
ABOUT PHASING WITH NATIVE SAD

JIAN YU
HOKKAIDO UNIVERSITY

2021.11.13

Phasing with Native-SAD

- Merit

1. Only native crystal is needed.
2. Homolog structure is not necessary.

- Demerit

1. Anomalous signal is weak.
2. Need the X-ray beam of long wavelength.

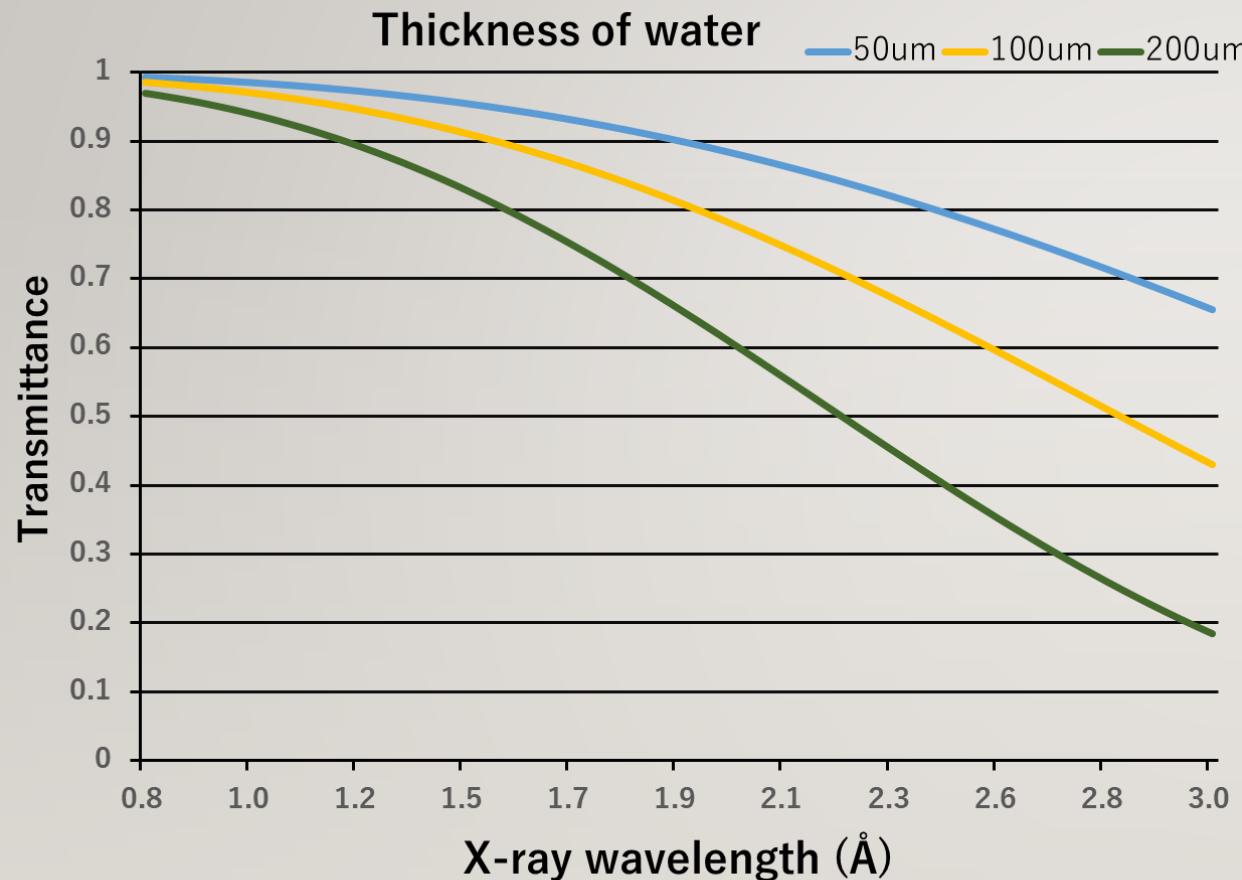
Outline

- Before data collection
- During data collection
- After data collection

Outline

- Before data collection
- During data collection
- After data collection

Phasing with Native-SAD



Henke et al., 1993.

Absorption

Transmittance through water decreases with increasing X-ray wavelength and the thickness of the water solution.

