

Are my data good enough for experimental phasing?

(or, 'Have I solved my structure?')



Gordon Leonard
ESRF Structural Biology
Group

Some Interesting Statistics

Beamline	2007			2008		
	Datasets	PDB Depositions	Depositions/Dataset	Datasets	PDB Depositions	Depositions/Dataset
ID14-1	2027	107	19	2268	131	17
ID14-2	1648	131	13	1677	142	12
ID14-4	1401	133	11	3275	80	41
ID23-1	4391	94	47	4476	129	35
ID23-2	2442	45	54	2264	33	69
ID29	3635	117	31	3554	138	26

Beamline	2008 (Excluding Industry)		
	Datasets	PDB Depositions	Depositions/Dataset
ID14-1	2179	131	17
ID14-2	1576	142	11
ID14-4	2510	80	31
ID23-1	2628	129	20
ID23-2	1924	33	58
ID29	2574	138	19

Average BL requires ~50 data sets to produce 1 PDB deposit (Holton, 2005; Minor 2006, 2008)



So, tunable beamlines are > twice as 'inefficient' as fixed wavelength beamlines. Why?

European Synchrotron Radiation Facility

(More) experimental phasing on tunable beamlines

- Lots of 'derivatives' are not derivatives
 - If no back soaking, then will always see an absorption edge, peak in XFE spectrum.
 - Need to check that heavy atom is bound (how?) before collecting more than one data set
- Many users less experienced/confident when faced with experimental phasing
 - Don't always use best crystals first – frightened of radiation damage (?)
 - Rarely solve structure on beamline, so to make sure collect multiple data sets
 - Not always aware of:
 - Best experiment to do (MAD vs. SAD vs. IR)
 - Indicators 'showing' that structure is solved
 - Tools to rapidly (i.e. *in minutes*) indicate whether structure is solved



Summary

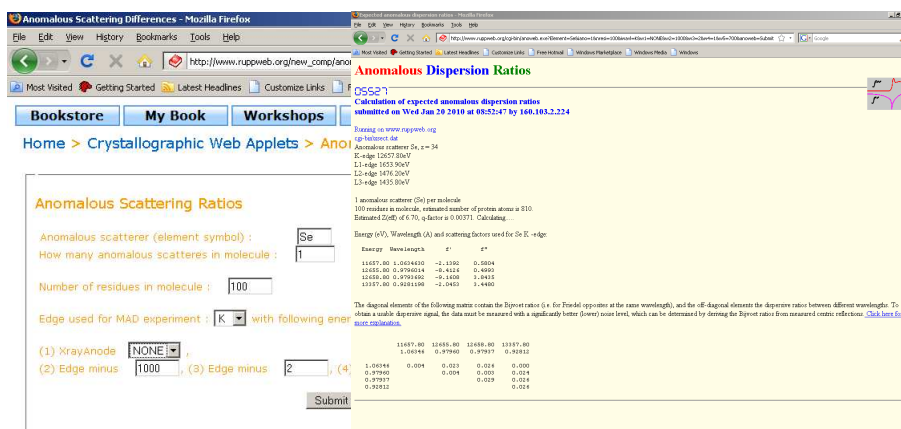
European Synchrotron Radiation Facility

- **There are plenty of rapidly available indicators as to whether data are good enough for experimental phasing. The best one is to actually solve the structure!**
- **Process/analyse data as you go along and take your structure home with you.**
- **Repeat experiments if you have to - but only if you have to.**
- **Once you have left the beamline, its too late.**

- Before arriving at beamline
 - What is the expected anomalous signal?
 - If possible, check that heavy atom is bound
- Before data collection
 - What will the data quality be?
- After data integration and scaling
 - What is the real data quality?
 - Is there any anomalous signal in the data?
 - How far (to what resolution) does the anomalous signal extend?
 - Are more data needed to try to improve things?
- Substructure determination
 - Correct/reasonable number of heavy atom sites
 - Correlation coefficients
- Phasing
 - One hand significantly better than the other?
 - What does 'best' electron density map look like?
 - Does auto-building work?

$$\langle \Delta I / I \rangle_{anom} \approx \frac{1}{\sqrt{2}} \frac{\sqrt{N_A} 2 f''}{\langle |F_T| \rangle}$$

(J. Smith, *Curr. Opin. Struct. Biol.* **1**, 1991)



Anomalous Dispersion Ratios

Calculation of expected anomalous dispersion ratios
submitted on Wed Jun 28 2010 at 08:52:47 by 166.103.2.224

Tracing to www.esrf.org
go to: home page
Accelerator status: On, x = 34
E-edge: 1257.26 keV
L1-edge: 1453.90 keV
L2-edge: 1478.20 keV
L3-edge: 1435.90 keV

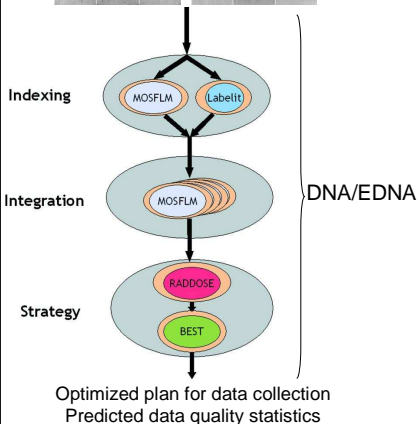
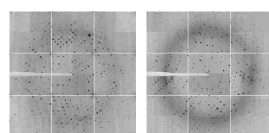
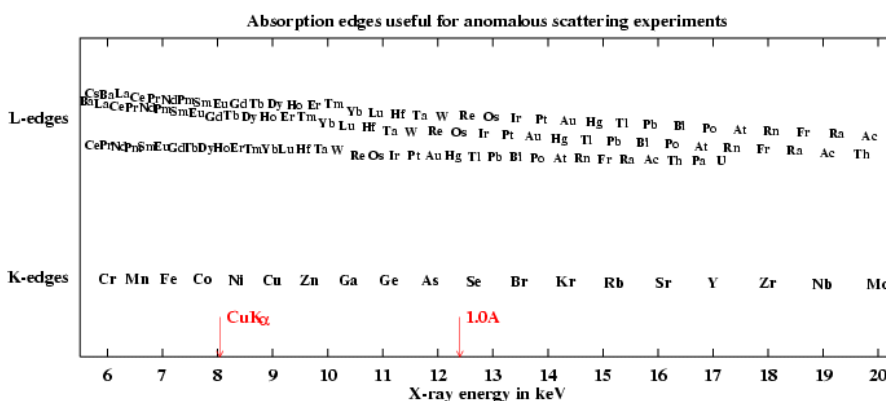
1 anomalous scatterer (Se) per molecule
100 residues in molecule, estimated number of protein atoms is 510.
Estimated SAD of 0.7% at factor is 0.0071. Calculating...

Energy (keV), Wavelength (Å) and scattering factors used for Se K-edge

Energy	Wavelength	f'	f''
11.657 keV	1.064830	-2.1200	0.0004
12.657 keV	0.979244	-0.1212	0.0002
14.058 keV	0.879400	-0.1408	0.0010
14.537 keV	0.853100	-0.1403	0.0002

The diagonal elements of the following matrix contain the Bijou ratio (i.e. for Friedel opposites at the same wavelength), and the off-diagonal elements the dispersion ratios between different wavelengths. To obtain a single dispersion signal, the data must be measured with a significantly better (lower) noise level, which can be determined by dividing the Bijou ratio from measured centrosymmetric reflections. [Click here for more explanation.](#)

	11.657 keV	12.657 keV	14.058 keV	14.537 keV
11.657 keV	1.00000	0.97924	0.87940	0.85310
12.657 keV	0.0004	1.00000	0.0002	0.0002
14.058 keV	0.0010	0.0002	1.00000	0.0002
14.537 keV	0.0002	0.0002	0.0002	1.00000



Optimal Plan of data collection

Resolution is selected according to the user's choice

Attenuation = 1.0000

N	Phi_start	N. of images	Rot. width	Exposure	Distance	Overlap
1	0.00	267	1.35	0.19	232.4	No

Resolution limit : 2.20 Angstrom
Anomalous data : Yes
Phi_start - Phi_finish : 0.00 - 360.45
Total rotation range : 360.45 degree
Total # of images : 267
Overall Completeness : 97.32
Redundancy : 2.00
R-factor (outer shell) : 9.12 (27.12)
I/Sigma (outer shell) : 25.8 (2.8)
Total Exposure time : 50.4 sec (0.014 hour)
Total Data Collection time : 717.9 sec (0.199 hour)

Data collection statistics according to the plan

Resolution	Comp.	Average	I/Sigma	R-factor	Ranom	Overload
Lower Upper	%	Intensity	Sigma	%	%	%
12.00 8.00	100.0	25535.8	505.1	50.6	1.6	1.4
8.00 6.50	100.0	14347.6	294.9	48.7	1.6	1.5
6.50 5.50	100.0	10253.5	221.1	46.4	1.7	1.5
5.50 4.90	98.2	12453.9	244.6	47.3	1.7	1.5
4.90 4.50	98.5	15372.8	319.1	48.2	1.7	1.5
4.50 4.10	97.0	15590.1	323.5	48.2	1.7	1.5
4.10 3.91	100.0	12589.8	271.5	46.4	1.8	1.5
3.91 3.68	100.0	10870.1	242.0	44.9	1.9	1.6
3.68 3.49	99.1	8174.5	203.5	40.2	2.1	1.8
3.49 3.32	98.8	6649.2	180.2	36.9	2.3	1.9
3.32 3.18	96.2	5016.9	161.1	31.1	2.8	2.3
3.18 3.05	97.4	3693.7	144.5	25.6	3.3	2.8
3.05 2.94	97.1	2893.8	127.3	20.6	4.1	3.5
2.94 2.84	96.5	2323.8	132.5	17.5	4.7	4.1
2.84 2.75	98.4	1769.9	131.9	13.6	6.2	6.3
2.75 2.67	97.5	1435.5	120.5	11.0	7.6	6.5
2.67 2.59	95.7	1277.4	131.4	9.7	8.5	7.3
2.59 2.52	96.8	1042.8	129.0	7.5	11.0	9.5
2.52 2.46	97.1	878.7	136.0	6.5	12.7	11.1
2.46 2.40	94.0	788.1	141.6	5.6	14.7	12.9
2.40 2.34	97.2	702.2	147.1	4.9	17.0	15.0
2.34 2.29	96.3	606.1	148.0	4.1	19.0	17.5
2.29 2.24	95.1	549.7	157.2	3.5	22.6	20.5
2.24 2.20	95.7	470.3	159.3	3.0	26.4	24.3
All data	97.3	4532.8	175.7	25.8	3.1	2.8

$$R_{\text{int}} = \frac{\sum_i |I_{\text{int}} - \langle I_{\text{int}} \rangle|}{\sum_i \langle I_{\text{int}} \rangle}$$

