
Experimental Phasing with SHARP/autoSHARP

Clemens Vornrhein
Global Phasing Ltd.

DLS/CCP4 2020

SHARP/autoSHARP:

- principles
- the “pipelines”
- tools around it

Examples:

- Radiation damage
- Improve on first results
- SAD, MAD, MIR ... alphabet soup

Non-automated structure solution:

- Why bother?
- **SHARP** – advanced phasing
- (Some) tips ...

Tutorials:



<http://www.globalphasing.com/sharp/wiki/>
<http://www.globalphasing.com/autoproc/wiki/>
<http://www.globalphasing.com/buster/wiki/>

SHARP/autoSHARP Design

SHARP:

- getting **best phases** possible
- applicable to large variety of **phasing scenarios**:
MAD+native, SAD+SIR, ...
- extensive user control
- interfaced to downstream phase improvement
- online help

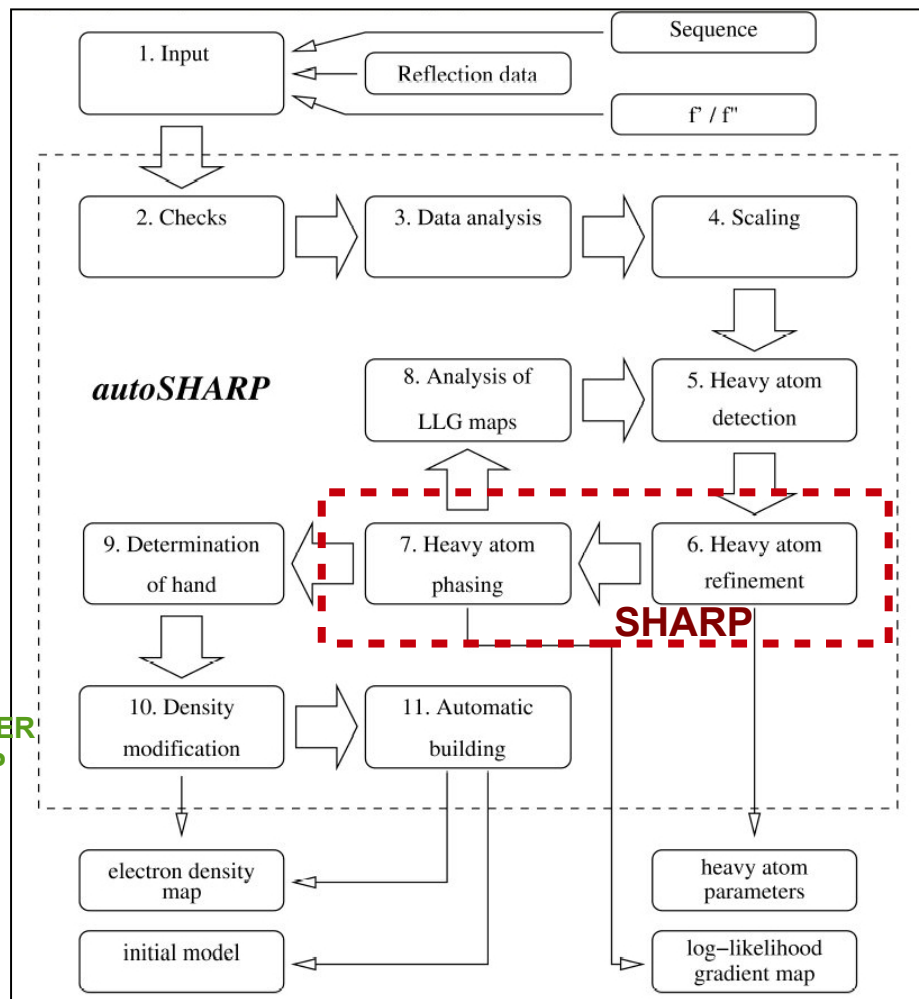
Bricogne, G., Vonnrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Cryst. D59, 2023-2030.

autoSHARP:

- centred around **SHARP**
- supports common phasing scenarios: **MAD, SAD, SIR(AS), MIR(AS)**
- minimum input
- explanation of results
- online help

Vonnrhein, C, Blanc, E., Roversi, P and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol Biol 364, 215-30.

autoSHARP “pipeline”



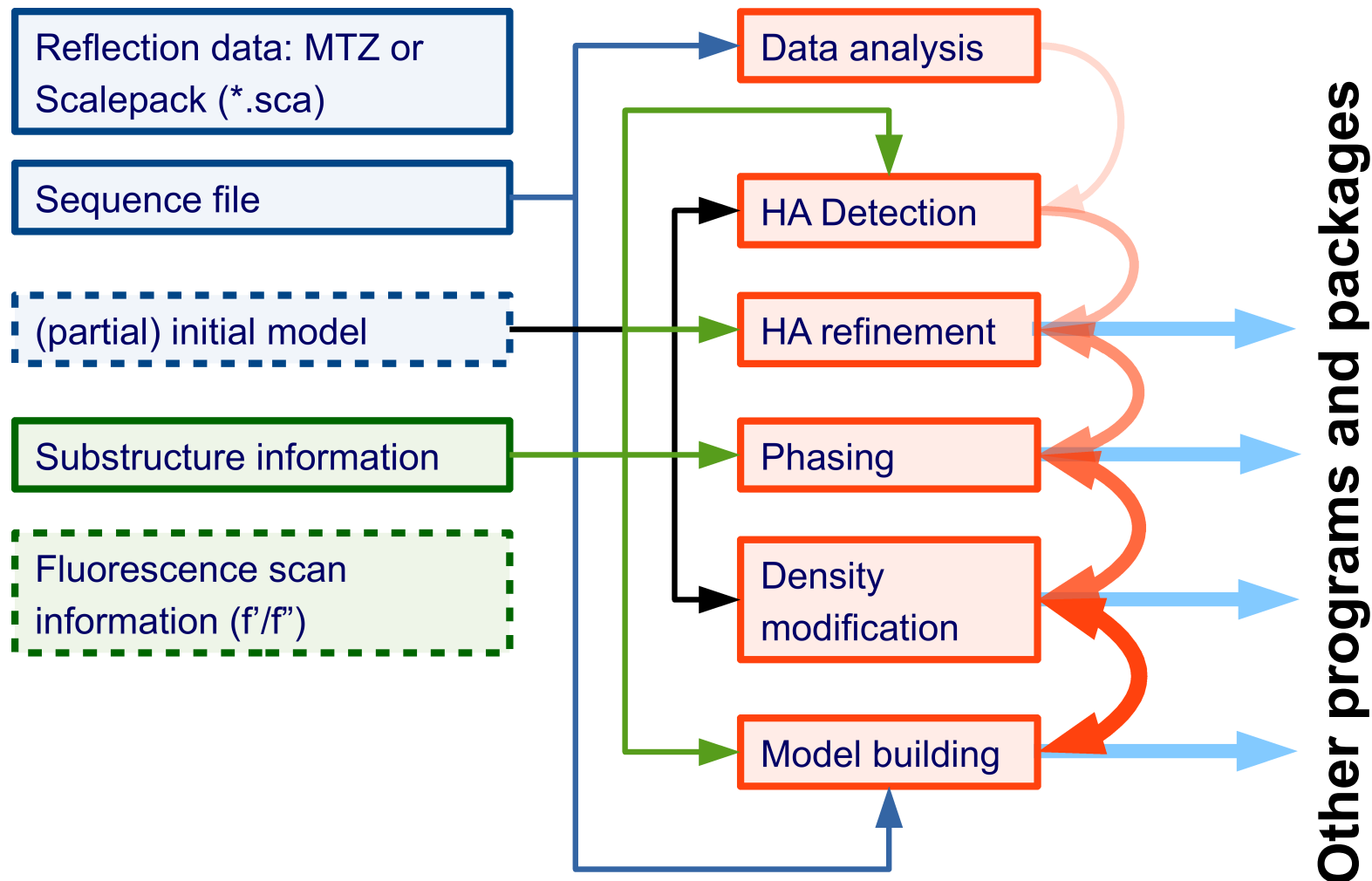
minimum amount of input

feedback:

internally and to the user

several levels of results

In short: what do I need and get?



autoSHARP - Input

- experiment type: **SAD, MAD, SIR(AS), MIR(AS)**
- **data files:** MTZ, Scalepack (merged or unmerged)
- additional information:
 - **no. of HA sites** (estimate)
 - **type of HA** (predominant one)
- optional information:
 - sequence
 - wavelength, f'/f''
 - space group

The screenshot displays the CCP4i autoSHARP input window. At the top, a note states: "NOTE This task uses the SHARP/autoSHARP programs which are not distributed with CCP4 - see: www.globalphasing.com". The interface is organized into several sections:

- General parameters:** Includes fields for Job title (autoSHARP run), Project identifier (test), Run a (SAD/MAD), experiment (experiment), Input file(s) (MTZ), format with spacegroup (None), Speed vs. accurate setting of (5 - accurate), Run up to ((5)), + Automatic model building, using SHELXD, and for HA detection.
- Asymmetric Unit:** Includes Molecular weight (0.0), No. of residues (0), Sequence (oBUSTER-CCP4), Atom type (Se), No. of sites (2), and Known HA sites (oBUSTER-CCP4).
- Dataset parameters:** This section contains three identical blocks for Dataset no. 1, 2, and 3. Each block includes:
 - Identifier: w1
 - Wavelength: 1.5418
 - f': 0.0
 - f'': 0.0
 - Datafile: oBUSTER-CCP4
 - FMID: F, SMID: SIGF
 - DANO: DANO, SANO: SIGDANO
 - ISYM: ISYM

At the bottom, there are buttons for "Run", "Save or Restore", and "Close".

CCP4i interface

Command-line interface

run_autoSHARP.sh -h

**writes HTML output
(view in browser)**

**creates scripts ready to
start in Coot**

SAD:

```
run_autoSHARP.sh -seq 1o22.pir -ha "Se" \
-wvl 0.9778 peak -7 5 -sca 1o22_peak.sca
```

MAD:

```
run_autoSHARP.sh -seq 3isy.pir -ha "Se" \
-wvl 0.97934 infl -11 3.3 -sca 3isy_aimless_0.97934A.sca \
-wvl 0.91162 hrem -1.8 3.3 -sca 3isy_aimless_0.91162A.sca
```

SIR(AS):

```
run_autoSHARP.sh -seq 1GXT.pir -nat -mtz 1GXT_nat.mtz \
-ha "Hg" -nsit 2 \
-wvl 0.99970 peak -16 10 -mtz 1GXT_hg.mtz
```

MIR(AS):

```
run_autoSHARP.sh -seq 3zft.pir -nat -mtz 3zft_nat.mtz \
-ha "Hg" -nsit 1 -wvl 1.54179 -mtz 3zfq_Hg.mtz \
-ha "Ir" -nsit 2 -wvl 1.54179 -mtz 3zfr_Ir.mtz
```

partial model:

```
run_autoSHARP.sh -seq 3get.pir -ha "Se" \
-pdb 3ffh_ala_MR.pdb \
-wvl 0.9789 peak -8 4 -sca 3get.sca
```

SAD with Ta6Br12 cluster:

```
run_autoSHARP.sh -seq 4cv5.pir -ha "Ta6Br12:Ta" \
-wvl 1.25472 peak -mtz 4cv5.mtz
```

<https://www.globalphasing.com/sharp/wiki/index.cgi?TutorialDIs2020>

autoSHARP - Analysis

- Consistency (N>1):
 - cell (significant changes)
 - spacegroup (screw axes)
 - indexing (alternate possibilities)
 - sequence (Se-Met, S-SAD)

