Automation of Structure Determination by Macromolecular Crystallography

Tom Terwilliger
Los Alamos National Laboratory

Automation of structure determination

Automation…

makes straightforward cases accessible to a wider group of structural biologists
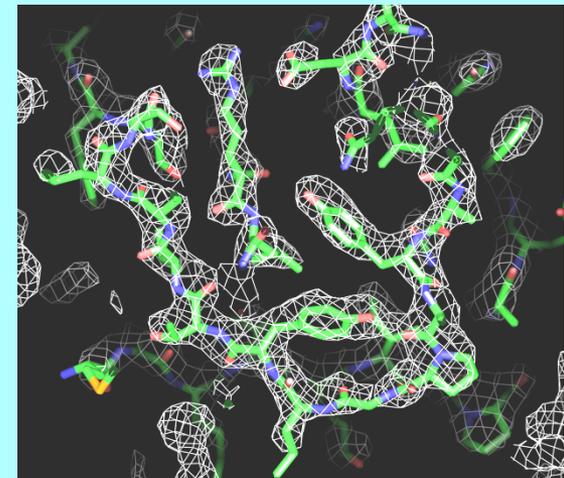
makes difficult cases more feasible for experts

can speed up the process

can help reduce errors

Automation also allows you to…

try more possibilities

estimate uncertainties

Requirements for automation of structure determination of macromolecules by X-ray crystallography
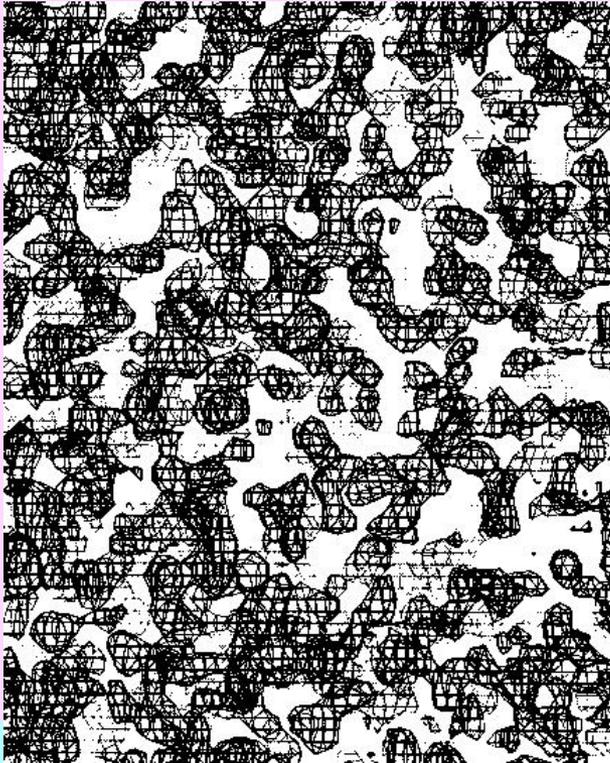
(1) Software carrying out individual steps

(2) Seamless connections between steps

(3) A way to decide what is good

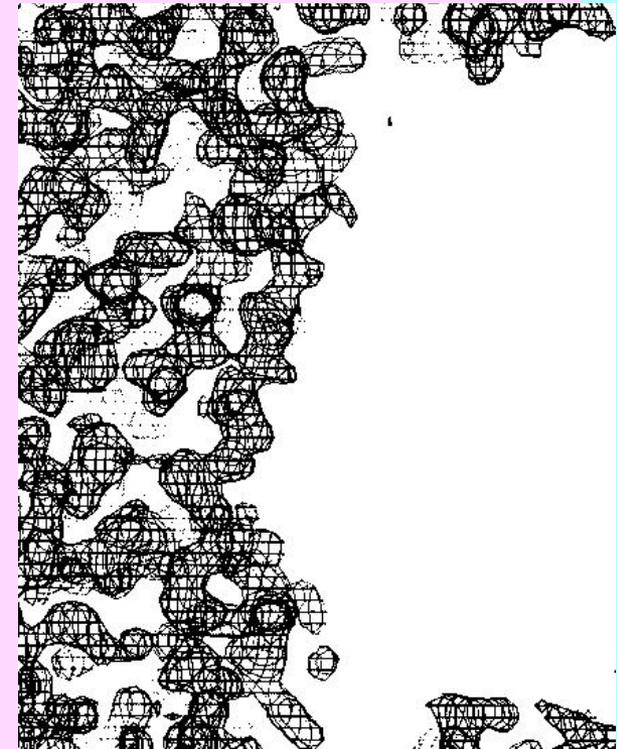(4) Strategies for structure determination and decision-making
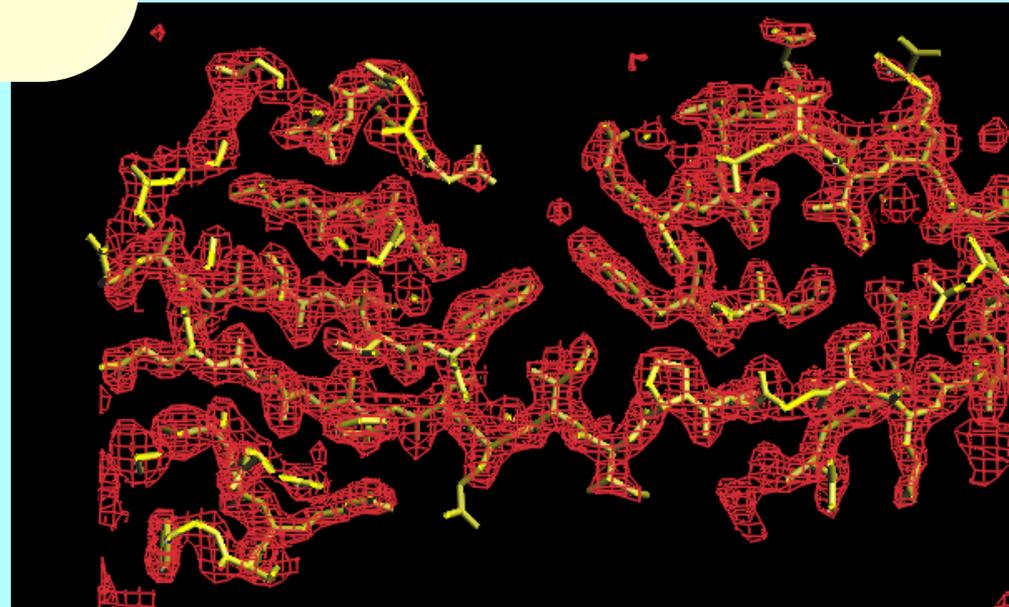
Why we need good measures of the quality of an electron-density map:

Which solution is best?
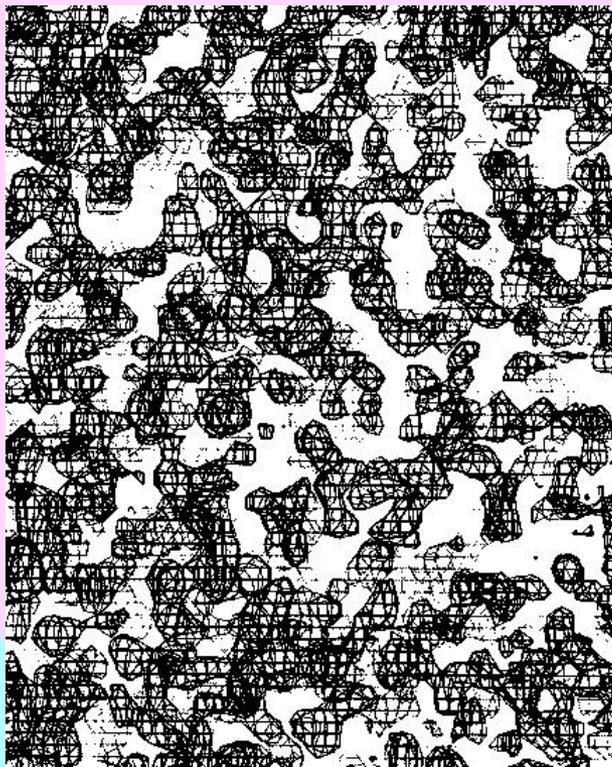
Are we on the right track?
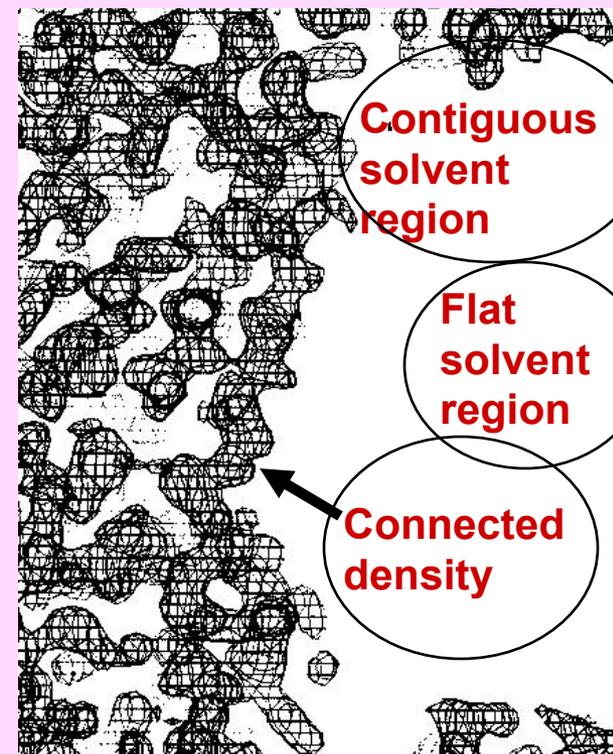
If map is good:

It is easy

Why we need good measures of the quality of an electron-density map:
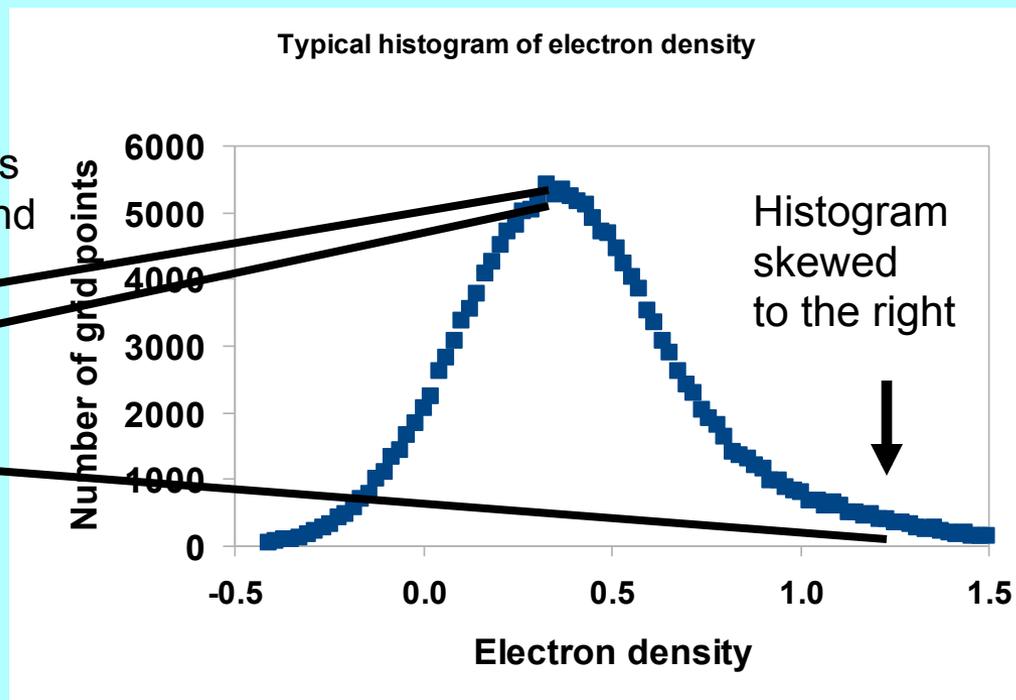
Which solution is best?

