
Experimental Phasing with SHARP/autoSHARP

Clemens Vonnrhein
Global Phasing Ltd.

CCP4/DLS workshop
05.12.2018

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

autoSHARP:

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

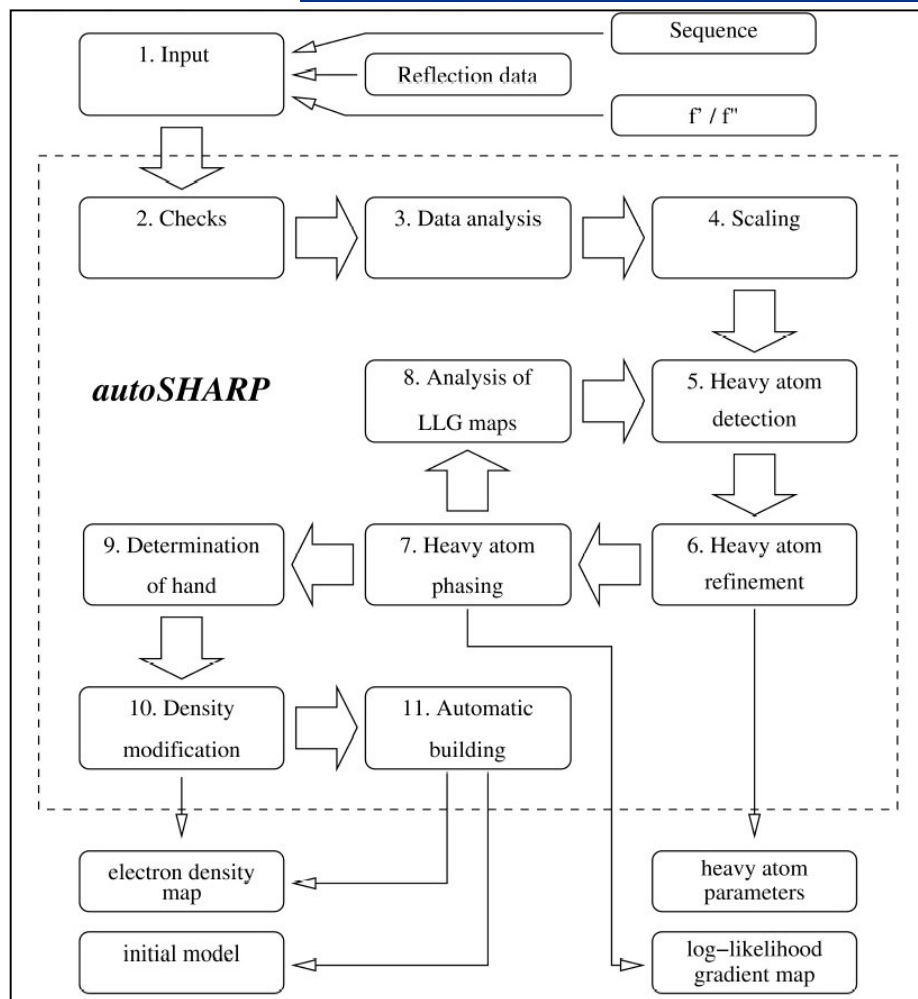
Vonrhein, C, Blanc, E., Roversi, P and Bricogne, G. (2007). Automated structure solution with autoSHARP. *Methods Mol Biol* 364, 215-30.

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., Vonrhein, 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. D* 59, 2023-2030.

autoSHARP “pipeline”



minimum amount of input

feedback:

internally and to the user

several levels of results

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 shows the CCP4i autoSHARP input interface. 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 divided into several sections:

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

At the bottom, there are buttons for Run, Save or Restore, and Close, along with an Edit list button and an Add another dataset button.

