ESRF | The European Synchrotron

(INDICATORS FOR A) SUCCESSFUL PHASING EXPERIMENT

GORDON LEONARD
ESRF STRUCTURAL BIOLOGY GROUP

PHASING EXPLOITING ANOMALOUS SCATTERING

1. Introduce a small number of anomalously scattering atoms into crystals
2. **Collect both F^+ and F^-** at one (or more) wavelength(s). If more than one wavelength make sure these are chosen using absorption edge scan.
3. Derive anomalous and (if more than one wavelength) dispersive differences.
4. Use anomalous and/or dispersive differences to locate anomalously scattering atoms
5. Use these positions to initiate phase calculations

COLLECTING BOTH F(+) AND F(-) (AFTER DAUTER, Z. *METHODS IN ENZYMOLOGY*, 276)

Crystal System	Laue Group	Total Oscillation required for	
		Standard data collection	**Anomalous data collection**
Triclinic	-1	180°	**360°**
Monoclinic	2/m	180° (b*), 90° (a*,c*)	**180° (a*,b*,c*)**
Orthorhombic	mmm	90°(a*, b*,c*)	**90°(a*,b*,c*)**
Tetragonal	4/m	90°(a*, b*,c*)	**90°(c*), 180°(a*, b*)**
	4/mmm	45°(c*), 90° (a*, b*)	**45°(c*), 90°(a*,b*)**
Trigonal	3/m	60°(c*), 90°(a*,b*)	**120°(c*), 180°(a*, b*)**
	-31m	30°(c*), 90°(a*,b*)	**60°(c*), 180°(a*, b*)**
	-3m1	30°(c*), 90°(a*,b*)	**30°(c*), 180°(a*, b*)**
Hexagonal	6/m	60°(c*), 90°(a*,b*)	**60°(c*), 180°(a*, b*)**
	6/mmm	30°(c*), 90°(a*,b*)	**60°(c*), 90°(a*, b*)**
Cubic	23	~60°	**~70°**
	432	~35°	**~45°**

(S)MAD: GOOD & BAD

1. No 'native' crystal – no problems with isomorphism when using heavy atom derivatives
2. Selenomethionine can be introduced into many protein sequences and can be used as a "magic bullet"
3. The 'method of choice' for *de novo* structure solution in macromolecular crystallography
4. Signals are small in comparison to those from isomorphous replacement – data have to be good!
5. In MAD have to collect a lot of data from one crystal – radiation damage is a problem

(S)MAD INTENSITY DIFFERENCES ARE MUCH SMALLER THAN FOR IR

$$\Delta' \text{Dispersive} \approx \frac{1}{\sqrt{2}} \frac{\sqrt{N_A}\,|\,f'_{\lambda_i} - f'_{\lambda_j}\,|_{\max}}{<|\,F_T\,|>}$$

$$\Delta' \text{Anomalous} \approx \frac{1}{\sqrt{2}} \frac{\sqrt{N_A}\,2\,f''_{\lambda_{i,\max}}}{<|\,F_T\,|>}$$

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

K absorption edges: f" ~ **4e⁻** ; can rise to **6-7e⁻** with 'white line'. Δf' usually has a maximum around **7e⁻** ; can rise to **~10e⁻** if far from edge.

L absorption edges: f" usually has a maximum around **12e⁻**; can rise to **~30e⁻** with 'white line'. Δf' usually has a maximum around **13e⁻**; can rise to **> 20e⁻** if far from edge

For comparison, Hg has Z = 78 e⁻.

In comparison to signals available in IR, AD signals are ***small***

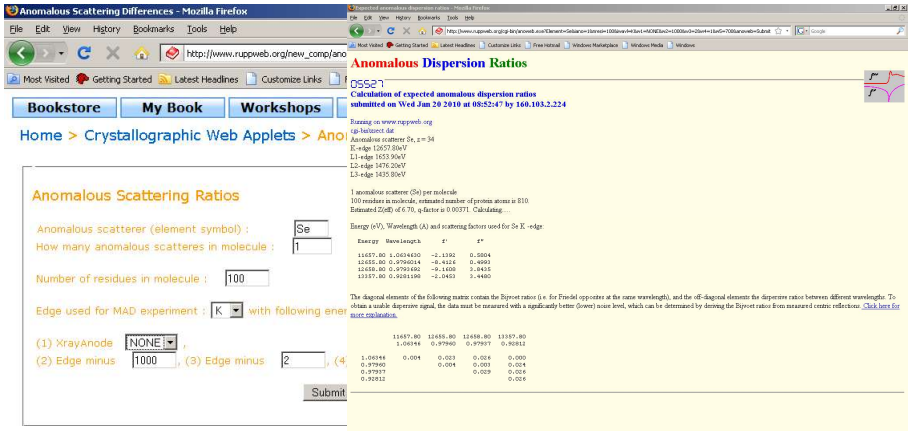
This knowledge must inform all decisions when collecting data for AD experiments

Need to: MAXIMISE SIGNAL and/or REDUCE ERRORS

EXPECTED ANOMALOUS SIGNAL

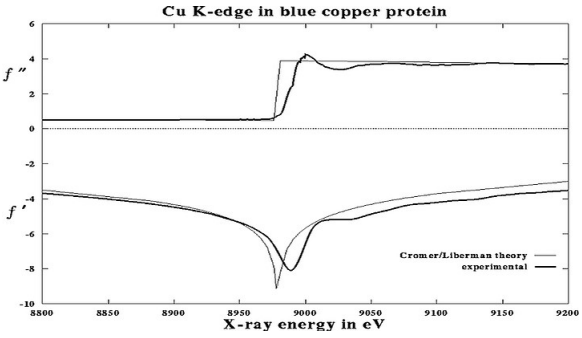
$$< \Delta I / I >_{anom} \approx \frac{1}{\sqrt{2}} \frac{\sqrt{N_A}\, 2 f''}{<|F_T|>}$$

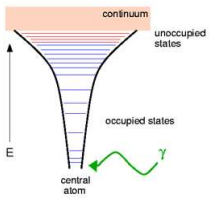
(J. Smith, *Curr. Opin. Struct. Biol.* **1**, 1991)

