Structure solution from weak anomalous data

CCP4 Study Weekend

Nottingham, UK

Jan. 8-9, 2015

Tom Terwilliger, Li-Wei Hung (Los Alamos National Laboratory)
Gábor Bunkóczi, Randy Read (Cambridge University)
Nat Echols, Ralf Grosse-Kunstleve, Paul Adams (Lawrence Berkeley National Laboratory)
James Holton (Lawrence Berkeley National Laboratory, UCSF)
David Akey, Janet Smith (University of Michigan)

Phenix

Structure solution from weak anomalous data

Problems with weak signal

Quantifying the anomalous signal

Solving the anomalous sub-structure and phasing with weak signal

Estimating the anomalous signal in a dataset

Scaling and merging SAD data

Will I solve the anomalous sub-structure?

Problems with weak anomalous signal

Why would I have a weak anomalous signal?

Few anomalous scatterers, sulfur SAD, weak diffraction, wavelength far from peak

Why is this a problem?

Sub-structure identification is difficult

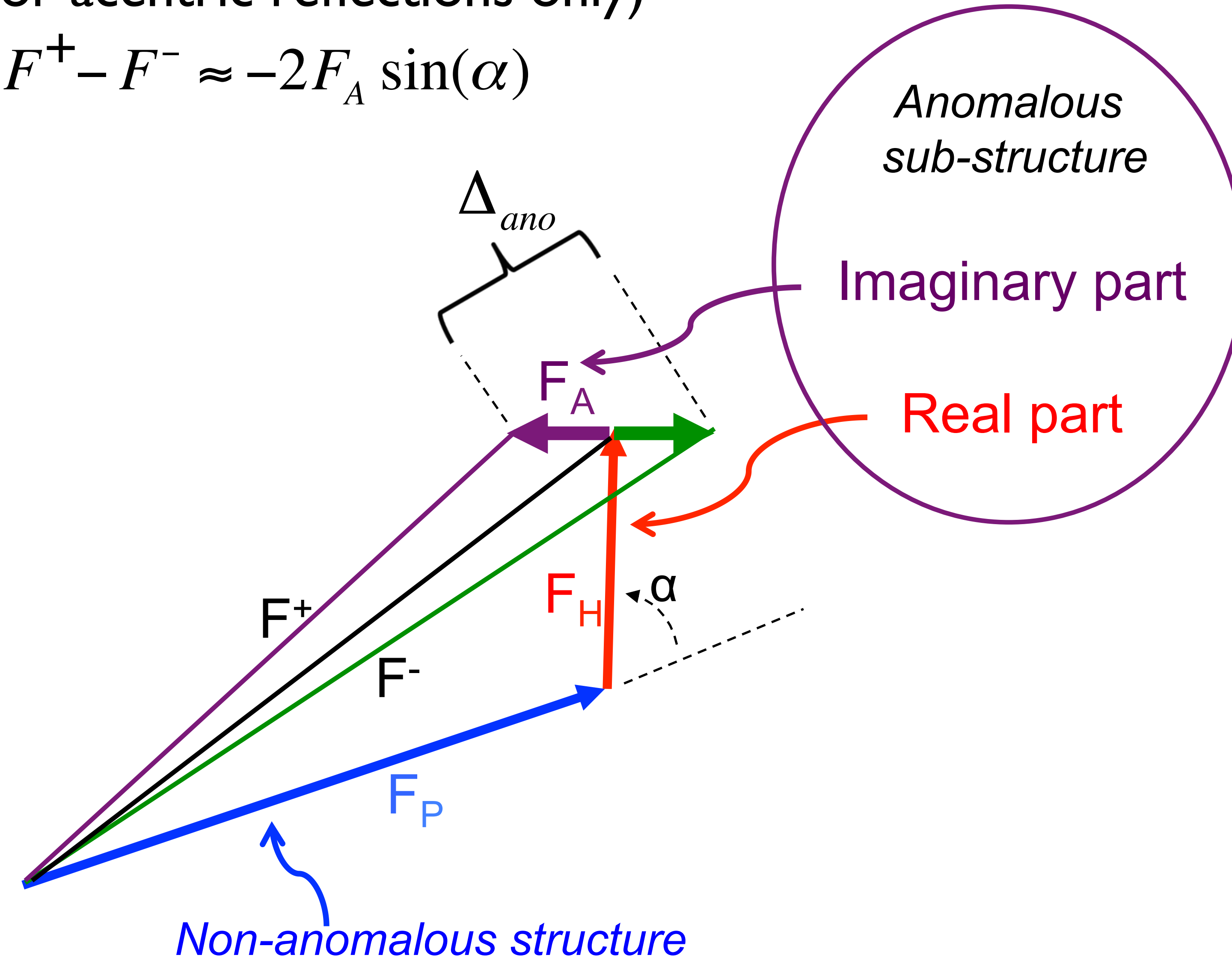
Phasing is poor

Iterative density modification, model-building and refinement works poorly

Quantifying the anomalous signal

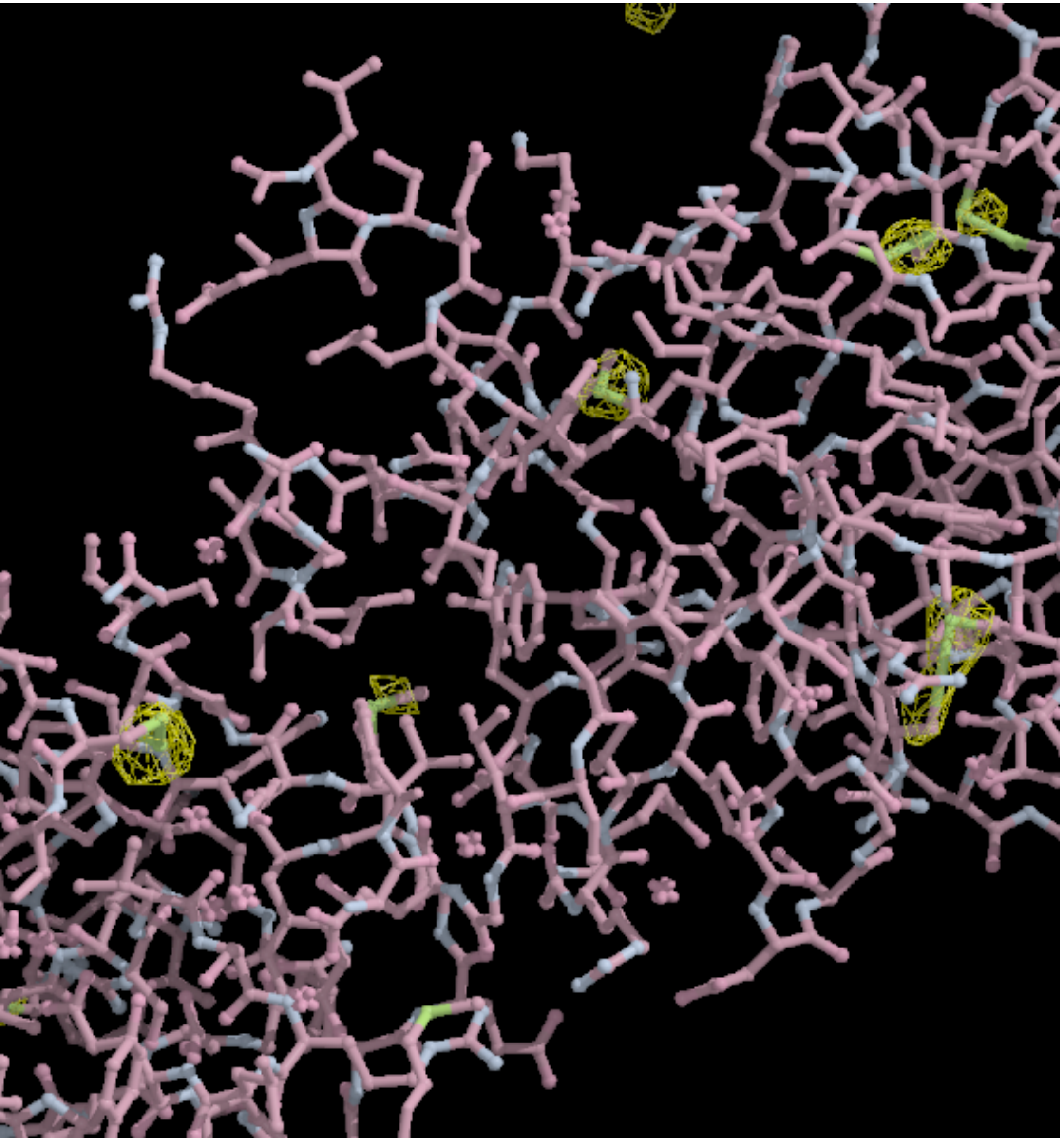
Anomalous differences
(Present for acentric reflections only)

$$\Delta_{ano} \equiv F^{+} - F^{-} \approx -2F_A \sin(\alpha)$$



Anomalous difference Fourier with observed data and model phases

$$\rho(x) = \frac{1}{V} \sum_h \Delta_{ano,h}^{obs} e^{i(\varphi_h^c - \frac{\pi}{2})} e^{-2\pi i (h \bullet x)}$$



Anomalous signal:
Peak height at coordinates of anomalously-scattering atoms

$$S_{ano} = \frac{< \rho_{ano}(x_j) >}{< \rho_{ano}^2 >^{1/2}}$$

Typical values of S_{ano} for solved datasets: 10-20

Contributions to measured anomalous differences

$$\Delta_{ano}^{obs} = \Delta_{ano} + \Delta_{ano}^{other} + \varepsilon$$

Measured

Sub-structure anomalous difference

Minor sites, C,N,O atoms...

Measurement errors

How similar are my anomalous differences to model differences?

