

Optimising MAD & SAD Experiments in MX

Gordon Leonard
Structural Biology Group
European Synchrotron Radiation Facility,
Grenoble,
France.



Page 1 | 18th September 2014 | Gordon Leonard

The European Synchrotron | 

CONTENTS

1. Anomalous Scattering

1. What is it?
2. How can we use it?

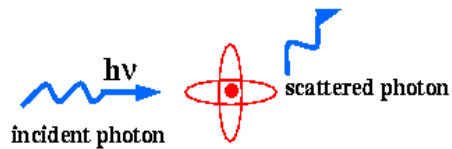
2. 'Tips' for data collection for SAD/MAD

3. Indicators for successful structure solution

The European Synchrotron | 

NORMAL SCATTERING

X-ray photon interacts with the electrons in the atoms of the sample. Atomic scattering factor f proportional to atomic number, Z , of atoms



$$f = f_o$$

$$F_{hkl} = \sum_j f_j \exp(-2\pi i(hx_j + ky_j + lz_j))$$

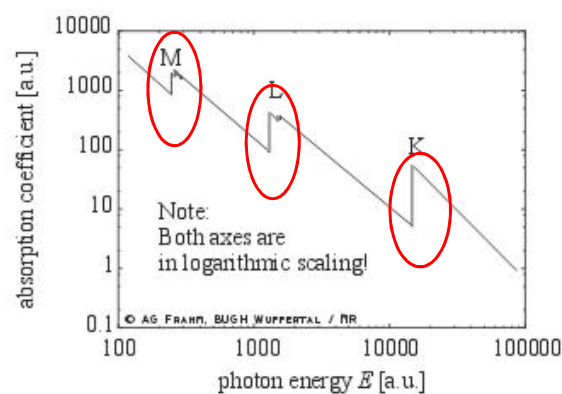
$$F_{hkl} = \sum_j |F_{hkl}| \exp(-2\pi i(\phi))$$

All atoms scatter X-rays in the same way. Friedel's Law is obeyed.

$$|F_{hkl}| = |F_{\bar{h}\bar{k}\bar{l}}|$$

The European Synchrotron | ESRF

ELECTRON SHELLS IN ATOMS HAVE X-RAY ABSORPTION EDGES

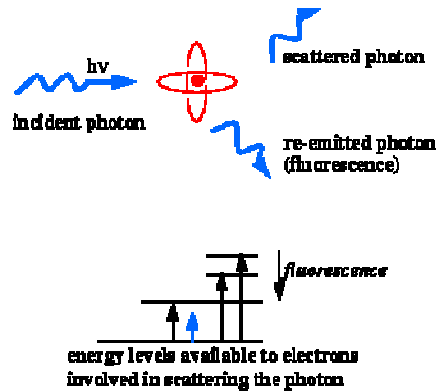


At incident X-ray energies close to absorption edges scattering from atoms is no longer normal

The European Synchrotron | ESRF

NORMAL SCATTERING IS MODIFIED CLOSE TO ABSORPTION EDGE

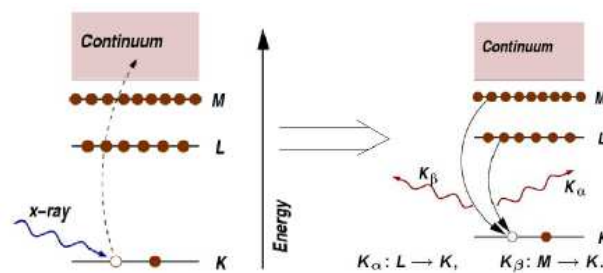
- At photon energies close to absorption edges of a given atom, some photons are absorbed by the atom and electrons are ejected from its inner shell.
- When this happens, electrons from higher energy shells decay to lower energy shells emitting fluorescence as they do so.



http://www.bmsc.washington.edu/scatter/AS_tutorial.html

The European Synchrotron | ESRF

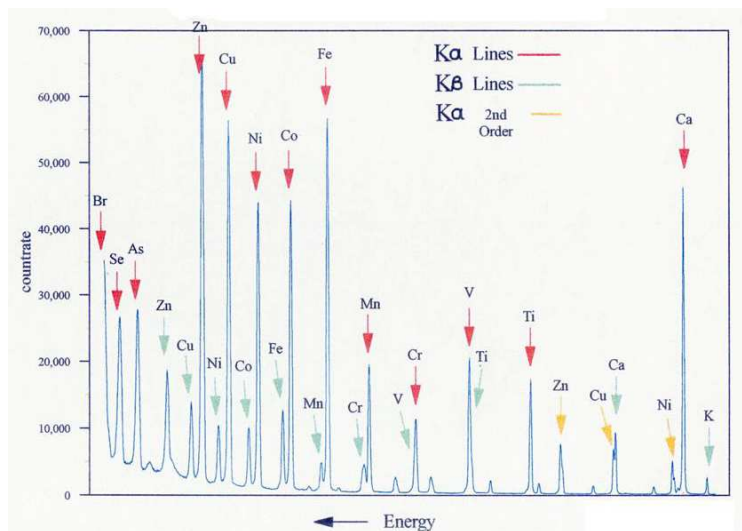
A BIT MORE ABOUT FLUORESCENCE



Incident X-rays eject electrons from the inner shell of atoms (left), higher shell electrons decay down, then photons are emitted via fluorescence. **Fluorescence emission for different atoms has a characteristic energy**

