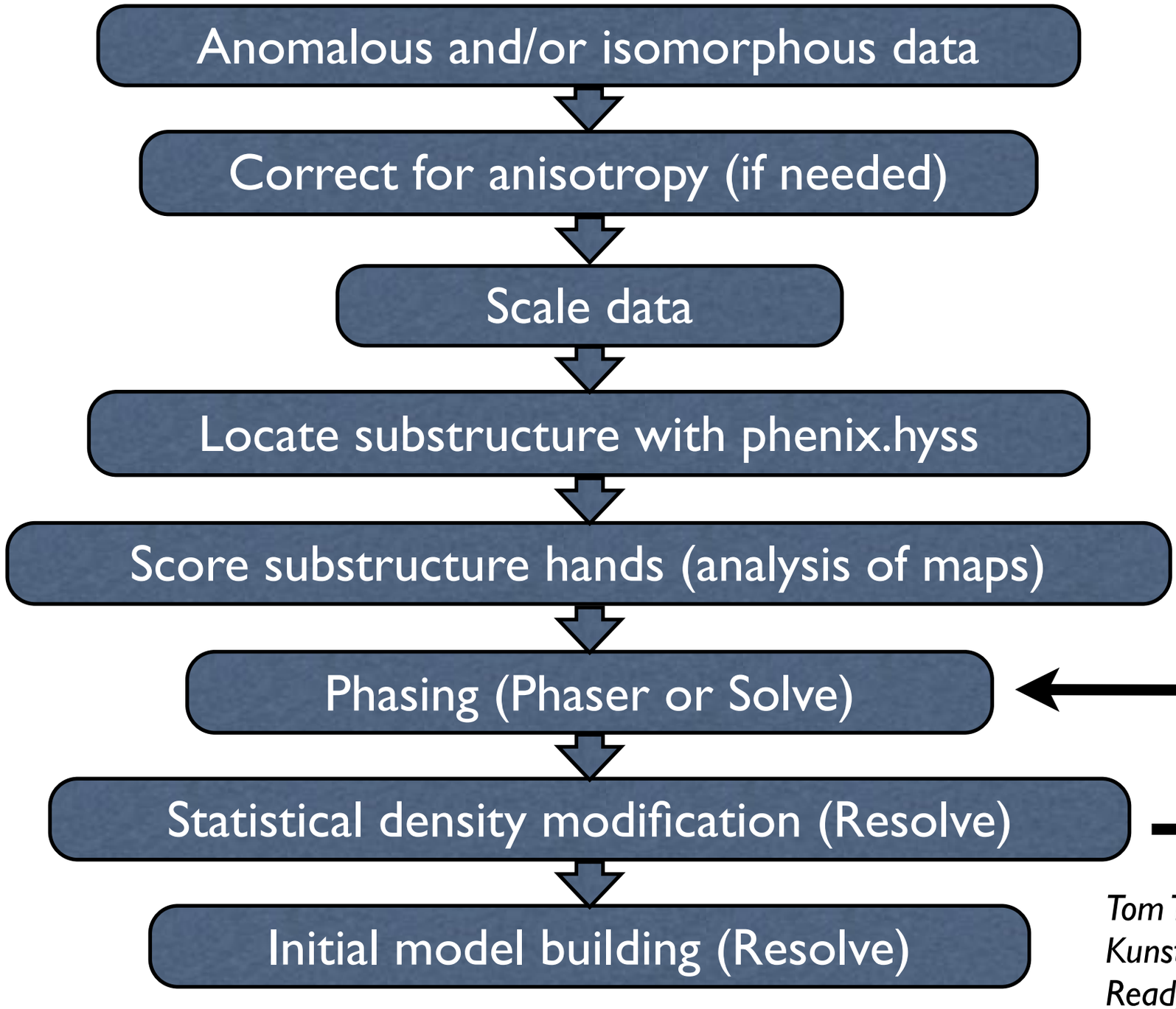
THE UNIVERSITY OF CALIFORNIA BERKELEY
LET THERE BE LIGHT
1868

Why Automation?

- Can speed up the process and can help reduce errors
- Software can try more possibilities than we are typically willing to bother with
- Makes difficult cases more feasible for experts
- Routine structure solution cases are accessible to a wider group of (structural) biologists
- Multiple trials or use of different parameters can be used to estimate uncertainties

- What is required:
  - Software carrying out individual steps
  - Integration between the steps (collaboration between developers)
  - Algorithms to decide which is best from a list of possible results
    - The computer has to make the decisions
  - Strategies for structure determination and decision-making

AutoSol Procedure



Tom Terwilliger, Ralf Grosse-Kunstleve, Airlie McCoy, Randy Read, Pavel Afonine

Terwilliger et al: Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Cryst. 2009, D65:582-601.