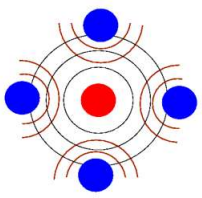


http://www.ruppweb.org/new_comp/anomalous_scattering.htm

MAXIMUM SIGNAL: EXPERIMENT DIFFERS FROM THEORY. DON'T RELY ON TABLES





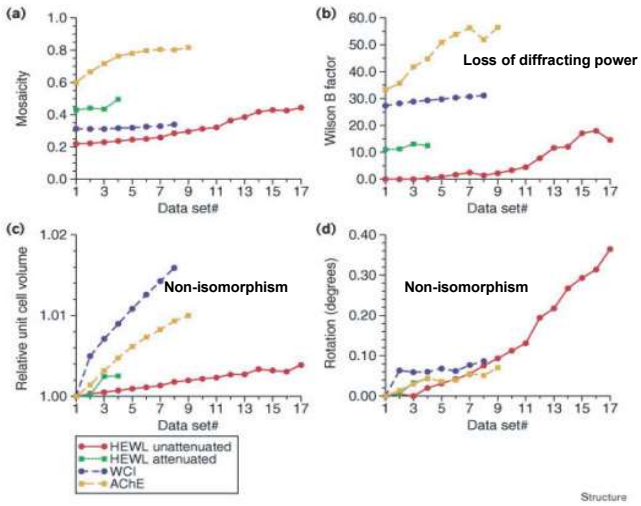


***JM Guss, EA Merritt, RP Phizackerley, B Hedman, M Murata, KO Hodgson, & HC Freeman (1989)**. "Phase determination by multiple-wavelength X-ray diffraction: crystal structure of a basic blue copper protein from cucumbers". Science 241, 806-811.*

We're not dealing with free atoms. Interaction of anomalous scatterer with other atoms causes deviation from ideality, as do changes in oxidation state. So, we MUST measure absorption edge experimentally if we want to be sure to obtain maximum 'signal'

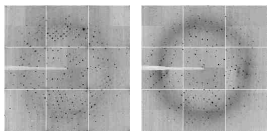
RADIATION DAMAGE IS (USUALLY) OUR ENEMY

Global radiation damage

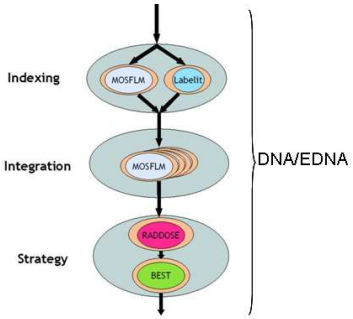


Ravelli & McSweeney (2000). The 'fingerprint' that X-rays can leave on structures. *Structure*, **8**, 315-328; https://doi.org/10.1016/S0969-2126(00)00109-X

DATA COLLECTION STRATEGIES AND DATA QUALITY PREDICTIONS



Flux = absorbed dose



Optimized plan for data collection
Predicted data quality statistics

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 N.of images	: 267
Overal Completeness	: 97.3%
Redundancy	: 2.00
R-factor (outer shell)	: 3.1% (27.1%)
I/Sigma (outer shell)	: 25.8 (2.0)
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

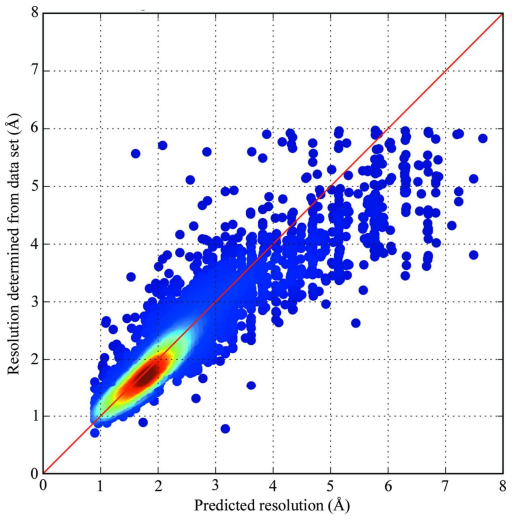
Resolution Lower	Upper	Compl. %	Average Intensity	Sigma	I/Sigma	R-fact %	Rsym %	Overload %
12.00	8.08	100.0	25535.8	505.1	50.6	1.6	1.4	0.00
8.08	6.50	100.0	14347.6	294.9	48.7	1.6	1.5	0.00
6.50	5.59	100.0	10253.5	221.1	46.4	1.7	1.5	0.00
5.59	4.98	98.2	12453.9	264.6	47.1	1.7	1.5	0.00
4.98	4.53	98.5	15372.8	319.1	48.2	1.7	1.5	0.00
4.53	4.18	97.0	15590.1	323.5	48.2	1.7	1.5	0.00
4.18	3.91	100.0	12589.8	271.5	46.4	1.8	1.5	0.00
3.91	3.68	100.0	10870.1	242.0	44.9	1.9	1.6	0.00
3.68	3.49	99.1	8174.5	203.5	40.2	2.1	1.8	0.00
3.49	3.32	98.8	6649.2	180.2	36.9	2.3	1.9	0.00
3.32	3.18	96.2	5016.9	161.1	31.1	2.8	2.3	0.00
3.18	3.05	97.4	3693.7	144.5	25.6	3.3	2.8	0.00
3.05	2.94	97.1	2833.8	137.3	20.6	4.1	3.5	0.00
2.94	2.84	96.5	2323.8	132.5	17.5	4.7	4.1	0.00
2.84	2.75	98.4	1769.9	131.9	13.4	6.2	5.3	0.00
2.75	2.67	97.5	1435.5	130.5	11.0	7.6	6.5	0.00
2.67	2.59	95.7	1277.4	131.4	9.7	8.5	7.3	0.00
2.59	2.52	96.8	1042.8	139.0	7.5	11.0	9.5	0.00
2.52	2.46	97.1	878.7	136.0	6.5	12.7	11.1	0.00
2.46	2.40	94.0	786.1	141.6	5.6	14.7	12.9	0.00
2.40	2.34	97.2	702.2	147.1	4.8	17.0	15.0	0.00
2.34	2.29	96.3	606.1	148.0	4.1	19.0	17.5	0.00
2.29	2.24	95.1	549.7	157.2	3.5	22.6	20.5	0.00
2.24	2.20	95.7	470.3	159.3	3.0	26.4	24.3	0.00
All data		97.3	4532.8	175.7	25.8	3.1	2.8	0.00

Bourenkov, G.P. & Popov, A.N. (2010). Optimization of data collection taking radiation damage into account. *Acta Cryst.* **D66**, 409-419.

$$R_{sym} = \Sigma_{hkl} \Sigma_i | I_{(hkl)i} - \langle I_{(hkl)} \rangle | / \Sigma_{hkl} \Sigma_i I_{(hkl)i}$$

$$R_{predicted} = \Sigma_{hkl} | \langle I_+ \rangle - \langle I_- \rangle | / \Sigma_{hkl} (\langle I_+ \rangle + \langle I_- \rangle)$$

Scatter plot of predicted resolutions for data collections against the resolutions determined by autoprocessing for all crystals processed so far on ESRF MASSIF-1



Svensson *et al. (2018). Acta Cryst D*74, 433–440.