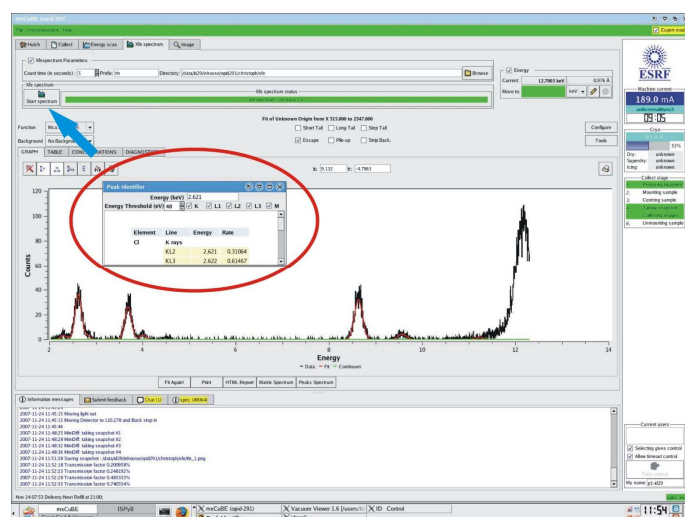
The European Synchrotron | ESRF

ENERGIES OF SOME FLUORESCENCE EMISSION LINES



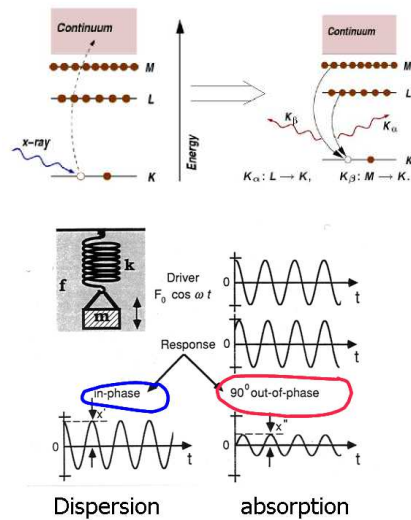
The European Synchrotron ESRF

GUI FOR ON-BEAMLINE X-RAY FLUORESCENCE EMISSION ANALYSIS



The European Synchrotron ESRF

Scattering factors modified when incident X-ray energy is close to that of absorption edge



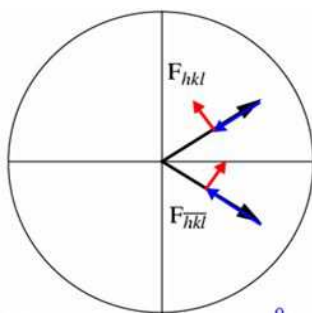
Treat electron shells as *harmonic oscillators* with resonance frequencies close to energy of shell absorption edge. Now we get:

$$f = f_o + f'(\lambda) + if''(\lambda)$$

Scattering factor modified in such a way that it has additional, wavelength(energy)-dependent dispersive (f') (real) and absorptive (f'') (imaginary) terms

The European Synchrotron | ESRF

WHAT HAPPENS TO STRUCTURE FACTORS IN THE PRESENCE OF ANOMALOUS SCATTERING?



$$f = f_o + f'(\lambda) + if''(\lambda)$$

$$|F_{hkl}| \neq |F_{hkl}^*|$$

Friedel's law is no longer true.

There is an intensity difference, ΔF_{ano}

$$\Delta F_{ano} = |F_{hkl}| - |F_{hkl}^*| = |F^+| - |F^-|$$

$$|F_{mean}| = (|F_{hkl}| + |F_{hkl}^*|) / 2$$

We can use this intensity difference to solve the phase problem in MX.

Need to measure both $|F_{hkl}|$ ($|F^+|$) and $|F_{hkl}^*|$ ($|F^-|$)

The European Synchrotron | ESRF

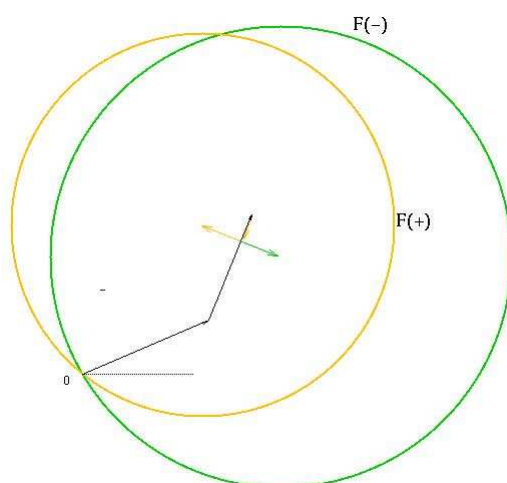
Total rotation ranges (MAD, SAD)

