Substructure Determination using SHELX

SHELXD: From *Ab initio* to substructure solution

Phasing and Density Modification with SHELXE

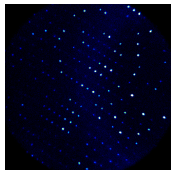
ARCIMBOLDO

Isabel Usón, ICREA, IBMB-CSIC, Barcelona
uson@ibmb.csic.es

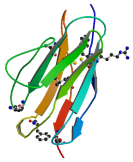
Crystallographic Phase problem

$$\rho_{xyz} = (1/V) \sum_{hkl} F_{hkl} \exp[-2\pi i (hx+ky+lz)]$$

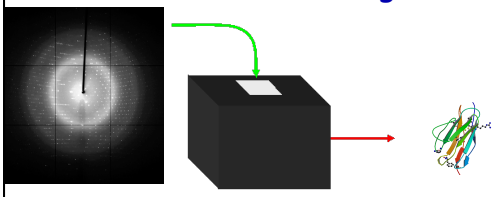
Soft X-rays for experimental phasing



From data to structure



SHELXD + SHELXE George Sheldrick

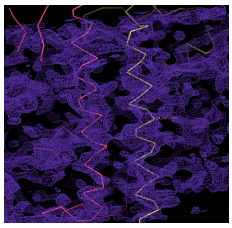


- SHELXD: to determine a heavy atom or anomalous scatterer substructure from ΔF or MAD F_A data.
- SHELXE: once the substructure is known, it can be used to calculate reference phases, and from them phases for the whole structure but usually, density modification is required

Ab Initio: Phasing by Direct methods

Most small-molecule structures are solved by *direct methods*. Approximate phases for the strongest reflections are derived from the measured intensities

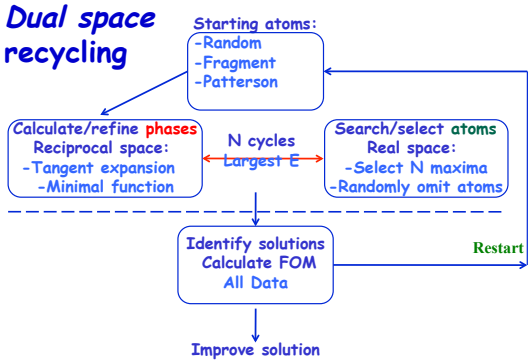
This is enough to find most atoms, and then complete the structure



Making some general assumptions, it is possible to exploit the high redundancy of the diffracted intensities: (100 independent reflections / atom)

Equal atom structures with > 200 atoms are a problem

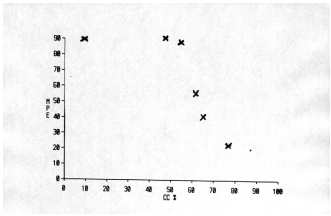
Dual space recycling



Shake & Bake // SHELXD

Sheldrick, Hauptman, Weeks, Miller & Usón (2001), *International Tables for Crystallography Vol. F*, eds. Arnold & Rossmann, pp. 333-351

Figures of merit: CC, PATFOM, minimal function, CCweak...



$$CC = \frac{[\Sigma w E_o^2 E_c^2 . \Sigma w - \Sigma w E_o^2 . \Sigma w E_c^2]}{\{[\Sigma w E_o^4 . \Sigma w - (\Sigma w E_o^2)^2] . [\Sigma w E_c^4 . \Sigma w - (\Sigma w E_c^2)^2]\}^{1/2}}$$

Fujinaga & Read. (1987). J.Appl.Cryst., 20, 517-521.

Structures solved by SHELXD

Compound	Spacegroup	N(+solv)	HA	dÅ
Actinomycin X2	P1	273(305)	-	0.9
Actinomycin Z3	$P2_12_12_1$	186(307)	2Cl	0.96
Vancomycin	$P4_32_12$	202(312)	6Cl	1.09
Actinomycin D	P1	270(314)	-	0.94
Ristocetin A	$P2_1$	294(420)	-	1.03
Hirustasin	$P4_32_12$	402(467)	10S	1.2
Cyclodextrin	$P2_1$	448(467)	-	0.88
Decaplanin	$P2_1$	448(635)	4Cl	1.00
Cyclodextrin	P1	483(562)	-	1.00
Bucandin	C2	516(634)	10S	1.05
Amilose CA26	P1	572(719)	-	1.1
Viscotoxin B2	$P2_12_12_1$	722(812)	12S	1.05
Mersacidin	$P3_2$	750(826)	24S	1.04
rc-WT CvHiPIP	$P2_12_12_1$	1264(1599)	8Fe	1.2
Cytochrome c3	$P3_1$	2024(2208)	8Fe	1.2

Triclinic Lysozyme, biggest equal atom structure.