STRONG ANOMALOUS SIGNAL

- *The Random Orientation 'Technique'* – Crystal is mounted in a random orientation with respect to the X-ray beam and goniometer axes. Data are then collected in one contiguous oscillation range such **that complete and high multiplicity anomalous data** are obtained.
- The rotation range required can be obtained using various data collection strategy programs (i.e. EDNA/BEST).
- Takes no real account of factors that can adversely affect intensities (absorption, different detector position for diffraction spots)

SMALLER ANOMALOUS SIGNAL

- *The Inverse-beam Method* – Crystal mounted in a random orientation. Data are collected in one contiguous oscillation range (from ϕ = X° to ϕ = Y°) such that data is as complete as possible for a 'standard' data collection. The Freidel mates of these reflections are then collected by inverting the orientation of the crystal and collecting the equivalent data (i.e. by collecting from ϕ=180+X° to ϕ=180+Y°). Can also collect in smaller wedges.

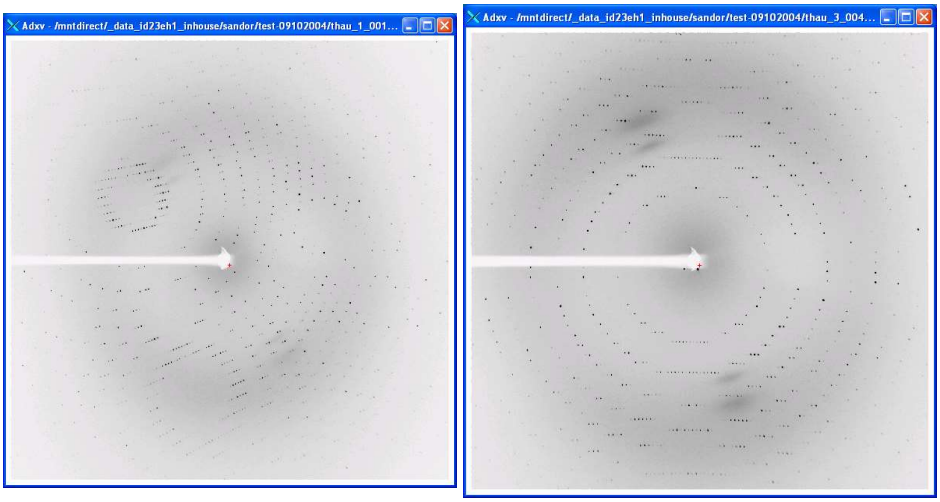
- **Freidel pairs are measured with the same path-length through crystal (absorption).**
- **Freidel pairs are collected close together in time (if we use small wedges).**
- **Freidel pairs measured with the same multiplicity**
- **For MAD experiments, best to interleave collection of wedges**
- Complicated to set up manually if using small wedges

SMALL ANOMALOUS SIGNAL

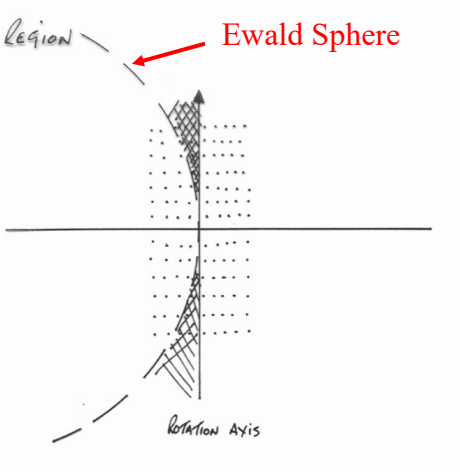
- *'Set-Crystal' Method* – The crystal is mounted such that the principle rotation axis of the crystal point group is parallel to the spindle axis of the goniometer. If this rotation axis is even-fold (i.e. 2, 4, 6) then:

- Bijvoet mates are collected on the same oscillation image.
- Bijvoet mates are collected after same total exposure time.
- Bijvoet mates suffer from 'same' absorption and radiation damage.
- 'Complicated' to align crystal in the desired manner

USING KAPPA GONIOMETERS



A 'PROBLEM' WITH SET CRYSTAL METHOD



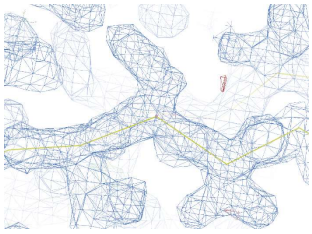
The 'Blind' Region:

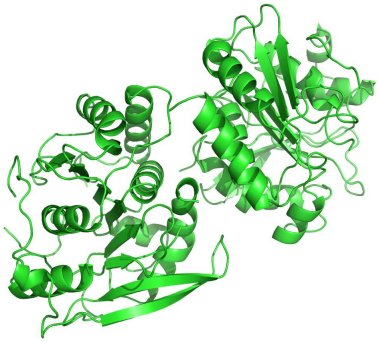
Shaded region will never pass through Ewald Sphere. Need to collect data in a second orientation to ensure data is properly complete.

Size of blind region: $\lambda^2 / 2d^2_{\min}$

HOW DO WE KNOW THAT OUR PHASING EXPERIMENT HAS WORKED?

Electron density map is the best metric





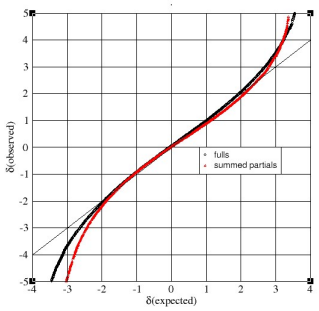
AFTER DATA PROCESSING – ARE THE DATA GOOD?

N	1/d^2	Dmin(A)	Rmerg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
$$																	
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

$$R_{merge} = \Sigma_{hkl} \Sigma_i | I_i(hkl) - < I(hkl) > | / \Sigma_i \Sigma_{hkl} I_i(hkl)$$

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

Normal probability plot




```
================================================
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:  94913     0.884     0.013

====== Run number:   1, fulls against fulls only ======
                              Number     Slope Intercept
                    All data: 158207     1.003     0.012
Data within expected delta 0.90:  99967     0.932     0.021

========================================
```

BUT, 'SUCCESS' INDICATORS FOR EXPERIMENTAL PHASING PROTOCOLS ARE AVAILABLE AT ALL STAGES OF THE EXPERIMENT

- **Before arriving at beamline**

If possible, check that heavy atom is bound
What is the expected anomalous signal?