(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°

The European Synchrotron ESRF

PHASING USING ANOMALOUS SCATTERING

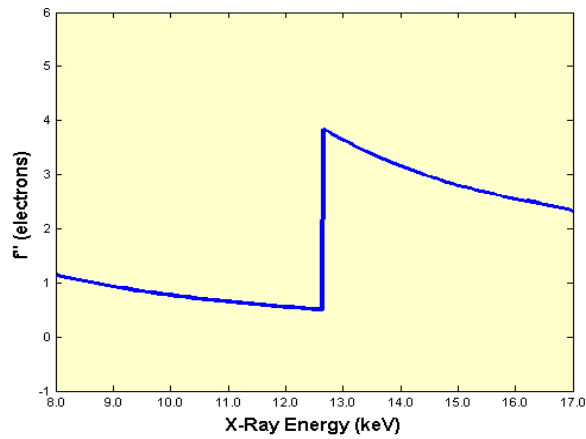


SAD results (like SIR) in phase ambiguity. But this can usually be straightforwardly resolved.

The larger f'' , the larger ΔF_{ano} and the easier it is solve the phase problem. So ideally want to collect data where f'' is at a maximum.

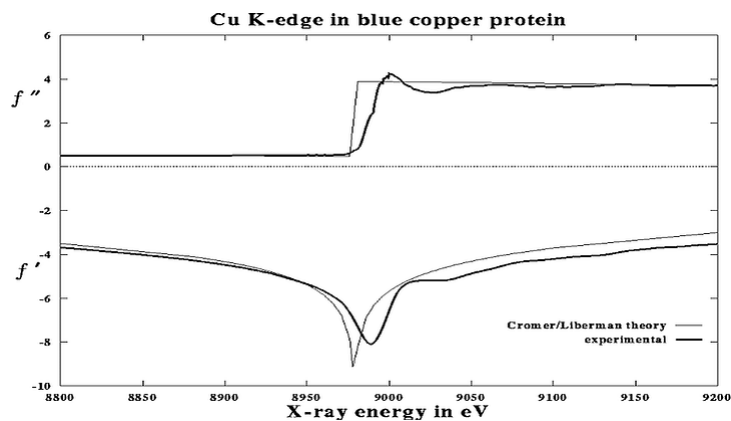
The European Synchrotron ESRF

CALCULATION SHOWS THAT F'' IS AT A MAXIMUM JUST ABOVE AN ABSORPTION EDGE



The European Synchrotron | ESRF

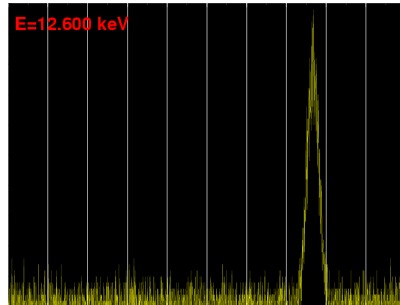
EXPERIMENT DIFFERS FROM THEORY. DON'T RELY ON TABLES!



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.

The European Synchrotron | ESRF

MEASURING AN ABSORPTION EDGE

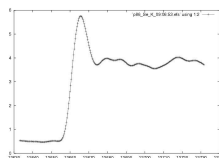


'Incident X-ray peak' ~static (i.e. energy change not big). We don't really care about this.

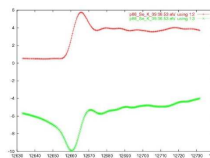
Measure intensity contained in 'region of interest' around expected position (i.e. energy) of emission line as we scan from low to high energy around the absorption edge.

Automatic on all tunable MX beamlines at SR sources

$$f = f_o + f'(\lambda) + if''(\lambda)$$



Experimental fluorescence scan



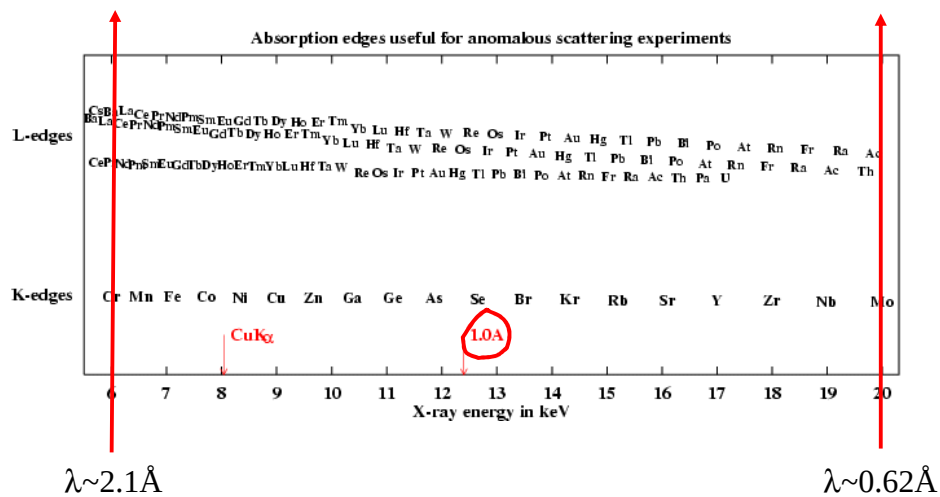
Experimentally determined anomalous scattering factors

$$f''(\omega) = \left(m\omega c / 4\pi Ne^2 \right) \mu(\omega)$$

$$f'(\omega) = (2/\pi) \int_0^\infty [\omega' f''(\omega') / (\omega^2 - \omega'^2)] d\omega'$$