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

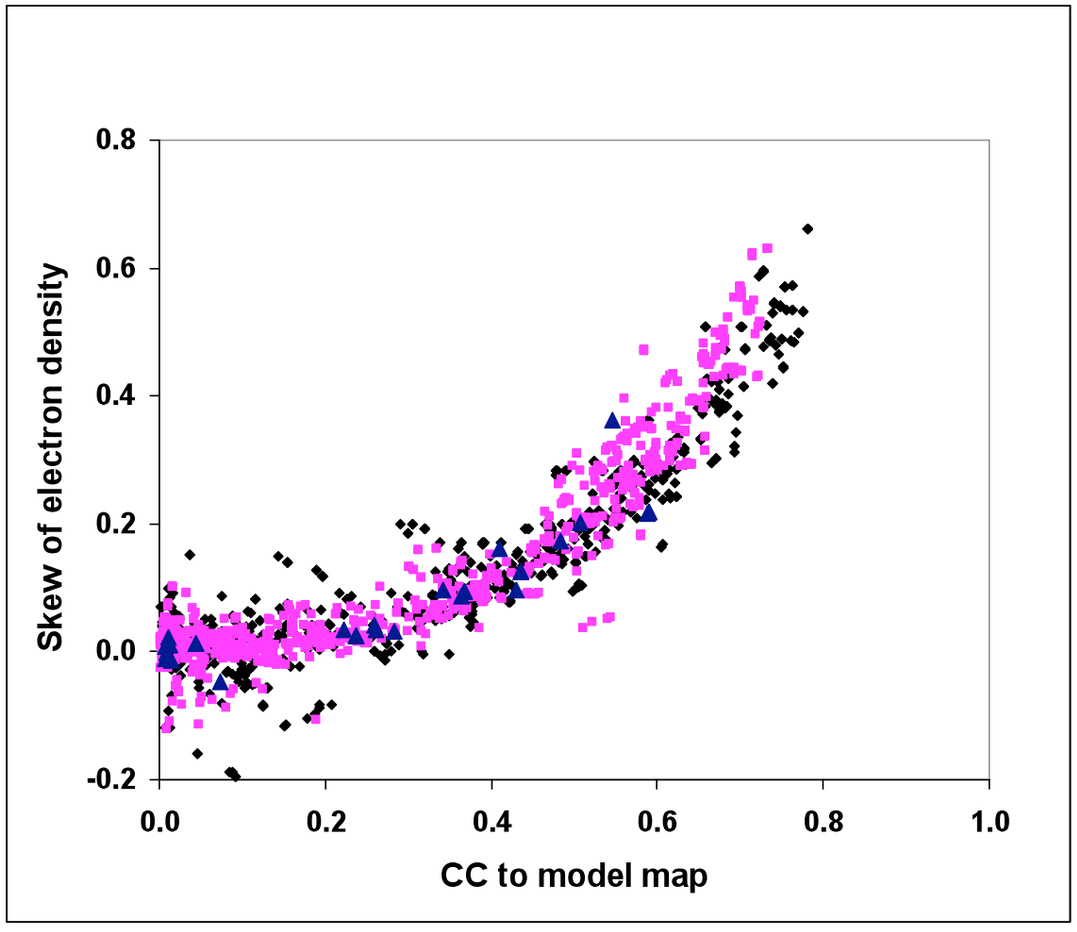
Automated Assessment of Map Quality

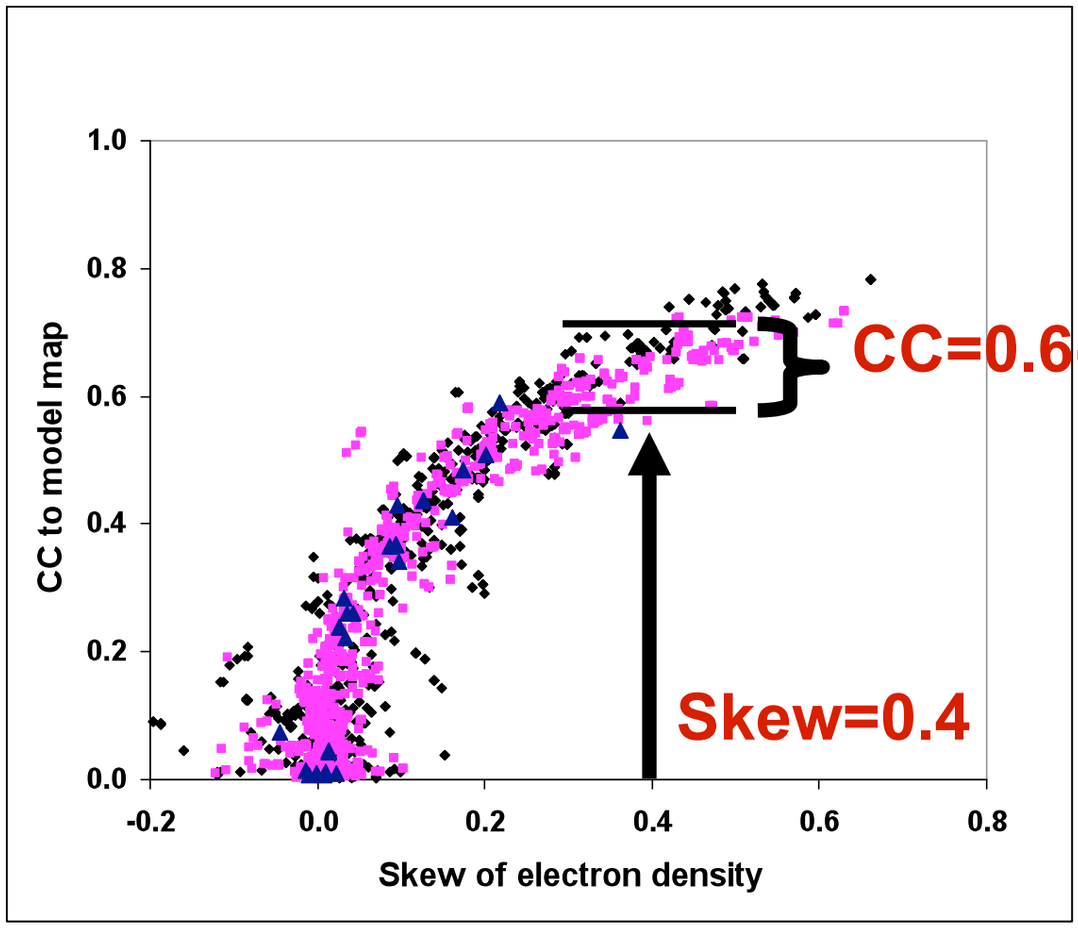
- 246 MAD, SAD, MIR datasets with final model available:
  - PHENIX library & JCSG publicly-available data
- Run AutoSol Wizard on each dataset
- Generate statistics for each solution considered:
  - Opposing hands, Additional sites, Inclusion of various derivatives for MIR

Tom Terwilliger, Los Alamos National Laboratory

Using Scores to Estimate Map Quality

- Measure skew of electron density map
- Calculate correlation of map to "correct" map
- Create lookup table to estimate correlation and standard deviation for any new map





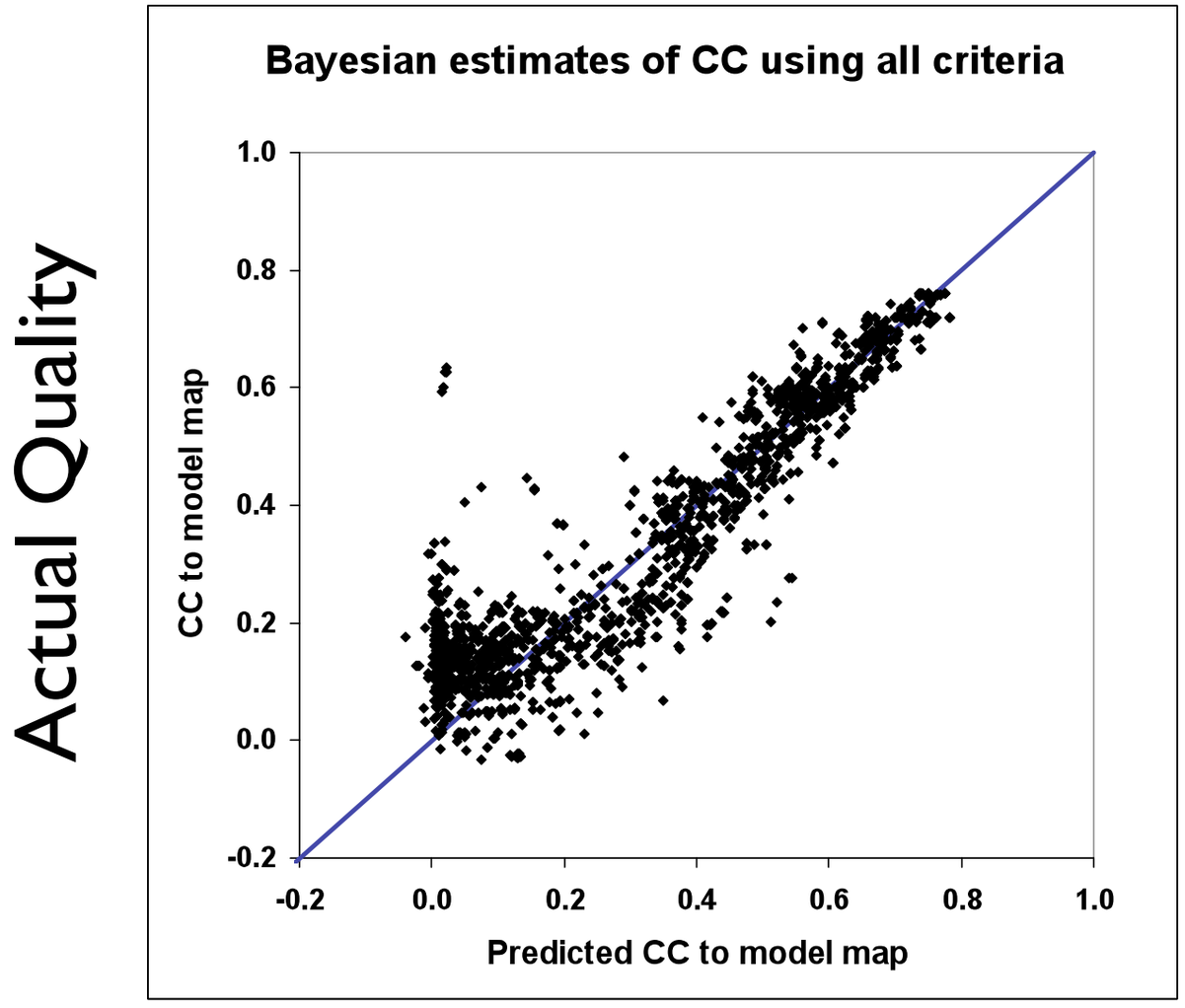
Tom Terwilliger, Los Alamos National Laboratory

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY
LET THERE BE LIGHT
1868

How Accurate are the Estimates of Quality?