Experimental phasing of macromolecules

Find the small molecule in the macromolecule:

heavy atoms provide *reference phases*

calculate the phases ϕ_T of the full structure by:

$$\phi_T = \phi_A + \alpha$$

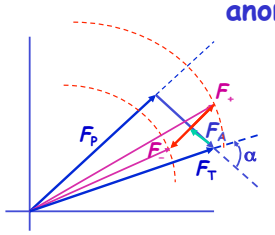
ϕ_A is the calculated phase of the heavy atom substructure

α can be estimated from the experimental data

The phase determination requires the following stages:

1. Location of the heavy atoms.
2. (Refinement of heavy atoms and) calculation of ϕ_A.
3. Calculation of starting protein phases using $\phi_T = \phi_A + \alpha$.
4. Improvement of these phases by density modification.

Substructures of heavy atoms and anomalous scatterers



$F_T = F_P + F_A$

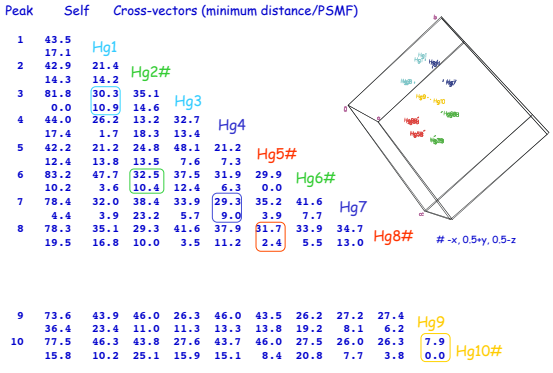
$f = f^0 + f' + f''$

If $|F_+| >> |F_-|$, adding $\alpha = 90°$ to the phase of $|F_A|$ gives a good approximation to the phase of $|F_T|$.

Substructure solution with SHELXD

Protein	#sites	kDa	SG	d_{min}	CC	P ratio
ApD	3/3 Se	16	$C222_1$	2.2	45	16
RRF	3/4 Se	20	$P4_32_12$	4.0	60	1.4
ModE	6/6 Se	57	$P2_12_12$	3.0	66	7.3
9hem	18/18 Fe	64	$P2_1$	2.9	73	4.0
AT	32/32 Se	160	C2	3.5	49	5.1
Cyanase	40/40 Se	170	P1	2.4	57	1.0
TH	51/60 Se	161	$P2_1$	2.5	52	2.8
AEP	66/66 Se	270	$P2_1$	2.55	61	13
KPHMT	145/160 Se	567	$P2_1$	2.8	49	35

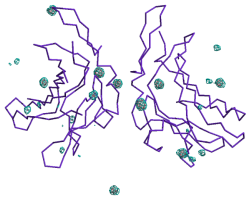
- Anomalous scatterers were traditionally found with Patterson or Direct Methods.
- SHELXD solve larger substructures combining Patterson, direct methods and geometry.

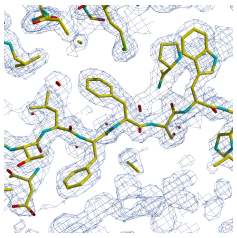


$HgCl_2$-soak of AtsK, Ilka Müller

I^--substructure and phases

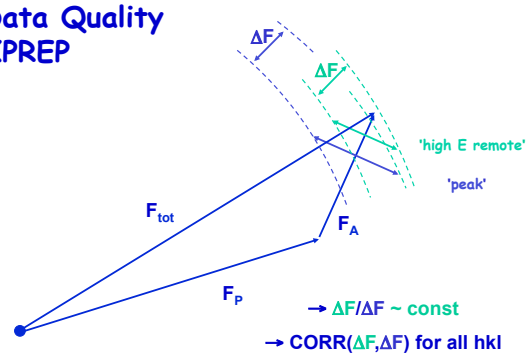
- R11 C2, $P2_12_12_1$, 210 aa
- SAS / SIRAS 16 sites
- Phasing with SHARP/dm