The European Synchrotron | ESRF

Energy of absorption edges that can commonly be accessed on a tunable SR beamline

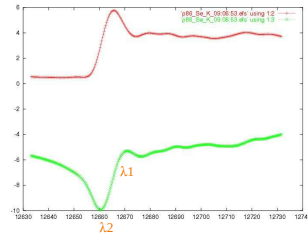


$\lambda \sim 2.1 \text{ \AA}$

$\lambda \sim 0.62 \text{ \AA}$

The European Synchrotron | ESRF

WHY DO WE CARE ABOUT F'' ?



If we collect data at different wavelengths about absorption edge then F_{mean} at different λ changes

$$\lambda^1 F \neq \lambda^2 F$$

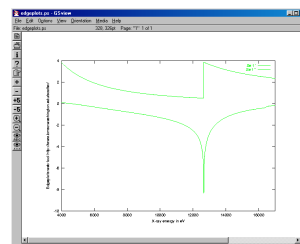
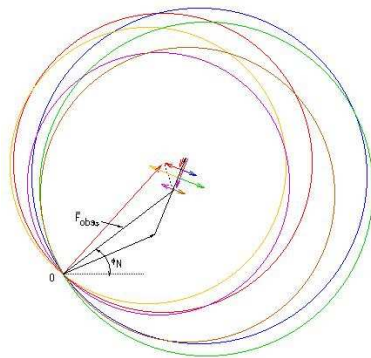
DISPERSIVE difference:

$$\Delta F_{\text{disp}} = |\lambda^1 F_{hkl}| - |\lambda^2 F_{hkl}|$$

By combining measurement of anomalous and dispersive differences we can solve the phase problem using HA derivatives in MX. Technique is called MAD.

The European Synchrotron | ESRF

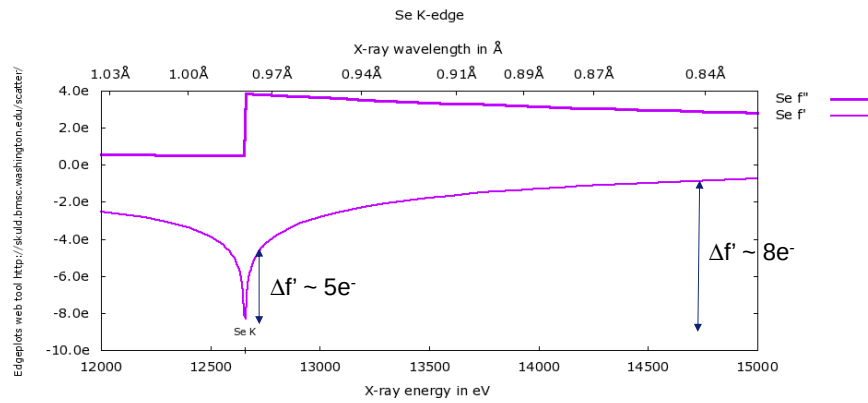
MAD as Quasi-MIR



"From a practical point of view a larger f'' will give more accurate values for the possible solutions [of the phase]....while a larger difference in f' will provide better discrimination in resolving the [phase] ambiguity" Woolfson & Fan, "Solving Crystal Structures" (1995) Cambridge University Press.

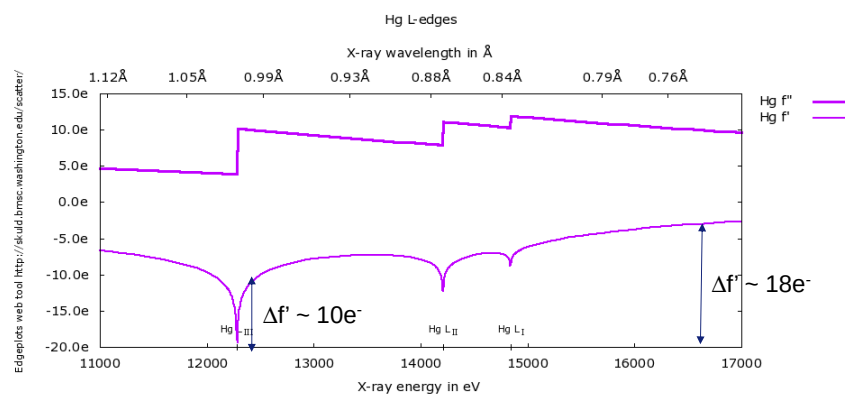
So, a typical MAD experiment uses three wavelengths (peak, inflection point, remote). Peak maximizes f'' , inflection point and (high energy) remote maximize $\Delta f'$.

