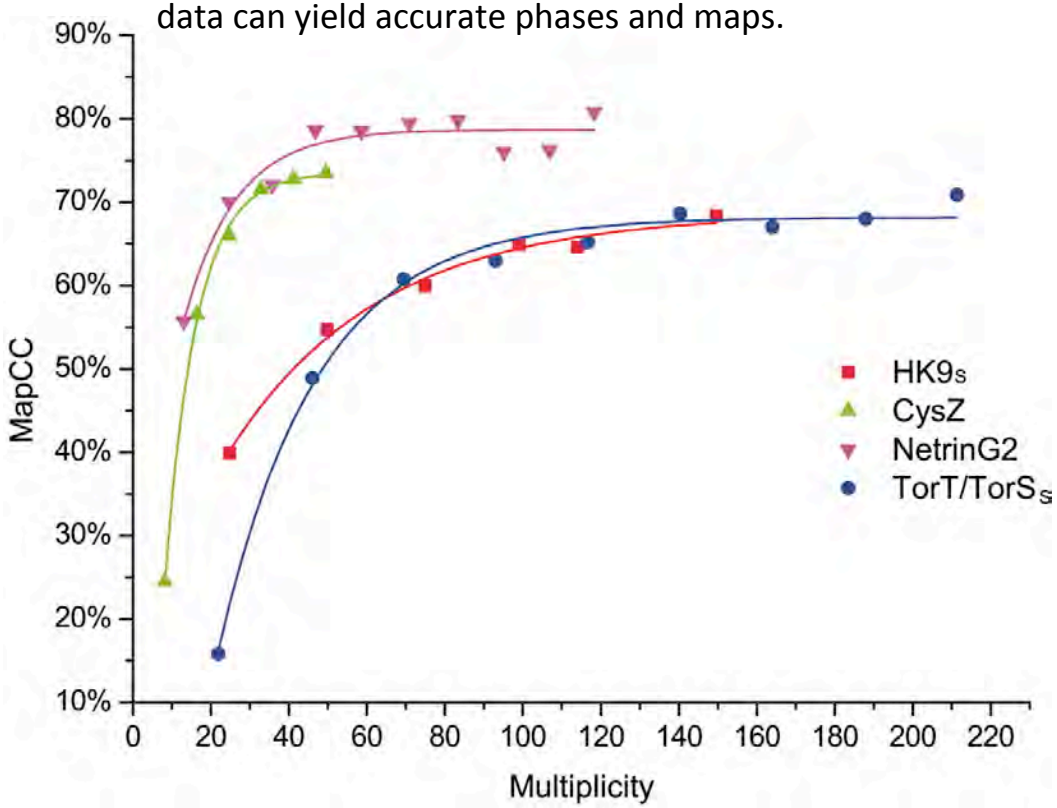
Ultra-High Mulitplicity to Detect a Weak Phasing Signal

Janet L. Smith
GM/CA @ APS
University of Michigan

New Science Opportunities Provided by a Multi-bend Achromat Lattice
Advanced Photon Source
October 21, 2013

Phasing from sulfur anomalous scattering using data far from the S edge

Wow! – if the sulfur sites are known, high-multiplicity data can yield accurate phases and maps.



Structures from Anomalous Diffraction of Native Biological Macromolecules

Qun Liu,[1] Tassadite Dahmane,[2] Zhen Zhang,[2] Zahra Assur,[2] Julia Brasch,[2] Lawrence Shapiro,[2] Filippo Mancia,[3] Wayne A. Hendrickson[1,2,3,4*]

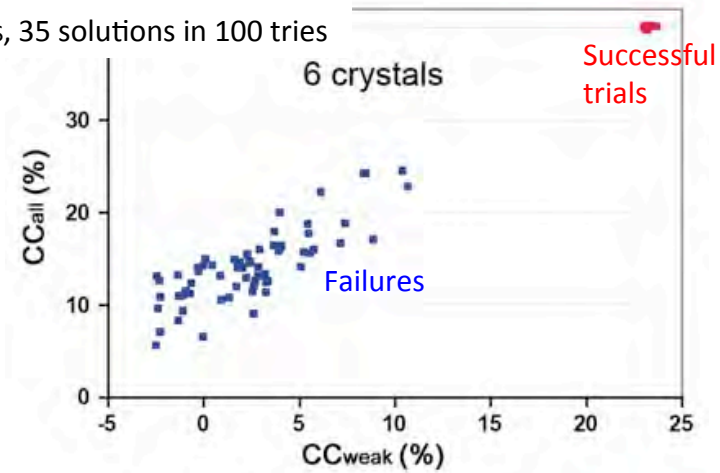
Science **336**,1033-1037 (2012)

- Multiple crystals
- Partial data sets
- Ultra-high multiplicity

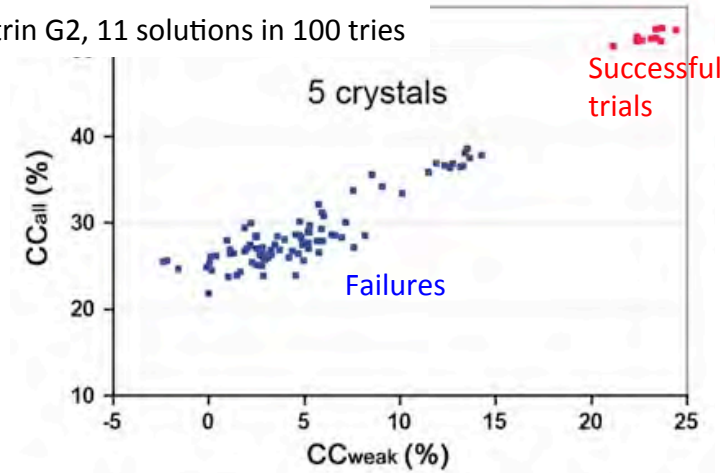
Sulfur sub-structure can be solved with many phasing trials (100-2000) in SHELXD

SHELXD correlation coefficients (CC) between observed and calculated Bijvoet differences