Are we on the right track?
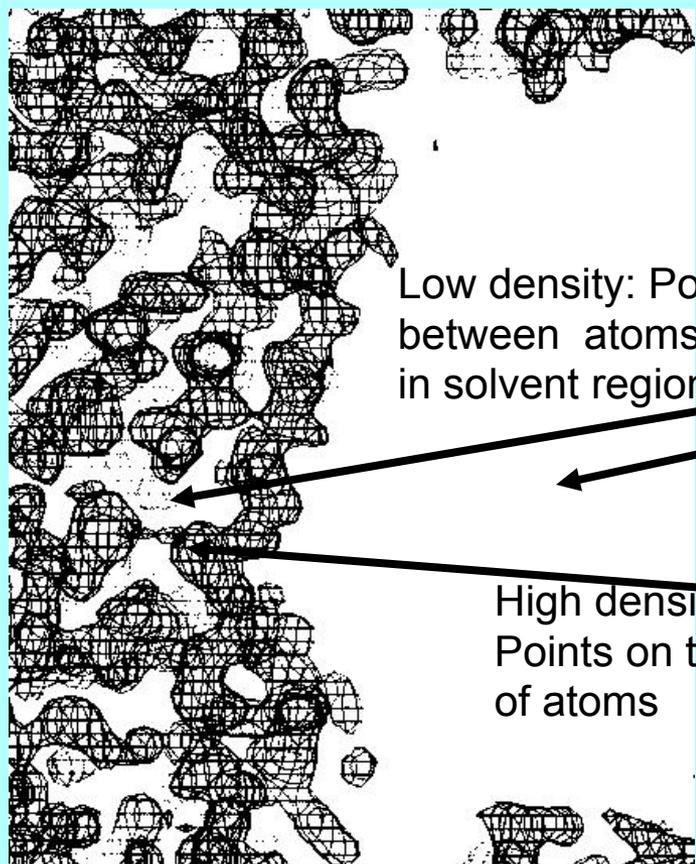

If map is good:

It is easy

Contiguous solvent region

Flat solvent region

Connected density

Histogram of electron density values has a positive "skew"

Low density: Points between atoms and in solvent region

High density: Points on top of atoms

Typical histogram of electron density

Histogram skewed to the right

Number of grid points

6000
5000
4000
3000
2000
1000
0

-0.5   0.0   0.5   1.0   1.5

Electron density

Skew of electron density for poor and good maps
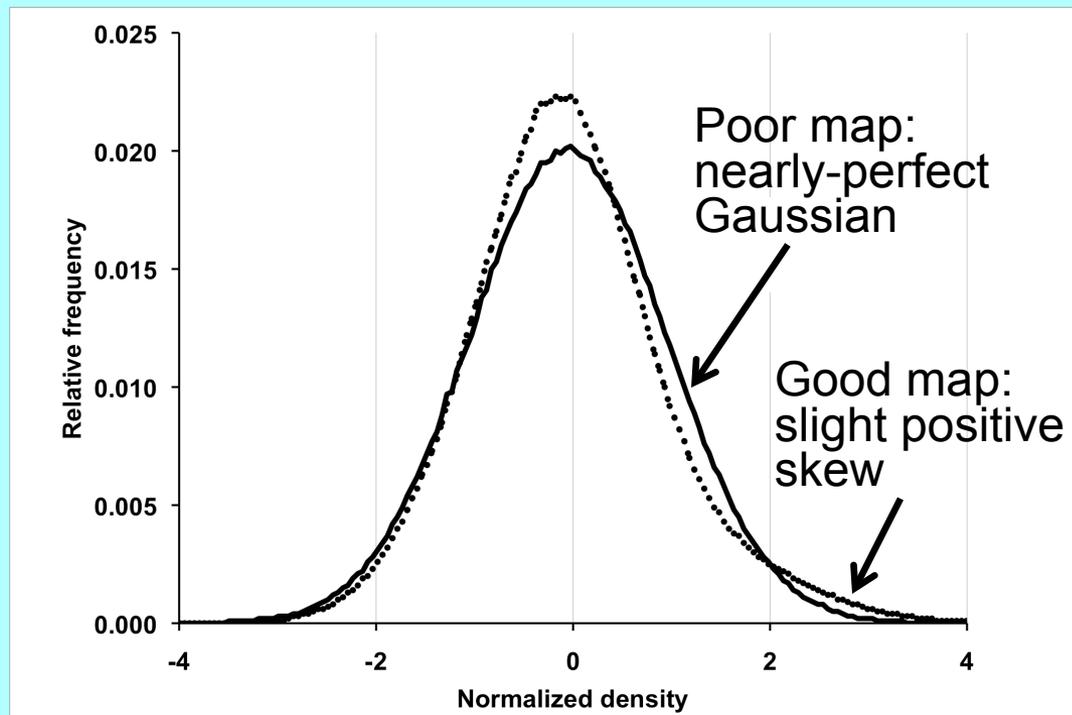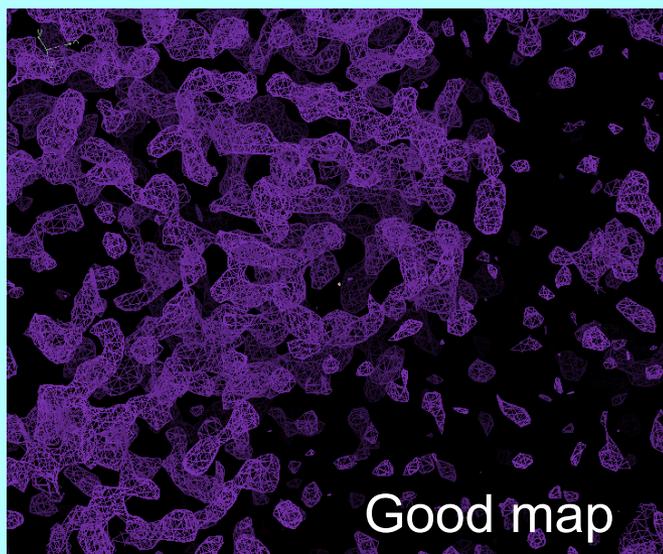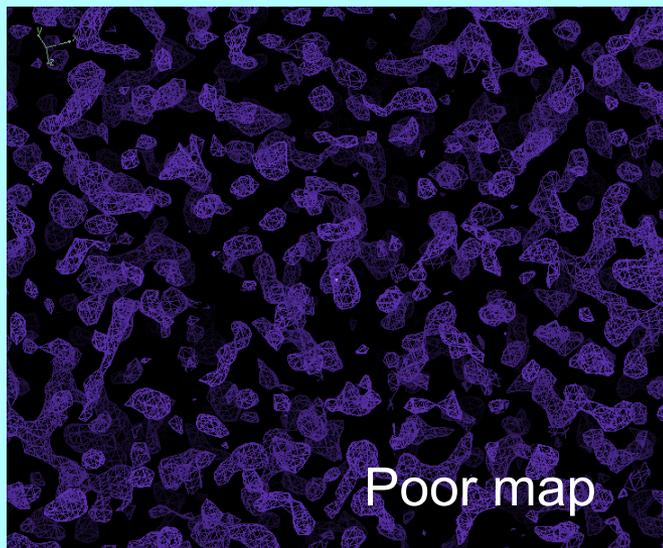
Evaluating electron density maps