The European Synchrotron | ESRF

Maximising $\Delta f'$ 

Page 19

The European Synchrotron | ESRF

Maximising $\Delta f'$ 

NB: Efficiency of Pilatus detectors is not optimal at high energy

Page 20

The European Synchrotron | ESRF

(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. SAD is most common method for *de novo* structure solution in MX.
4. Signals are small in comparison to those from isomorphous replacement – data have to be good. How you might want to do a SAD/MAD experiment often depends on signal expected.
5. In MAD have to collect a lot of data from one crystal – radiation damage is a problem!

S/MAD INTENSITY DIFFERENCES

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

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

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

K- absorption edges: f'' usually has a maximum around **4e⁻**. Can rise to **6-7e⁻** with 'white line'. $\Delta f'$ usually has a maximum around **5 - 6e⁻**. 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'. $\Delta f'$ usually has a maximum around **10e⁻**. Can rise to **-20e⁻** if far from edge

A single Hg atom has $Z = 78 e^-$ (i.e. MAD signals are smaller than those from isomorphous replacement. But generally use only one crystal to collect all data so non-isomorphism not a problem.)

From tables: signal usually underestimated except in the case of remote SAD

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

METHODS FOR S/MAD DATA COLLECTION - I

The Random Orientation 'Technique'

Crystal is mounted in a random orientation with respect to the X-ray beam and goniometer axes.

- Data collection strategy required can be obtained using various data collection strategy programs (i.e. BEST). Look at R_{Friedel}
- Takes no real account of any factors that can adversely affect measurement of intensities (i.e. absorption).
- Works very well when we have a relatively large signal ($\geq 2.5\%$ - 3.0%) and crystal diffracts reasonably well.
- If signal is small (i.e. S-SAD) can aim for higher data multiplicity

METHODS FOR S/MAD DATA COLLECTION - II

The Inverse-beam Method

Crystal mounted in a random orientation. Data are collected from $\phi = X^\circ$ to $\phi = Y^\circ$ such that data is as complete as possible for a 'standard' data collection. The Friedel mates of these reflections are then collected by inverting the orientation of the crystal and collecting the equivalent data (i.e. by collecting from $\phi = 180 + X^\circ$ to $\phi = 180 + Y^\circ$). Can collect in a series of small wedges or in contiguous complete sweeps.

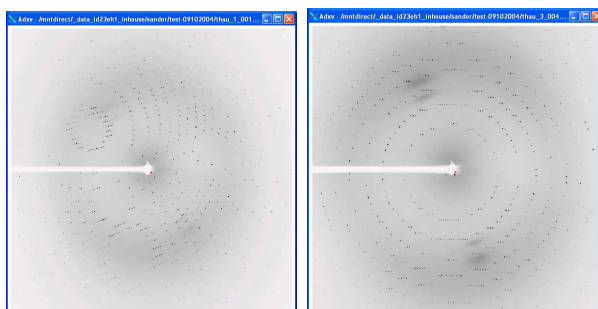
- Friedel pairs are measured with the same path-length through crystal (absorption).
- Friedel pairs are collected close together in time (if we use small wedges).
- Friedel pairs measured with the same multiplicity.
- Helps reduce systematic errors

METHODS FOR S/MAD DATA COLLECTION - III

The 'Aligned-Crystal' Method

Crystal is mounted such that the principle rotation axis of the crystal point group is parallel to the spindle axis of the goniometer. If the principle rotation axis is even-fold then:

- Reflections in a Friedel pair are collected on the same image.
- Reflections in a Friedel pair are collected after same total exposure time/dose.
- Reflections in a Friedel pair suffer from 'same' absorption and radiation damage when they are measured.



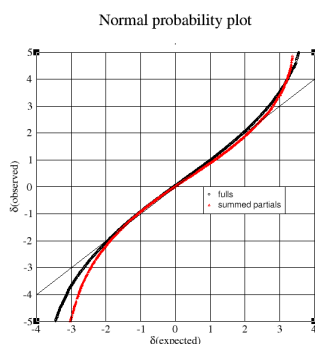
- Mini-Kappa goniometers on many synchrotron BLs make crystal alignment straightforward.
- BEST can tell you which alignment is most useful for optimal experiment.

How can I tell whether my (S)MAD data are good enough for structure solution?

AFTER DATA PROCESSING – ARE THE DATA GOOD?

