Experimental Phasing with SHARP/autoSHARP

Clemens Vonrhein
Global Phasing Ltd.

CCP4/DLS workshop
12/2019

GΦL
Global Phasing Limited

Overview

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/

GΦL
Global Phasing Limited

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.

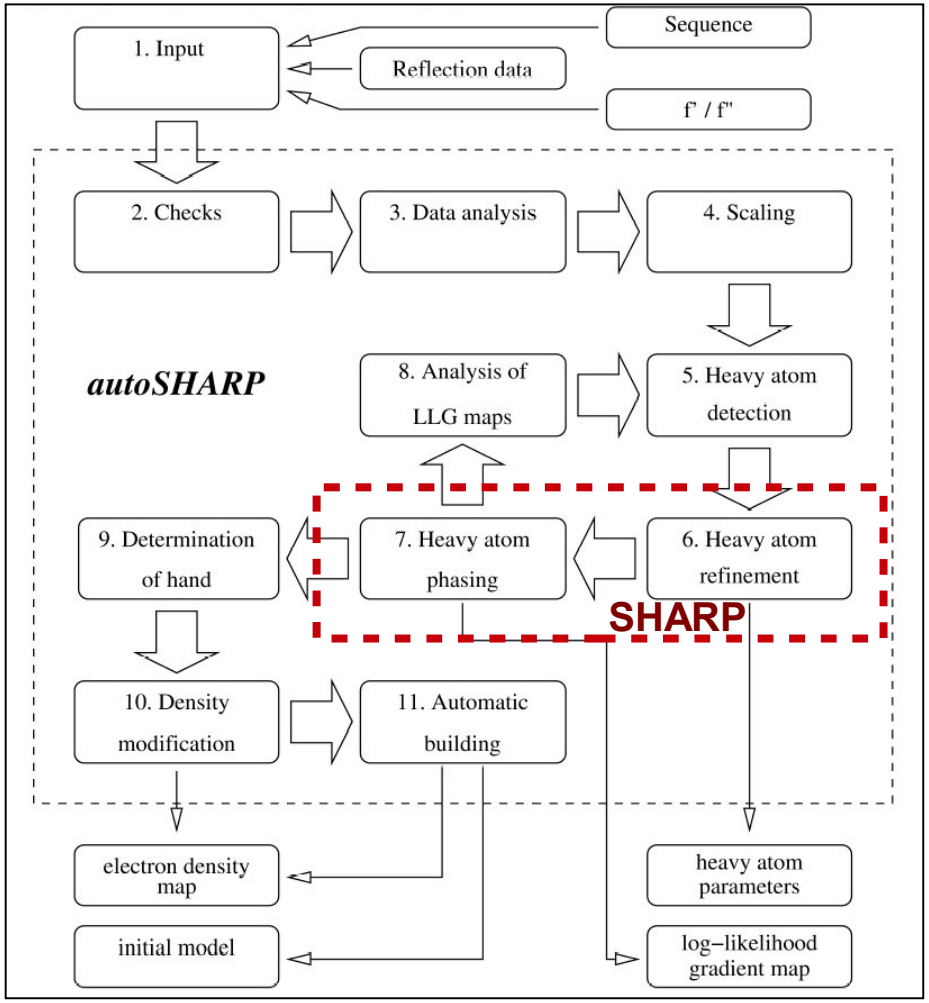
SHARP:

- 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. D59, 2023-2030.

GΦL
Global Phasing Limited

autoSHARP "pipeline"



minimum amount of input

feedback:

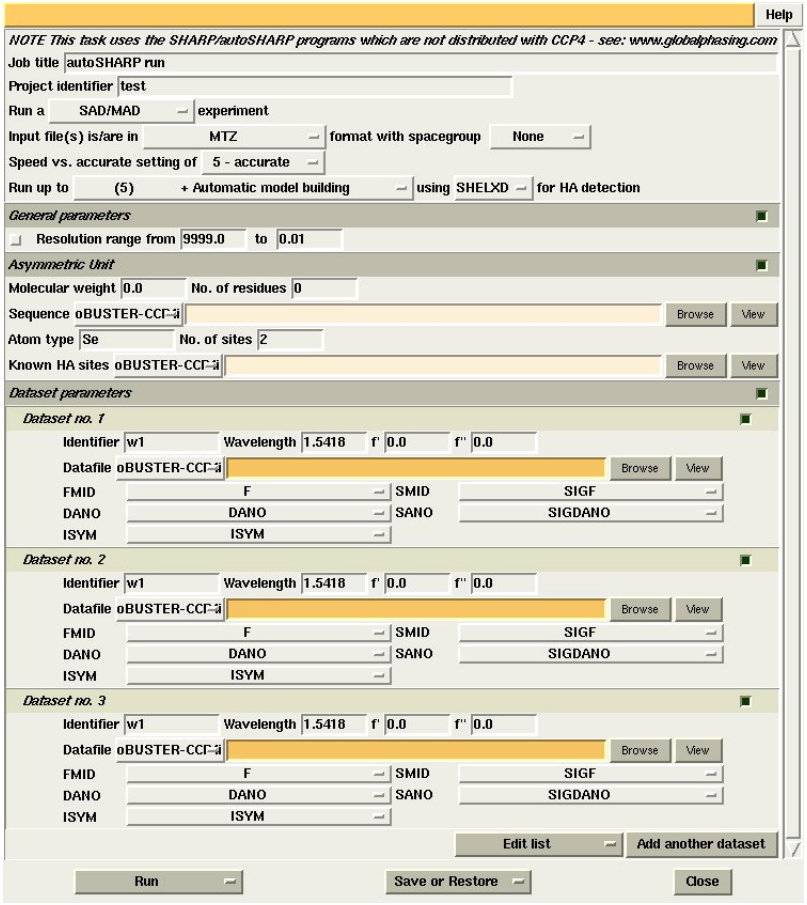
internally and to the user

several levels of results

GΦL
Global Phasing Limited

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



CCP4i interface

GΦL
Global Phasing Limited

Command-line interface

run_autoSHARP.sh -h

writes HTML output (view in browser)

