

The Evaluation of Sulfur's Anomalous Signal in Diffraction Data and Beyond

Zhi-Jie Liu

National Lab of Biomacromolecules

Institute of Biophysics, Chinese Academy of Sciences

**4th Winter School on Soft X-rays in Macromolecular Crystallography
Grenoble, France, February 6-8, 2012**



Outline

- (1) Background
- (2) Evaluate the readiness of data collection systems
- (3) Evaluate the anomalous data quality
- (4) A better way to collect high quality anomalous data
- (5) A powerful tool for S-SAD phasing

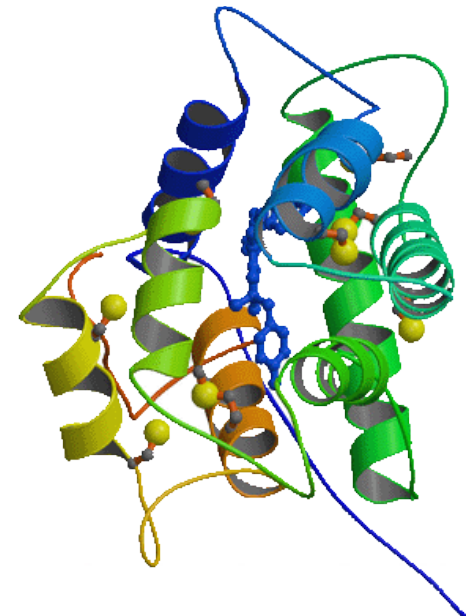
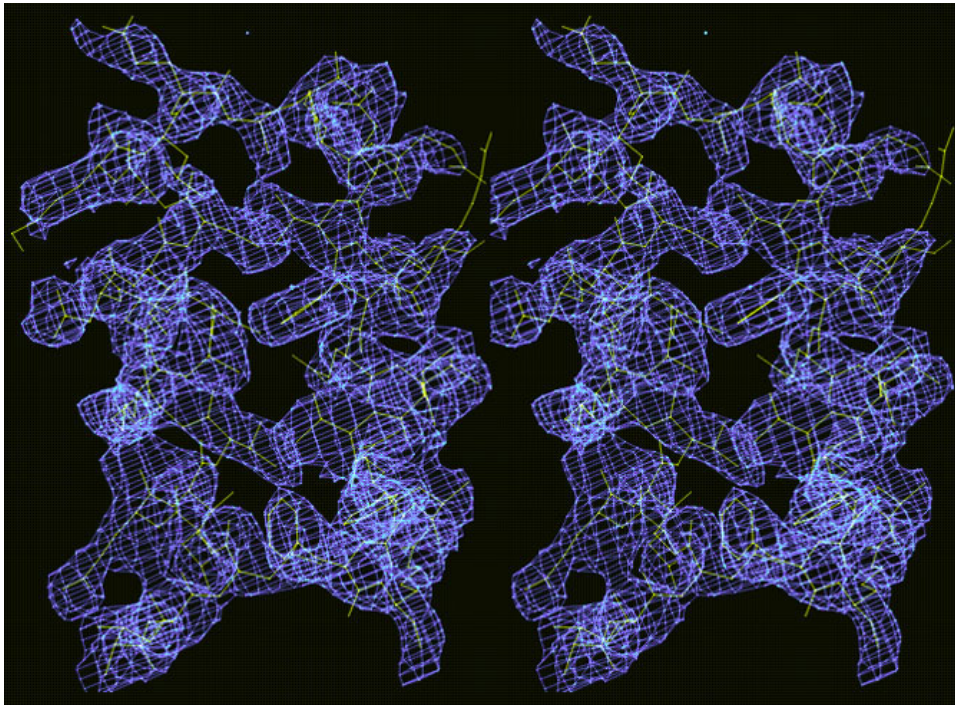
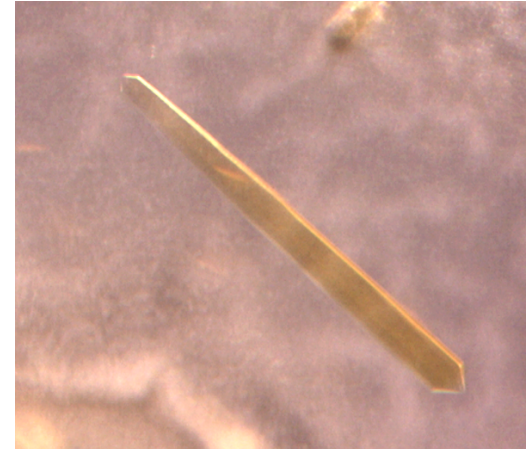


Background



1999: S-SAD New Structure Obelin

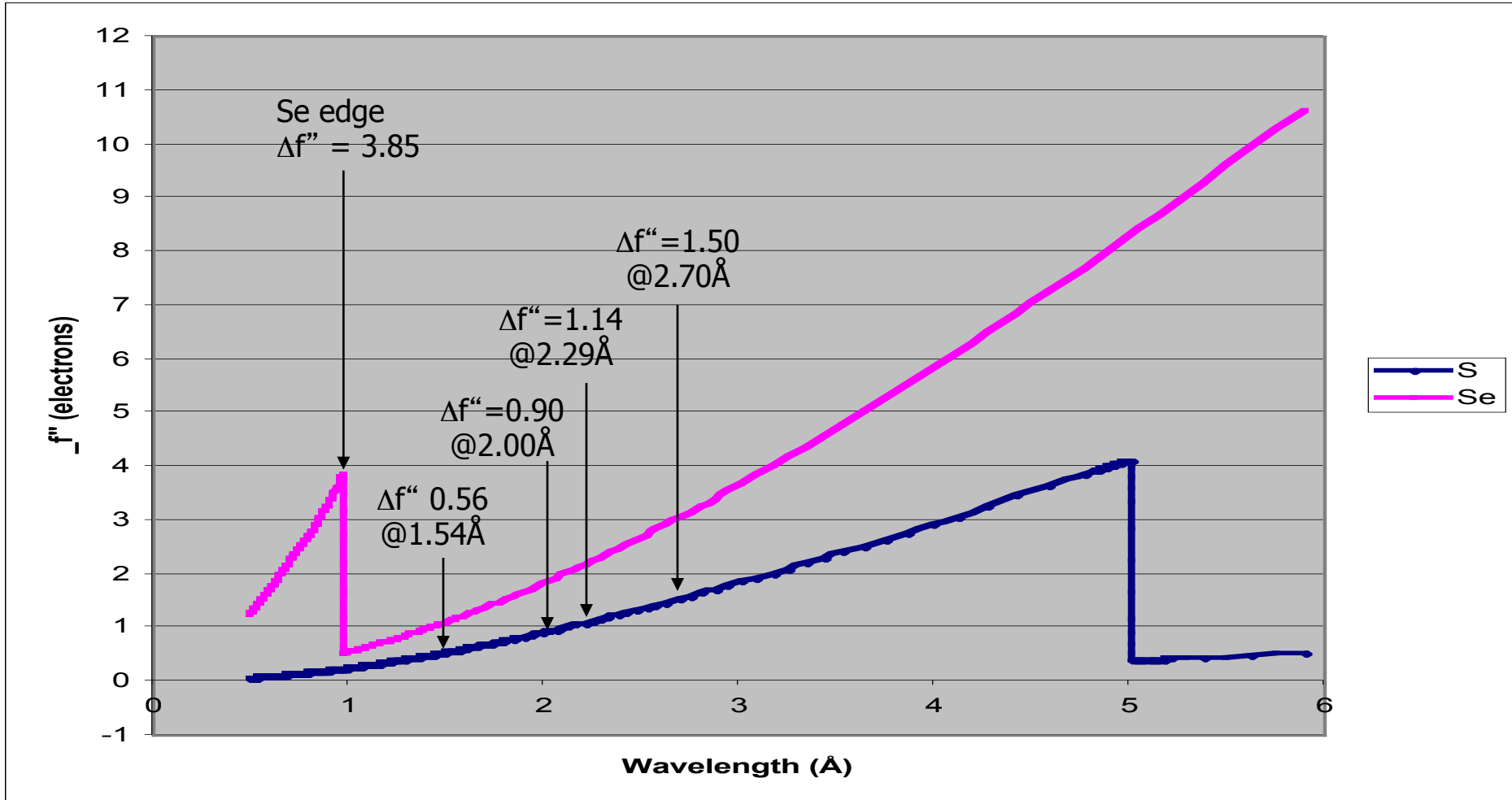
Molecular Weight:	22.2 kD
Diffraction limit:	1.70Å
Heavy atoms:	8S, 1Cl
X-ray:	1.74Å (17ID)
Phasing Resolution:	3.0Å