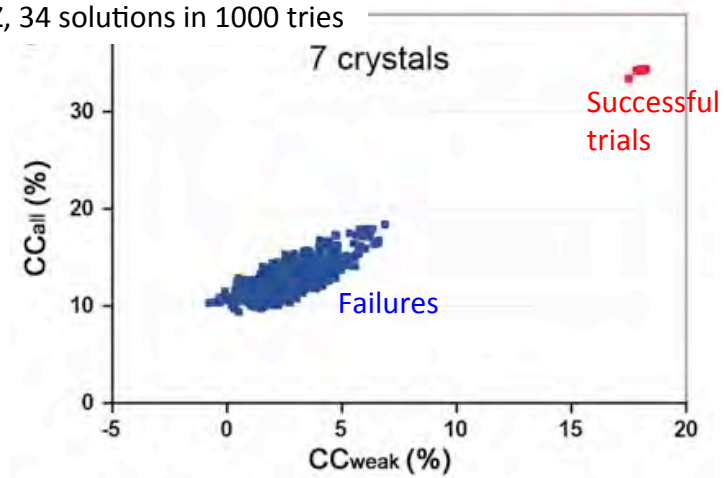
HK9s, 35 solutions in 100 tries



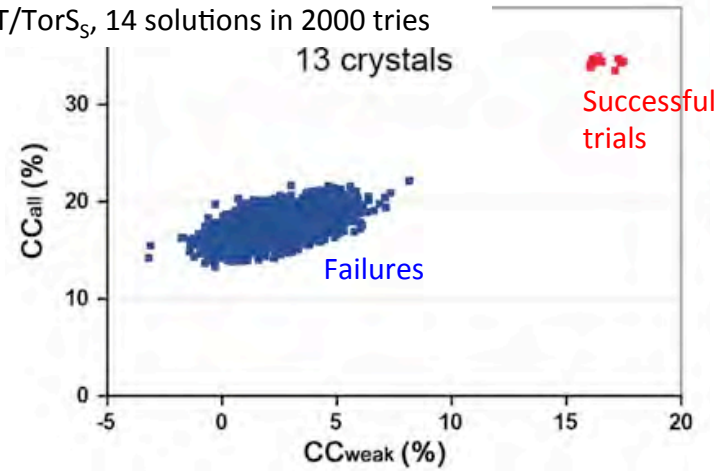
Netrin G2, 11 solutions in 100 tries



CysZ, 34 solutions in 1000 tries

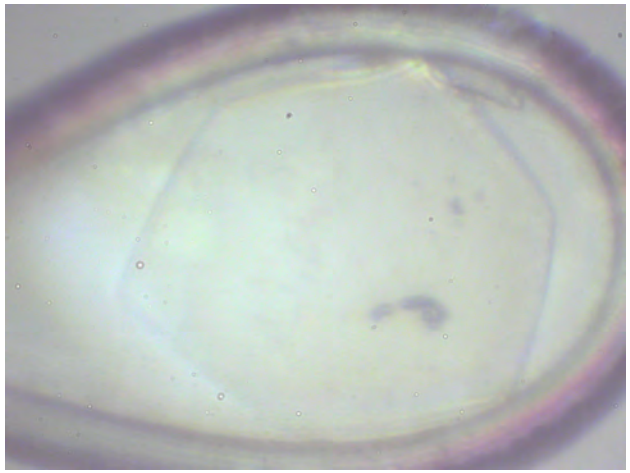


TorT/TorS$_S$, 14 solutions in 2000 tries

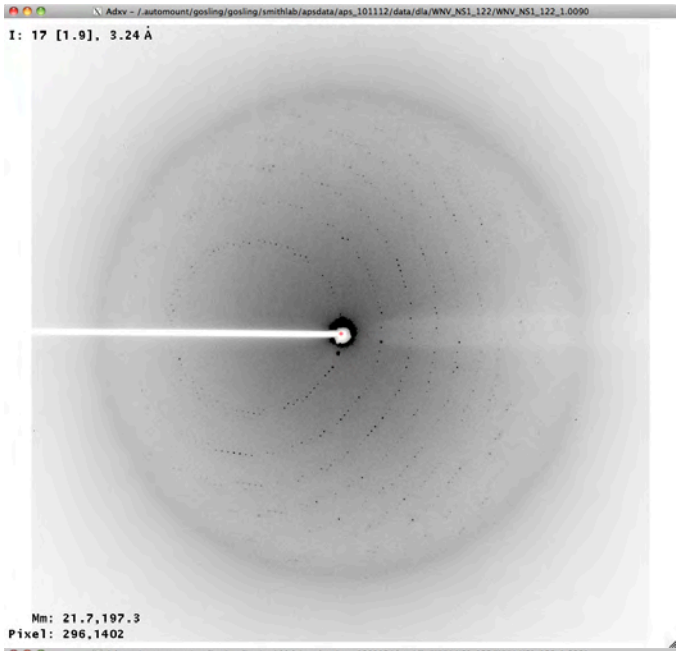


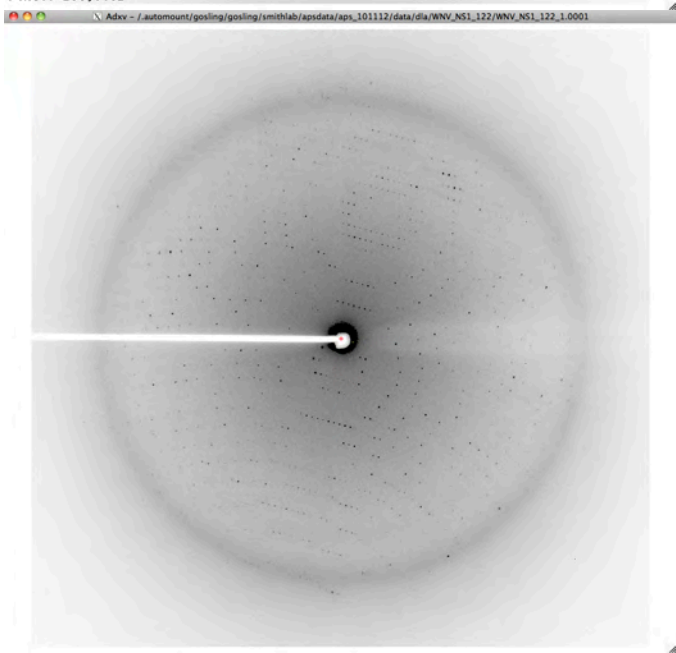
Liu, Q., *et al.*, *Science* **336**,1033-1037 (2012).

Crystals of our protein diffract to only ~3.1 Å



*P*321 or $P3_121$ or $P3_221$
168 Å x 168 Å x 94 Å
Typical d_{min} ~3.1 Å (aggressive)





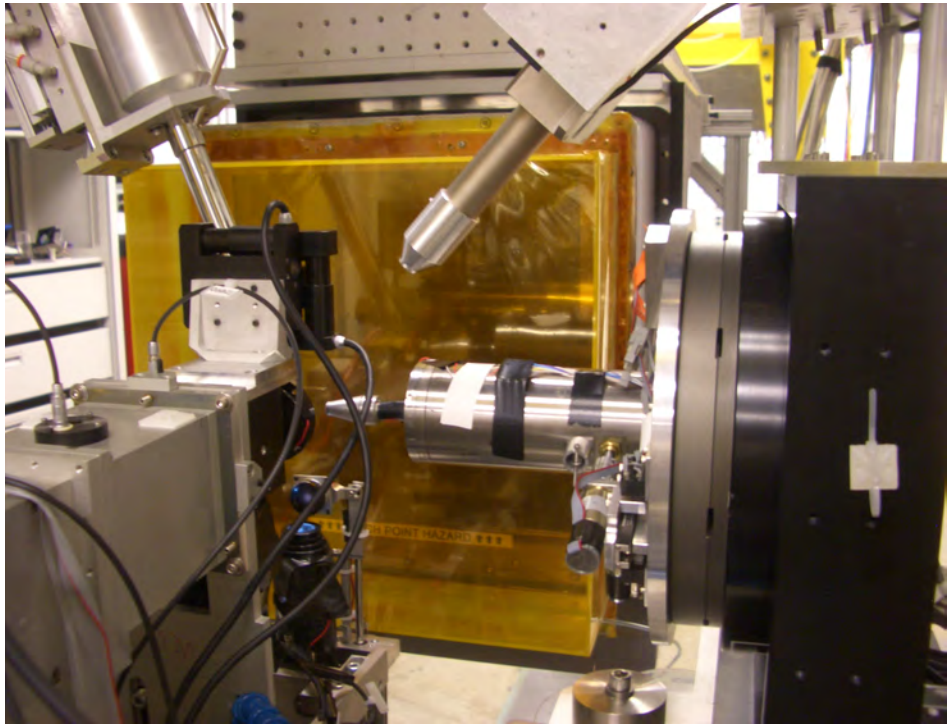
Data collection (E = 7.100 keV, λ = 1.7462 Å)
- 0.5° images
- 90° data
- "Inverse beam" with 5° wedges
- He path between sample & detector

Start at different positions on crystals (face, edge)

David Akey & Craig Ogata

Experimental setup – APS 23-IDD (GM/CA CAT)

View along goniometer axis

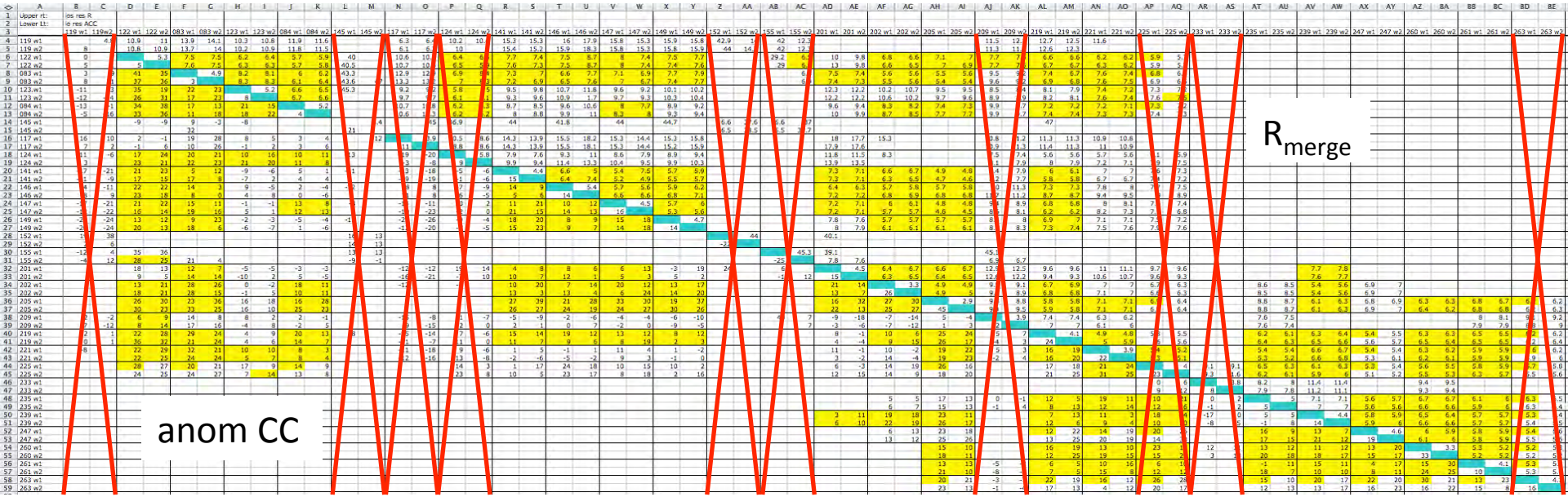


View along beam

David Akey @ U Michigan & Craig Ogata @ GM/CA

Pairwise comparison of 28 data sets – which are compatible?

Criteria: R_{merge} < 7.5% and anom CC > 0 for the LOWEST resolution bin



For all crystals: 90° data collected in inverse beam mode; 0.5°/image; 5° wedges, 7.1 keV, He box

Data from 18 crystals combined in final set (from 28 total crystals collected).

Yellow boxes have R_{merge} < 7.5% and anom CC > 0 for the LOWEST resolution bin only.

Crystals eliminated if (consistently) R_{merge} > 7.5% and anom CC < 0 when scaled individually with other crystals.