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
```

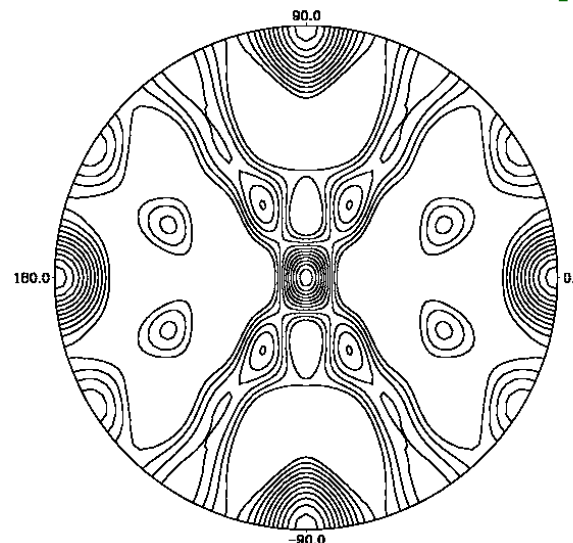
autoSHARP - Analysis

- 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)
- self-rotation function
- native Patterson
- Matthews parameters (probability, Rupp)

?/ASU

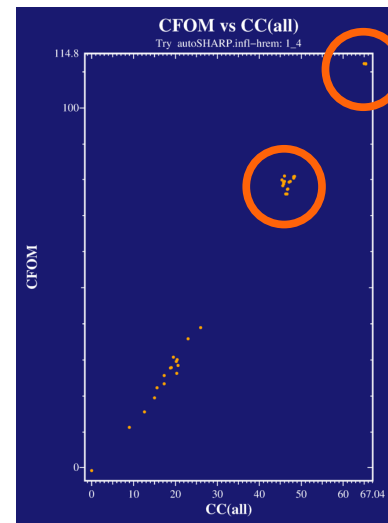
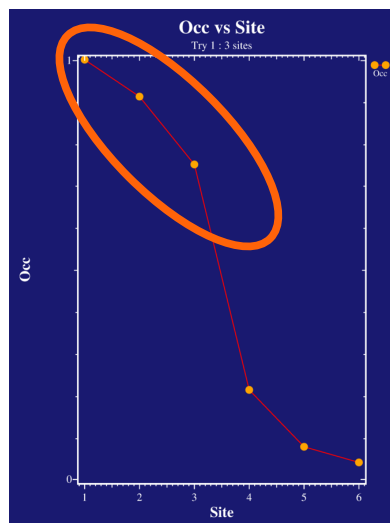
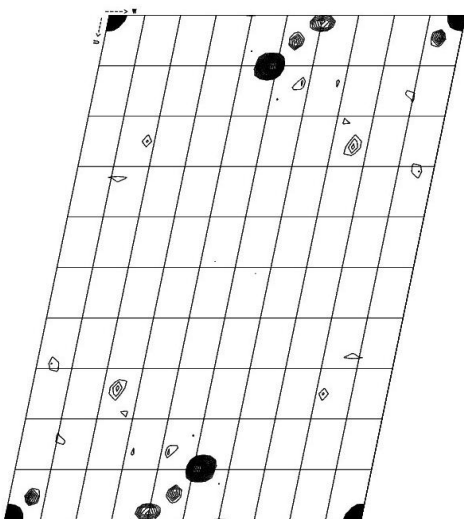
SIGNAL

NOISE



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

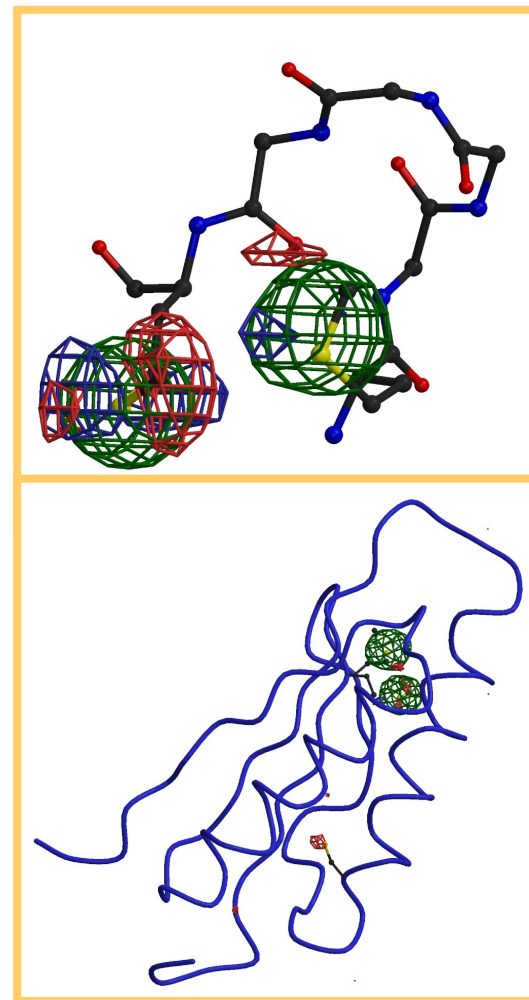
- HA refinement and phasing
- overall anisotropic scaling
- analysis of LLG (residual) maps:

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**

- phases in two hands (enantiomorphs)

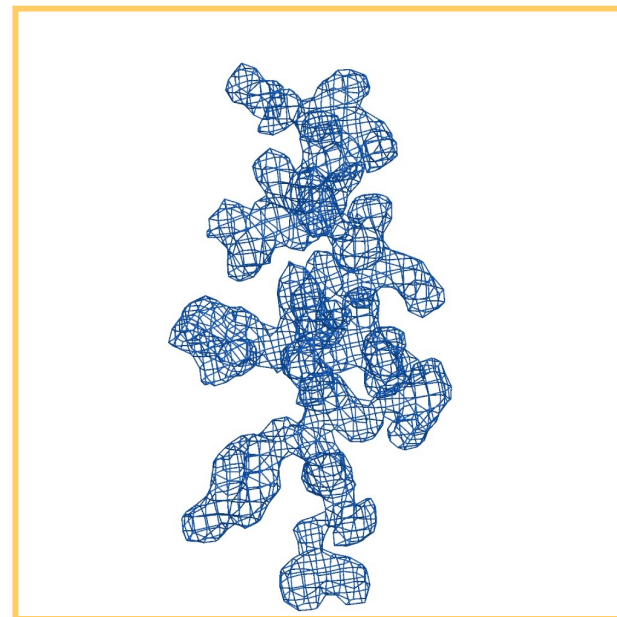
**Best
possible
phases**



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 correct no. of molecules

Similar to MR (test enantiomorphs)



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

autoSHARP – automatic building

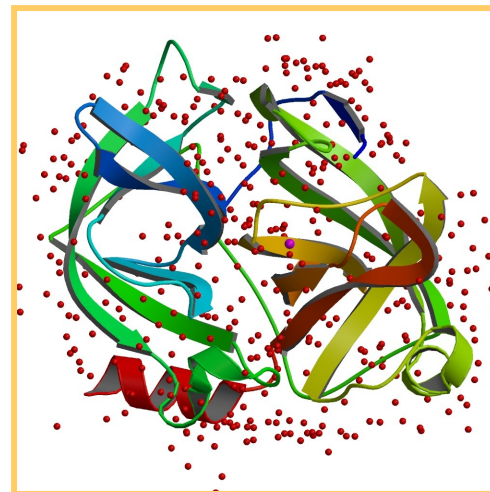
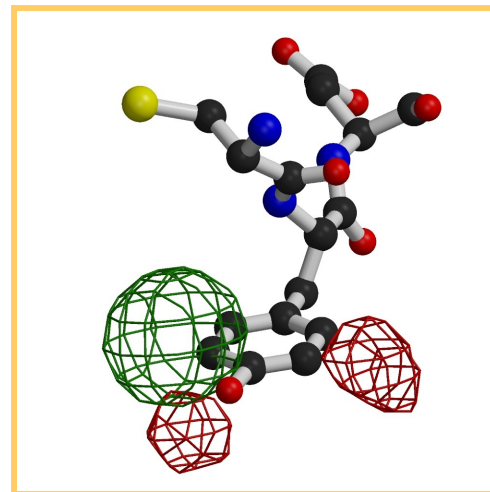
- **ARP/wARP** ¹⁾ suite of programs (latest ARP/wARP as distributed via CCP4)