at end: 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
```

GΦL
Global Phasing Limited

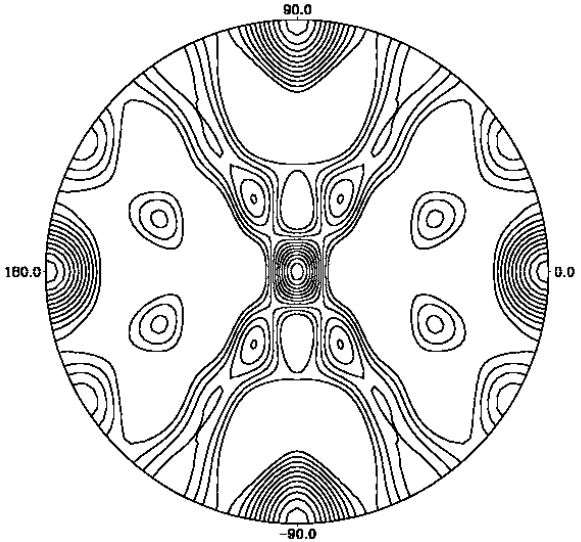
autoSHARP - Analysis

- Consistency (N>1):
  - cell (significant changes)
  - spacegroup (screw axes)
  - indexing (alternate possibilities)
  - sequence (Se-Met, S-SAD)
- Data quality:
  - CC(Δano), R-values, completeness
  - outliers (E-values)

SIGNAL

NOISE

- self-rotation function



- native Patterson
- Matthews parameters (probability, Rupp)

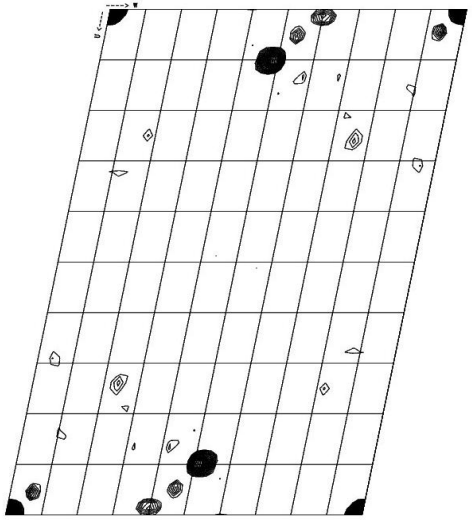
?/ASU

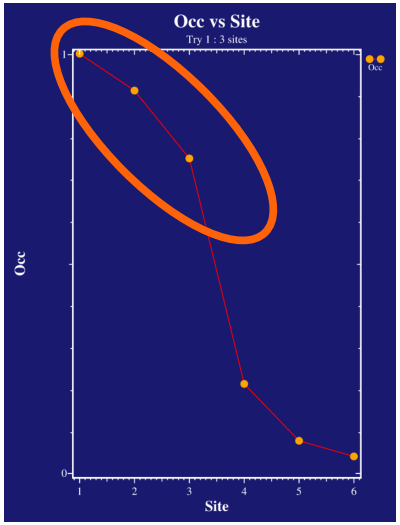
autoSHARP – HA detection

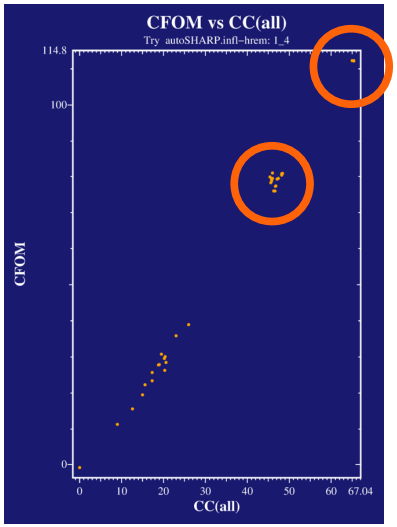
- only data with significant signal (from *Analysis* step)
- try different scenaria (MAD, peak-SAD, 2-wvl MAD, ...)
- adjust number of HA sites (e.g. if exact number not known)
- use solution with best score

- **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.

GΦL
Global Phasing Limited

autoSHARP - SHARP

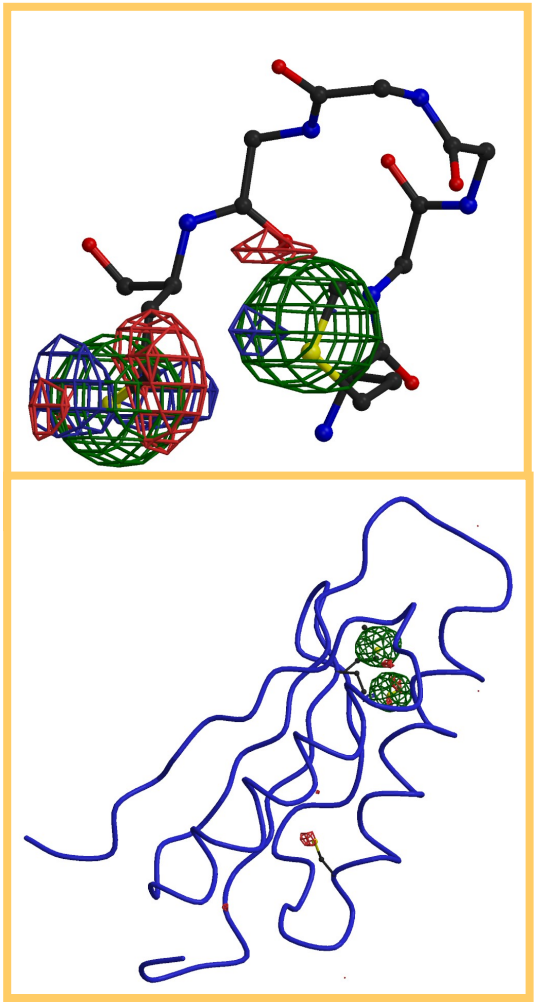
- HA refinement and phasing
- overall anisotropic scaling
- analysis of log-likelihood gradient maps:

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**