Contradiction

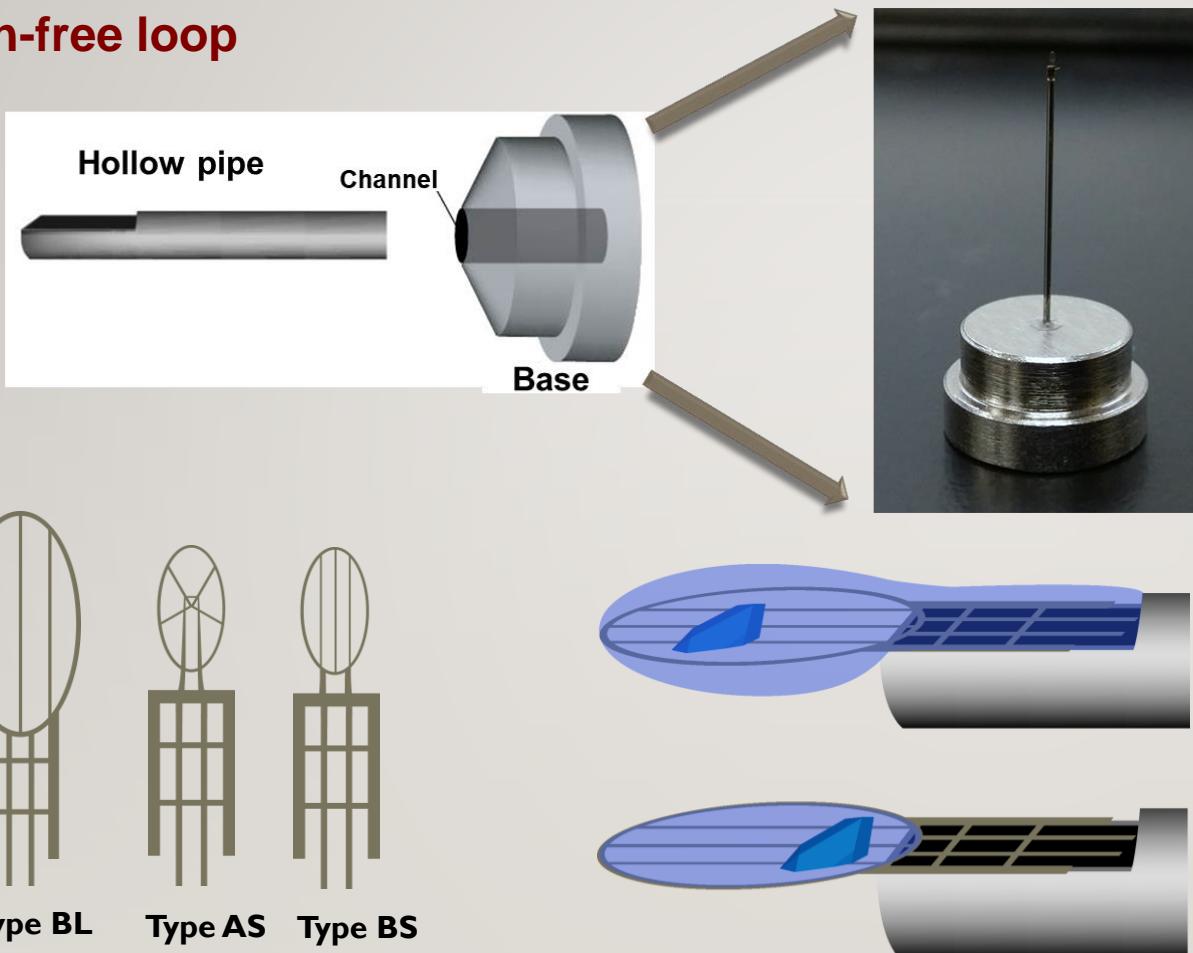
Long wavelength leads to strong anomalous signal

vs

Easy to be absorbed by the materials

Freeze your crystal

➤ Solution-free loop



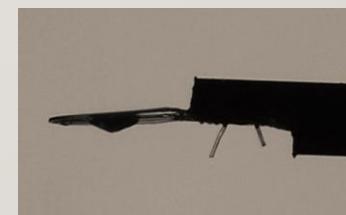
Type AL Type BL Type AS Type BS

Decrease the absorption from solution

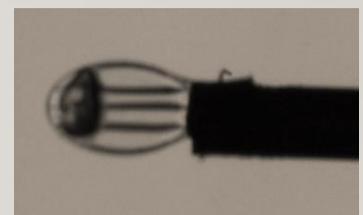
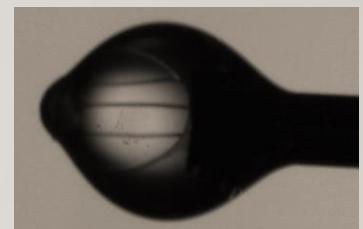
One way to reduce the unwanted absorption is to use a **Solution-free loop** to remove the solution around the crystal.



Before suction



After suction



Check the anomalous signal of your target protein

- Calculate the anomalous scattering ratio by the following equation

$$\langle|\Delta F|\rangle/\langle F\rangle = \sqrt{2}(\sqrt{N_A} \cdot \Delta f_A'') / (\sqrt{N_P} \cdot Z_{eff})$$

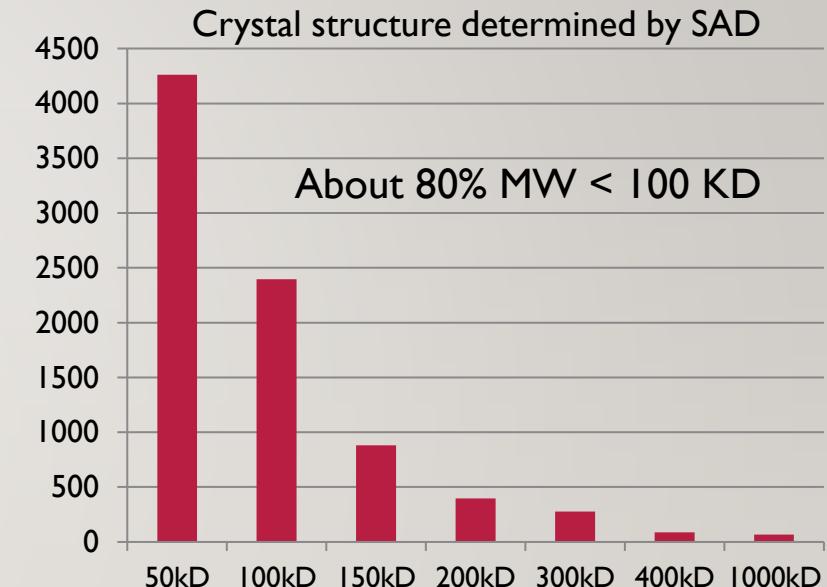
Nature, 290, P107–113 (1981)

Z_{eff} = effective atomic number, 6.7 for non-hydrogen

N_P : number of atoms (5C, 1.5O, 1.35N/per residue)

N_A : number of anomalous scattering atoms

$\Delta f_A''$: anomalous scattering factor depended on wavelength



Example : 423 residues, 17 (Met+Cys) → $\langle\Delta F\rangle/\langle F\rangle = 1.24\% (\lambda=1.9\text{\AA})$ 2.3% ($\lambda=2.7\text{\AA}$)

- Using the X-ray of longer wavelength will get stronger anomalous while leads to more absorption and low-resolution data.

