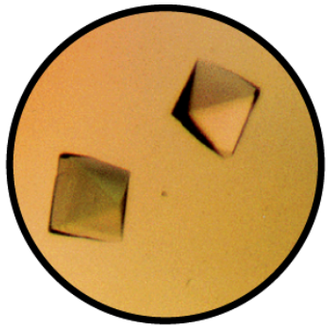
Phase Improvement

Macromolecular Crystallography School
Madrid, May 2017

Paul Adams

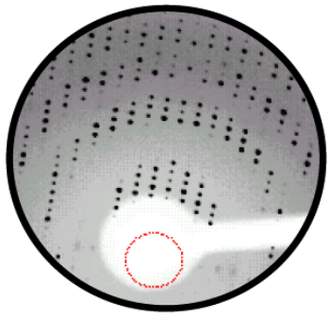
Lawrence Berkeley Laboratory and
Department of Bioengineering UC Berkeley

The Crystallographic Process

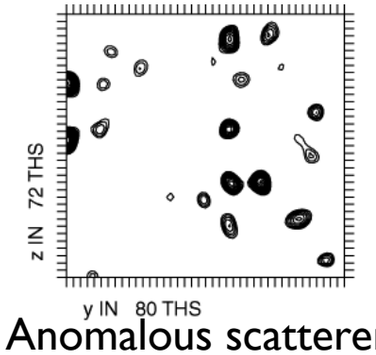


Crystallization

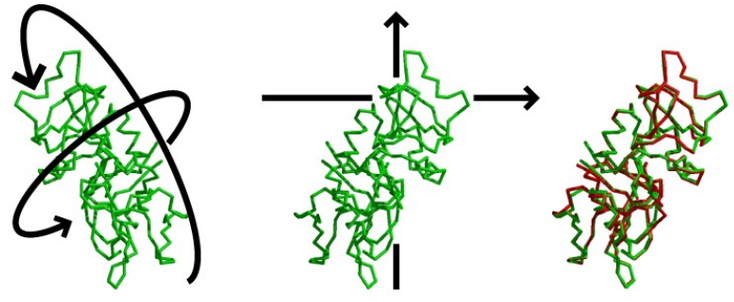
Data collection



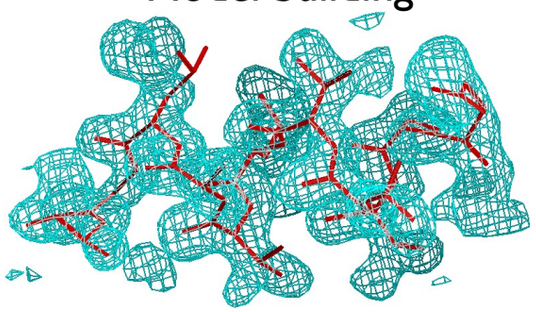
Data processing



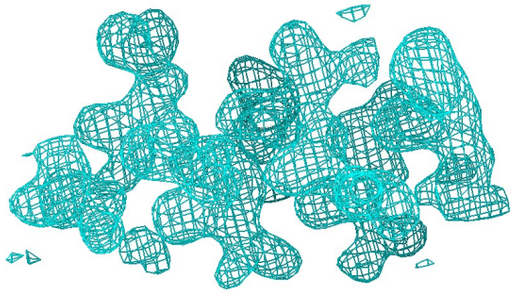
Anomalous scatterer location



Molecular replacement

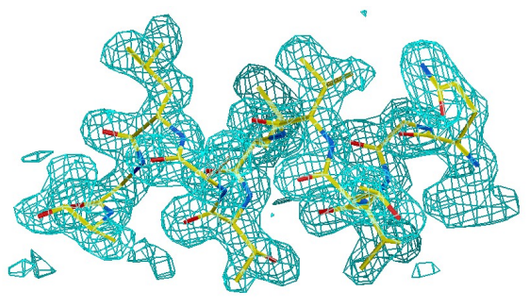


Model building

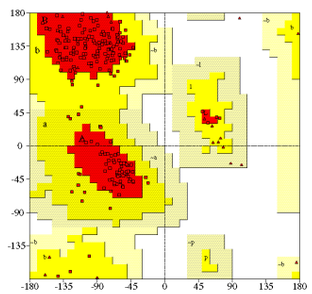


Phase improvement

Phase determination



Model refinement



Validation

BERKELEY LAB

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Phase Improvement

- Experimental phases (and those from molecular replacement) typically contain errors
- The experimental phases can be improved by the application of real space constraints
- The phases are modified to produce a map most consistent with what we know about macromolecular structures:
  - Solvent density distribution (Solvent flattening)
  - Atomicity and positivity (Sayre's equation)
  - Macromolecular density distributions (histogram matching)
  - Similarity between molecules (NCS averaging)

Phenix

The Basics

- Method to identify solvent versus macromolecular density in map
- Methods to determine relationships between different regions of the asymmetric unit
- Method to combine phase probability distributions (e.g. experimental phases with calculated phases)

Solvent flattening: Wang, B.-C. (1985). Methods Enzymol. 115, 90-112

NCS Averaging: Bricogne, G. (1974). Acta Cryst. A30, 395-405.

DM Program: Cowtan, K.D. & Main, P. Acta Cryst. (1993). D49, 148-157

BERKELEY LAB

Phenix

Identifying the Solvent Region

- Experimental and MR-phased maps usually contain some information about the boundary of the macromolecule
  - SAD and SIR maps are the combination of the correct map (made with the correct phase choice) and noise (a map made with the incorrect phase choice)
- The envelope can be recovered by looking at the local standard deviation (the variance) of the electron density at each grid point in the map
  - The standard deviation will be high in the macromolecular region and low in the solvent

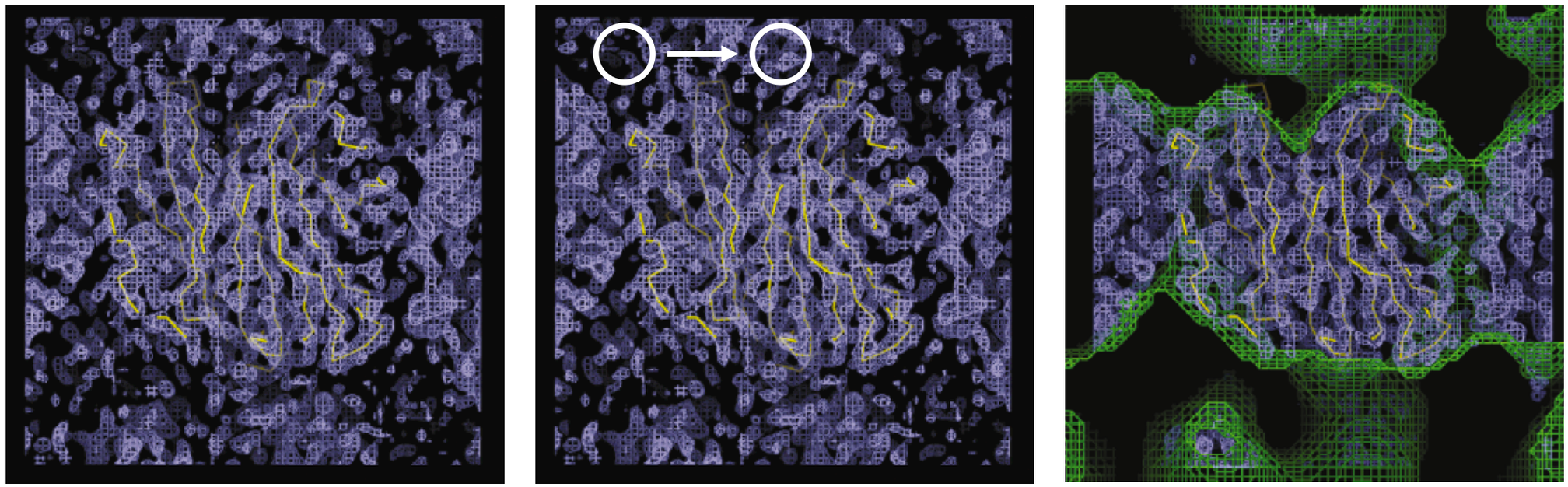


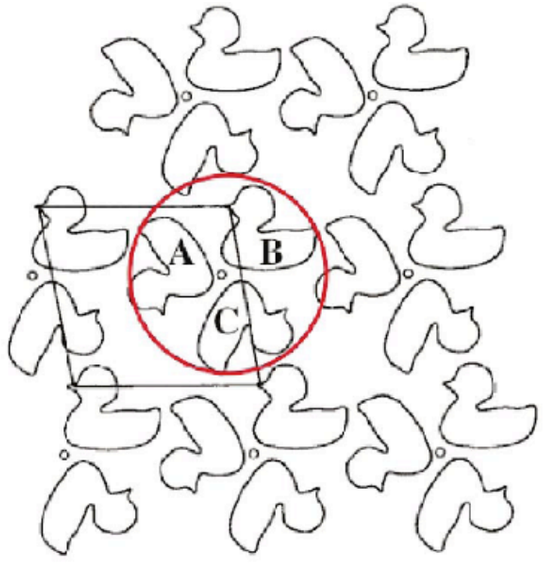
Image from G.Taylor, Acta Cryst. D, 59, 1881-1890 (2003)

BERKELEY LAB

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Non-crystallographic Symmetry



$$\rho(\mathbf{x}_B) = \mathbf{R}_B\rho(\mathbf{x}_A) + \mathbf{t}_B$$
$$\rho(\mathbf{x}_C) = \mathbf{R}_C\rho(\mathbf{x}_A) + \mathbf{t}_C$$

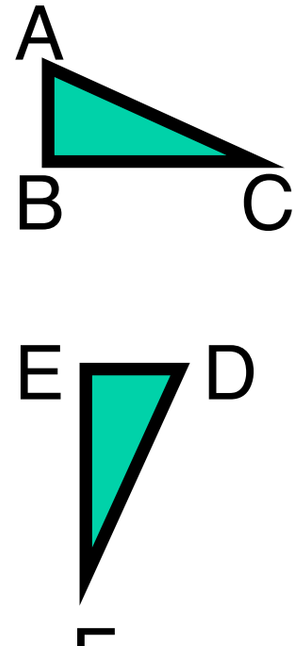
NCS symmetry operators

- The presence of multiple copies of the same molecule in the asymmetric unit provides additional information in phase improvement
  - Electron density can be averaged to enforce the NCS relationship
  - The similarity of the related regions can be used as an indicator of the success of phase improvement
- The relationship between molecules and the mask around them must be defined
  - NCS is often referred to as proper (2-fold, 3-fold, 4-fold etc.) or improper (an arbitrary relationship between molecules
  - NCS is quite common