Correlation of observed and sub-structure anomalous differences

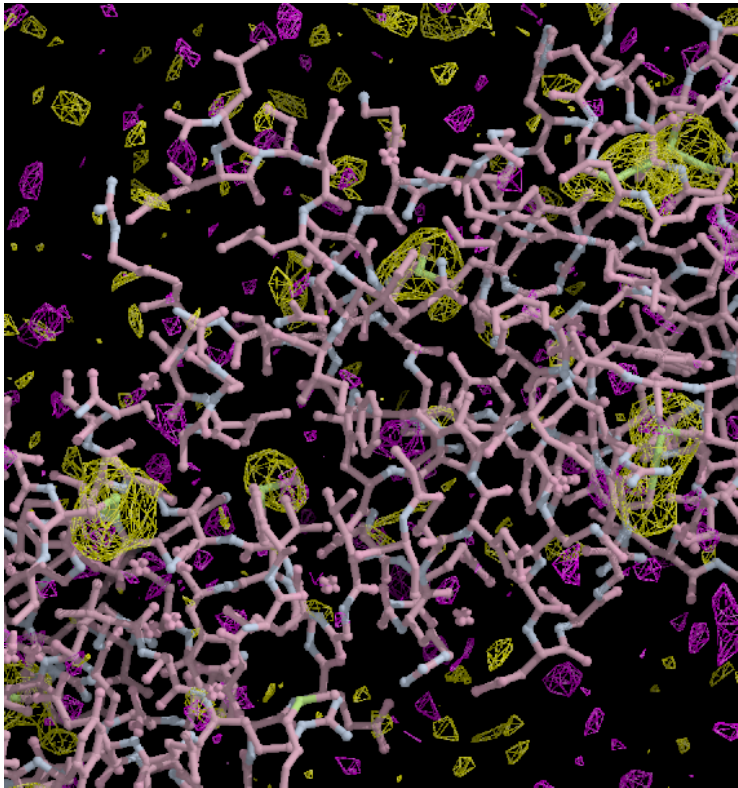
$$CC_{ano} \equiv \frac{<\Delta_{ano,j}\Delta^{obs}_{ano,j}>}{<\Delta_{ano}>^{1/2}<\Delta^{2,obs}_{ano}>^{1/2}}$$

Expected value of CC_{ano}

$$<CC_{ano}> = \frac{rms(\Delta_{ano})}{rms(\Delta^{obs}_{ano})}$$

CC_{ano} indicates how much of each anomalous difference is useful (on average)

Anomalous difference Fourier map at 2.5 σ (with noise)



How big will my anomalous signal be?

(Based on our simple model for anomalous differences)

Expected value of anomalous signal S_{ano}

$$< S_{ano} >= CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$

f is 2nd moment of the anomalous scattering factor

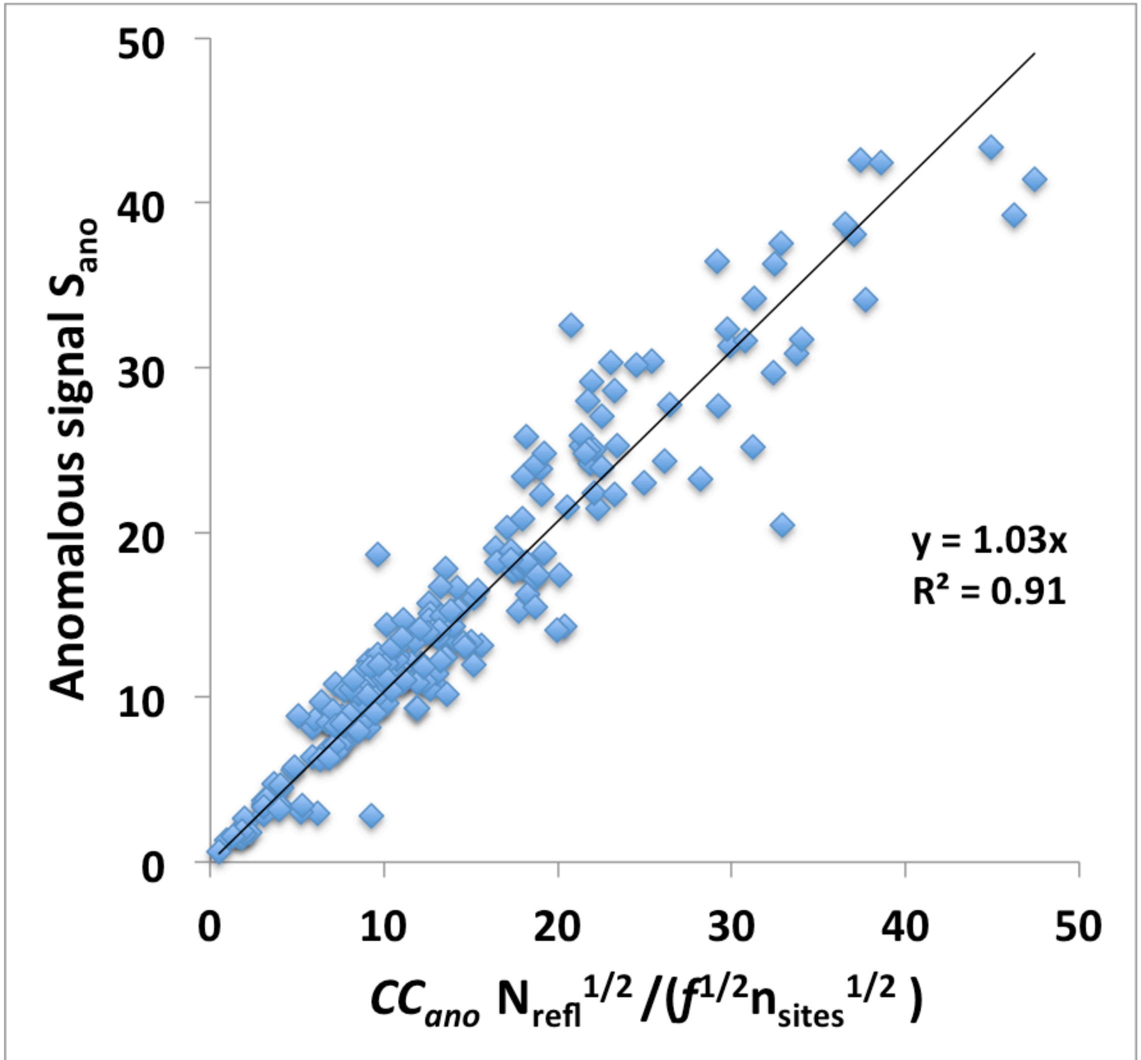
$$f = \frac{< (f^h)^2 >}{< f^h >^2}$$

Anomalous scattering factor

$$f^h \equiv f'' e^{-B\ (\sin^2 \theta_h / \lambda^2)}$$

Perfect data (20,000 reflections, 8 sites): S_{ano} = $(20000/8)^{1/2}$ = 50
Good data (overall CC_{ano}=0.36 f=2.0): S_{ano} = 12.6

Checking our simple model for anomalous signal



$$< S_{ano} >= CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$

CC_{ano}: Correlation of anomalous differences with model differences

S_{ano}: Peak height in model-phased difference Fourier

215 SAD datasets 1.2 – 4.5 Å

What affects the anomalous signal?

Anomalous correlation

Number of reflections

Anomalous signal

$$< S_{ano} > = CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$

Number of sites

B-value for anomalous sub-structure

Solving the anomalous sub-structure with weak signal

The SAD likelihood function

The likelihood of measuring the observed anomalous data given a partial model

Most powerful source of information about the sub-structure before phases are known

Using the SAD likelihood function to find the anomalous sub-structure

Start with guess about the anomalous sub-structure

From anomalous difference Patterson

Random

Any other source

Find additional sites that increase the likelihood

LLG completion based on log-likelihood gradient maps*

Iterative addition of sites

Related to using an anomalous difference Fourier—but 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.