Basis	Good map	Random map
Skew of density (Podjarny, 1977)	Highly skewed (very positive at positions of atoms, zero elsewhere)	Gaussian histogram
Connectivity of regions of high density (Baker, Krukowski, & Agard, 1993)	A few connected regions can trace entire molecule	Many very short connected regions
Correlation of local rms densities (Terwilliger, 1999)	Neighboring regions in map have similar rms densities	Map has uniform rms density
R-factor in 1st cycle of density modification (Cowtan, 1996)	Low R-factor	High R-factor

Which scoring criteria best reflect the quality of a map?

Create real maps

Score the maps with each criteria

Compare the scores with the actual quality of the maps

Creating real maps

247 MAD, SAD, MIR datasets with final model available
(PHENIX library and JCSG publicly-available data)

Run AutoSol Wizard on each dataset.

Calculate maps for each solution considered
(opposing hands, additional sites, including various derivatives
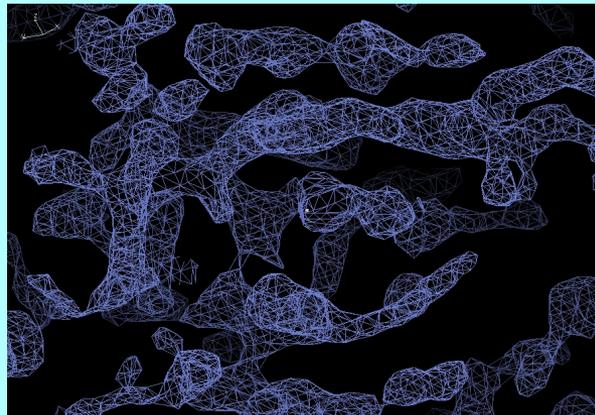for MIR)

Score maps based on each criteria

Calculate map correlation coefficient (CC) to model map
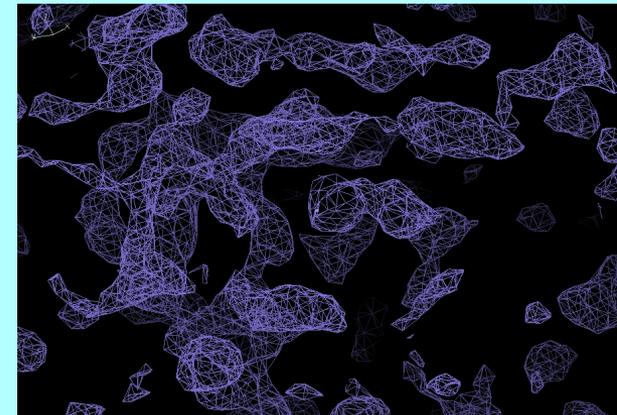(no density modification, shift origin if necessary)
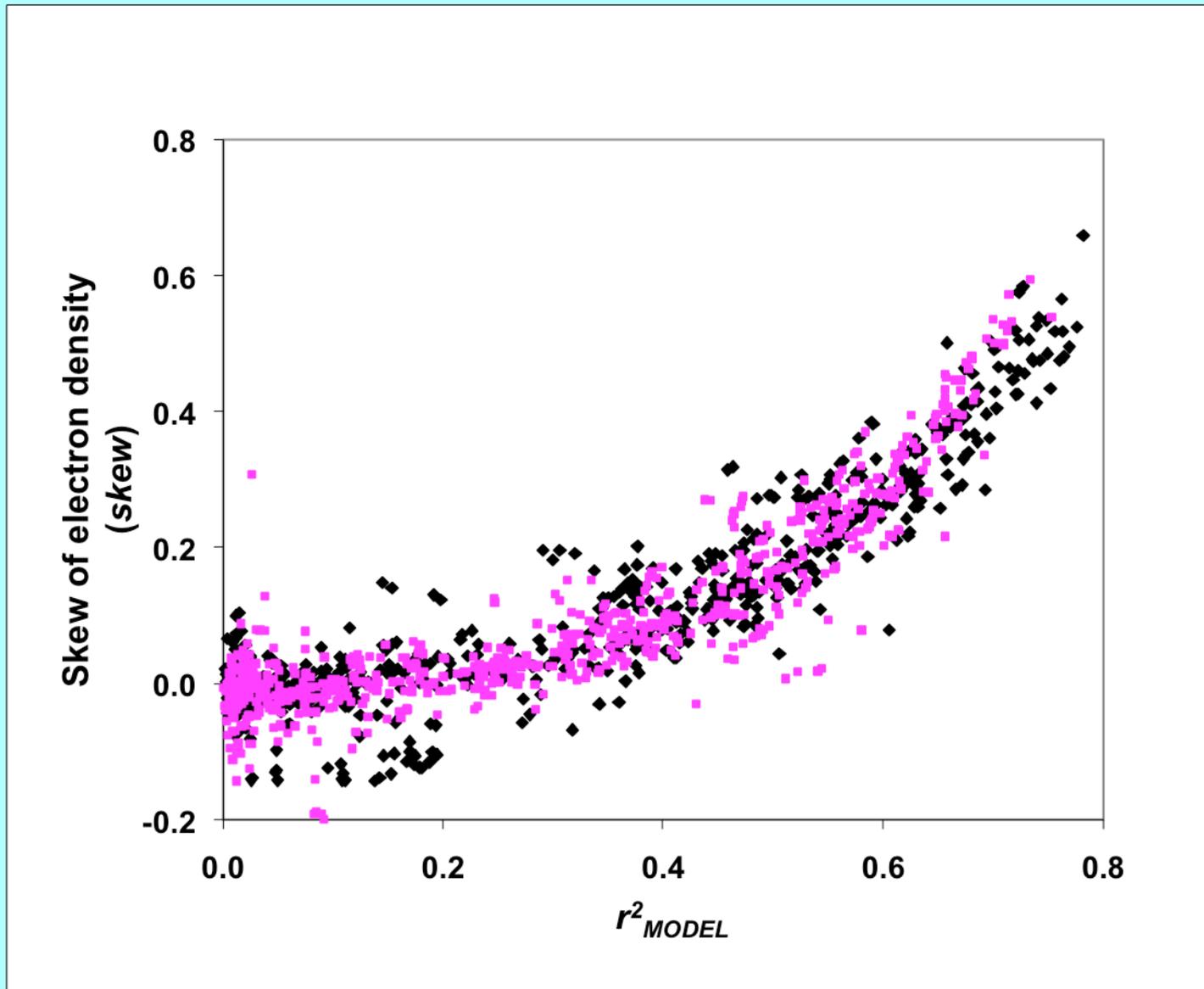
Model map
1VQB, 2.6 Å, SG C2
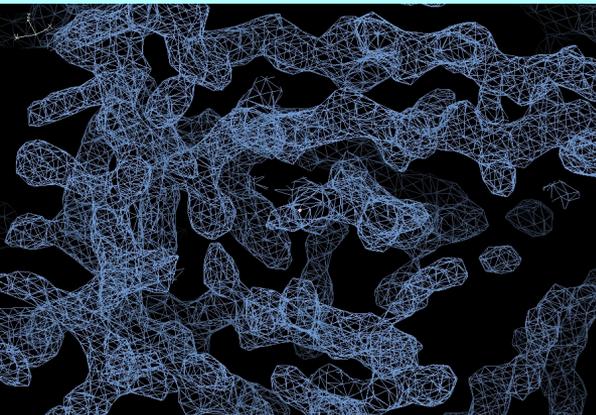
SOLVE MAD map
CC=0.62