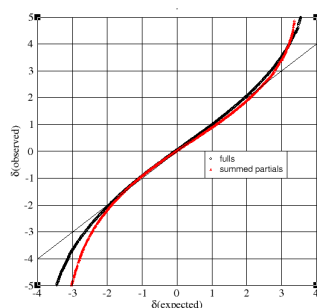
$$R_{\text{merge}} = \frac{\sum_i |I_{\text{int}} - \langle I_{\text{int}} \rangle|}{\sum_i (\langle I_{\text{int}} \rangle + \langle I_{\text{int}} \rangle)}$$

F AE; Se-SAD; S.G. = P2₁2₁2₁; 160° data; λ = 0.97 Å
(ΔF/F ~5.4%)

	N	1/d ²	Dmin(Å)	R _{meq}	R _{full}	R _{cum}	R _{anom}	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	N _{meas}	N _{ref}	N _{cent}	FRCBIAS	N _{bias}
1	0.0250	6.32	0.023	0.019	0.023	0.065	1368	3534	188.0	21.0	159	51.9	10844	1696	334	-0.003	4164	
2	0.0500	4.47	0.026	0.021	0.025	0.058	2693	3951	185.1	21.4	191	47.7	18062	3044	400	-0.004	6979	
3	0.0750	3.65	0.029	0.023	0.027	0.051	3545	5140	269.1	19.1	263	46.0	24252	3980	422	-0.005	9294	
4	0.1000	3.16	0.028	0.023	0.027	0.053	4320	3845	193.5	19.9	202	43.9	29460	4742	463	-0.001	11352	
5	0.1250	2.83	0.029	0.024	0.028	0.063	4930	2328	111.6	20.9	127	41.1	33306	5365	488	0.002	12929	
6	0.1500	2.58	0.034	0.028	0.029	0.072	5515	1654	89.3	18.5	95	38.5	37261	5971	484	0.005	14544	
7	0.1750	2.39	0.037	0.031	0.029	0.072	6015	1384	77.1	17.9	84	35.4	40623	6472	498	0.002	15866	
8	0.2000	2.24	0.039	0.033	0.030	0.072	6481	1241	72.3	17.2	79	33.5	43624	6934	495	0.006	17079	
9	0.2250	2.11	0.044	0.038	0.031	0.075	6852	1098	71.5	15.3	73	31.9	46198	7304	487	0.003	18090	
10	0.2500	2.00	0.054	0.046	0.033	0.080	7265	855	66.1	12.9	61	29.2	49015	7709	484	0.006	19251	

$$R_{\text{merge}} = \sum_{\text{all}} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{\text{all}} I_i(hkl)$$

Normal probability plot



$$\delta_{hkl} = (I_{hkl} - \langle I_{hkl} \rangle) / \sigma^2(I_{hkl})$$

Slope = 1; Intercept = 0

```

=====
Normal probability analysis, by run & partiality

===== Run number: 1, fulls =====
Number      Slope Intercept
All data: 182820 1.013 0.048
Data within expected delta 0.90: 115520 0.941 0.058

===== Run number: 1, summed partials =====
Number      Slope Intercept
All data: 150209 1.024 -0.024
Data within expected delta 0.90: 94813 0.884 0.013

===== Run number: 1, fulls against fulls only =====
Number      Slope Intercept
All data: 158267 1.003 0.012
Data within expected delta 0.90: 99967 0.932 0.021
=====

```

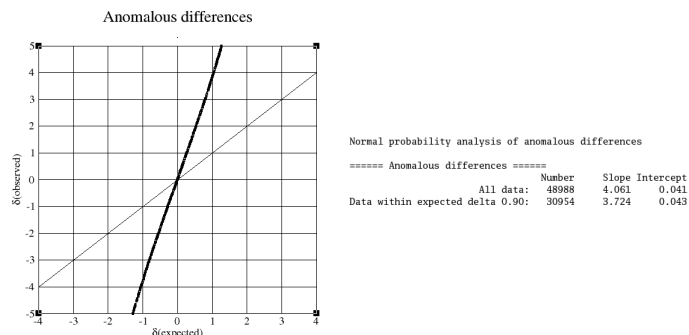
#	1/d ²	Dmin(Å)	R _{avg}	R _{full}	R _{cum}	R _{anom}	Nanom	Av_I	SIGMA	I/sigma	sd	ln(I/sd)	N _{mean}	N _{ref}	N _{cent}	FRCBIAS	N _{bias}
1	0.0250	6.32	0.023	0.019	0.023	0.065	1368	3534	168.0	21.0	159	51.9	10844	1696	334	-0.003	4164
2	0.0500	4.47	0.026	0.021	0.025	0.058	2693	3951	185.1	21.4	191	47.7	18062	3044	400	-0.004	6979
3	0.0750	3.65	0.029	0.023	0.027	0.051	3545	5140	269.1	19.1	263	46.0	24252	3980	422	-0.005	9294
4	0.1000	3.16	0.028	0.023	0.027	0.053	4320	3845	193.5	19.9	202	43.9	29460	4742	463	-0.001	11352
5	0.1250	2.83	0.029	0.024	0.028	0.063	4930	2328	111.6	20.9	127	41.1	33306	5365	488	0.002	12929
6	0.1500	2.58	0.034	0.028	0.029	0.072	5515	1654	89.3	18.5	95	38.5	37261	5971	484	0.005	14544
7	0.1750	2.39	0.037	0.031	0.029	0.072	6015	1384	77.1	17.9	84	35.4	40623	6472	498	0.002	15866
8	0.2000	2.24	0.039	0.033	0.030	0.072	6481	1241	72.3	17.2	79	33.5	43624	6934	495	0.006	17079
9	0.2250	2.11	0.044	0.038	0.031	0.075	6852	1098	71.5	15.3	73	31.9	46198	7304	487	0.003	18090
10	0.2500	2.00	0.054	0.046	0.033	0.080	7265	855	66.1	12.9	61	29.2	49015	7709	484	0.006	19251

$$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$$

$$R_{anom} = \sum_{hkl} |I^+(hkl) - I^-(hkl)| / \sum_{hkl} |I^+(hkl) + I^-(hkl)| / 2$$

$$R_{anom} > R_{merge} ?$$

The presence of anomalous signal can be deduced from a normal probability plot of the normalized differences $\delta_{anom} = (I^+ - I^-) / [(\sigma^2(I^+) + \sigma^2(I^-))^{1/2}]$. The slope of the central region of this plot will be > 1 if the anomalous differences are larger than expected from their standard deviations (i.e. if there is significant anomalous signal)

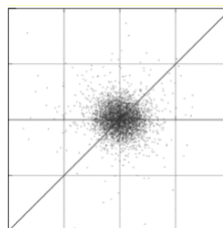
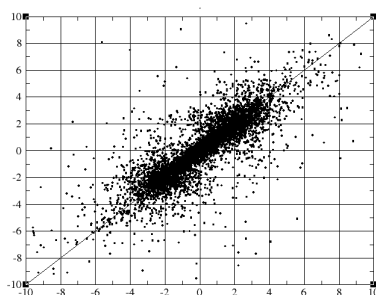


(Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86; Evans, P. (2006). *Acta Cryst.* **D62**, 72-82.)

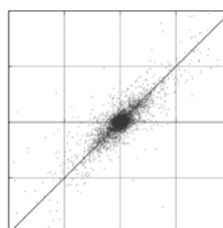
What is the correlation coefficient between anomalous differences in two half data sets chosen at random? What do scatter plots look like?

N	1/resol ²	dmax	CC_anom	N_anom	CC_cen	N_cen	RCR_anom	N_anom	RCR_cen	N_cen	CC_1mean	N_1mean
1	0.0250	6.32	0.930	1353	-0.108	187	5.245	1353	0.891	187	0.999	1558
2	0.0500	4.47	0.853	2526	-0.024	215	3.597	2526	1.022	215	0.999	2869
3	0.0750	3.65	0.760	3555	-0.090	312	2.708	3555	0.915	312	0.998	3809
4	0.1000	3.16	0.821	4117	-0.129	336	3.191	4117	0.880	336	0.999	4571
5	0.1250	2.83	0.868	4710	-0.145	364	3.759	4710	0.866	364	0.999	5225
6	0.1500	2.58	0.877	5362	0.304	378	3.900	5362	1.362	378	0.999	5863
7	0.1750	2.39	0.865	5848	-0.058	384	3.719	5848	0.940	384	0.998	6374
8	0.2000	2.24	0.855	6279	0.052	373	3.580	6279	1.053	373	0.998	6823
9	0.2250	2.11	0.828	6647	-0.011	379	3.259	6647	0.991	379	0.998	7186
10	0.2500	2.00	0.795	7042	-0.091	383	2.956	7042	0.910	383	0.997	7606

DelAnom scatter plot



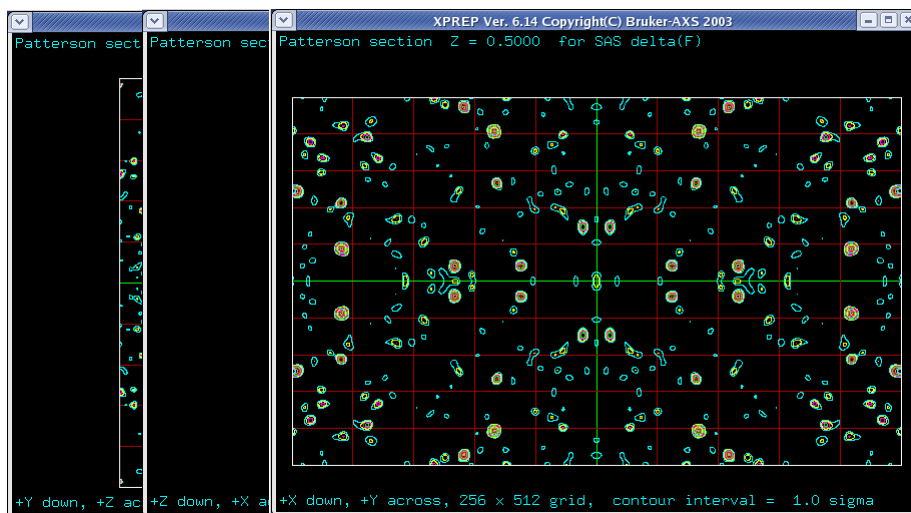
no
anomalous
signal



Strong
anomalous
signal

Summary data for Project: Unspecified Crystal: Unspecified Dataset: Unspecified

	Overall	InnerShell	OuterShell
Low resolution limit	39.81	39.81	2.11
High resolution limit	2.00	6.32	2.00
Rmerge	0.033	0.023	0.054
Rmerge in top intensity bin	0.025	-	-
Rmeas (within I+/I-)	0.039	0.027	0.065
Rmeas (all I+ & I-)	0.075	0.073	0.103
Rpim (within I+/I-)	0.021	0.015	0.035
Rpim (all I+ & I-)	0.030	0.028	0.040
Fractional partial bias	0.000	-0.003	0.006
Total number of observations	335117	10954	49295
Total number unique	53920	1724	7768
Mean(I)/sd(I)	35.6	54.3	24.7
Completeness	98.1	93.4	97.6
Multiplicity	6.2	6.4	6.3
Anomalous completeness	98.0	97.5	97.5
Anomalous multiplicity	3.2	3.6	3.3
DelAnom correlation between half-sets	0.833	0.930	0.795
Mid-Slope of Anom Normal Probability	3.724	-	-



HKL2MAP: Pape T & Schneider TR (2004).
HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.* **37**:843-844.