David Akey

Scaling statistics suggest anomalous signal to only low resolution, useful data to ~3.1 Å

Combined data; 18 crystals; separate I^+ & I^-

RESOLUTION LIMIT	# REFLECTIONS OBSERVED	UNIQUE	MULTI-PLICITY	COMPLETENESS OF DATA	Rmerge observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
8.57	249,133	2547	98	99.5%	9.2%	10.6%	249133	70.05	9.3%	100.0*	64*	1.719	1133
6.11	449,294	4476	100	100.0%	13.4%	13.4%	449294	52.18	13.4%	99.9*	31*	1.276	2095
5.00	585,535	5807	101	100.0%	19.5%	18.6%	585535	39.21	19.6%	99.9*	-4	0.905	2762
4.34	675,539	6721	101	100.0%	20.9%	20.0%	675539	36.63	21.0%	99.9*	-16	0.760	3222
3.88	787,769	7835	101	100.0%	36.8%	36.3%	787769	23.45	37.0%	99.7*	-16	0.685	3771
3.55	836,250	8331	100	100.0%	69.5%	71.9%	836250	13.04	69.9%	99.4*	-11	0.681	4030
3.28	961,534	9610	100	100.0%	134.2%	147.7%	961534	6.92	134.9%	97.7*	-5	0.680	4657
3.07	987,954	9909	100	100.0%	286.8%	335.8%	987954	3.16	288.2%	90.9*	-1	0.653	4813
2.90	318,782	10164	31	99.0%	780.5%	925.3%	318681	0.55	794.0%	11.7*	1	0.559	4869
total	5,851,790	65400	89	99.8%	31.1%	32.9%	5851689	19.60	31.3%	99.9*	-1	0.763	31352

Data from 1 “good” crystal – 90°; 0.5° oscillation; separate I^+ & I^-

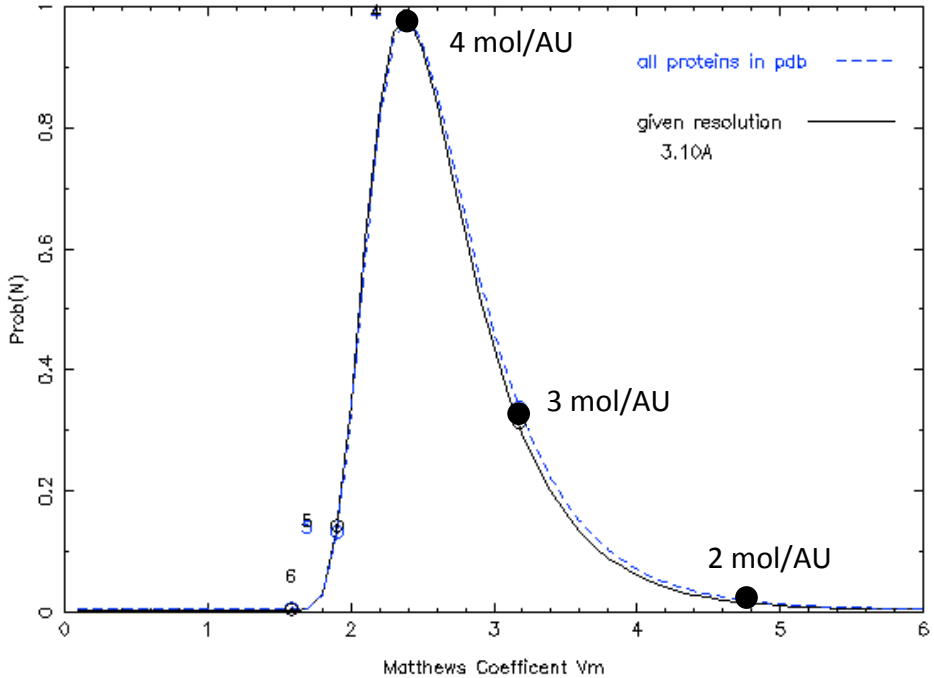
RESOLUTION LIMIT	# REFLECTIONS OBSERVED	UNIQUE	MULTI-PLICITY	COMPLETENESS OF DATA	Rmerge observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
8.87	6401	2154	3.0	94.2%	3.8%	4.2%	6111	23.29	4.6%	99.7*	24*	0.995	746
6.33	11283	3787	3.0	95.1%	6.0%	5.5%	10787	16.69	7.3%	99.3*	11*	1.003	1375
5.18	14829	4939	3.0	95.9%	9.2%	8.4%	14165	11.48	11.0%	98.7*	9	0.938	1794
4.49	17434	5872	2.9	96.2%	9.2%	8.4%	16644	10.87	11.1%	98.9*	4	0.895	2140
4.02	19768	6667	3.0	96.9%	14.0%	13.5%	18905	7.58	16.9%	97.2*	5	0.856	2418
3.67	21835	7422	2.9	97.7%	27.9%	28.5%	20884	3.94	33.7%	90.9*	1	0.773	2676
3.40	23772	8144	2.9	98.4%	48.3%	51.5%	22722	2.16	58.3%	77.8*	4	0.739	2909
3.18	25287	8773	2.9	99.0%	89.3%	99.3%	24083	1.11	107.8%	52.6*	-1	0.651	3062
3.00	22428	8371	2.7	88.3%	173.9%	202.8%	20793	0.49	211.3%	22.1*	2	0.598	2657
total	163037	56129	2.9	95.8%	16.0%	16.6%	155094	6.15	19.3%	99.0*	5	0.788	19777

Scaling statistics indicate an anomalous signal to ~5.4 Å.

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
20.13	17358	194	207	93.7%	11.4%	9.5%	17358	79.88	11.4%	100.0*	82*	2.474	74
14.23	34625	359	359	100.0%	7.6%	9.7%	34625	77.11	7.7%	100.0*	80*	1.996	152
11.62	45519	460	460	100.0%	8.0%	10.0%	45519	73.75	8.1%	100.0*	66*	1.874	205
10.06	54412	549	549	100.0%	8.8%	10.1%	54412	73.52	8.9%	100.0*	54*	1.840	246
9.00	62151	627	627	100.0%	8.8%	10.3%	62151	70.93	8.9%	100.0*	37*	1.424	288
8.22	70007	704	704	100.0%	10.4%	10.8%	70007	66.04	10.4%	100.0*	44*	1.458	327
7.61	74872	747	747	100.0%	10.7%	11.5%	74872	60.95	10.8%	100.0*	28*	1.290	345
7.12	77435	776	776	100.0%	13.4%	13.2%	77435	54.03	13.4%	99.9*	21*	1.131	362
6.71	88572	881	881	100.0%	14.1%	14.7%	88572	48.44	14.1%	99.9*	31*	1.194	417
6.36	93704	928	928	100.0%	15.0%	15.7%	93704	46.24	15.1%	99.9*	20*	1.096	436
6.07	95640	946	946	100.0%	16.8%	18.1%	95640	41.25	16.9%	99.9*	26*	1.061	446
5.81	97628	967	967	100.0%	17.9%	19.4%	97628	38.87	18.0%	99.9*	4	0.880	458
5.58	106568	1053	1053	100.0%	19.6%	21.3%	106568	36.28	19.7%	99.9*	9	0.873	503
5.38	109605	1088	1088	100.0%	18.8%	20.5%	109605	37.09	18.9%	99.9*	-3	0.801	513
5.20	110529	1097	1097	100.0%	20.0%	21.4%	110529	34.70	20.1%	99.9*	-5	0.750	524
5.03	119499	1185	1185	100.0%	19.8%	21.1%	119499	35.45	19.9%	99.9*	-7	0.707	565
4.88	118762	1180	1180	100.0%	19.6%	21.1%	118762	35.48	19.7%	99.9*	-6	0.725	563
4.74	124479	1239	1239	100.0%	20.1%	21.2%	124479	35.30	20.2%	99.9*	-18	0.679	594
4.62	126948	1266	1266	100.0%	19.2%	20.6%	126948	35.54	19.3%	99.9*	-9	0.689	604
4.50	131613	1307	1307	100.0%	20.1%	21.6%	131613	34.83	20.2%	99.9*	-18	0.653	630
total	1759926	17553	17566	99.9%	14.1%	15.2%	1759926	45.17	14.2%	100.0*	17*	0.994	8252

18-crystal combined data: low resolution only (4.5 Å)

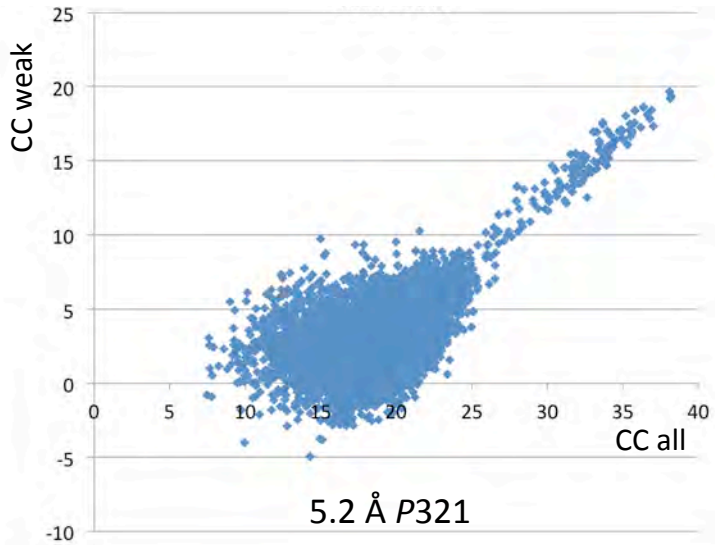
Matthews probability suggests 3 or 4 copies per AU
(3-Å diffraction limit suggests <u>not</u> 4, maybe 3, perhaps 2)