For example, at PF-BL1A, around 2.6 \AA is the highest resolution when 2.7 \AA wavelength is used

Outline

- Before data collection
- During data collection
- After data collection

Work at beamline

I. Take one shot to check exposure time and resolution.

PF-BLIA: the highest resolution is around 2.6 Å when 2.7 Å wavelength is used

2. Data collection for one dataset, and then use XDS to process the data.

Check XDS statistics: **Resolution, Redundancy, I/s(I), Rmeas and Sigano**

3. After that, check the radiation damage with “xdsstat”.

```
$ xdsstat > XDSSTAT.LP  
$ lograph XDSSTAT.LP
```

4. Adjust the exposure time to collect multiple datasets.

Data: 2019-11-25, PF-BLIA, $\lambda=2.7 \text{ \AA}$, exposure 0.1S



Check the quality of your data with graph

\$ xdsstat > XDSSTAT.LP; \$ lograph XDSSTAT.LP

$$R_d = \frac{\sum_{hkl} \sum_{|i-j|=d} |I_{hkl,i} - I_{hkl,j}|}{\sum_{hkl} \sum_{|i-j|=d} (I_{hkl,i} + I_{hkl,j})/2}$$

Diederichs K. (2006) *Acta Cryst D62*, 96-101

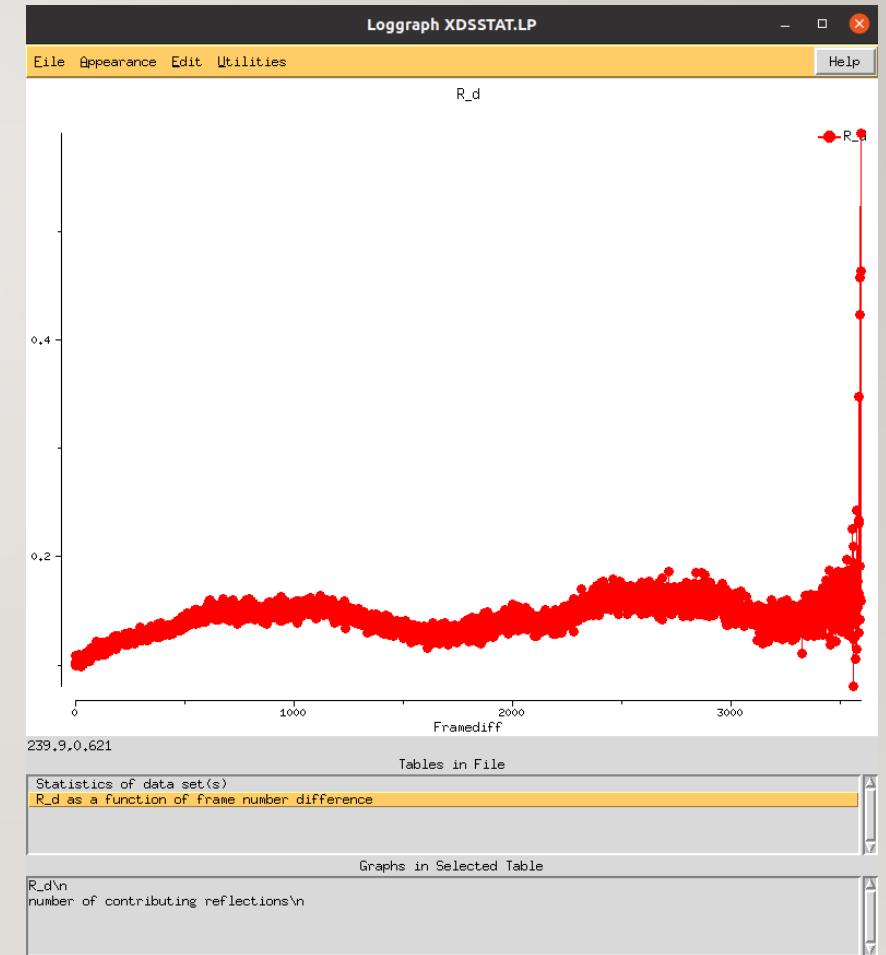
R_d : R-factors as a function of frame number difference

- The horizontal axis is the difference between frames.

Indicators for measuring radiation damage

- Flat → No damage
- Right shoulder upward → damage
- The variation at the end is not very important, since it contributes little to the reflection.

Data: 2019-11-25, PF-BL1A, $\lambda=2.7$ Å, exposure 0.1S



S-SAD Data collection protocol (PF-BLIA)

- 360-3600 deg. of data from multiple possible position under 100% transmittance, 0.1-0.5 deg./sec.
- XDS results

Resolution > 3.0 Å

Redundancy > 10

I/s(I) > 2.0

Sig_{ano} > 0.7

Decay <5% (delete Rmeas>1.5*avgRmeas)

Outline

- Before data collection
- During data collection
- After data collection

After data collection

- Single dataset
 - 1. Process data with XDS
 - 2. *Phenix.AutoSol or SHELX C/D/E*
 - 3. Phenix.AutoBuild
 - Multiple datasets
 - 1. Process every single dataset with XDS
 - 2. Merge data with XSCALE
 - 3. *Phenix.AutoSol or SHELX C/D/E*
 - 4. Phenix.AutoBuild
 - SHELX C: set up files for Shelx D
 - SHELX D: locate heavy atoms
 - SHELX E: phasing and density modification
- 
- 