N	1/d ²	Dmin(Å)	R _{meq}	R _{full}	R _{cum}	R _{anom}	Nanom	Av_I	SIGMA	I/sigma	sd	ln(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	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

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



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

Normal probability analysis, by run & partiality

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

==== Run number: 1, summed partials =====
 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 =====
 All data: 158207 1.003 0.012
 Data within expected delta 0.90: 99967 0.932 0.021

The European Synchrotron | ESRF

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

N	1/d ²	Dmin(Å)	R _{meq}	R _{full}	R _{cum}	R _{anom}	Nanom	Av_I	SIGMA	I/sigma	sd	ln(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	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

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

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

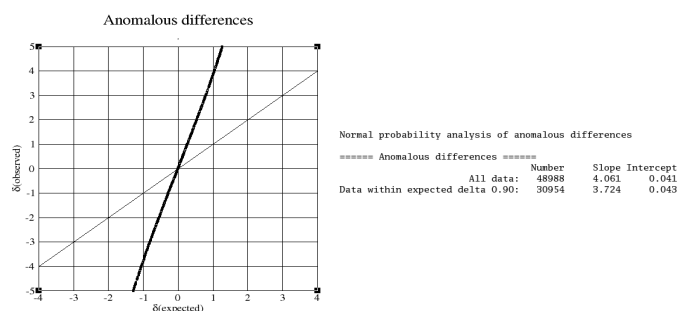
$$R_{anom} > R_{merge} ?$$

The European Synchrotron | ESRF

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

The presence of anomalous signal can usually 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)

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



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

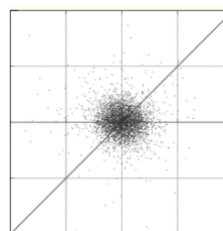
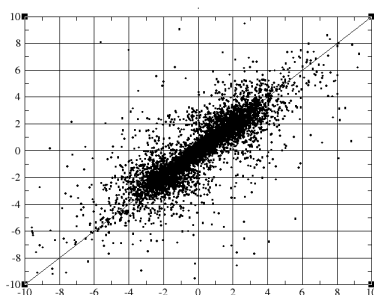


AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

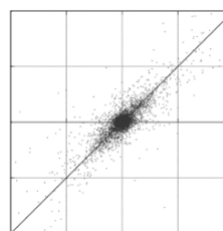
What is the correlation coefficient between anomalous differences in two half data sets chosen at random? What do scatter plots look like?

N	1/resol ²	dmax	CC_anom	N_anom	CC_cen	N_cen	RCR_anom	N_anom	RCR_cen	N_cen	CC_Inean	N_Inean
1	0.0250	6.32	0.990	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	0.621	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	2.30	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	7942	-0.091	383	2.856	7942	0.910	383	0.997	7606

DelAnom scatter plot



no
anomalous
signal



Strong
anomalous
signal

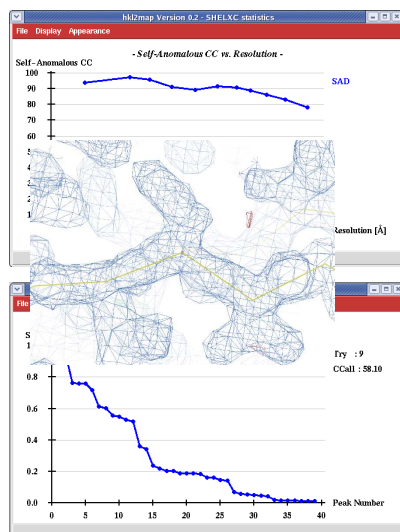
DATA PROCESSING – SUMMARY

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	97.5	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	-	-

F4E; $\Delta F/F \sim 5.4\%$; $\lambda = 0.97 \text{ \AA}$ (S.G. = $P2_12_12_1$; 160° data)

AFTER DATA PROCESSING – SHELXC/SHELXD/SHELXE OUTPUT



EXAMPLE 2

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

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

N	1/d ²	Δmin(Å)	R _{avg}	R _{full}	R _{sum}	R _{anom}	R _{anom}	Av ₁	SIGMA	I/sigma	sd	Rh(I/sd)	N _{obs}	N _{ref}	N _{cent}	FCBIAS	N _{bias}
1	0.0100	9.62	0.049	0.045	0.049	0.049	337	29765	2709	10.7	2329	30.0	2401	319	0	-0.035	905
2	0.0216	6.80	0.044	0.024	0.046	0.054	666	19015	1600	12.4	1819	27.1	4927	666	0	-0.023	1643
3	0.0324	5.55	0.044	0.047	0.051	0.062	981	10446	1175	9.2	1027	22.2	8216	884	0	-0.009	2163
4	0.0433	4.81	0.040	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	1382	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	1366	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	1694	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	1806	0	0.011	5947
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	11298	2002	0	0.014	4265
15	0.1622	2.48	0.130	0.515	0.075	0.134	1953	2189	413	5.3	549	6.4	7504	1976	0	-0.014	2268
16	0.1731	2.40	0.157	0.311	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	0.416	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	0.597	0.082	0.247	2191	1075	508	2.1	642	3.2	7916	2108	0	-0.018	2487