Actual Quality

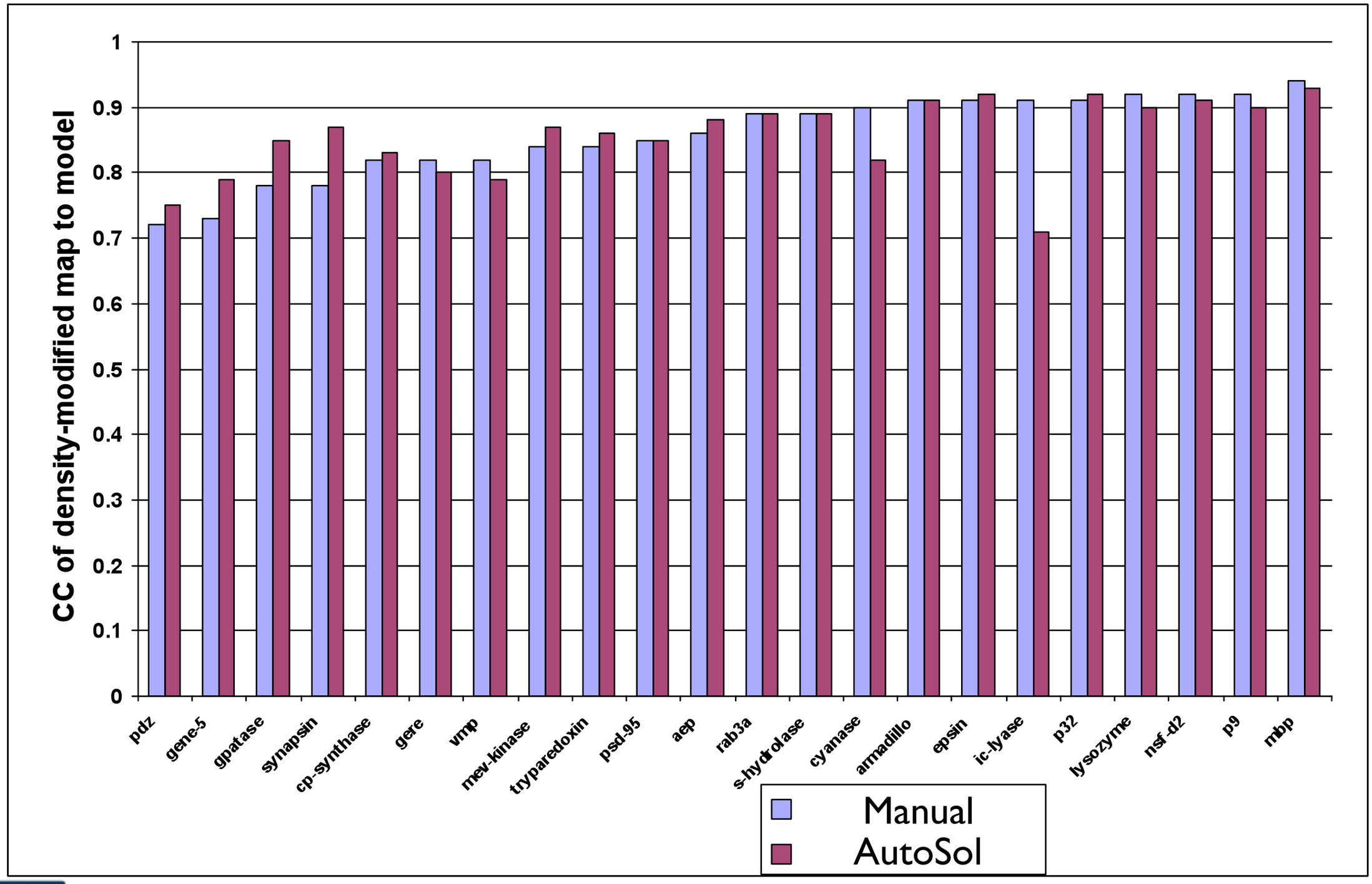
Estimated Quality

- By considering multiple scoring criteria it is possible to generate a reliable automated scoring mechanism

Tom Terwilliger, Los Alamos National Laboratory

Phenix

How Competitive is Automated Solution?



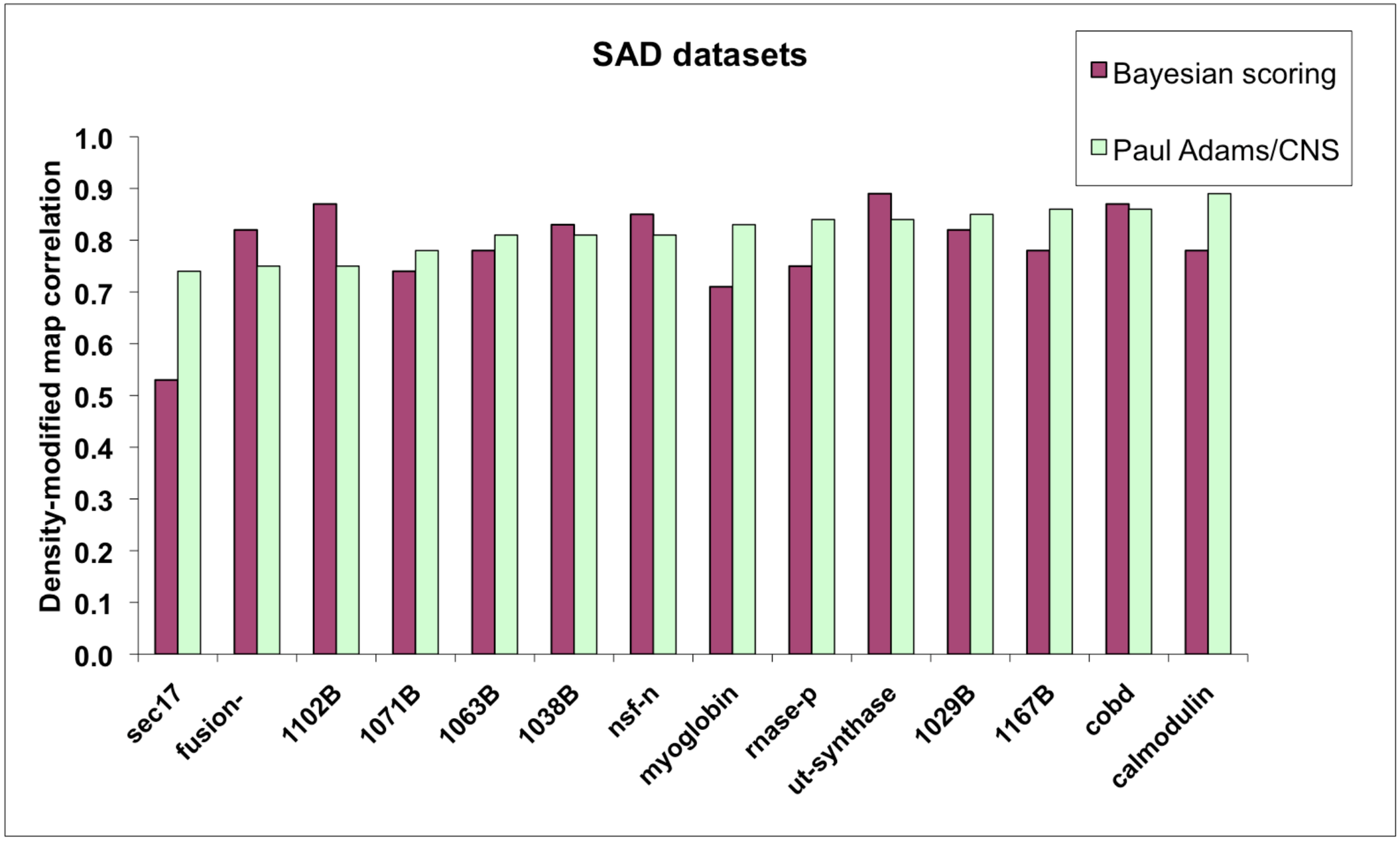
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Tom Terwilliger, Paul Adams

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

How Well Does This Work?