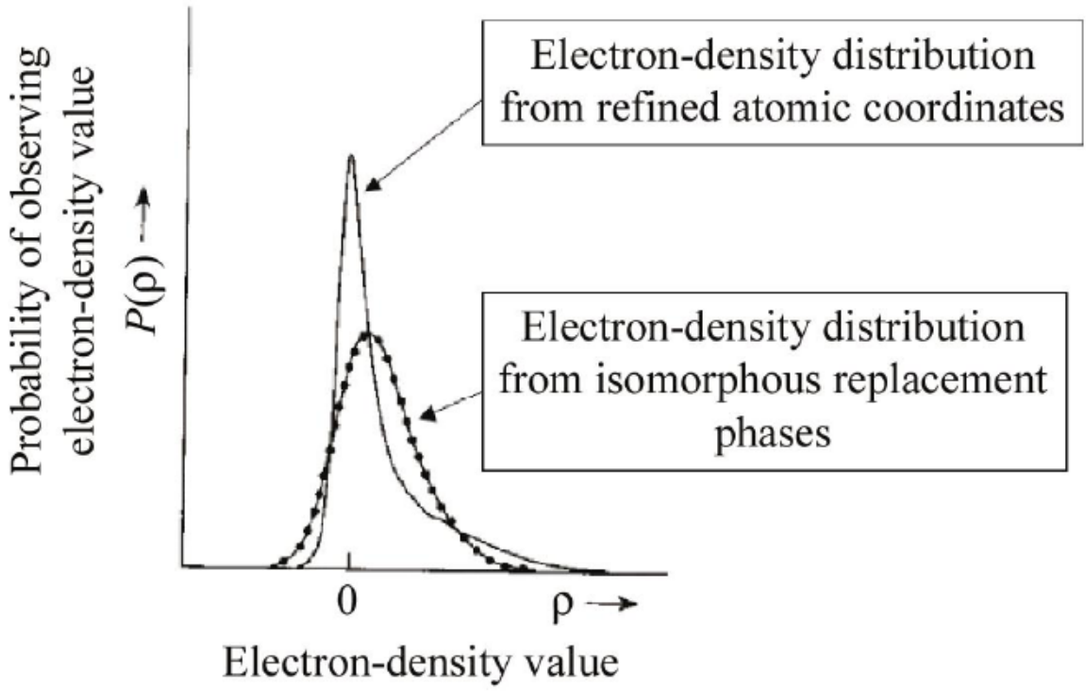
Image from G.Taylor, Acta Cryst. D, 59, 1881-1890 (2003)

Determining NCS Relationships

- Non-crystallographic symmetry can typically be determined:
  - From substructure sites
  - From real space correlation searches
  - From the MR solution
- From substructure sites:
  - Expand heavy-atom sites within radius R of origin
  - Make list of all pairs of sites, sorted by distance between sites d
  - Choose any 3 HA sites forming a triangle ABC
  - Find all other sets of 3 HA sites that form the same triangle
    - If some exist (DEF) -> this might correspond to NCS
    - If none exist then try another set of 3 HA sites
  - Test the electron density for each possible NCS operator to see if they show some correlation



Histogram Matching



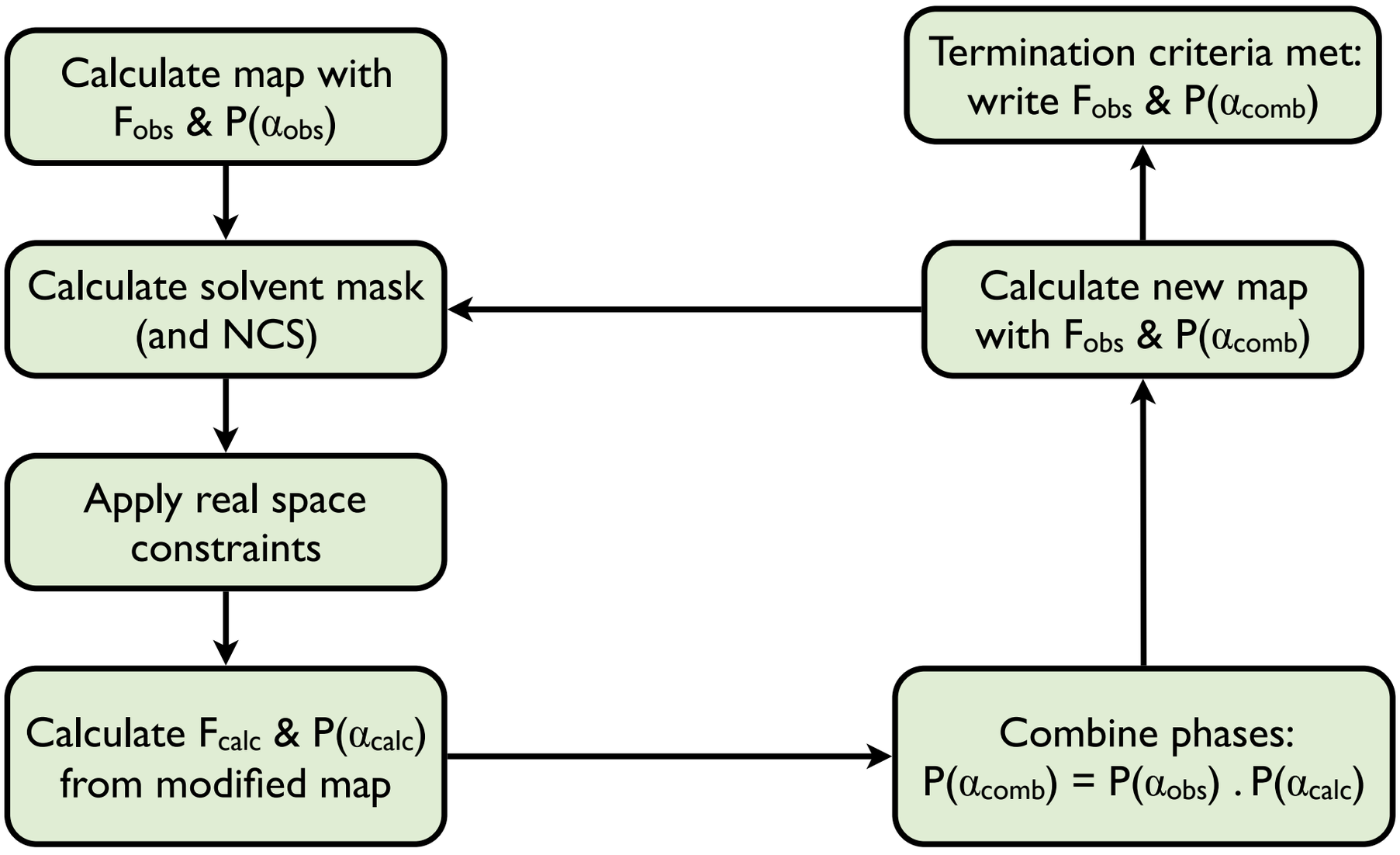
- The electron density of macromolecules have fairly similar distributions (but are dependent on the type of molecule and the resolution)
- This information can be used to match the observed histogram of densities to an ideal histogram
- This is one of the most powerful constraints on the density (and hence in phase improvement)
- The histogram matching method is not unique to crystallography
  - Used in many different image processing applications

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

BERKELEY LAB

Phenix

Classical Density Modification



- This approach works, but there is a bias problem
  - The observed and modified phases (and amplitudes) are correlated – we used the observed phases to calculate the map that we modified to make the new phases

BERKELEY LAB

Phenix

The γ-correction to reduce bias

- Solvent flattening is the multiplication of the original map with a mask
- This can be expressed in reciprocal space as a convolution of a reciprocal space mask function (G-function) with experimental structure factors
- A term in the G-function results in a component of the original map always being present in the modified map
- This component can be subtracted to minimize this bias term
- In practice the result is multiplication of the solvent density by a negative factor (flipping the solvent density)

$$\rho(x)_{new} = g(x) \times \rho(x)_{old}$$

$$F(h)_{new} = G(h) \otimes F(h)_{old}$$

$$F(h)_{new} = G(h \neq 0) \otimes F(h)_{old} + G(h = 0) \otimes F(h)_{old}$$

$$\rho(x)_{new} = FT[G(h \neq 0) \otimes F(h)_{old}] + FT[G(h = 0) \otimes F(h)_{old}]$$

but $FT[G(h = 0)]$ is constant

$$\rho(x)'_{new} = FT[G(h) \otimes F(h)_{old}] - FT[G(h = 0) \otimes F(h)_{old}]$$

$$\rho(x)'_{new} = g(x) \times \rho(x)_{old} - g_{const} \times \rho(x)_{old}$$

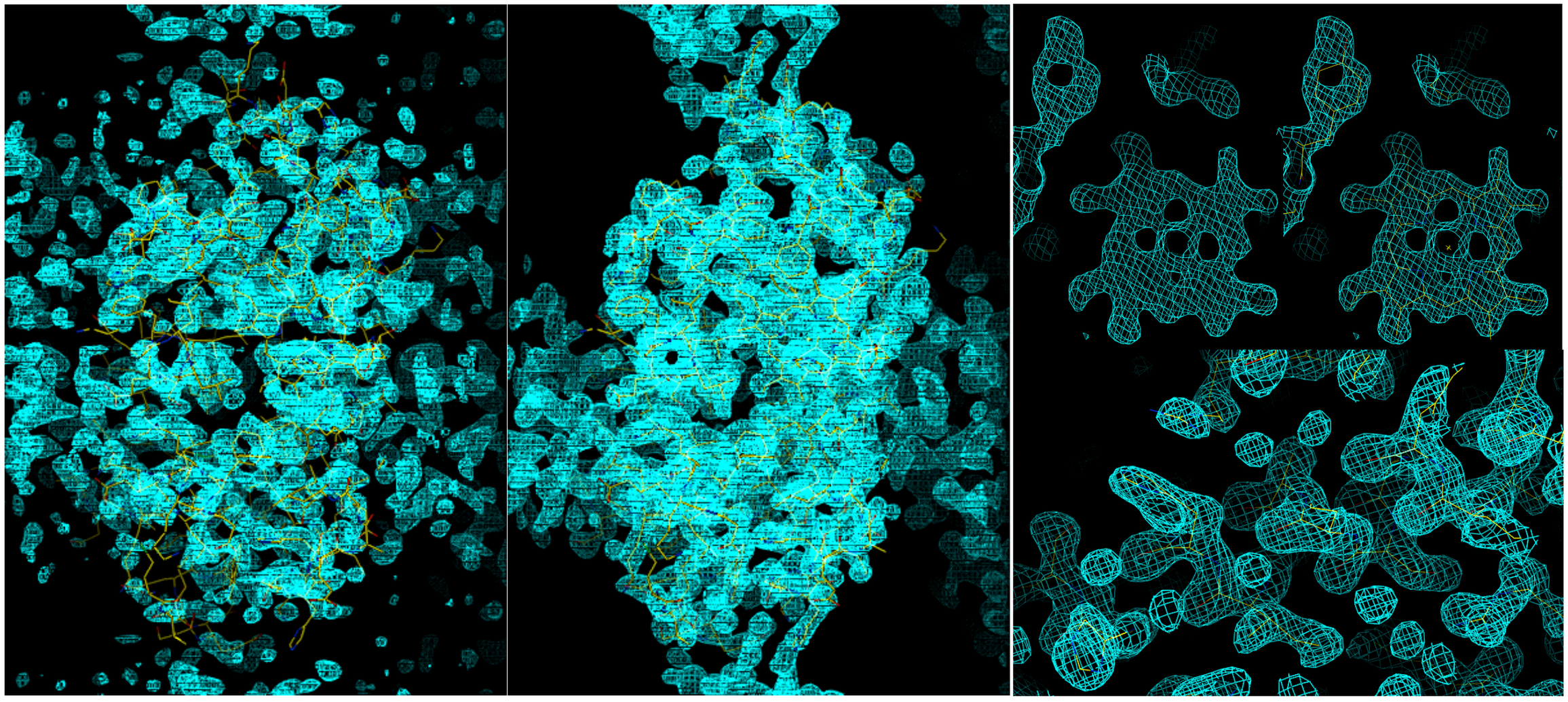
$$= [g(x) - g_{const}] \times \rho(x)_{old} \qquad g_{const} = \frac{V_{protein}}{V_{total}}$$