SHELXC: Sheldrick GM (2003). Goettingen University.

SHELXD: Schneider TR, Sheldrick GM (2002). Substructure solution with SHELXD. *Acta Cryst. D* **58**:1772-1779.

SHELXE: Sheldrick GM (2002). Macromolecular phasing with SHELXE. *Z. Kristallogr.* **217**:644-650.


```

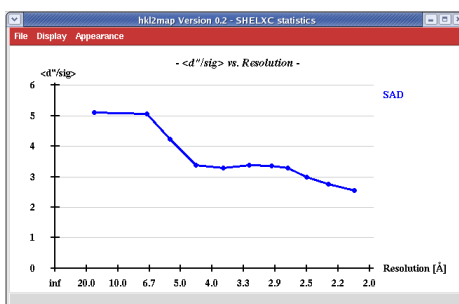
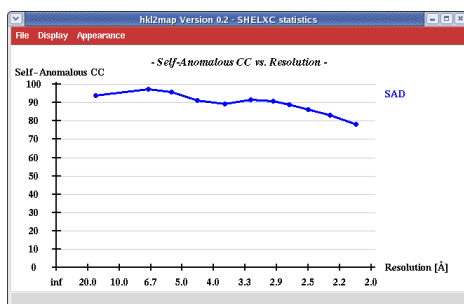
335117 Reflections read from SAD file fae-pk.sca
53892 Unique reflections, highest resolution 2.000 Angstroms
263.5 Friedel pairs used on average for local scaling

Resl.   Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.00
N(data) 858 1162 1458 3316 3364 5981 3685 4938 6644 9256 13230
Chi-sq   3.76 1.61 0.98 1.65 1.44 1.04 0.84 0.84 0.80 0.75 0.95
<I/sig>  58.1 53.0 50.5 53.1 51.3 44.8 38.6 35.0 32.1 29.7 26.2
RComplete 85.6 93.4 93.8 95.9 97.7 98.3 98.5 98.6 98.3 98.2 97.7
<d*/sig>  5.11 5.07 4.23 3.37 3.30 3.37 3.36 3.28 2.98 2.76 2.55
CC(anom)  94.1 97.4 95.8 91.4 89.5 91.5 90.9 88.9 86.4 83.4 78.3

For zero signal <d*/sig> and <d*/sig> should be about 0.80

47418 Reflections written to file FAE-pk_fa.hkl for input to SHELXD/E
53892 Reflections written to file FAE-pk.hkl for input to SHELXE

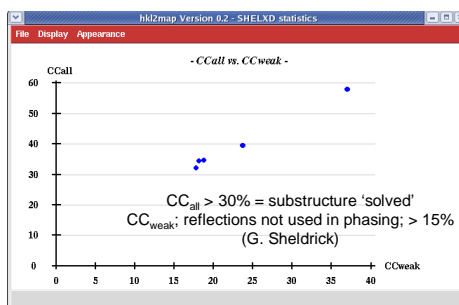
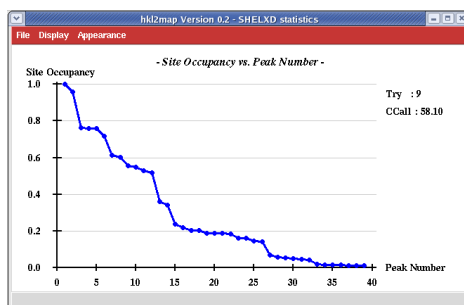
```



```

PSUM 33.80 PSMP Peaks: 182 139 121 101 101 99 85 77 72 69 61 57 46 41 31 26
Try 64:20 Peaks 99 95 76 75 74 71 60 59 55 54 52 51 35 34 23 21 20 19 18 18
R = 0.163, Min.fun. = 0.299, <cos> = 0.561, Ra = 0.192
Try 64, CC All/Weak 58.09 / 37.01, best 58.10 / 37.04, best PATFOM 7.35
PATFOM 7.34

```



$$CC = \frac{100 [\sum (wE_o E_c) \sum w - \sum (wE_o) \sum (wE_c)]}{\{ [\sum (wE_o^2) \sum w - (\sum wE_o)^2] \cdot [\sum (wE_c^2) \sum w - (\sum wE_c)^2] \}^{1/2}}$$

Fujinaga & Read, *J. Appl. Cryst.* 20 (1987) 517-521.

```

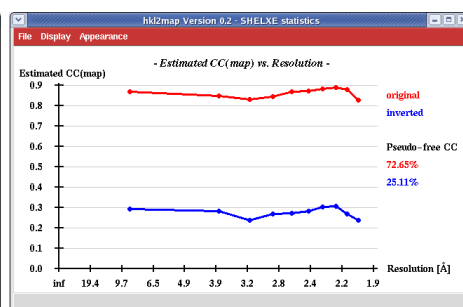
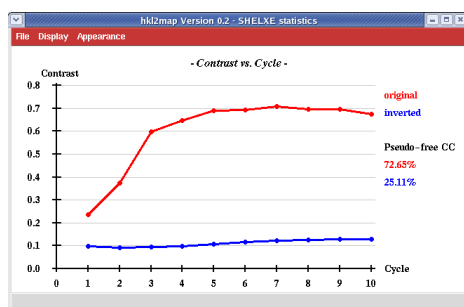
/SHELXC/SHELXD/SHELXE original/SHELXE inverted
<vt> = 0.630 after including heavy atoms
<vt> = 0.300, Contrast = 0.236, Connect = 0.872 for dens mod cycle 1
<vt> = 0.300, Contrast = 0.375, Connect = 0.898 for dens mod cycle 2
<vt> = 0.300, Contrast = 0.597, Connect = 0.913 for dens mod cycle 3
<vt> = 0.300, Contrast = 0.648, Connect = 0.917 for dens mod cycle 4
Pseudo-free CC = 71.96 %
<vt> = 0.300, Contrast = 0.691, Connect = 0.918 for dens mod cycle 5
<vt> = 0.300, Contrast = 0.693, Connect = 0.919 for dens mod cycle 6
<vt> = 0.300, Contrast = 0.708, Connect = 0.920 for dens mod cycle 7
<vt> = 0.300, Contrast = 0.696, Connect = 0.921 for dens mod cycle 8
<vt> = 0.300, Contrast = 0.699, Connect = 0.921 for dens mod cycle 9
<vt> = 0.300, Contrast = 0.676, Connect = 0.921 for dens mod cycle 10
Mean weight and estimated mapCC as a function of resolution
d inf - 4.32 - 3.43 - 3.00 - 2.73 - 2.53 - 2.38 - 2.26 - 2.16 - 2.08 - 2.01
<vt> 0.688 0.659 0.645 0.666 0.696 0.704 0.712 0.709 0.691 0.643
<mapCC> 0.889 0.851 0.831 0.846 0.869 0.873 0.885 0.890 0.879 0.828
N 5459 5434 5429 5339 5327 5443 5459 5448 5193 5203
Pseudo-free CC = 72.65 %

```

```

/SHELXC/SHELXD/SHELXE original/SHELXE inverted
<vt> = 0.577 after including heavy atoms
<vt> = 0.300, Contrast = 0.699, Connect = 0.689 for dens mod cycle 1
<vt> = 0.300, Contrast = 0.692, Connect = 0.726 for dens mod cycle 2
<vt> = 0.300, Contrast = 0.695, Connect = 0.783 for dens mod cycle 3
<vt> = 0.300, Contrast = 0.699, Connect = 0.815 for dens mod cycle 4
Pseudo-free CC = 16.50 %
<vt> = 0.300, Contrast = 0.108, Connect = 0.835 for dens mod cycle 5
<vt> = 0.300, Contrast = 0.115, Connect = 0.850 for dens mod cycle 6
<vt> = 0.300, Contrast = 0.122, Connect = 0.859 for dens mod cycle 7
<vt> = 0.300, Contrast = 0.126, Connect = 0.867 for dens mod cycle 8
<vt> = 0.300, Contrast = 0.129, Connect = 0.868 for dens mod cycle 9
<vt> = 0.300, Contrast = 0.127, Connect = 0.875 for dens mod cycle 10
Mean weight and estimated mapCC as a function of resolution
d inf - 4.32 - 3.43 - 3.00 - 2.73 - 2.53 - 2.38 - 2.26 - 2.16 - 2.08 - 2.01
<vt> 0.219 0.230 0.200 0.223 0.226 0.234 0.247 0.249 0.218 0.190
<mapCC> 0.282 0.282 0.219 0.268 0.272 0.281 0.303 0.308 0.269 0.236
N 5459 5434 5429 5339 5327 5443 5459 5448 5193 5203
Pseudo-free CC = 25.11 %

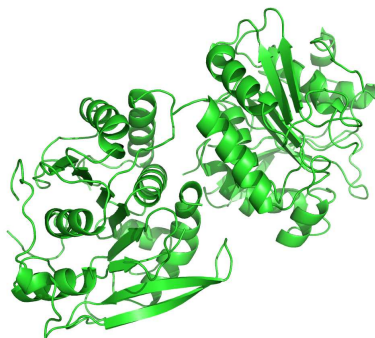
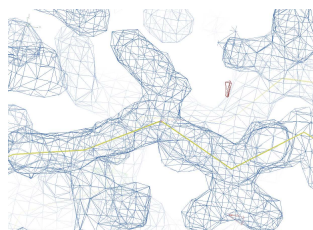
```

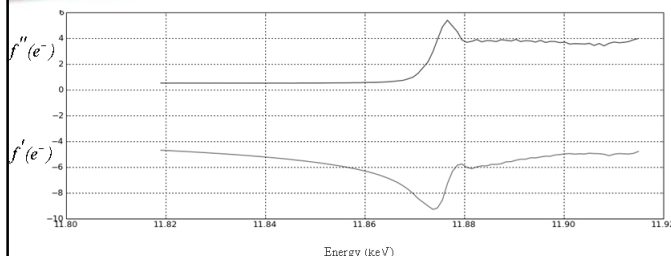


```

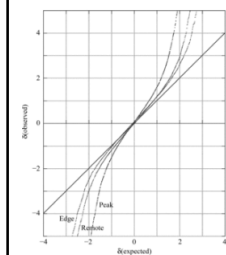
CCP4I Refinement 1.47f - map.mrg
Found 53 (53 requested) and removed 9 (48 requested) atoms.
Cycle 43: After refine, R = 0.210 (Rfree = 0.000).
Found 35 (35 requested) and removed 15 (16 requested) atoms.
Cycle 44: After refine, R = 0.207 (Rfree = 0.000).
Found 35 (35 requested) and removed 18 (18 requested) atoms.
Cycle 45: After refine, R = 0.204 (Rfree = 0.000).
Found 35 (35 requested) and removed 19 (18 requested) atoms.
Building cycle 5: atomic shape factors 1.02 1.40
Found 1: 546 peptides in 7 chains. The longest chain comprises 177 peptides.
Found 2: 546 peptides in 18 chains. The longest chain comprises 155 peptides.
Taking the entire row found 1
Chain 10: Residues 136. Estimated correctness of the model 99.4 %
----> Final restrained refinement block. No atom update.
SADIR reflections ( 99.77 % complete ) and 2139 restraints for refining 1271 atoms.
(Observed/parameters ratio is 2.53)
Cycle 46: After refine, R = 0.229 (Rfree = 0.000).
Found 0 (52 requested) and removed 0 (39 requested) atoms.
Cycle 47: After refine, R = 0.223 (Rfree = 0.000).
Found 0 (52 requested) and removed 7 (38 requested) atoms.
Cycle 48: After refine, R = 0.223 (Rfree = 0.000).
Found 0 (52 requested) and removed 6 (36 requested) atoms.
Cycle 49: After refine, R = 0.225 (Rfree = 0.000).