- **Before data collection**

Experimentally measure absorption edge
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?

AFTER DATA PROCESSING – IS THERE ANY 'SIGNAL' IN THE DATA?

N	1/d^2	Dmin(Å)	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
$$																	
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

$$R_{merge} = \Sigma_{hkl}\Sigma_i \mid I_i(hkl) - < I(hkl) > \mid / \Sigma_i\Sigma_{hkl} I_i(hkl)$$

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

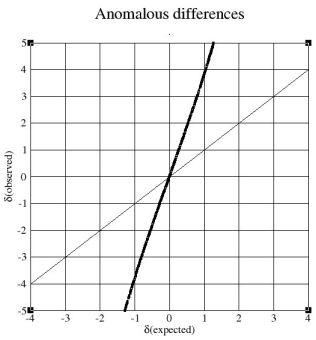
$$R_{anom} > R_{merge} ?$$

AFTER DATA PROCESSING – IS THERE ANY 'SIGNAL' IN THE DATA?

The presence of anomalous signal can be deduced from a normal probability plot of the normalized measured anomalous differences.

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 measured anomalous signal)

$$\delta_{obs} = (I^+ - I^-)/[\sigma^2(I^+) + \sigma^2(I^-)]^{1/2}$$

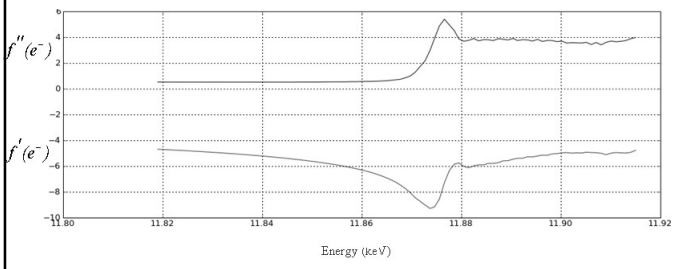



```
Normal probability analysis of anomalous differences

====== Anomalous differences ======
                                 Number     Slope Intercept
                       All data:   48988     4.061     0.041
Data within expected delta 0.90:   30954     3.724     0.043
```

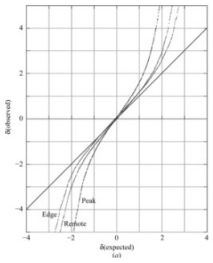
(Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86; Evans, P. (2006). *Acta Cryst.* **D62**, 72-82.)

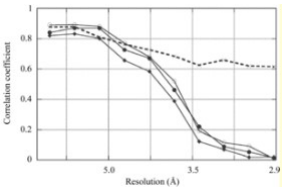
MAD DATA SETS



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

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





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

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

NORMAL PROBABILITY PLOTS AND ISOMORPHOUS REPLACEMENT

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

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

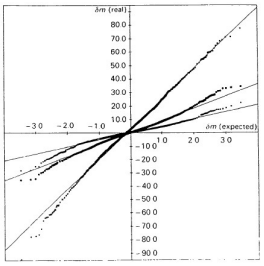


Fig. 4. Plot of Δm(real) *versus* Δm(expected) for native macromonycin data and three heavy-atom derivatives; × is $K_2Pt(NO_2)_4$; ■ is $HgCl_2$; ● is $(NH_4)_2IrCl_6$.

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

The residual is defined as $\sum|F_1 - F_2|/\sum F_1$. The slope and intercept of the least-squares straight line are defined as Δ(real) = slope × Δ(expected) + intercept. Numbers in parentheses are the slope and intercept of the least-squares straight line calculated from those data with |Δ(expected)| < 0.674 (50% of the data).

Resolution (Å)	Number of data	Residual	Slope	Intercept
(1) $K_2Pt(NO_2)_4$				
10–6	198	0.325	27.77 (27.06)	4.62 (5.06)
10–5	389	0.317	26.63 (26.32)	3.70 (4.04)
10–5*	111	0.265	23.57 (21.67)	3.61 (4.18)
10–4	795	0.267	25.55 (25.04)	3.30 (3.54)
10–3	1927	0.250	22.56 (22.14)	2.75 (2.90)
8–3	1870	0.246	22.35 (21.86)	2.66 (2.78)
8–3*	488	0.220	20.71 (20.21)	2.66 (2.84)
5–3	1538	0.231	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	390	0.151	11.47 (10.96)	0.73 (0.72)
10–5*	112	0.146	11.03 (10.64)	1.07 (0.99)
10–4	797	0.125	10.32 (9.77)	0.62 (0.57)
10–3	1929	0.117	9.06 (8.58)	0.39 (0.36)
8–3	1872	0.116	8.92 (8.49)	0.32 (0.30)
8–3*	489	0.112	8.73 (8.29)	0.22 (0.31)
5–3	1539	0.107	8.35 (8.04)	0.30 (0.27)
(3) $(NH_4)_2IrCl_6$				
10–6	199	0.141	9.00 (8.81)	0.43 (0.52)
10–5	390	0.127	8.14 (7.98)	0.09 (0.23)
10–5*	112	0.115	7.58 (6.90)	0.18 (0.39)
10–4	797	0.097	6.63 (5.99)	0.10 (0.25)
10–3	1929	0.092	5.26 (4.72)	0.08 (0.19)
8–3	1872	0.090	5.06 (4.61)	0.05 (0.14)
8–3*	489	0.087	4.87 (4.36)	0.12 (0.23)
5–3	1539	0.083	4.30 (4.15)	0.07 (0.08)

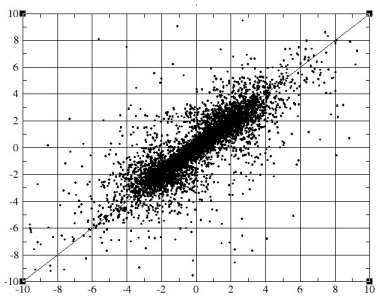
Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86

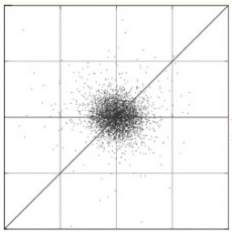
AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

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

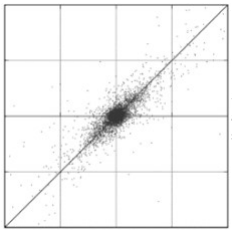
N	1/resol^2	dmax	CC_anom	N_anom	CC_cen	N_cen	RCR_anom	N_anom	RCR_cen	N_cen	CC_Imean	N_Imean
$$												
1	0.0250	6.32	0.930	1353	-0.108	187	5.245	1353	0.891	187	0.999	1558
2	0.0500	4.47	0.853	2526	0.024	215	3.597	2526	1.022	215	0.999	2869
3	0.0750	3.65	0.760	3355	-0.090	312	2.708	3355	0.915	312	0.998	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