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

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

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

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

#	N	I/d ²	Delta(A)	R _{eq}	R _{full}	R _{anom}	R _{anom}	N _{anom}	Av. I	SIGMA I/sigma	sd	Rn(I/sd)	N _{mean}	N _{ref}	N _{cent}	FBCEIAS	N _{bias}
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	19935	1600	12.4	1639	27.1	4927	666	0	-0.023	1643
3	0.0324	5.55	0.044	0.047	0.051	0.061	981	10446	1175	8.1	1037	21.2	6226	984	0	-0.009	2165
4	0.0432	4.81	0.040	0.048	0.054	0.060	1013	14543	1550	9.4	1364	21.7	6922	1009	0	-0.023	2464
5	0.0541	4.30	0.035	0.071	0.054	0.050	1162	20951	1979	10.5	1839	25.6	8416	1182	0	-0.033	3332
6	0.0649	3.93	0.046	0.069	0.057	0.052	1292	17160	1971	8.7	1531	24.4	8413	1289	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	1386	0	-0.020	4040
8	0.0865	3.40	0.072	0.080	0.061	0.050	1489	12524	1491	8.4	1324	20.9	10969	1485	0	-0.018	4412
9	0.0973	3.21	0.082	0.075	0.061	0.057	1600	8687	1206	7.2	1044	18.3	11887	1599	0	-0.005	4741
10	0.1082	3.04	0.085	0.069	0.065	0.064	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	12286	1774	0	0.003	5510
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	5847
13	0.1406	2.67	0.161	0.168	0.071	0.080	1880	2811	735	3.9	641	9.8	14094	1879	0	0.014	6029
14	0.1514	2.57	0.170	0.204	0.075	0.103	1992	2337	614	3.8	610	7.9	11290	2002	0	0.014	6265
15	0.1622	2.48	0.130	0.315	0.075	0.114	1953	2189	413	5.3	548	6.4	7504	1976	0	-0.014	6268
16	0.1731	2.40	0.157	0.311	0.074	0.131	1984	1861	415	4.5	551	5.6	7462	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	2138	0	-0.019	2567
18	0.1947	2.27	0.223	0.434	0.079	0.170	2115	1391	442	3.1	585	4.3	8145	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	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	2208	0	-0.018	2487

$R_{p.i.m.}$ (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.

#	N	I/cos ²	Delta	N _{anom}	N _{ref}	N _{cent}	I _{anom}	I _{ref}	I _{cent}	I _{anom} /I _{ref}	I _{anom} /I _{cent}	I _{ref} /I _{cent}	R _{anom}	R _{ref}	R _{cent}	R _{anom} /R _{ref}	R _{anom} /R _{cent}	R _{ref} /R _{cent}
1	0.011	9.62	5520	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	0.024	0.028	0.059	0.086
3	0.032	5.55	6351	899	0	97.4	96.4	7.0	96.1	98.2	3.5	0.075	0.098	0.044	0.040	0.037	0.084	0.115
4	0.043	4.81	6961	1019	0	97.0	96.7	6.8	96.5	99.2	3.4	0.071	0.090	0.040	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	0.035	0.034	0.030	0.077	0.098
6	0.065	3.93	9448	1301	0	96.9	96.7	7.3	94.2	98.3	3.6	0.077	0.091	0.040	0.040	0.034	0.093	0.113
7	0.076	3.63	10172	1397	0	96.1	96.5	7.3	95.4	98.8	3.6	0.082	0.093	0.070	0.041	0.034	0.095	0.114
8	0.087	3.40	10994	1459	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	11517	1609	0	96.0	96.4	7.4	95.5	99.1	3.7	0.086	0.100	0.082	0.045	0.040	0.112	0.133
10	0.108	3.04	12563	1693	0	95.6	96.3	7.4	95.1	99.2	3.7	0.111	0.126	0.093	0.058	0.046	0.133	0.156
11	0.119	2.90	13119	1798	0	95.7	96.2	7.5	95.0	99.0	3.7	0.111	0.147	0.112	0.068	0.054	0.156	0.195
12	0.130	2.78	13795	1844	0	95.4	96.1	7.5	94.9	99.1	3.7	0.151	0.163	0.128	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	0.208	0.220	0.161	0.096	0.073	0.228	0.234
14	0.151	2.57	14514	2030	0	94.4	95.8	5.6	91.7	97.7	2.8	0.209	0.221	0.170	0.110	0.090	0.237	0.262
15	0.162	2.48	14949	2044	0	94.1	95.6	3.8	89.9	94.9	1.9	0.189	0.205	0.130	0.130	0.102	0.191	0.234
16	0.173	2.40	15011	2056	0	93.4	95.4	3.8	88.8	94.5	1.9	0.221	0.232	0.157	0.157	0.121	0.232	0.281
17	0.184	2.33	14494	2222	0	93.0	95.3	3.8	90.0	95.6	2.0	0.285	0.301	0.201	0.201	0.152	0.301	0.357
18	0.195	2.27	13882	2194	0	93.5	95.2	3.8	90.1	96.0	1.9	0.316	0.328	0.223	0.223	0.167	0.316	0.392
19	0.206	2.21	12770	2304	0	92.8	95.0	3.8	88.8	95.1	2.0	0.392	0.406	0.278	0.278	0.236	0.419	0.491
20	0.216	2.15	11447	2290	0	90.7	94.7	3.7	86.5	94.0	1.9	0.461	0.473	0.324	0.324	0.243	0.501	0.573

R _{anom}	R _{p.i.m.}
0.048	0.029
0.054	0.026
0.062	0.040
0.060	0.039
0.050	0.034
0.052	0.040
0.050	0.042
0.050	0.044
0.057	0.049
0.046	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	0.324
0.267	0.354

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.)