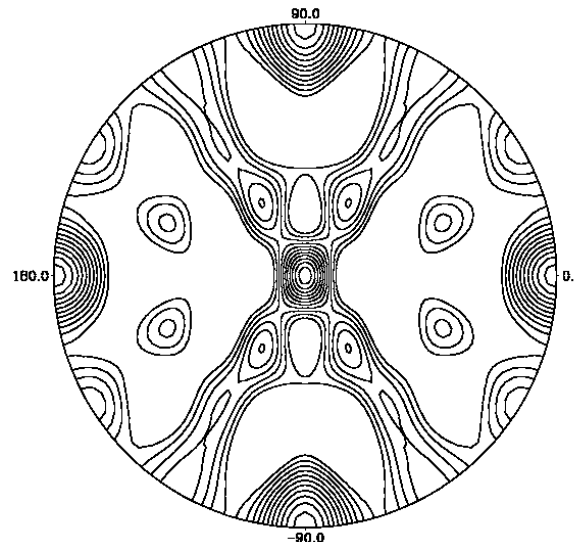
- Data quality:

- CC(Δ ano), R-values, completeness
- outliers (E-values)

SIGNAL

NOISE

- self-rotation function



- native Patterson
- Matthews parameters (probability, Rupp)

?/ASU

Also from data processing:

- Consistency ($N > 1$):
 - cell (significant changes)
 - spacegroup (screw axes)
 - indexing (alternate possibilities)
 - sequence (Se-Met, S-SAD)

SIGNAL

- Data quality:
 - CC(Δ ano), R-values, completeness
 - outliers (E-values)

NOISE

- self-rotation function

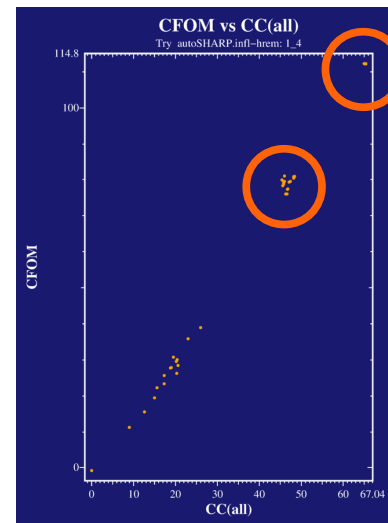
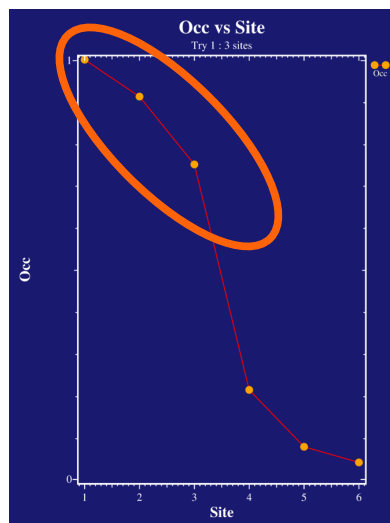
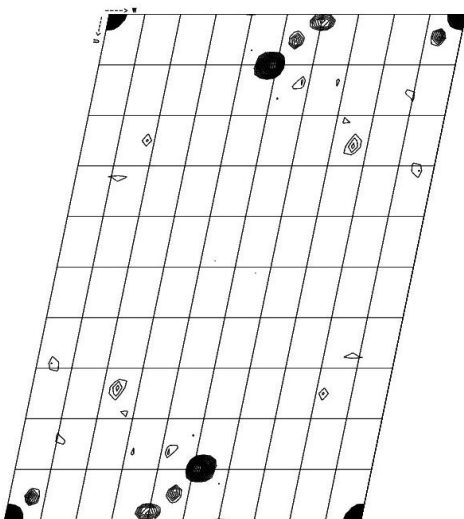
- Are there convincing indications of heavy atom signal?
- Did the fluorescence scan show existence of HA in sample?

- native Patterson
- Matthews parameters (probability, Rupp)

?/ASU

autoSHARP – HA detection

- only data with significant signal (from *Analysis* step)
 - try different scenaria (MAD, peak-SAD, 2-wvl MAD, ...)
 - adjust number of HA sites (e.g. if exact number not known)
 - use solution with best score
- **SHELXC/SHELXD** ¹⁾ :
 - stop if significant solution found
 - present results graphical
 - try detecting ‘jump’ in occupancy
- Alternative:** give your own set of initial sites to autoSHARP



1) Schneider, T. R. & Sheldrick, G. M. (2002). Acta Cryst. D58, 1772-1779.

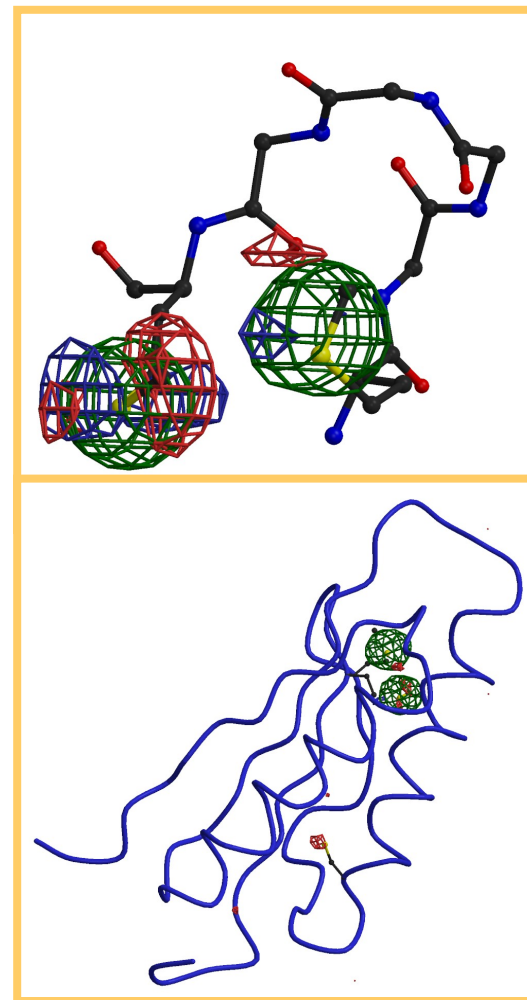
autoSHARP - SHARP

- HA refinement and phasing
- overall anisotropic scaling
- analysis of log-likelihood gradient maps:

G. Bricogne (1992). A Statistical Formulation of the Molecular Replacement and Molecular Averaging Methods. In: "The Molecular Replacement Method, Proceedings of the CCP4 Study Weekend 31 January - 1st February 1992" (W. Wolf, E.J. Dodson & S. Gover, eds.) pp. 62-75. Daresbury Laboratory.

- remove spurious sites
 - add additional sites
 - switch on refinement of f' and/or f''
 - feed-back into **SHARP**
- Best possible phases**

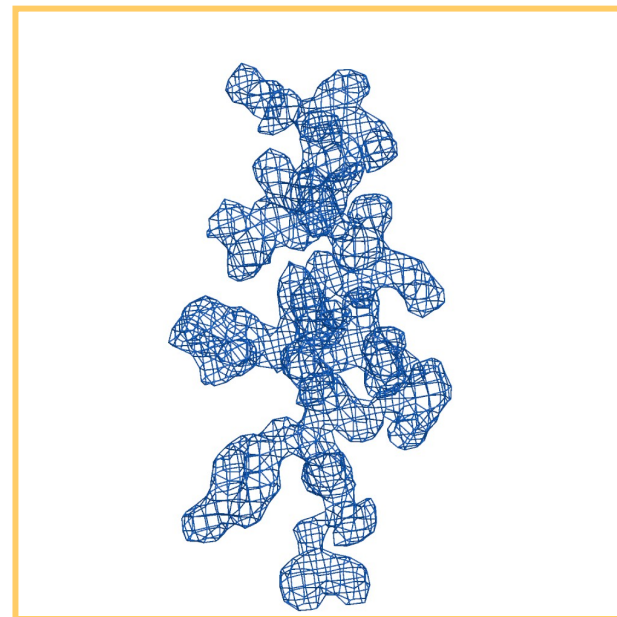
- phases in two hands (enantiomorphs)



autoSHARP – density modification

- **SOLOMON**¹⁾ program
- start from resolution with significant phase information
- take heavy-atoms into account
- detect correct hand (based on $CC(E^{**2})$ and “contrast”)
- optimize solvent content
- adjust no. of molecules/ASU

Similar to MR (test enantiomorphs)



1) J. P. Abrahams and A. G. W. Leslie (1996). Acta Cryst. D52, 30-42.

autoSHARP – automatic building

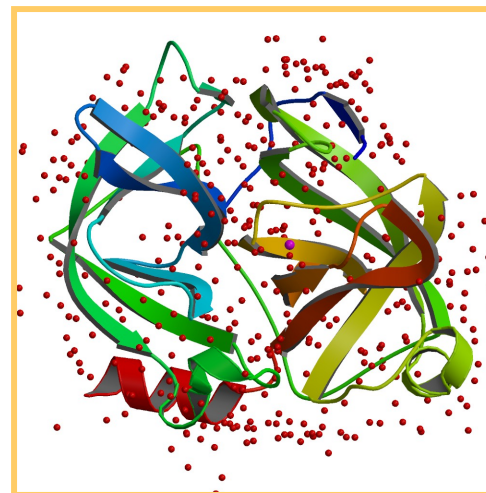
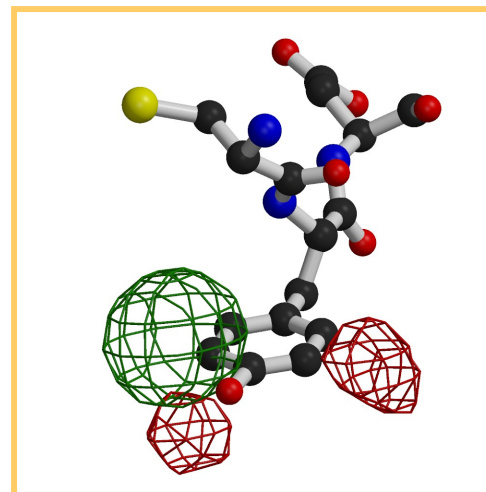
- **ARP/wARP**¹⁾ suite of programs (latest ARP/wARP as distributed via CCP4)
- **PARROT/BUCCANEER**²⁾ iterative density-modification and automatic building tool (*long_john_silver.sh -h*)
- Results from heavy atom phasing (SHARP) plus density modification (SOLOMON) can easily be feed into other building programs:
 - Good experience with BUCCANEER at medium/low resolution and for quick/initial building
 - Taking advantage of NCS information (if possible): very powerfull

1) Perrakis, A., Morris, R. & Lamzin, V.S. (1999). Nat. Struct. Biol. 6, 458-63.

2) K. Cowtan (2006) Acta Cryst. D62, 1002-1011.

autoSHARP - Output

- self-rotation plots, native Patterson, difference Patterson (Harker sections), most meaningful/important analysis plots/tables
- scaled datasets (MAD, SIR, MIR)
- initial HA sites
- refined (completed) HA sites (→ e.g. ACORN)
- **SHARP** phases (→ e.g. SHELXE building)
- LLG (residual) maps
- density modified phases/maps (→ e.g. NCS-detection via GETAX)
- initial model(s) (→ completion, refinement etc)



Tools around SHARP/autoSHARP

“web-based” interface (local installation)

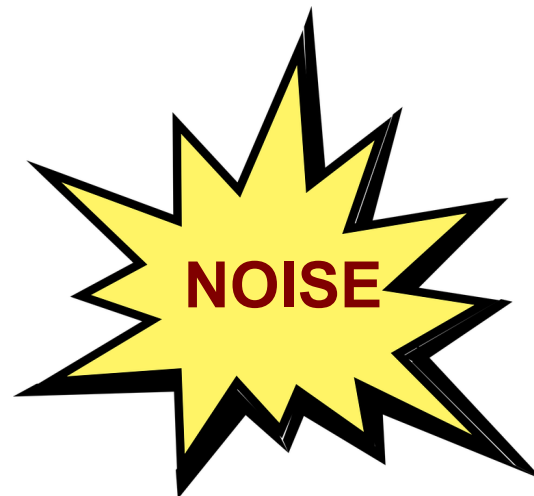
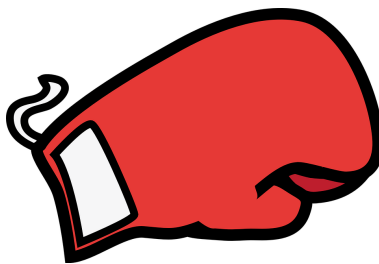
- Any phasing scenario with **SHARP**, e.g.:
 - MAD+native
 - MAD+derivative+native
 - radiation damage (refining exponential occupancy decay)
 - cluster compounds (spherically averaged)
 - complete control
- separate density modification control panel:
 - inclusion of partial model
 - NCS detection (**GETAX**¹⁾) and averaging (**DM**²⁾)
 - extraction of HLs for **SHARP** (from PDB or MTZ)
 - difference Fourier maps (detection of additional sites)
 - **ARP/wARP** control panel
 - view results (maps, model, skeleton, ...)
- LLG (residual) map panel: viewing, analysis, peak-picking

1) C. Vornrhein and G. E. Schulz, Acta Cryst., D55, 225 - 229 (1999)

2) K. Cowtan (1994), Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, 31, p34-38.