Abrahams, J.P. Acta Cryst. (1997). D53, 371-376

BERKELEY LAB

Phenix

Density Modification (SAD Phases)



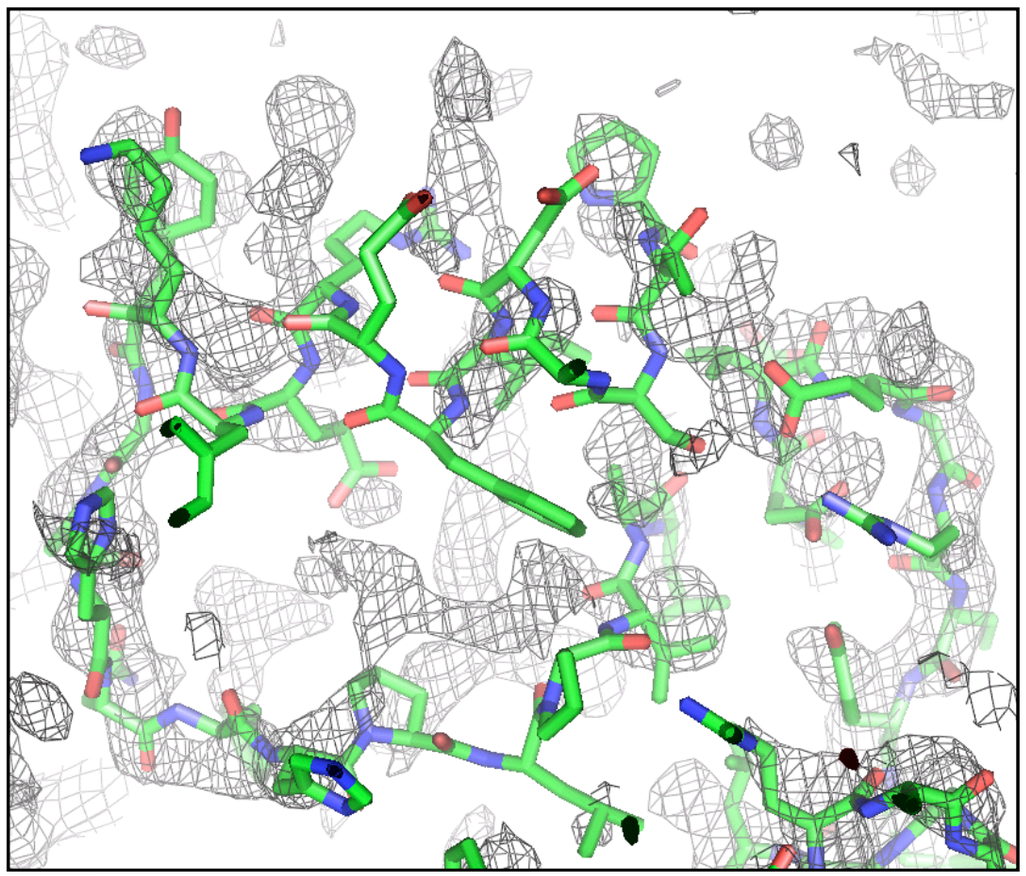
- Myoglobin, phasing from 1 Fe, solvent content=58%

BERKELEY LAB

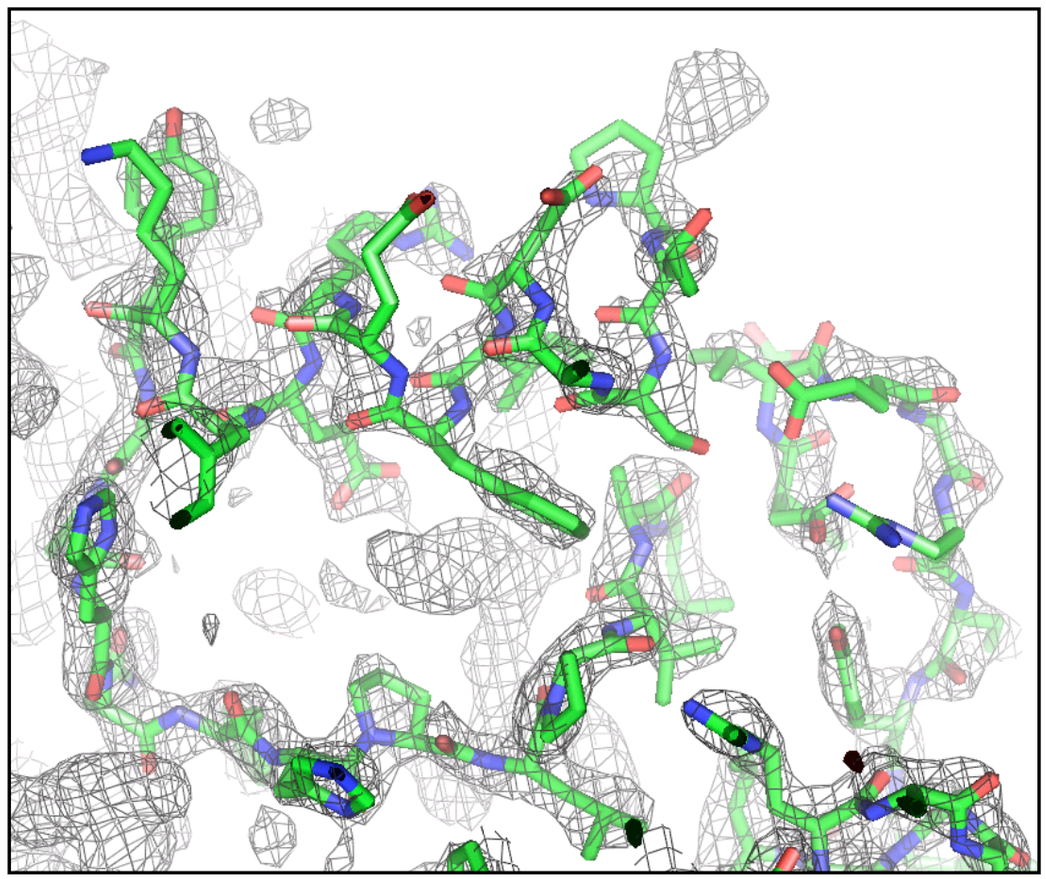
Phenix

Phase Extension with NCS

- Sometimes high resolution native data are available in addition to the data from the phasing experiment
- Phases can be extended to higher resolution, especially in the presence of NCS
- Phase extension works because long-range relationships in the electron density (such as NCS) lead to short range relationships in reciprocal space. Determining the phases at a given resolution limit also generates some useful information about reflections at a slightly higher resolution.



σ_A *weighted map from MR solution*



density modified map (3-fold NCS)

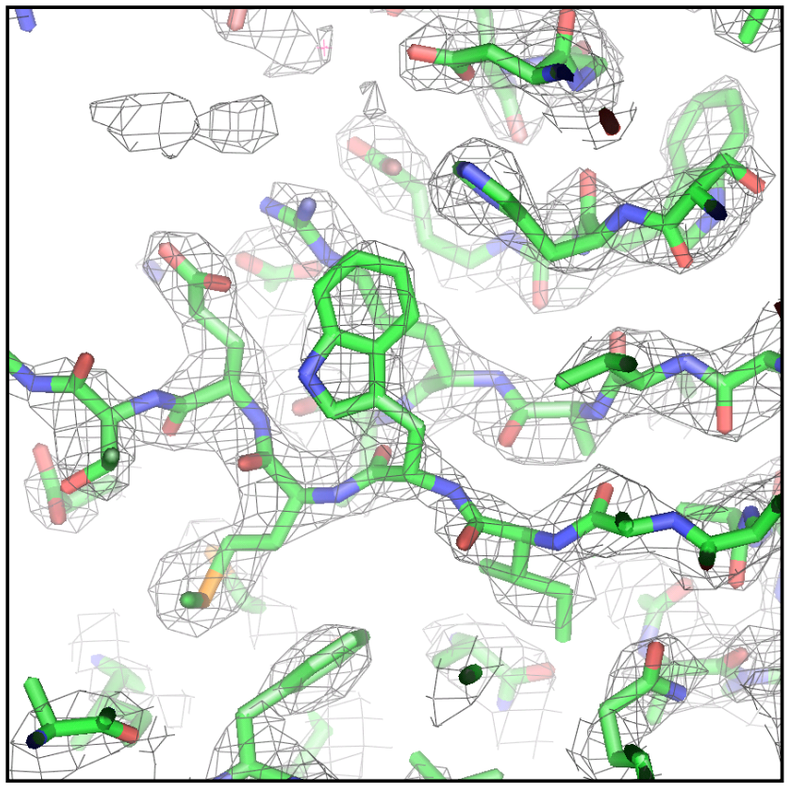
BERKELEY LAB

Phenix

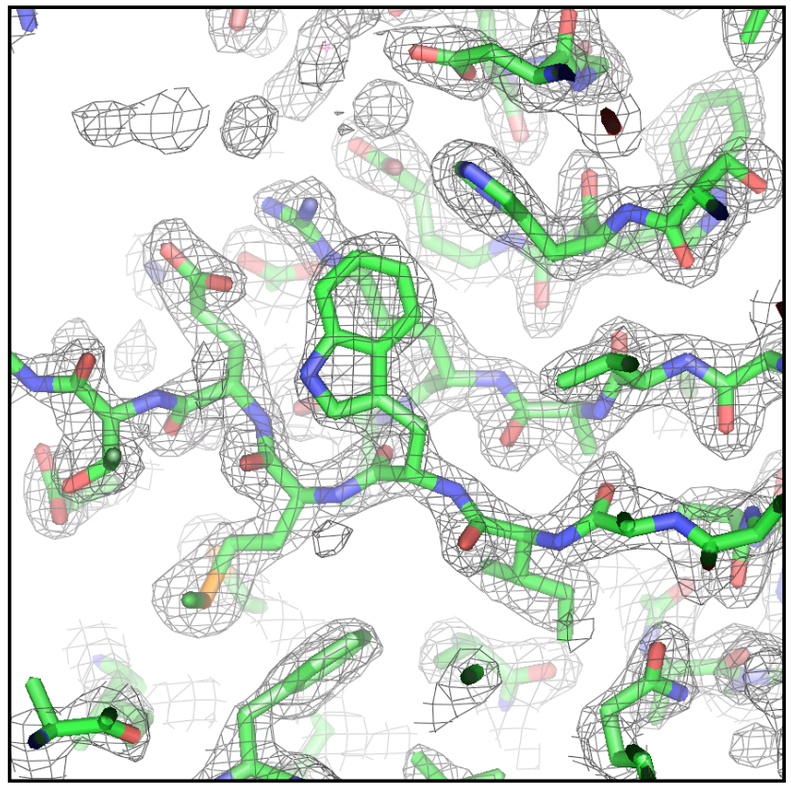
THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Phase Extension