- **PARROT/BUCCANEER** ²⁾ iterative density-modification and automatic building tool (“Long John Silver”:
long_john_silver.sh -h)
- Results from heavy atom phasing (SHARP) plus density modification (SOLOMON) can easily be feed 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
- **SHARP** phases
- LLG (residual) maps
- density modified phases/maps
- initial model



Tools around SHARP/autoSHARP

- 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**)
 - 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)

How to improve on first results

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

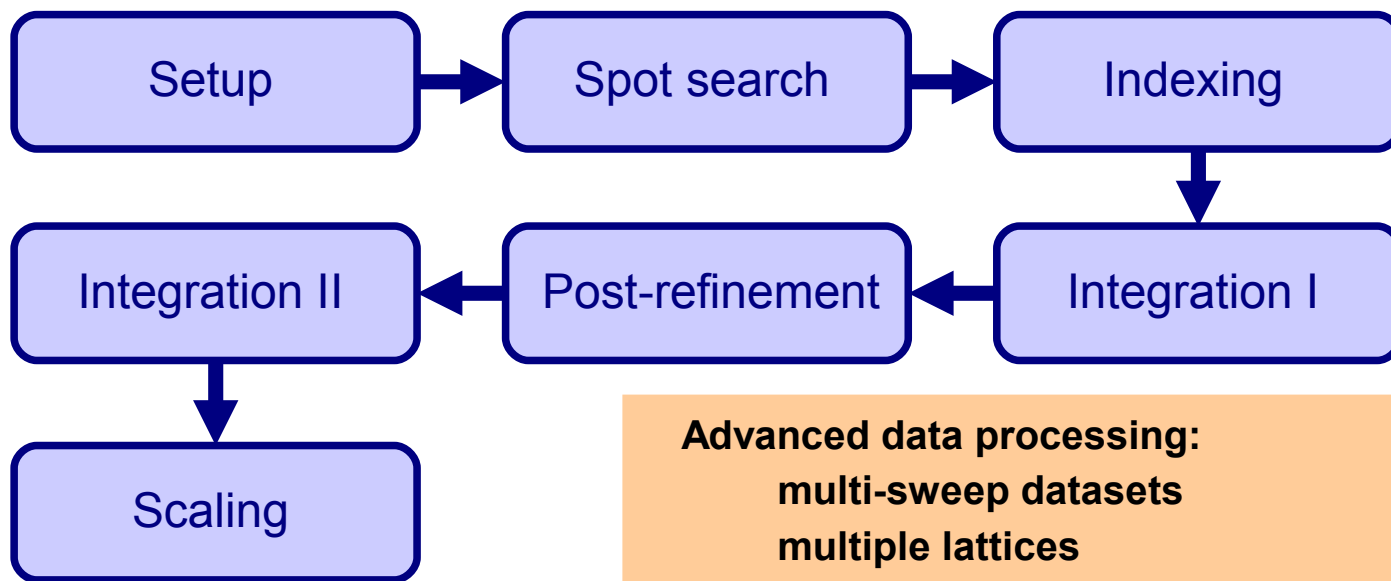
autoPROC - behind the scenes

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T. & Bricogne, G. (2011). Data processing and analysis with the **autoPROC** toolbox. *Acta Cryst. D67*, 293-302.

❑ Uses external programs

- **XDS/XSCALE**: Kabsch, W. XDS. *Acta Cryst. D66*, 125-132 (2010).
- **POINTLESS, AIMLESS**: Evans, P. (2006). Scaling and assessment of data quality. *Acta Cryst. D62*, 72-82.
- **CCP4**: Collaborative Computational Project, Number 4. 1994. "The CCP4 Suite: Programs for Protein Crystallography". *Acta Cryst. D50*, 760-763.