Background

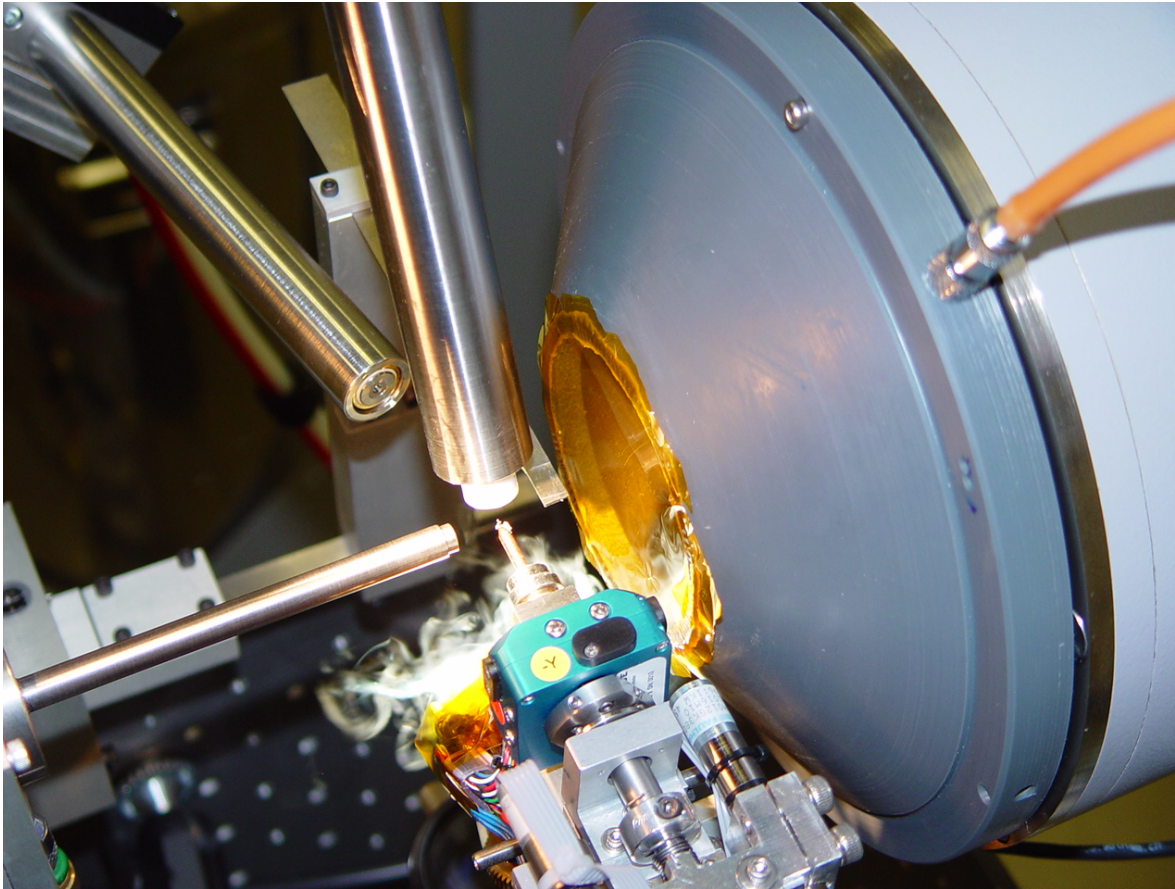


Anomalous signal of sulfur at different wavelengths



Background

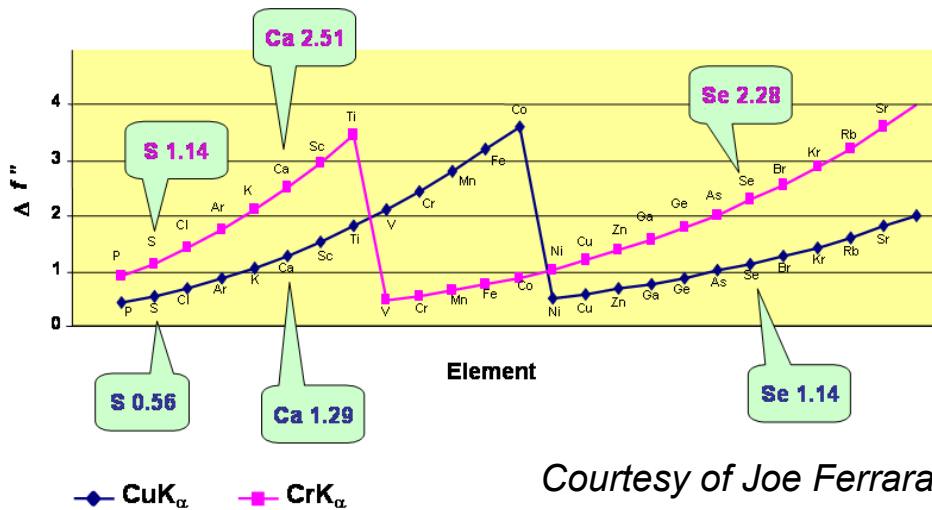
Sulfur SAD phasing with
soft synchrotron X-rays