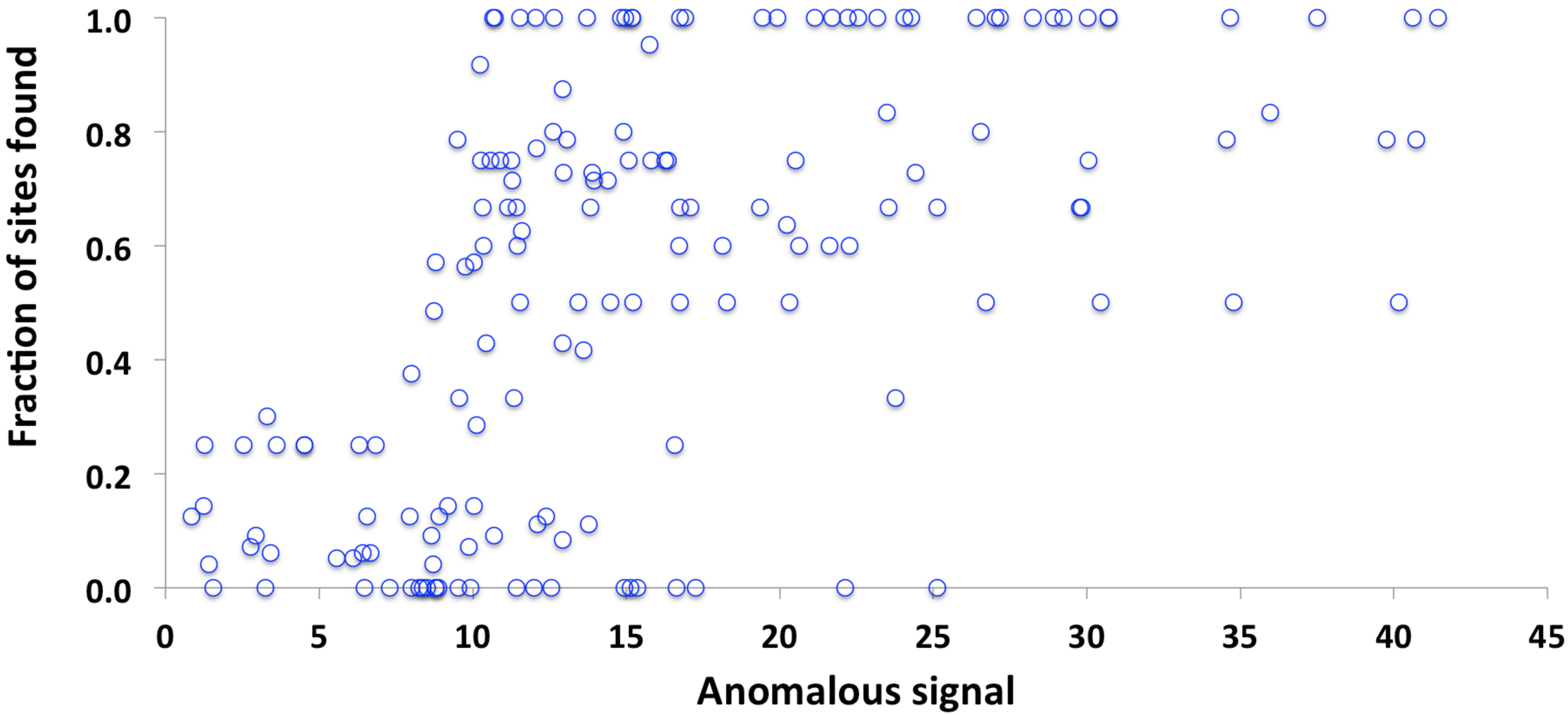
LLG sub-structure searches in HySS

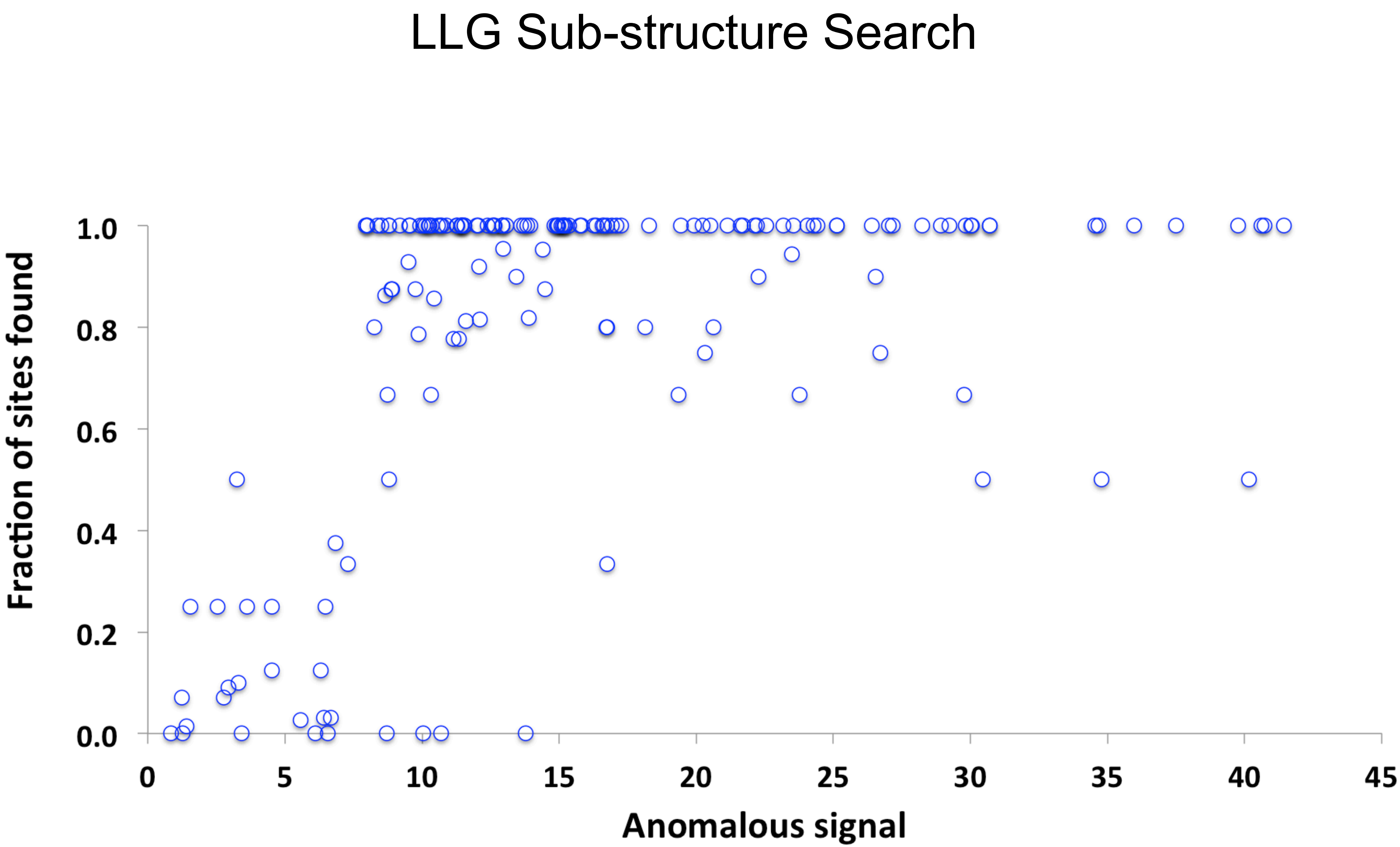
Test cases

164 SAD datasets from PDB (largely JCSG MAD data)

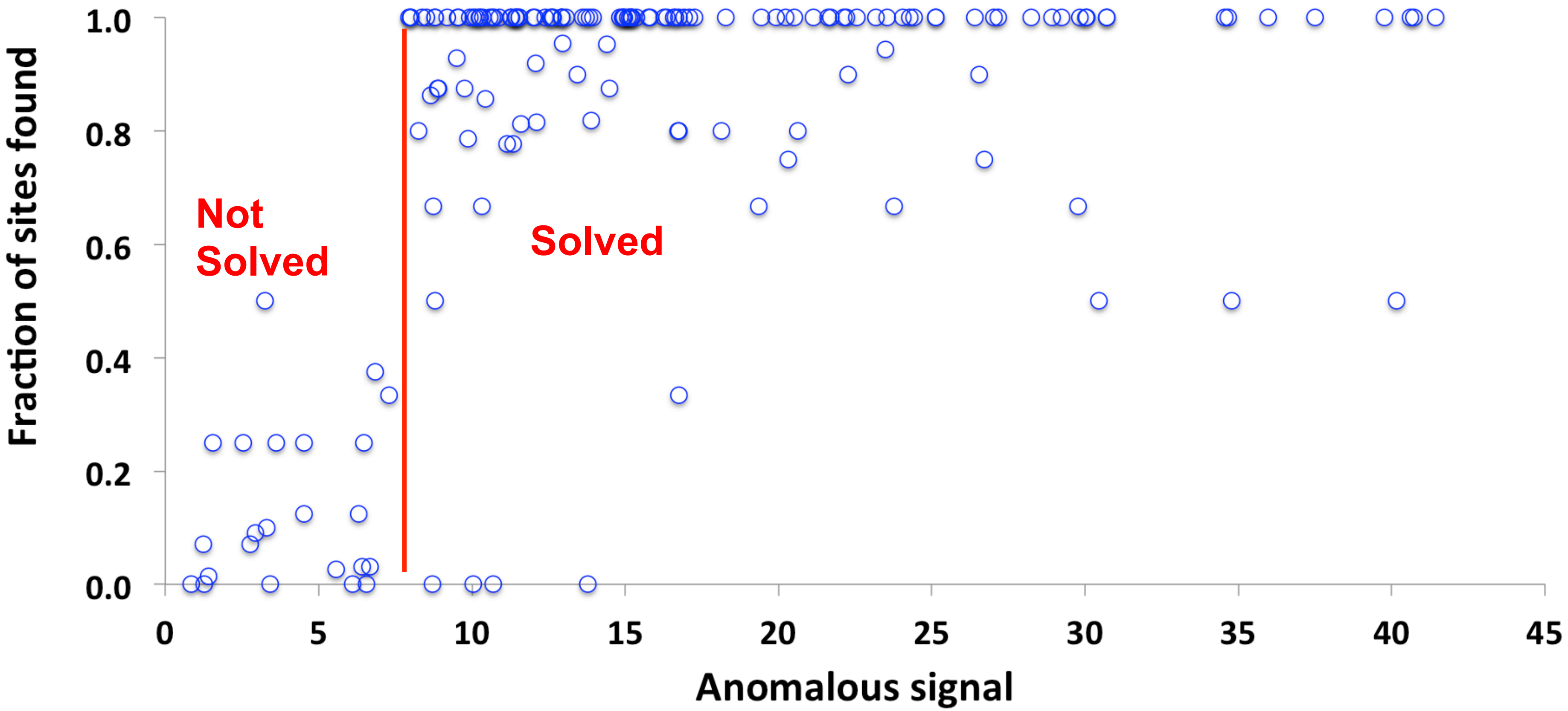
Using peak, remotes, inflection as available to include data with low anomalous signal

Dual Space Sub-structure Completion




LLG Sub-structure Search
Fraction of sites found
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
35
40
45
Anomalous signal

Anomalous signal indicates if a dataset can be solved



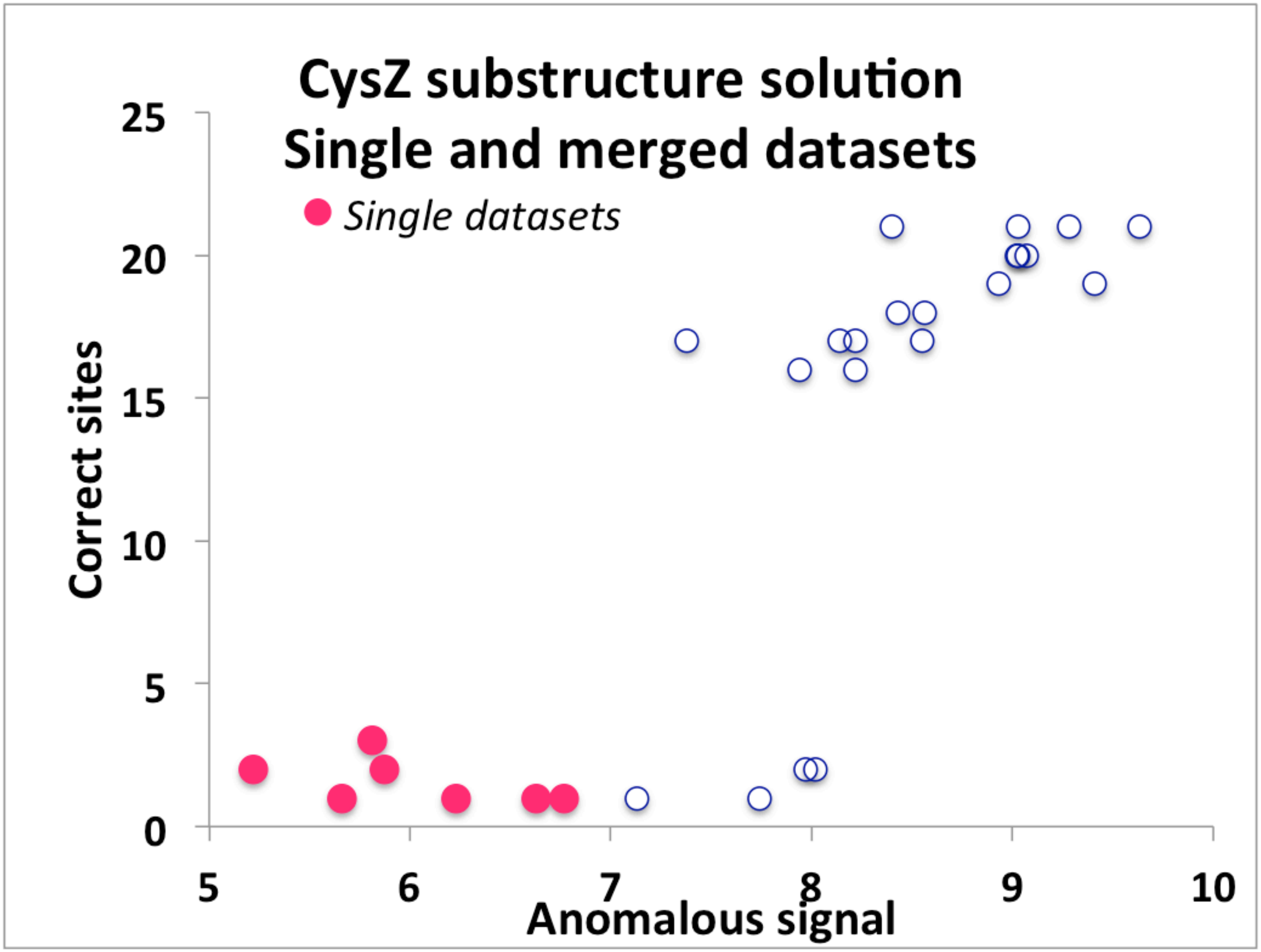
CysZ multi-crystal sulfur-SAD data

Qun Liu, Tassadite Dahmane, Zhen Zhang, Zahra Assur, Julia Brasch, Lawrence Shapiro, Filippo Mancia, Wayne Hendrickson (2012). Science 336, 1033-1037