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

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

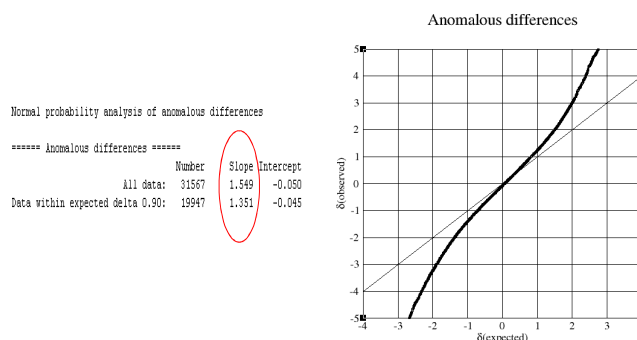
$$R_{p.i.m.} = \sum_{hkl} [1/(N-1)^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)]$$

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

The European Synchrotron | ESRF

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

Normal probability plot - anomalous differences



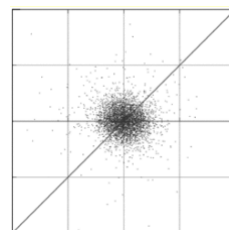
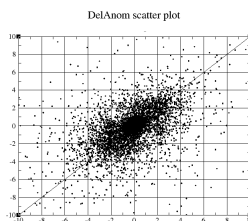
(Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86; Evans, P. (2006). *Acta Cryst.* **D62**, 72-82.)

The European Synchrotron | ESRF

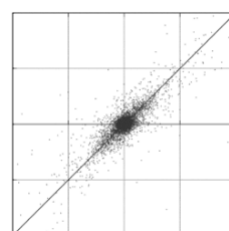
AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

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	delta _{cc}	CC_anom	N_anom	CC_cen	N_gen	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	0.0432	4.81	0.578	974	-	0	1.916	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	2.76	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	-



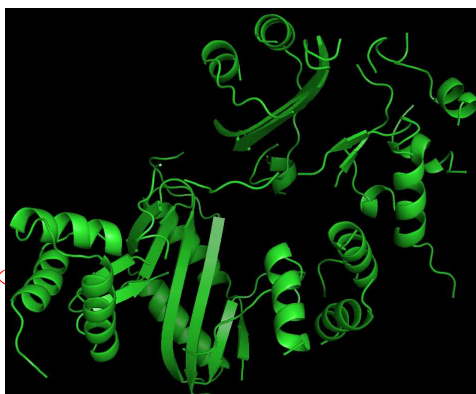
no
anomalous
signal



Strong
anomalous
signal

The European Synchrotron | ESRF

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



ON	MPARED	I/SIGMA	R-meas	Rmrgd-F	Anomal Corr	SigAno	Nano
2315	31.07	4.4%	2.4%	88%	3.023	302	
4849	30.13	4.4%	2.6%	89%	3.162	646	
6017	27.51	5.0%	3.1%	87%	3.019	808	
6900	27.19	5.1%	3.2%	84%	2.917	933	
7921	27.92	4.7%	2.9%	78%	2.307	1062	
8760	26.59	5.0%	3.1%	76%	2.200	1183	
9480	24.86	5.2%	3.6%	72%	2.086	1274	
10555	23.39	5.6%	3.9%	67%	1.935	1408	
11447	21.10	6.3%	4.6%	66%	1.868	1512	
12271	17.85	7.7%	6.3%	63%	1.771	1612	
12750	15.51	9.3%	7.8%	59%	1.597	1661	
13399	13.58	10.9%	9.2%	53%	1.445	1745	
13944	11.32	13.6%	12.4%	49%	1.333	1809	
7389	7.84	15.7%	21.0%	34%	1.134	1669	
6964	6.86	16.3%	23.7%	28%	1.023	1643	
7308	6.09	18.1%	25.9%	29%	1.003	1720	
7274	5.14	22.4%	33.3%	21%	0.933	1710	
7160	4.61	24.6%	36.9%	19%	0.921	1677	
7494	3.93	29.9%	44.3%	12%	0.851	1725	
4662	3.47	32.9%	46.3%	12%	0.840	1057	
168859	14.16	6.6%	10.3%	60%	1.569	27156	

The European Synchrotron | ESRF

- Decide how you want to collect your data. **Don't overcomplicate things unless you have to.**
- There are plenty of rapidly available indicators as to whether:
 - Data are good enough
 - Experimental phases might be useful.
 - The best indicator is to actually solve the structure!
- Process/analyse data as you go along
- Take your structure home with you.
- Repeat experiments if you have to - but only if you have to.