ΔF/PHI-90 5 sigma
10 sigma

Data Quality XPREP



→ ΔF/ΔF ~ const

→ CORR(ΔF,ΔF) for all hkl

Fe-substructure, Aneta Oleksi

·3 wavelength MAD to 2.6 Å, K-Edge 1.74Å

·$P4_132$ or $P4_332$

·Anomalous signal/noise ratios (1.0 is random). The first line is based on input sigmas, the second on variances of F+ and F- (if not already averaged):

```
Inf- 8.0 - 6.0 - 5.0 - 4.2 - 4.0 - 3.8 - 3.6 - 3.4 - 3.2 - 3.0 - 2.8 -2.6Å
 3.65  2.90  3.40  3.01  2.47  2.28  2.01  1.74  1.93  1.59  1.24  1.26
```

Anomalous correlation coefficients (%) against previous datasets

```
Inf- 8.0 - 6.0 - 5.0 - 4.2 - 4.0 - 3.8 - 3.6 - 3.4 - 3.2 - 3.0 - 2.8 -2.6Å
   95.8  90.1  92.5  88.8  83.9  76.2  78.6  75.1  70.0  55.7  28.4  13.5
   84.0  79.6  87.2  84.8  80.6  77.6  77.9  67.3  65.6  51.5  20.5  10.6
```

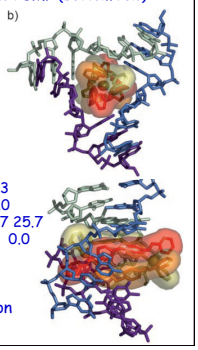
·SHELXD finds no solution (?)

Oleksi, Blanco, Boer, Usón, Aymami, Rodger, Hannon and Coll
Angew Chem, 45 (2006), 1227-1231

Fe-on special positions

Minimum distances (top row, 0 if special position) and PSMF (bottom row)

```
Peak   x      y      z      self  cross-vectors
99.9  0.5212 0.9788 0.0212   0.0
                             128.5
89.3  0.8750 0.8750-0.1250   0.0  27.8
                             164.1 100.2
82.0  0.4283 1.0717-0.0717   0.0  11.3 34.4
                             62.2 128.1127.0
43.0  0.7176 1.0772-0.0679   9.2  16.6 18.4 19.0
                             0.0  84.3 28.0 57.5
26.3  0.7823 0.7823-0.2177   0.0  28.4 11.3 25.6 19.3
                             2.5  29.8 28.0 10.7  0.0
20.0  0.5907 1.0237-0.0019  10.4   6.0 24.1 12.8 10.7 25.7
                             0.0   7.0  0.0  0.0  0.0  0.0
```



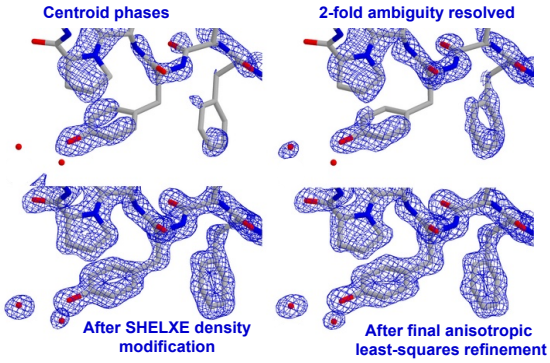
Oleksi, Blanco, Boer, Usón, Aymami, Rodger, Hannon and Coll (2006). Angew Chem, **45**, 1227-1231

SAD phasing procedure in SHELXE

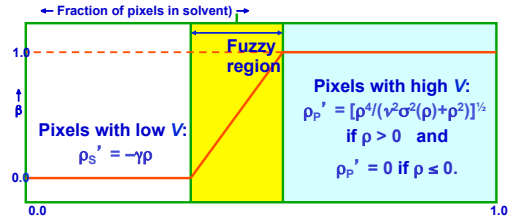
1. Calculate *centroid phases* (α = 90 or 270°) with weights that depend on the (normalized) $|\Delta F|$.
2. Use these phases, foms and $|F_T|$ to calculate a map, apply *low density elimination* (Woolfson, Giacovazzo) and reinvert map to estimate phases for small $|\Delta F|$ and centrics. This is an alternative to B.C.Wang's method of solvent flattening to *resolve two-fold ambiguities*.
3. If heavy atoms are present in the native, include their σ_A-weighted direct estimates of ϕ_T (as in the heavy-atom method for small molecules, does not use $|\Delta F|$).

These three steps produce independent phase estimates (2 and 3 are orthogonal to 1) so can be combined using σ_A- weights (Read).