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Tom Terwilliger, Paul Adams

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Structure Solution with Weak SAD Signal

- Tom Terwilliger (Los Alamos National Laboratory)
- Gábor Bunkóczi, Airlie McCoy, Randy Read (Cambridge University)
- Nat Echols, Ralf Grosse-Kunstleve (Lawrence Berkeley Lab)

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Structure Solution from Weak Anomalous Data

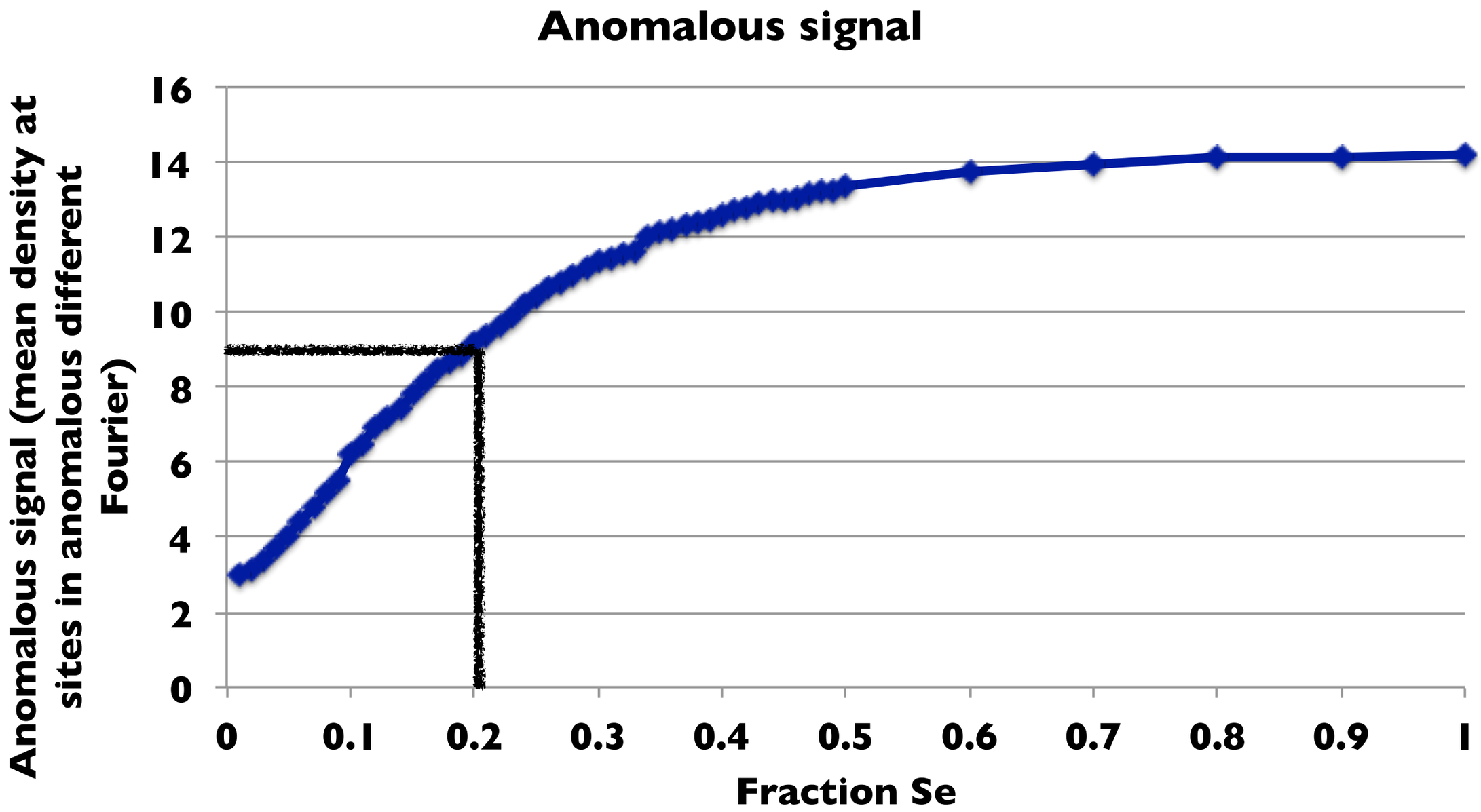
- Low anomalous signal-to-noise:
  - Few anomalous scatterers
  - Sulfur SAD
  - Weak diffraction
  - Wavelength far from peak
- Impact:
  - Substructure identification is difficult
  - Phasing is poor
  - Iterative density modification, model-building and refinement works poorly

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868 FIAT LUX

Anomalous Signal-to-noise



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Locating the Substructure

- Current approaches:
  - Anomalous Difference Patterson seeding
  - Direct methods (Rantan)
  - Dual-space methods (Shelxd, HySS, Crunch2, SnB)
  - Difference Fourier (Solve)

- Instead, most powerful source of information about the substructure before phases are known is the SAD likelihood function:
  - The likelihood of measuring the observed anomalous data given a partial model

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Using the SAD Likelihood Function

- Start with a guess about the anomalous sub-structure
  - From anomalous difference Patterson
  - Random
  - Any other source
- Find additional sites that increase the likelihood
  - Completion based on log-likelihood gradient maps*
  - Iterative addition of sites
  - Related to using a difference Fourier - but much better

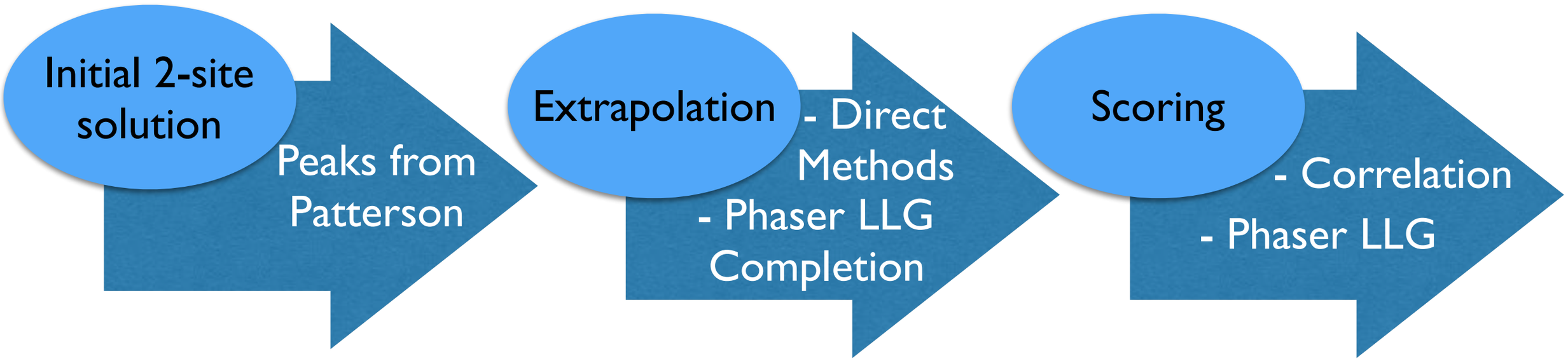
* *La Fortelle, E. de & Bricogne, G. (1997). Methods Enzymol. 276, 472-494*
McCoy, A. J. & Read, R. J. (2010). Acta Cryst. D66, 458-469.

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Making use of LLG in HySS



- Range of Resolutions
- Number of Patterson Peaks

- Adjust LLG Sigma (cutoff for peak height)

- Run quick direct methods first
- LLG scoring
- Terminate early if same solution found several times

Grosse-Kunstleve RW, Adams PD: Substructure search procedures for macromolecular structures. Acta Cryst. 2003, D59:1966-1973