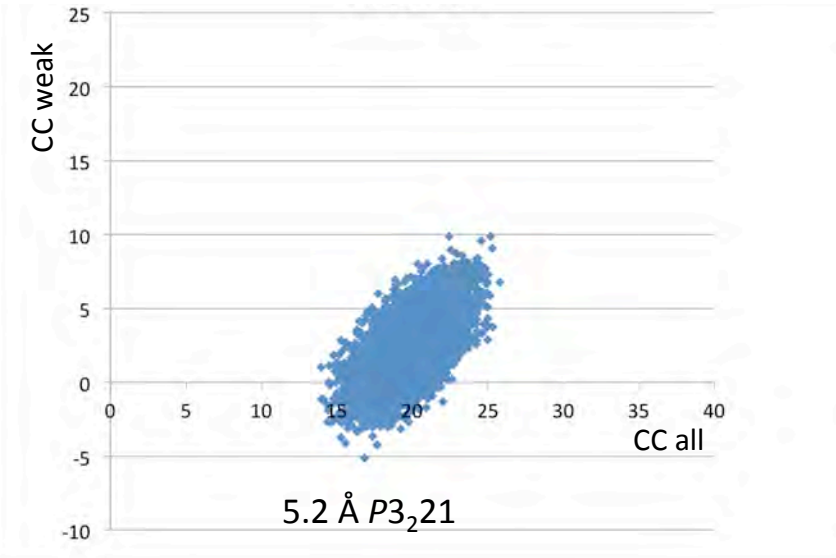


N(mol)	Prob(N) for d-min	Prob(N) overall	Vm (A**3/Da)	Volume solvent(%)	MW Da
1	0.002	0.004	9.52	87	38045
2	0.011	0.013	4.76	74	76090
3	0.213	0.228	3.17	61	114135
4	0.674	0.664	2.38	48	152180
5	0.097	0.088	1.90	35	190225
6	0.003	0.003	1.59	22	228270

http://www.ruppweb.org/mattprob/

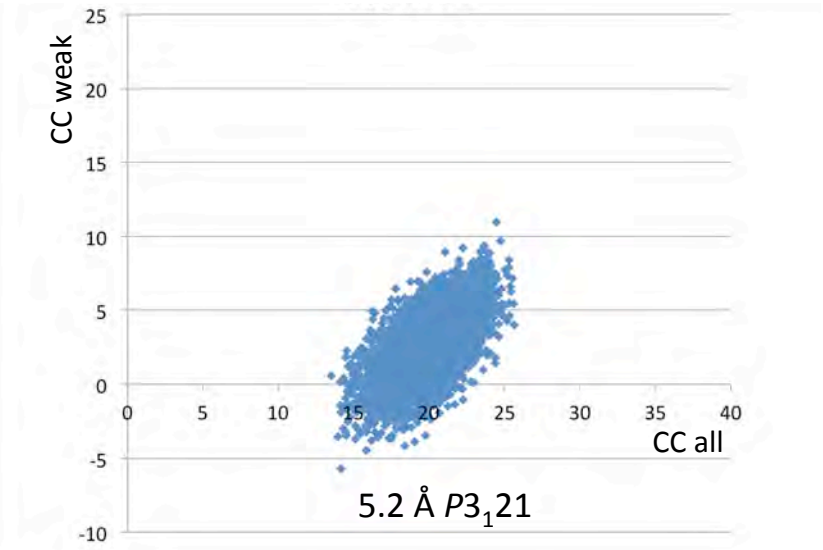
Phasing trials: clear indication of space group *P*321 – this may be working!



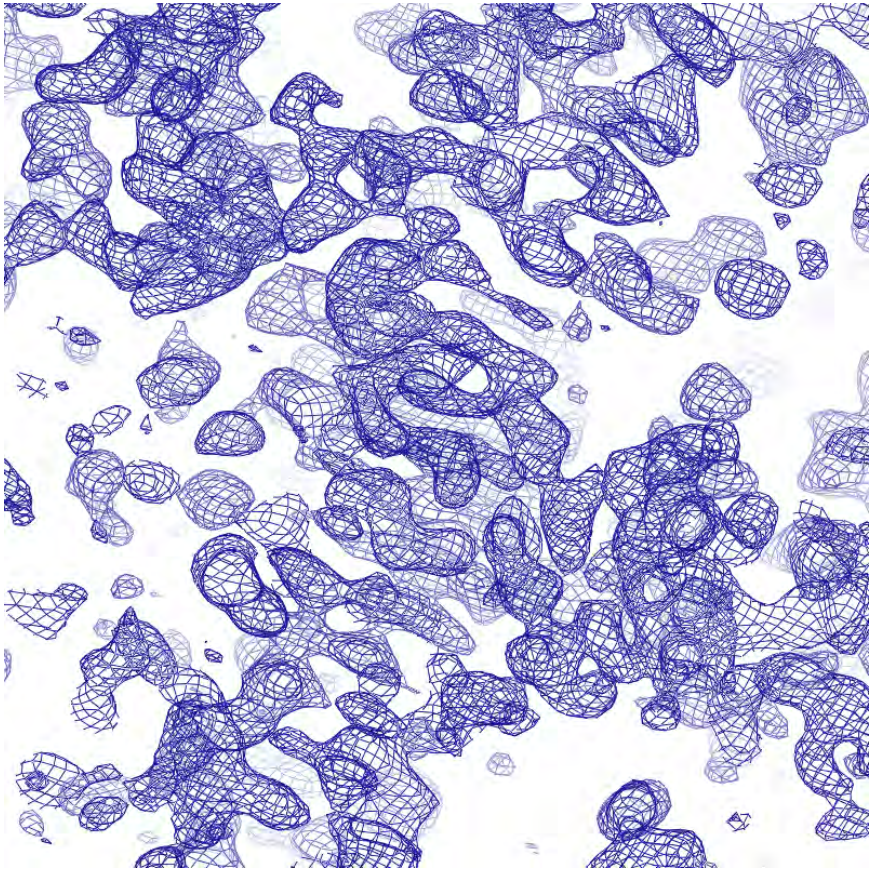


CC(all) *vs*. CC(weak) for various SHELXD trials

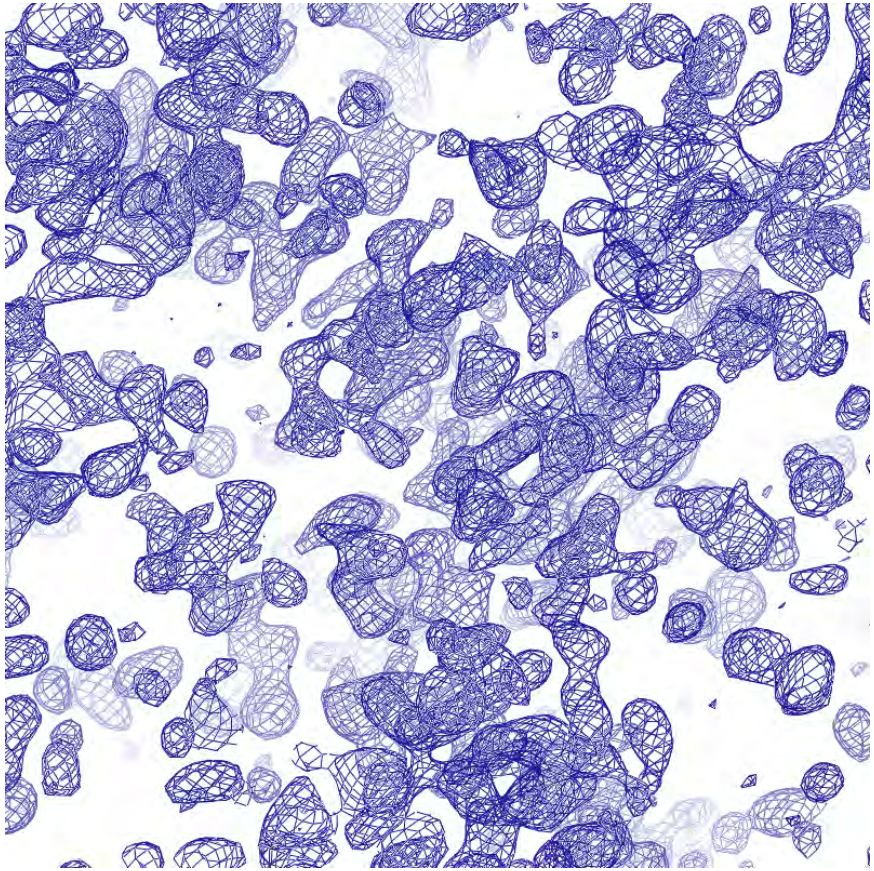
10,000 trials per SHELXD run
Search for 3x17 sites per AU