After data collection

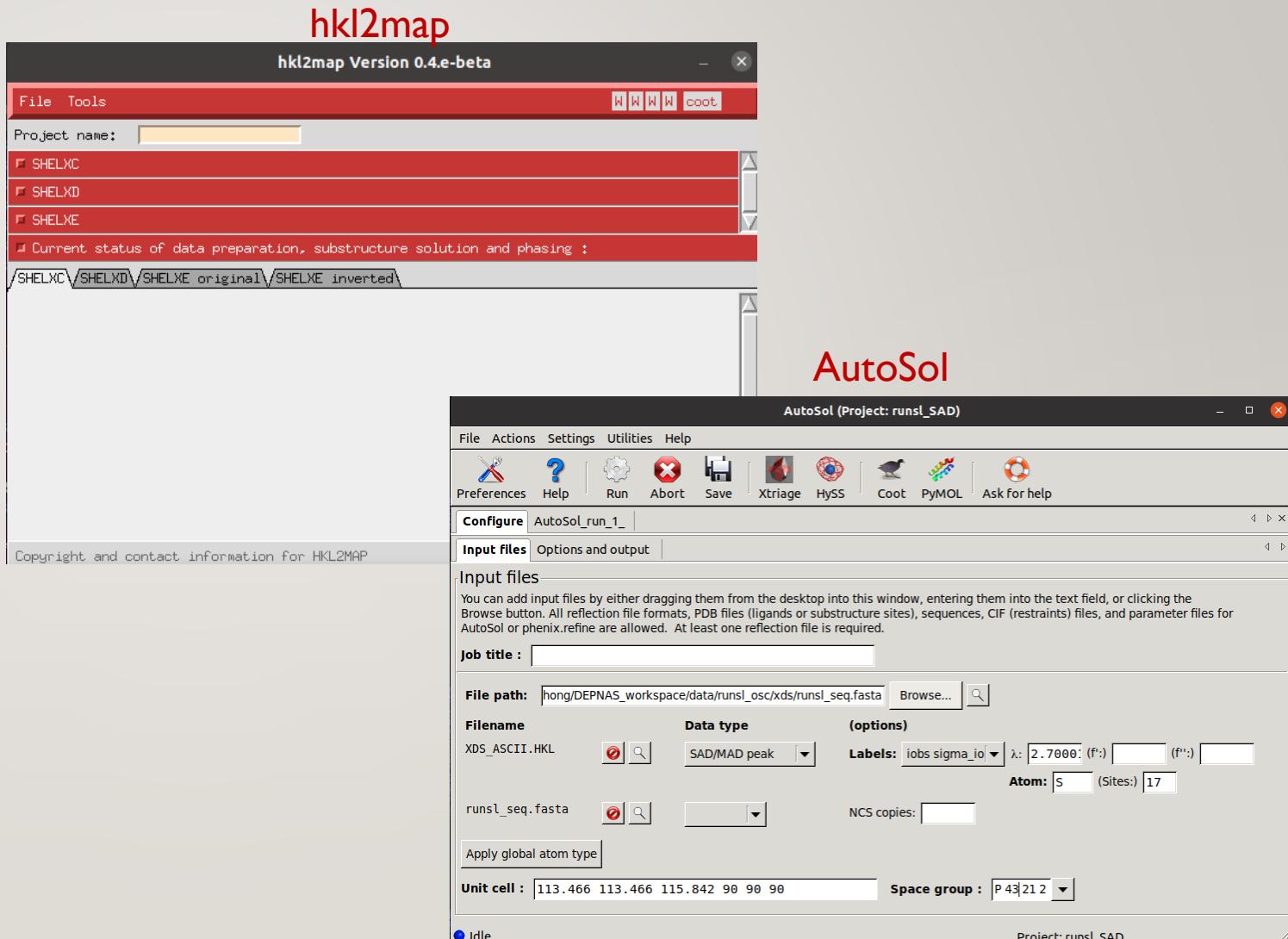
➤ Single dataset

1. Process data with XDS
2. *Phenix.AutoSol* or *SHELX C/D/E*
3. *Phenix.AutoBuild*

➤ Multiple datasets

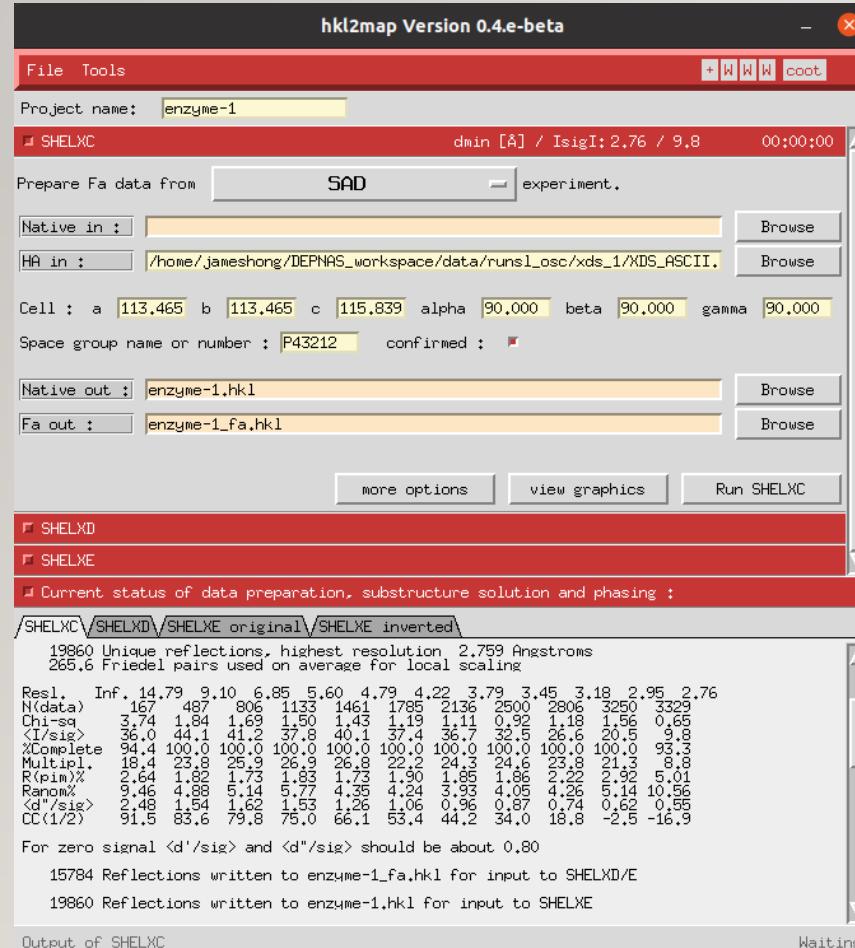
1. Process every single dataset with XDS
2. Merge data with XSCALE
3. *Phenix.AutoSol* or *SHELX C/D/E*
4. *Phenix.AutoBuild*

- *SHELX C*: set up files for *Shelx D*
- *SHELX D*: locate heavy atoms
- *SHELX E*: phasing and density modification