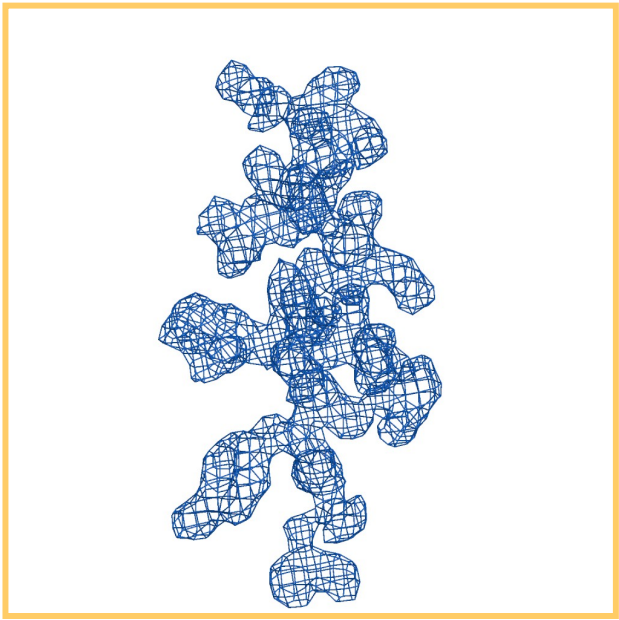
Best possible phases

- 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 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.

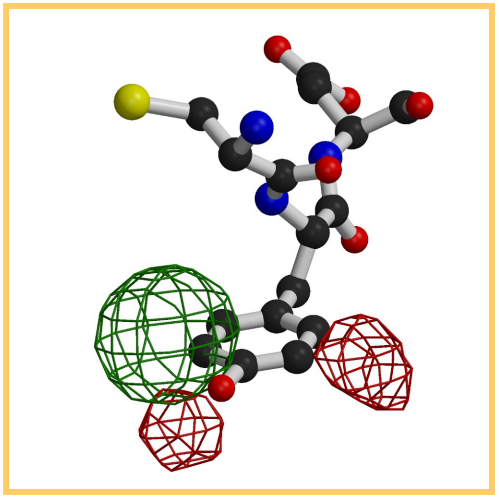
GΦL
Global Phasing Limited

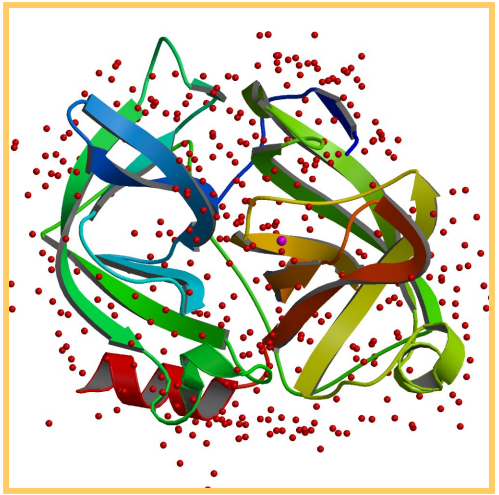
autoSHARP - Output

- 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 (→ SHELXE)
- LLG (residual) maps

- density modified phases/maps
- initial model(s)





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**[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

"web-based" interface (local installation)

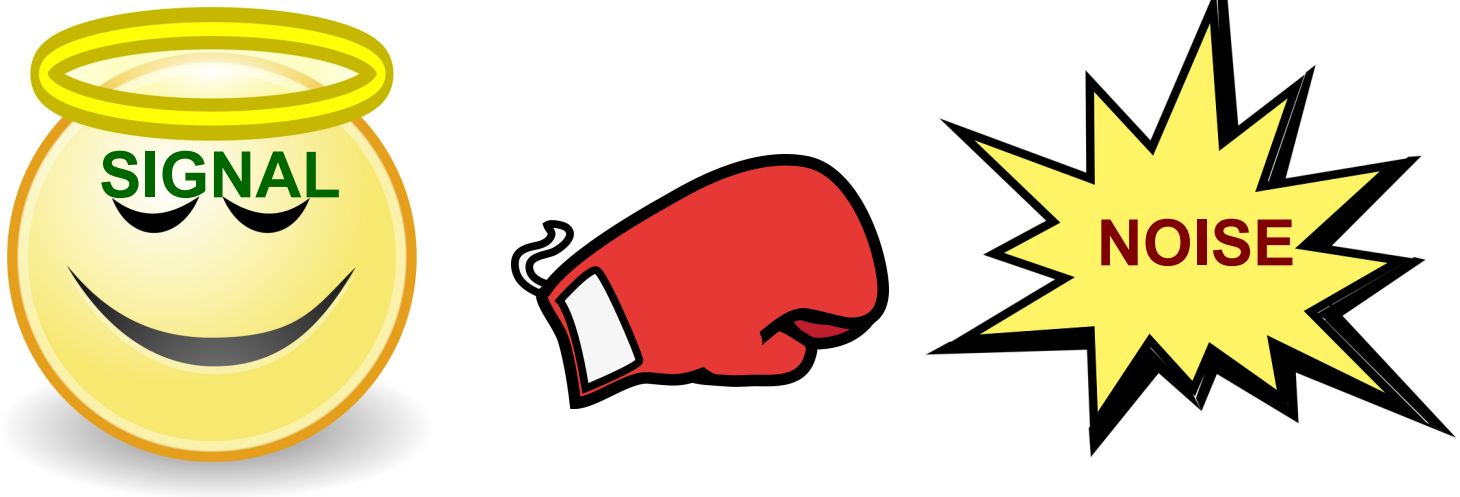
1) C. Vonrhein and G. E. Schulz, Acta Cryst., D55, 225 - 229 (1999)

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

GΦL
Global Phasing Limited

How to improve on first results

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



GΦL
Global Phasing Limited

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**	*

GΦL
Global Phasing Limited

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 *

GΦL
Global Phasing Limited

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.
```