Inverse-hand map
 CC=0.55

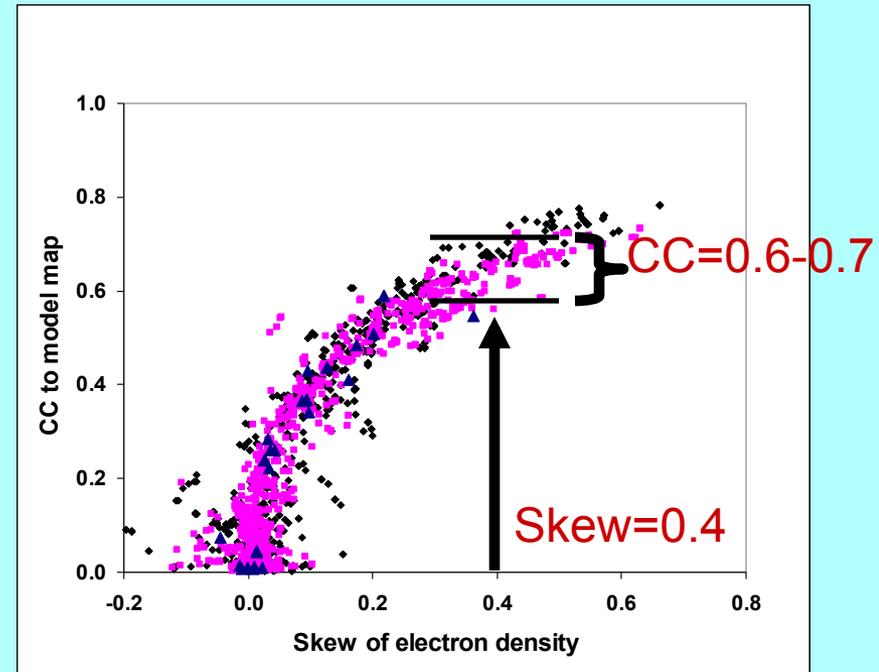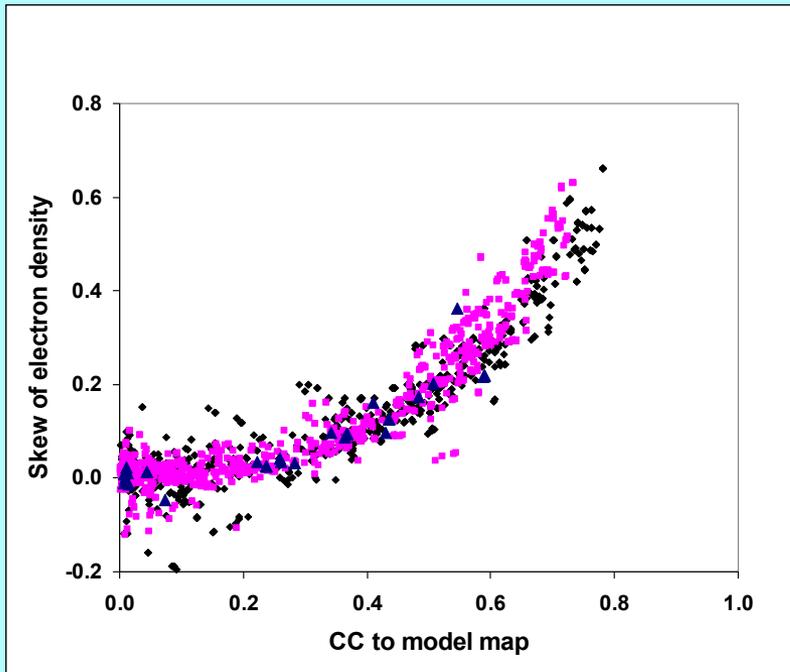# Skew of electron density – positive skew of density values