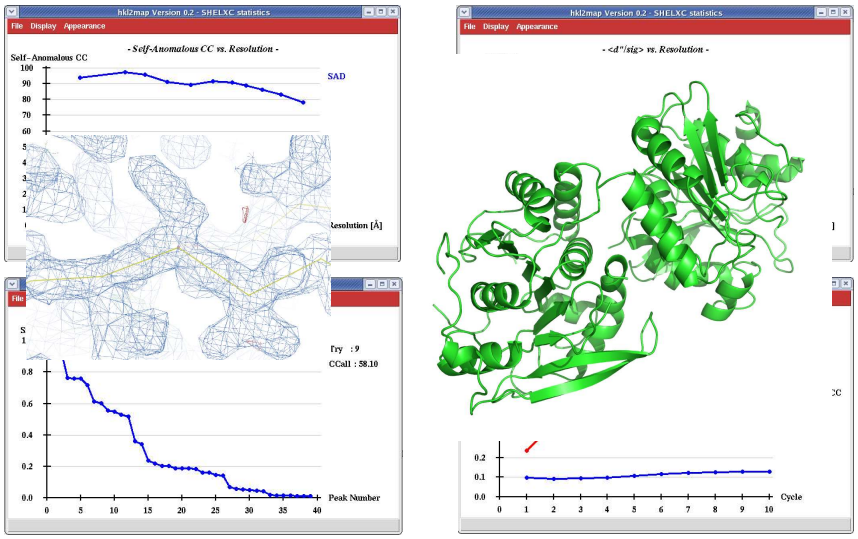
DATA PROCESSING – SUMMARY

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	97.3	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	-	-

FAE; $\Delta F/F$ ~5.4%; λ = 0.97 Å (S.G. = $P2_12_12_1$; 160° data)

AFTER DATA PROCESSING – SUBSTRUCTURE DETERMINATION, PHASE CALCULATIONS AND ELECTRON DENSITY MAPS



A slightly less clear-cut example….

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

N	1/d^2	Dmin(Å)	Rmerg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
$$																	
1	0.0108	9.62	0.048	0.045	0.048	0.048	337	29765	2789	10.7	2329	30.0	2481	329	0	-0.035	905
2	0.0216	6.80	0.044	0.024	0.046	0.054	666	19835	1600	12.4	1639	27.1	4927	666	0	-0.023	1643
3	0.0324	5.55	0.064	0.047	0.051	0.062	881	10446	1275	8.2	1027	22.2	6226	884	0	-0.009	2163
4	0.0433	4.81	0.060	0.068	0.054	0.060	1013	14543	1550	9.4	1366	23.7	6922	1009	0	-0.023	2664
5	0.0541	4.30	0.055	0.071	0.054	0.050	1162	20691	1979	10.5	1838	25.6	8416	1182	0	-0.033	3335
6	0.0649	3.93	0.066	0.069	0.057	0.052	1292	17160	1971	8.7	1612	24.4	9413	1289	0	-0.022	3777
7	0.0757	3.63	0.070	0.062	0.059	0.050	1386	14246	1712	8.3	1427	22.6	10143	1386	0	-0.020	4040
8	0.0865	3.40	0.072	0.083	0.061	0.050	1488	12524	1491	8.4	1324	20.9	10968	1485	0	-0.018	4412
9	0.0973	3.21	0.082	0.075	0.063	0.057	1600	8687	1206	7.2	1044	18.3	11887	1599	0	-0.005	4741
10	0.1082	3.04	0.095	0.069	0.065	0.066	1684	5916	925	6.4	840	15.3	12531	1682	0	0.001	5102
11	0.1190	2.90	0.112	0.165	0.067	0.073	1772	4206	764	5.5	719	13.1	13286	1774	0	0.003	5518
12	0.1298	2.78	0.129	0.133	0.070	0.075	1834	3612	727	5.0	700	11.8	13764	1836	0	0.011	5547
13	0.1406	2.67	0.161	0.169	0.073	0.089	1880	2831	725	3.9	661	9.8	14094	1879	0	0.014	5829
14	0.1514	2.57	0.170	0.204	0.075	0.103	1992	2337	614	3.8	610	7.9	11398	2002	0	0.014	4265
15	0.1622	2.48	0.130	0.515	0.075	0.114	1953	2189	413	5.3	548	6.4	7504	1976	0	-0.014	2268
16	0.1731	2.40	0.157	0.331	0.076	0.131	1984	1861	415	4.5	551	5.6	7662	2020	0	-0.022	2474
17	0.1839	2.33	0.201	0.552	0.077	0.157	2134	1489	427	3.5	559	4.7	8256	2158	0	-0.019	2567
18	0.1947	2.27	0.223	0.436	0.079	0.170	2115	1391	442	3.1	585	4.3	8142	2138	0	-0.029	2561
19	0.2055	2.21	0.278	0.278	0.080	0.208	2203	1209	485	2.5	618	3.7	8526	2234	0	-0.024	2592
20	0.2163	2.15	0.326	0.567	0.082	0.247	2191	1075	508	2.1	642	3.2	7916	2108	0	-0.018	2487
$$																	

$$R_{merge} = \Sigma_{hkl}\Sigma_i \mid I_i(hkl) - < I(hkl) > \mid / \Sigma_i \Sigma_{hkl} I_i(hkl)$$