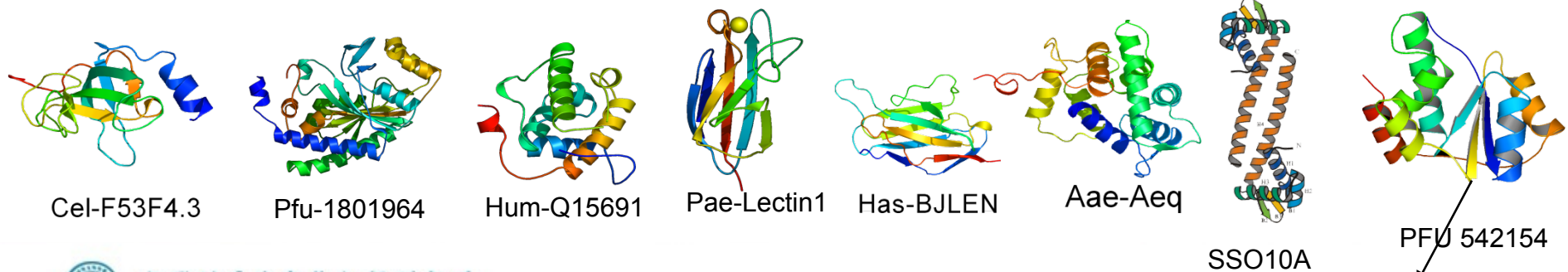
Wavelength: 2.0\AA

Background

Structures determined by Sulfur-SAD phasing using Cr X-rays



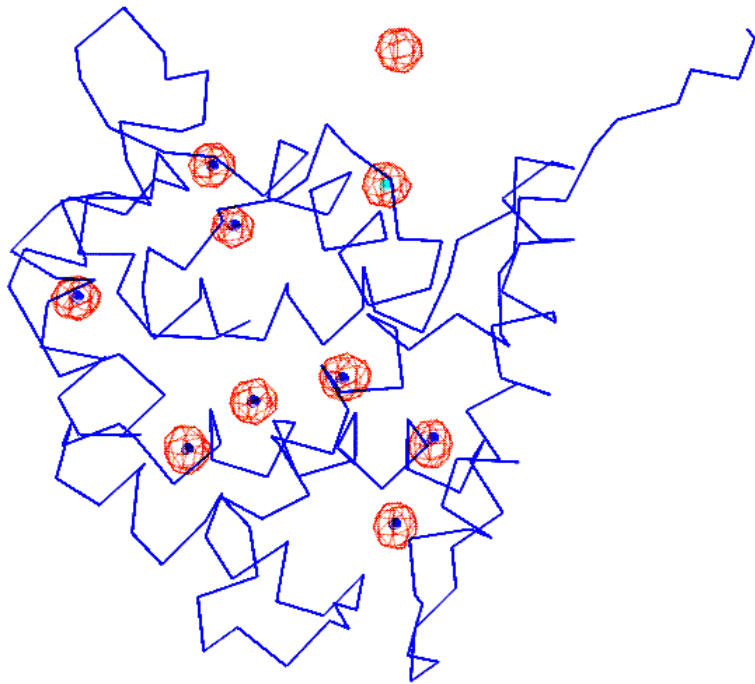
Courtesy of Joe Ferrara



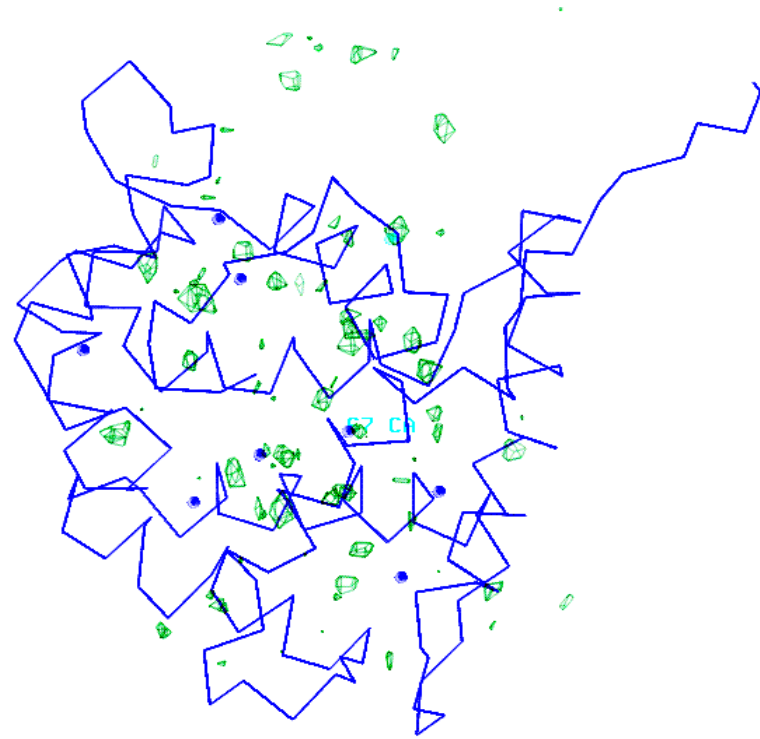
Evaluate the readiness of data collection systems



What difference a data collection system could make?



Data collection system 1



Data collection system 2

Evaluate the readiness of data collection systems



- (1) The advantage of synchrotron x-ray beamlines
High flux, tunable wavelength, monochromatic;
- (2) Synchrotron x-ray beamlines is a very complicated system and it is composed of thousands of optical, electronic and mechanical parts. It is a great challenge to keep them at peak performance simultaneously;
- (3) A monitoring mechanism is need to access the readiness of data collection systems.

Evaluate the readiness of data collection systems



What could go wrong with the beamline?

X-Ray beam: instability of beam intensity, beam position, divergence etc.

Goniometer system: mechanical accuracy of goniometer and shutter synchronization.

Detector system: Dark current correction, balance of different mosaic chips, sensitivity, dynamic range, DQE, cooling T.

Evaluate the readiness of data collection systems



What could go wrong with the user?

X-Ray beam: Choice of wavelength, beam size, attenuation

Goniometer system: Crystal centering.

Detector system: Detector distance, exposure time, oscillation width, scan range.





Evaluate the readiness of data collection systems

A quick and easy way to evaluate the readiness of data collection systems

Requirements:

1. Evaluate the readiness of the entire data collection system
2. Quick and easy to use
3. Objective and consistent



A quick and easy way to evaluate the readiness of data collection systems



Solutions:

1. Evaluate the readiness of the entire data collection system

Use sulfur's anomalous signal in collected data

2. Quick and easy to use

Use Zn-free insulin crystal

3. Objective and consistency

Develop objective parameters from the collected data.

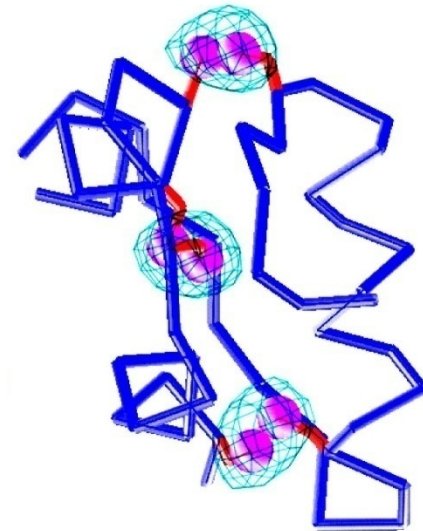


A quick and easy way to evaluate the readiness of data collection systems



Why Zn-free insulin?

- High symmetry ($I2_13$): quick data collection
- High crystal quality: reliable
- Cheap, easy to grow crystals



A quick and easy way to evaluate the readiness of data collection systems



Parameters used in the readiness test:

1. Relative peak height in anomalous difference Fourier map

$$\text{RPH} = \text{AVG}(\text{peak1} + \text{peak2} + \text{peak3}) / \text{AVG}(\text{peak7} + \text{peak8} + \text{peak9})$$

2. Mean phase difference between model phases and

S-SAD phases: $\Delta\phi$

2. Map correlation coefficient between model phased 2Fo-Fc

map and S-SAD map: MapCC

A quick and easy way to evaluate the readiness of data collection systems



Systems used in readiness test:

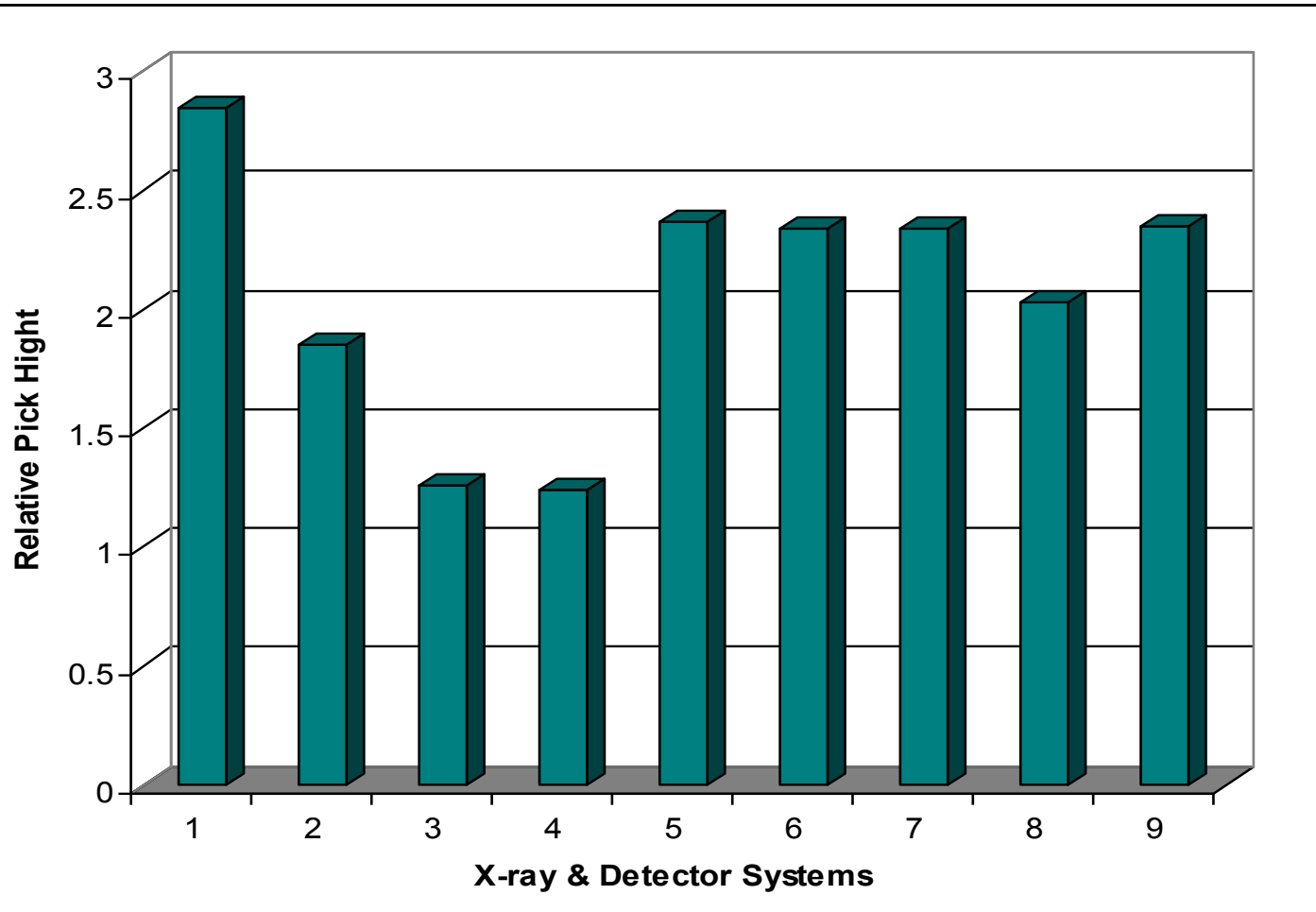
X-ray System	Detector	X-ray source	Wavelength(Å)
1	CCD1	Synchrotron1	1.74Å
2	CCD1	Home X-ray Cu	1.54Å
3	CCD2	Synchrotron2	1.74Å
4	CCD3	Synchrotron1	1.74Å
5	CCD4	Home X-ray Cu	1.54Å
6	Image Plate	Home X-ray Cu	1.54Å
7	Image Plate	Home X-ray Cr [§]	2.29Å
8	CCD2	Synchrotron3	1.70 Å
9	CCD4	Home X-ray Cu	1.54Å

§: Chromium anode.

A quick and easy way to evaluate the readiness of data collection systems



Relative Peak Height in Anomalous Difference Fourier Map:

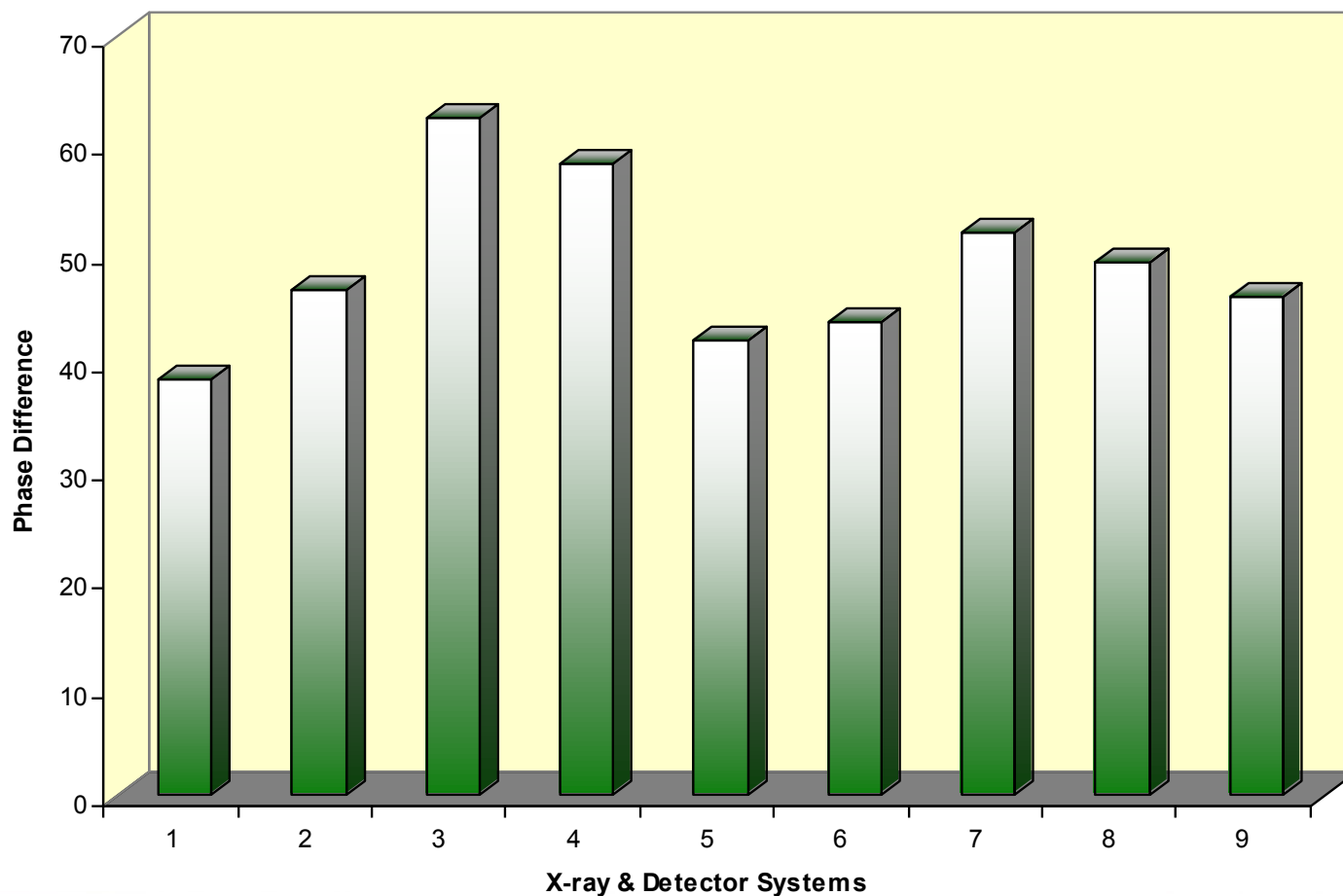


- 1.CCD1, Syn1
- 2.CCD1, Cu1
- 3.CCD2, Syn2
- 4.CCD3, Syn1
- 5.CCD4, Cu1
- 6.ImgPL, Cu1
- 7.ImgPL, Cr
- 8.CCD2, Syn3
- 9.CCD4, Cu2

A quick and easy way to evaluate the readiness of data collection systems



Phase different in the readiness test:



A quick and easy way to evaluate the readiness of data collection systems



How can we make it quicker and easier?

A web based Readiness Server is being built at IBP

~10 min insulin crystal data, ~5 min process data, ~2 min submit data. Total 17 min obtain Readiness parameters.