How to improve on first results

- after **autoSHARP**:
 - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)



Warning messages - 1

WARNING : No f'/f'' values defined for atom of type "I" in wavelength 1: automatically set to calculated values for wavelength 0.8900 Å (-0.4308,2.6934).

WARNING : completeness of only 65.00% (below 90%) for native dataset in the range 18.47 - 10.0 Å. This might give serious problems later! Maybe you haven't included your **low-resolution data** into this dataset? Or there were a lot of overloads in the reflection patter? Anyway, it is **never** good to miss this many low resolution reflections.

WARNING : there are serious differences between 2 amplitudes from different datasets (as judged by analysing E values). If these appear only in specific resolution ranges or shells you might be able to improve results by restricting e.g. low resolution. Here is a list of the reflections that look **suspicious**:

H	K	L	Reso	infl	hrem (all)
1	0	2	42.02	425.45	3.15 *
1	1	1	45.00	539.98	3.82 *

Warning messages - 1

WARNING : No f'/f'' values defined for atom of type "I" in wavelength 1: automatically set to calculated values for wavelength 0.8900 Å (-0.4308,2.6934).

Take fluorescence scan home (and do one in the first place)

WARNING : completeness of only 65.00% (below 90%) for native dataset in the range 18.47 - 10.0 Å. This might give serious problems later! Maybe you haven't included your **low-resolution data** into this dataset? Or there were a lot of overloads in the reflection patter? Anyway, it is **never** good to miss this many low resolution reflections.

Overloads?

WARNING : there are serious differences between 2 amplitudes from different datasets (as judged by analysing E values). If these appear only in specific resolution ranges or shells you might be able to improve results by restricting e.g. low resolution. Here is a list of the reflections that look **suspicious**:

H	K	L	Reso	infl	hrem (all)
1	0	2	42.02	425.45	3.15 *
1	1	1	45.00	539.98	3.82 *

Beamstop masking?

Warning messages - 2

WARNING : there are serious differences between 49 amplitudes from different datasets (as judged by analysing E values). If these appear only in specific resolution ranges or shells you might be able to improve results by restricting e.g. low resolution. Here is a list of the reflections that look **suspicious**:

H	K	L	Reso	natder1_Hg_peak (all)		
5	2	9	7.30	252.60	613.66	*
8	4	-62	2.22	727.50	3.86	*
9	9	51	2.22	684.00	3.80	*
10	0	-11	4.74	80.20	332.09	*
10	2	-61	2.22	453.10	4.00	*
10	6	55	2.22	817.30	4.03	*
10	10	45	2.22	1732.80	6.44	*
11	0	-61	2.23	919.30	4.14	*
11	8	48	2.22	1050.60	5.26	*
11	8	-48	2.22	530.20	3.86	*
12	7	44	2.30	35.40	108.02	*
12	10	38	2.22	1086.80	5.63	*
13	0	13	3.68	1516.30	16.43	*
13	4	51	2.23	1306.20	11.32	*
13	6	-47	2.22	1448.40	3.94	*
13	8	41	2.22	1032.70	5.22	*
13	10	33	2.22	1070.80	5.81	*
13	11	-28	2.22	1466.30	8.43	*
13	13	9	2.22	1339.00	9.46	*

Warning messages - 2

WARNING : there are serious differences between 49 amplitudes from different datasets (as judged by analysing E values). If these appear only in specific resolution ranges or shells you might be able to improve results by restricting e.g. low resolution. Here is a list of the reflections that look **suspicious**:

H	K	L	Reso	natder1_Hg_peak (all)		
5	2	9	7.30	252.60	613.66	*
8	4	-62	2.22	727.50	3.86	*
9	9	51	2.22	684.00	3.80	*
10	0	-11	4.74	80.20	332.09	*
10	2	-61	2.22	453.10	4.00	*
10	6	55	2.22	817.30	4.03	*
10	10	45	2.22	1732.80	6.44	*
11	0	-61	2.23	919.30	4.14	*
11	8	48	2.22	1050.60	5.26	*
11	8	-48	2.22	530.20	3.86	*
12	7	44	2.30	35.40	108.02	*
12	10	38	2.22	1086.80	5.63	*
13	0	13	3.68	1516.30	16.43	*
13	4	51	2.23	1306.20	11.32	*
13	6	-47	2.22	1448.40	3.94	*
13	8	41	2.22	1032.70	5.22	*
13	10	33	2.22	1070.80	5.81	*
13	11	-28	2.22	1466.30	8.43	*
13	13	9	2.22	1339.00	9.46	*

Ice-rings?

Warning messages - 3

NOTE : 4 reflexions have large anomalous differences:

H	K	L	FMID	SMID	DANO	SANO	ABS(DANO)/FMID
-3	1	12	203.27	5.43	398.98	7.68	1.96 *
-2	2	12	79.28	2.86	-150.88	4.05	1.90 *
-2	6	17	177.12	8.44	-259.31	11.94	1.46
-1	1	2	28.28	2.17	-43.53	3.07	1.54

WARNING : We will remove 2 reflexions with a ABS(DANO)/FMID ratio > 1.9

WARNING : the two hands don't show a significant difference in the two scores! This might be due to a very weak phasing signal, wrong heavy atom sites or something else (single site in polar spacegroups?). Please check your heavy atom sites solution carefully (does it look convincing?) as well as the LLG/residual maps from SHARP (for minor sites or something pathologically going on). However, if you start seeing meaningful features in the electron density map after density modification (e.g. secondary structures) this might not be a cause for concern.

Warning messages - 3

NOTE : 4 reflexions have large anomalous differences:

H	K	L	FMID	SMID	DANO	SANO	ABS(DANO)/FMID
-3	1	12	203.27	5.43	398.98	7.68	1.96 *
-2	2	12	79.28	2.86	-150.88	4.05	1.90 *
-2	6	17	177.12	8.44	-259.31	11.94	1.46
-1	1	2	28.28	2.17	-43.53	3.07	1.54

Beamstop?
Outlier rejection?

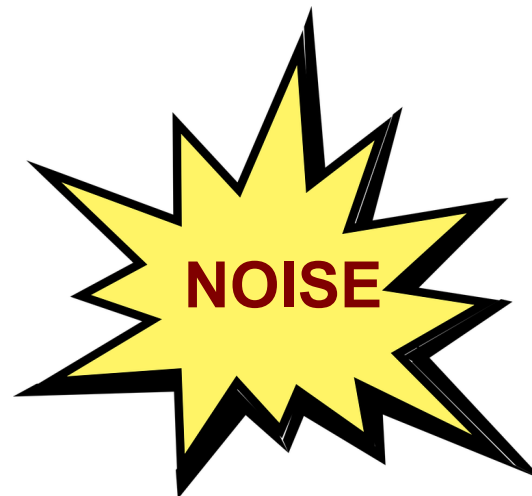
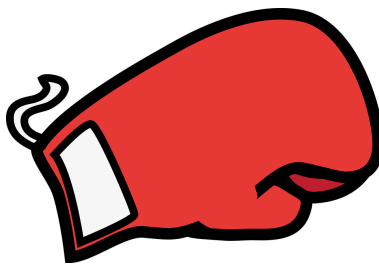
WARNING : We will remove 2 reflexions with a ABS(DANO)/FMID ratio > 1.9

WARNING : the two hands don't show a significant difference in the two scores! This might be due to a very weak phasing signal, wrong heavy atom sites or something else (single site in polar spacegroups?). Please check your heavy atom sites solution carefully (does it look convincing?) as well as the LLG/residual maps from SHARP (for minor sites or something pathologically going on). However, if you start seeing meaningful features in the electron density map after density modification (e.g. secondary structures) this might not be a cause for concern.

Not good ... no
solution?

How to improve on first results

- after **autoSHARP**:
 - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)



How to improve on first results

- after **autoSHARP**:
 - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)
- **autoPROC (2005-)**: www.globalphasing.com/autoproc/
 - expert system (using XDS, POINTLESS, AIMLESS, XSCALE)
 - multiple lattices, multi-wavelength, multi-sweep, multi-orientation
 - anisotropy (STARANISO)

Auto Processing

Fast DP: ✓ Xia2/3dii ✓ DIALS: ✓ Xia2/MultiProc ✓ **autoPROC: ✓**

Type	Resolution	Spacegroup	Mn<I/sig(I)>	Rmeas Inner	Rmeas Outer	Completeness	Cell	Status
xia2 dials	45.25 - 2.14	19.5	0.060	1.311	100.0			processing successful
xia2 3dii	64.09 - 2.12	20.0	0.057	2.519	99.9			processing successful
autoPROC	90.66 - 1.96	15.4	0.062	26.207	99.9			processing successful
fast_dp	28.66 - 2.48	34.8	0.037	0.958	99.8			processing successful
autoPROC+STARANISO	90.66 - 2.06	18.6	0.056	5.020	95.7			processing successful

xia2 dials xia2 3dii **autoPROC** fast_dp **autoPROC+STARANISO**

Beam Centre X Y

Start 156.14 167.62

Refined 156.21 167.64

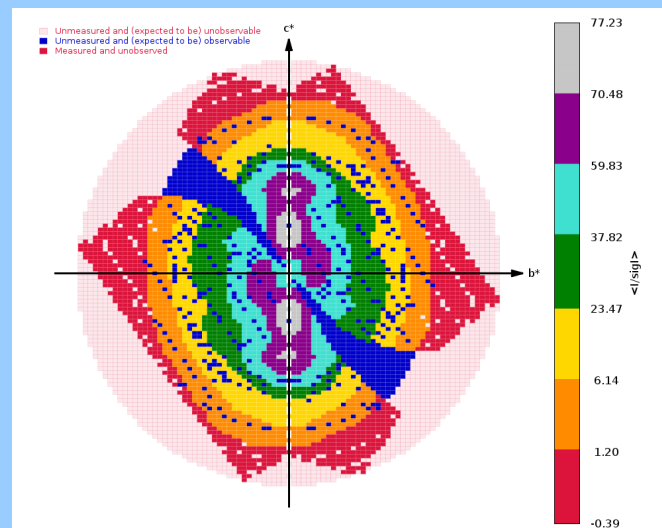
Δ -0.07 -0.02

Space Group A B C α β γ

Shell	Observations	Unique	Resolution	Rmeas	I/sig(I)	CC Half	Completeness	Multiplicity	Anom Completeness	Anom Multiplicity	CC Anom
outerShell	63907	786	2.06 - 2.12	5.020	1.4	0.7	54.4	81.3	54.2	43.3	0.1
innerShell	53654	784	6.00 - 90.66	0.056	75.8	1.0	100.0	68.4	100.0	44.9	0.6
overall	1210684	15729	2.06 - 90.66	0.266	18.6	1.0	95.7	77.0	95.4	42.3	0.1