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

$R_{anom} > R_{merge}$? Hmmmm……

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

N	1/d^2	Dmin(A)	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
$$																	
1	0.0108	9.62	0.048	0.045	0.048	0.048	337	29765	2789	10.7	2329	30.0	2481	329	0	-0.035	905
2	0.0216	6.80	0.044	0.024	0.046	0.054	666	19835	1600	12.4	1639	27.1	4927	666	0	-0.023	1643
3	0.0324	5.55	0.064	0.047	0.051	0.062	881	10446	1275	8.2	1027	22.2	6226	884	0	-0.009	2163
4	0.0433	4.81	0.060	0.068	0.054	0.060	1013	14543	1550	9.4	1366	23.7	6922	1009	0	-0.023	2664
5	0.0541	4.30	0.055	0.071	0.054	0.050	1162	20691	1979	10.5	1838	25.6	8416	1182	0	-0.033	3335
6	0.0649	3.93	0.066	0.069	0.057	0.053	1292	17160	1971	8.7	1612	24.4	9413	1289	0	-0.022	3777
7	0.0757	3.63	0.070	0.062	0.059	0.050	1386	14246	1712	8.3	1427	22.6	10143	1386	0	-0.020	4040
8	0.0865	3.40	0.072	0.083	0.061	0.050	1488	12524	1491	8.4	1324	20.9	10968	1485	0	-0.018	4412
9	0.0973	3.21	0.082	0.075	0.063	0.057	1600	8687	1206	7.2	1044	18.3	11887	1599	0	-0.005	4741
10	0.1082	3.04	0.095	0.069	0.065	0.066	1684	5916	925	6.4	840	15.3	12531	1682	0	0.001	5102
11	0.1190	2.90	0.112	0.165	0.067	0.073	1772	4206	764	5.5	719	13.1	13286	1774	0	0.003	5518
12	0.1298	2.78	0.129	0.133	0.070	0.075	1834	3612	727	5.0	700	11.8	13764	1836	0	0.011	5547
13	0.1406	2.67	0.161	0.169	0.073	0.089	1880	2811	725	3.9	661	9.8	14094	1879	0	0.014	5829
14	0.1514	2.57	0.170	0.204	0.075	0.103	1992	2337	614	3.8	610	7.9	11298	2002	0	0.014	4265
15	0.1622	2.48	0.130	0.515	0.075	0.114	1953	2189	413	5.3	548	6.4	7504	1976	0	-0.014	2268
16	0.1731	2.40	0.157	0.331	0.076	0.131	1984	1861	415	4.5	551	5.6	7662	2020	0	-0.022	2474
17	0.1839	2.33	0.201	0.552	0.077	0.157	2134	1489	427	3.5	559	4.7	8256	2158	0	-0.019	2567
18	0.1947	2.27	0.223	0.436	0.079	0.170	2115	1391	442	3.1	585	4.3	8142	2138	0	-0.029	2561
19	0.2055	2.21	0.278	0.278	0.080	0.208	2203	1209	485	2.5	618	3.7	8526	2234	0	-0.024	2592
20	0.2163	2.15	0.326	0.567	0.082	0.247	2191	1075	508	2.1	642	3.2	7916	2108	0	-0.018	2407
$$																	

N	1/resol^2	Dmin	Nmeas	Nref	Ncent	%poss	C%poss	Miplct	AnoCmp	AnoFrc	AnoMlt	Rmeas	RmeasO	(Rsym)	Rpim	RpimO	PCV	PCVO
$$	$$																	
1	0.011	9.62	2520	349	0	92.9	92.9	7.2	89.7	91.6	3.7	0.056	0.075	0.048	0.029	0.028	0.065	0.084
2	0.022	6.80	4944	671	0	97.1	95.6	7.4	96.4	99.0	3.7	0.051	0.075	0.044	0.026	0.028	0.059	0.088
3	0.032	5.55	6252	893	0	97.4	96.4	7.0	96.1	98.2	3.5	0.075	0.098	0.064	0.040	0.037	0.086	0.115
4	0.043	4.81	6961	1019	0	97.0	96.7	6.8	96.5	99.2	3.4	0.071	0.093	0.060	0.038	0.035	0.081	0.112
5	0.054	4.30	8471	1190	0	96.7	96.7	7.1	94.5	97.2	3.6	0.065	0.080	0.055	0.034	0.030	0.077	0.098
6	0.065	3.93	9448	1301	0	96.9	96.7	7.3	96.2	99.1	3.6	0.077	0.091	0.066	0.040	0.034	0.091	0.111
7	0.076	3.63	10172	1397	0	96.1	96.6	7.3	95.4	98.8	3.6	0.082	0.093	0.070	0.042	0.034	0.095	0.114
8	0.087	3.40	10994	1493	0	96.1	96.5	7.4	95.8	99.3	3.7	0.084	0.095	0.072	0.044	0.035	0.099	0.117
9	0.097	3.21	11917	1609	0	96.0	96.4	7.4	95.5	99.1	3.7	0.096	0.108	0.082	0.049	0.040	0.112	0.133
10	0.108	3.04	12363	1693	0	95.6	96.3	7.4	95.1	99.2	3.7	0.111	0.126	0.095	0.058	0.046	0.133	0.156
11	0.119	2.90	13318	1784	0	95.7	96.2	7.5	95.0	99.0	3.7	0.131	0.147	0.112	0.068	0.054	0.156	0.181
12	0.130	2.78	13795	1844	0	95.4	96.1	7.5	94.9	99.1	3.7	0.151	0.163	0.129	0.078	0.059	0.181	0.204
13	0.141	2.67	14130	1891	0	94.8	96.0	7.5	94.2	99.0	3.7	0.188	0.200	0.161	0.096	0.073	0.228	0.254
14	0.151	2.57	11414	2030	0	94.4	95.8	5.6	92.7	97.7	2.8	0.209	0.221	0.170	0.119	0.090	0.227	0.262
15	0.162	2.48	7749	2044	0	94.1	95.6	3.8	89.9	94.9	1.9	0.183	0.202	0.130	0.130	0.102	0.191	0.234
16	0.173	2.40	7901	2086	0	93.4	95.4	3.8	88.8	94.5	1.9	0.221	0.239	0.157	0.157	0.121	0.232	0.281
17	0.184	2.33	8484	2222	0	93.8	95.3	3.8	90.0	95.4	2.0	0.285	0.301	0.201	0.201	0.152	0.301	0.357
18	0.195	2.27	8382	2196	0	93.5	95.2	3.8	90.1	96.0	1.9	0.316	0.328	0.223	0.223	0.167	0.336	0.395
19	0.206	2.21	8770	2304	0	92.9	95.0	3.8	88.9	95.1	2.0	0.392	0.406	0.278	0.278	0.206	0.419	0.491
20	0.216	2.15	8447	2298	0	90.7	94.7	3.7	86.5	94.8	1.9	0.461	0.473	0.326	0.326	0.243	0.501	0.573
$$																		

$$R_{merge} = \Sigma_{hkl} \Sigma_i | I_i(hkl) - < I(hkl) > | / \Sigma_{hkl} \Sigma_i I_i(hkl)$$

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

$$R_{p.i.m} = \Sigma_{hkl} [1/(N-1)^{1/2}] \Sigma_i | I_i(hkl) - < I(hkl) > | / \Sigma_i \Sigma_{hkl} I_i(hkl)$$

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

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

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

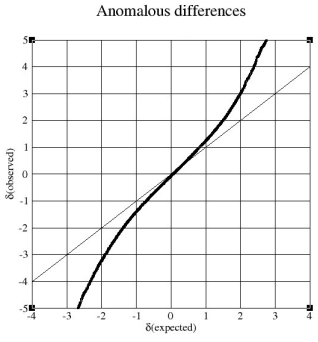
AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

Normal probability plot - anomalous differences

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

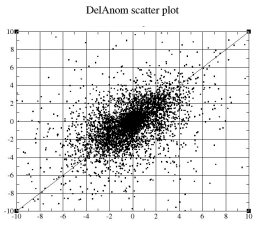


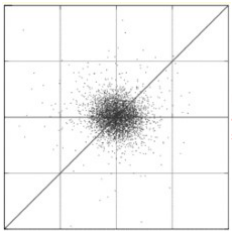
(Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86; Evans, P. (2006). *Acta Cryst.* **D62**, 72-82.)