Initial maps – choice between two hands is clear



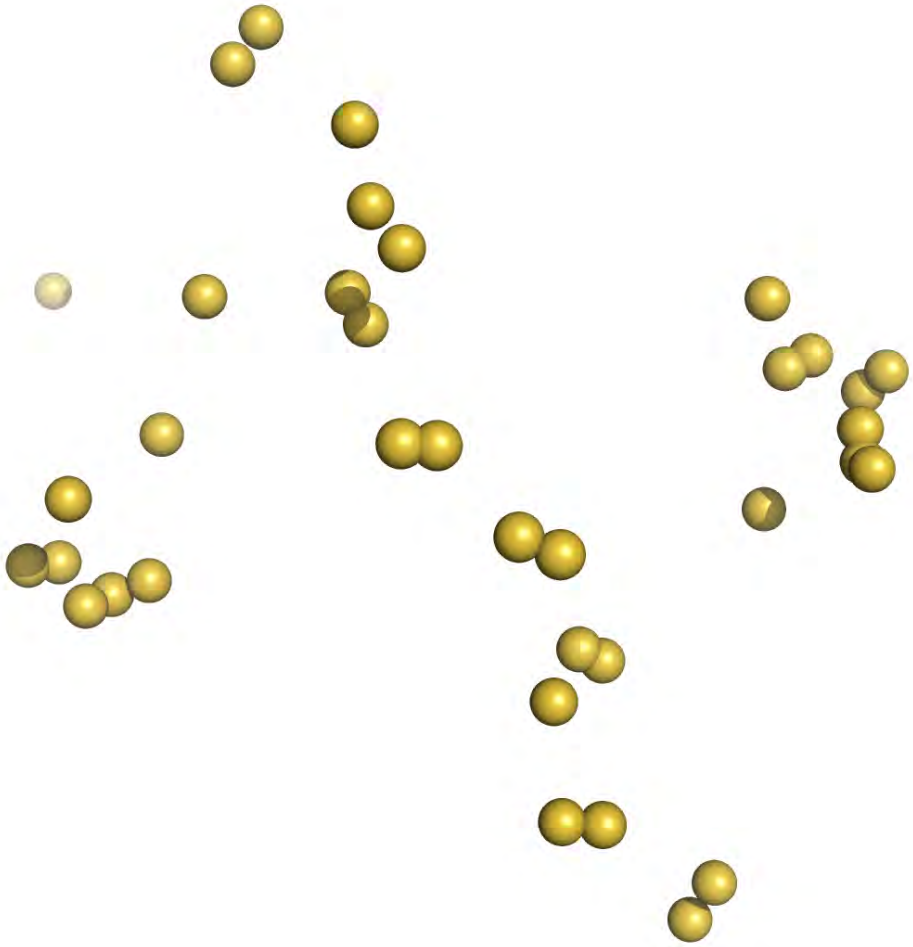
Contrast = 0.442
Connect = 0.823
Pseudo-free CC = 42.56 %



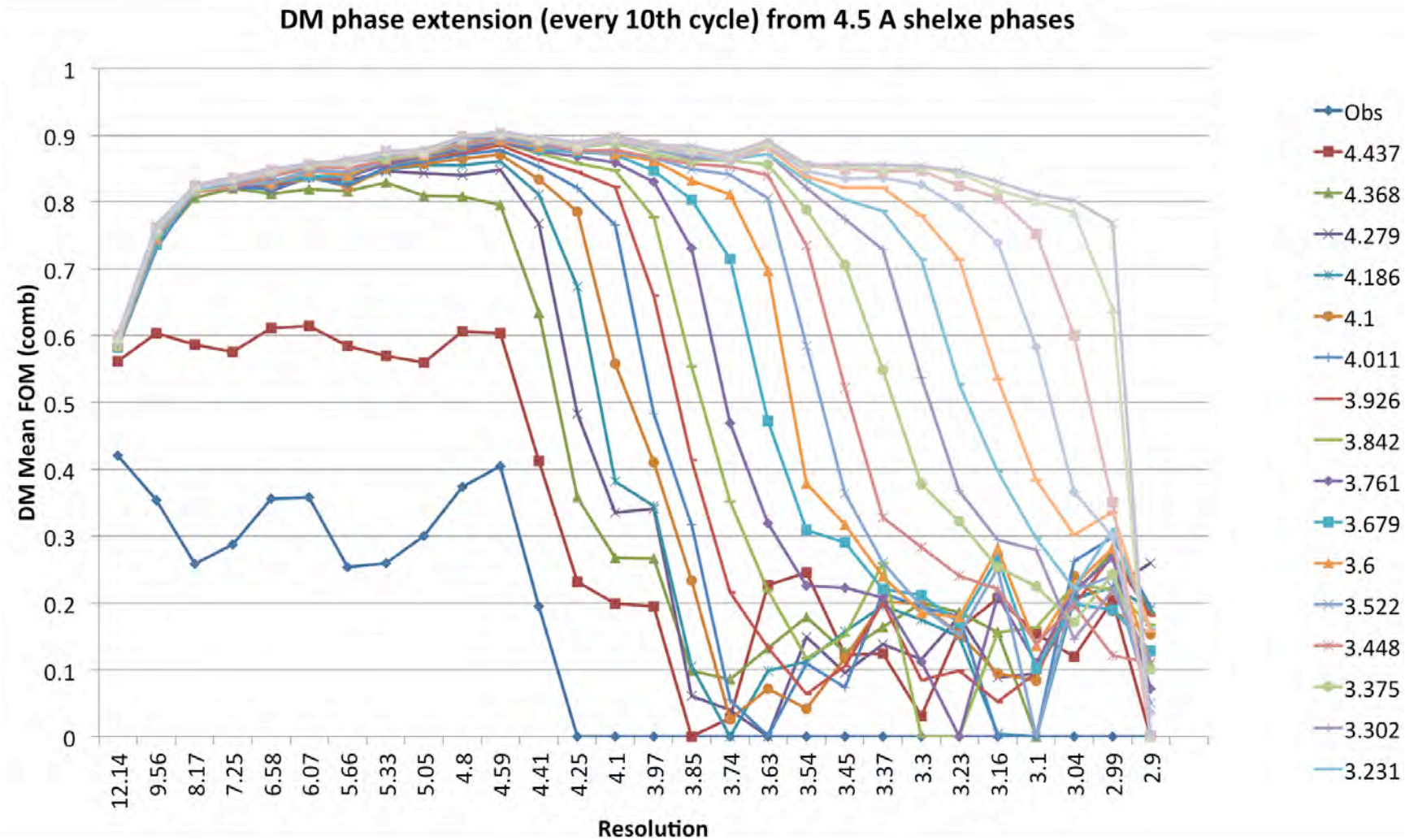
Contrast = 0.338
Connect = 0.792
Pseudo-free CC = 35.91 %

SHELXE phases and maps calculated to 4.5 Å, contoured at 1.5 σ

First inspection of the S sub-structure – clearly 2-fold NCS



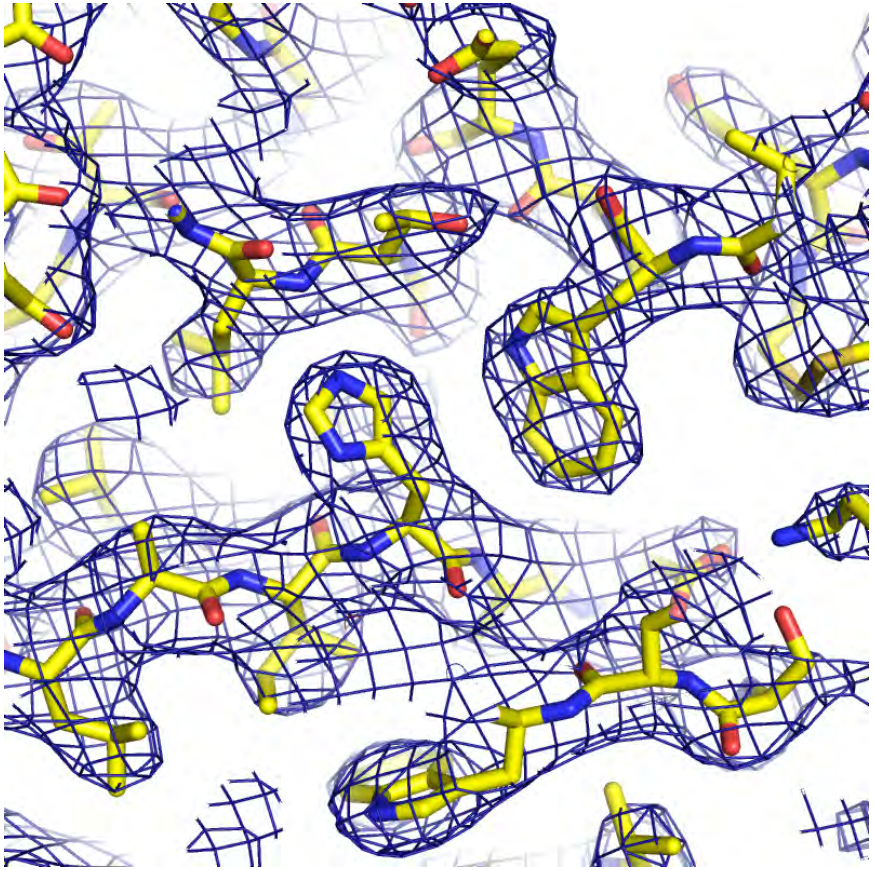
Phase extension and refinement from 4.5 Å to 3.0 Å in 200 steps