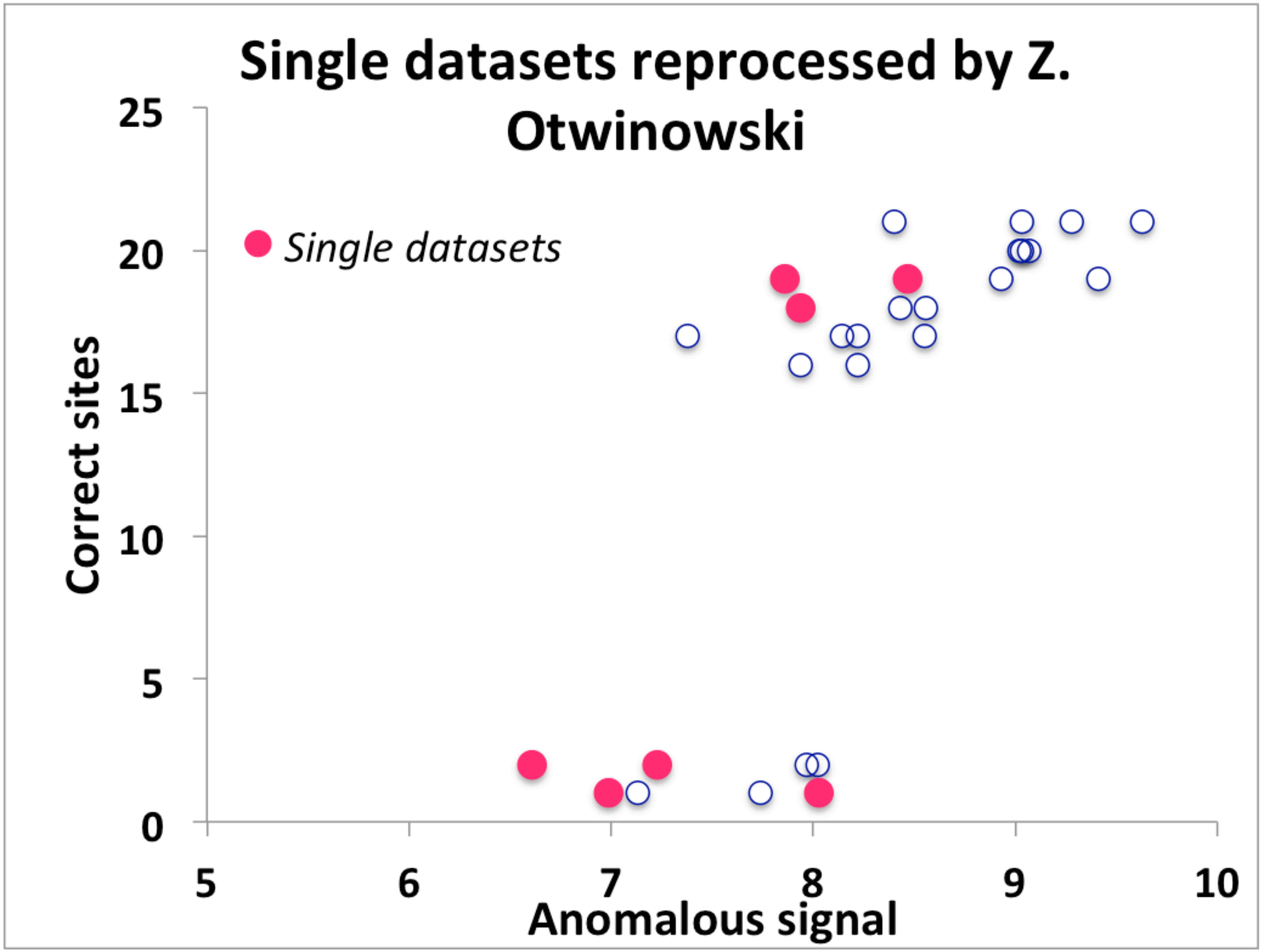
Data from 7 crystals collected at wavelength of 1.74 Å to resolution of 2.3 Å

Can anomalous signal tell us which merged datasets will be solved?