- Phases can be extended to higher resolution even without NCS
- Phase extension still works because long-range relationships in the electron density (such as the solvent region) lead to short range relationships in reciprocal space. Determining the phases at a given resolution limit also generates some useful information about reflections nearby in reciprocal space.
- The effect of the solvent is less powerful than NCS, but significant improvements in map quality can by obtained



Density modified map at 3Å



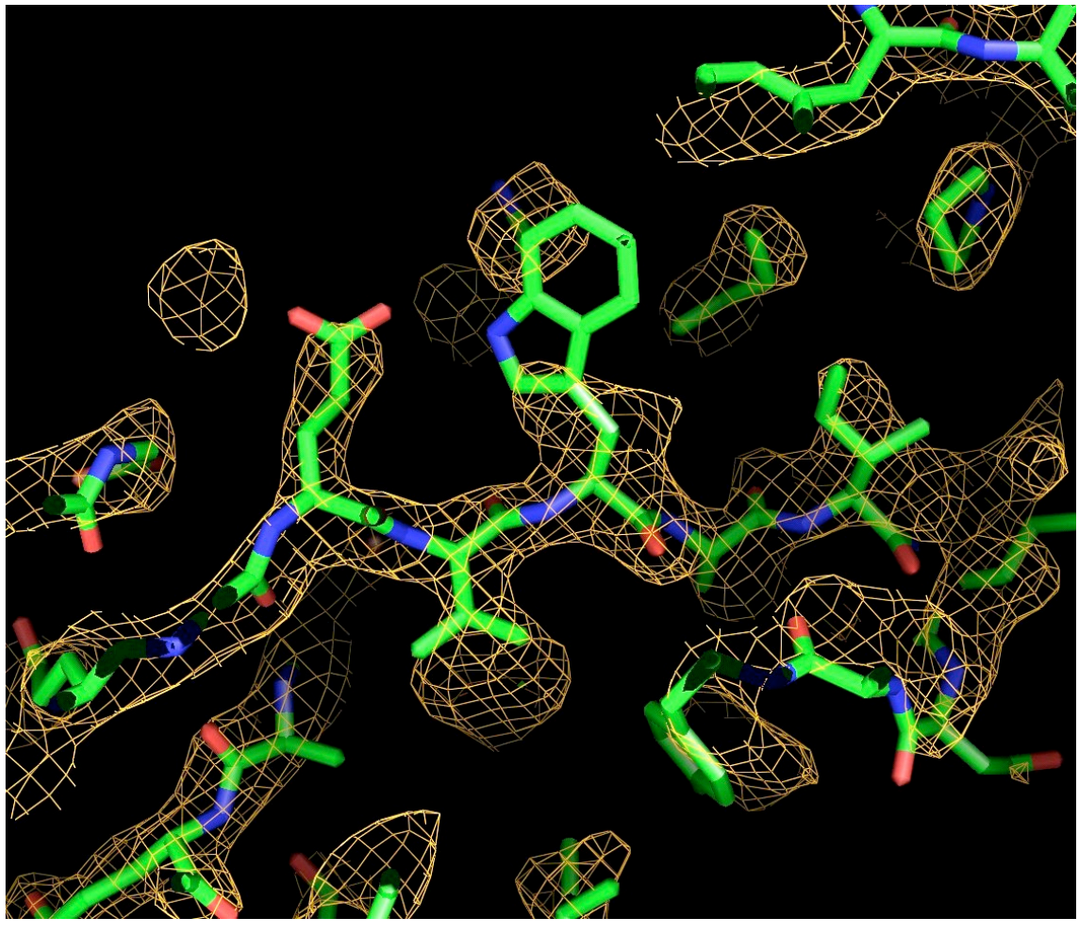
Density modified map at 2Å

BERKELEY LAB

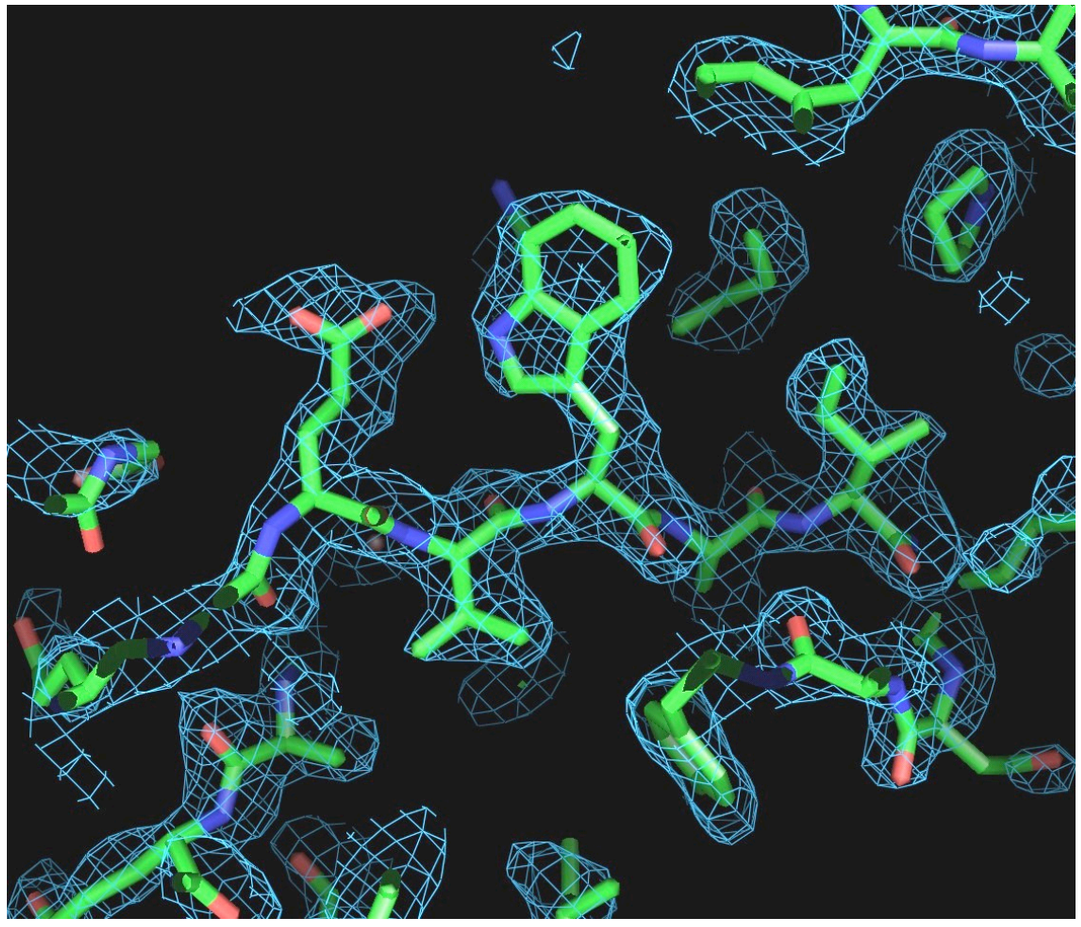
Phenix

Bias Removal

Before



After



Phasing from MR model (FOM=0.27), solvent content=58%

- Model bias is a significant issue with molecular replacement phases
  - The map looks like the input model
- By generating phases consistent with the observed amplitudes the bias can be reduced