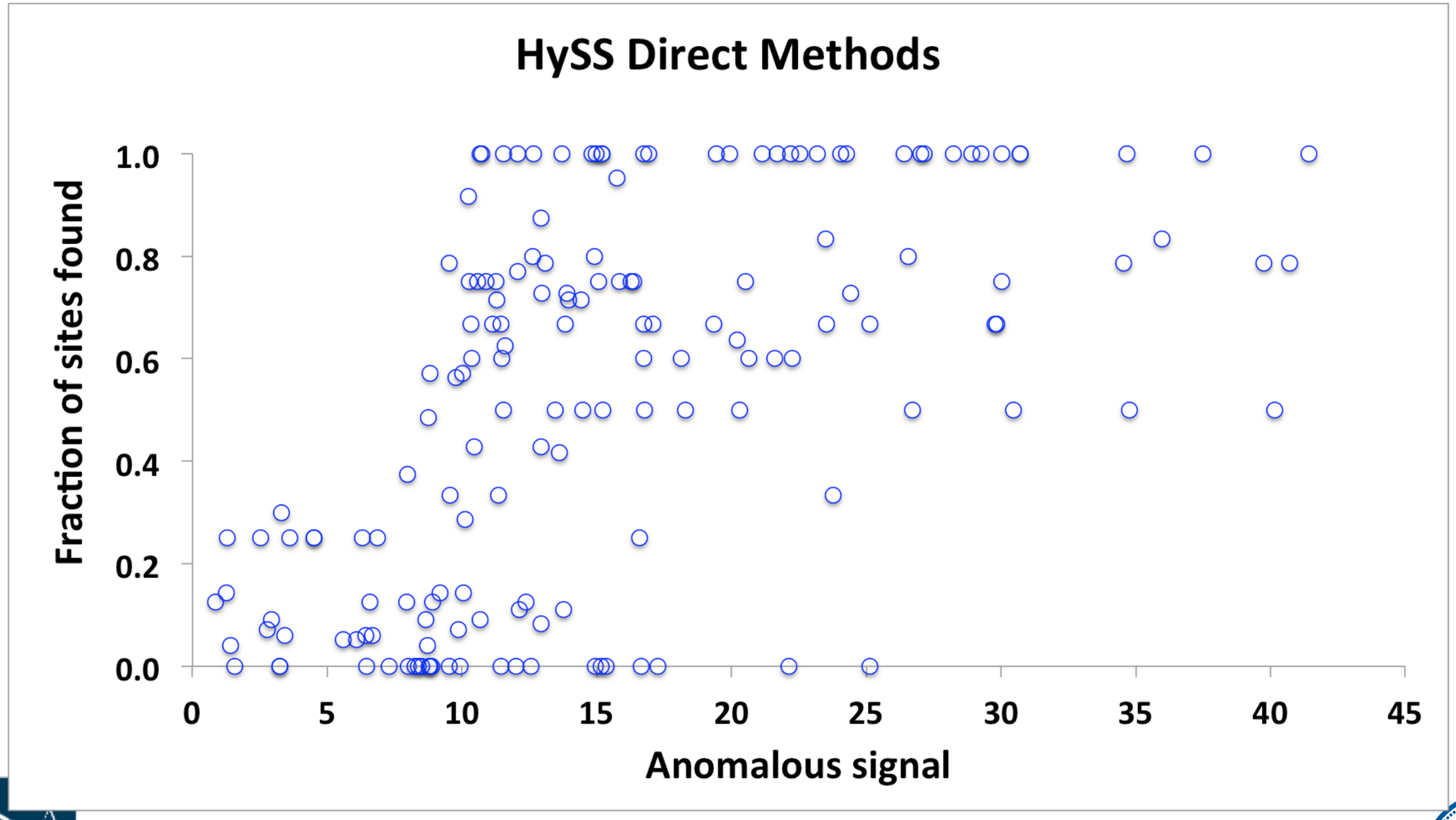
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Direct methods vs LLG completion

- 164 SAD datasets from PDB (JCSG)



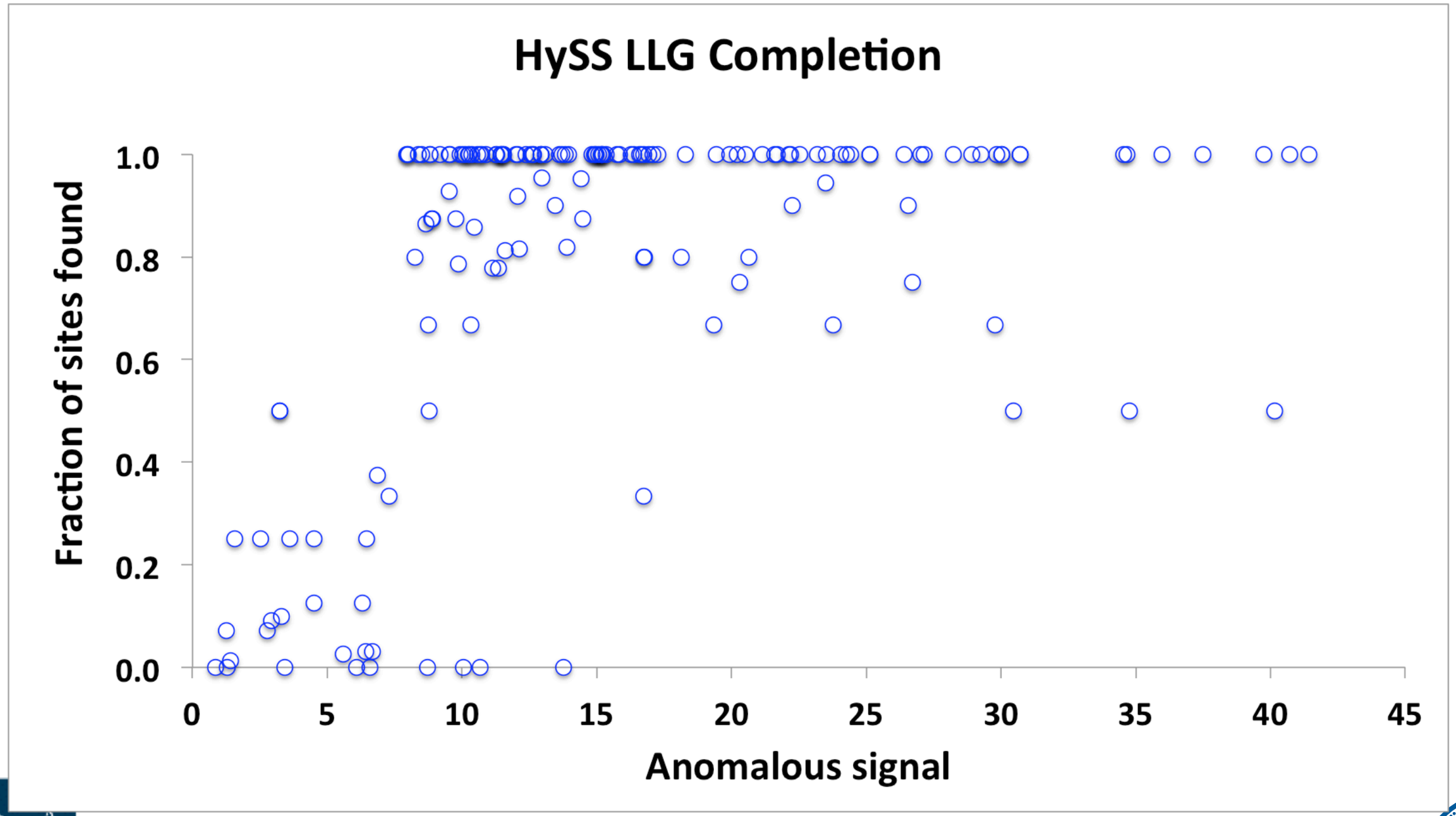
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Direct methods vs LLG completion

- 164 SAD datasets from PDB (JCSG)



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Summary of New Features in HySS

- Initiation of search with Patterson solutions, input sites, or randomized input sites
- LLG completion from Patterson solutions or direct methods solutions
- Parallel execution of searches
- Automation of search over resolution, direct methods, and Phaser completion
- Termination if same solution is found from different Patterson seeds at same resolution

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Structure Solution with Weak Signal

- AutoSol
  - Substructure solution, phasing, density modification, preliminary model-building
- AutoBuild
  - Iterative model-building, refinement, density modification
- Parallel AutoBuild
  - Parallel runs of AutoBuild with map averaging and picking best models

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Structure Solution with AutoSol

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

Find heavy-atom sites with HySS direct methods

Calculate phases (Phaser)

Improve phases, find NCS, build model

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

AutoSol Enhancements for Weak Data

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

Find heavy-atom sites with HySS direct methods and LLG Completion

Calculate phases (Phaser)