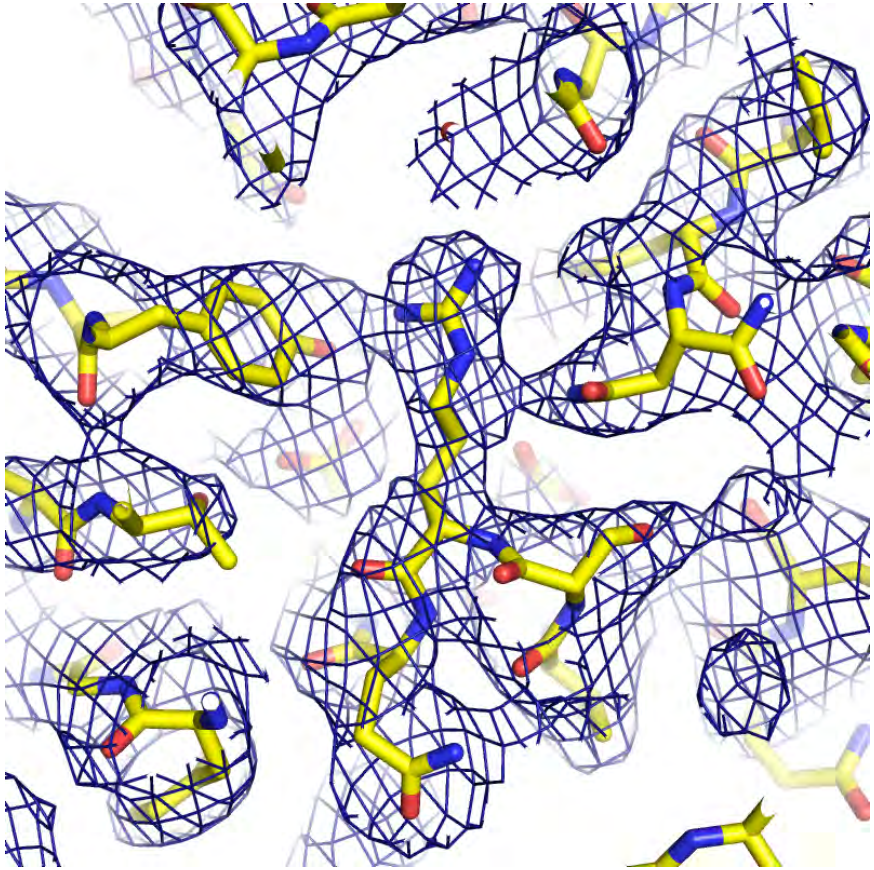


DM strategy: Calculate phases & weights to 4.5 Å
extend to 3 Å in 200 steps
2-fold NCS (operator refined in LSQKAB)
75% solvent

High quality 3.0-Å map after phase refinement & extension

3.0-Å map (1.5 σ) with DM <u>experimental</u> phases (refined model superimposed)





Final refinement at d_{min} = 2.9 Å: R_{work} = 17.2%, R_{free} = 21.3%

Success of Native Sulfur SAD Phasing – The Envelope Expands

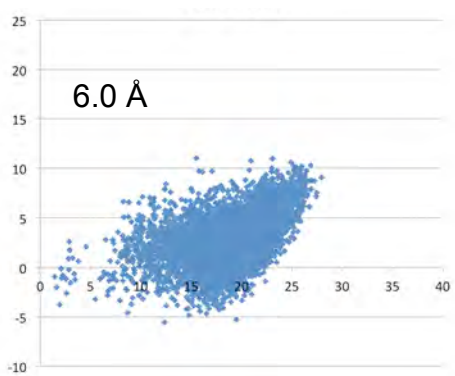
	HK9s	**CysZ**	**Netrin G2**	**TorT/TorSs**	**POI** (Akey/Smith)
Number of crystals	6	7	5	13	18
d_{min} (Å)	2.3	2.3	2.3	2.8	3.1
Unique residues	157	492	332	1162	740
Unique sulfur atoms	3	22	26	28	34
$\langle\lvert\Delta F_{\pm h}\rvert\rangle/\langle\lvert F\rvert\rangle(\%)$	0.74	1.12	1.49	0.83	1.49
Substructure success rate	35/100	34/1000	11/100	14/2000	few/10,000
Substructure determination resolution (Å)	3.0	3.0	3.0	3.9	5.2
Auto-built residues	98/127	408/453	291/312	965/1148	
R	0.187	0.2	0.203	0.184	0.199
R_{free}	0.227	0.238	0.241	0.242	0.246
PDB code	3VA9	3TX3	3TBD	3O1I	

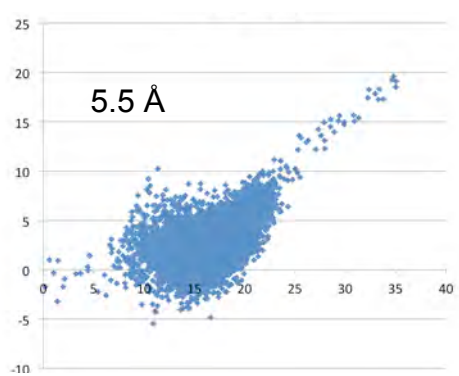
Poorer data
Lower resolution
Greater multiplicity

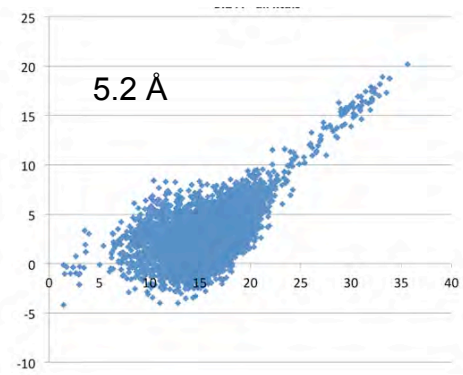
Liu, Q., *et al.*, *Science* **336**,1033-1037 (2012).

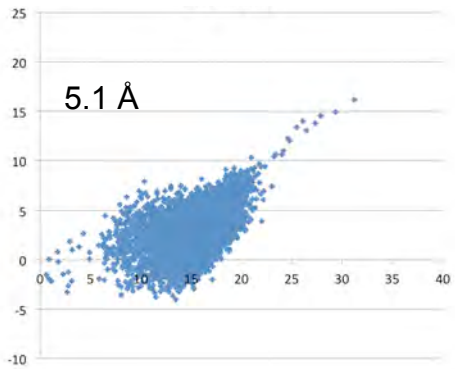
Post mortem of S sub-structure determination: Low-resolution data only

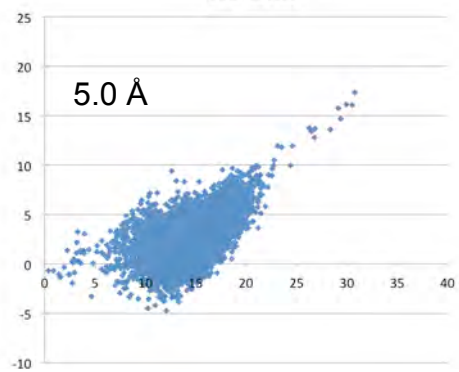
CC(all) *vs*. CC(weak) for various SHELXD trials – resolution

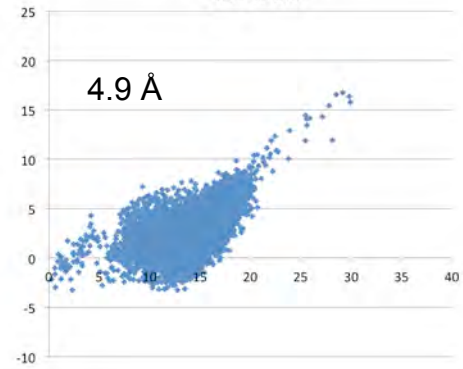


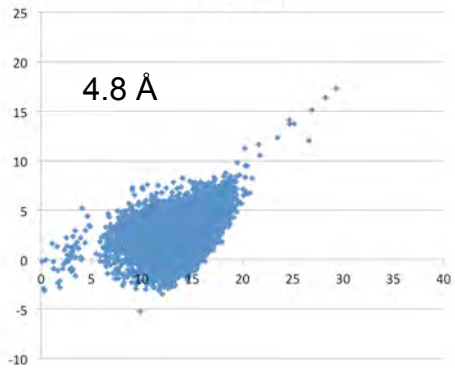


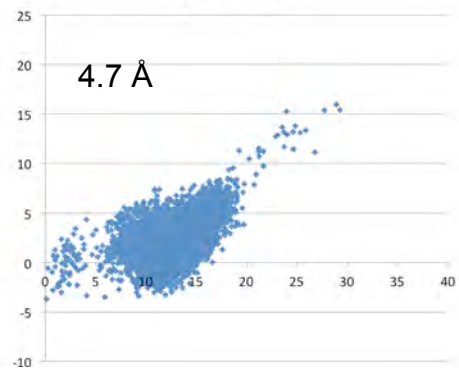










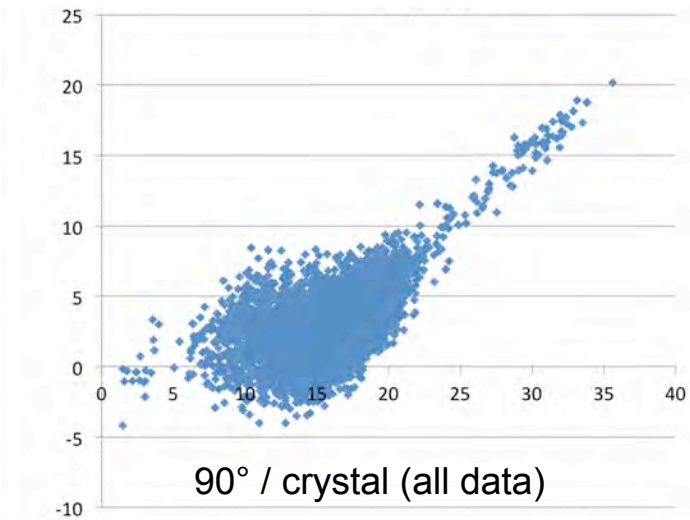


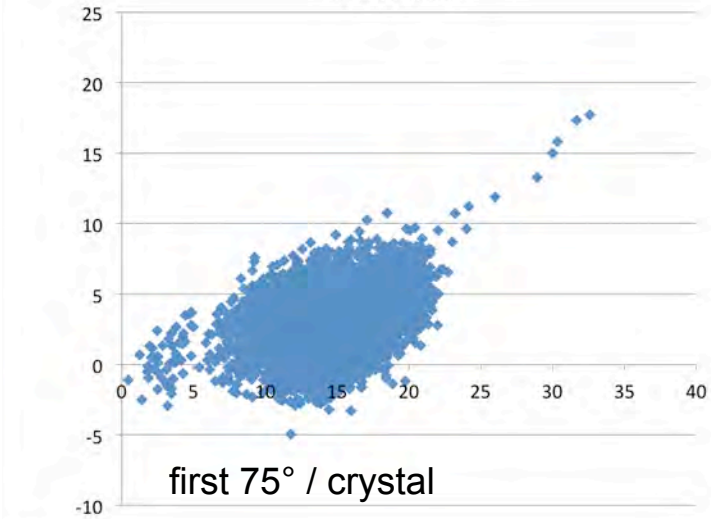
Search for 2 x (6 S-S + 5 S) sites per AU