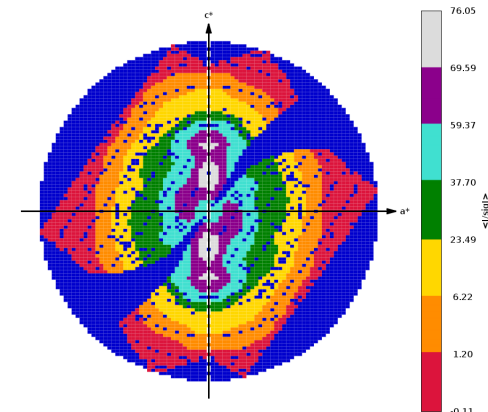
Typical steps for a single set of consecutive images:



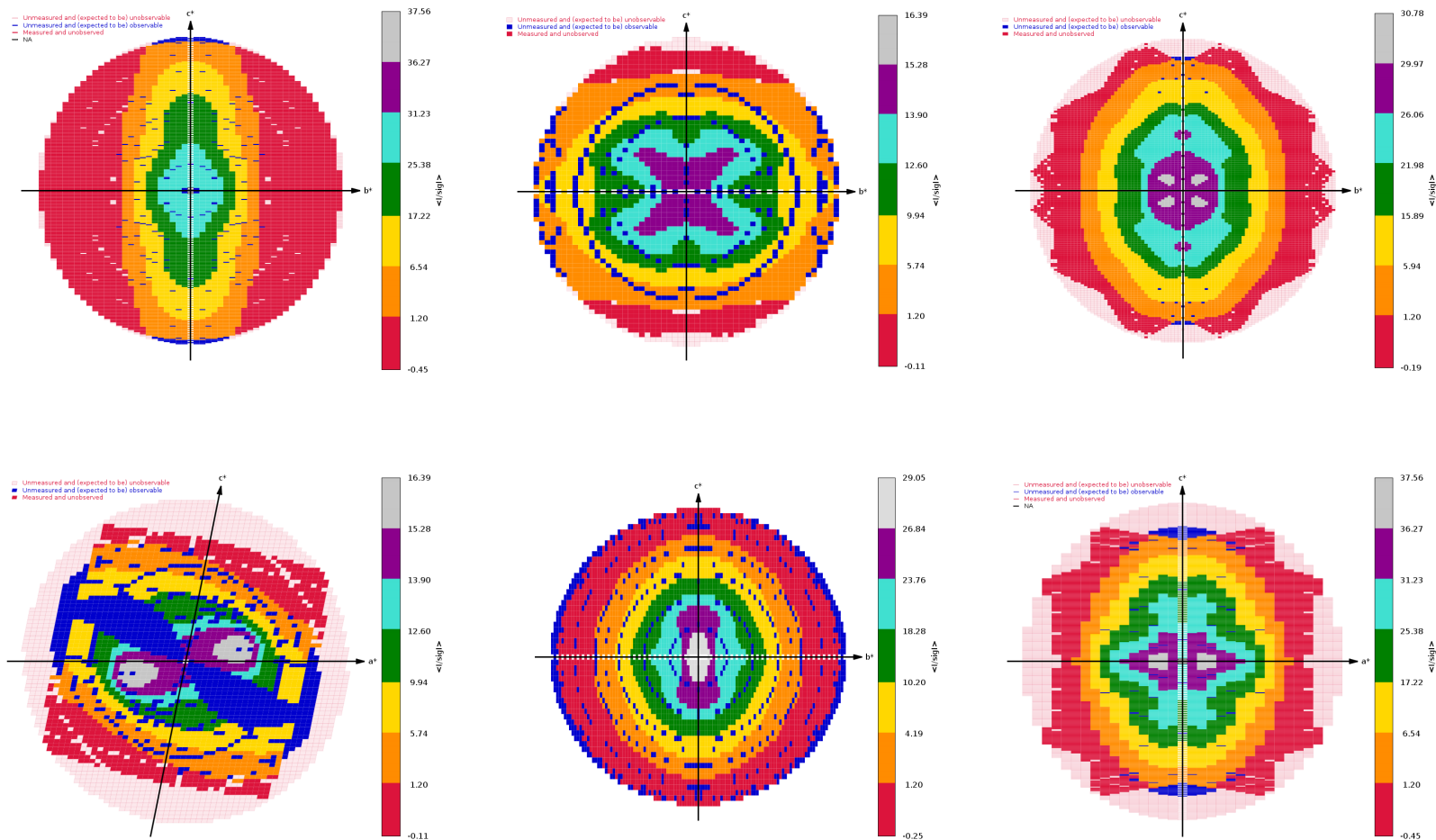
autoPROC (1) adds expert knowledge at and between each step and (2) combines separate sweeps/datasets

Advanced data processing:
multi-sweep datasets
multiple lattices
multi-axis goniostats
ice-rings
resolution limits, cut-offs etc
reporting (HTML, XML, PDF)
integration (ISPyB, Crims)

- ❑ main author: Ian Tickle (Global Phasing)
- ❑ **anisotropic** cut-off of merged intensity data
 - avoids problems of isotropic resolution cut-offs
 - throwing away good data (in one direction)
 - including poor data (in another direction)
- ❑ Bayesian estimation of structure factor amplitudes
 - more rigorous treatment of prior and anomalous data
 - for more details see:
http://staraniso.globalphasing.org/staraniso_about.html
- ❑ (optional) anisotropic correction of data
- ❑ Part of autoPROC ('process', 'aP_scale')
 - user is given classical, isotropic (TRUNCATE) output
 - but also data after **anisotropic** analysis (STARANISO)



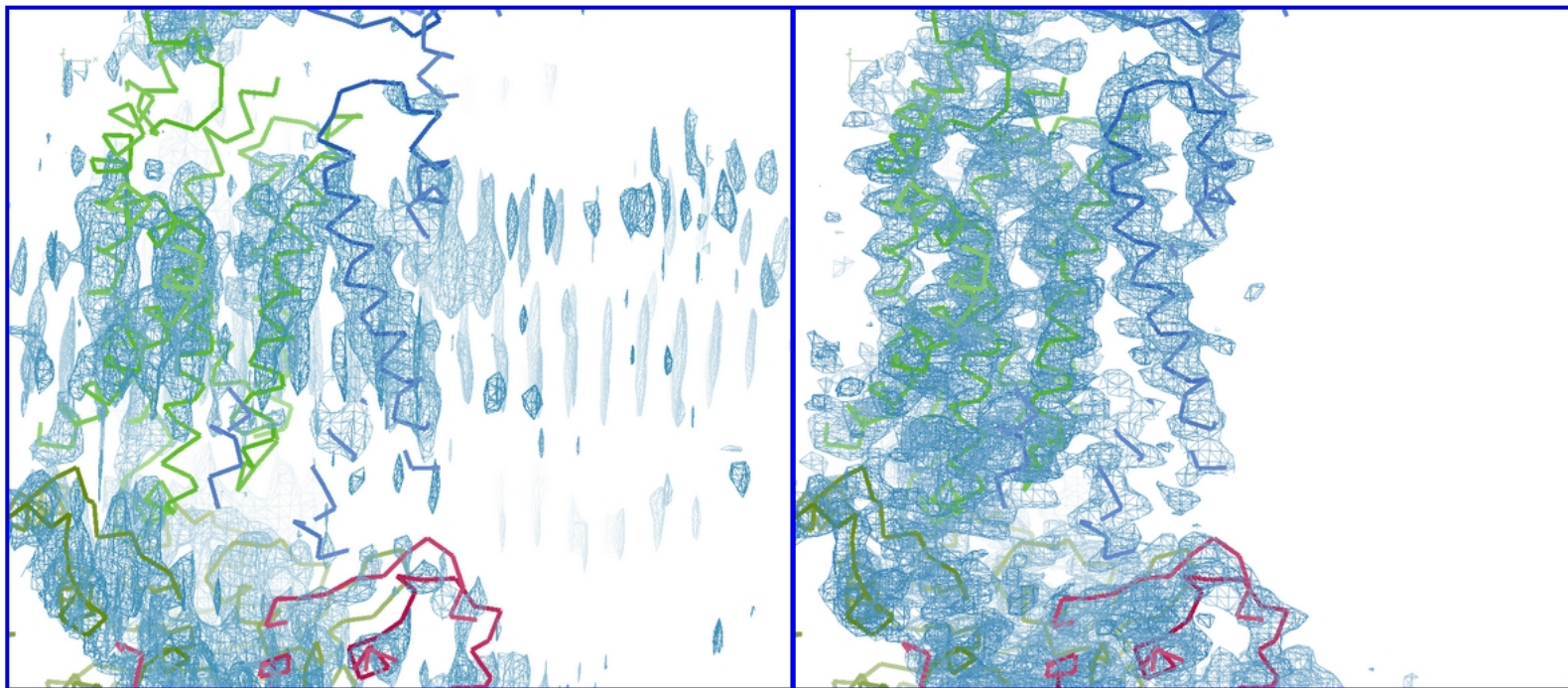
Not necessarily ellipsoidal!



- The following results were communicated to us by Neil Paterson at Diamond Light Source.