Downstream Processing

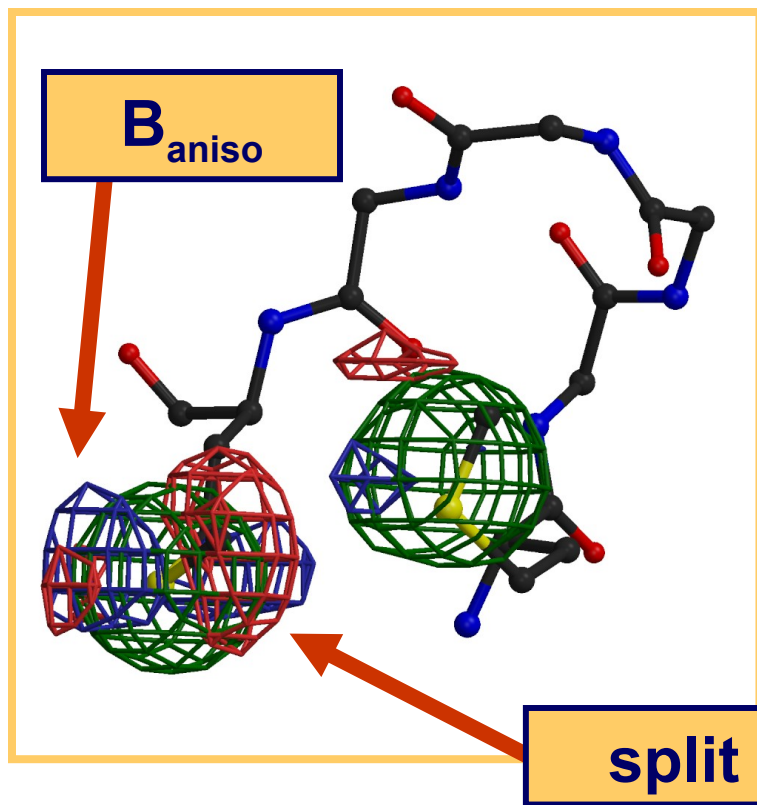
Fast EP: ? Dimple ? MrBUMP ? Big EP/XDS ? Big EP/DIALS ?



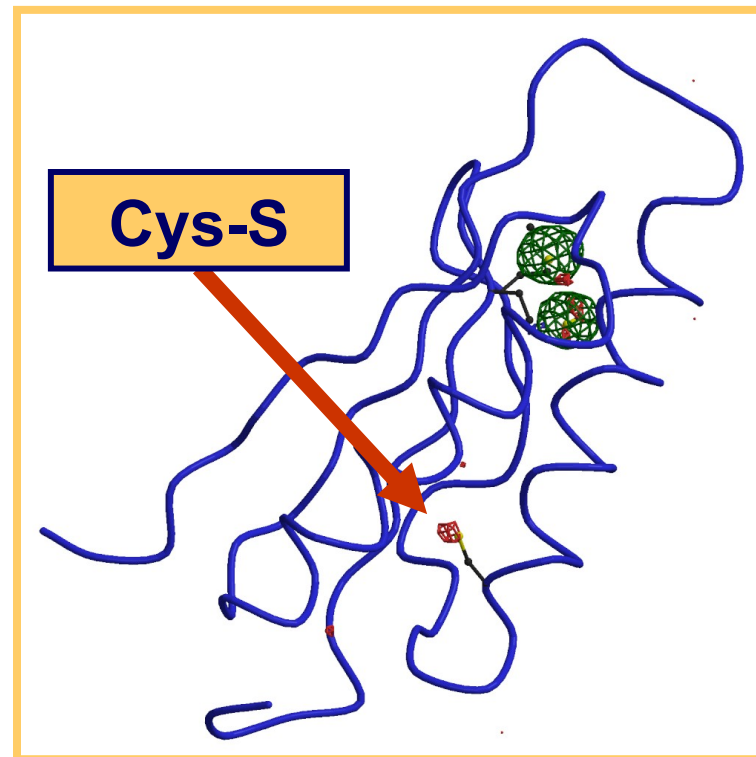
How to improve on first results

- after **autoSHARP**:
 - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)
 - analyse **log-likelihood gradient** (residual) maps: additional sites, f'/f'' refinement, anisotropic B-factor for sites, ...

Example (LLG maps)



Se sites (LLG peak)



Anomalous LLG

How to improve on first results

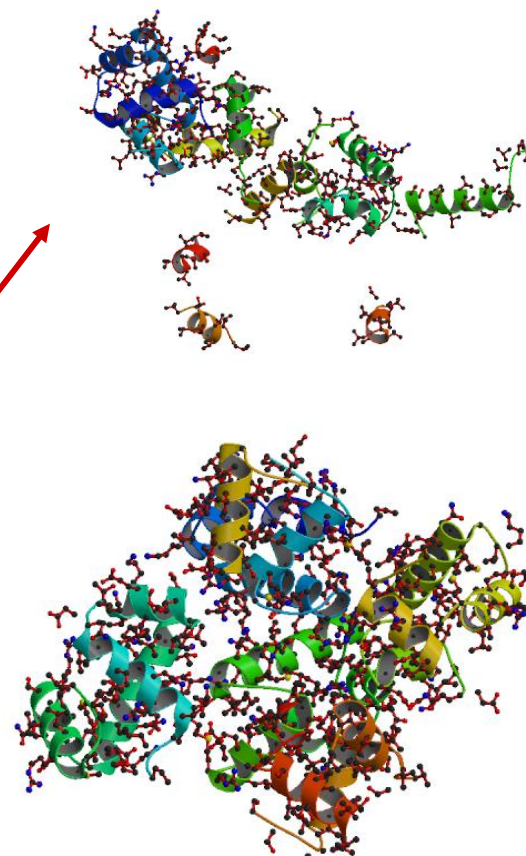
- after **autoSHARP**:
 - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)
 - analyse **log-likelihood gradient** (residual) maps: additional sites, f'/f'' refinement, anisotropic B-factor for sites, ...
 - significant difference between score of **original/inverted** hand?
 - feed-back of density modified phases into **SHARP** (improving LLG maps)
 - feed-back of **partial model** into density modification and/or **SHARP**

Example (feed-back of partial model)

- GerE (3wvl-MAD, 2.66 Å, 444 aa, 6 mol)
- cycling ARP/wARP with DenMod:
using model in first DenMod cycle(s)

Cycle	DenMod	ARP/wARP		
		built	docked	R / Rfree
0	33.8%	251 / 13	103 / 2	0.249 / 0.513
1	40.3%	262 / 9	128 / 2	0.251 / 0.465
2	37.3%	310 / 10	239 / 4	0.236 / 0.432
3	36.5%	315 / 11	246 / 5	0.232 / 0.432

Morris, R.J., Zwart, P.H., Cohen, S., Fernandez, F.J., Kakaris, M., Kirillova, O., Vonnrhein, C., Perrakis, A. & Lamzin, V.S. (2004) Breaking good resolutions with ARP/wARP. *J. Synchr. Rad.* 11, 56–59



External Phase Information

- **SHARP** allows use of additional phase information in likelihood function

E de la Fortelle & G Bricogne (1997)
 "Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods". Methods in Enzymology 276, 472-494.

- often encoded as HL-coefficients
- from density-modification, non-isomorphous crystal, model, etc
- Interface (SHARP and autoSHARP) allows easy use of this feature
- Beneficial for:
 - Stabilising parameter refinement
 - Sensitivity of log-likelihood gradient maps
 - Overcoming non-isomorphism issues

"MR-SAD"

HA detection

- **autoSHARP** can do:

"ExtPhas-SAD"
"ExtPhas-MAD"

"ExtPhas-SIR"
"ExtPhas-MIR"

"ExtPhas-MIRAS"
"ExtPhas-SIRAS"

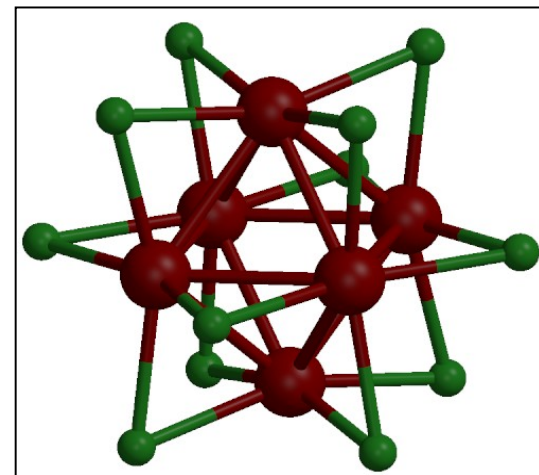
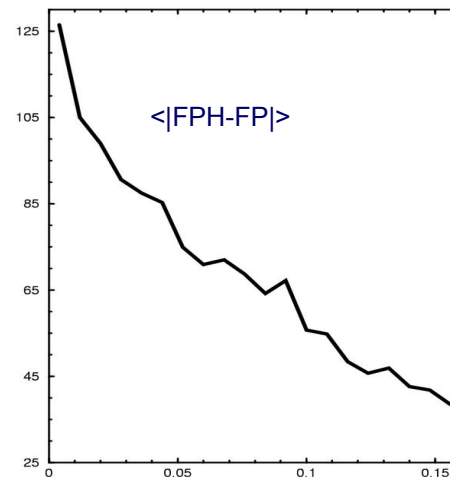
- **SHARP** can also do:

"ExtPhas-MAD-NATIVE"
"ExtPhas-XYZ"

Why not MIR?

- **Non-isomorphism**
 - datasets from different crystals
 - aggressive soaks, co-crystallization
- no. of sites (and occupancy) not well known
- 'old-fashioned'

- **BUT:**
 - spectacular signal
 - soaking can be very fast
 - no need for Se-Met expression
 - soaks can be used for SAD/MAD too
 - quick soaks (Dauter) more gentle to crystals
 - good statistics:
 $\langle |FPH-FP| \rangle$



“Advanced” experimental phasing

- **Spherical average clusters** (getting a hold on very large complexes)
- **Exponential occupancy decay** (turning site-specific radiation damage into something positive)
- **Anisotropy of anomalous scattering** (expect the unexpected...)
- **Breaking of symmetries** (... and the even more unexpected)

RESEARCH PAPERS

Acta Cryst. (2004). **D60**, 1024-1031
<https://doi.org/10.1107/S0907444904006377>



Phasing in the presence of severe site-specific radiation damage through dose-dependent modelling of heavy atoms

M. Schiltz, P. Dumas, E. Ennifar, C. Flensburg, W. Paciorek, C. Vornrhein and G. Bricogne

The case of a brominated RNA crystal structure determination in which standard three-wavelength MAD phasing was unsuccessful because of fast X-ray-induced debromination was reinvestigated [Ennifar *et al.* (2002), *Acta Cryst.* **D58**, 1262-1268]. It was found that if the data are kept unmerged and if a dose-stamp is associated with each reflection measurement, dose-dependent occupancies can be refined for the Br atoms. Such a parametrization has been implemented in the macromolecular phasing program *SHARP*. Refining such dose-dependent occupancies on an unmerged data set gave a dramatic improvement, even for SAD phases from only the first wavelength (peak), and resulted in a good electron-density map after solvent flattening. The adverse effect of radiation damage has been turned into a beneficial one. The crucial difference is made by the use of unmerged data: phasing power is generated through the intensity differences of symmetry-related reflections recorded at different doses, *i.e.* corresponding to different states of the X-ray-induced debromination. This approach should prove useful in all situations of experimental phasing where site-specific radiation damage occurs unavoidably and undesirably and not only in cases in which radiation damage is purposely being created in order to demonstrate its potential usefulness.