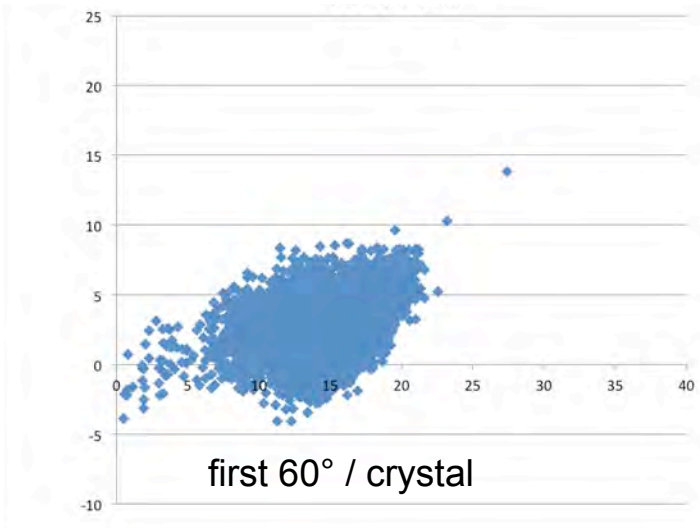
Post mortem of S sub-structure determination: Ultra-high multiplicity needed

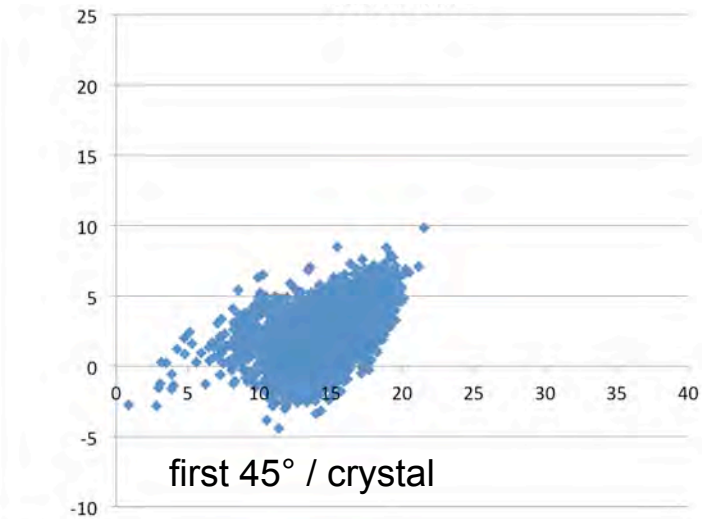
CC(all) *vs*. CC(weak) for various SHELXD trials – “freshest” data from each crystal

Search for 2 x (6 S-S + 5 S) sites per AU with data to 5.2 Å





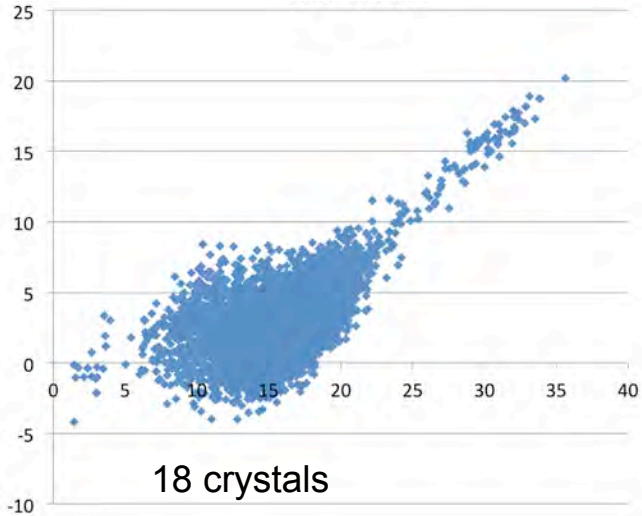




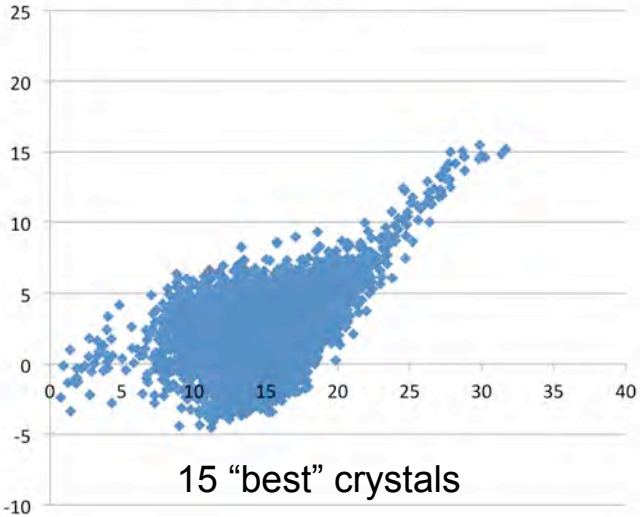
Post mortem of S sub-structure determination: Ultra-high multiplicity needed

CC(all) *vs*. CC(weak) for various SHELXD trials – best subset of crystals

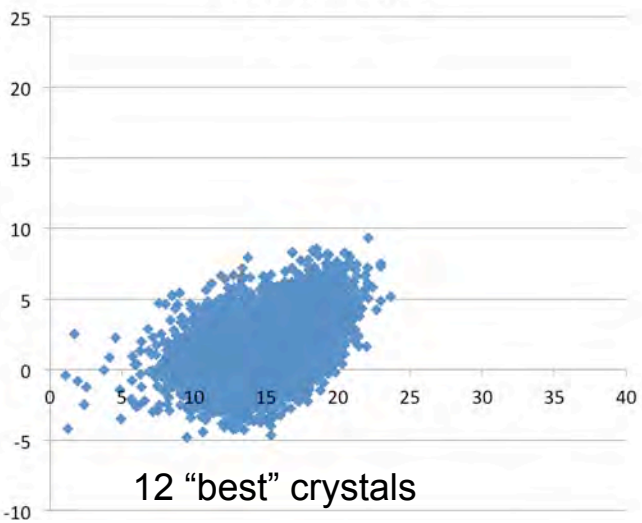
Search for 2 x (6 S-S + 5 S) sites per AU with data to 5.2 Å