The panels show the maps after density modification (CCP4 DM program) using (left) the standard protocol with an isotropic resolution cut-off, and (right) the STARANISO protocol with anisotropic resolution cut-off and correction. Compare in particular the density in the more mobile loop regions at the ends of the helices, and also the flatness of the solvent region (right-hand half of each map).

Neil comments: "STARANISO has been key to solving our structure. We've got a severely anisotropic case that goes from around 7 Å along a^* to 3.0-3.5 Å along b^*/c^* , with lots of artifacts in the maps. We can find the substructure using Shel-X without correction but the maps are pretty bad; once corrected it works a treat (77% solvent!)."



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.

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 - 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 - 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

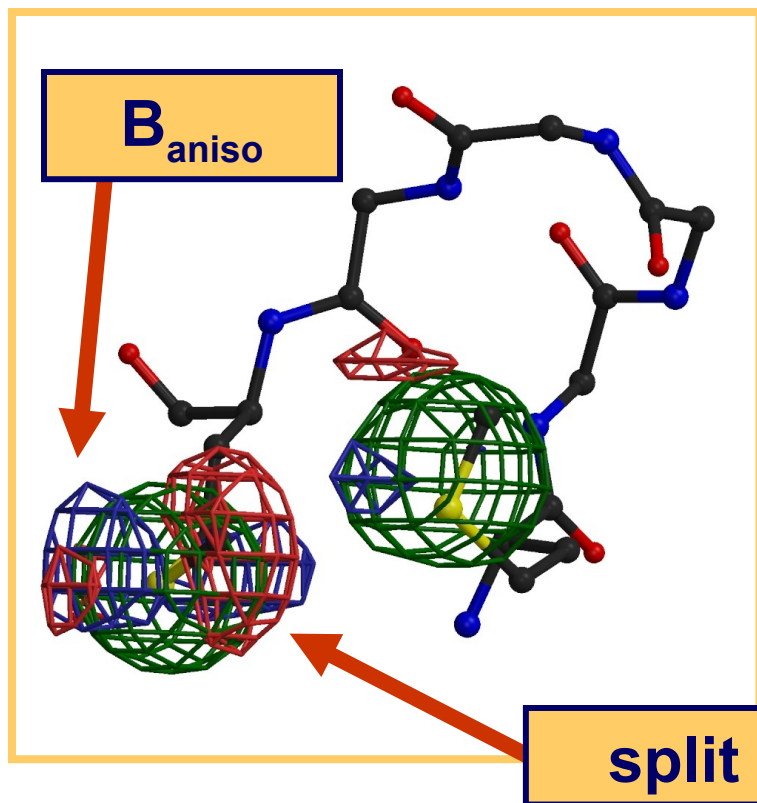
WARNING : We will remove 2 reflexions with a ABS(DANO)/FMID ratio > 1.9

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.