GΦL
Global Phasing Limited

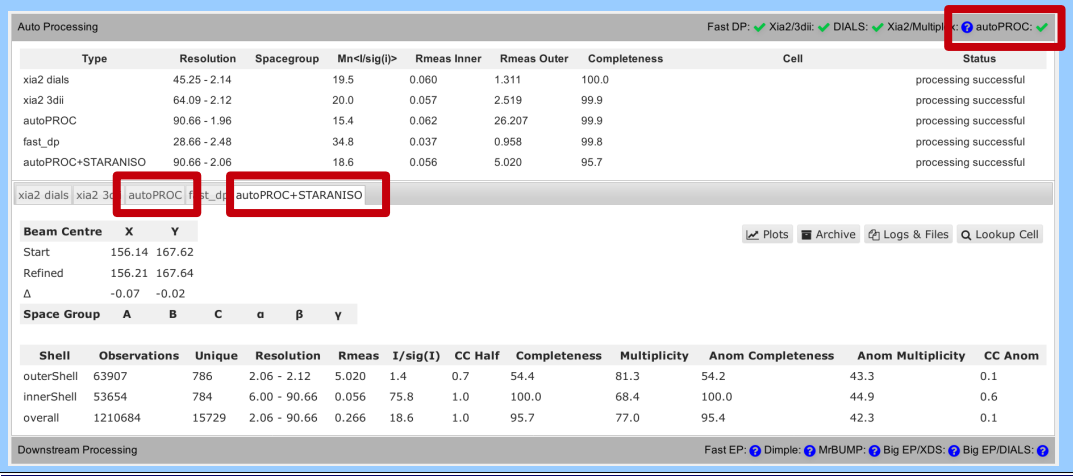
How to improve on first results

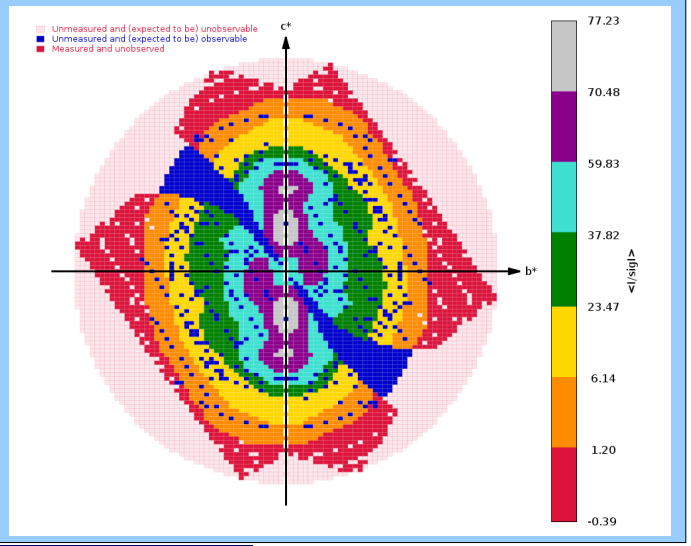
- after **autoSHARP**:
  - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)
  - analyse **log-likelihood gradient** (residual) maps: additional sites, f'/f'' refinement, anisotropic B-factor for sites, ...

Global Phasing Limited

How to improve on first results

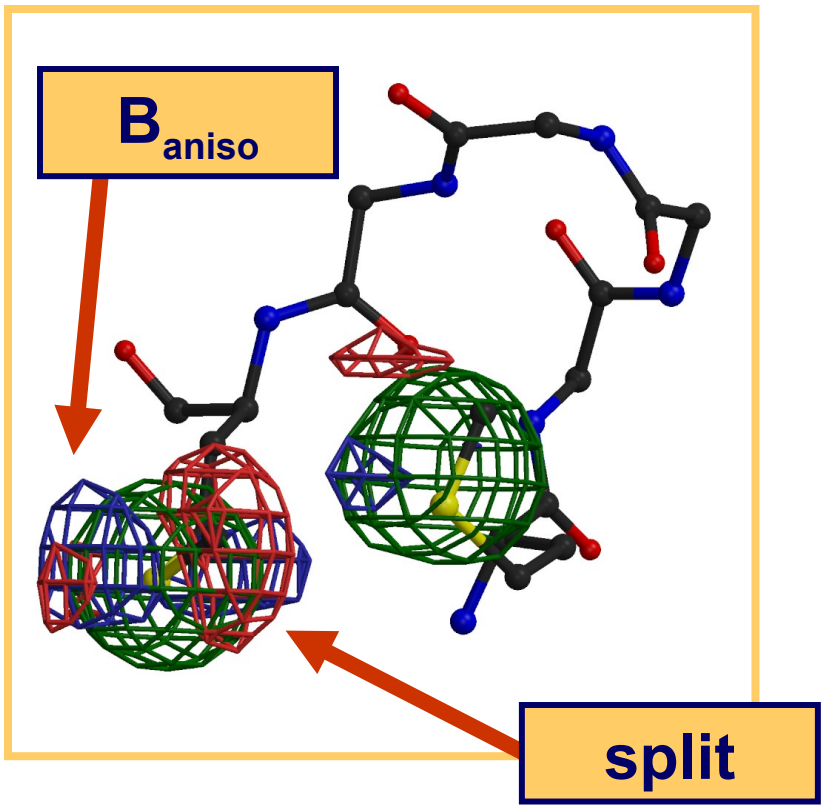
- after **autoSHARP**:
  - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)
- **autoPROC (2005-)**: *www.globalphasing.com/autoproc/*
  - expert system (using XDS, POINTLESS, AIMLESS, XSCALE)
  - multiple lattices, multi-wavelength, multi-sweep, multi-orientation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