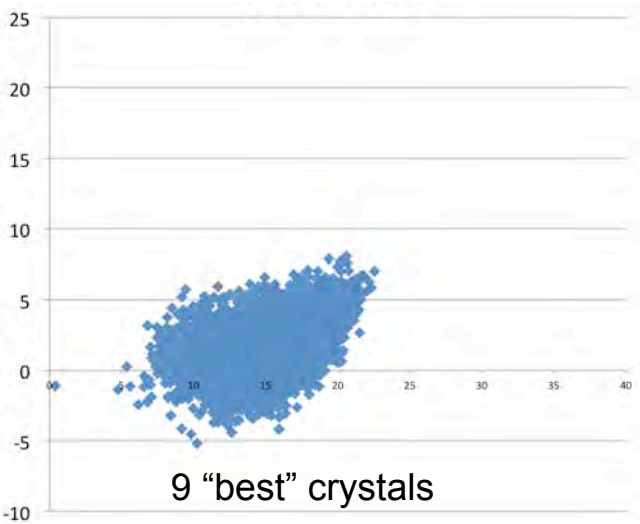
18 crystals



15 “best” crystals



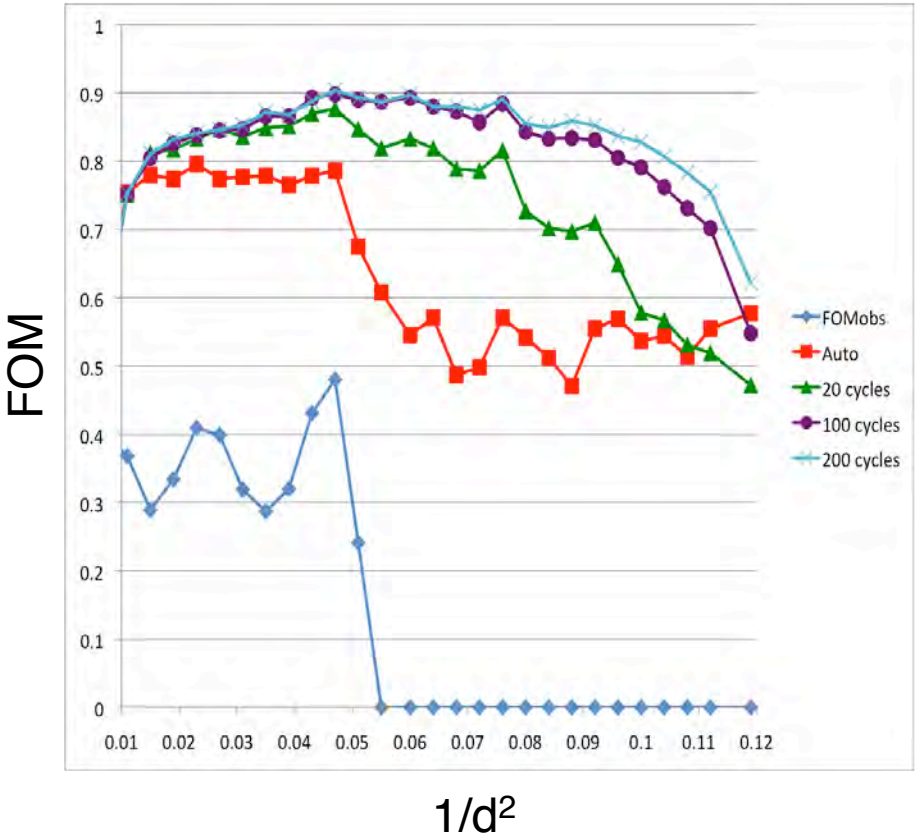
12 “best” crystals



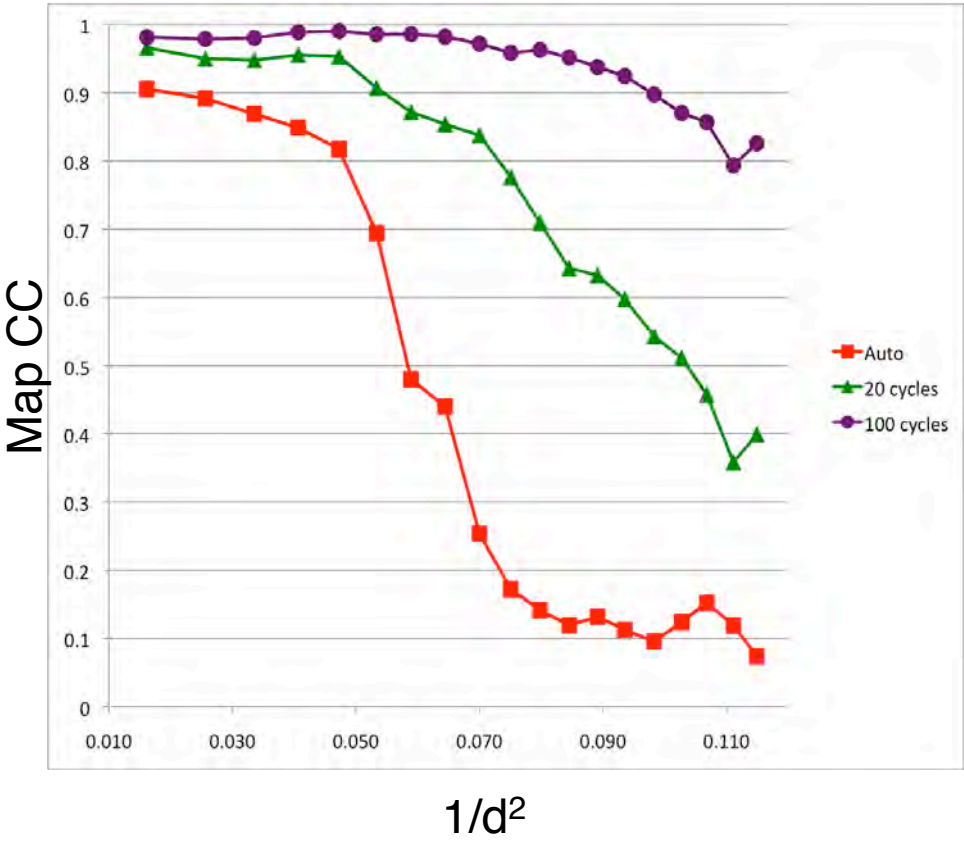
9 “best” crystals

Phase improvement: many steps of phase extension needed

DM FOM

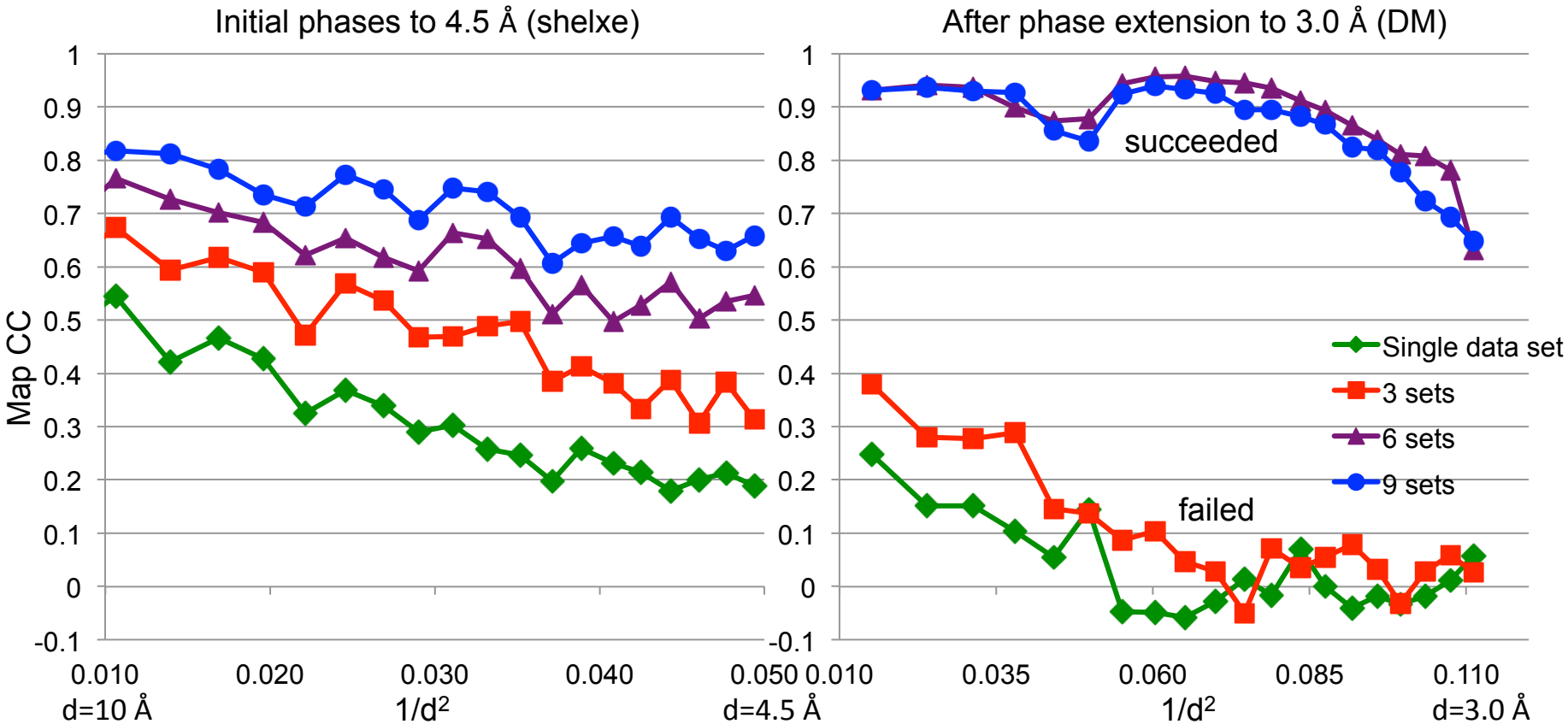


CC of DM map to final map



Successful phase calculation at 4.5 Å (from known sites): high multiplicity needed (data from at least 6 crystals)

Map CC with "18 data set" maps after phase calculation to 4.5 Å and phase extension to 3.0 Å
(S sub-structure solved with all data to 5.2 Å)



Three steps of diminishing difficulty

Difficulty

Solving the S sub-structure

Low-resolution data only (d_{min} 5.2 Å OK; 4.5 Å failed)

Ultra-high multiplicity (100-fold OK; 50-fold failed)

Phase calculation from the S sub-structure

Moderate-resolution data (d_{min} 4.5 Å OK; 3.0 Å failed)

High multiplicity (33-fold OK; 17-fold failed)

Phase refinement (2-fold NCS, 75% solvent)

Extension in small steps (100 steps OK; 20 steps marginal)

Low multiplicity (6-fold OK)

Detect weak phasing signals with ultra-high multiplicity data

Structures from Anomalous Diffraction of Native Biological Macromolecules

Qun Liu,[1] Tassadite Dahmane,[2] Zhen Zhang,[2] Zahra Assur,[2] Julia Brasch,[2] Lawrence Shapiro,[2] Filippo Mancia,[3] Wayne A. Hendrickson[1,2,3,4]*

Science **336**, 1033-1037 (2012)

Reach the true limit of our data

Linking Crystallographic Model and Data Quality

P. Andrew Karplus[1] and Kay Diederichs[2]*

Science **336**, 1030-1033 (2012)

Synergies leading to a new era for macromolecular crystallography

- Extreme multiplicity
- Improved measures of data quality
- Fast, low-noise detectors
- Beamline optics to deliver micro-focus beams
- Software to interrogate samples with a micro-beam
- Automation to screen many samples
- Hot, low-divergence sources, *i.e*. multi-bend achromat

Acknowledgements

Clay Brown – protein production
David Akey – crystallography

EM @ U Michigan: Georgios Skiniotis, Somnath Dutta

Purdue: Richard Kuhn

GM/CA @ APS: Craig Ogata, Steve Corcoran

Photons: APS

Funding: Smith lab @ U Michigan: NIH NIAID, UM Ludwig Professorship
GM/CA: NIH NIGMS & NCI