RESEARCH PAPERS

Acta Cryst. (2008). **D64**, 711-729
<https://doi.org/10.1107/S0907444908010202>



Exploiting the anisotropy of anomalous scattering boosts the phasing power of SAD and MAD experiments

M. Schiltz and G. Bricogne

The X-ray polarization anisotropy of anomalous scattering in crystals of brominated nucleic acids and selenated proteins is shown to have significant effects on the diffraction data collected at an absorption edge. For conventionally collected single- or multi-wavelength anomalous diffraction data, the main manifestation of the anisotropy of anomalous scattering is the breakage of the equivalence between symmetry-related reflections, inducing intensity differences between them that can be exploited to yield extra phase information in the structure-solution process. A new formalism for describing the anisotropy of anomalous scattering which allows these effects to be incorporated into the general scheme of experimental phasing methods using an extended Harker construction is introduced. This requires a paradigm shift in the data-processing strategy, since the usual separation of the data-merging and phasing steps is abandoned. The data are kept unmerged down to the Harker construction, where the symmetry-breaking is explicitly modelled and becomes a source of supplementary phase information. These ideas have been implemented in the phasing program *SHARP*. Refinements using actual data show that exploitation of the anisotropy of anomalous scattering can deliver substantial extra phasing power compared with conventional approaches using the same raw data. Examples are given that show improvements in the phases which are typically of the same order of magnitude as those obtained in a conventional approach by adding a second-wavelength data set to a SAD experiment. It is argued that such gains, which come essentially for free, *i.e.* without the collection of new data, are highly significant, since radiation damage can frequently preclude the collection of a second-wavelength data set. Finally, further developments in synchrotron instrumentation and in the design of data-collection strategies that could help to maximize these gains are outlined.

RESEARCH PAPERS

Acta Cryst. (2010). **D66**, 447-457
<https://doi.org/10.1107/S0907444909053578>

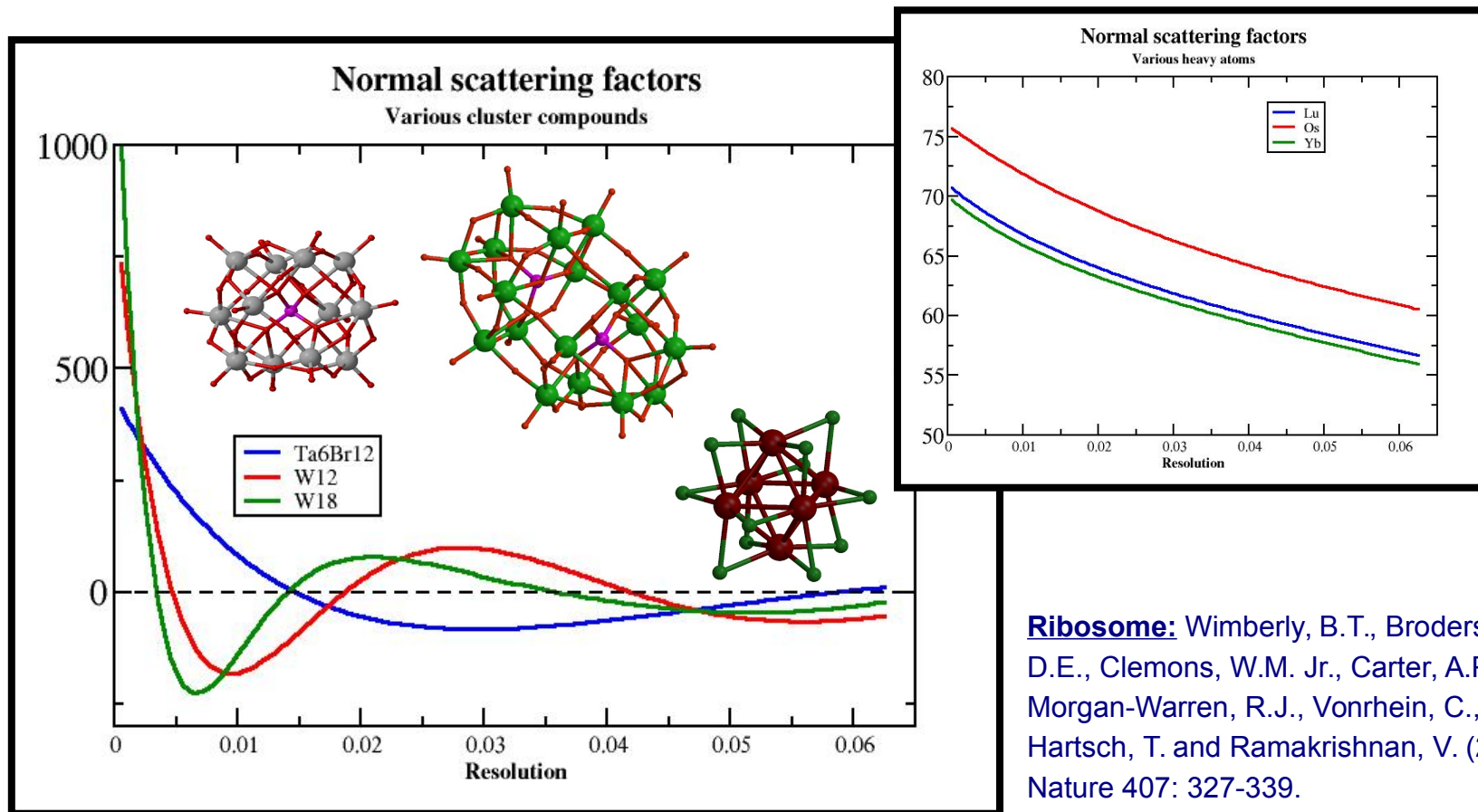


'Broken symmetries' in macromolecular crystallography: phasing from unmerged data

M. Schiltz and G. Bricogne

The space-group symmetry of a crystal structure imposes a point-group symmetry on its diffraction pattern, giving rise to so-called symmetry-equivalent reflections. Instances in macromolecular crystallography are discussed in which the symmetry in reciprocal space is broken, *i.e.* where symmetry-related reflections are no longer equivalent. Such a situation occurs when the sample suffers from site-specific radiation damage during the X-ray measurements. Another example of broken symmetry arises from the polarization anisotropy of anomalous scattering. In these cases, the genuine intensity differences between symmetry-related reflections can be exploited to yield phase information in the structure-solution process. In this approach, the usual separation of the data merging and phasing steps is abandoned. The data are kept unmerged down to the Harker construction, where the symmetry-breaking effects are explicitly modelled and refined and become a source of supplementary phase information.

Spherical averaged cluster compounds



Ribosome: Wimberly, B.T., Brodersen, D.E., Clemons, W.M. Jr., Carter, A.P., Morgan-Warren, R.J., Vornrhein, C., Hartsch, T. and Ramakrishnan, V. (2000). Nature 407: 327-339.

SHARP: special **SPHCLUSTER** keyword – also in autoSHARP

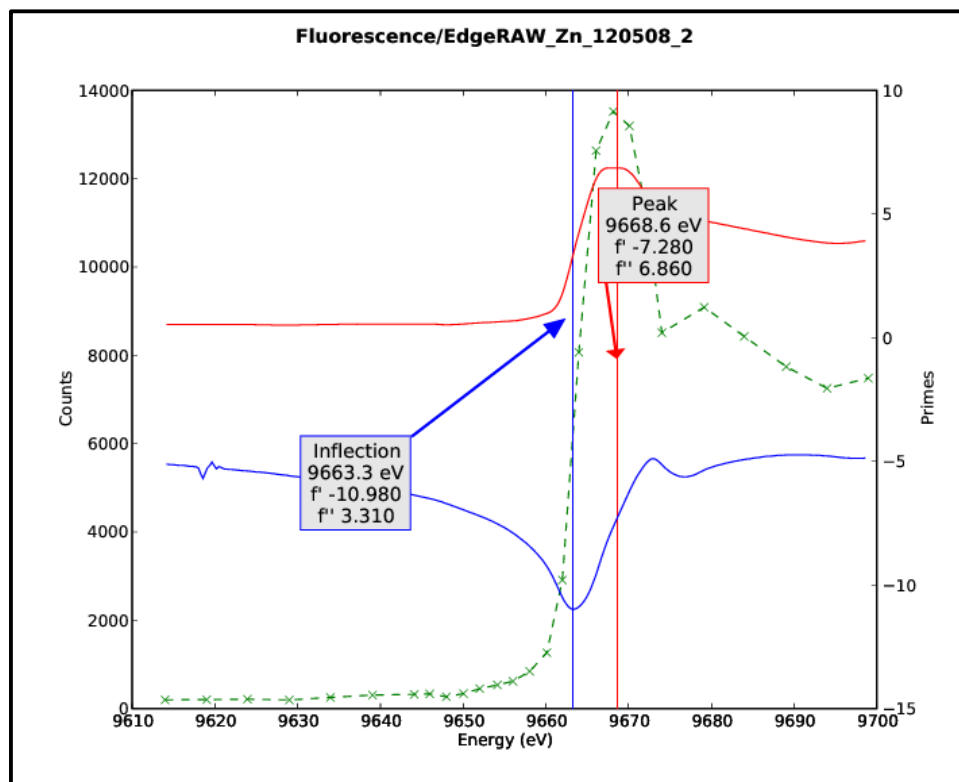
SAD or 2-wvl MAD?

A. González (2003). Optimizing data collection for structure determination. Acta Cryst. D59, 1935-1942

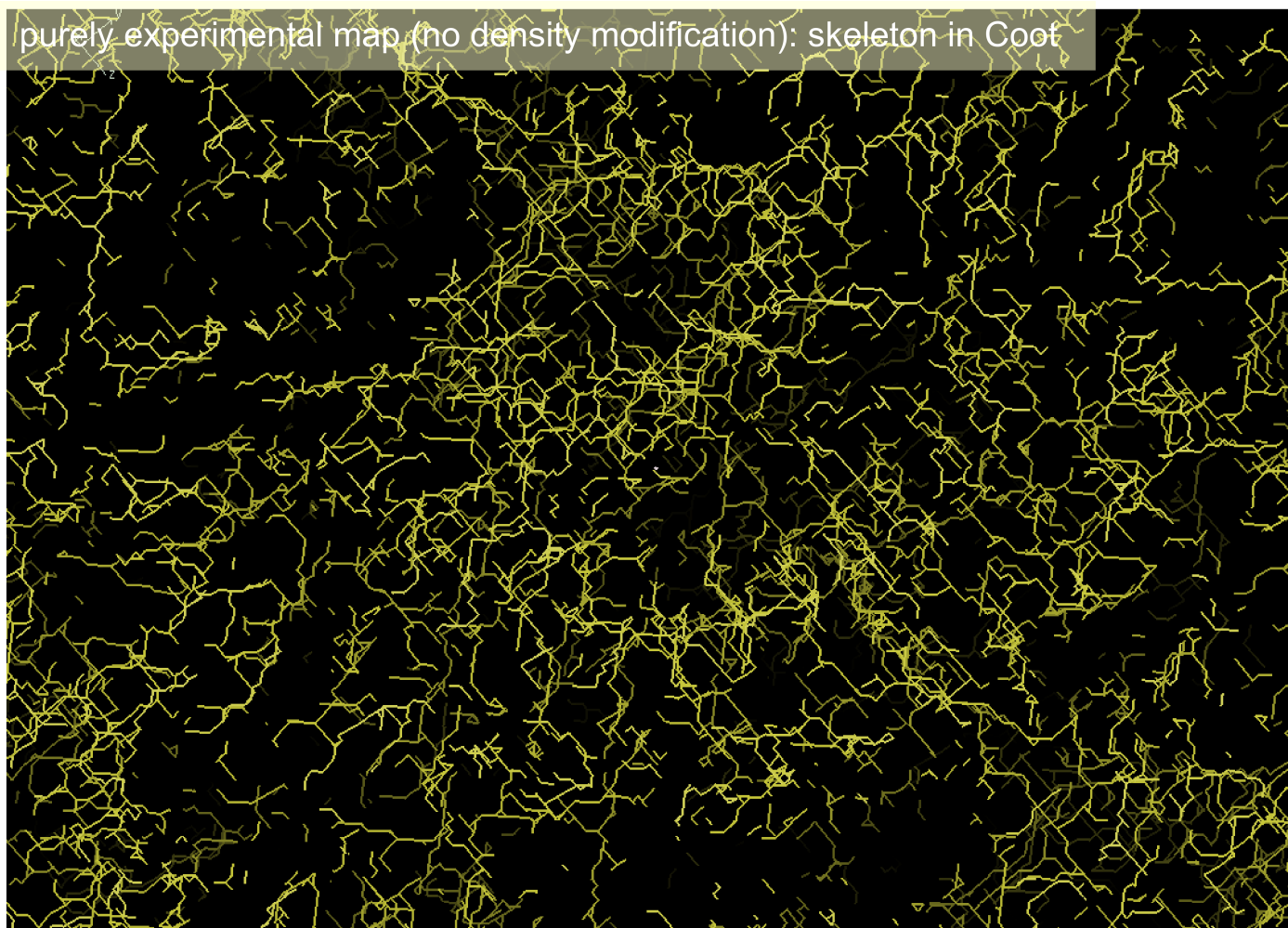
- ❑ SLS (M. Wang, V. Olieric), Soleil (A. Thompson, P. Legrand), May 2012