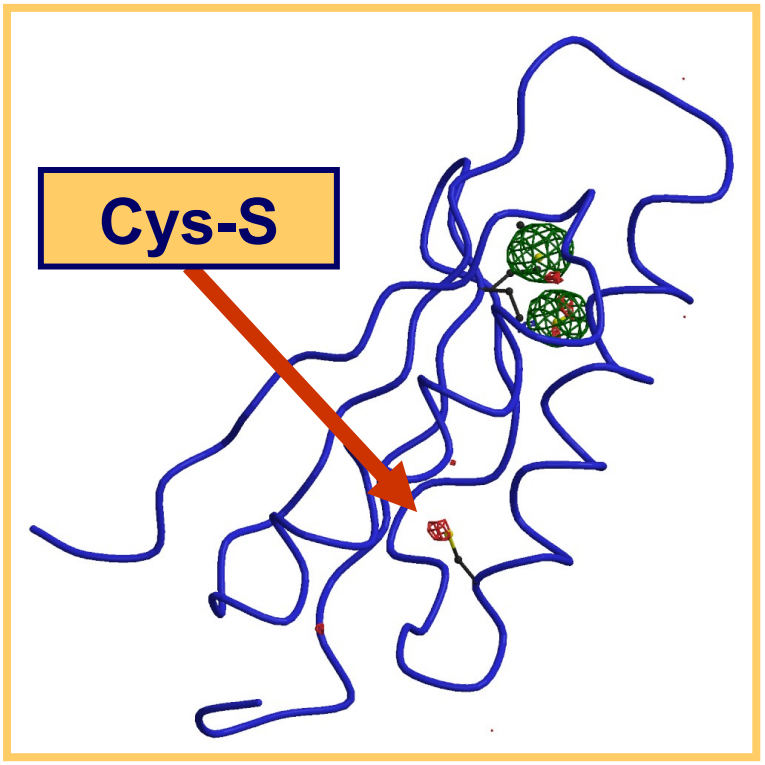


GΦL
Global Phasing Limited

Example (LLG maps)



Se sites (LLG peak)



Anomalous LLG

GΦL
Global Phasing Limited

How to improve on first results

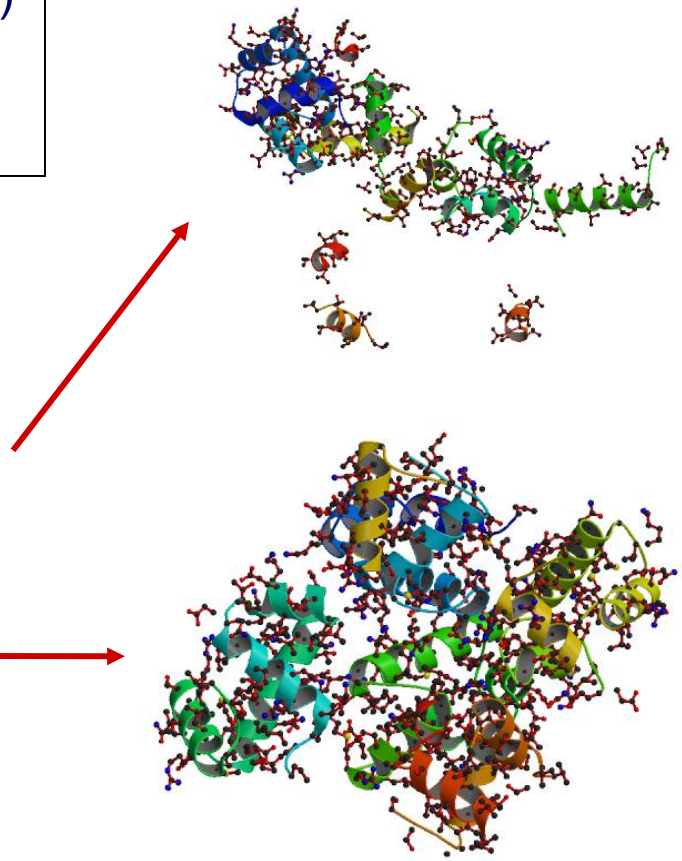
- after **autoSHARP**:
  - **warning messages** point back to data processing/scaling problems (beamstop, ice-rings, decay, ...)
  - analyse **log-likelihood gradient** (residual) maps: additional sites, f'/f'' refinement, anisotropic B-factor for sites, ...
  - significant difference between score of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