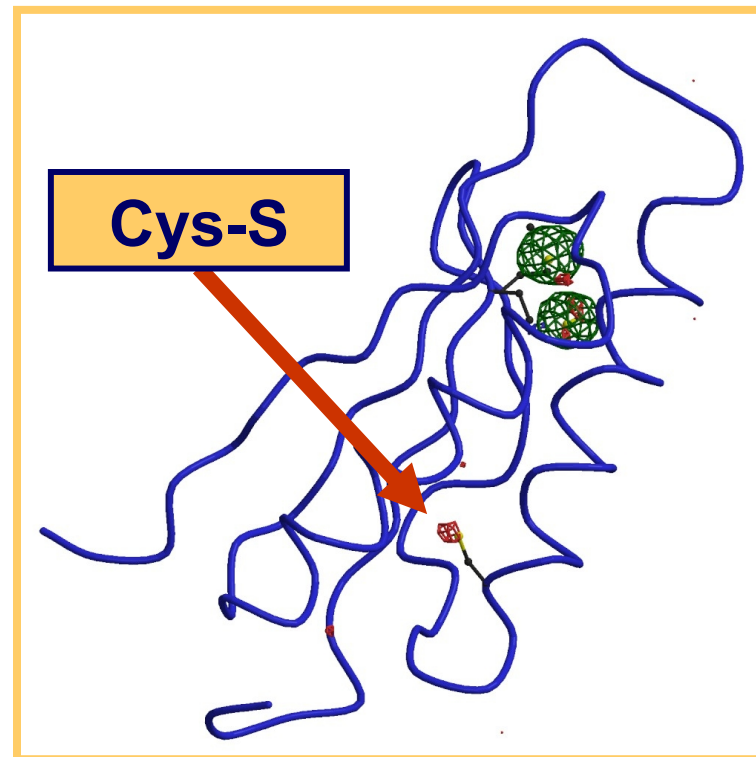
How to improve on first results

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

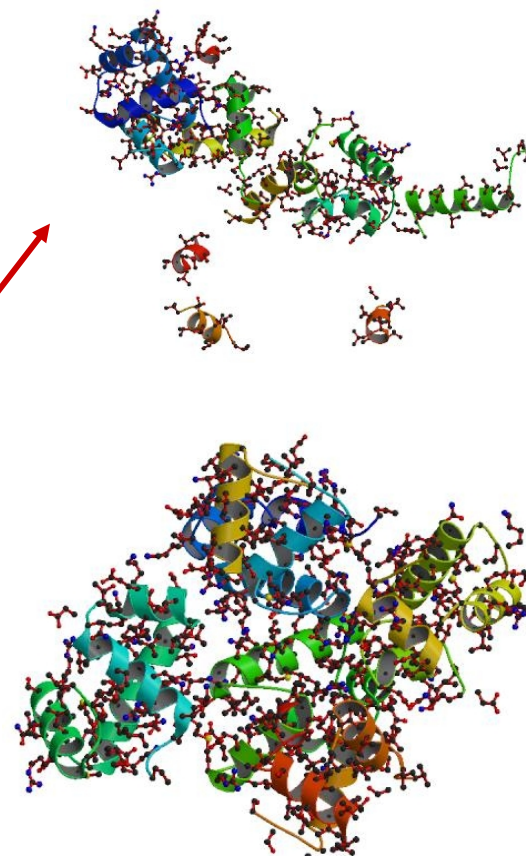
- after **autoSHARP**:
 - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)
 - analyse **LLG** (residual) maps: additional sites, f'/f'' refinement, anisotropic B-factor for sites, ...
 - significant difference between score of **original/inverted** 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 LLG residual maps → HA detection
 - Overcoming non-isomorphism issues

"MR-SAD"

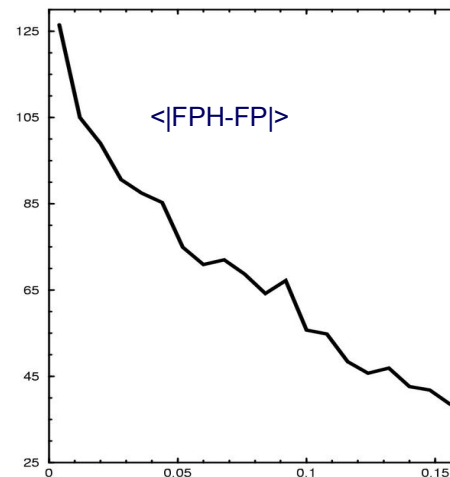
- **autoSHARP** can do:

"ExtPhas-SAD"	"ExtPhas-SIR"	"ExtPhas-MIRAS"
"ExtPhas-MAD"	"ExtPhas-MIR"	"ExtPhas-SIRAS"
- **SHARP** can also do:

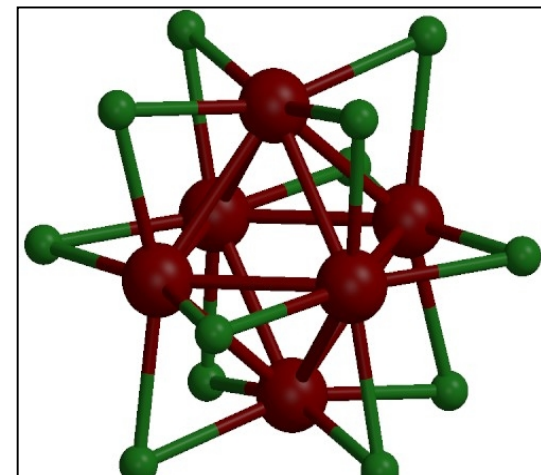
"ExtPhas-MAD-NATIVE"
"ExtPhas-XYZ"

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:
 $\langle |FPH-FP| \rangle$

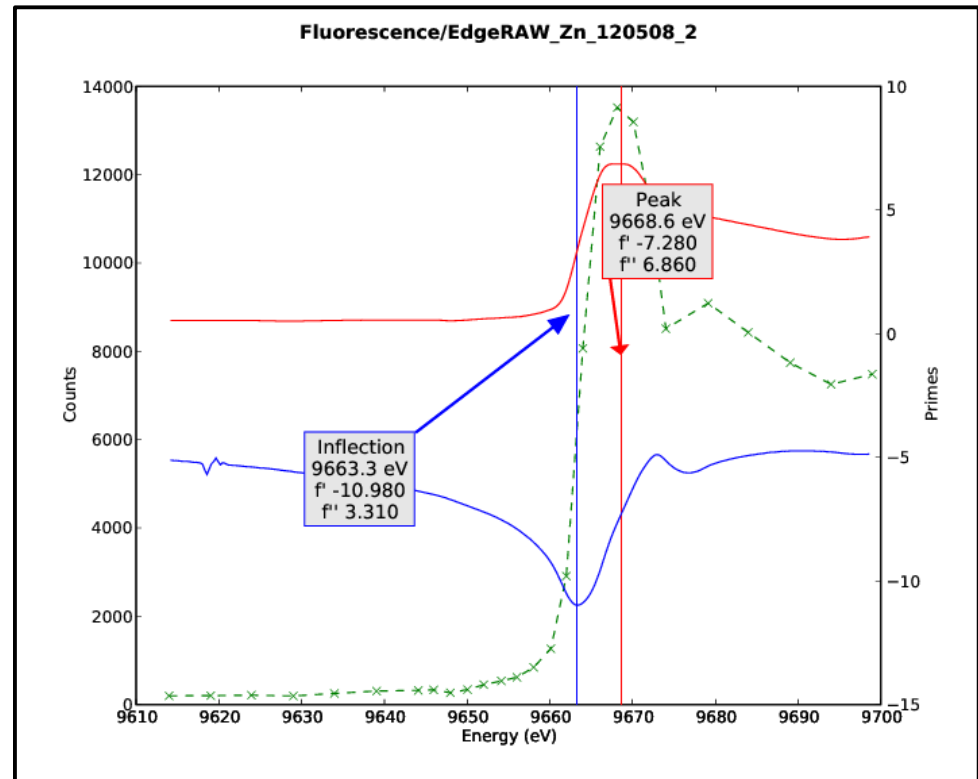


$\text{Ta}_6\text{Br}_{12}$

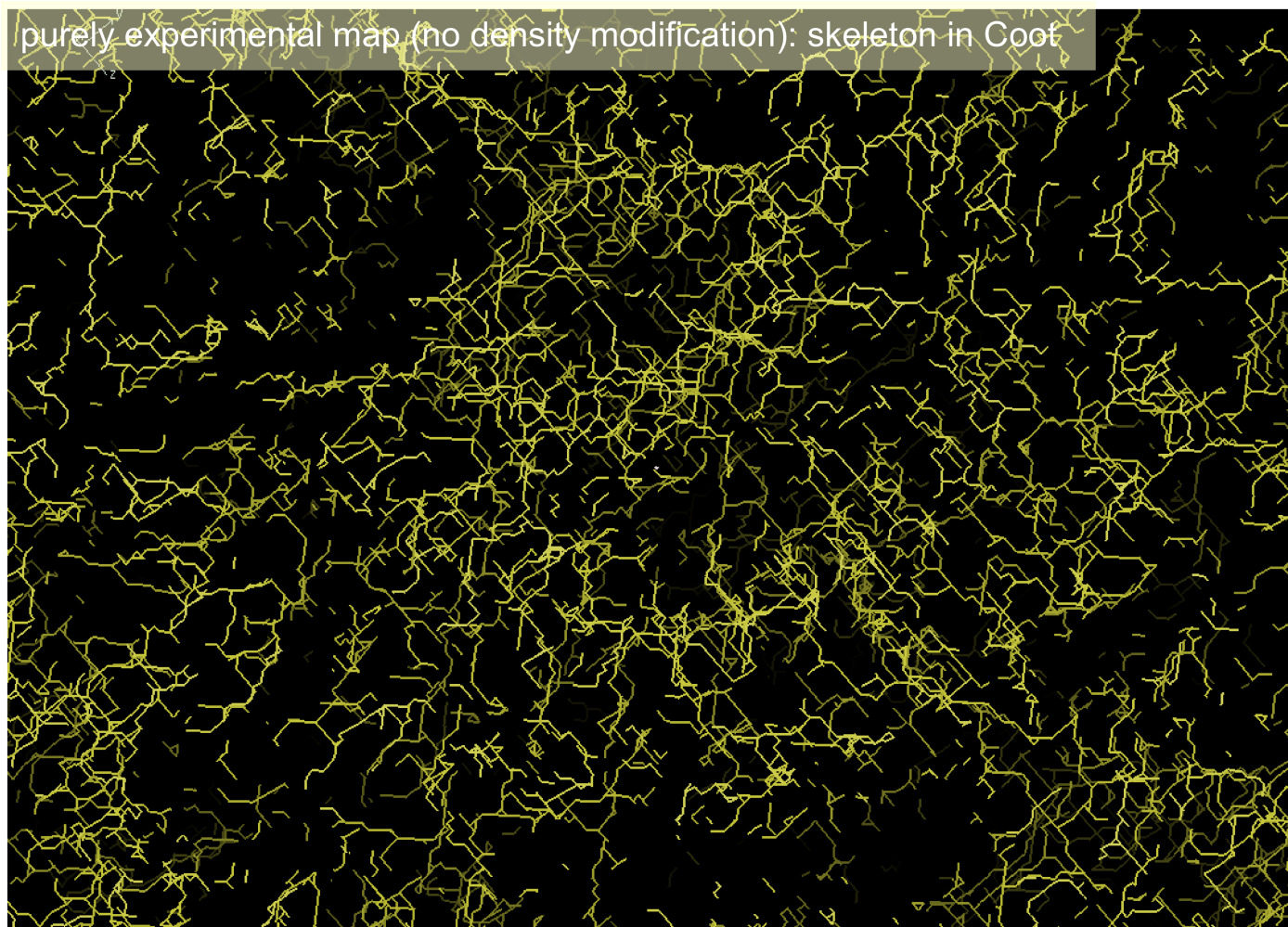
SAD or 2-wvl MAD?

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

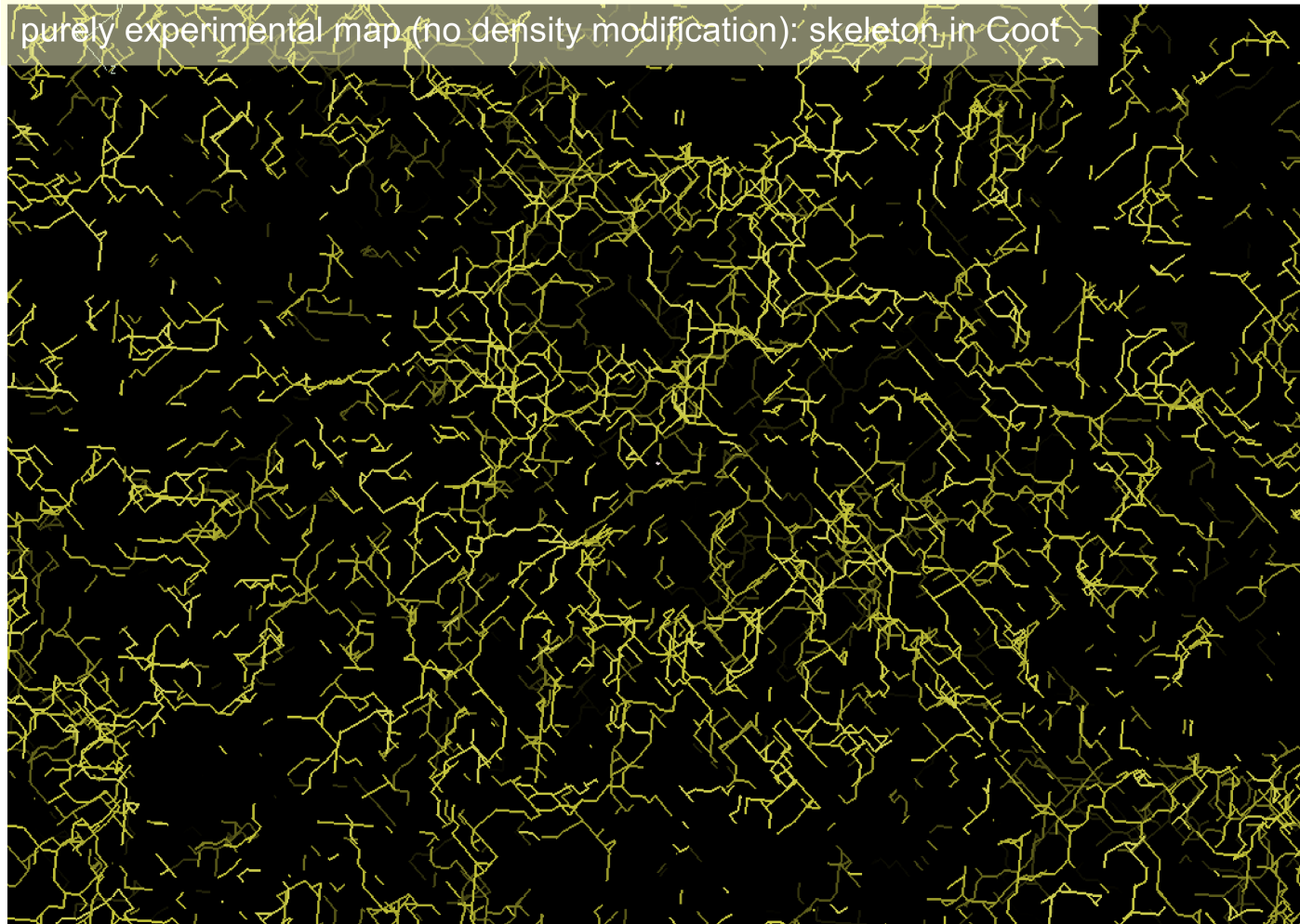
- ☐ SLS (M. Wang, V. Olieric), Soleil (A. Thompson, P. Legrand), May 2012