- ❑ Zn-containing protein
- ❑ tetragonal SG

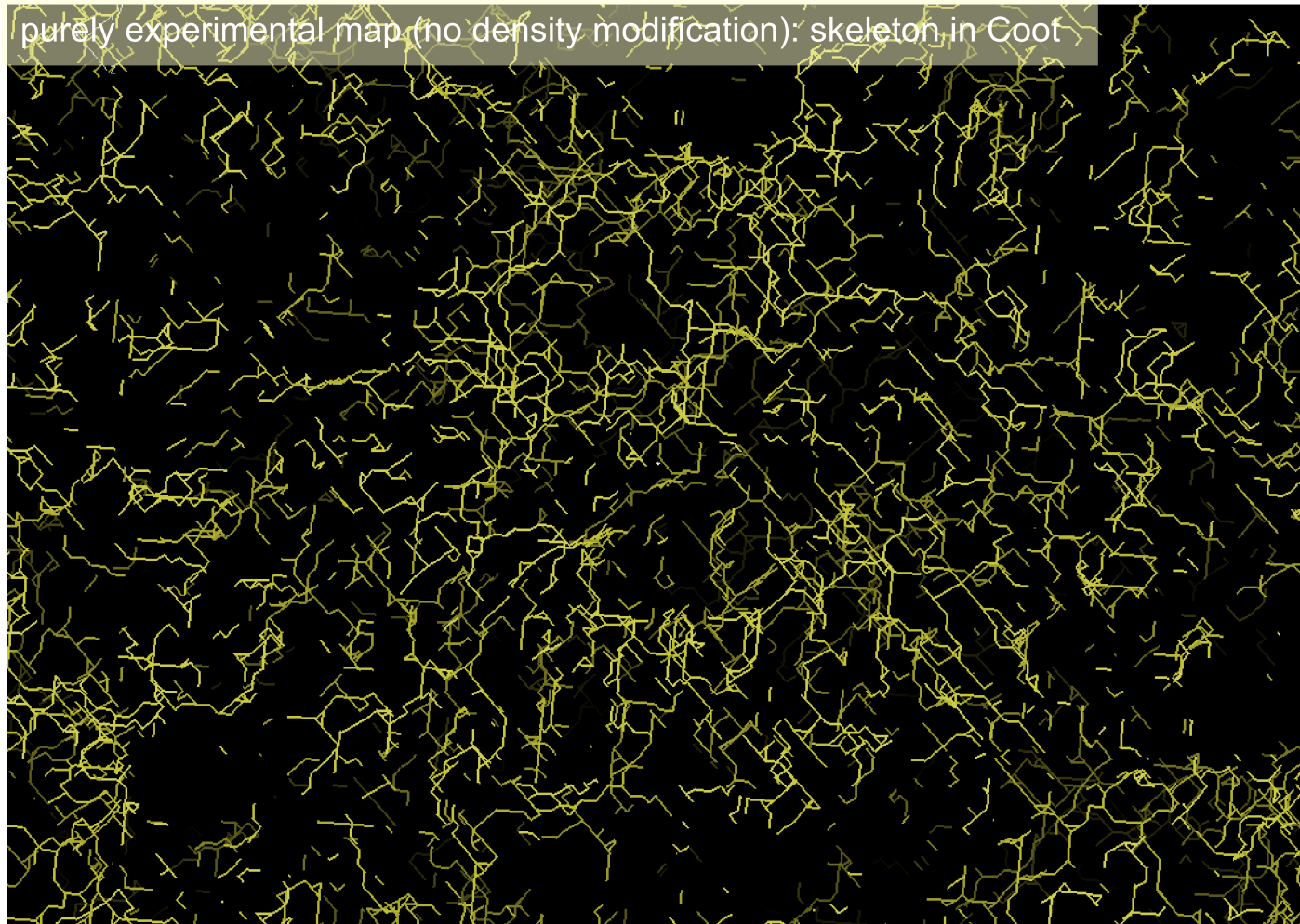
- ❑ Data collected at very low dose (Pilatus 2M)



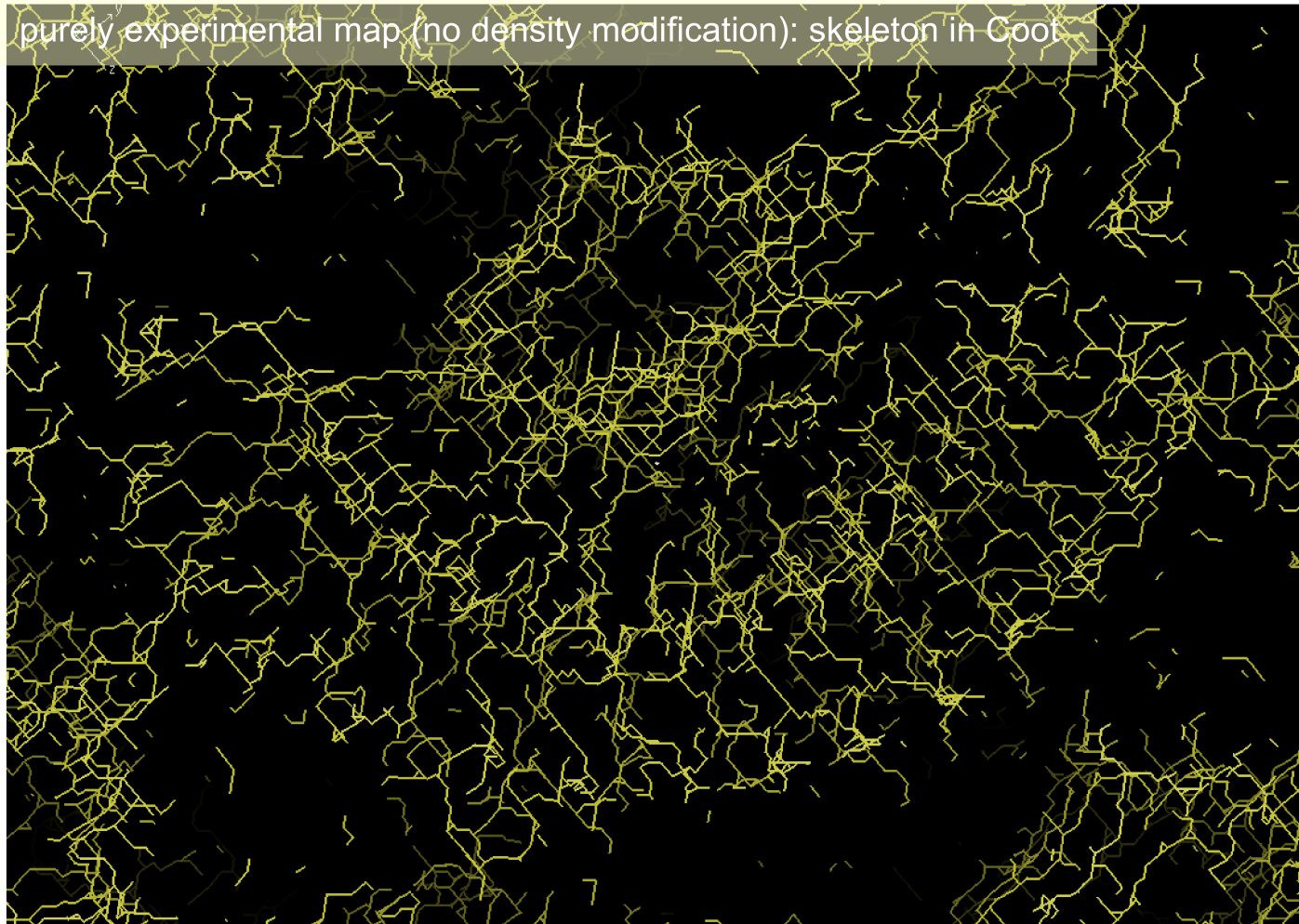
SAD (PEAK, 3600 images, multiplicity 25)



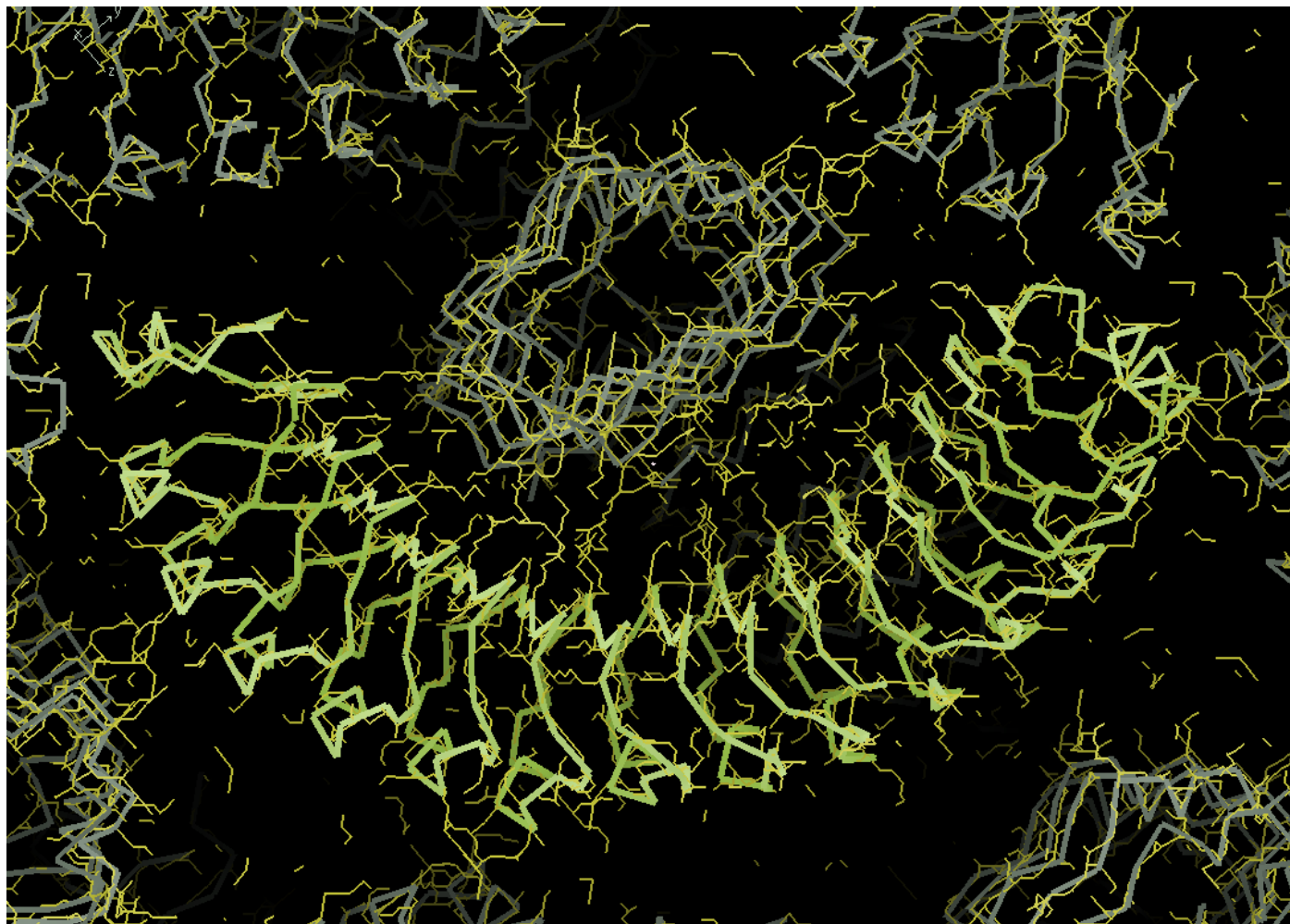
SAD, inverse-beam (1800+1800 images, multiplicity = 25)