Improve phases, find NCS, build model

Use map and model in LLG Completion

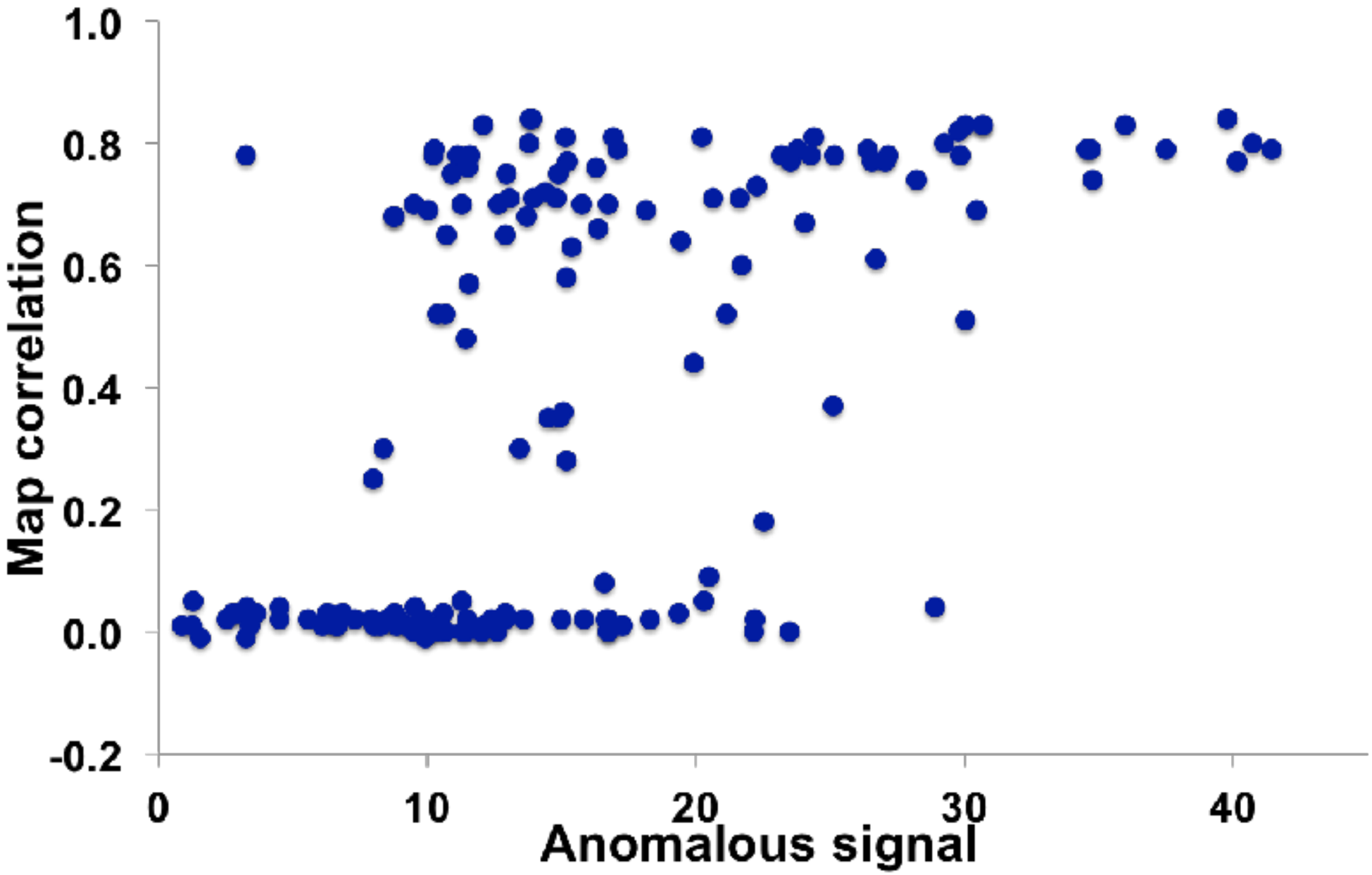
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

AutoSol structure solution

- 164 SAD datasets from PDB (JCSG)



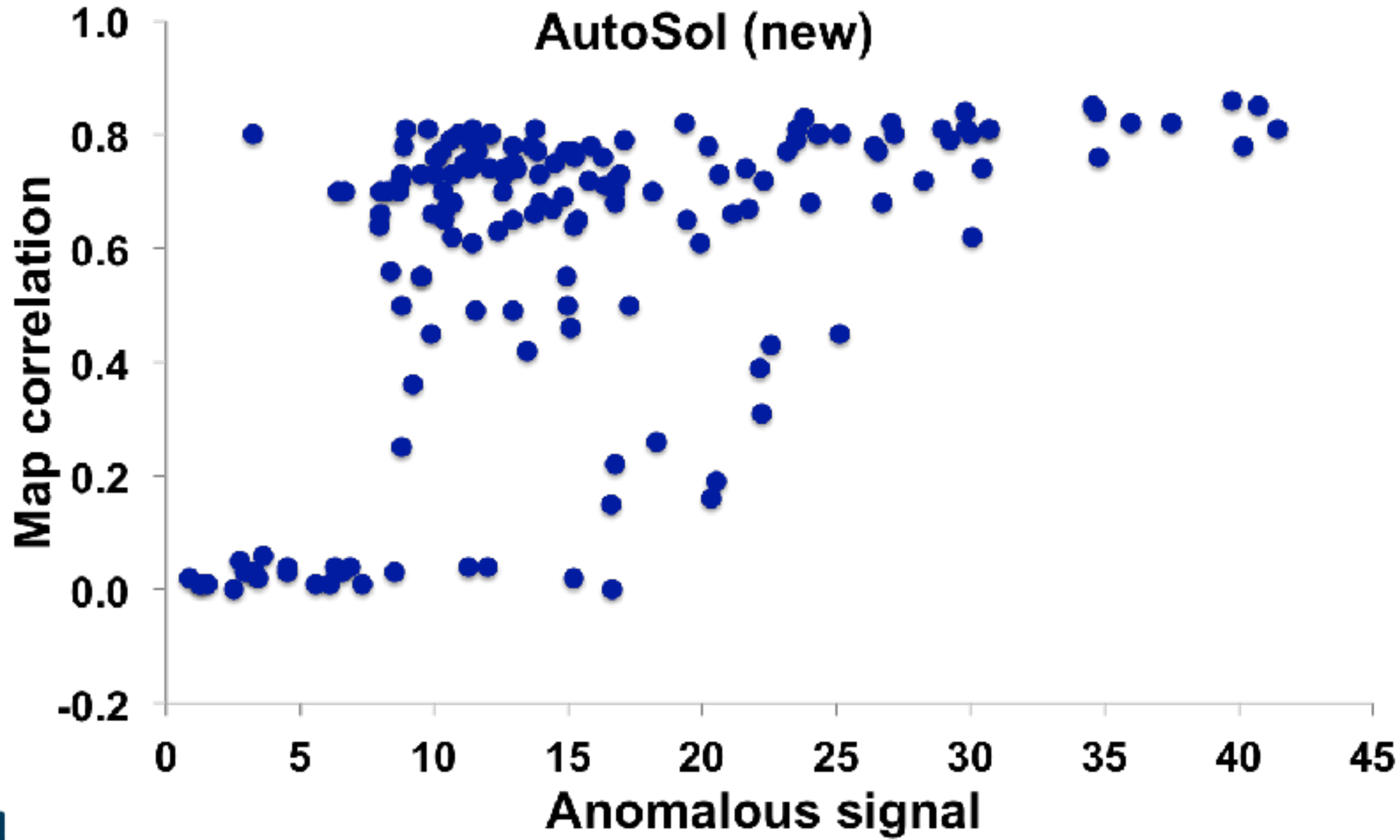
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