CysZ multi-crystal sulfur-SAD data

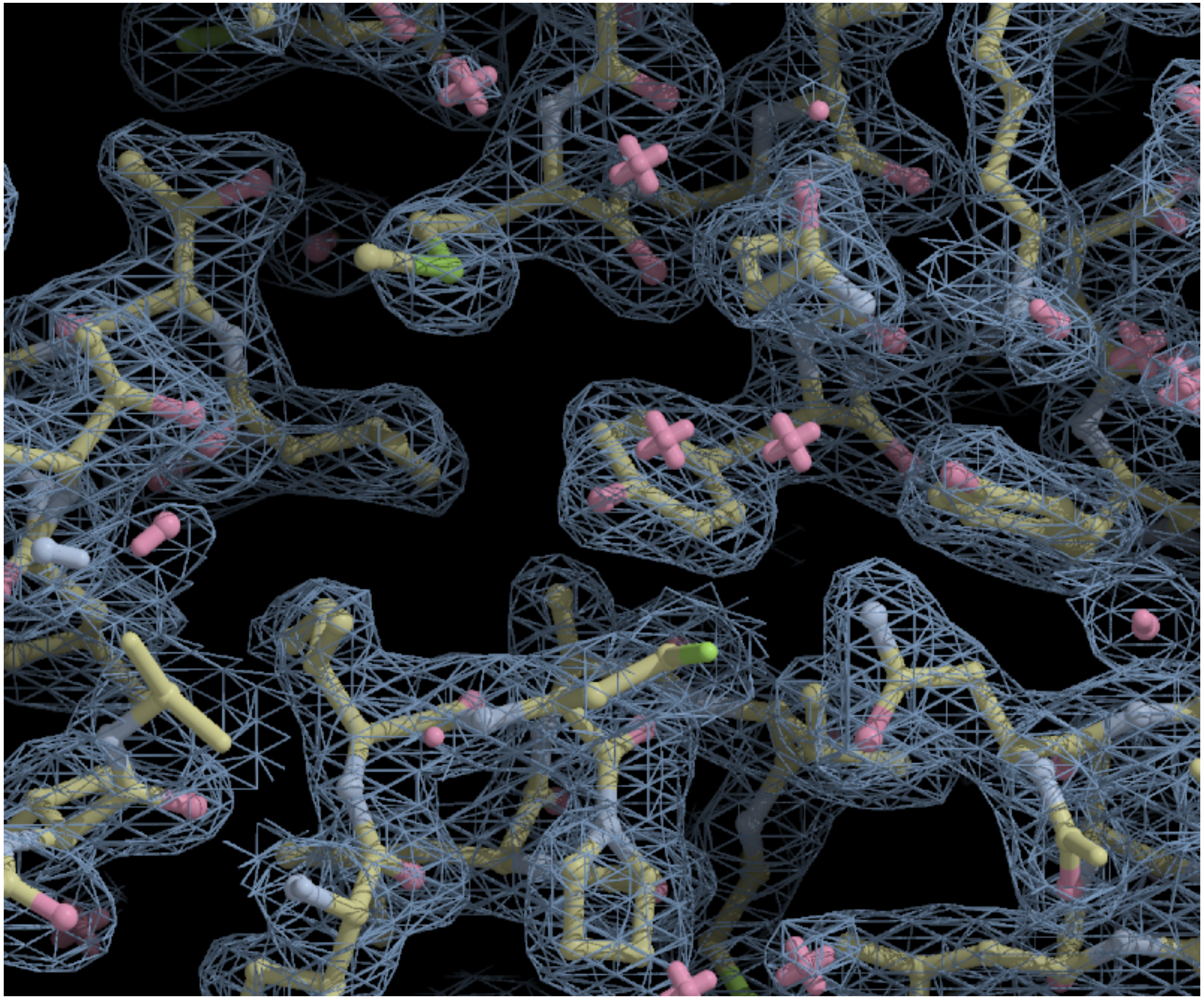


CysZ multi-crystal sulfur-SAD data



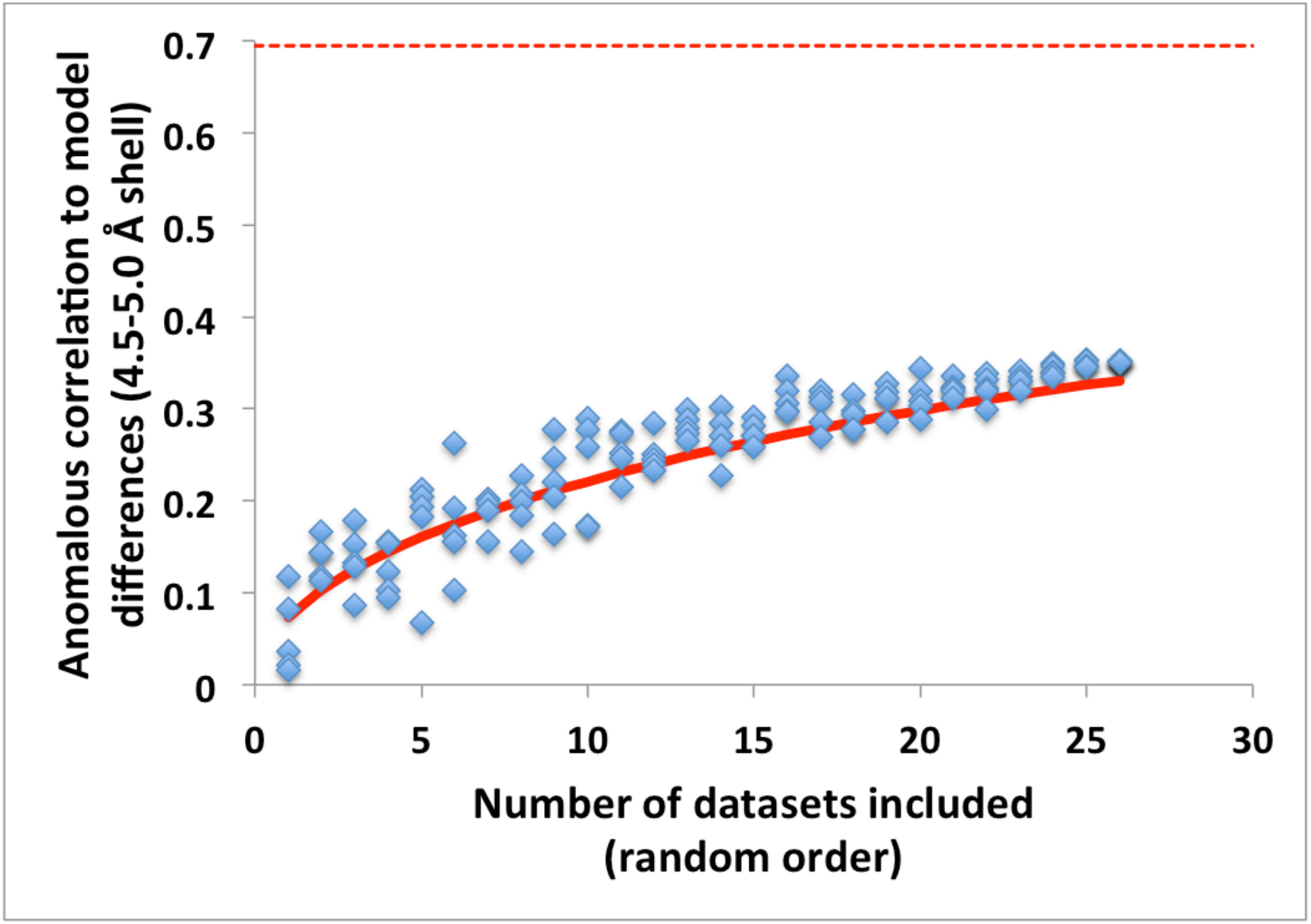
CysZ single-crystal sulfur-SAD data

Crystal 6 *AutoSol R/Rfree=0.24/0.27*



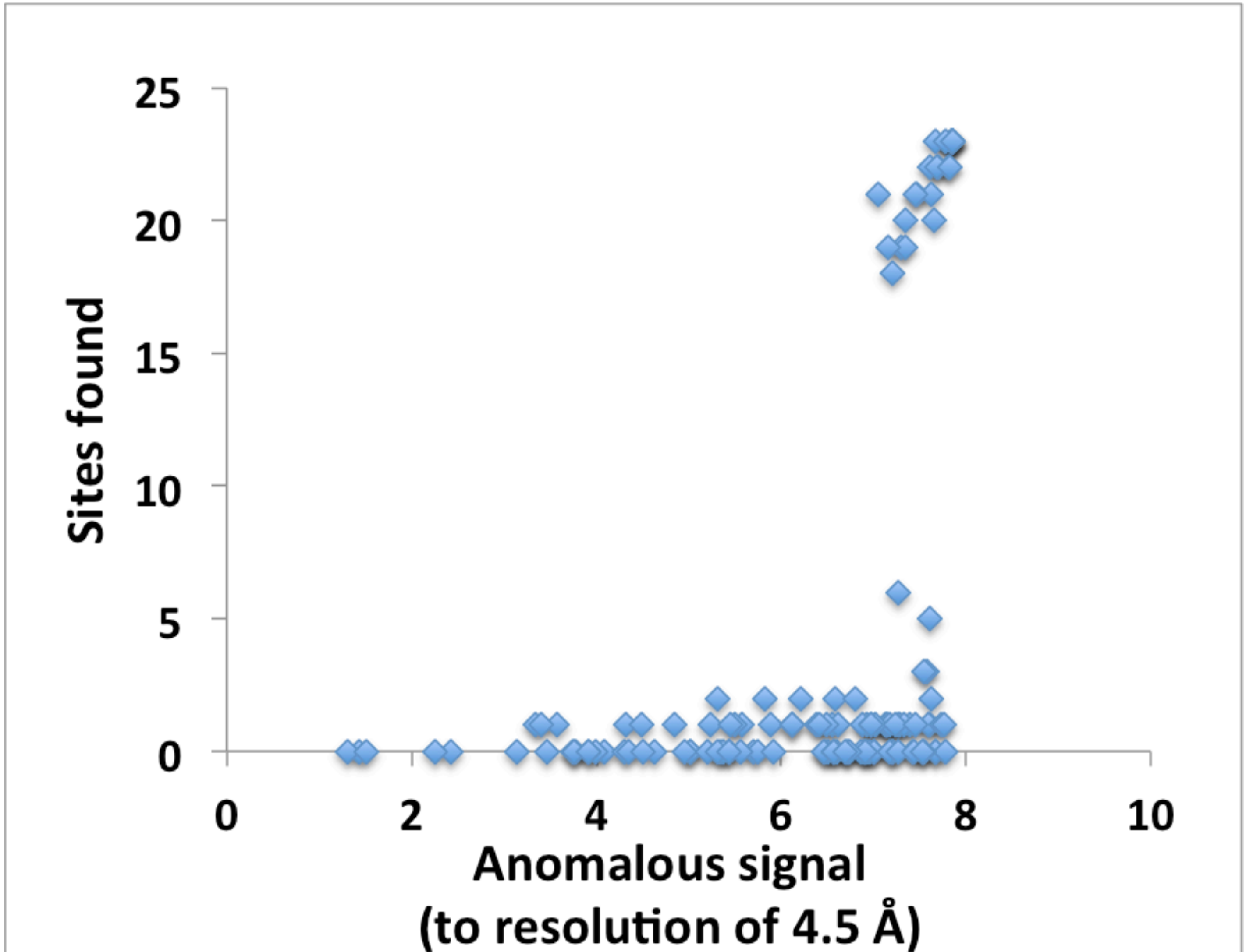
Flavivirus NS1 multi-crystal sulfur-SAD data

Akey et al., (2014) Science 343: 881- 885



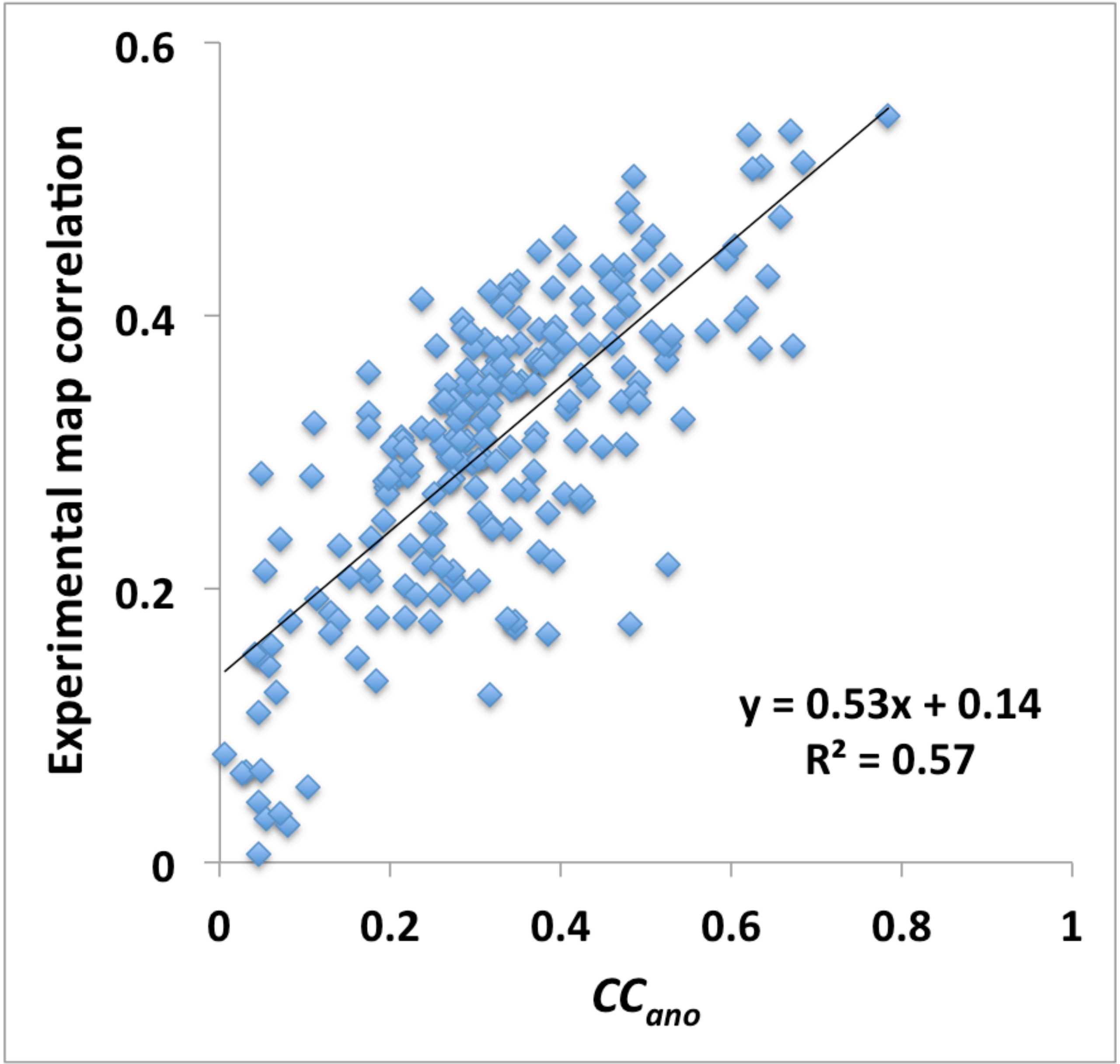
Flavivirus NS1 multi-crystal sulfur-SAD data

Akey et al., (2014) Science 343: 881- 885



Phasing with weak signal

Quality of phasing depends on the anomalous correlation (CC_{ano})



Estimating the anomalous signal before and after collecting the data

What affects the anomalous signal?