After data collection

➤ SHELXC: set up files for SHELXD



Data: 2019-11-25, PF-BLIA, $\lambda=2.7 \text{ \AA}$, exposure 0.1S

/SHELXC\SHELXD\SHELXE original\SHELXE inverted													
19860 Unique reflections, highest resolution 2.759 Angstroms													
265.6 Friedel pairs used on average for local scaling													
Resl.	Inf.	14.79	9.10	6.85	5.60	4.79	4.22	3.79	3.45	3.18	2.95	2.76	
N(data)	167	487	806	1133	1461	1785	2136	2500	2806	3250	3329		
Chi-sq	3.74	1.84	1.69	1.50	1.43	1.19	1.11	0.92	1.18	1.56	0.65		
$\langle I/\sigma \rangle$	36.0	44.1	41.2	37.8	40.1	37.4	36.7	32.5	26.6	20.5	9.8		
%Complete	94.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	93.3	
Multipl.	18.4	23.8	25.9	26.9	26.8	22.2	24.3	24.6	23.8	21.3	8.8		
R(pim)%	2.64	1.82	1.73	1.83	1.73	1.90	1.85	1.86	1.85	2.22	2.92	5.01	
Ranom%	9.46	4.88	5.14	5.77	4.35	4.24	3.93	4.05	4.26	5.14	10.56		
$\langle d''/\sigma \rangle$	2.48	1.54	1.62	1.53	1.26	1.06	0.96	0.87	0.74	0.62	0.55		
CC(1/2)	91.5	83.6	79.8	75.0	66.1	53.4	44.2	34.0	18.8	-2.5	-16.9		

There is an output table. It gives against resolution:

- The number of reflections read in
- The average intensity divided by its standard deviation σ
- Completeness (in %)
- d'' divided its estimated standard deviation σ , giving you a good indication for the strength of the anomalous signal, which should asymptote to 0.8 in the outer shell, if the data are processed well.
- The self-correlation coefficient for the anomalous signal. It should be above 25% for a significant signal.

SHELX D

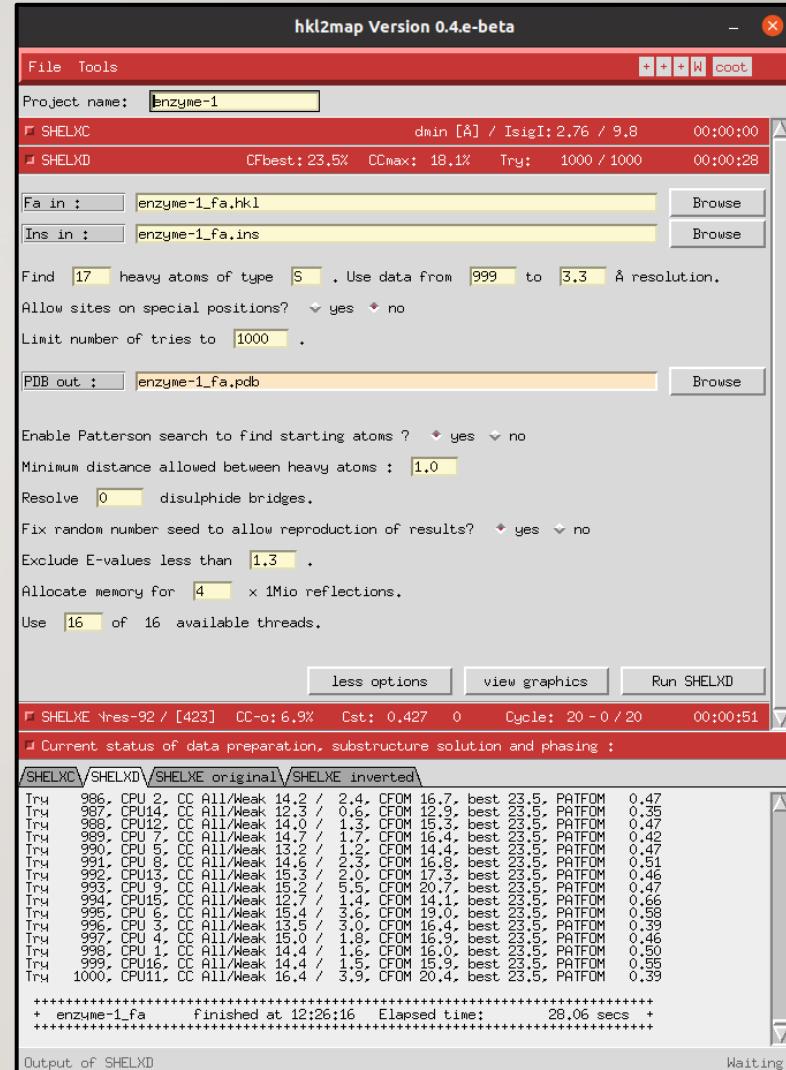
➤ SHELX D: locate heavy atoms

- Critical parameters for SHELXD

1. Truncation of the data at a particular resolution, typically in the range **3.0-3.5 Å**, can be critical to success.
2. The **resolution** at which the data are truncated, e.g. where the internal CC between the signed anomalous differences of two randomly chosen reflection subsets falls below 30%.
3. The **number of sites** requested should be within about 20% of the true value so that the occupancy refinement works well (and reveals the true number).
4. In difficult cases it may be necessary to run more trials (say 10000). The **multiple-CPU version** of SHELXD is recommended!

- Values to look at in SHELXD

1. A high value for the correlation coefficient CC and CC(weak) indicates a correct solution (**eg. 30 and 15**).
2. The CC-values of correct solutions are usually well separated from the ones from wrong solutions.
3. The best way to check is to run SHELXE with the best solution and look at the electron density map.