- ☐ Zn-containing protein
- ☐ tetragonal SG
- ☐ Data collected at very low dose (Pilatus 2M)



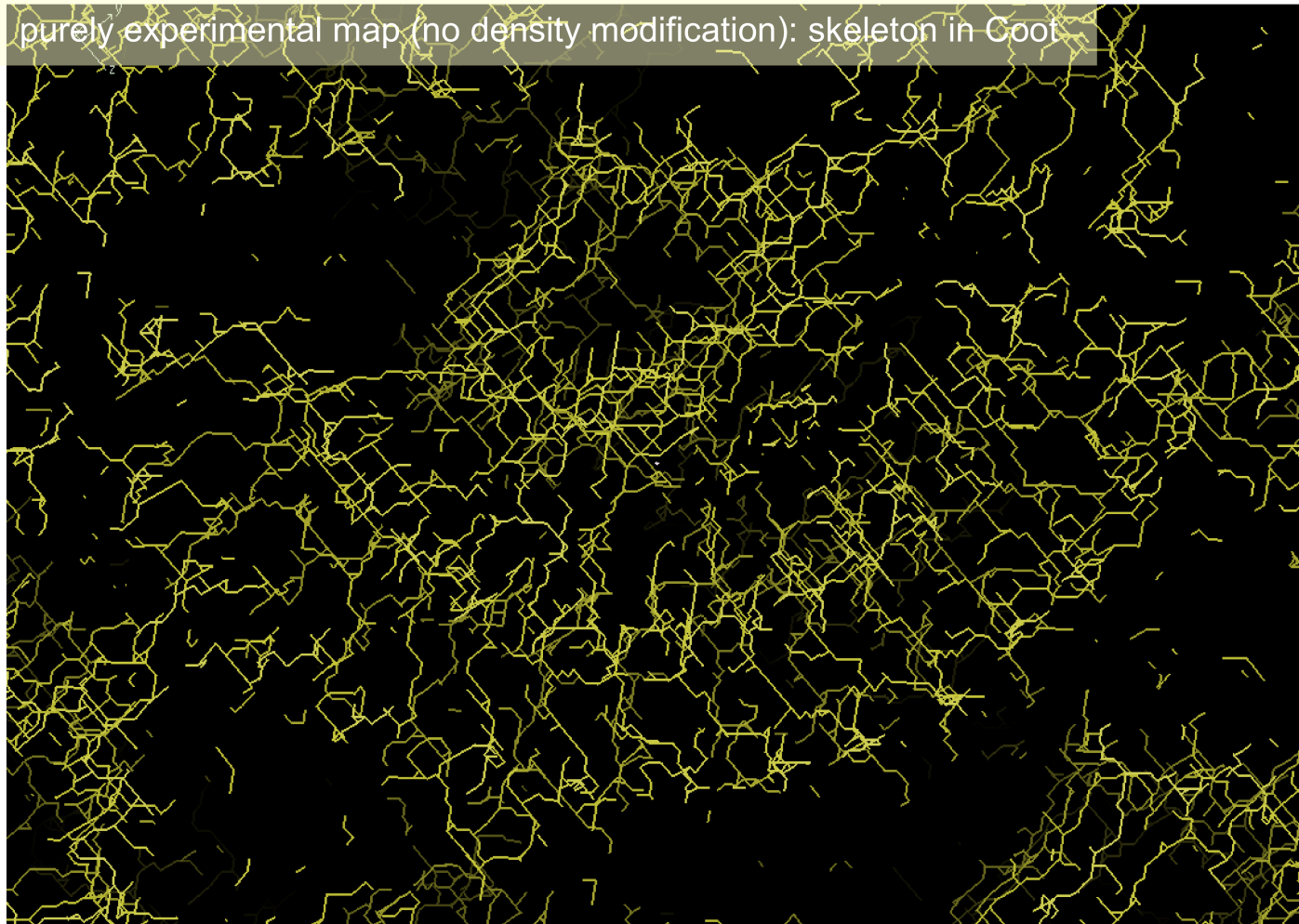
SAD (PEAK, 3600 images, multiplicity 25)



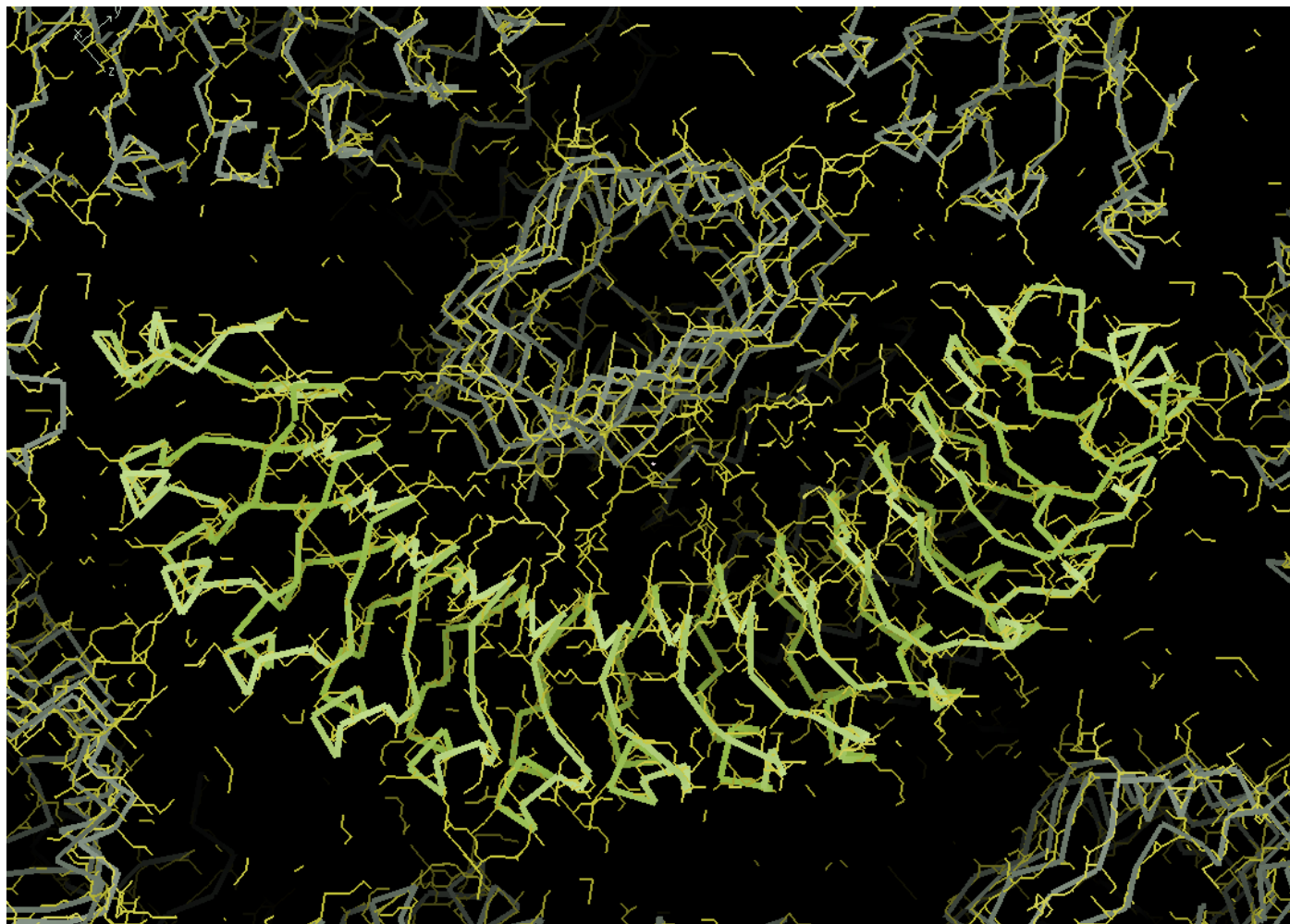
SAD, inverse-beam (1800+1800 images, multiplicity = 25)



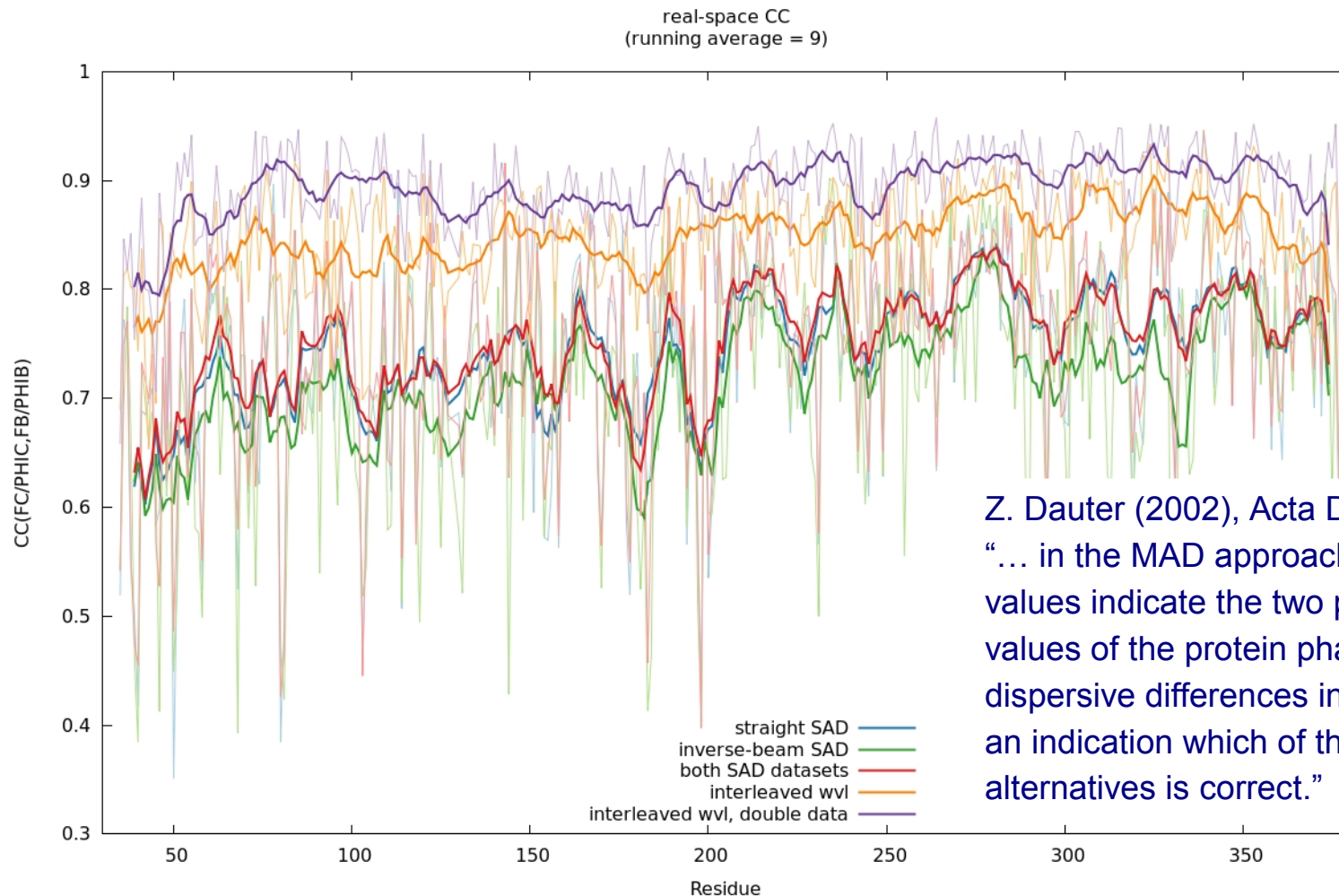
Double-inflection, interleaved-wvl (1800+1800 images, multiplicity = 12.5 + 12.5)



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



Phase ambiguity - SAD-vs-MAD

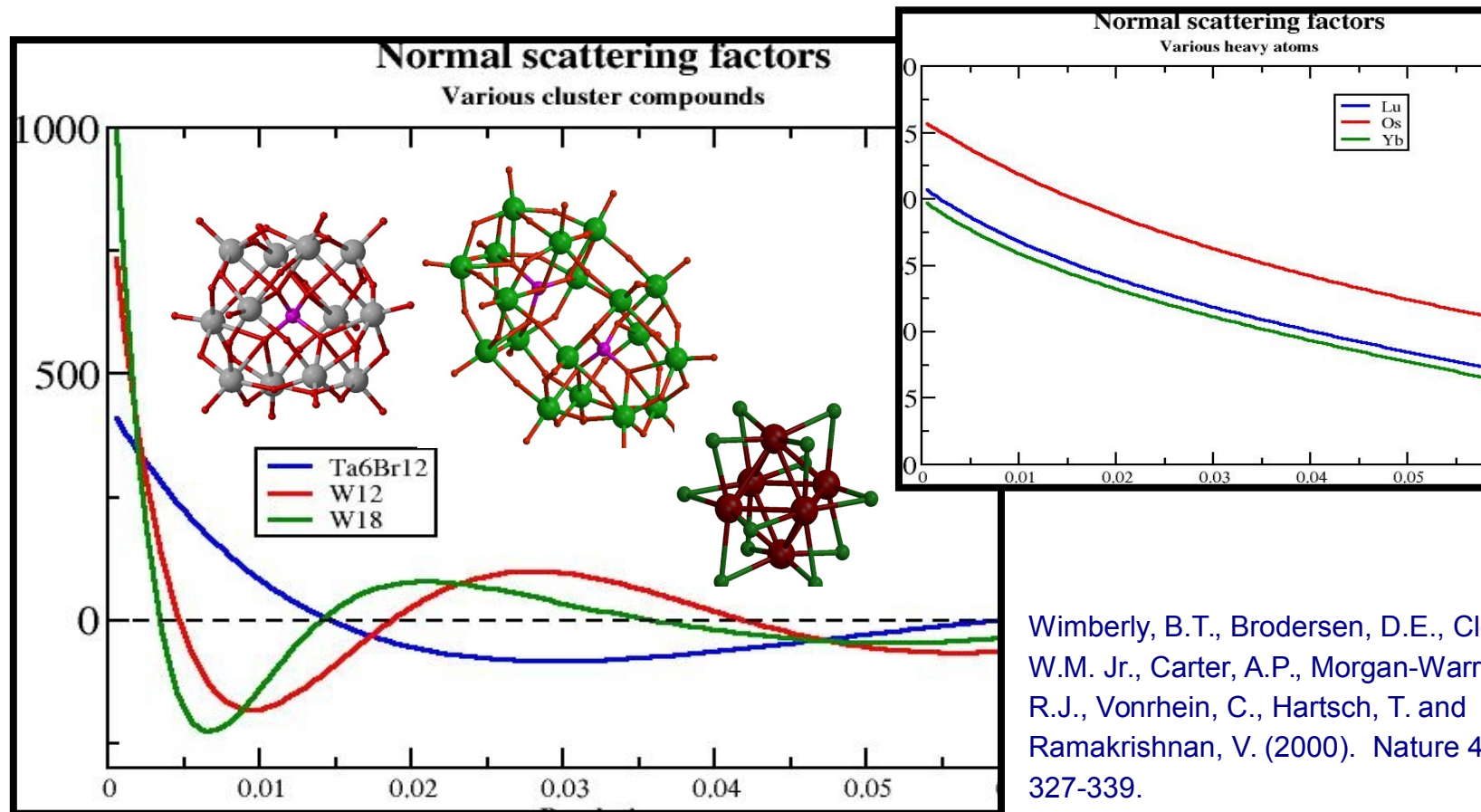


Z. Dauter (2002), Acta D58, 1958:
“... in the MAD approach the f'' values indicate the two possible values of the protein phase and the dispersive differences in f' provide an indication which of the two alternatives is correct.”

“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)

Spherical averaged cluster compounds



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

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
- 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

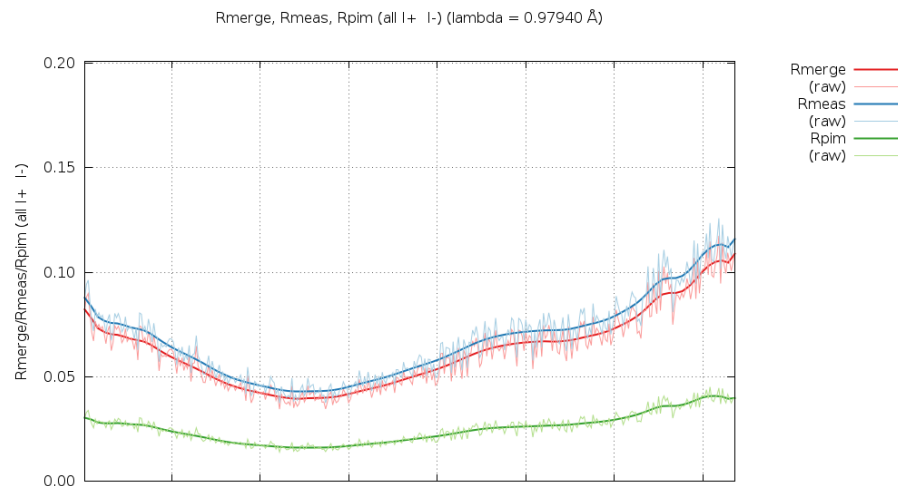
❑ 1y13

- Medium resolution
- 2-wvl MAD
- Small phasing signal