```

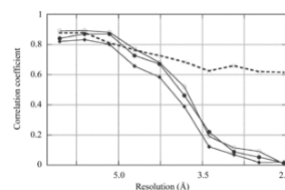




$$f''_{peak} > f''_{remote} > f''_{edge}$$



$$Slope_{peak} > Slope_{remote} > Slope_{edge}$$



$$CC_{anom} (pk/ip) \\ CC_{anom} (pk/rm) \\ CC_{anom} (ip/rm)$$

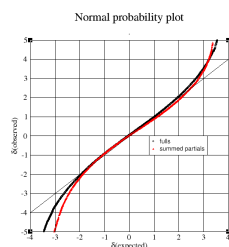
Evans, P. (2006). *Acta Cryst.* **D62**, 72-82

$$\left\langle \frac{\Delta I}{I} \right\rangle = \sqrt{\frac{N_H}{2N_p}} f_H / f_p$$

N_H, N_p : # Heavy, # (equal) light atoms of scattering factors f_H, f_p respectively

$$\delta = (|F_{pH}| - |F_p|) / (\sigma^2(|F_{pH}| + \sigma^2(|F_p|)))$$

$$Slope \gg 1; Intercept > 0$$



$$\delta_{hkl} = (I_{hkl} - \langle I_{hkl} \rangle) / \sigma^2(I_{hkl}) \\ Slope = 1; Intercept = 0$$

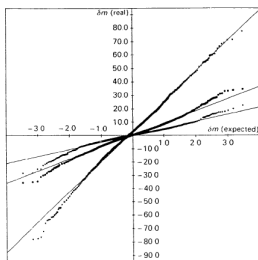


Fig. 4. Plot of δ_{real} versus $\delta_{expected}$ for native macromycin data and three heavy-atom derivatives; x is $K_2Pt(NO_3)_4$; ■ is $HgCl_2$; ♦ is $(NH_4)_2HgCl_4$.

Table 3. A comparison of three heavy-atom derivatives with native data for macromycin as a function of resolution

The residual is defined as $\sum (F_p - F_H) / \sum F_H$. The slope and intercept of the least-squares straight line are defined as $Slope = \Delta I / \Delta F_H$ and $Intercept = I - Slope \cdot F_H$. Numbers in parentheses are the slope and intercept of the least-squares straight line calculated from these data with $(Intercept) < 0.074$ (10% of the data).

Resolution (Å)	Number of data	Residual	Slope	Intercept
(1) $K_2Pt(NO_3)_4$				
10-6	198	0.325	27.77 (27.66)	4.62 (5.06)
10-5	189	0.317	26.63 (26.32)	3.70 (4.04)
10-4	111	0.280	23.75 (23.47)	3.61 (4.11)
10-3	795	0.307	23.50 (23.04)	3.30 (3.91)
10-2	1927	0.259	22.58 (22.54)	2.75 (2.96)
10-1	8910	0.306	22.50 (22.46)	2.66 (2.78)
10-0	488	0.220	20.73 (20.33)	2.66 (2.84)
5-0	1558	0.251	21.38 (21.05)	2.46 (2.49)
(2) $HgCl_2$				
10-6	199	0.146	11.43 (10.92)	1.34 (1.41)
10-5	190	0.151	11.47 (10.96)	0.73 (0.72)
10-4	112	0.148	11.03 (10.64)	1.07 (0.95)
10-3	797	0.155	10.72 (10.77)	0.63 (0.57)
10-2	1929	0.157	9.96 (9.96)	0.39 (0.36)
10-1	1872	0.116	8.92 (8.49)	0.32 (0.30)
10-0	489	0.112	8.75 (8.59)	0.22 (0.21)
5-0	1539	0.107	8.35 (8.94)	0.30 (0.27)
(3) $(NH_4)_2HgCl_4$				
10-6	199	0.141	9.90 (9.81)	0.43 (0.52)
10-5	200	0.127	8.14 (7.90)	0.09 (0.25)
10-4	112	0.115	7.58 (6.50)	0.18 (0.30)
10-3	797	0.097	4.63 (5.09)	0.10 (0.25)
10-2	1929	0.092	5.26 (4.72)	0.09 (0.19)
10-1	1872	0.090	5.08 (4.61)	0.07 (0.16)
10-0	489	0.087	4.87 (4.56)	0.12 (0.25)
5-0	1539	0.083	4.30 (4.15)	0.07 (0.06)

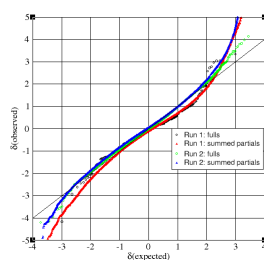
Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86

XXX; Zn-SAD; S.G. = P1; 360° data; $\lambda = 1.28 \text{ \AA}$
($\Delta F/F \sim 4.2\%$)

N	1/d ²	hmin(A)	hmax	Full	Sum	Random	Mean	AV_I	SIGMA	I/sigma	ed	No(I/ed)	Mean	Nref	Ncent	FRC(IAS)	Nbins
44																	
1	0.0108	9.62	0.048	0.045	0.048	0.048	337	29765	2789	10.7	2329	30.0	2481	329	0	-0.035	905
2	0.0216	6.80	0.044	0.024	0.046	0.054	666	19835	1600	12.4	1639	27.1	4927	666	0	-0.023	1643
3	0.0324	5.58	0.044	0.047	0.051	0.062	881	10446	1275	8.2	1027	22.2	6226	884	0	-0.009	2163
4	0.0432	4.81	0.060	0.048	0.054	0.060	1013	14543	1550	9.4	1366	23.7	6923	1009	0	-0.033	2464
5	0.0541	4.30	0.055	0.071	0.054	0.050	1162	20691	1979	10.5	1830	25.6	8416	1182	0	-0.033	3335
6	0.0649	3.90	0.066	0.069	0.057	0.052	1192	17160	1971	8.7	1412	24.4	9413	1239	0	-0.022	3777
7	0.0757	3.63	0.070	0.062	0.059	0.050	1186	14246	1712	8.3	1427	22.6	10143	1186	0	-0.020	4040
8	0.0865	3.40	0.072	0.063	0.061	0.050	1480	12324	1491	8.4	1324	20.9	10960	1485	0	-0.018	4412
9	0.0973	3.21	0.082	0.075	0.063	0.057	1400	8687	1206	7.2	1044	18.3	11887	1599	0	-0.005	4761
10	0.1082	3.04	0.095	0.069	0.065	0.066	1684	5916	925	6.4	840	15.3	11531	1692	0	0.001	5102
11	0.1190	2.90	0.112	0.165	0.087	0.073	1772	4408	744	5.5	719	13.1	11286	1774	0	0.003	5510
12	0.1308	2.78	0.129	0.133	0.070	0.075	1834	3612	727	5.0	700	11.8	11769	1836	0	0.011	5547
13	0.1406	2.67	0.161	0.169	0.073	0.089	1880	2831	725	3.9	661	9.8	14094	1879	0	0.014	5829
14	0.1514	2.57	0.170	0.204	0.075	0.103	1992	2337	614	3.6	610	7.9	11398	2002	0	0.014	6265
15	0.1622	2.48	0.130	0.515	0.075	0.114	1953	2189	433	5.3	548	6.4	7504	1976	0	-0.014	2268
16	0.1731	2.40	0.157	0.311	0.078	0.121	1984	1851	415	4.8	551	5.8	7882	2020	0	-0.022	2474
17	0.1839	2.33	0.201	0.552	0.077	0.157	2134	1489	427	3.5	559	4.7	8256	2150	0	-0.019	2567
18	0.1947	2.27	0.223	0.436	0.079	0.170	2115	1391	442	3.1	585	4.3	8142	2138	0	-0.029	2561
19	0.2055	2.21	0.278	0.278	0.080	0.208	2203	1209	495	2.6	618	3.7	8526	2214	0	-0.024	2592
20	0.2163	2.15	0.326	0.567	0.082	0.247	2191	1075	508	2.1	642	3.2	7916	2108	0	-0.018	2487

$$R_{\text{merge}} = \frac{\sum_i \sum_j |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_i \sum_j I_i(hkl)}$$

Normal probability plot



Normal probability analysis, by run & partiality

===== Run number: 1, Run 1: fulls =====	Number	Slope	Intercept
All data:	364	0.898	-0.076
Data within expected delta 0.90:	230	0.896	-0.084
===== Run number: 1, Run 1: summed partials =====	Number	Slope	Intercept
All data:	6022	0.981	-0.148
Data within expected delta 0.90:	4178	0.937	-0.129
===== Run number: 1, Run 1: fulls against fulls only =====	Number	Slope	Intercept
All data:	80	0.878	-0.065
Data within expected delta 0.90:	38	0.823	-0.069
===== Run number: 2, Run 2: fulls =====	Number	Slope	Intercept
All data:	5252	0.987	0.013
Data within expected delta 0.90:	3318	0.921	-0.010
===== Run number: 2, Run 2: summed partials =====	Number	Slope	Intercept
All data:	115672	0.990	0.060
Data within expected delta 0.90:	73090	0.881	0.060
===== Run number: 2, Run 2: fulls against fulls only =====	Number	Slope	Intercept
All data:	640	0.993	0.056
Data within expected delta 0.90:	404	0.885	0.004

$$R_{anom} = \sum_{hkl} |I^+(hkl) - I^-(hkl)| / \sum_{hkl} |I^+(hkl) + I^-(hkl)| / 2$$

 $R_{anom} > R_{merge} ?$ Hmmm.....