Ins entrance

http://zjliulab.ibp.ac.cn/ins/phpweb/ins.html

Google

HHpred Robetta Phenix 科学网 CCP4BB Liu lab IBP Apple Yahoo! Google Maps YouTube Wikipedia

DT System Readiness

User Password

[Request New Account](#)



Evaluate the anomalous data quality



Objectives:

- Identify if the data is suitable for S-SAD
- *In situ* evaluation enables signal based data collection

Anomalous signal strength indicators:

- $d''/\text{sig}(d'')$: SHELX C
- CC (All/Weak): SHELX D
- $R_{\text{ano}}/R_{\text{pim}}$ ratio: R_{ano} (anomalous R factor) must be greater than R_{pim} (error in observed intensities). Weiss, Acta Cryst D, 2001
- RAS: $\Delta_a/\Delta_c = (\text{signal} + \text{noise}) / (\text{noise})$, Fu, Rose, Wang, Acta Cryst D, 2004



Structure determination using PF soft x-rays



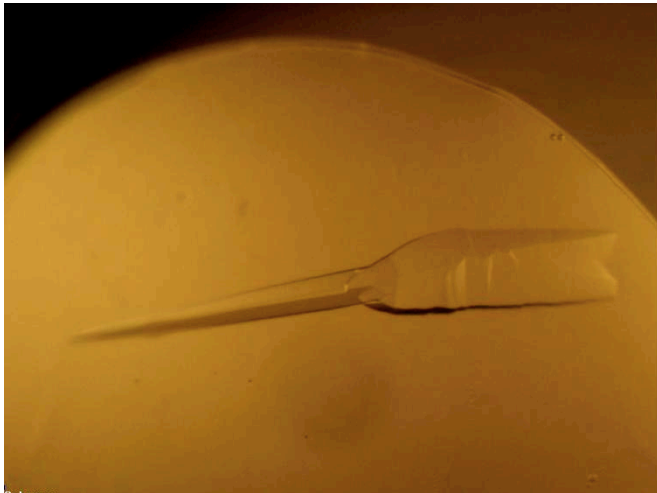
Case study:

MW: 40 kDa

of S: 4 Met, 9 disulfides

Resolution @home: 3.0Å

Space group: P6₁22



Data collection

Beamline: PF 17A, 1A

Wavelength: 2.0Å, 2.7Å,

Detector: ADSC 315 CCD

Oscillation: 0.5°, 0.75°

Scan range: 360°, 270°

Helium Path: Yes for $\lambda=2.7\text{\AA}$

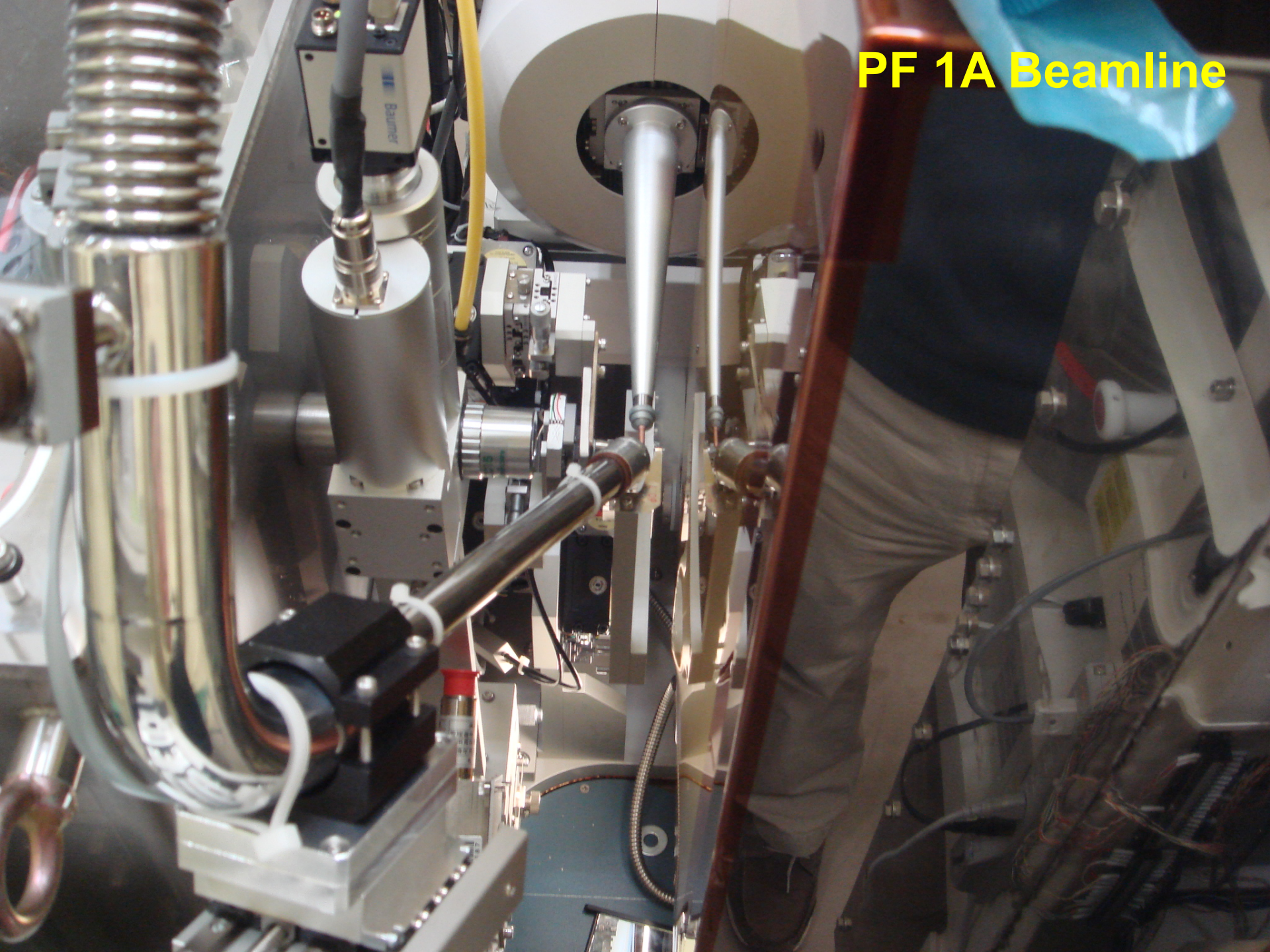
Cooling: Helium gas for

$\lambda=2.7\text{\AA}$

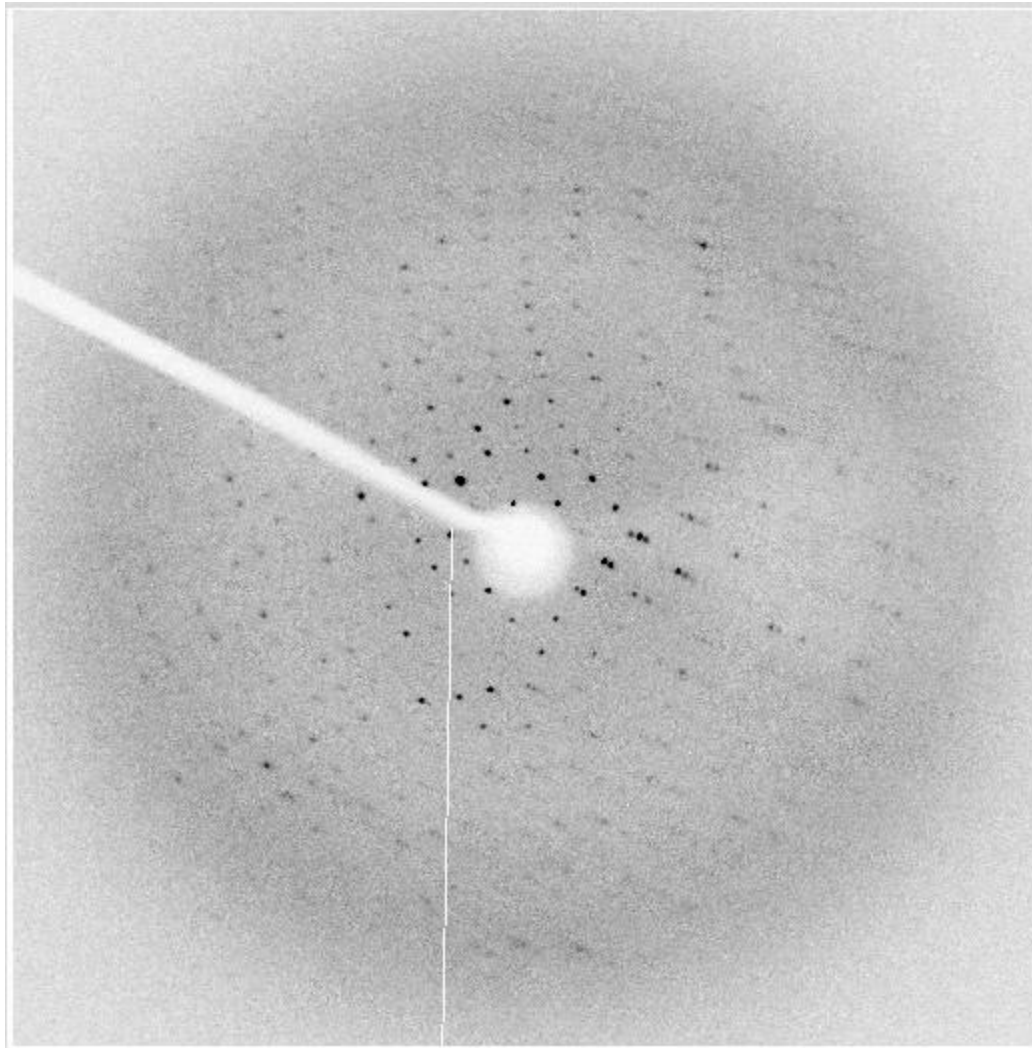
PF 1A Beamline



PF 1A Beamline



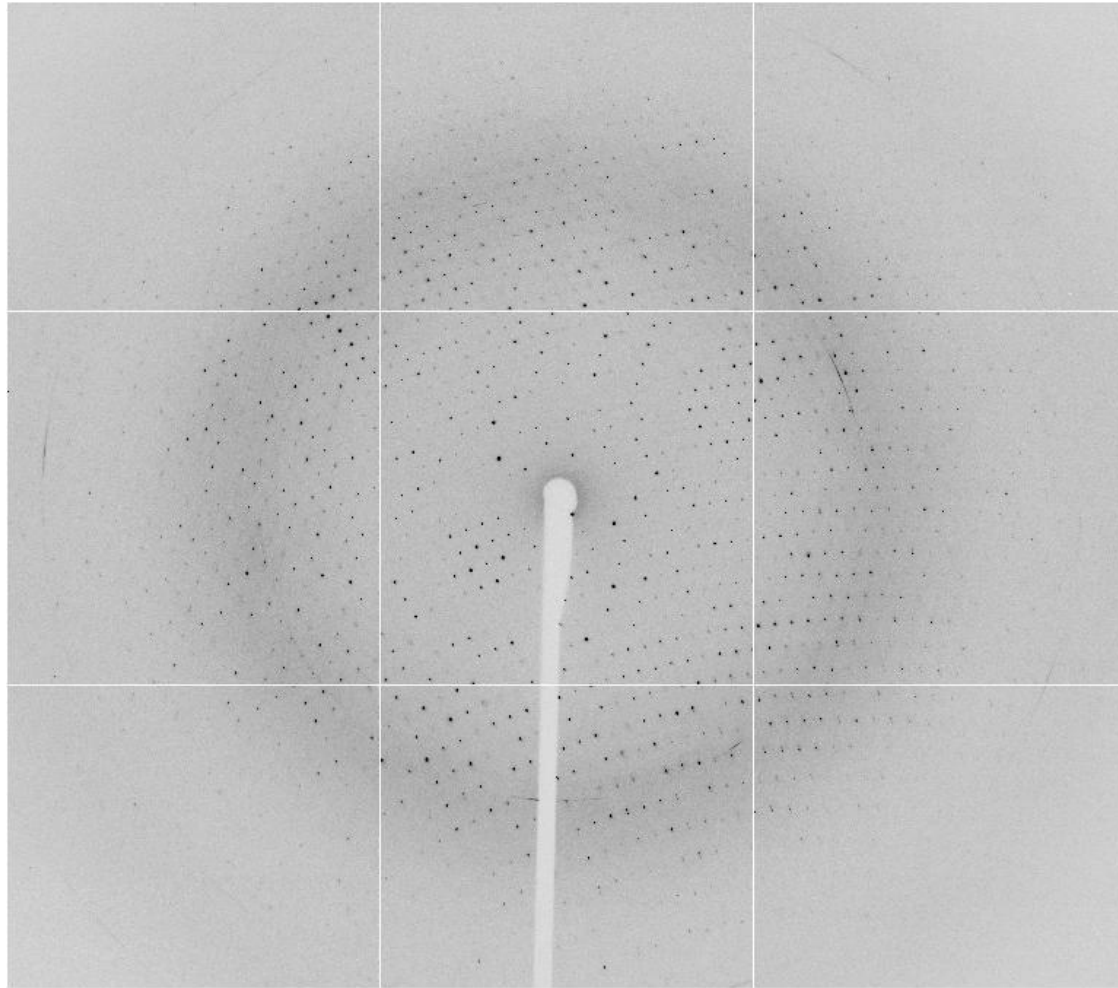
Structure determination using PF soft x-rays