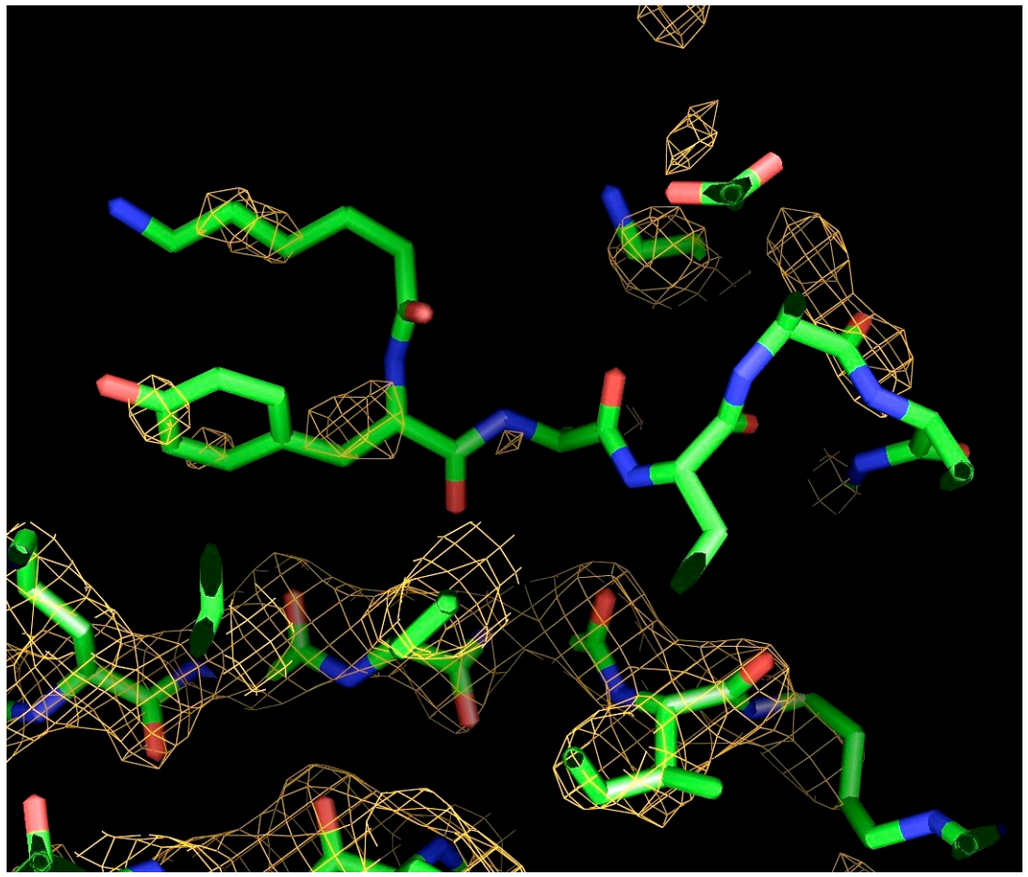
BERKELEY LAB

Phenix

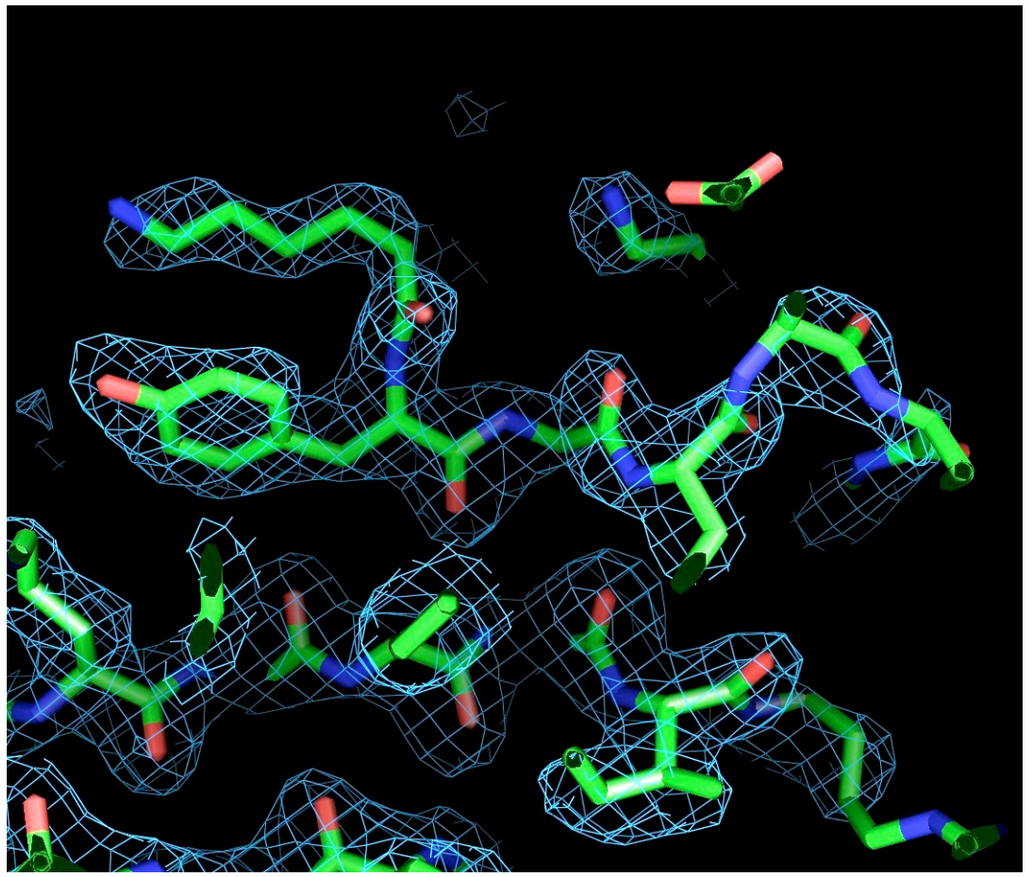
THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Recovery of Missing Information

Before



After



Phasing from MR model (FOM=0.27), solvent content=58%

- Model bias, noise and phase errors can contribute to missing features in the map
- Density modification can retrieve features (if they are not too weak)

BERKELEY LAB

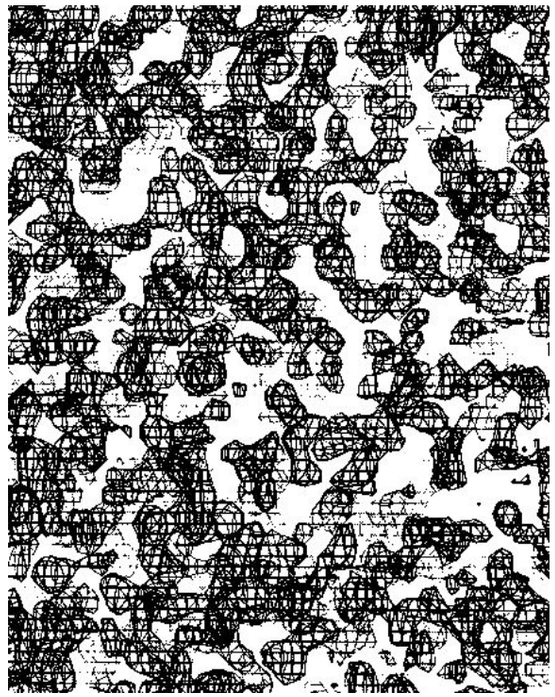
Phenix

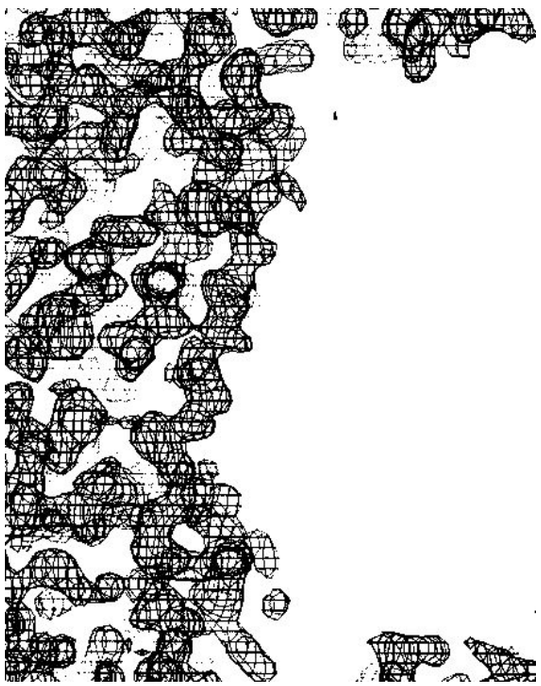
Improving Phase Improvement

- The traditional phase improvement method has been used very successful to solve many structures. However, there are still some problems:
  - Relative weights in phase combination
  - When to terminate the procedure
  - Unequal uncertainties in different parts of the map
- The traditional method has no way to measure the "correctness" of the modified map

Statistical Phase Improvement

- Principle: phase probability information from probability of the map and from experiment:
- $P(\varphi) = P_{\text{map probability}}(\varphi)\, P_{\text{experiment}}(\varphi)$
- Phases that lead to a believable map are more probable than those that do not
- A believable map is a map that has…
  - a relatively flat solvent region
  - NCS (if appropriate)
  - A distribution of densities like those of model proteins
- Method:
  - calculate how map probability varies with electron density ρ
  - deduce how map probability varies with phase φ
  - combine with experimental phase information



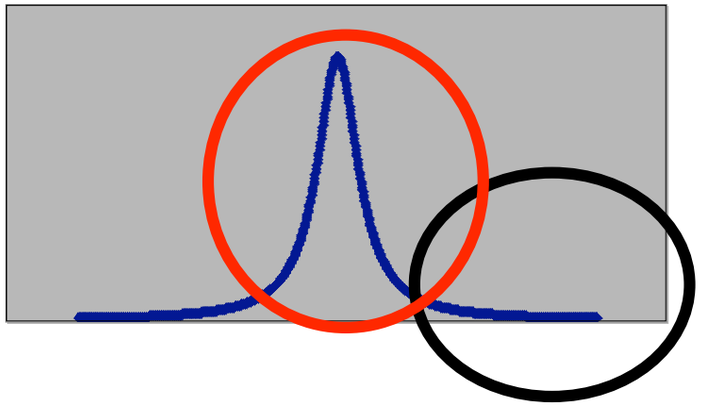


BERKELEY LAB

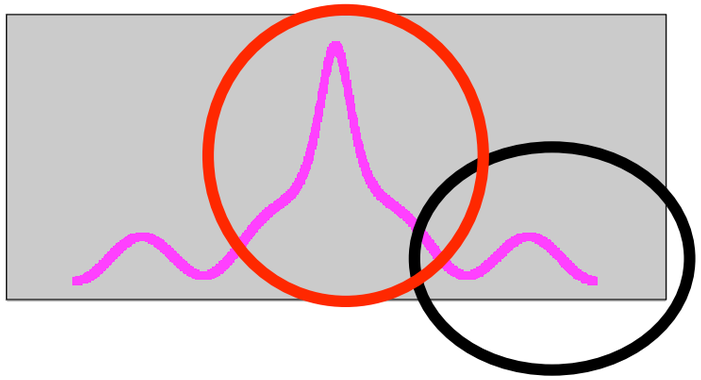
Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

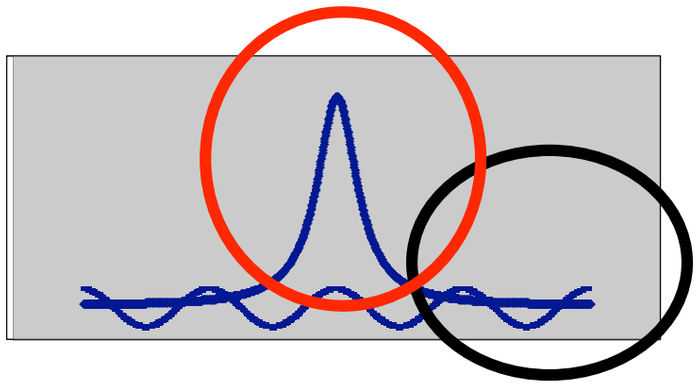
Map Probability Phasing



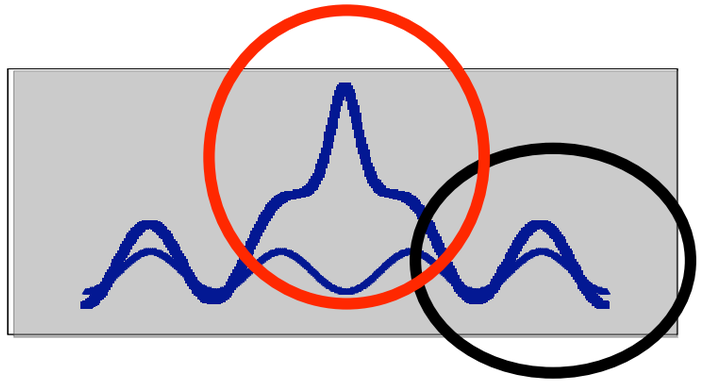
A function that is (relatively) flat far from the origin



Function calculated from estimates of all structure factors but one (k)



Test each possible phase of structure factor k. P(φ) is high for phase that leads to flat region



- Test all possible phases φ for structure factor k (for each phase, calculate new map including k)
- Probability of phase φ estimated from agreement of map with expectations
- Phase probability of reflection k from map is independent of starting phase probability because reflection k is omitted from the map