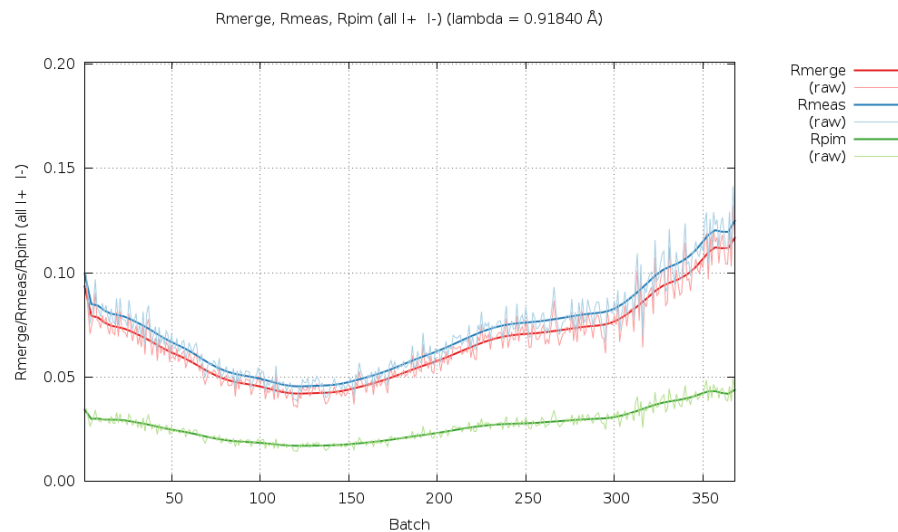
<http://www.globalphasing.com/autoproc/wiki/index.cgi?ACA2011Tutorial>

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

infl



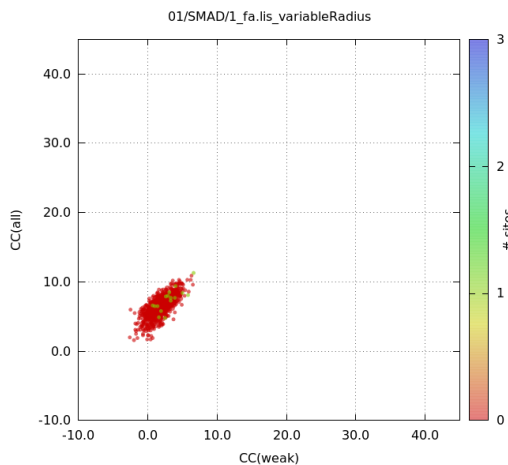
hrem



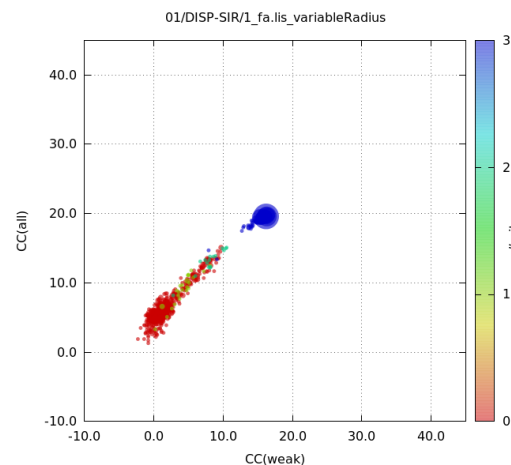
Radiation damage?

1Y13 – anomalous/dispersive differences

Anomalous 2.8Å

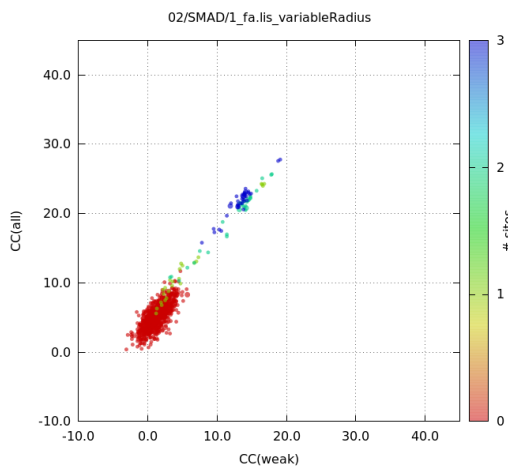


Dispersive 2.8Å

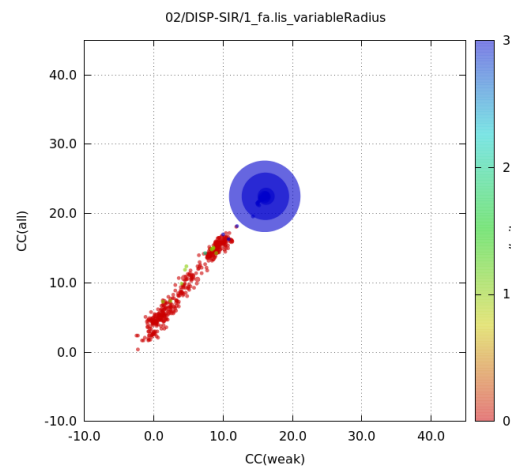


No Odose

Anomalous 2.8Å



Dispersive 2.8Å

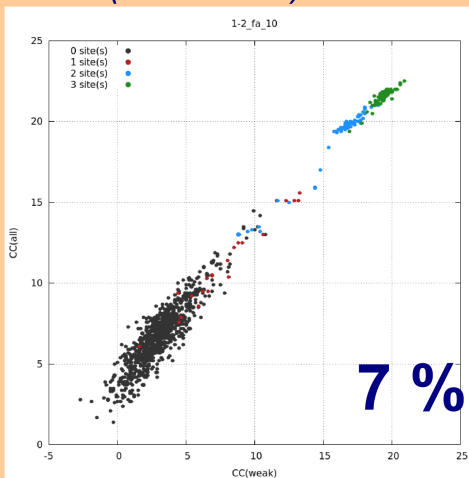


Odose

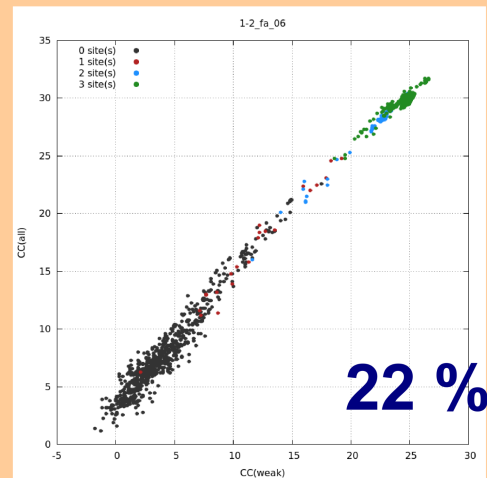
1Y13 - SHELXD (MAD) success rate

AIMLESS data (not shown): 13%

XSCALE
no 0-dose

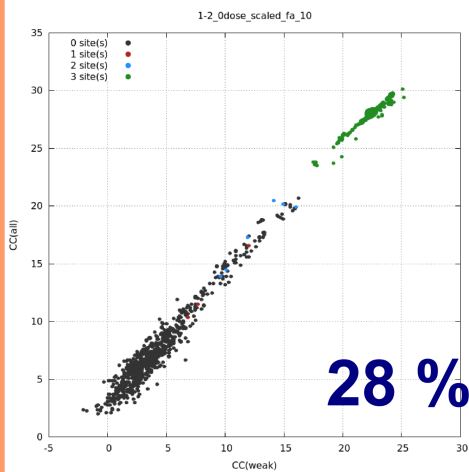


XSCALE
0-dose

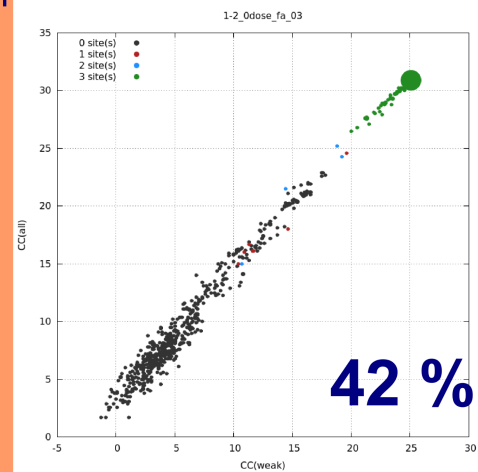


single 0dose on combined datasets with correct dose-stamp:

XSCALE*
0-dose

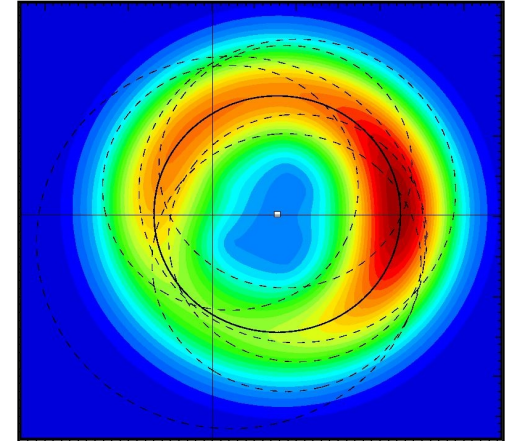


AIMLESS-XSCALE*
0-dose



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
- 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

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

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

<http://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