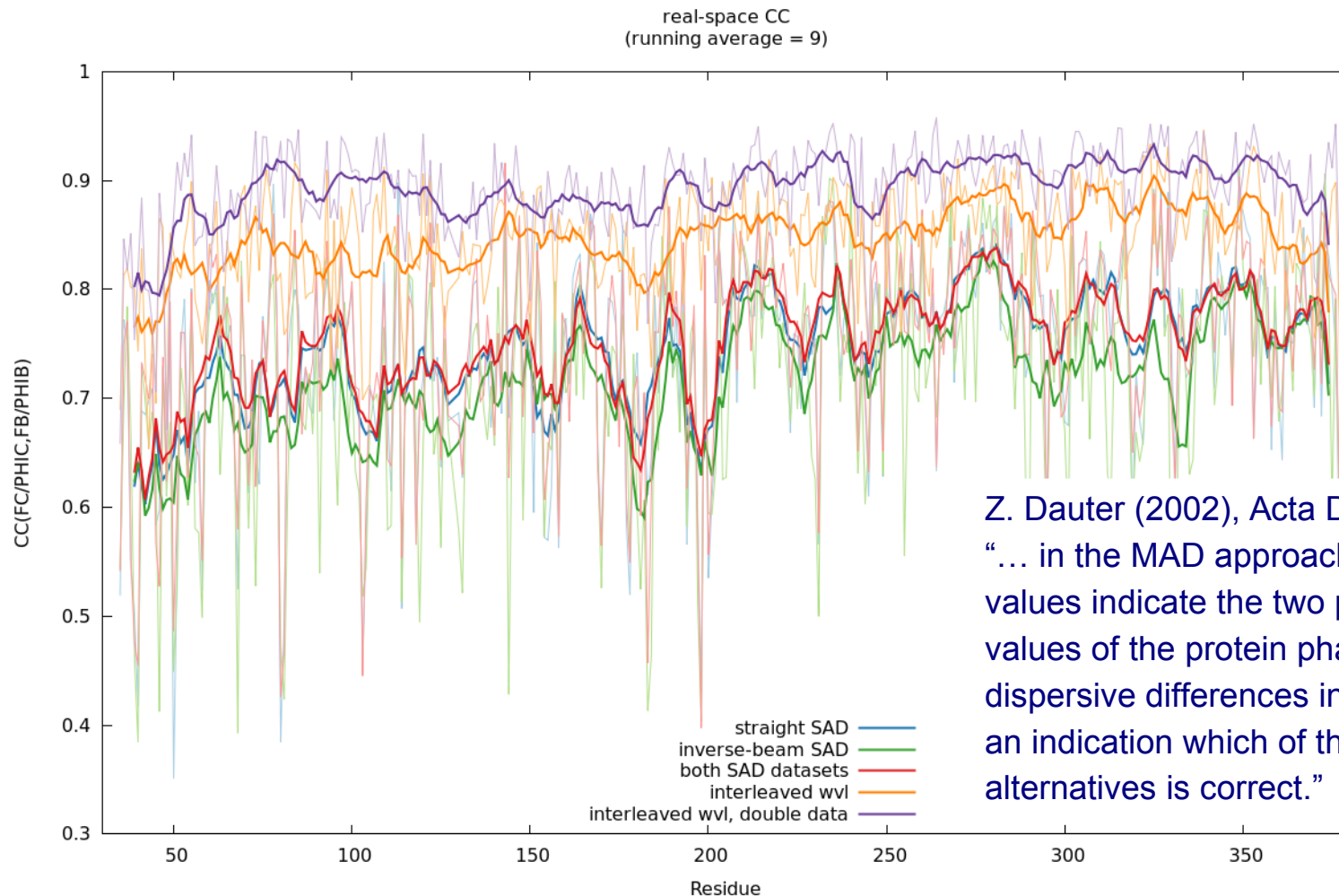
Double-inflection, interleaved-wvl (1800+1800 images, multiplicity = 12.5 + 12.5)



MAD, unaligned crystal, 2 x 180 degree (interleaved-wvl)



Phase ambiguity - SAD-vs-MAD



Z. Dauter (2002), Acta D58, 1958:
“... in the MAD approach the f''
values indicate the two possible
values of the protein phase and the
dispersive differences in f' provide
an indication which of the two
alternatives is correct.”

Session example datasets

❑ 2vb1

- High resolution
- P1 Lysozyme

❑ 2qvo

- Medium resolution
- S-SAD phasing

❑ 3csl

- Low resolution
- Se-MET
- Difficult spots

❑ 1y13

- Medium resolution
- 2-wvl MAD (interleaved wavelengths)
- Small phasing signal

<http://www.globalphasing.com/autoproc/wiki/index.cgi?ACA2011Tutorial>

<http://www.globalphasing.com/sharp/wiki/index.cgi?ACA2011Tutorial>

Session example datasets

❑ 2vb1

- High resolution
- P1 Lysozyme

❑ 2qvo

- Medium resolution
- S-SAD phasing

❑ 3csl

- Low resolution
- Se-MET
- Difficult spots

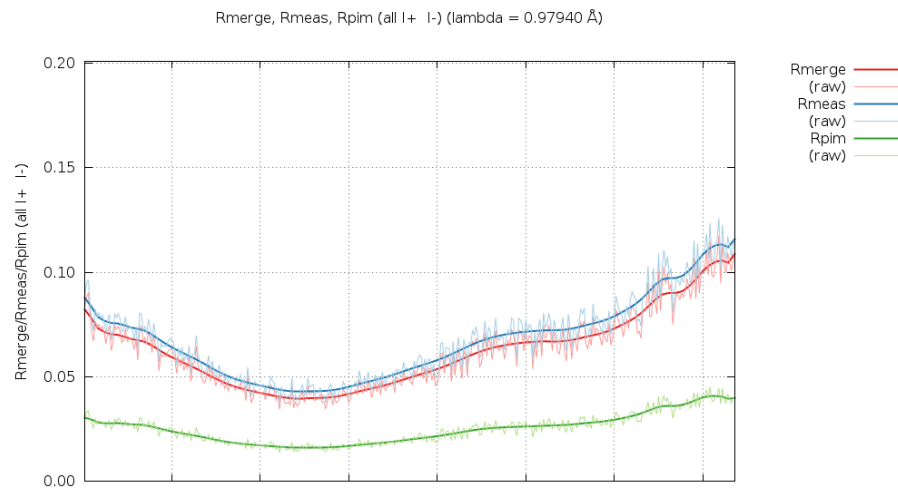
❑ 1y13

- Medium resolution
- 2-wvl MAD (interleaved wavelengths)
- Small phasing signal

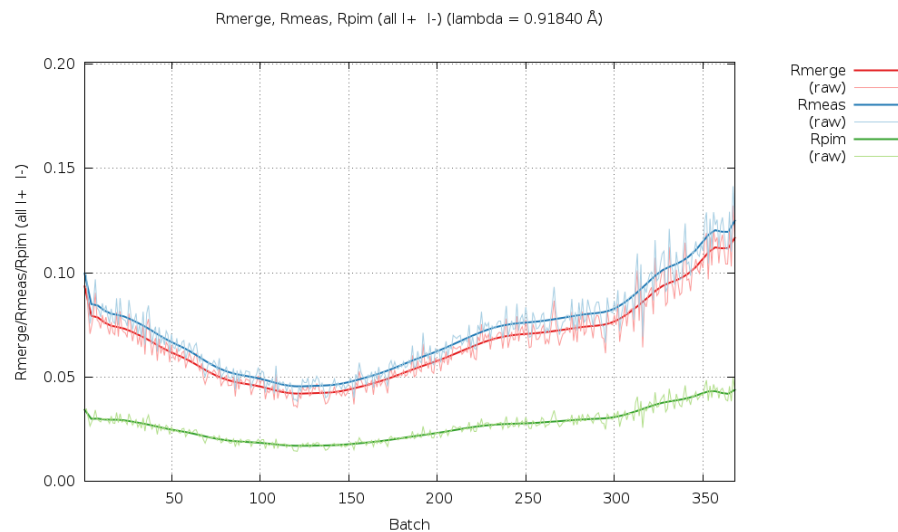
<http://www.globalphasing.com/autoproc/wiki/index.cgi?ACA2011Tutorial>

<http://www.globalphasing.com/sharp/wiki/index.cgi?ACA2011Tutorial>

infl



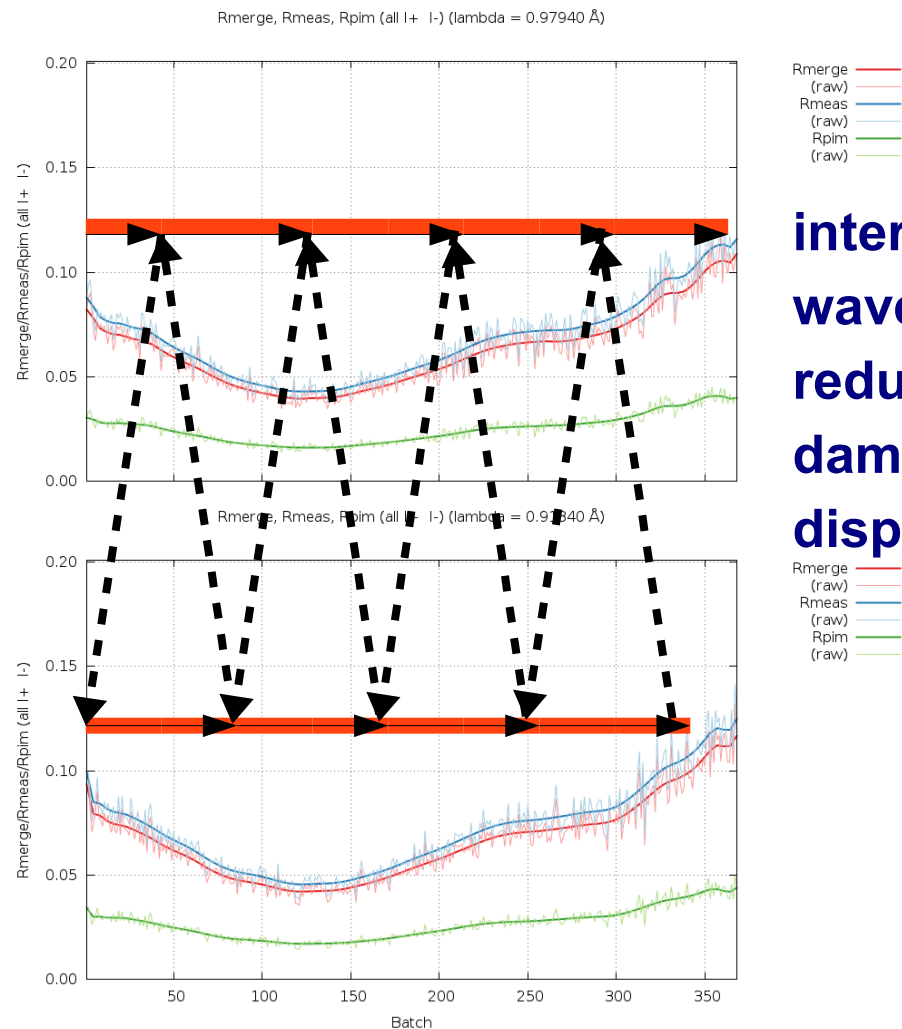
hrem



Radiation damage?