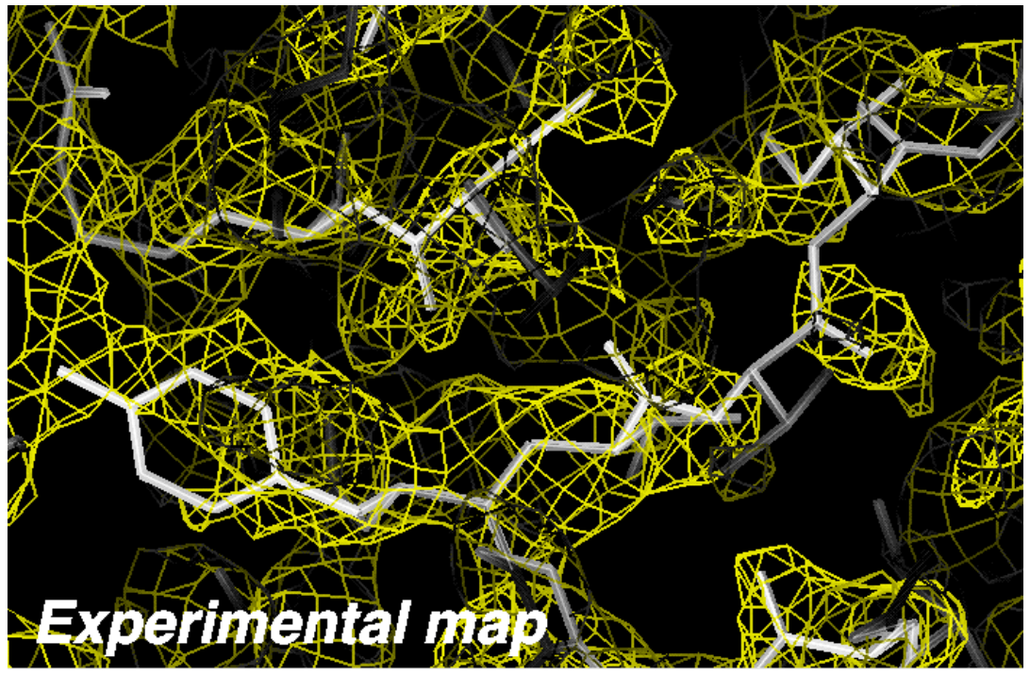
Image from Tom Terwilliger, Los Alamos National Laboratory

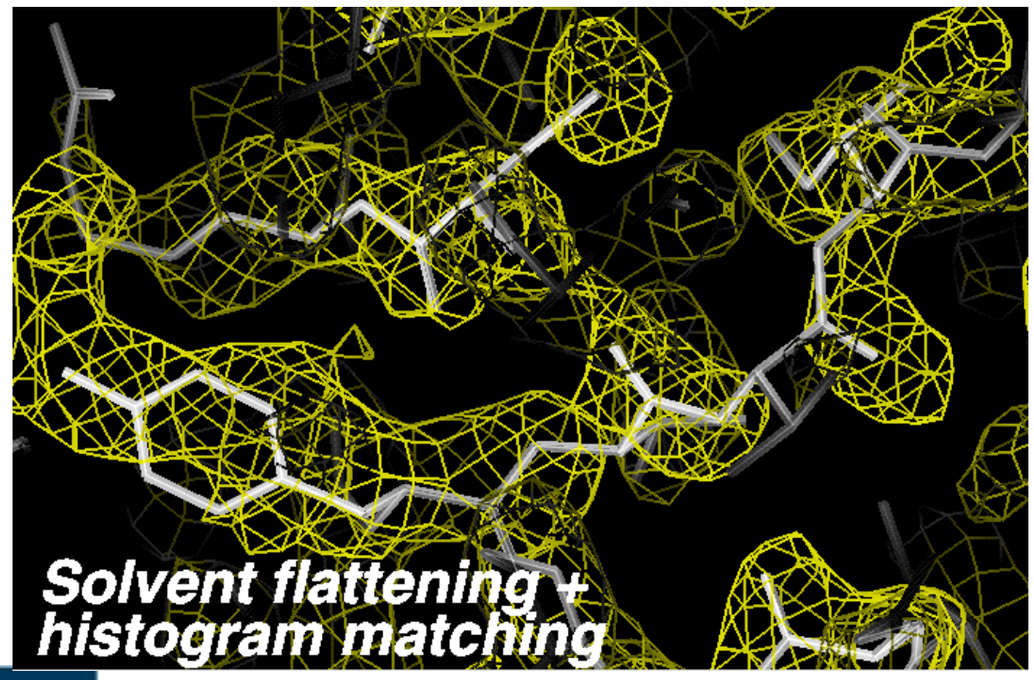
BERKELEY LAB

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Statistical Phase Improvement





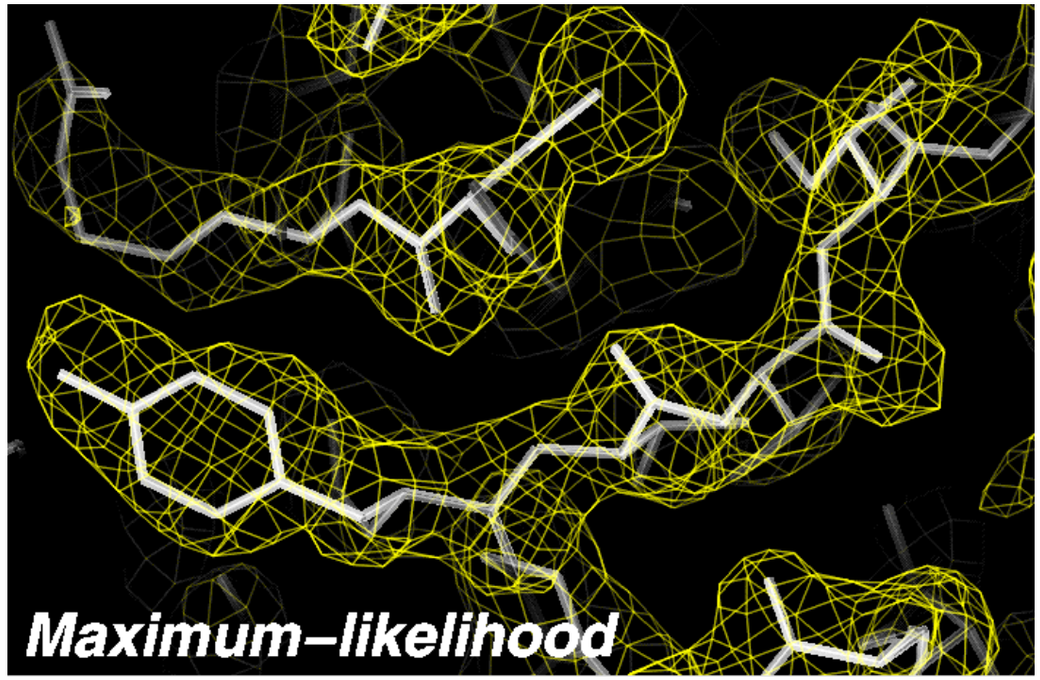


Image from Tom Terwilliger, Los Alamos National Laboratory

BERKELEY LAB

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Statistical Phase Improvement

- Prime-and-switch phasing (RESOLVE):
  - Start with σ_A-weighted map
  - Identify solvent region (or other features of map)
  - Adjust the phases to maximize the likelihood of the map – without biasing towards the model phases