GΦL
Global Phasing Limited

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., Vonrhein, C., Perrakis, A. & Lamzin, V.S. (2004) Breaking good resolutions with ARP/wARP. J. Synchr. Rad. 11, 56-59

GΦL
Global Phasing Limited

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 → HA detection
  - Overcoming non-iomorphism issues

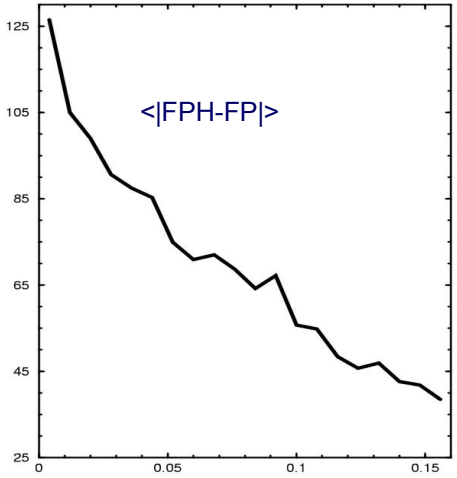
"MR-SAD"

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

GΦL
Global Phasing Limited

Why not MIR?

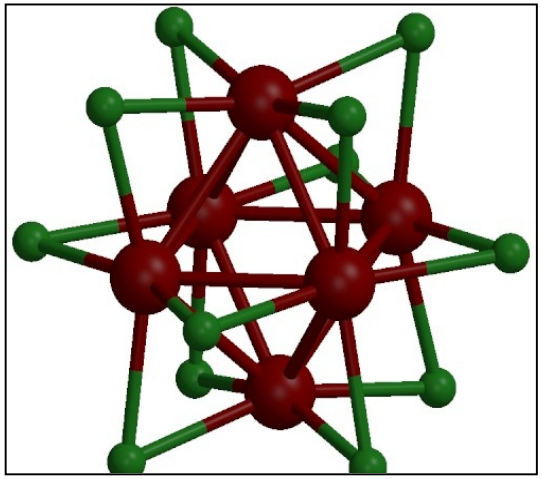
- **Non-isomorphism**
  - **datasets from different crystals**
  - **agressive soaks, co-crystallization**
- **no. of sites (and occupancy) not well known**
- **'old-fashioned'**