infl

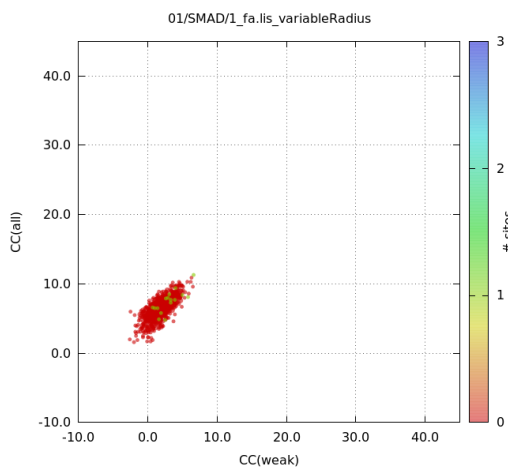
hrem



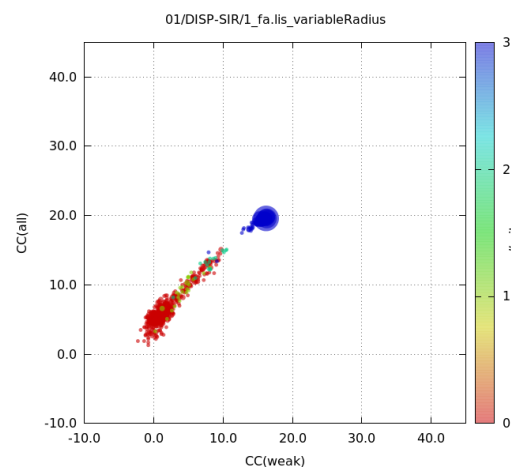
**interleaved
wavelengths (to
reduce radiation
damage effect on
dispersive differences)**

1Y13 – anomalous/dispersive differences

Anomalous 2.8Å

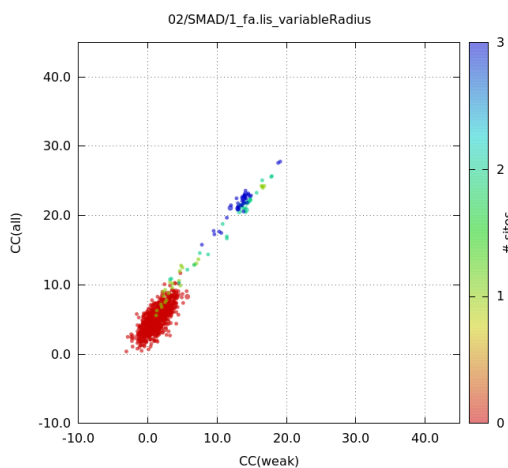


Dispersive 2.8Å

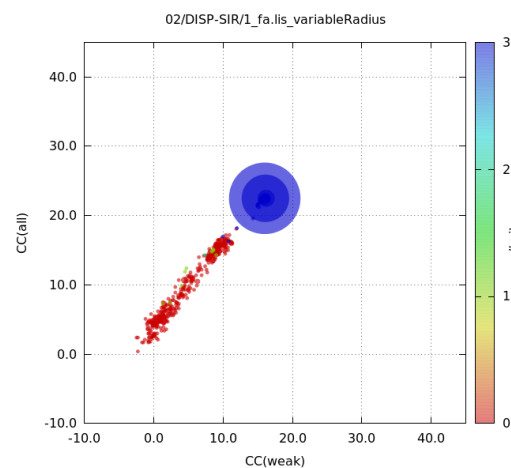


No Odose

Anomalous 2.8Å



Dispersive 2.8Å

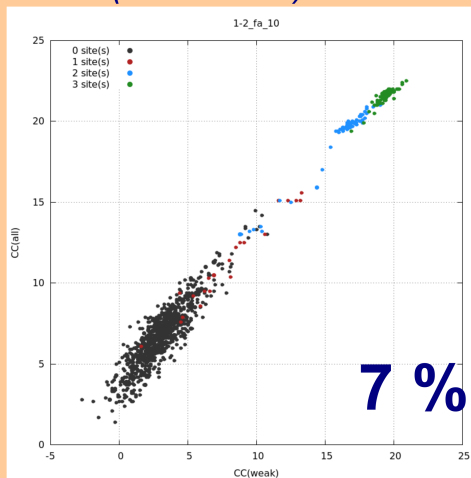


Odose

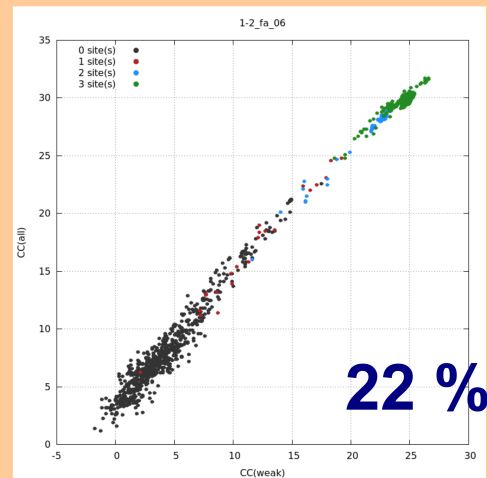
1Y13 - SHELXD (MAD) success rate

AIMLESS data (not shown): 13%

XSCALE
no 0-dose

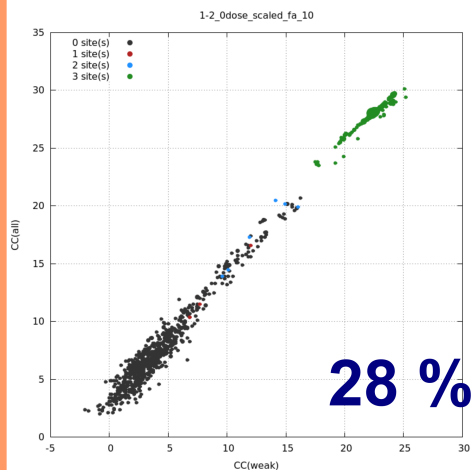


XSCALE
0-dose

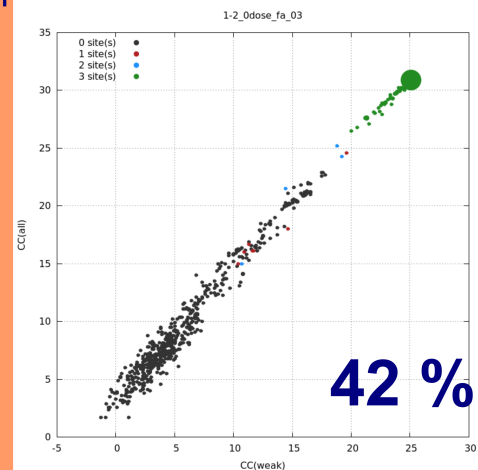


single 0dose on combined datasets with correct dose-stamp:

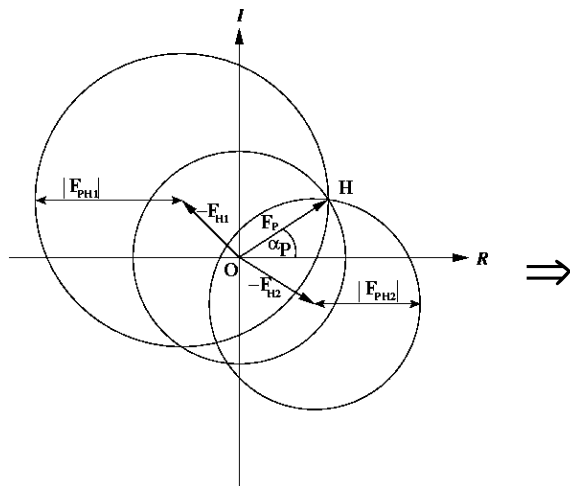
XSCALE*
0-dose



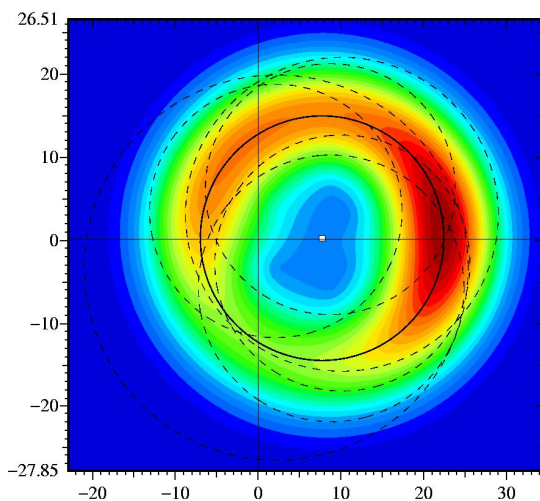
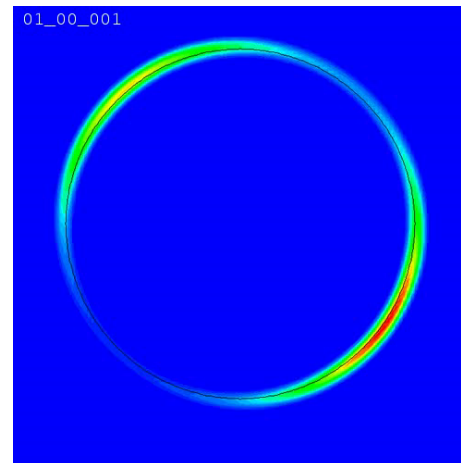
AIMLESS-XSCALE*
0-dose



Phase probability



2D phase probability



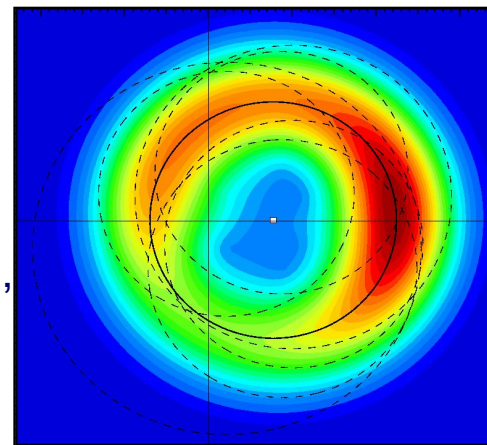
real life more complicated, since angular distribution (HLs) related to:

- radius R (and variance V)
- offset (x,y)

G. Bricogne, C. Vonrhein, C. Flensburg, M. Schiltz, and W. Paciorek (2003). *Acta Crystallogr* **D59**, 2023-30.

Summary

- Automation is great ... if it works
- Meaningful feedback important!
- Improving on initial results often involves going back to scaling and/or data processing and/or collection strategy.
- Think about **differences** (and how to measure, scale, analyse and describe them): anomalous, dispersive, isomorphous, radiation damage, MAD+native etc.
- It's all about **signal-over-noise**: HA signal, noise in data ... MIR still very useful ... so is MAD
- High multiplicity (good): maybe split between different wavelengths instead of staying with SAD
- Nomenclature can be confusing between programs
- Goal: best set of phases from well measured and processed data



2D phase probability in SHARP

Acknowledgement

Global Phasing, Cambridge (UK):

- G rard Bricogne, Claus Flensburg, Andrew Sharff, Wlodek Paczior, Peter Keller, Rasmus Fogh, Ian Tickle, Leigh Carter
- Previous group-members: M. Schiltz, E. Blanc, P. Roversi, G. Evans, R. Morris, T. Womack, Oliver Smart

<https://www.globalphasing.com/> (also: BUSTER, autoPROC)

<https://www.globalphasing.com/autoproc/wiki/>

<https://www.globalphasing.com/sharp/wiki/>


Tutorials, FAQ,
documentation,
examples, ...

ARP/WARP (Tassos Perrakis, Victor Lamzin, Richard Morris)

PARROT/BUCCANEER (Kevin Cowtan)

SHELXC/SHELXD: (George Sheldrick)

CCP4

MOSFLM (Andrew Leslie & Harry Powell)

SCALA/POINTLESS/AIMLESS (Phil Evans)

XDS (Wolfgang Kabsch & Kay Diederichs)

Test data: JCSG, SBGrid, Global Phasing Consortium members,
SLS/PSI, Soleil, DLS, ESRF, IMCA-CAT, SSRL, APS, ALS, PF, LNLS, NSLS,
... many others and users