Using scoring criteria to estimate the quality of a map
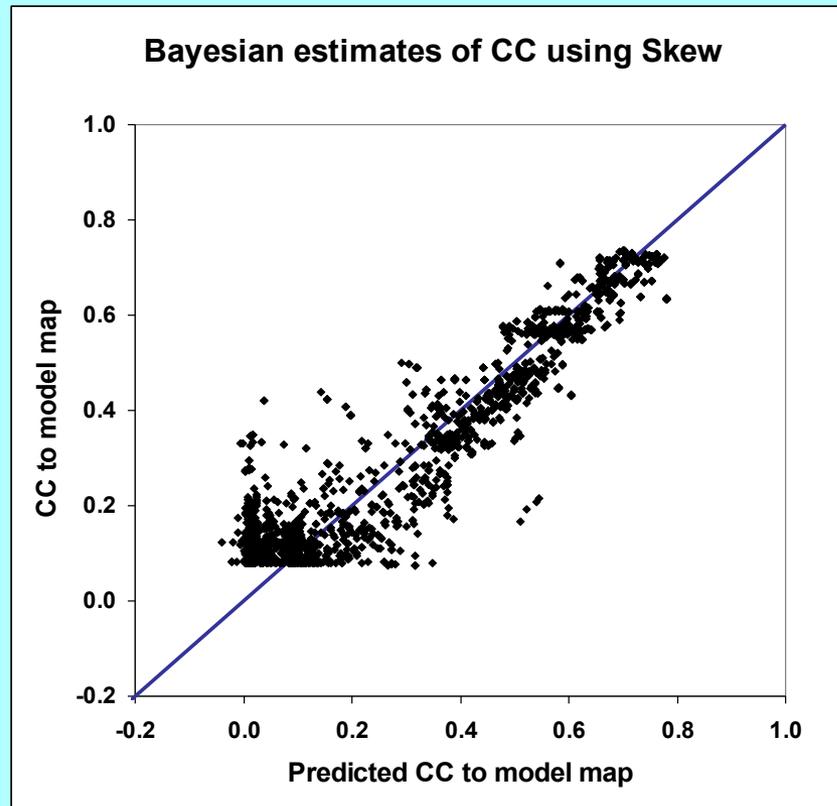

Skew depends on CC → Estimate CC from skew

How accurate are estimates of map quality?

Actual quality

Bayesian estimates of CC using Skew

Estimated quality

Cross-validated estimates of quality

Estimated map quality in practice
Evaluating solutions to a 2-wavelength MAD experiment
(JCSG Tm3681, 1VPM, SeMet 1.6 Å data)

Data for HYSS	Sites	Estimated CC ± 2SD	Actual CC
Peak	12	0.73 ± 0.04	0.72 ←
Peak (inverse hand)	12	0.11 ± 0.43	0.04
F_A	12	0.73 ± 0.03	0.72
F_A (inverse)	12	0.11 ± 0.42	0.04
Sites from diff Fourier	9	0.70 ± 0.17	0.69

Structure solution with phenix.autosol

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

<u>Decisions to be made</u>

Find heavy-atom sites with direct methods (HYSS)

Multiple solutions, different derivatives or wavelengths

Calculate phases (Phaser/Solve)