AutoSol structure solution

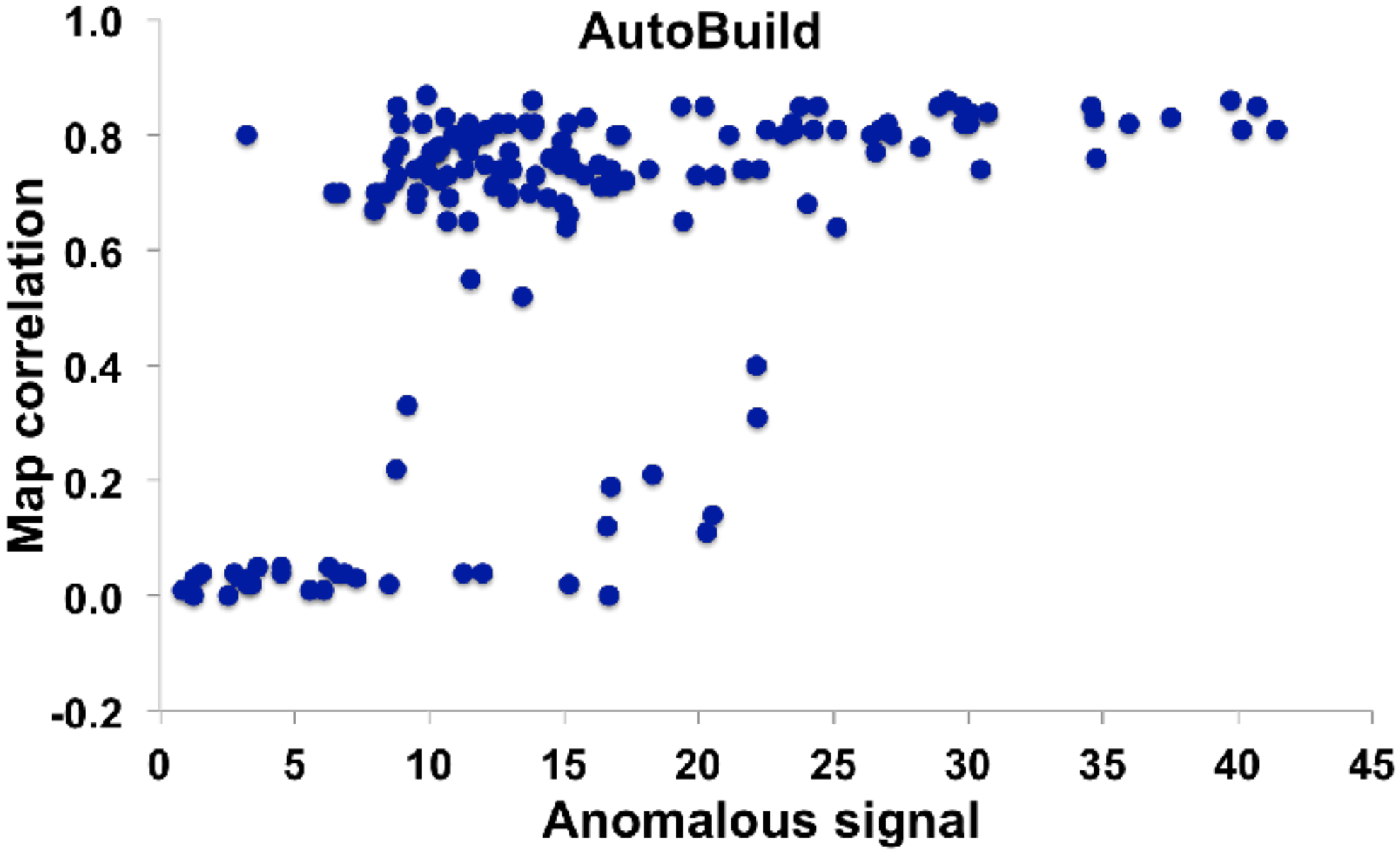
- 164 SAD datasets from PDB (JCSG)



Phenix

AutoBuild Model Building

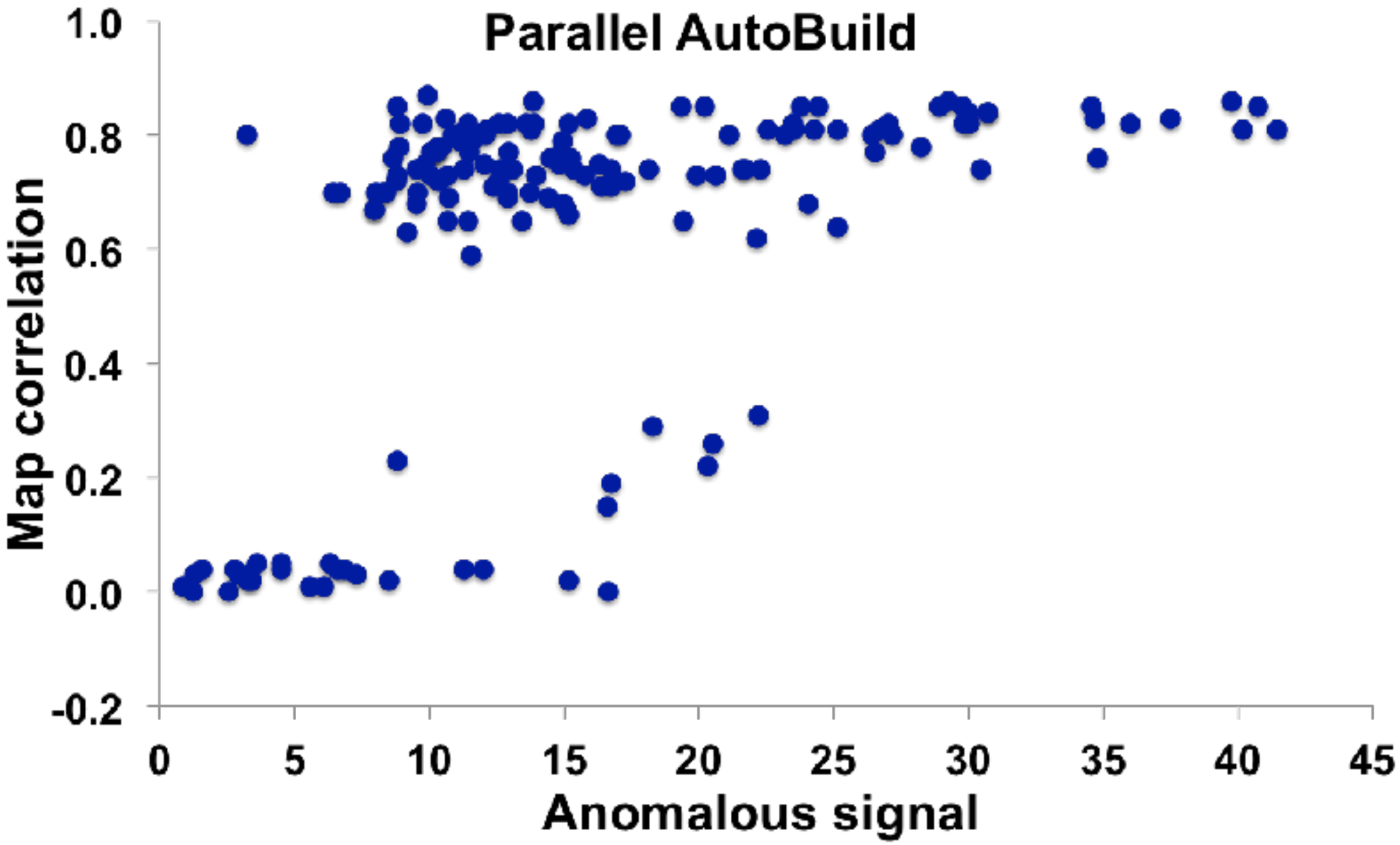
- 164 SAD datasets from PDB (JCSG)