#	W (Å)	σ_{rel}	Effail	Rate	Score	Av	1/20th	1/10th	50th	Med	90th	Score	Ref	Score	Ref	Score	
0	0.0100	6.41	0.0000	0.000	0.000	37	20705	10705	7	23329	10300	2481	319	0	-0.005	905	
0	0.0210	6.41	0.0000	0.000	0.000	66	19935	10000	4	1429	271	907	466	0	-0.003	1643	
0	0.0234	5.55	0.0044	0.001	0.001	66	19446	1275	0	1077	72.2	8216	804	0	-0.009	2163	
0	0.0234	5.55	0.0044	0.001	0.001	66	19446	1275	0	1077	72.2	8216	804	0	-0.009	2163	
0	0.0234	5.55	0.0044	0.001	0.001	66	19446	1275	0	1077	72.2	8216	804	0	-0.009	2163	
0	0.0541	4.30	0.0055	0.001	0.004	1052	11629	10919	1079	105	1839	25.6	8416	1382	0	-0.003	3335
0	0.0449	3.29	0.0048	0.001	0.002	1282	17100	1071	107	1162	24.4	9413	1389	0	-0.022	3779	
0	0.0449	3.29	0.0048	0.001	0.002	1282	17100	1071	107	1162	24.4	9413	1389	0	-0.022	3779	
0	0.0865	3.40	0.0092	0.003	0.004	1488	12524	1041	64	1324	20.9	10988	1485	0	-0.018	4412	
0	0.0865	3.40	0.0092	0.003	0.004	1488	12524	1041	64	1324	20.9	10988	1485	0	-0.018	4412	
0	0.0865	3.40	0.0092	0.003	0.004	1488	12524	1041	64	1324	20.9	10988	1485	0	-0.018	4412	
0	1.0002	3.04	0.0095	0.005	0.006	1664	5916	925	64	140	15.3	12351	1602	0	-0.001	5102	
0	1.1010	2.50	0.0112	0.005	0.007	1772	4200	764	5	719	13.1	11326	1774	0	-0.003	5518	
0	1.1010	2.50	0.0112	0.005	0.007	1772	4200	764	5	719	13.1	11326	1774	0	-0.003	5518	
0	1.1010	2.50	0.0112	0.005	0.007	1772	4200	764	5	719	13.1	11326	1774	0	-0.003	5518	
0	1.1406	2.47	0.1161	0.169	0.073	1890	2831	125	6	961	9.8	10494	1879	0	0.014	5829	
0	1.1406	2.47	0.1161	0.169	0.073	1890	2831	125	6	961	9.8	10494	1879	0	0.014	5829	
0	1.1406	2.47	0.1161	0.169	0.073	1890	2831	125	6	961	9.8	10494	1879	0	0.014	5829	
0	1.1622	2.48	0.1130	0.1515	0.075	1914	2963	1289	413	5	548	6.4	7504	1076	0	-0.014	2268
0	1.1731	2.40	0.1571	0.331	0.076	2114	1894	1861	115	6	551	5.6	7662	2000	0	-0.024	2474
0	1.1731	2.40	0.1571	0.331	0.076	2114	1894	1861	115	6	551	5.6	7662	2000	0	-0.024	2474
0	1.1731	2.40	0.1571	0.331	0.076	2114	1894	1861	115	6	551	5.6	7662	2000	0	-0.024	2474
0	1.1947	2.27	0.2123	0.416	0.079	2110	1715	1159	142	3	595	4.3	8142	2138	0	-0.029	2581
0	1.1947	2.27	0.2123	0.416	0.079	2110	1715	1159	142	3	595	4.3	8142	2138	0	-0.029	2581
0	1.1947	2.27	0.2123	0.416	0.079	2110	1715	1159	142	3	595	4.3	8142	2138	0	-0.029	2581
0	2.1610	2.15	0.3256	0.687	0.282	2431	1075	1008	4	642	3.2	7916	2308	0	-0.018	2487	

R_{p.i.m.} (Precision-Indicating Merging R factor; Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130-135; N = number of observations of a given reflection)
is a measure of the quality of the data after averaging the multiple measurements.

$$R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_i \sum_{hkl} I_i(hkl)$$

$$R_{anom} = \sum_{hkl} |I^+(hkl) - I^-(hkl)| / \sum_{hkl} |I^+(hkl) + I^-(hkl)| / 2$$

$$R_{n/m} = \sum_{hkl} [1/(N-1)^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_i \sum_{hkl} I_i(hkl)]$$

Ranom	Rp.i.m
0.048	0.029
0.054	0.026
0.062	0.040
0.060	0.038
0.050	0.034
0.052	0.040
0.050	0.042
0.050	0.044
0.057	0.049
0.066	0.058
0.073	0.068
0.075	0.078
0.089	0.096
0.103	0.119
0.114	0.130
0.111	0.157
0.157	0.201
0.170	0.223
0.208	0.278
0.247	0.326
0.067	0.054

A better indicator of anomalous signal is, perhaps, the ratio $R_{\text{anom}}/R_{\text{p.i.m.}}$; R_{anom} = signal, $R_{\text{p.i.m.}}$ = noise. (Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771-777; Panjikar, S. & Tucker, P. (2002). *J. Appl. Cryst.* **35**, 261-266.)

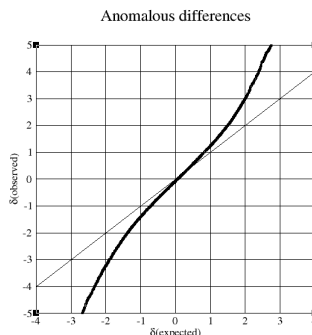
$$R_{\text{anom}} / R_{\text{p.i.m.}} > 1.2?$$

The presence of anomalous signal can be deduced from a normal probability plot of the normalized differences $\delta_{\text{anom}} = (I^+ - I^-) / [(\sigma^2(I^+) + \sigma^2(I^-))]^{1/2}$. The slope of the central region of this plot will be > 1 if the anomalous differences are larger than expected from their standard deviations (i.e. if there is significant anomalous signal)

Normal probability analysis of anomalous differences

```

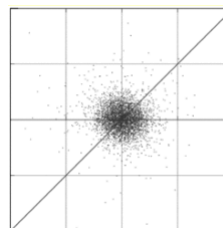
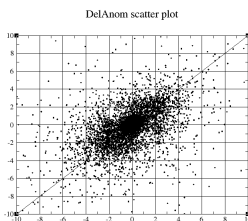
===== Anomalous differences =====
Number      Slope Intercept
All data:   31567  1.549  -0.050
Data within expected delta 0.90: 19947  1.351  -0.045
  
```



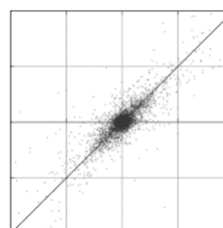
(Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86; Evans, P. (2006). *Acta Cryst.* **D62**, 72-82.)

What is the correlation coefficient between anomalous differences in two half data sets chosen at random? Where is the point that CC_{anom} drops below 30%?

N	1/resol ²	dhkl	CC_anom	N_anom	CC_cen	N_cen	RCR_anom	N_anom	RCR_cen
1	0.0108	9.62	0.749	316	-	0	2.627	316	-
2	0.0216	6.80	0.697	651	-	0	2.165	651	-
3	0.0324	5.55	0.511	855	-	0	1.758	855	-
4	0.0432	4.81	0.578	974	-	0	1.916	974	-
5	0.0540	4.30	0.484	1099	-	0	1.701	1099	-
6	0.0648	3.93	0.393	1257	-	0	1.517	1257	-
7	0.0757	3.63	0.361	1356	-	0	1.464	1356	-
8	0.0865	3.40	0.404	1454	-	0	1.539	1454	-
9	0.0973	3.21	0.439	1563	-	0	1.600	1563	-
10	0.1082	3.04	0.408	1643	-	0	1.579	1643	-
11	0.1190	2.90	0.375	1731	-	0	1.481	1731	-
12	0.1298	2.78	0.317	1783	-	0	1.388	1783	-
13	0.1406	2.67	0.304	1835	-	0	1.399	1835	-
14	0.1514	2.57	0.226	1881	-	0	1.259	1881	-
15	0.1622	2.48	0.236	1760	-	0	1.274	1760	-
16	0.1731	2.40	0.193	1800	-	0	1.217	1800	-
17	0.1839	2.33	0.136	1955	-	0	1.146	1955	-
18	0.1947	2.27	0.121	1915	-	0	1.119	1915	-
19	0.2055	2.21	0.077	2003	-	0	1.081	2003	-
20	0.2163	2.15	0.098	1828	-	0	1.103	1828	-

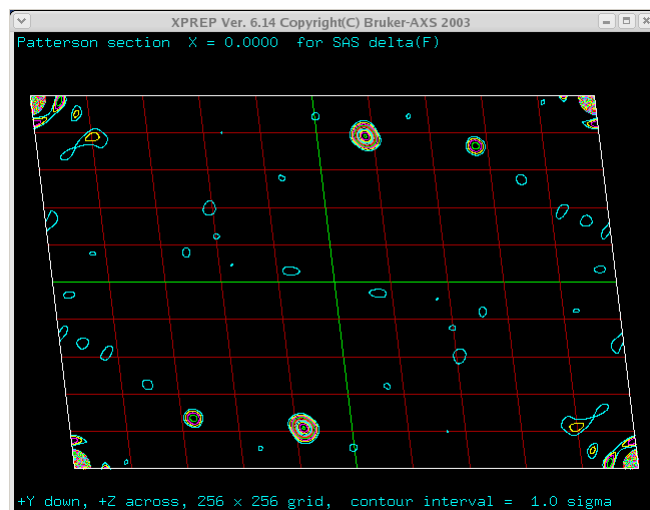


no anomalous signal



Strong anomalous signal

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= 0.0 AS FUNCTION OF RESOLUTION										
RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA		R-FACTOR observed	R-FACTOR expected	COMPARED I/SIGMA	R-meas	Rmrgd-F
	OBSERVED	UNIQUE	POSSIBLE							
9.77	2351	654	718	91.1%	3.8%	4.0%	2315	31.07	4.4%	2.4%
6.91	4854	1319	1352	97.6%	3.8%	4.1%	4849	30.13	4.4%	2.6%
5.64	6028	1673	1750	95.6%	4.3%	4.3%	6017	27.51	5.0%	3.1%
4.88	6924	1949	2038	95.6%	4.3%	4.3%	6900	27.19	5.1%	3.2%
4.37	7948	2216	2314	95.8%	4.0%	4.2%	7921	27.92	4.7%	2.9%
3.99	8783	2457	2604	94.4%	4.2%	4.3%	8760	26.59	5.0%	3.1%
3.69	9501	2642	2774	95.2%	4.4%	4.6%	9480	24.86	5.2%	3.6%
3.45	10579	2899	3034	95.6%	4.7%	4.8%	10555	23.39	5.6%	3.9%
3.26	11467	3092	3250	95.1%	5.4%	5.4%	11447	21.10	6.3%	4.6%
3.09	12284	3276	3426	95.6%	6.6%	6.7%	12271	17.85	7.7%	6.3%
2.94	12761	3350	3538	94.7%	8.0%	8.3%	12750	15.51	9.3%	7.8%
2.82	13405	3501	3706	94.5%	9.4%	10.1%	13399	13.58	10.9%	9.2%
2.71	13954	3626	3886	93.3%	11.7%	12.8%	13944	11.92	13.6%	12.4%
2.61	1491	3628	4084	88.8%	11.4%	11.4%	7389	7.84	15.7%	21.0%
2.52	7084	3602	4128	87.3%	11.6%	12.0%	6964	6.86	16.3%	23.7%
2.44	7454	3800	4382	86.7%	12.8%	13.8%	7308	6.09	18.1%	25.9%
2.37	7440	3803	4462	85.2%	15.9%	16.7%	7274	5.14	22.4%	33.3%
2.30	7359	3779	4526	83.5%	17.4%	18.6%	7160	4.61	24.6%	36.9%
2.24	7779	4032	4876	82.7%	21.2%	22.5%	7494	3.93	29.9%	44.3%
2.18	5153	2822	4738	59.6%	23.3%	24.9%	4662	3.47	32.9%	46.3%
total	170599	58120	65586	88.6%	5.5%	5.7%	168859	14.16	6.6%	10.3%