Alternative hands of space-group and substructure
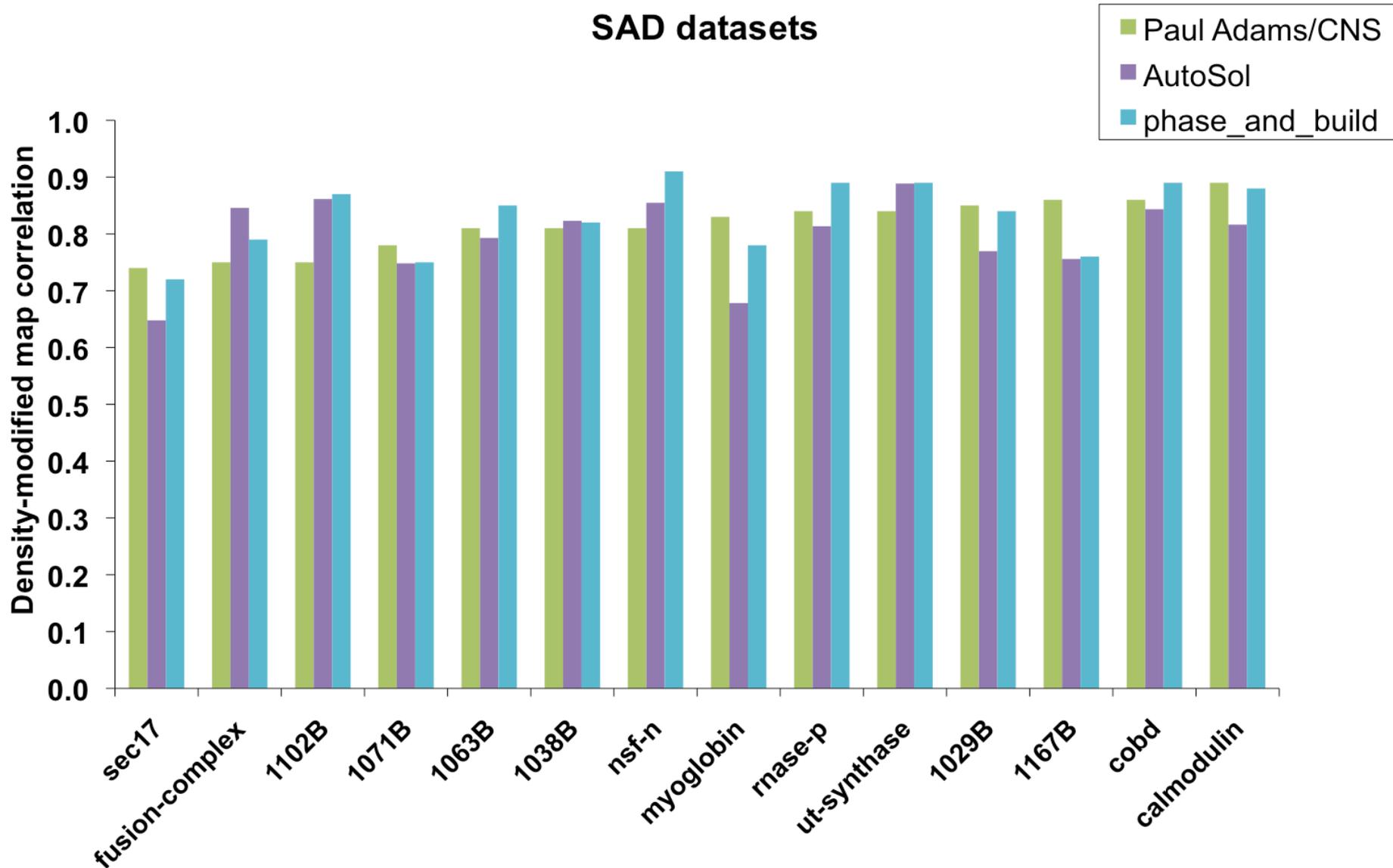
Improve phases, find NCS, build model (phase_and_build)
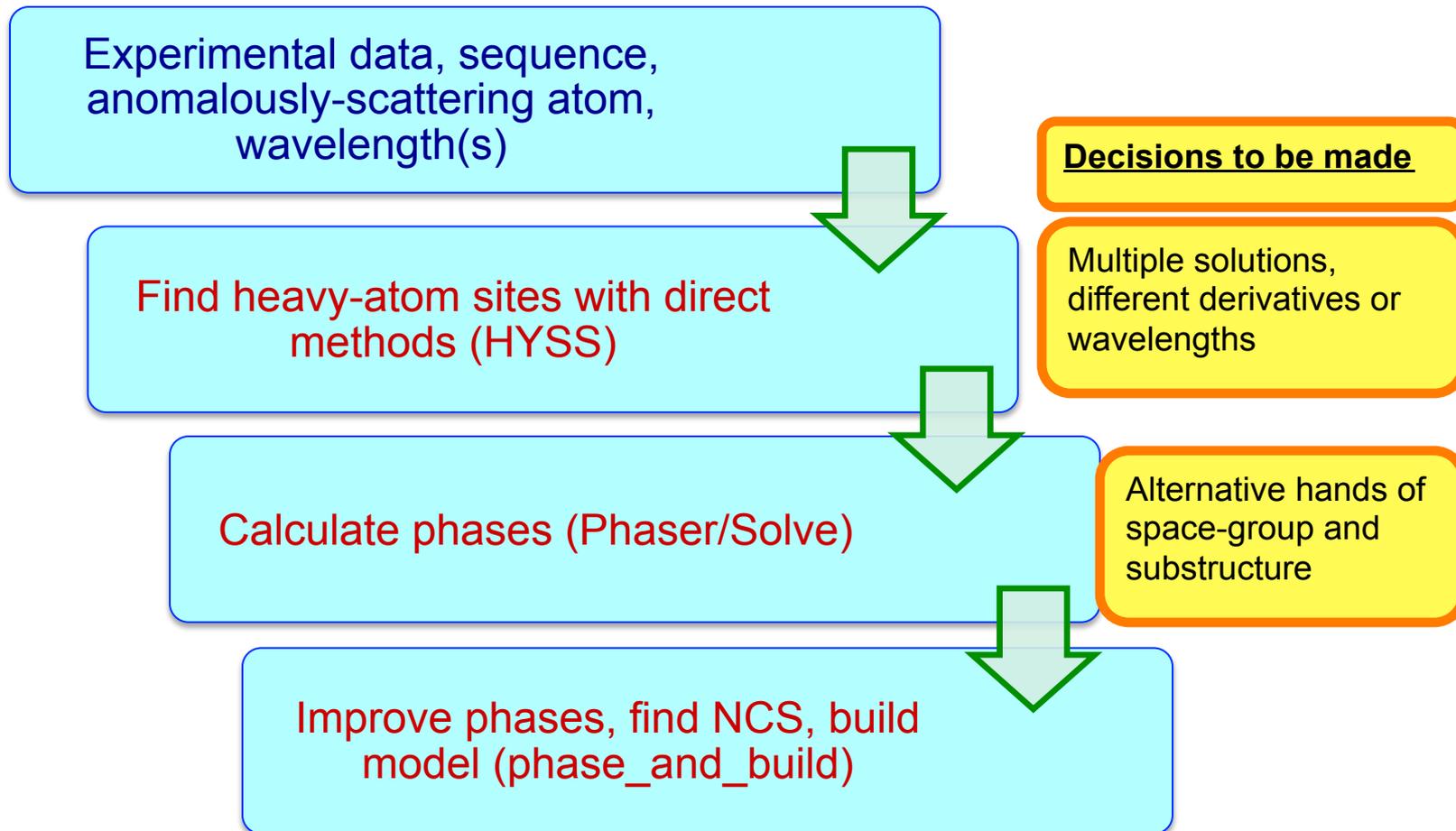
AutoSol – fully automatic tests with structure library
(MAD datasets, HYSS search, SOLVE)
RESOLVE/ phase_and_build maps
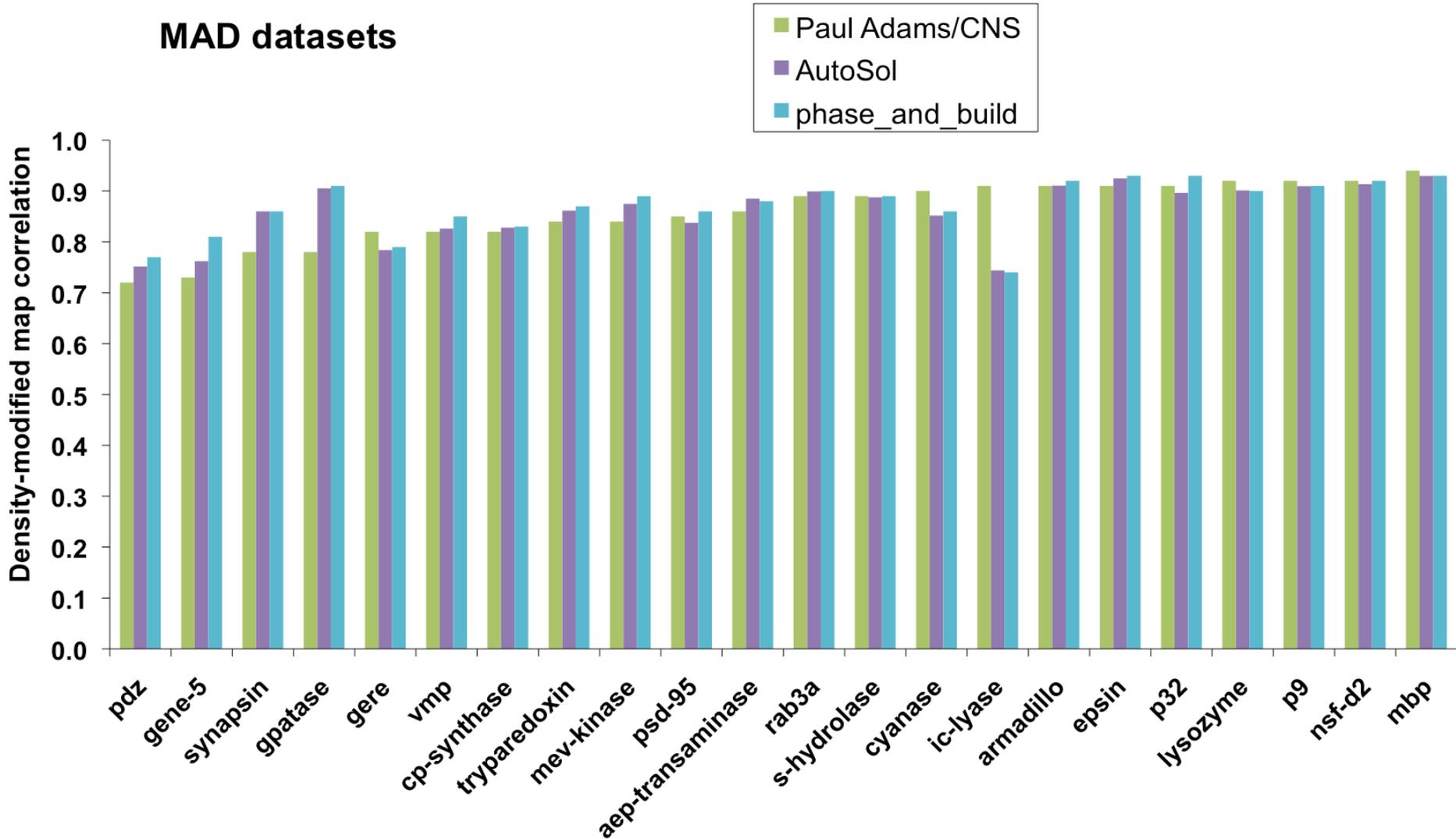
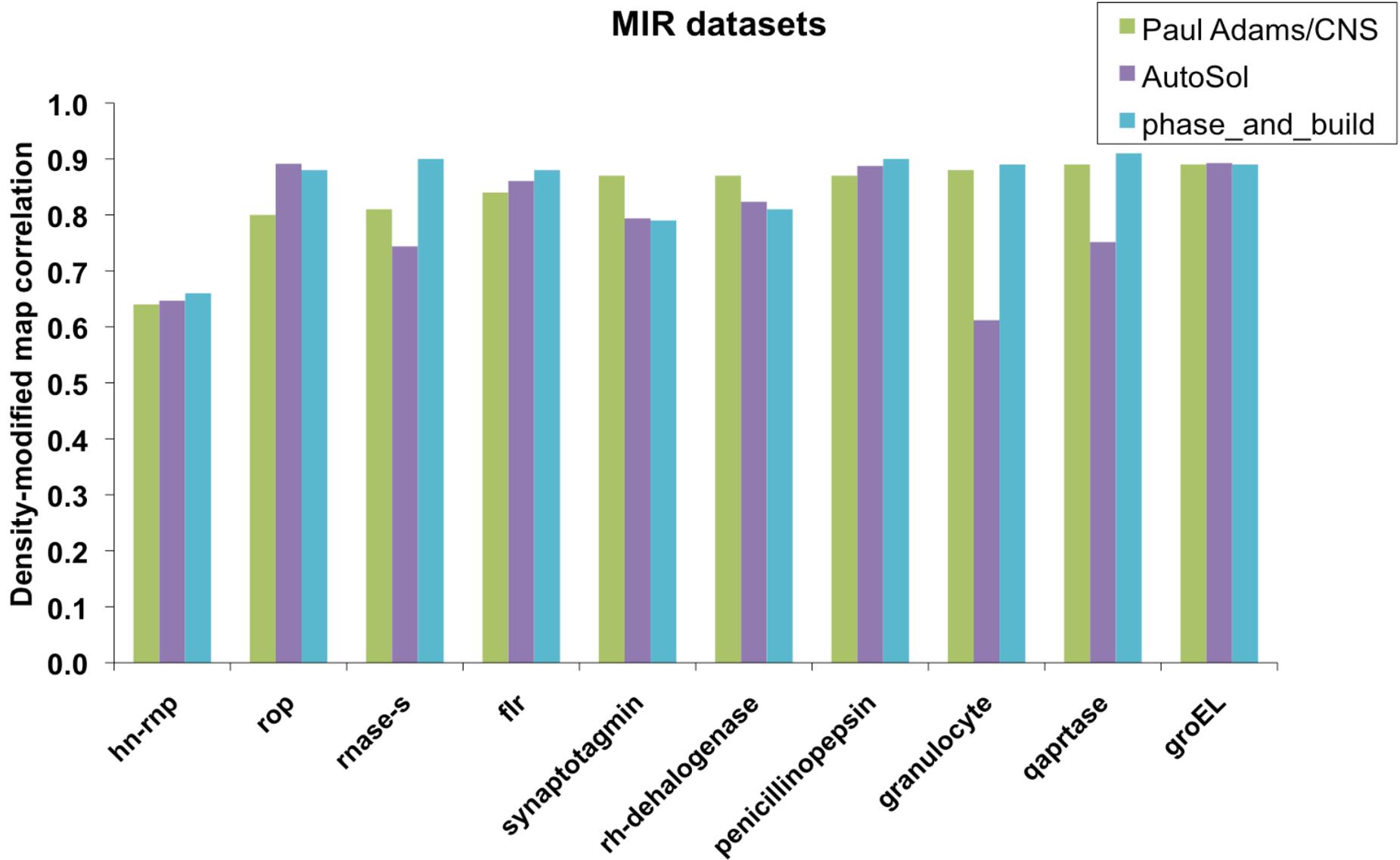AutoSol – fully automatic tests with structure library
(SAD datasets, HYSS, Phaser)
RESOLVE/ phase_and_build maps

SAD datasets

AutoSol – fully automatic tests with structure library
(SAD datasets, HYSS, Phaser)
RESOLVE/ phase_and_build maps

The PHENIX Project

Lawrence Berkeley Laboratory

Paul Adams, Ralf Grosse-Kunstleve, Pavel Afonine, Nat Echols, Nigel Moriarty, Jeff Headd, Nicholas Sauter, Peter Zwart

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Duke University

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner

Cambridge University

Jane & David Richardson, Vincent Chen, Chris Williams, Bryan Arendall, Laura Murray

An NIH/NIGMS funded Program Project