Anomalous correlation

Number of reflections

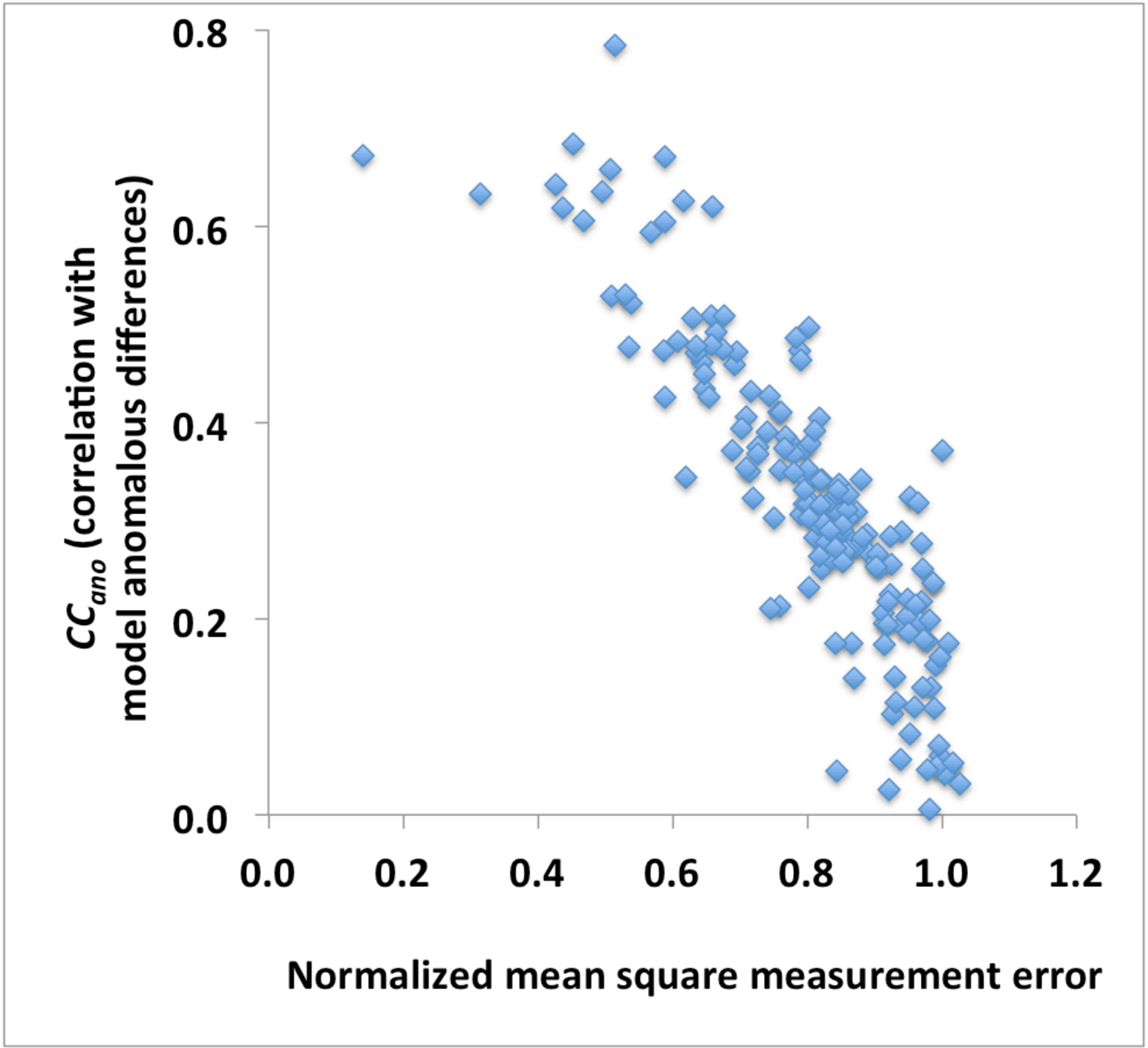
Anomalous signal

$$< S_{ano} >= CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$

Number of sites

B-value for anomalous sub-structure

Anomalous correlation decreases if the data are not accurately measured



Estimating the anomalous signal before collecting the data

Choose I/sigI, estimate normalized errors and CC_{ano}

Anomalous correlation

Number of reflections

Guess from sequence

Anomalous signal

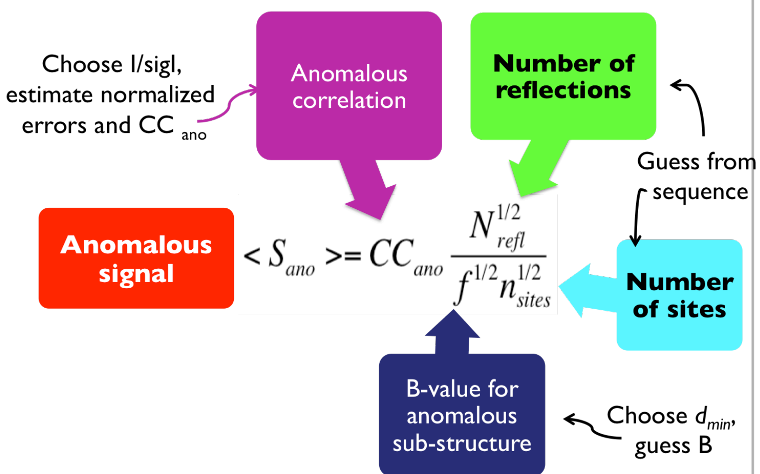
$$< S_{ano} > = CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$

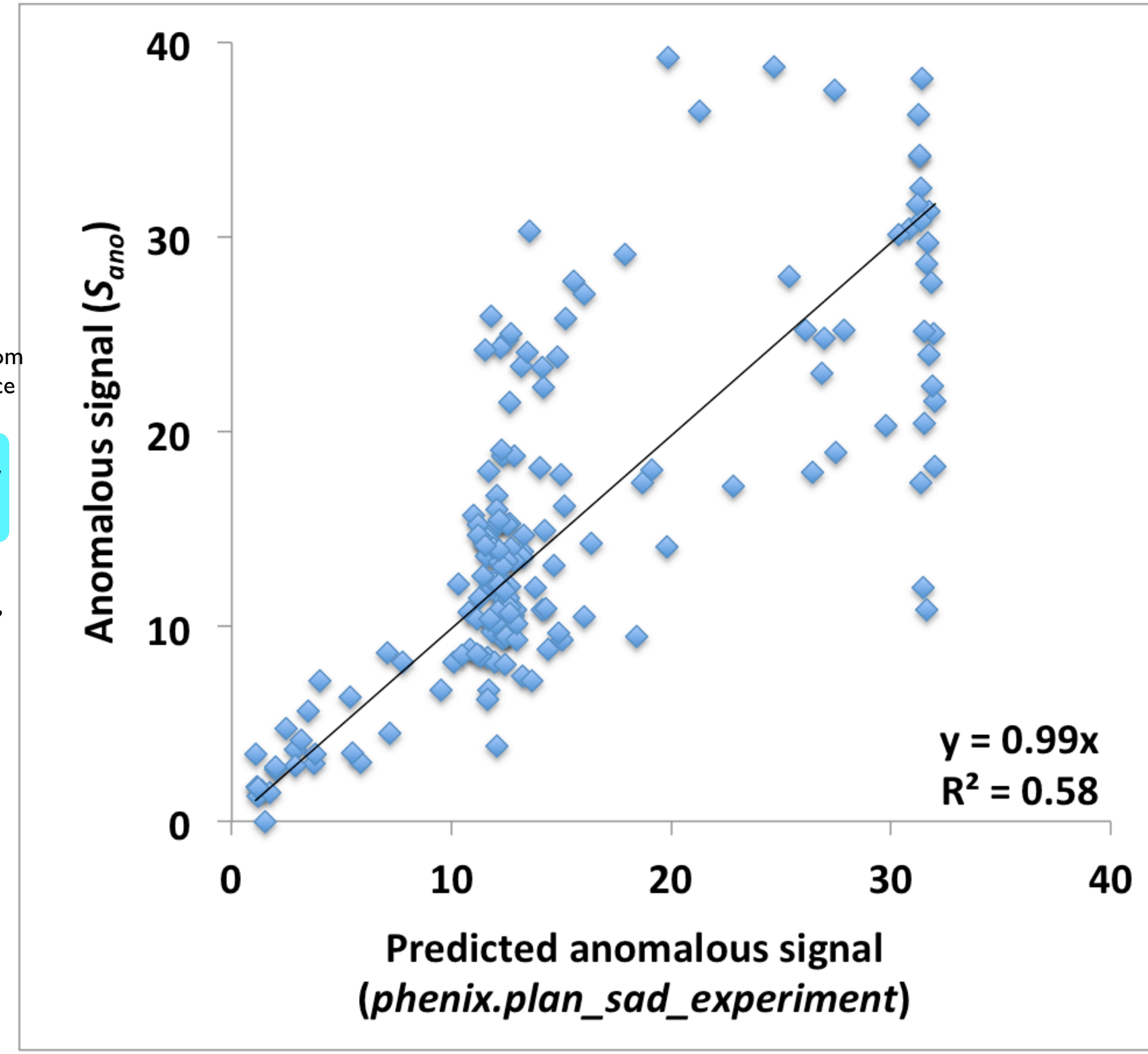
Number of sites

B-value for anomalous sub-structure

Choose d_{min}, guess B

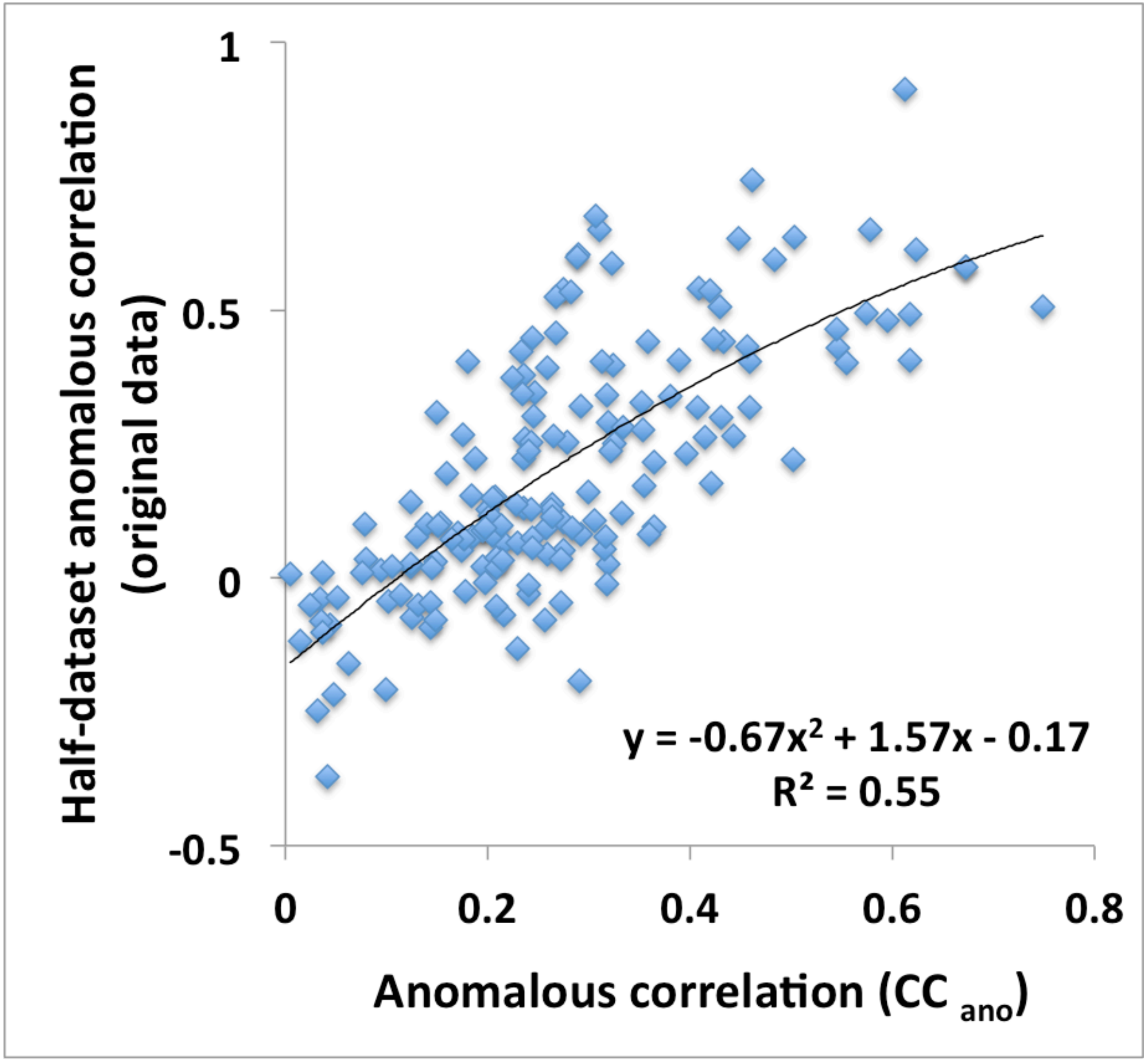
Estimating the anomalous signal before collecting the data