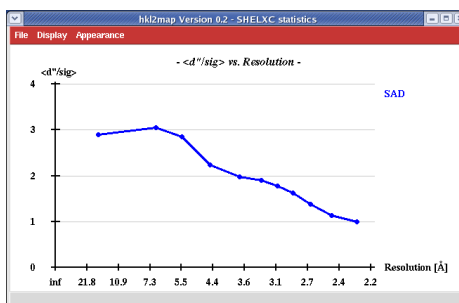
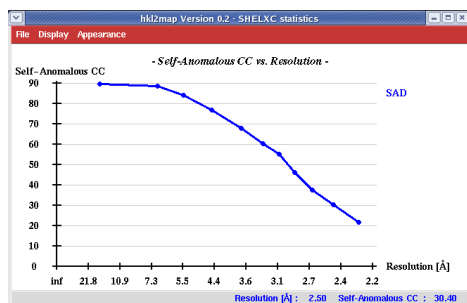
```
176322 Reflections read from SAD file HpFur_910.sca
31027 Unique reflections, highest resolution 2.185 Angstroms
189.6 Friedel pairs used on average for local scaling

Resl.   Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.2 - 3.0 - 2.8 - 2.6 - 2.4 - 2.18
N(data) 628 899 1150 2517 2568 2367 2155 2815 3742 5005 7181
Chi-sq   0.80 1.18 1.37 0.97 1.02 1.07 1.12 1.19 1.43 1.27 1.30
<I/sig>  44.2 40.9 38.4 38.6 34.4 29.5 23.8 19.2 12.4 8.0 5.0
%Complete 95.7 98.5 98.0 97.4 97.1 97.0 97.0 96.4 96.1 95.5 89.4
<d*/sig>  2.90 3.05 2.85 2.24 1.98 1.90 1.78 1.62 1.38 1.13 1.00
CC(snom) 89.8 88.6 84.4 77.0 67.9 60.4 55.2 46.1 37.5 30.4 21.8
```

For zero signal <d*/sig> and <d*/sig> should be about 0.80

21144 Reflections written to file HpFur_fa.hkl for input to SHELXD/E

31027 Reflections written to file HpFur.hkl for input to SHELXE

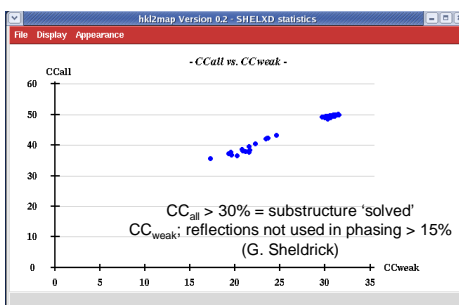
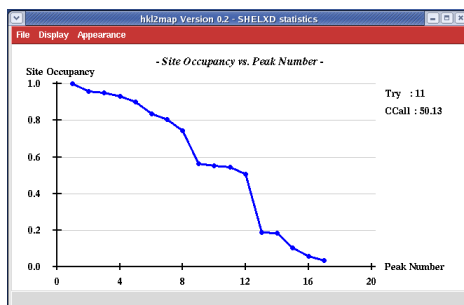




Finding Substructure – ShelxD output

European Synchrotron Radiation Facility

```
PSUM 157.14 PSMP Peaks: 96 25 24 23 16 16 15 15 14 13 13 12 12 11 11 11 11
Try 100.20 Peaks 99 91 88 86 84 83 77 61 57 55 52 44 13 12 7 7 3
R = 0.293, Min. fun. = 0.262, <cos> = 0.659, Ra = 0.226
Try 100, CC All/Weak 48.54 / 30.27, best 50.13 / 31.49, best PATFOM 45.24
PATFOM 43.39
=====
```



$$CC = \frac{100 [\sum (wE_o E_c) \sum w - \sum (wE_o) \sum (wE_c)]}{\{ [\sum (wE_o^2) \sum w - (\sum wE_o)^2] \cdot [\sum (wE_c^2) \sum w - (\sum wE_c)^2] \}^{1/2}}$$

Fujinaga & Read, *J. Appl. Cryst.* 20 (1987) 517-521.

The European Light Source
Slide: 33

Gordon Leonard, Getting the most from the ESRF MX beamlines, Grenoble, February 2010



Phasing quality – ShelxE output

European Synchrotron Radiation Facility

Current status of data preparation, substructure solution and phasing :

```
SHELXC/SHELXD/SHELXE original/SHELXE inverted
<vt> = 0.300, Contrast = 0.259, Connect = 0.817 for dens mod cycle 16
<vt> = 0.300, Contrast = 0.263, Connect = 0.818 for dens mod cycle 17
<vt> = 0.300, Contrast = 0.263, Connect = 0.821 for dens mod cycle 18
<vt> = 0.300, Contrast = 0.258, Connect = 0.820 for dens mod cycle 19
<vt> = 0.300, Contrast = 0.268, Connect = 0.823 for dens mod cycle 20

Mean weight and estimated mapCC as a function of resolution
d inf = 4.76 - 3.78 - 3.30 - 3.00 - 2.78 - 2.62 - 2.48 - 2.37 - 2.28 - 2.19
<vt> 0.258 0.336 0.493 0.516 0.524 0.538 0.572 0.565 0.557 0.543
<mapCC> 0.296 0.417 0.650 0.734 0.773 0.809 0.843 0.844 0.837 0.820
N 3107 3104 3119 3095 3189 3023 3223 3154 3043 2900

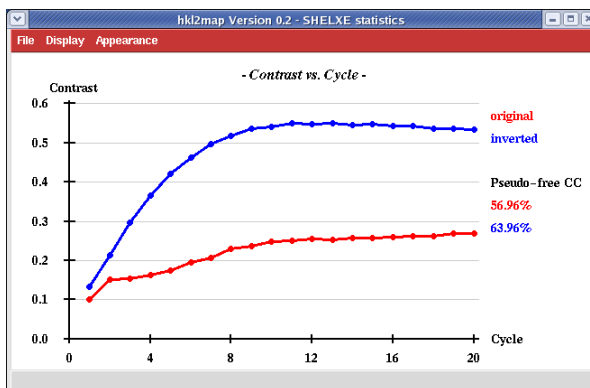
Pseudo-free CC = 56.96 %
```

Current status of data preparation, substructure solution and phasing :

```
SHELXC/SHELXD/SHELXE original/SHELXE inverted
<vt> = 0.300, Contrast = 0.548, Connect = 0.855 for dens mod cycle 15
<vt> = 0.300, Contrast = 0.545, Connect = 0.855 for dens mod cycle 16
<vt> = 0.300, Contrast = 0.545, Connect = 0.855 for dens mod cycle 17
<vt> = 0.300, Contrast = 0.537, Connect = 0.856 for dens mod cycle 18
<vt> = 0.300, Contrast = 0.537, Connect = 0.856 for dens mod cycle 19
<vt> = 0.300, Contrast = 0.533, Connect = 0.856 for dens mod cycle 20

Mean weight and estimated mapCC as a function of resolution
d inf = 4.76 - 3.78 - 3.30 - 3.00 - 2.78 - 2.62 - 2.48 - 2.37 - 2.28 - 2.19
<vt> 0.475 0.495 0.571 0.584 0.534 0.601 0.604 0.580 0.572 0.557
<mapCC> 0.628 0.681 0.771 0.807 0.845 0.864 0.872 0.864 0.860 0.840
N 3107 3104 3119 3095 3189 3023 3223 3154 3043 2900

Pseudo-free CC = 63.96 %
```



The European Light Source
Slide: 34

Gordon Leonard, Getting the most from the ESRF MX beamlines, Grenoble, February 2010

1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

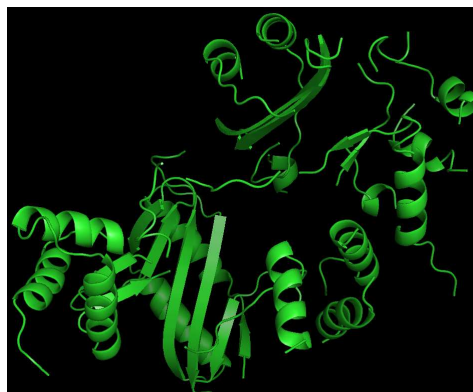
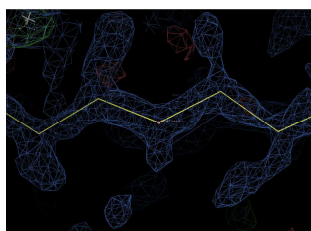
1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.

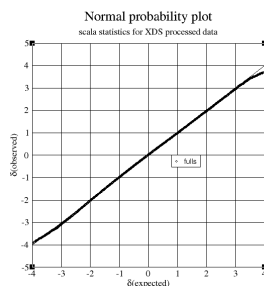
1000 reflections (1.00% of total) and 1000 reflections used, for reducing 40% noise.



Bovine Trypsin; S-SAD; S.G. = $P2_12_12_1$; 360° data; $\lambda = 1.77\text{\AA}$ (7.0 keV)
($\Delta F/F \sim 1.4\%$)

\$	N	1/d ²	Dmin(A)	R _{avg}	R _{full}	R _{cum}	R _{anom}	N _{anom}	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	N _{meas}	N _{ref}	N _{cent}	FRCBIAS	N _{bias}
1	0.0250	6.32	0.028	0.028	0.028	0.020	446	1582	76	20.7	64	74.2	8681	684	238	-	0	0
2	0.0501	4.47	0.028	0.028	0.028	0.015	923	1850	92	20.2	76	76.4	16192	1167	244	-	0	0
3	0.0751	3.65	0.029	0.029	0.029	0.014	1211	2288	116	19.7	96	72.5	19596	1446	234	-	0	0
4	0.1002	3.16	0.036	0.036	0.031	0.014	1442	1532	87	17.5	74	62.4	23354	1675	229	-	0	0
5	0.1252	2.83	0.048	0.048	0.033	0.018	1675	852	60	14.3	55	47.4	26943	1906	229	-	0	0
6	0.1503	2.58	0.069	0.069	0.037	0.023	1824	515	47	10.9	49	34.1	29266	2053	224	-	0	0
7	0.1753	2.39	0.090	0.090	0.041	0.029	1989	424	50	8.6	53	26.9	31619	2211	220	-	0	0
8	0.2003	2.23	0.127	0.127	0.046	0.039	2131	329	54	6.1	58	19.9	33541	2352	220	-	0	0
9	0.2254	2.11	0.200	0.200	0.053	0.057	2248	245	63	3.9	66	13.6	35181	2457	208	-	0	0
10	0.2504	2.00	0.378	0.378	0.062	0.109	2337	156	75	2.1	79	7.5	36106	2541	211	-	0	0

$$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} I_i(hkl)$$



Normal probability analysis, by run & partiality			
----- Run number: 1. fulls -----			
All data:	260479	Slope	0.992
Data within expected delta 0.90:	164591	Intercept	0.013