- **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:*

    <|FPH-FP|>

Ta_6Br_{12}



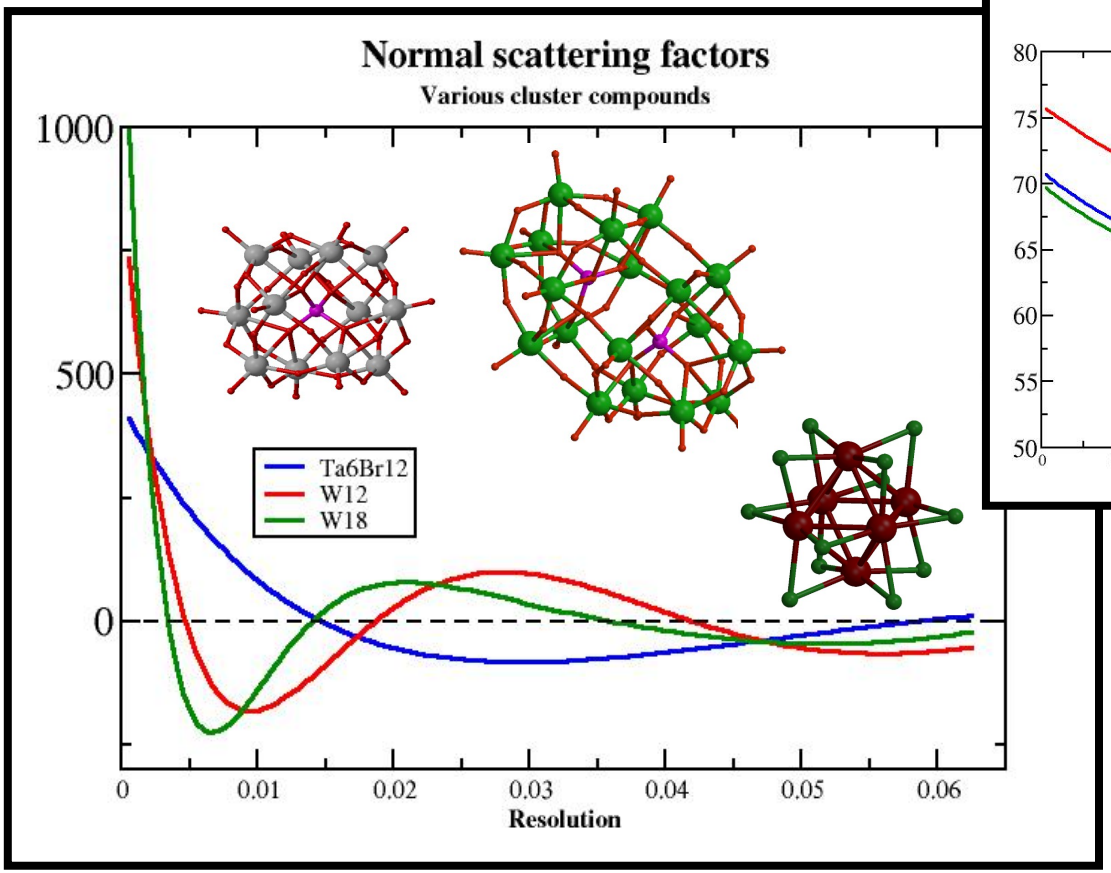
GΦL
Global Phasing Limited

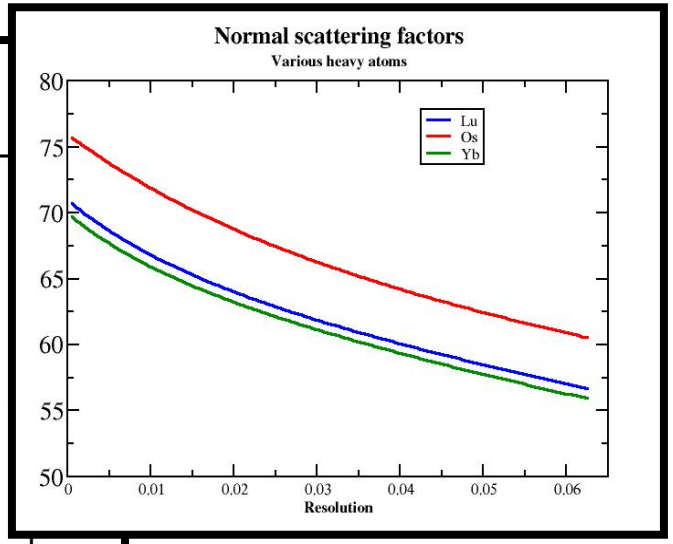
"Advanced" experimental phasing

- **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)

GΦL
Global Phasing Limited

Spherical averaged cluster compounds





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

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

GΦL
Global Phasing Limited

Summary

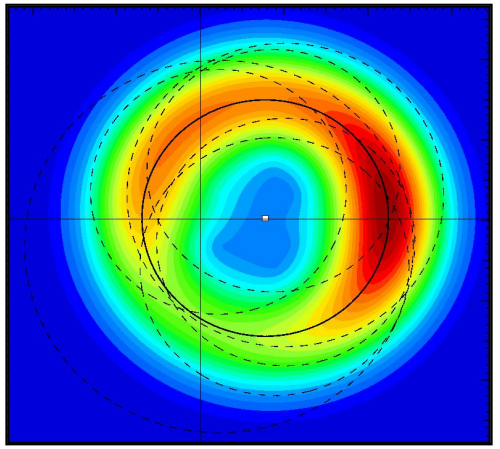
- Automation is great ... if it works
- Meaningful feedback important!
- Improving on initial results often involves going back to scaling and/or data processing and/or collection strategy.

- 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

GΦL
Global Phasing Limited

Acknowledgement

Global Phasing, Cambridge (UK):

• Gérard Bricogne, Claus Flensburg, Andrew Sharff, Wlodek Pacziorek, 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