Diffraction image @ home



Structure determination using PF soft x-rays



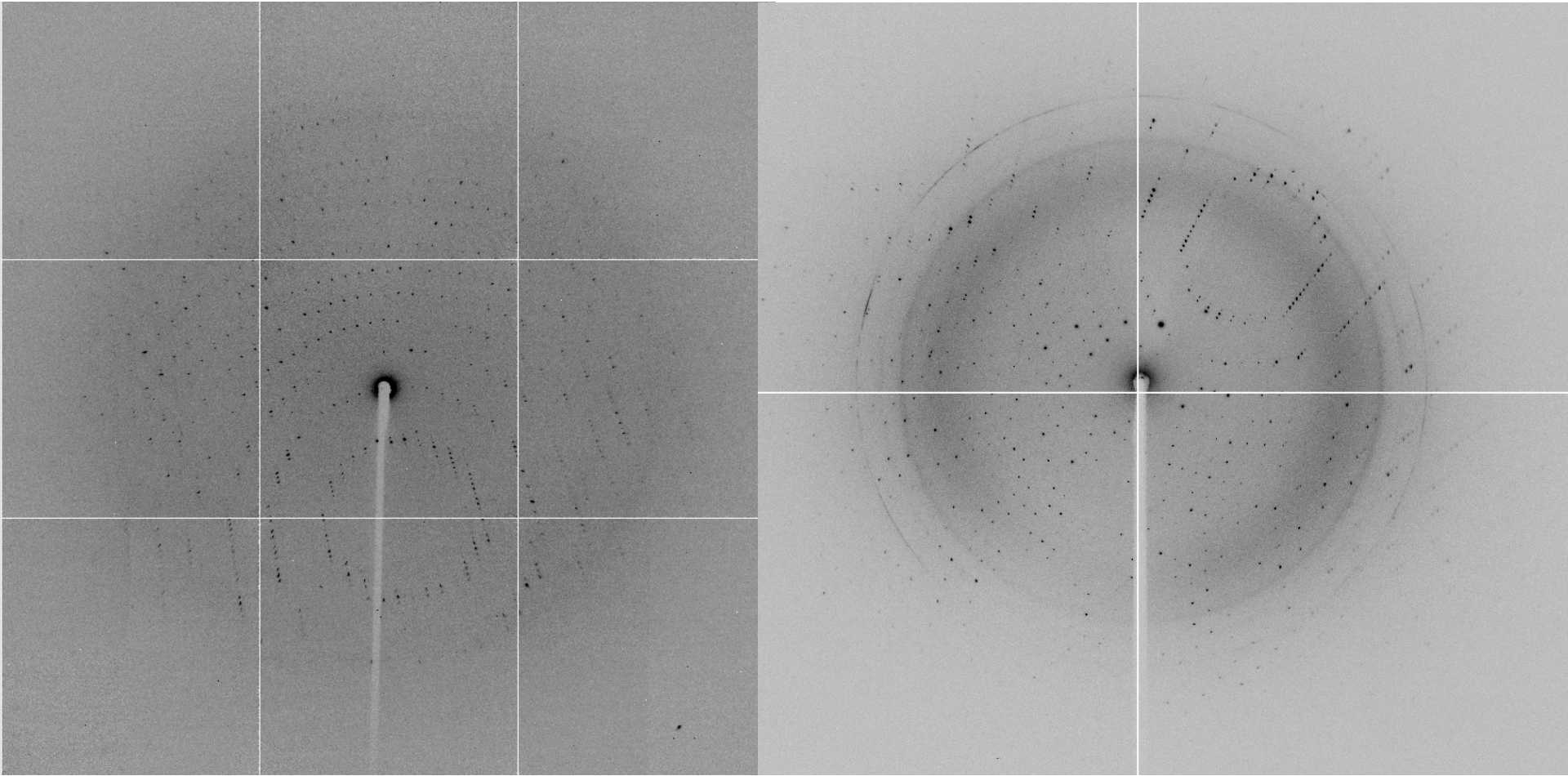
Diffraction image @ PF 17A



Structure determination using PF soft x-rays



Comparison between 2.0Å and 2.7Å wavelength

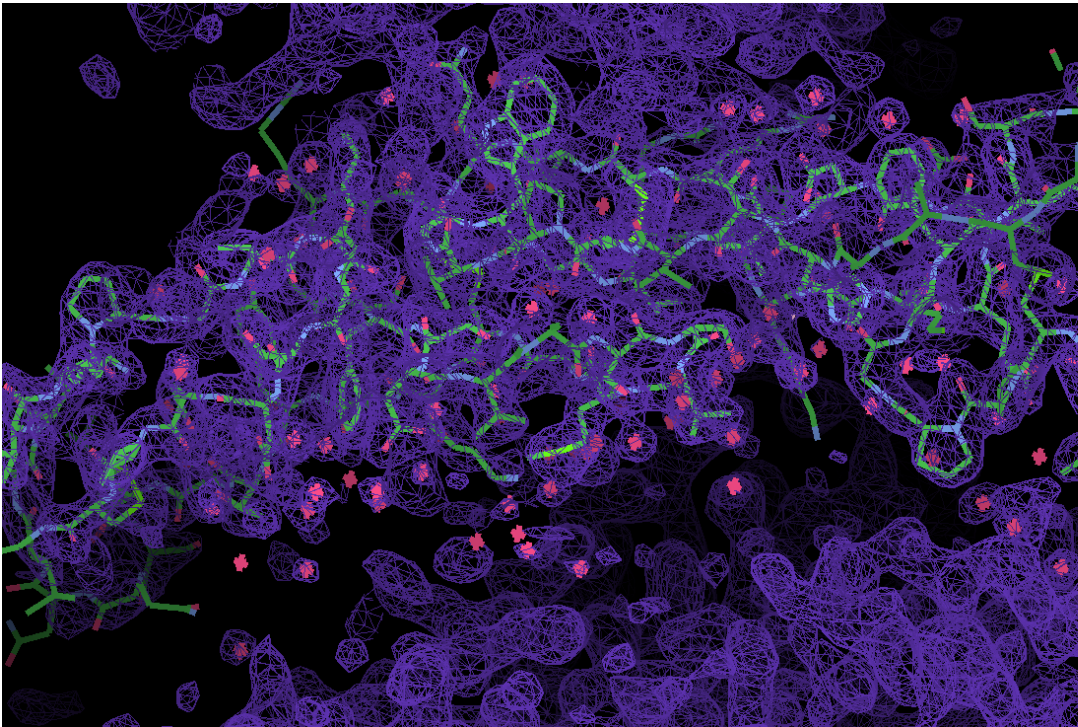


$\lambda=2.0\text{\AA}$

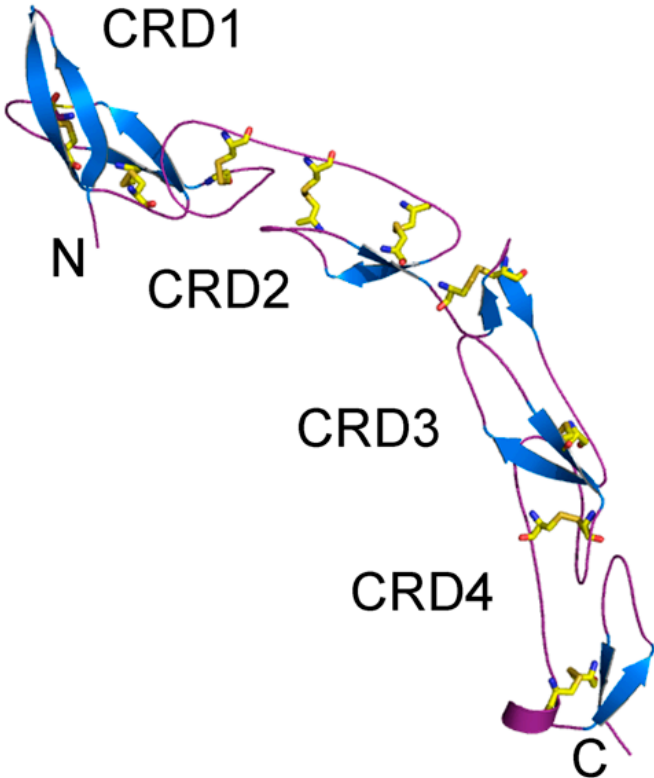
$\lambda=2.7\text{\AA}$



Structure determination using PF soft x-rays



Initial S-SAD phased map ($\lambda=2.0\text{\AA}$)



Evaluate the anomalous data quality



Data collection and processing

Crystal	A	A	B	C	D
Data Set	A1	A2	B1	C1	D1
X-ray source	PF17A	PF17A	PF17A	PF1A	PF1A
Distance (mm)	310	160	160	91	91
Number of images	360	360	360	248	281
Oscillation width (°)	0.5	1	1	1	0.75
Wavelength(Å)	0.98	2.00	2.00	2.70	2.70
Space group	P6 ₁ 22	P6 ₁ 22	P6 ₁ 22	P6 ₁ 22	P6 ₁ 22
a, c (Å)	77.86, 186.56	77.91, 186.42	77.68, 186.81	77.56, 187.41	77.47, 185.81
Mosaicity (°)	0.32	0.48	0.22	0.38	0.65
Resolution range(Å)	50.00-2.09 (2.16-2.09)	50.00-2.70 (2.82-2.70)	50.00-2.70 (2.82-2.70)	50.00-2.95 (3.08-2.95)	50.00-3.20 (3.34-3.20)
Rsym (%)	6.4(25.5)	7.5(37.3)	9.7(38.0)	11.4(42.4)	10.4(36.4)
Mean I/σ(I)	79.09(15.79)	80.782(21.0)	73.15(14.9)	60.70(12.76)	54.10(12.47)
Completeness (%)	99.9(100.0)	99.9(99.9)	99.9(99.7)	99.8(99.8)	99.8(99.9)
Redundancy	20.4	39.1	39.3	25.3	20.1



Evaluate the anomalous data quality



Calculations of different indicators

	A2	B1	C1	D1
Wavelength (Å)	2.00	2.00	2.70	2.70
Data Resolution (Å)	50.00-2.70	50.00-2.70	50.00-2.95	50.00-3.20
$\chi^{(1)}$	2.14%		3.61%	
$R_{\text{anom}}^{(2)}$	0.023	0.033	0.054	0.059
$R_{\text{p.i.m.}}^{(3)}$	0.015	0.033	0.049	0.051
$R_{\text{anom}}/R_{\text{p.i.m.}}$	1.53	1.00	1.10	1.16
$d''/\text{sig}(d'')^{(4)}$	1.80	1.32	1.15	0.88
CC (All/Weak) ⁽⁵⁾	47.40/23.42	30.16/9.55	35.79/11.22	34.19/10.83
$\Delta\theta^{(6)}$	66.56	69.73	70.90	74.04

(1): Estimated Bijvoet intensity ratio $\chi = \frac{\langle \Delta F_{\text{anom}} \rangle}{\langle F_{\text{p}} \rangle} = \left(\frac{2N_{\text{A}}}{N_{\text{p}}} \right)^{\frac{1}{2}} \times \left(\frac{\Delta f'_{\text{A}}}{Z_{\text{eff}}} \right)$

N_{A} : Number of anomalous scatters.