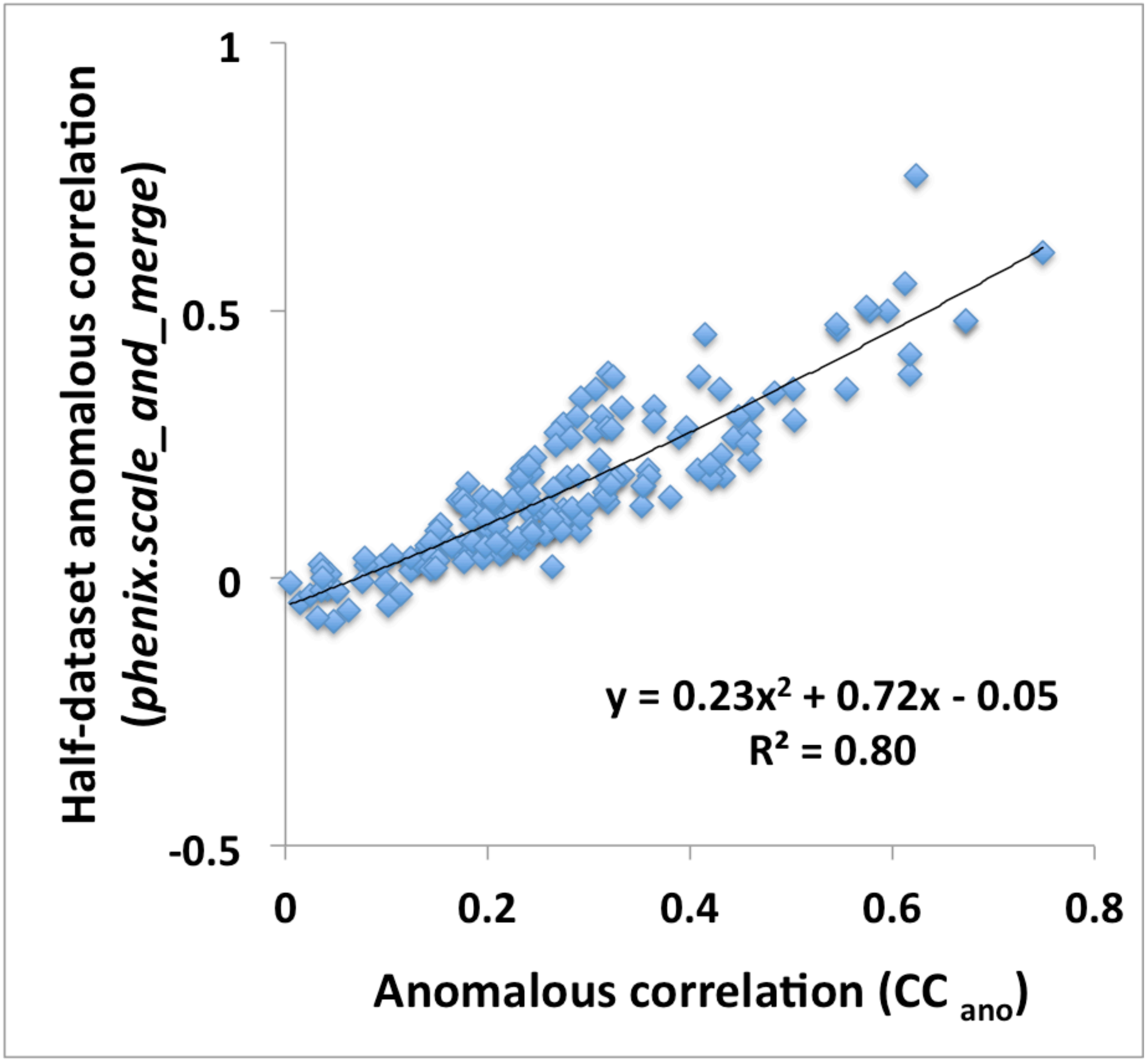
Scaling and merging SAD data

Anomalous correlation estimate from the measured data (deposited data)

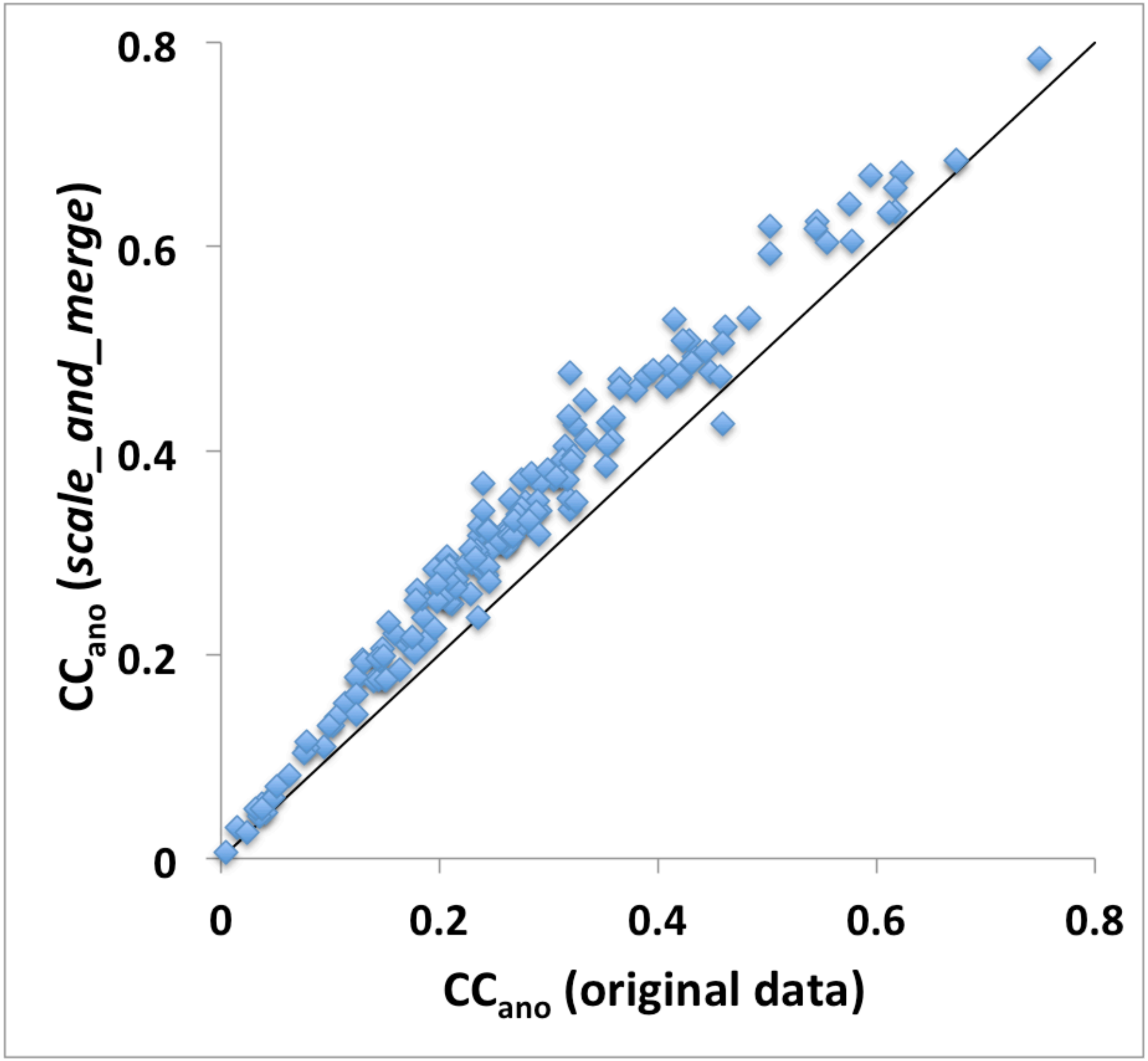


Anomalous correlation estimate from the measured data

(after local scaling with *phenix.scale_and_merge*)



Improvement in anomalous correlation using local scaling with *phenix.scale_and_merge*



Estimating the anomalous signal after collecting the data

Estimate from half-dataset anomalous CC, sigmas, skew

Anomalous correlation

Number of reflections

From data

Guess from sequence

Anomalous signal

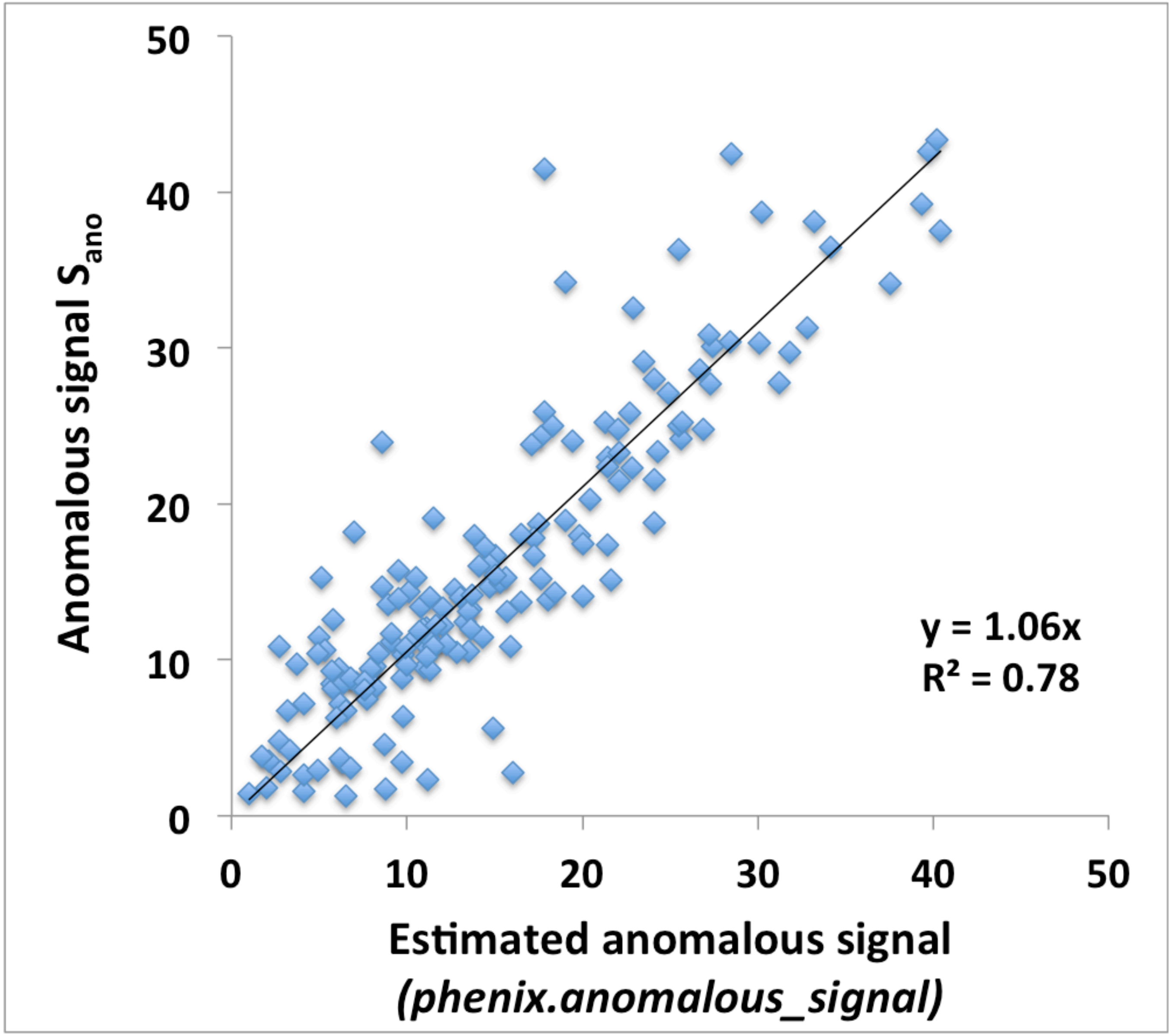
$$< S_{ano} > = CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$