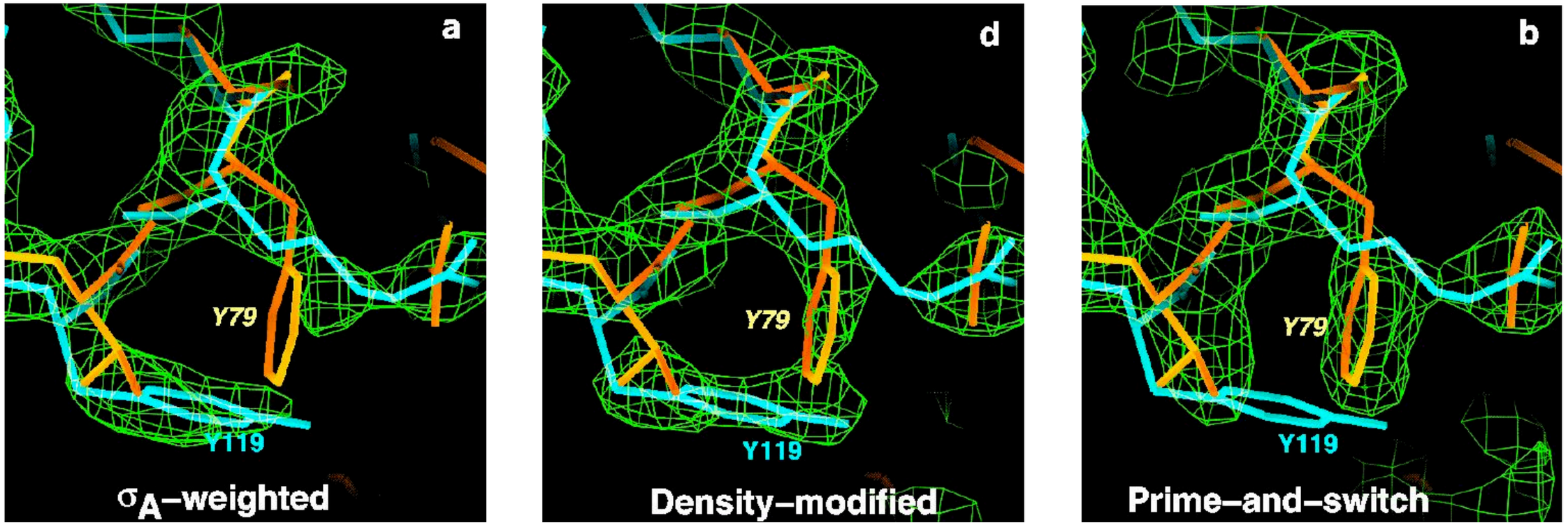


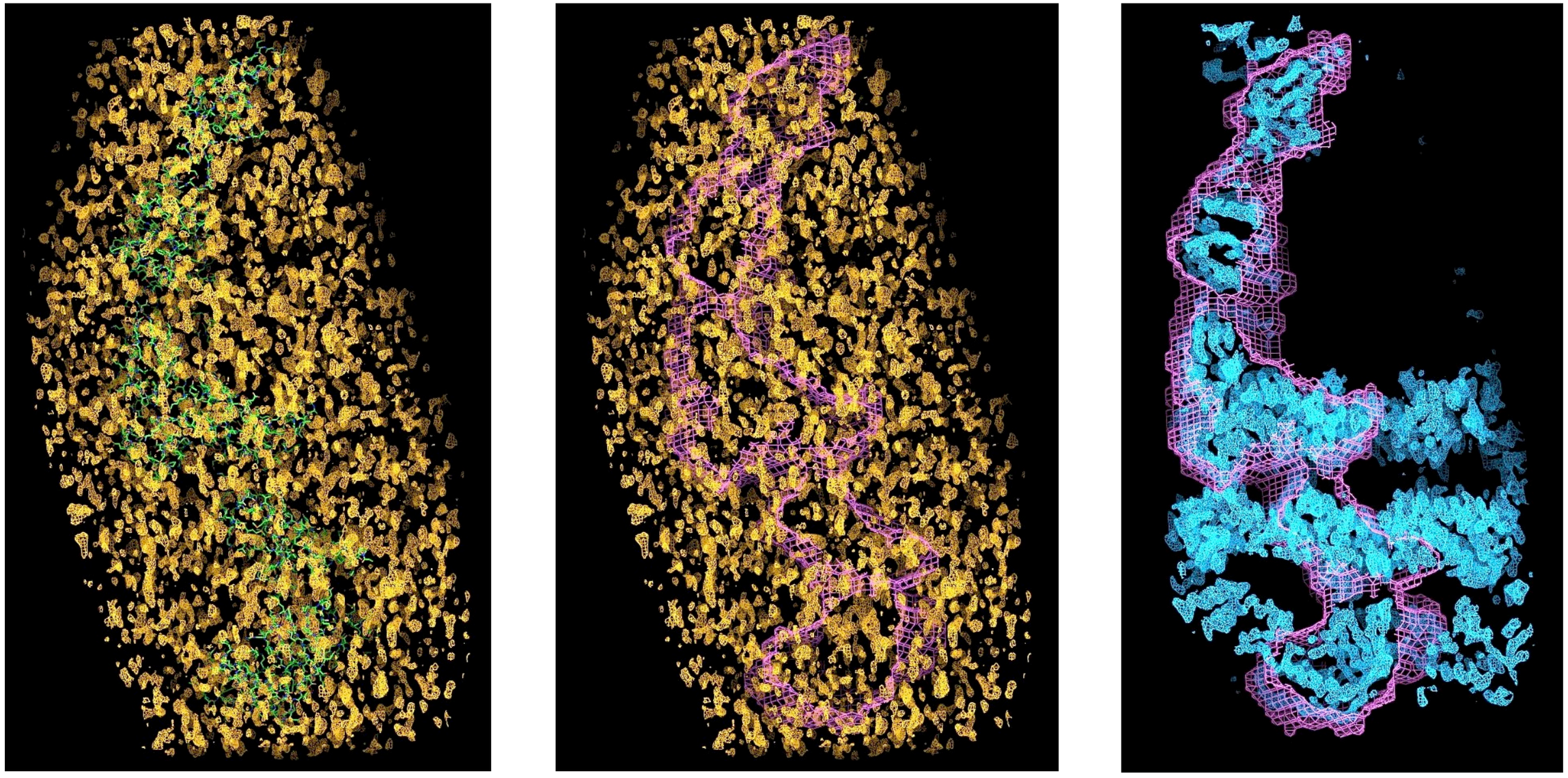
Image from Tom Terwilliger, Los Alamos National Laboratory

BERKELEY LAB

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Starting From Random Phases



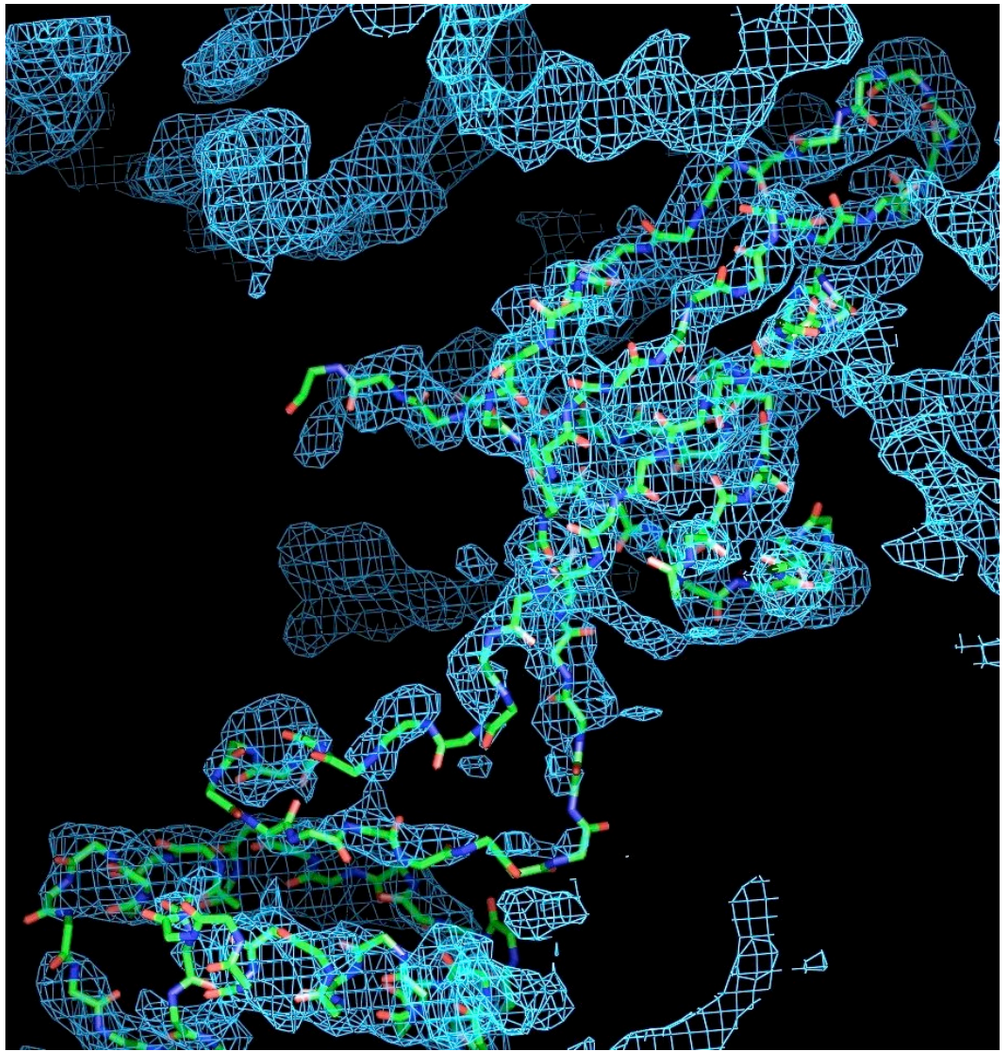
- GroEL, random phases (FOM=0.1), solvent content=60%
- 7-fold averaging using mask calculated from MR solution
- Starting high resolution limit=10Å, final=3.0Å, 170 modification steps

BERKELEY LAB

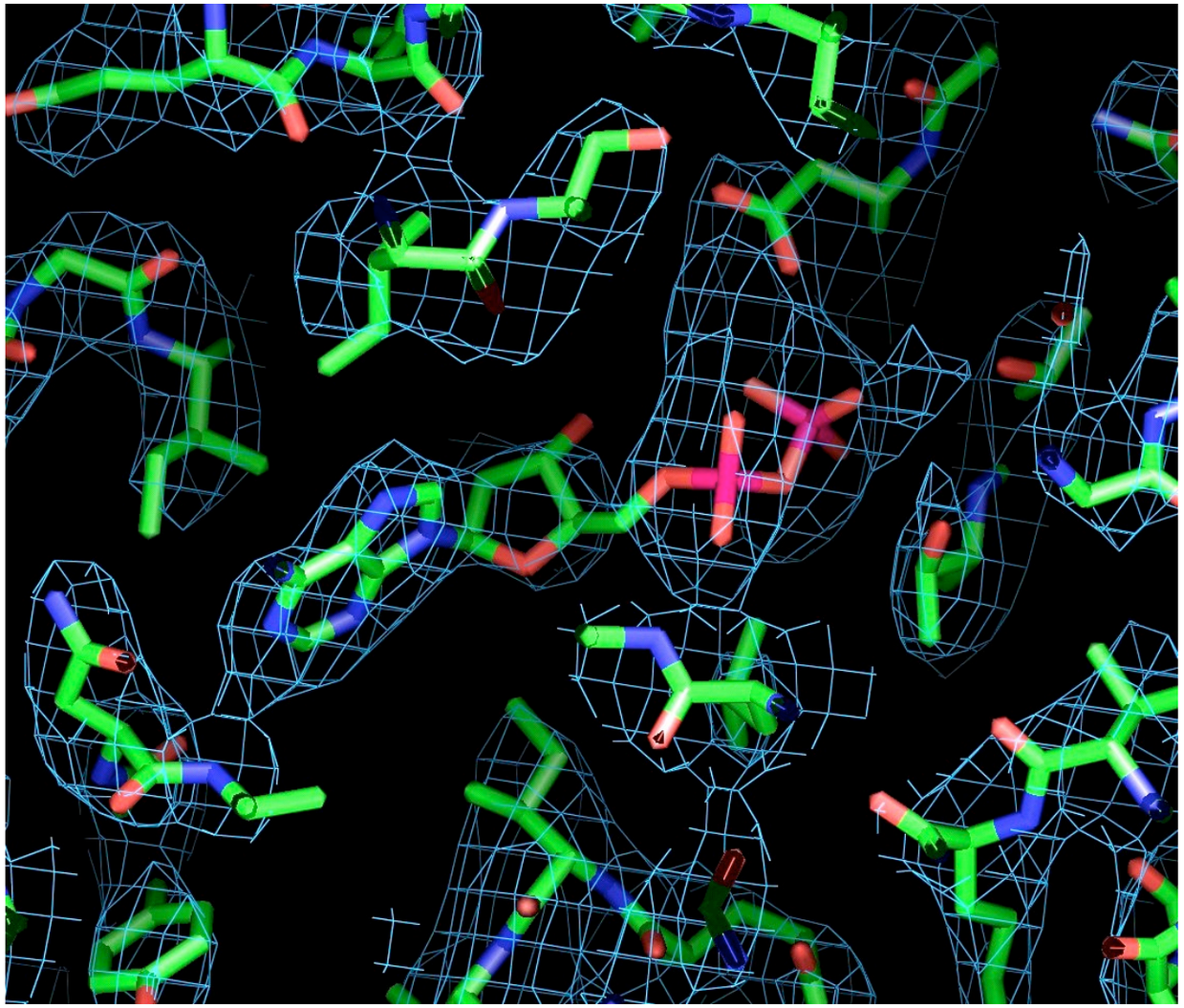
Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Starting from Random Phases



Density for GroES



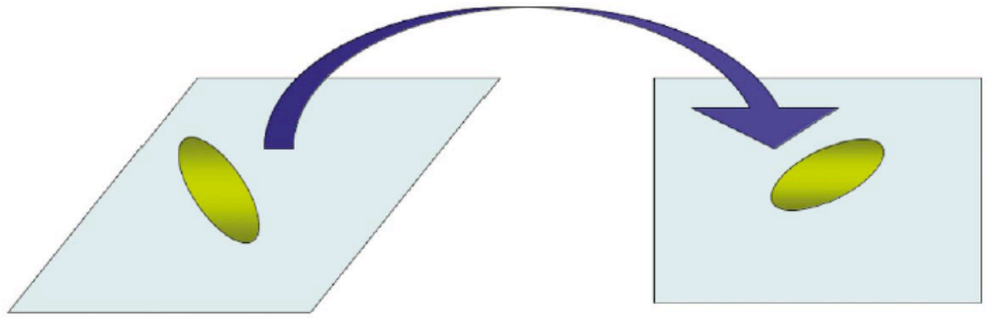
Nucleotide in Active Site

- The constraints imposed by the NCS are very powerful (there are very limited solutions for the phases)