Phenix

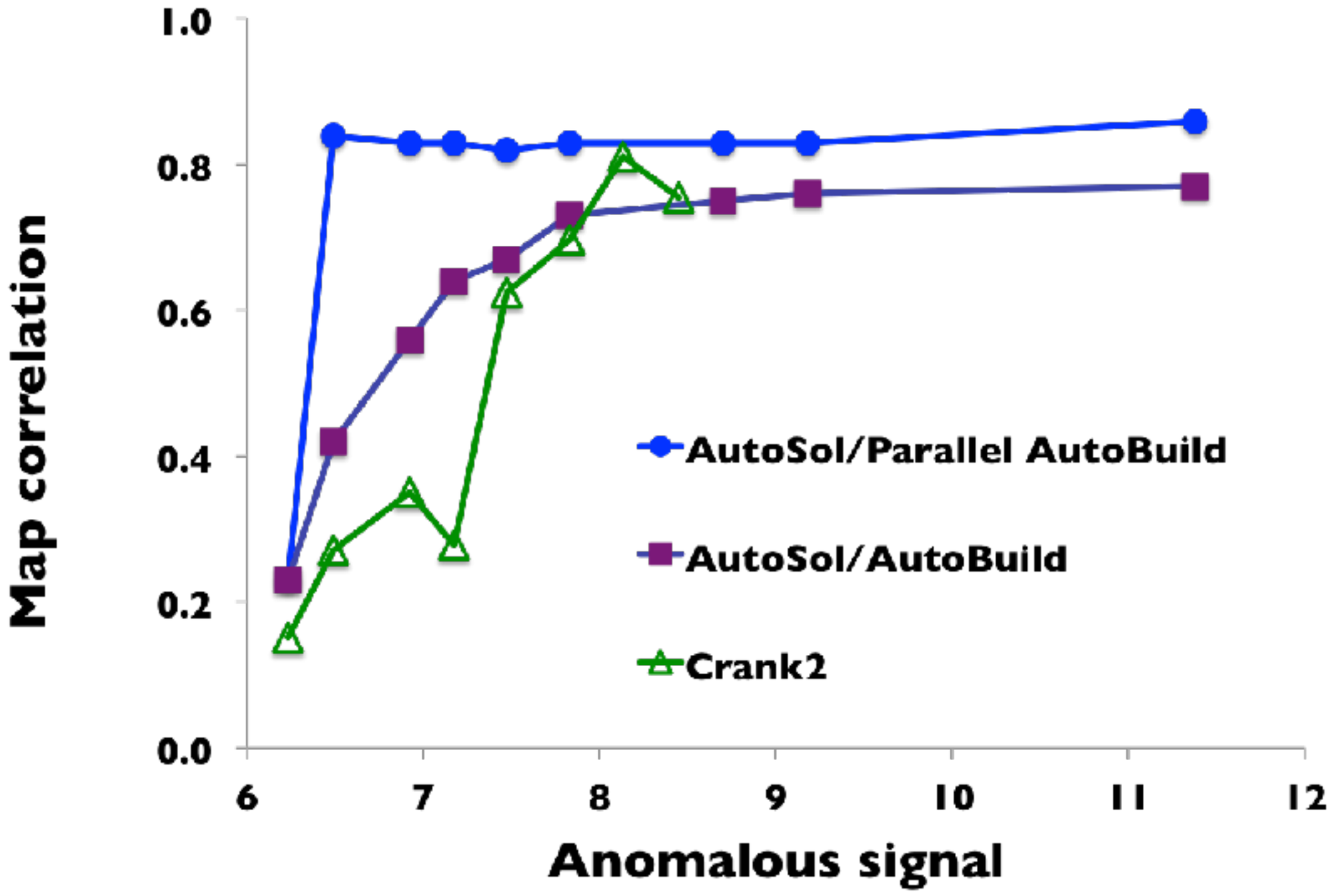
AutoBuild Model Building

- 164 SAD datasets from PDB (JCSG)



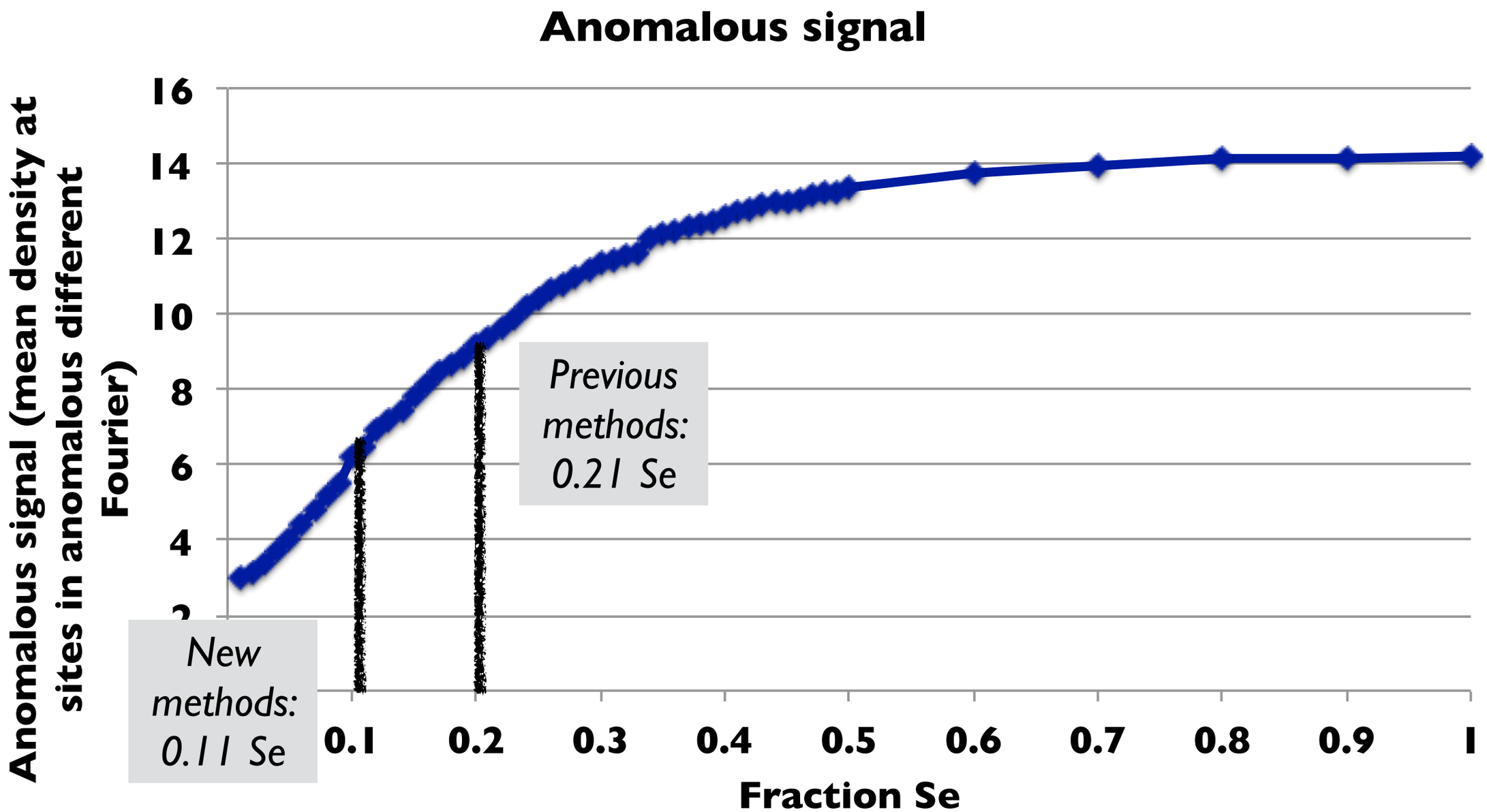
Phenix

Holton Challenge Data - Known Sites



Phenix

Progress



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY
1868

Acknowledgments

- **Lawrence Berkeley Laboratory**
  - Pavel Afonine, Youval Dar, Nat Echols, Jeff Headd, Richard Gildea, Ralf Grosse-Kunstleve, Dorothee Liebschner, Nigel Moriarty, Nader Morshed, Billy Poon, Ian Rees, Nicholas Sauter, Oleg Sobolev, Peter Zwart
- **Los Alamos National Laboratory**
  - Tom Terwilliger, Li-Wei Hung
- **Cambridge University**
  - Randy Read, Airlie McCoy, Laurent Storoni, Gabor Bunkoczi, Robert Oeffner
- **Duke University**
  - Jane Richardson & David Richardson, Ian Davis, Vincent Chen, Jeff Headd, Christopher Williams, Bryan Arendall, Laura Murray, Gary Kapral, Dan Keedy, Swati Jain, Bradley Hintze, Lindsay Deis, Lizbeth Videau
- **University of Washington**
  - Frank DiMaio, David Baker
- **Oak Ridge National Laboratory**
  - Marat Mustyakimov, Paul Langan
- **Others**
  - Alexandre Urzhumtsev & Vladimir Lunin
  - Garib Murshudov & Alexi Vagin
  - Kevin Cowtan, Paul Emsley, Bernhard Lohkamp
  - David Abrahams
  - PHENIX Testers & Users: James Fraser, Herb Klei, Warren Delano, William Scott, Joel Bard, Bob Nolte, Frank von Delft, Scott Classen, Ben Eisenbraun, Phil Evans, Felix Frolow, Christine Gee, Miguel Ortiz-Lombardia, Blaine Mooers, Daniil Prigozhin, Miles Pufall, Edward Snell, Eugene Valkov, Erik Vogan, Andre White, and many more
- **Funding:**
  - NIH/NIGMS:
    - *P01GM063210, P50GM062412, P01GM064692, R01GM071939*
  - Lawrence Berkeley Laboratory
  - PHENIX Industrial Consortium

BERKELEY LAB

Phenix