Stages in the phasing of cubic insulin

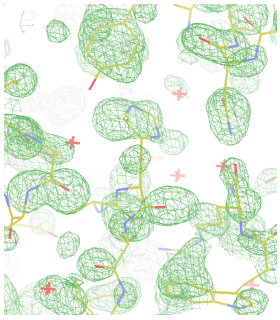


Density modification in SHELXE

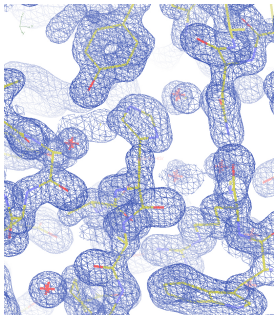


- The sphere of influence algorithm
- Data extrapolation (Giacovazzo) "free lunch algorithm"
- Mainchain autotracing
- Resolving super-sulfurs

Maps with no/a free lunch



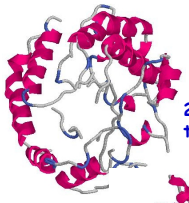
Best experimental phases after den. mod. (MapCC 0.57)



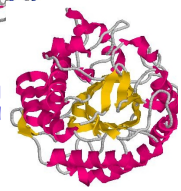
After expansion to 1.0 Å with virtual data (MapCC 0.94)

Autotracing in SHELXE

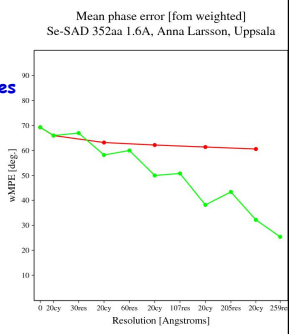
352aa protein 1.6Å, Anna Larsson, A. Jones, Uppsala



256 residues traced



Final refined structure

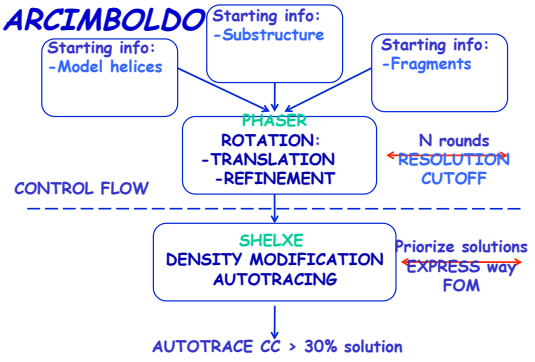


ARCIMBOLDO: fragments rather than atomicity

http://chango.ibmb.csic.es/ARCIMBOLDO

- ARCIMBOLDO: combines fragment search (PHASER, Read), experimental phasing, density modification, auto tracing (SHELXE, Sheldrick), within a multisolution frame. (2.0 Å)
- Condor Grid

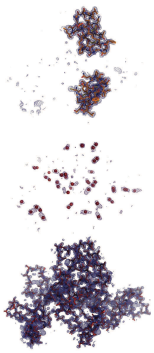
ARCIMBOLDO



ARCIMBOLDO
Rodríguez, Grosse, Himmel, Becker, M de Ilarduya, Sheldrick & Usón (2009), *Nature Methods*, 6, 651-653

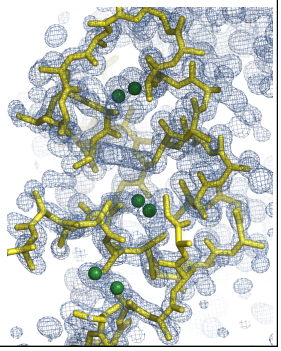
Hell D (Andrea Thorn)

- *up to 8 copies of a 30 aa peptide with 3 disulfide bridges, I422*
- 1.95Å Data collected at 1.9Å, 2Xtals
- Substructure could not be located (difficult even with calculated data…)
- Can be solved with ARCIMBOLDO with 2 helices from related structures
- Partial solution can be completed with anomalous signal



Viscotoxin A1, Aritra Pal

- $P4_32_12$ in House Data, 1.54Å, 1.3Å
- locate substructure against anomalous data with Phaser (six cysteins in three S-S bridges)
- Expand and autotrace



Torture Case: Gordon Leonard's DNA

- *8 base-pairs in $P6_1$: 14 P, data to 2.4 Å*
- *Can we solve it from the atoms? MPE = 70°*
- *From the anomalous data? MPE = 40°*
- *Can we find the substructure at 2.4 Å from calculated data and phase the structure?*

MPE = 33°

- *At 2.8Å?* *NO substructure*

Acknowledgements

- George Sheldrick, Andrea Thorn
- My group in Barcelona
- MICINN, ICREA, CSIC, AGAUR