BERKELEY LAB

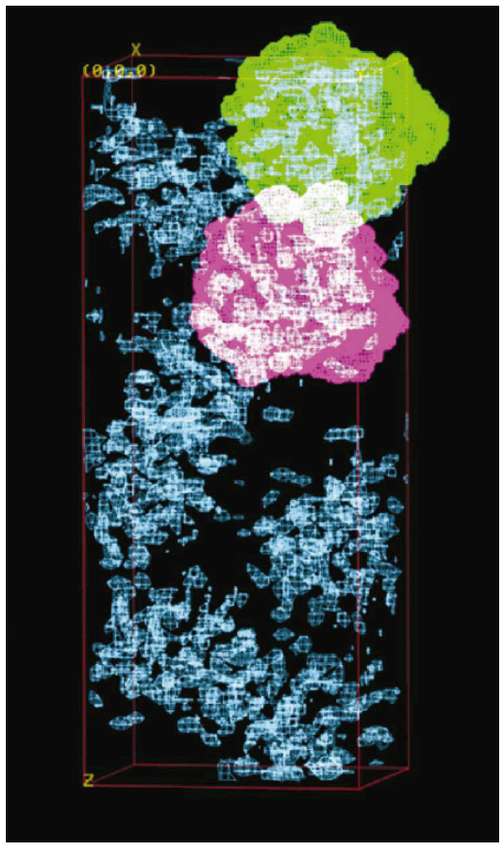
Phenix

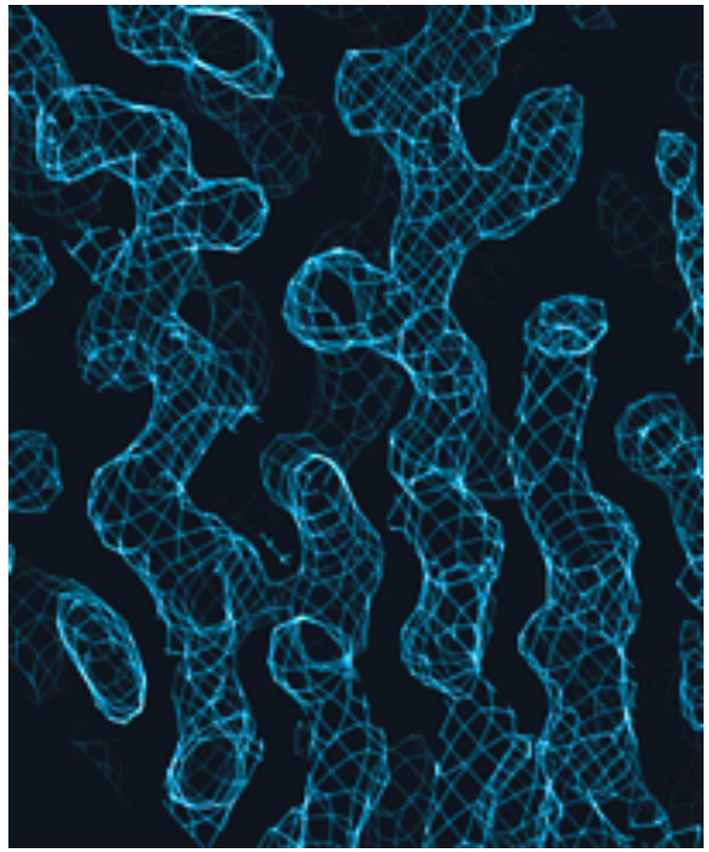
Multi-crystal Averaging



Find **R** and **t** that transform the molecule from *A* to *B*

Cross-crystal average and phase extend (*DMMULTI*)





Bootstrap from 6Å to 2Å

- Using the information from multiple crystals can be very powerful:
  - The different crystals sample the molecular transform in different places
  - With many different crystals this approaches direct recovery of the molecular transform
- The application of the method is not straight forward
  - Relationships between the different molecules need to be found

BERKELEY LAB

Phenix

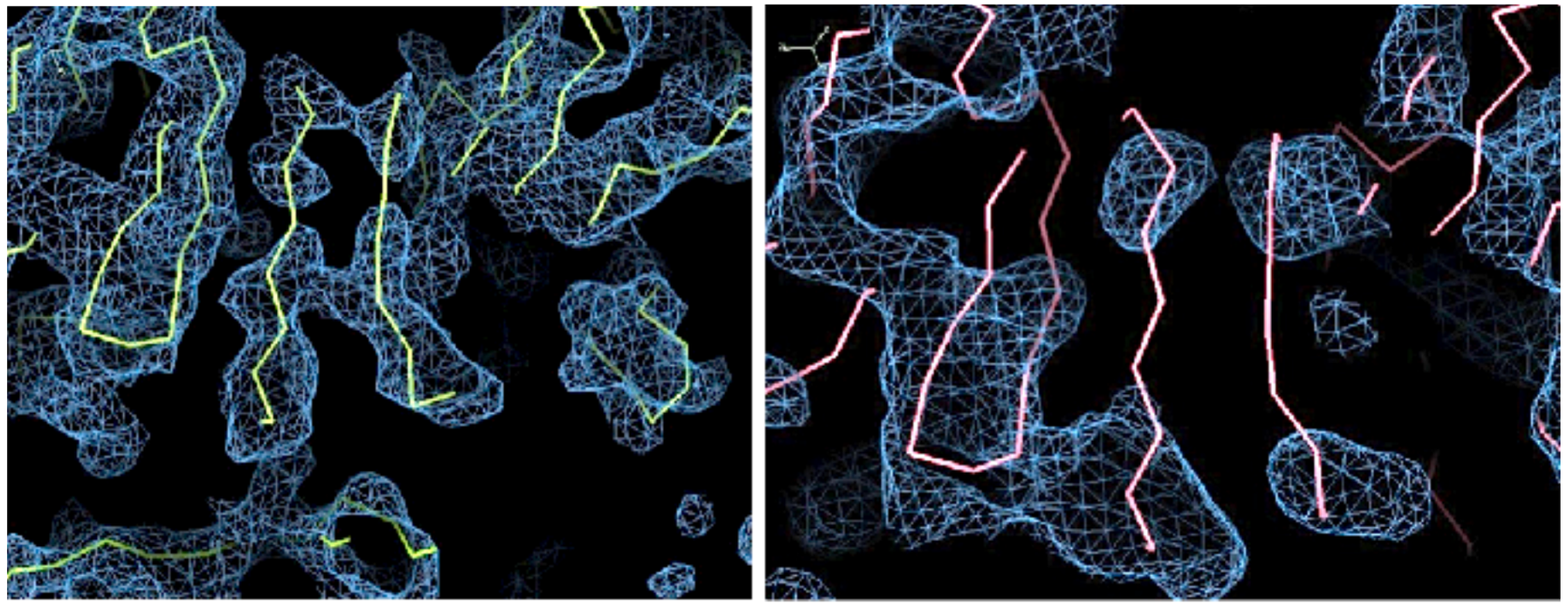
Image from G. Taylor, Acta Cryst. D, 59, 1881-1890 (2003)

Statistical Density Modification with Cross-Crystal Averaging

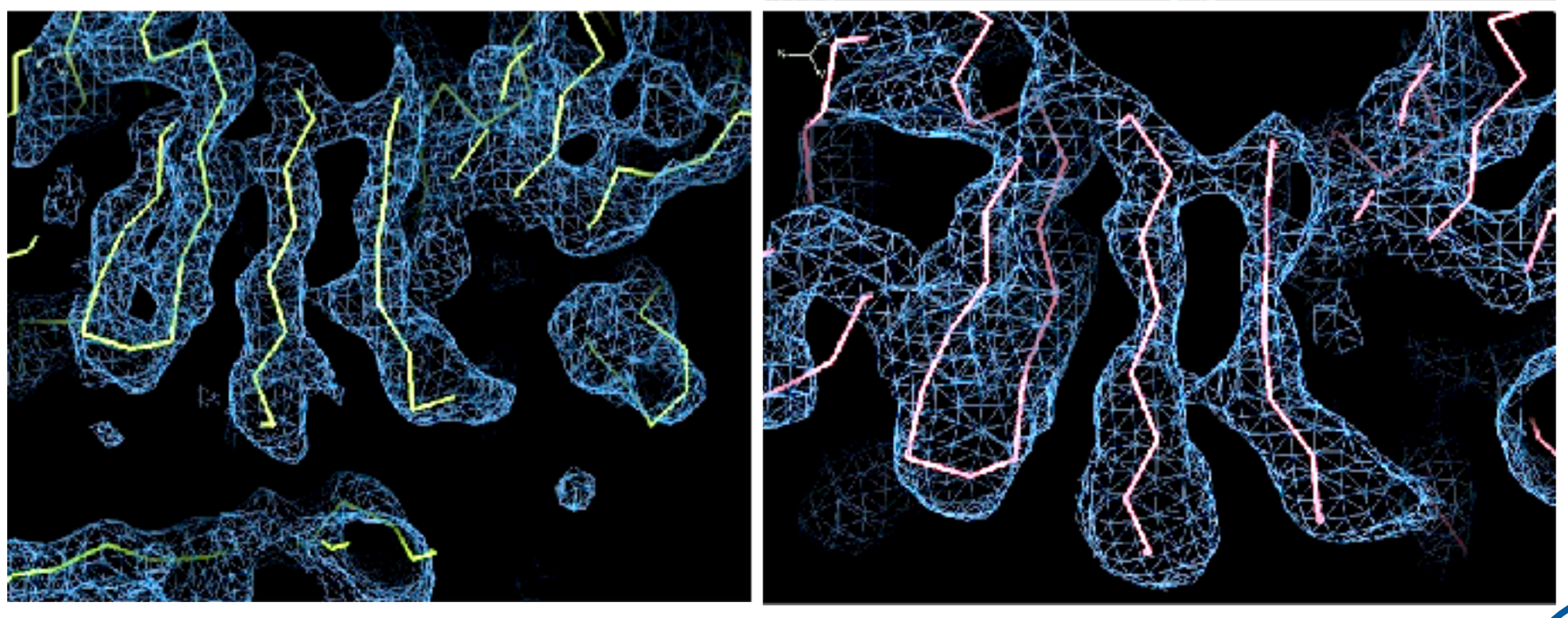
Crystal 1 (4 copies)

Crystal 2 (2 copies)

Single crystal statistical density modification



Cross-crystal statistical density modification



Phenix

Cell receptor at 3.5/3.7 Å. Data courtesy of J. Zhu

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