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

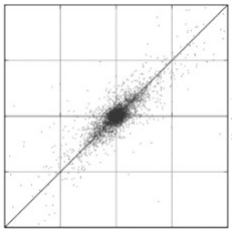
What is the correlation coefficient between anomalous differences in two half data sets chosen at random? Where is the point that CC_{anom} drops below 30%?

N	1/resol^2	dmax	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	2.365	651	-
3	0.0324	5.55	0.511	855	-	0	1.758	855	-
4	0.0433	4.81	0.578	974	-	0	1.936	974	-
5	0.0541	4.30	0.484	1099	-	0	1.701	1099	-
6	0.0649	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	0.428	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	0.324	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	1.129	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	-





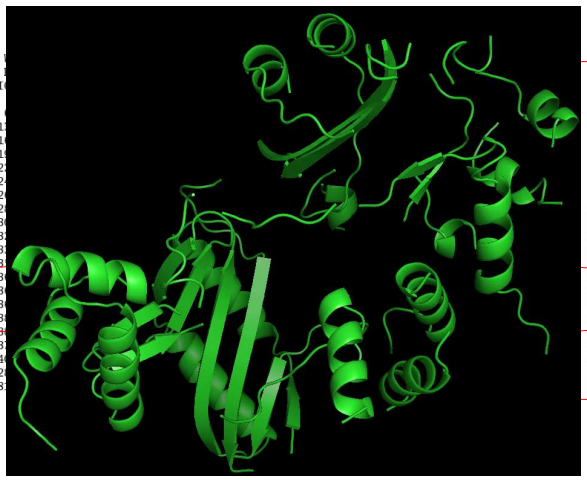
no anomalous signal



Strong anomalous signal

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA (XDS)?

SUBSET OF INTENSITY DATA					
RESOLUTION LIMIT	NUMBER OF OBSERVED	UNI		SigAno	Nano
9.77	2351			3.023	302
6.91	4854	1		3.162	646
5.64	6028	1		3.019	808
4.88	6924	1		2.917	933
4.37	7948	2		2.307	1062
3.99	8783	2		2.200	1183
3.69	9501	2		2.086	1274
3.45	10579	2		1.935	1408
3.26	11467	3		1.868	1512
3.09	12284	3		1.771	1612
2.94	12761	3		1.597	1661
2.82	13405	3		1.445	1745
2.71	13954	3		1.333	1809
2.61	7491	3		1.134	1669
2.52	7084	3		1.023	1643
2.44	7454	3		1.003	1720
2.37	7440	3		0.933	1710
2.30	7359	3		0.921	1677
2.24	7779	4		0.851	1725
2.18	5153	2		0.840	1057
total	170599	58		1.569	27156

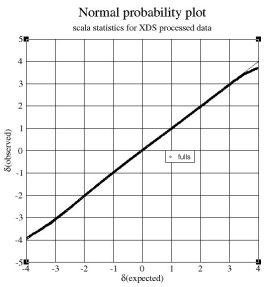


Bovine Trypsin; S-SAD; S.G. = $P2_12_12_1$; 360° data; λ = 1.77Å (7.0 keV)
(ΔF/F ~1.4%)

AFTER DATA PROCESSING – ARE THE DATA GOOD?

N	1/d^2	Dmin(A)	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
$$																	
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
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
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
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
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
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
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
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
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
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

$$R_{merge} = \Sigma_{hkl}\Sigma_i | I_i(hkl) - < I(hkl) > | / \Sigma_i \Sigma_{hkl} I_i(hkl)$$