\$\delta\$	N	1/d ²	Dmin(A)	R _{avg}	R _{full}	R _{cum}	R _{anom}	N _{anom}	Av. I	SIGMA	I/sigma	sd	Mn(I/sd)	N _{meas}	N _{ref}	N _{cent}	FRCBIAS	N _{bias}
1	0.0250	6.32	0.028	0.028	0.028	0.020	446	1582	76	20.7	64	74.2	8681	684	238	-	0	0
2	0.0501	4.47	0.028	0.028	0.028	0.015	923	1850	92	20.2	76	76.4	16192	1167	244	-	0	0
3	0.0751	3.65	0.029	0.029	0.029	0.014	1211	2288	116	19.7	96	72.5	19596	1446	234	-	0	0
4	0.1002	3.16	0.036	0.036	0.031	0.014	1442	1532	87	17.5	74	62.4	23354	1675	229	-	0	0
5	0.1252	2.83	0.048	0.048	0.033	0.018	1675	852	60	14.3	55	47.4	26943	1906	229	-	0	0
6	0.1503	2.58	0.069	0.069	0.037	0.023	1824	515	47	10.9	49	34.1	29266	2053	224	-	0	0
7	0.1753	2.39	0.090	0.090	0.041	0.029	1989	424	50	8.6	53	26.9	31619	2211	220	-	0	0
8	0.2003	2.23	0.127	0.127	0.046	0.039	2131	329	54	6.1	58	19.9	33541	2352	220	-	0	0
9	0.2254	2.11	0.200	0.200	0.053	0.057	2248	245	63	3.9	66	13.6	35181	2457	208	-	0	0
10	0.2504	2.00	0.378	0.378	0.062	0.109	2337	156	75	2.1	79	7.5	36106	2541	211	-	0	0

$$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} I_i(hkl)$$

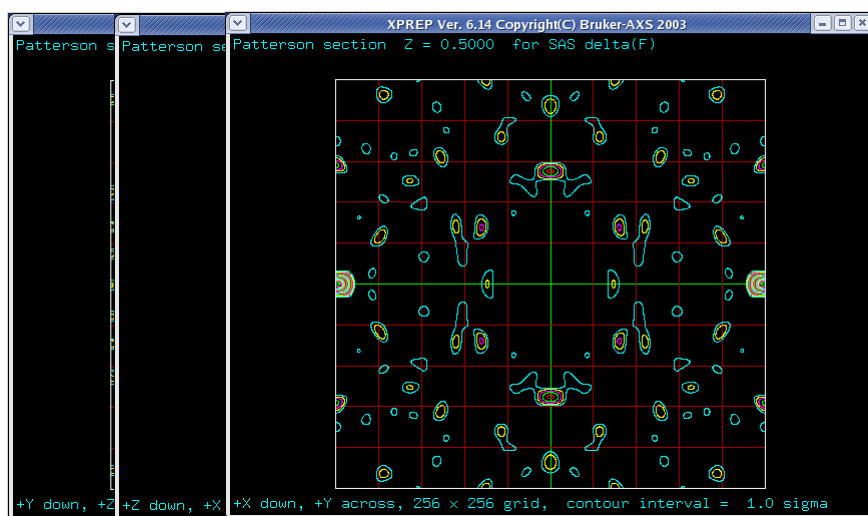
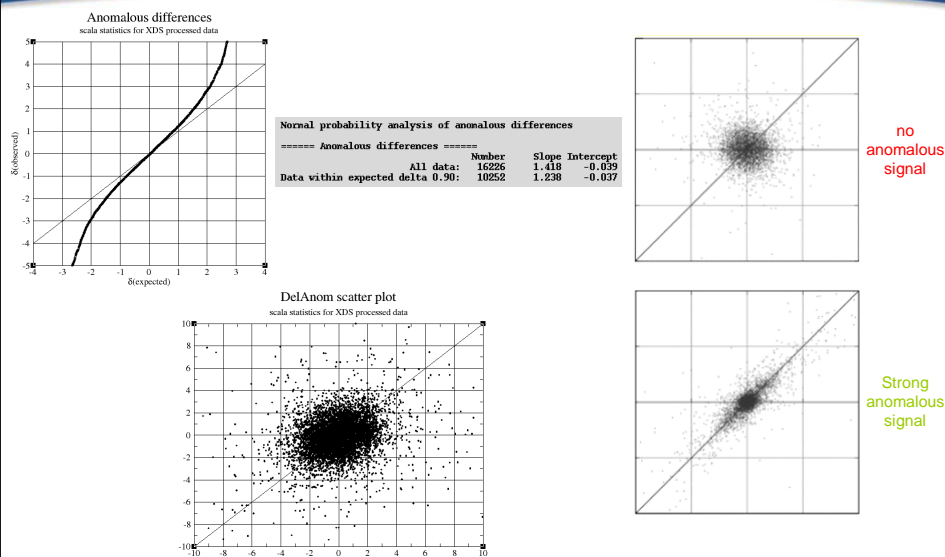
$$R_{anom} = \sum_{hkl} |I^+(hkl) - I^-(hkl)| / \sum_{hkl} |I^+(hkl) + I^-(hkl)| / 2$$

$$R_{p.i.m} = \sum_{hkl} [1/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} I_i(hkl)$$

$$R_{anom} < R_{merge}$$

R _{anom}	R _{p.i.m}	N	1/resol ²	d _{max}	CC _{anom}	N _{anom}	CC _{cen}	N _{cen}	R _{CR_{anom}}	N _{anom}	R _{CR_{cen}}	N _{cen}	CC _{I_{mean}}	N _{I_{mean}}
0.020	0.011	1	0.0250	6.32	0.653	442	-0.139	230	2.187	442	0.863	230	1.000	683
0.015	0.011	2	0.0501	4.47	0.566	912	0.014	232	1.898	912	1.009	232	1.000	1167
0.014	0.012	3	0.0751	3.65	0.333	1199	0.009	224	1.413	1199	1.010	224	0.999	1442
0.014	0.014	4	0.1002	3.16	0.168	1420	-0.028	211	1.185	1420	0.972	211	0.999	1671
0.010	0.019	5	0.1252	2.83	0.278	1653	0.065	217	1.331	1653	1.073	217	0.999	1903
0.023	0.027	6	0.1503	2.58	0.213	1809	-0.005	216	1.242	1809	0.995	216	0.999	2047
0.029	0.035	7	0.1753	2.39	0.172	1972	0.049	212	1.190	1972	1.044	212	0.998	2208
0.039	0.050	8	0.2003	2.23	0.107	2121	-0.046	214	1.115	2121	0.951	214	0.996	2351
0.057	0.078	9	0.2254	2.11	0.049	2232	-0.056	195	1.051	2232	0.952	195	0.992	2454
0.109	0.149	10	0.2504	2.00	0.064	2311	-0.043	209	1.066	2311	0.960	209	0.975	2534

$$R_{anom} / R_{p.i.m} < 1.2$$



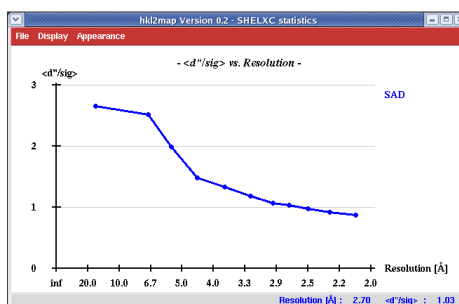
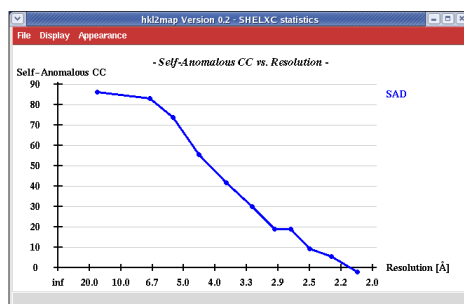
```

260535 Reflections read from SAD file rds.tryp.random.sca
18491 Unique reflections, highest resolution 1.999 Angstroms
224.3 Friedel pairs used on average for local scaling

Resl.   Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.00
N(data) 349  445  534 1203 1174 2077 1256 1699 2239 3131 4384
Chi-sq   0.75  0.75  0.65  0.83  1.07  0.98  1.01  0.98  1.03  1.08  1.16
<I/sig>  75.1  75.4  75.0  75.4  69.2  57.5  42.6  34.5  27.5  19.7  9.8
%Complete 96.4  98.9  99.1  98.8  98.2  98.0  98.0  96.7  96.5  95.7  93.5
<d"/sig>  2.66  2.52  1.99  1.48  1.33  1.18  1.07  1.03  0.98  0.92  0.87
CC(anom)  86.4  83.2  73.8  55.7  41.7  29.9  19.0  18.8  9.3  5.5  -2.0

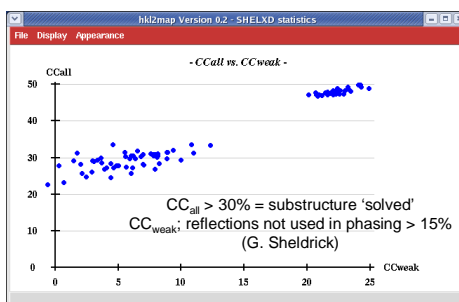
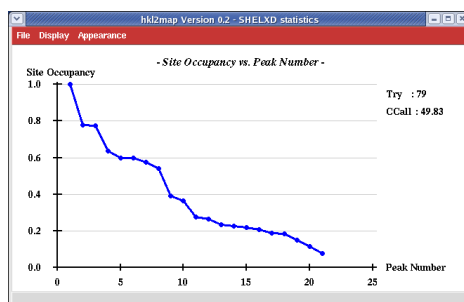
For zero signal <d"/sig> and <d"/sig> should be about 0.80

13759 Reflections written to file trypsin-random_fa.hkl for input to SHELXD/E
18491 Reflections written to file trypsin-random.hkl for input to SHELXE
    
```



```

PSUM 74.49 PSMF Peaks: 55 28 23 22 20 20 19 18 18 17 17 16 16 16 16 16
Try 100:20 Peaks 99 82 80 64 61 60 54 28 26 25 25 22 22 21 19 18 17 17 11
R = 0.300, Min.fun. = 0.411, <cos> = 0.415, Ra = 0.322
Try 100, CC All/Weak 47.56 / 21.91, best 49.83 / 24.93, best PATFOM 8.96
PATFOM 8.80
    
```



$$CC = \frac{100 [\sum (wE_o E_c) \sum w - \sum (wE_o) \sum (wE_c)]}{\{ [\sum (wE_o^2) \sum w - (\sum wE_o)^2] \cdot [\sum (wE_c^2) \sum w - (\sum wE_c)^2] \}^{1/2}}$$

Fujinaga & Read, *J. Appl. Cryst.* 20 (1987) 517-521.