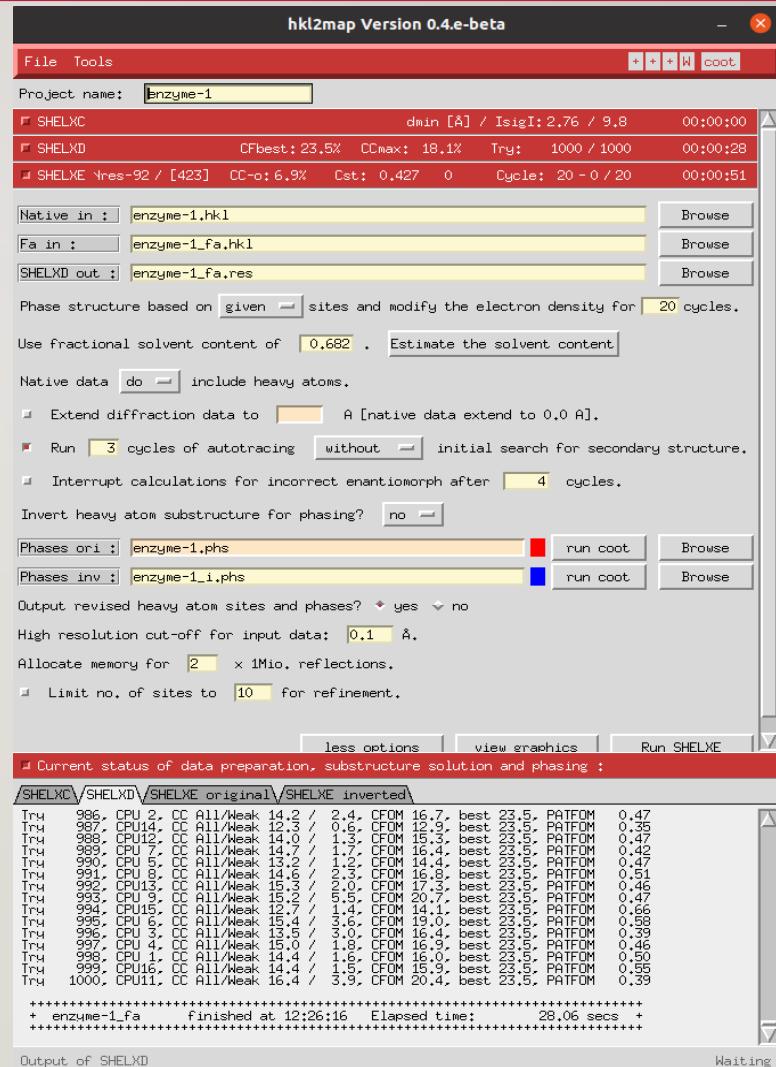
SHELXE

- **SHELXE:** phasing, density modification and auto-tracing of the protein backbone.

Critical parameters for SHELXE

- Number of cycles: 10 up to 200.
- Solvent content can be estimated by SHELXE
- Auto-tracing can help in difficult cases: 3 up to 20

The output files **name.pdb** (trace) and **name.phs** (phases) can be read directly into **COOT**.



Statistics of Native-SAD data

- In our group, Native-SAD data collected at PF/Spring-8 :
 - 1) 126 datasets; 2) 18 kinds of protein molecules

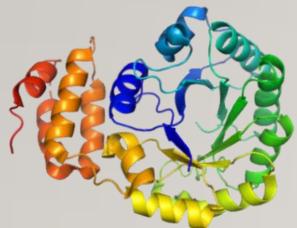
- Successful phasing:

- 32 datasets
- Statistics by XDS(average of 32 datasets) :
 - 1) completeness : 93%(73%)
 - 2) Redundancy : 20 (15)
 - 3) $I/\sigma(I)$: 36 (13)
 - 4) SigAno : 1.8(1.2)
- Results from SHELXD
 - CC all > 30
 - CC weak > 15

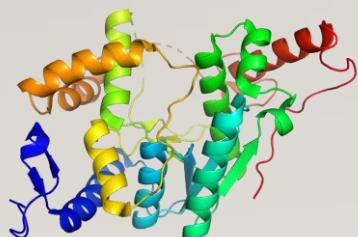
	Successful	Supposed to be successful	Failed
No. of dataset	32	14	80
redundancy	20.1 (14.9)	21.2(13.5)	11.6(7.8)
completeness	93.7(73.2)	95.1(76.3)	88.7(67.5)
$I/\sigma(I)$	36.4 (13.2)	28.5(8.3)	15.2(3.8)
SigAno	1.83(1.20)	1.48(1.23)	1.08(0.82)
CC all /CC weak	39.8/21.5	24.9(10.4)	21.4/8.7
$\langle \Delta F >/\langle F \rangle$ ($2.7\text{\AA}/1.9\text{\AA}$)	2.27/0.97	1.815/0.88	1.84/1.08

The structures solved by Native-SAD in our group

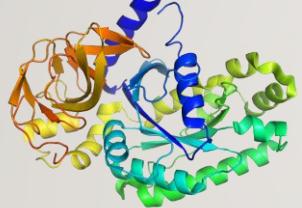
DusC (+1P)
MW: 36.2 kDa
S ratio: 12/322=3.7%
 $\langle|\Delta F|/\langle F\rangle: 1.19\%$
SG: P4₃2₁2 (Z'=1)
Resolution: 2.3 Å@1.9 Å
Data: 360°+600°
Model building: 97.8%



TtuA (+2Zn, 3P, 4Fe-4S)
MW: 36.21 kDa
S ratio: 15/321=4.7%
 $\langle|\Delta F|/\langle F\rangle: 2.48\%$
SG: P6₁2₂ (Z'=1)
Resolution: 3.3 Å@2.7 Å
Data: 180°x4-360°
Model building: 59%



EF2-domain I-II (+3P)
MW: 43.3 kDa
S ratio: 21/386=5.4%
 $\langle|\Delta F|/\langle F\rangle: 1.44\% \text{ or } 2.67\%$
SG: P3₂1 (Z'=1)
Resolution: 2.5 Å@1.9 or 2.7 Å
Data: 720°
Model building: 97.7%