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

====== Run number:   1, fulls ======
                                Number     Slope Intercept
                      All data:  260479     0.992     0.013
Data within expected delta 0.90:  164591     0.981     0.023
==========================================
```

AFTER DATA PROCESSING – ANY SIGNAL?

N	1/d^2	Dmin(A)	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
$$																	
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
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
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
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
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
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
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
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
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
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

$$R_{merge} = \Sigma_{hkl}\Sigma_i | I_i(hkl) - < I(hkl) > | / \Sigma_i \Sigma_{hkl} I_i(hkl)$$

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

$$R_{p.i.m} = \Sigma_{hkl} [1/(N-1)^{1/2} \Sigma_i | I_i(hkl) - < I(hkl) > | / \Sigma_i \Sigma_{hkl} I_i(hkl)$$

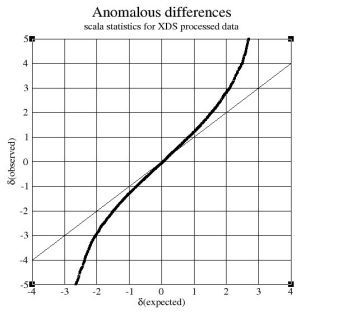
$R_{anom} < R_{merge}$

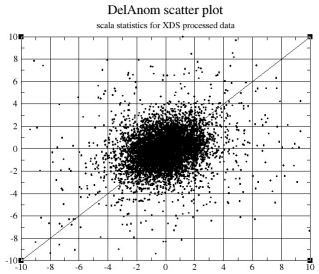
Ranom	Rpim
0.020	0.011
0.015	0.011
0.014	0.012
0.014	0.014
0.018	0.019
0.023	0.027
0.029	0.035
0.039	0.050
0.057	0.078
0.109	0.149

N	1/resol^2	dmax	CC_anom	N_anom	CC_cen	N_cen	RCR_anom	N_anom	RCR_cen	N_cen	CC_Imean	N_Imean
$$												
1	0.0250	6.32	0.653	442	-0.139	230	2.187	442	0.863	230	1.000	683
2	0.0501	4.47	0.566	912	0.014	232	1.898	912	1.009	232	1.000	1167
3	0.0751	3.65	0.333	1199	0.009	224	1.413	1199	1.010	224	0.999	1442
4	0.1002	3.16	0.168	1420	-0.028	211	1.185	1420	0.972	211	0.999	1671
5	0.1252	2.83	0.278	1653	0.065	217	1.331	1653	1.073	217	0.999	1903
6	0.1503	2.58	0.213	1809	-0.005	216	1.242	1809	0.995	216	0.999	2047
7	0.1753	2.39	0.172	1972	0.049	212	1.190	1972	1.044	212	0.998	2208
8	0.2003	2.23	0.107	2121	-0.046	214	1.115	2121	0.951	214	0.996	2351
9	0.2254	2.11	0.049	2232	-0.056	195	1.051	2232	0.952	195	0.992	2454
10	0.2504	2.00	0.064	2311	-0.043	209	1.066	2311	0.960	209	0.975	2534

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

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?



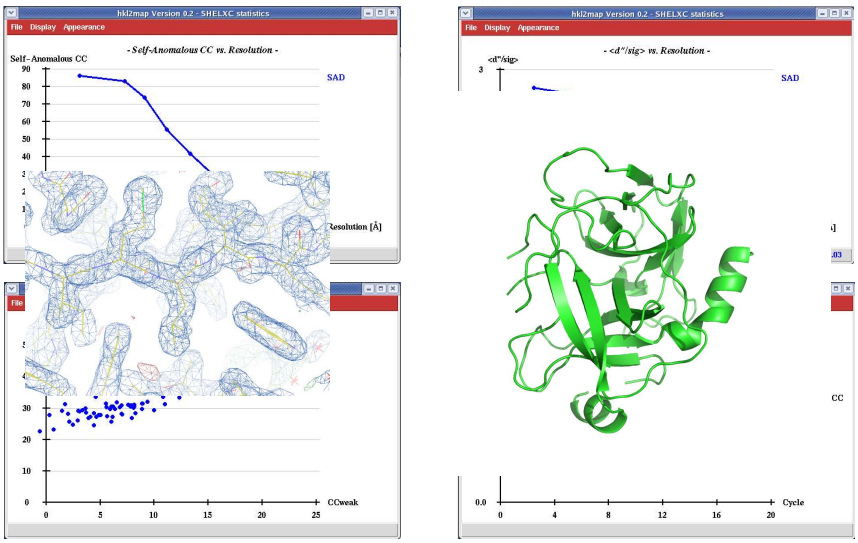


Normal probability analysis of anomalous differences

------ Anomalous differences ------

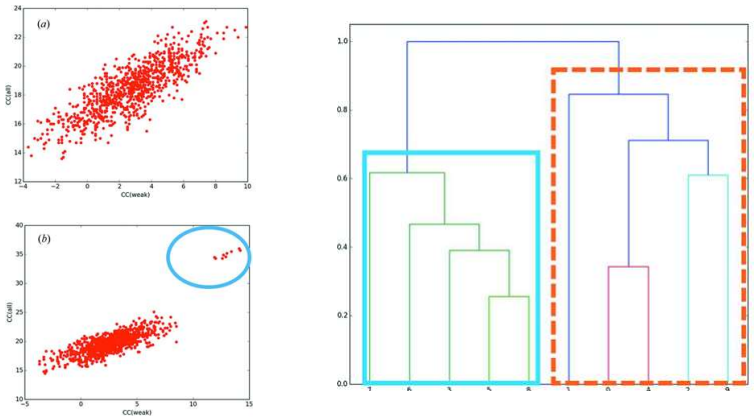
	Number	Slope	Intercept
All data:	16226	1.418	-0.039
Data within expected delta 0.90:	10252	1.238	-0.037

AFTER DATA PROCESSING – SHELXC/SHELXD/SHELXE OUTPUT



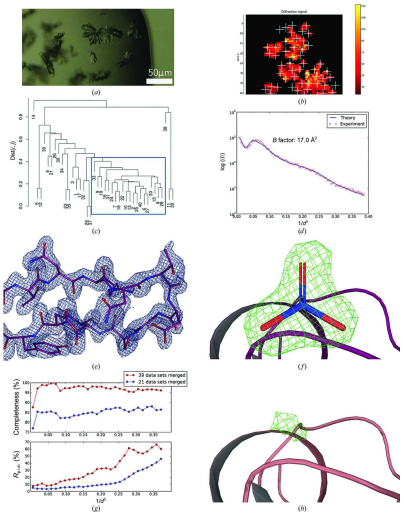
MULTI-CRYSTAL DATA COLLECTION

- Generate high multiplicity data sets by merging data collected from several crystals
- Increase effective absorbed dose that can be used during an experiment
- But there are pitfalls…..



JOURNAL OF APPLIED CRYSTALLOGRAPHY
Santoni *et al.*
Volume 50 | Part 6 | December 2017 | Pages 1844–1851 | 10.1107/S1600576717015229

ANOTHER EXAMPLE OF 'TOO MUCH DATA'



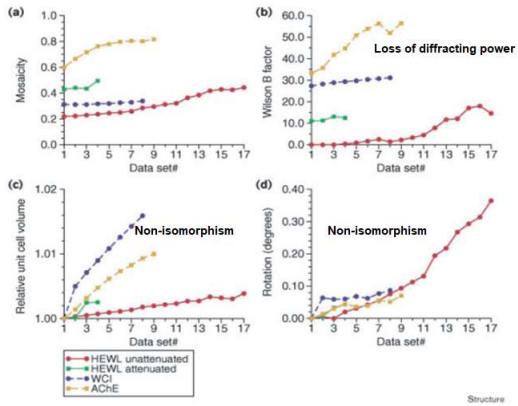
Acta Cryst. D STRUCTURAL BIOLOGY

Zander *et al.*

Volume 71 | Part 11 | November 2015 | Pages 2328–2343 | 10.1107/S1399004715017927

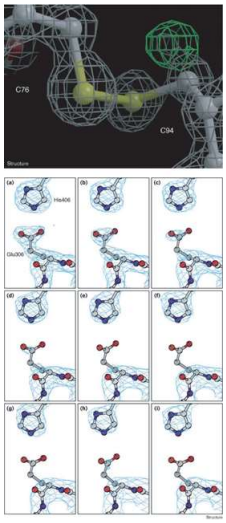
RADIATION DAMAGE INDUCED PHASING

Global radiation damage



Ravelli & McSweeney (2000). The 'fingerprint' that X-rays can leave on structures. *Structure*, **8**, 315-328; https://doi.org/10.1016/S0969-2126(00)00109-X

Specific radiation damage



Ravelli at al., (2003). Specific Radiation Damage Can Be Used to Solve Macromolecular Crystal Structures. Structure, 11, 217-224; https://doi.org/10.1016/S0969-2126(03)00006-6

Thanks for your attention