Number of sites

B-value for anomalous sub-structure

Estimate B for sub-structure from Wilson B

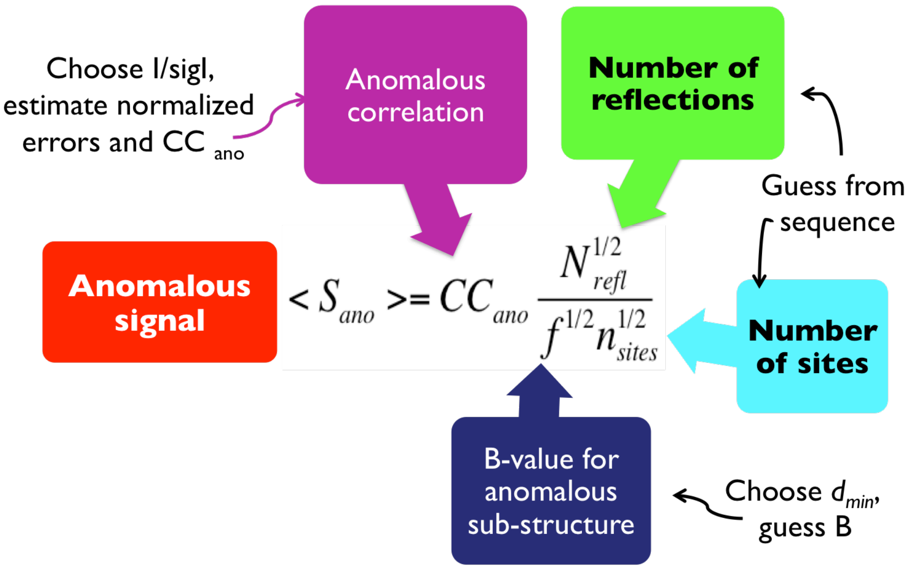
Estimating the anomalous signal after collecting the data



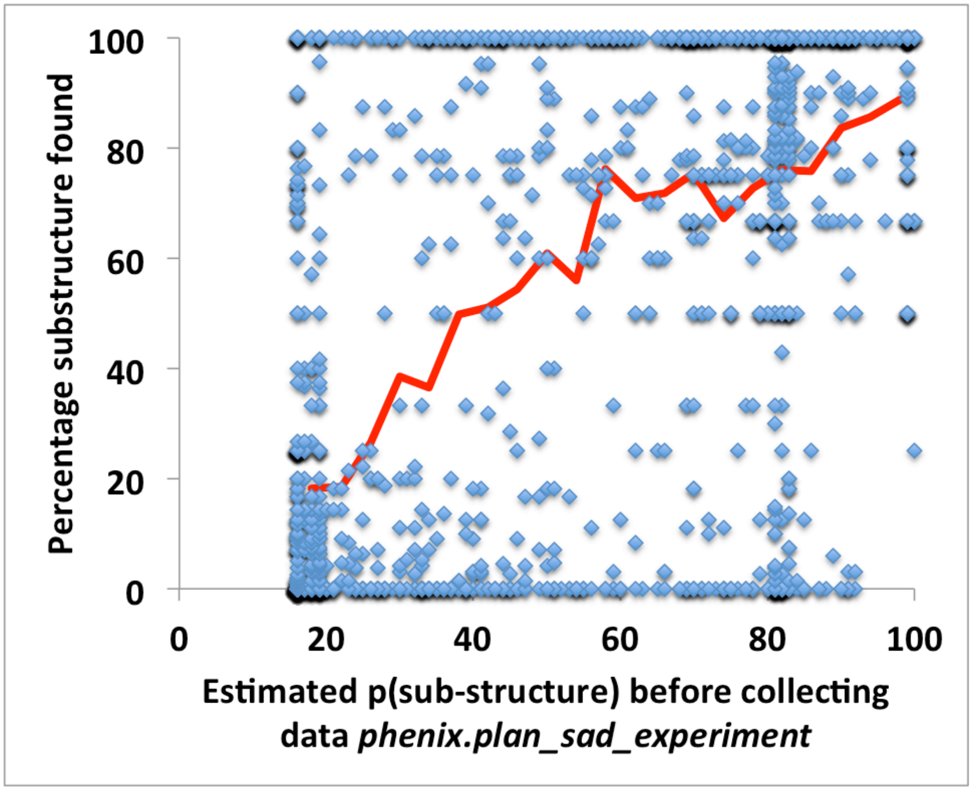
Will I solve the anomalous sub-structure?

Will I solve the anomalous substructure?

(Planning an experiment)

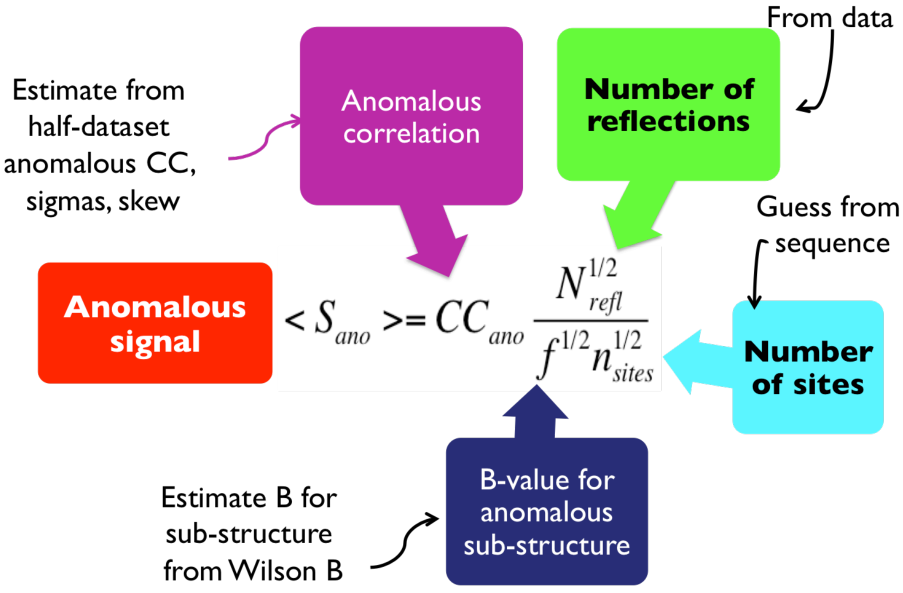


$$< S_{ano} > = CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$

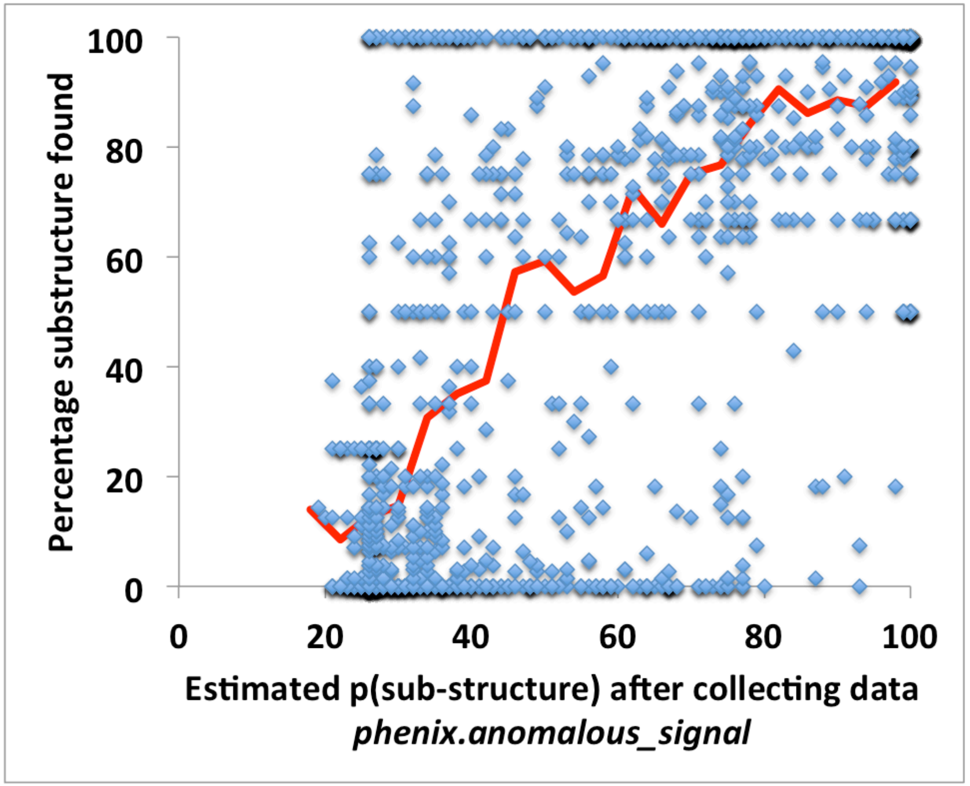


Will I solve the anomalous substructure?

(After collecting the data)



$$< S_{ano} >= CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$



Take-home message:

The anomalous signal is the key to solving your structure

Anomalous correlation

Number of reflections

Anomalous signal

$$<S_{ano}> = CC_{ano} \frac{N_{refl}^{1/2}}{f^{1/2} n_{sites}^{1/2}}$$

Number of sites

B-value for anomalous sub-structure

The PHENIX Project

Lawrence Berkeley Laboratory

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

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner, Richard Mifsud

Cambridge University

Duke University

Jane & David Richardson, Jeff Headd, Vincent Chen, Chris Williams, Bryan Arendall, Swati Jain, Bradley Hintze, Lizbeth Videau

An NIH/NIGMS funded Program Project

Phenix