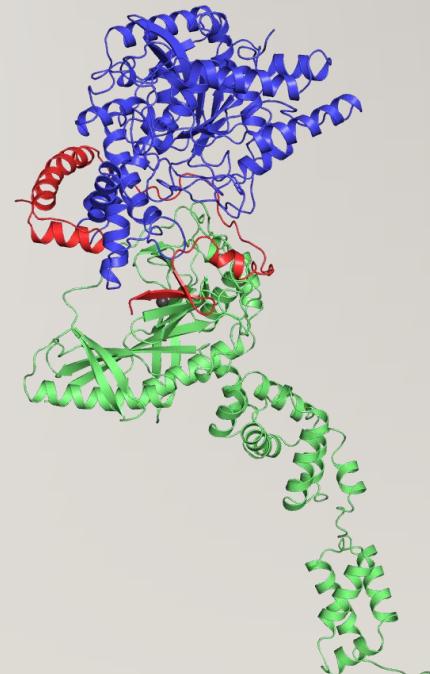
Kfla1896 (New)
MW: 43.97 kDa
S ratio: 9/405=2.2%
 $\langle|\Delta F|/\langle F\rangle: 1.1\% @ 2.1 \text{ Å}$
SG: P4₂1₂ (Z'=1)
Resolution: 2.5 Å
Data: 720° x 5
Model building: 84.7%



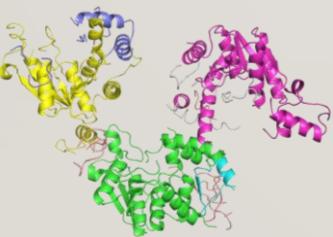
Protein (New·1st Screening)
MW: 24.7 kDa
S ratio: 22/229 = 9.62%
 $\langle|\Delta F|/\langle F\rangle: 2.27\% @ 2.1 \text{ Å}$
SG: P2₁2₁2 (Z'=1)
Resolution: 2.34 Å
Data: 720° x 8
Model building: 58%



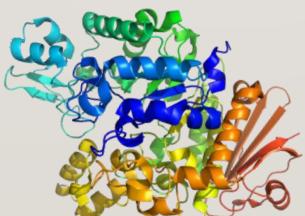
GatCAB (+2Mg)
MW: 117.7 kDa
S ratio: 32/1060 = 3.0%
 $\langle|\Delta F|/\langle F\rangle: 1.07\%$
SG: P2₁2₂1 (Z'=1)
Resolution: 2.2 Å@1.9 Å
Data: 180°x15 (90°shift)-112°
Model building: 77%



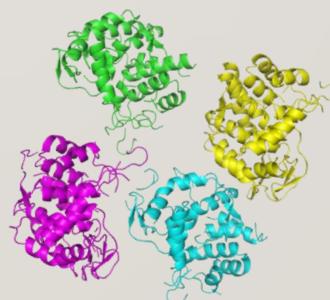
TtuAB (+2Zn, 4Fe-4S)
MW: 43.52 kDa
S ratio: 15/386=3.9%
 $\langle|\Delta F|/\langle F\rangle: 1.79\%$
SG: C2 (Z'=3)
Resolution: 2.8 Å@2.7 Å
Data: 720°x8
Model building: 74.4%



SmDG (+4Ca)
MW: 62.9 kDa
S ratio: 22/543 = 4.1%
 $\langle|\Delta F|/\langle F\rangle: 1.24\%$
SG: P2₁2₁2₁ (Z'=1)
Resolution: 2.4 Å@1.9 Å
Data: 180°x3(180°shift)
Model building: 79%



CesZ
MW: 41.2 kDa
S ratio: 7/346=2.0%
 $\langle|\Delta F|/\langle F\rangle: 1.63\%$
SG: P2₁ (Z'=4)
Resolution: 2.7 Å@2.7 Å
Data: 360°x10
Model Building: 85.3%



Example I --- Easy case

Protein A: 423 aa, 17 S atoms, 49 kDa

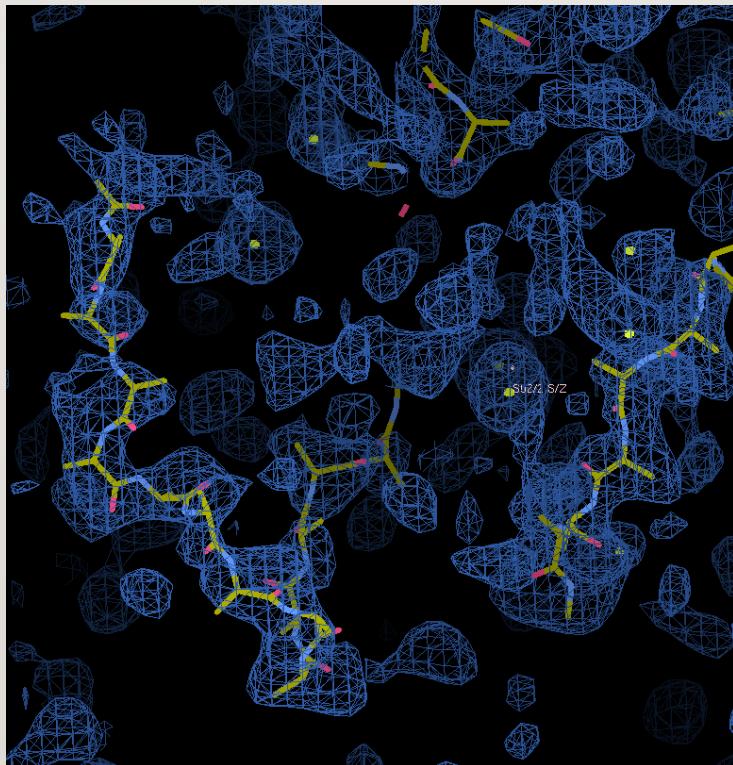
Phenix.AutoSol: 46 min with 8C16T Intel i7 CPU and 32G RAM