Mean weight and estimated mapCC as a function of resolution

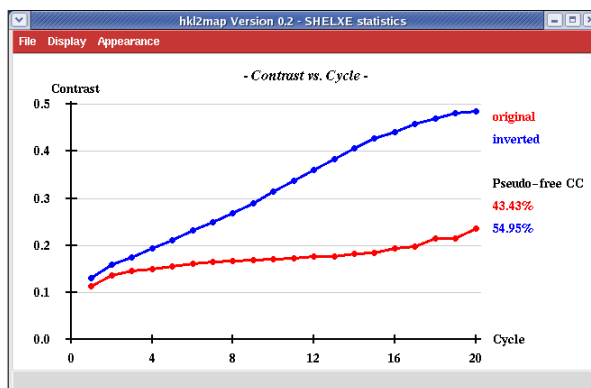
d	inf	4.46	3.51	3.05	2.76	2.55	2.40	2.27	2.17	2.08	2.00
<wt>	0.339	0.233	0.325	0.315	0.350	0.443	0.493	0.490	0.461	0.482	
<mapCC>	0.467	0.336	0.452	0.462	0.535	0.673	0.723	0.745	0.731	0.686	
N	1857	1848	1849	1879	1906	1787	1909	1833	1928	1695	

Pseudo-free CC = 43.43 %

Mean weight and estimated mapCC as a function of resolution

d	inf	4.46	3.51	3.05	2.76	2.55	2.40	2.27	2.17	2.08	2.00
<wt>	0.407	0.532	0.441	0.437	0.474	0.532	0.557	0.549	0.552	0.507	
<mapCC>	0.595	0.750	0.659	0.655	0.697	0.778	0.795	0.806	0.818	0.769	
N	1857	1848	1849	1879	1906	1787	1909	1833	1928	1695	

Pseudo-free CC = 54.95 %

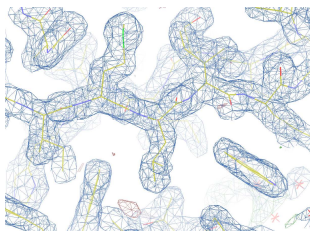


Starting model: R = 0.381 (Rfree = 0.310)

Cycle 1: After refine: R = 0.336 (Rfree = 0.288)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 2: After refine: R = 0.318 (Rfree = 0.288)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 3: After refine: R = 0.303 (Rfree = 0.284)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 4: After refine: R = 0.300 (Rfree = 0.283)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 5: After refine: R = 0.303 (Rfree = 0.283)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 6: After refine: R = 0.300 (Rfree = 0.282)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 7: After refine: R = 0.306 (Rfree = 0.291)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 8: After refine: R = 0.305 (Rfree = 0.288)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 9: After refine: R = 0.300 (Rfree = 0.288)
Found 19 (10 requested) and covered 11 (11 requested) atoms.
Cycle 10: After refine: R = 0.300 (Rfree = 0.292)
Found 19 (10 requested) and covered 11 (11 requested) atoms.

1718 reflections (94.70 % complete) and 0 constraint used, for refining 2000 atoms.
Observational/parameter ratio is 2.14

Building Cycle 1
Round 1: 184 peptides in 13 chains. The longest chain comprises 16 peptides.
Round 2: 211 peptides in 4 chains. The longest chain comprises 110 peptides.
Round 3: 211 peptides in 4 chains. The longest chain comprises 110 peptides.
Round 4: 211 peptides in 4 chains. The longest chain comprises 110 peptides.
Round 5: 211 peptides in 4 chains. The longest chain comprises 110 peptides.
Rounding results from round 4:
Chain 2, Residue 393 (out of 211), Connectivity Index 0.34

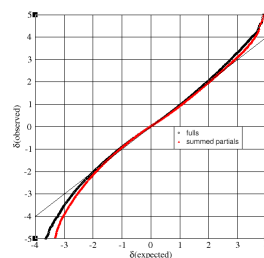


Gordon Leonard, Getting the most from the ESRF MX beamlines, Grenoble, February 2010

Gordon Leonard, Getting the most from the ESRF MX beamlines, Grenoble, February 2010

N	1/d*2	Dmin(A)	Ravg	Rfull	Rcum	Ranom	Rpin
1	0.0095	10.29	0.035	0.031	0.035	0.015	0.008
2	0.0189	7.27	0.037	0.034	0.036	0.013	0.008
3	0.0284	5.94	0.047	0.045	0.039	0.015	0.010
4	0.0378	5.14	0.058	0.057	0.045	0.015	0.013
5	0.0473	4.60	0.076	0.074	0.055	0.013	0.018
6	0.0567	4.20	0.081	0.078	0.063	0.014	0.018
7	0.0662	3.89	0.085	0.081	0.068	0.013	0.019
8	0.0756	3.64	0.083	0.076	0.071	0.013	0.018
9	0.0851	3.43	0.086	0.079	0.073	0.014	0.019
10	0.0945	3.25	0.090	0.081	0.075	0.014	0.020
11	0.1040	3.10	0.102	0.090	0.077	0.016	0.023
12	0.1134	2.97	0.116	0.101	0.079	0.017	0.025
13	0.1229	2.85	0.142	0.121	0.082	0.020	0.031
14	0.1323	2.75	0.152	0.132	0.085	0.021	0.034
15	0.1418	2.66	0.162	0.137	0.087	0.024	0.036
16	0.1512	2.57	0.183	0.153	0.090	0.026	0.040
17	0.1607	2.49	0.208	0.171	0.093	0.029	0.046
18	0.1701	2.42	0.250	0.203	0.096	0.037	0.055
19	0.1796	2.36	0.292	0.233	0.099	0.042	0.065
20	0.1890	2.30	0.352	0.282	0.103	0.049	0.078

Normal probability plot



Normal probability analysis, by run & partiality

```

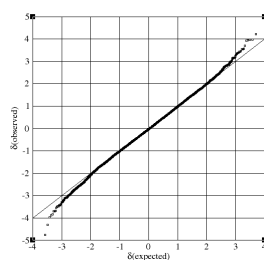
===== Run number: 1, fulls =====
Number      Slope Intercept
All data:   313243  0.980  0.027
Data within expected delta 0.90: 137931  0.903  0.025

===== Run number: 1, summed partials =====
Number      Slope Intercept
All data:   343001  0.996  0.006
Data within expected delta 0.90: 216735  0.888  0.014

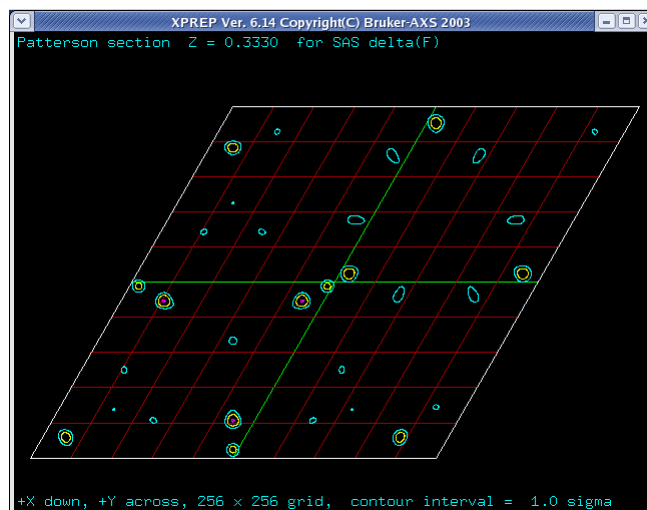
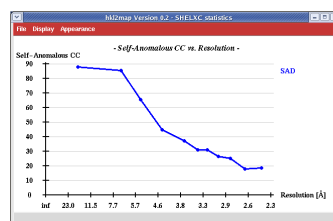
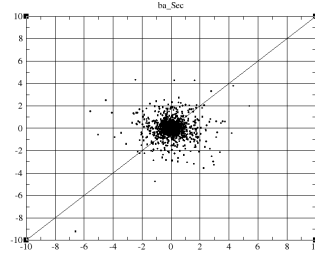
===== Run number: 1, fulls against fulls only =====
Number      Slope Intercept
All data:   313382  0.984  0.031
Data within expected delta 0.90: 197900  0.874  0.019

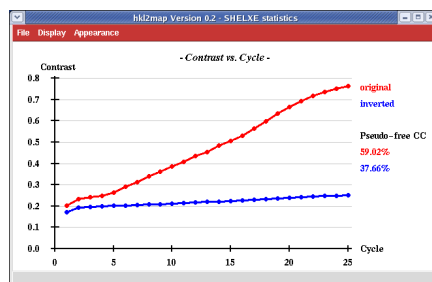
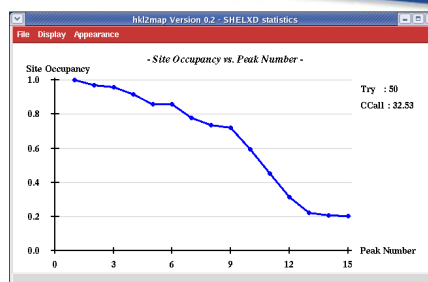
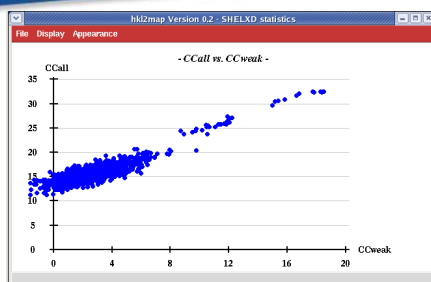
```

Anomalous differences



DelAnom scatter plot





15886 reflections (99.31 % complete) and 1259 restraints for refining 2109 atoms.
Observations/parameters ratio is 1.88

Cycle 41: After refine, R = 0.299 (Rfree = 0.000).
Found 6 (6 requested) and removed 50 (50 requested) atoms.

Cycle 42: After refine, R = 0.278 (Rfree = 0.000).
Found 6 (6 requested) and removed 34 (41 requested) atoms.

Cycle 43: After refine, R = 0.271 (Rfree = 0.000).
Found 6 (6 requested) and removed 10 (36 requested) atoms.

Cycle 44: After refine, R = 0.268 (Rfree = 0.000).
Found 6 (6 requested) and removed 8 (35 requested) atoms.

Cycle 45: After refine, R = 0.265 (Rfree = 0.000).
Found 6 (6 requested) and removed 9 (35 requested) atoms.

Building Cycle 9 Atomic shape factors 2.24 1.74
Round 1: 177 peptides in 6 chains. The longest chain comprises 44 peptides.
Round 2: 172 peptides in 7 chains. The longest chain comprises 45 peptides.
Taking the results from Round 1

Last building cycle:
Chain fragments will be rearranged
Chain 6: Residues 171. Estimated correctness of the model 95.1 %
5 chains (167 residues) have been docked in sequence



- **There are plenty of rapidly available indicators as to whether data are good enough for experimental phasing. The best one is to actually solve the structure!**
- **Once you have left the beamline, its too late.**
- **Process/analyse data as you go along and take your structure home with you.**
- **Repeat experiments if you have to - but only if you have to.**

Thanks for your attention!