Phenix.AutoBuild: 2h37m; Model building 99%; R/Rfree 0.16/0.20

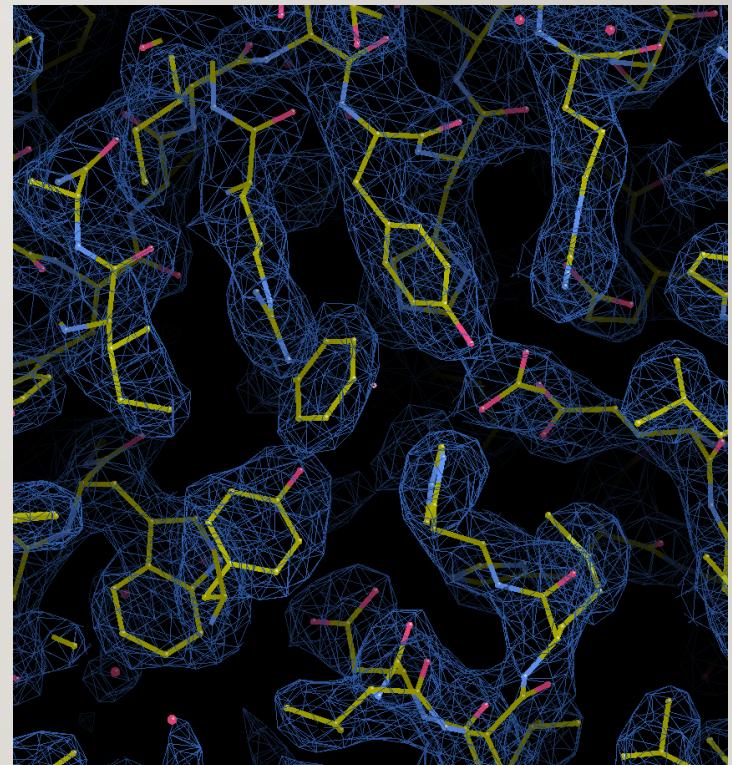
XDS statistics

No. of dataset	I
Wavelength	2.7 Å
Resolution	2.84 Å
Redundancy	11.6 (4.5)
Completeness	98(87.6)
I/s(I)	21.1 (6.9)
SigAno	0.94(0.58)
CCall / CCweak	16.3/7.1
$\langle \Delta F \rangle / \langle F \rangle$	2.23

SHELX C/D/E initial map



Phenix.AutoBuild



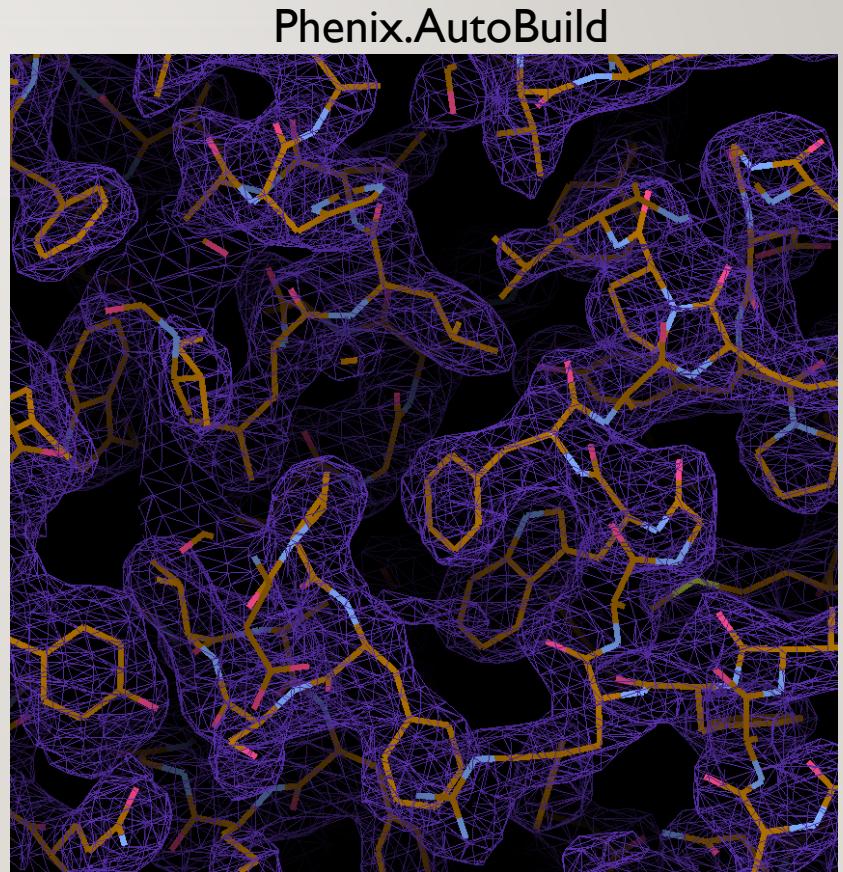
Example 2 --- Difficult case

Protein B: 350 aa, 7 S atoms, 41.2 kDa

Phenix.AutoSol: 6hours with 8C16T Intel i7 CPU and 32G RAM

Phenix.AutoBuild: 1 Day; Model building 85.3%; R/Rfree 0.25/0.28

XDS statistics			
No. of dataset	I(1*360)	6(6*720)	10(10*360)
Wavelength	1.9 Å	1.9 Å	2.7 Å
Resolution	2.11 Å	2.8 Å	2.7 Å
Redundancy	6.6 (6.2)	44.1 (26.0)	30.9(10.5)
Completeness	99.5(97.9)	99.9 (99.4)	91.4(32.3)
I/s(I)	9.39 (0.73)	8.1 (4.3)	43.6(8.8)
SigAno	0.88(0.67)	0.99 (1.2)	1.55(1.04)
CCall / CCweak	18.9/5.1	16.6/5.3	27.6/13.8
< ΔF >/<F>	0.88	0.88	1.63



S-SAD Data collection protocol (PF-BLIA)

- 360-3600 deg. of data from multiple possible position under 100% transmittance, 0.1-0.5 deg./sec.
- XDS results
 - Resolution > 3.0 Å
 - Redundancy > 10
 - I/s(I) > 2.0
 - Sigano > 0.7
 - Decay < 5% (delete frames Rmeas > 1.5 * avgRmeas)
- SHELX C/D result:
 - CCall/CCweak > 30 / 15

THANKS FOR
YOUR
ATTENTION