N_{p} : Approximate total number of non-hydrogen atoms.

Z_{eff} : the effective atomic number (~6.7 for non-hydrogen protein atoms).

(2): $R_{\text{anom}} = \frac{100 \times \sum_{\text{hkl}} |I(\text{hkl}) - I(-\text{h}-\text{k}-\text{l})|}{\sum_{\text{hkl}} \langle I(\text{hkl}) \rangle}$

(3): $R_{\text{p.i.m.}} = 100 \times \sum_{\text{hkl}} \left[\frac{1}{N-1} \right]^{1/2} \times \sum_i \left| \frac{I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle}{\sum_{\text{hkl}} \sum_i I_i(\text{hkl})} \right|$

(4): Calculated at 50.00 – 3.10 Å resolution range.

(5): Correlation coefficient in ShelxD.

(6): Average phase difference between model phases and S-SAD experimental phases

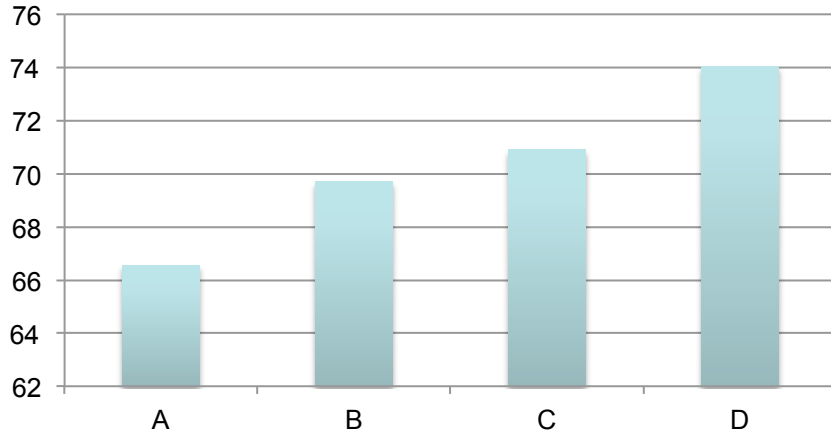


Evaluate the anomalous data quality

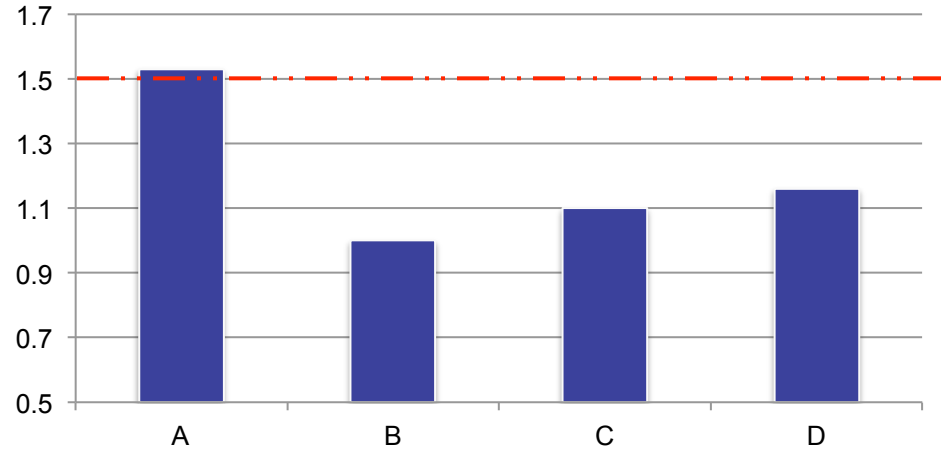


Comparison of indicators:

$\Delta\emptyset$



R_{ano}/R_{pim}

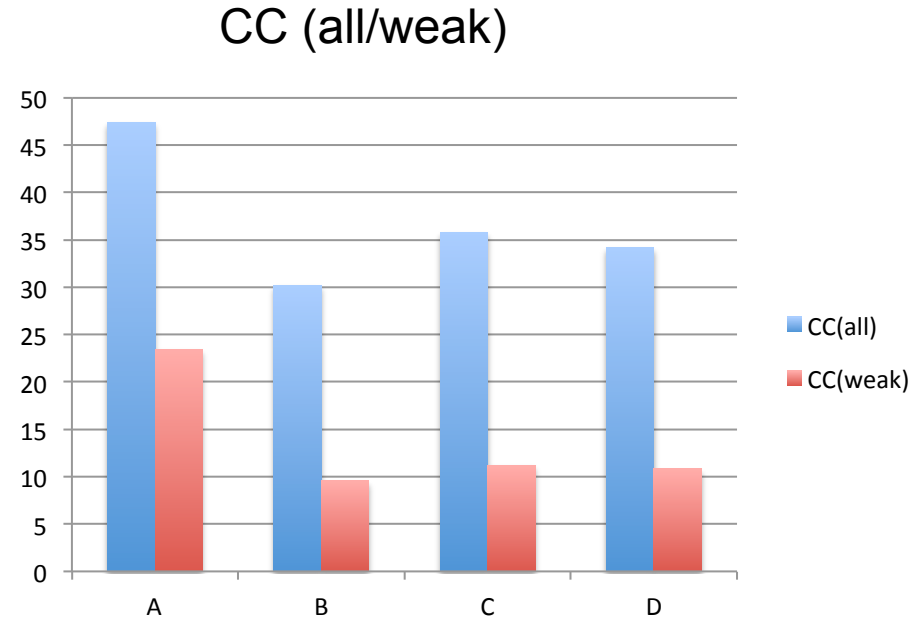
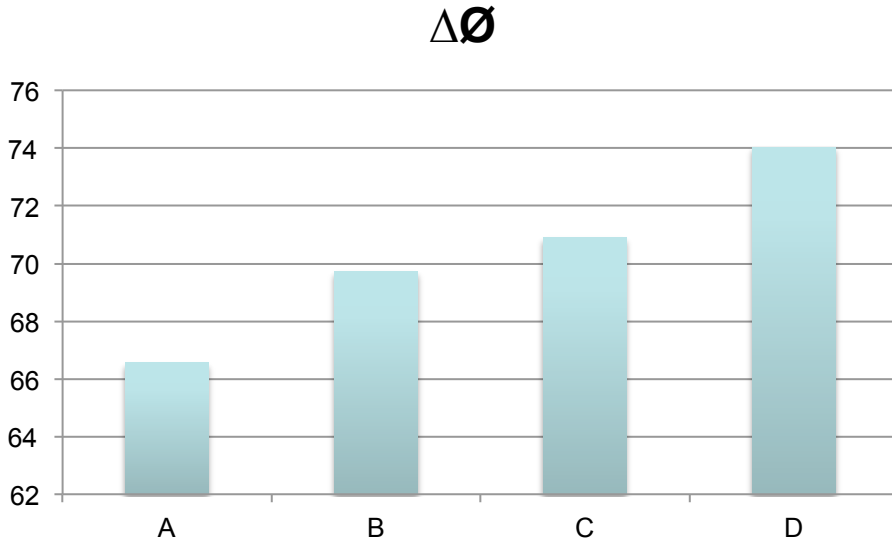


- R_{ano}/R_{pim} ratio followed Weiss prediction of **1.5** (Weiss, Acta Cryst D, 2001), only data A yield solution when S-SAD phasing was attempted manually.
- R_{ano}/R_{pim} ratio: Inconsistent with $\Delta\emptyset$ for weak anomalous data.

Evaluate the anomalous data quality



Comparison of indicators:

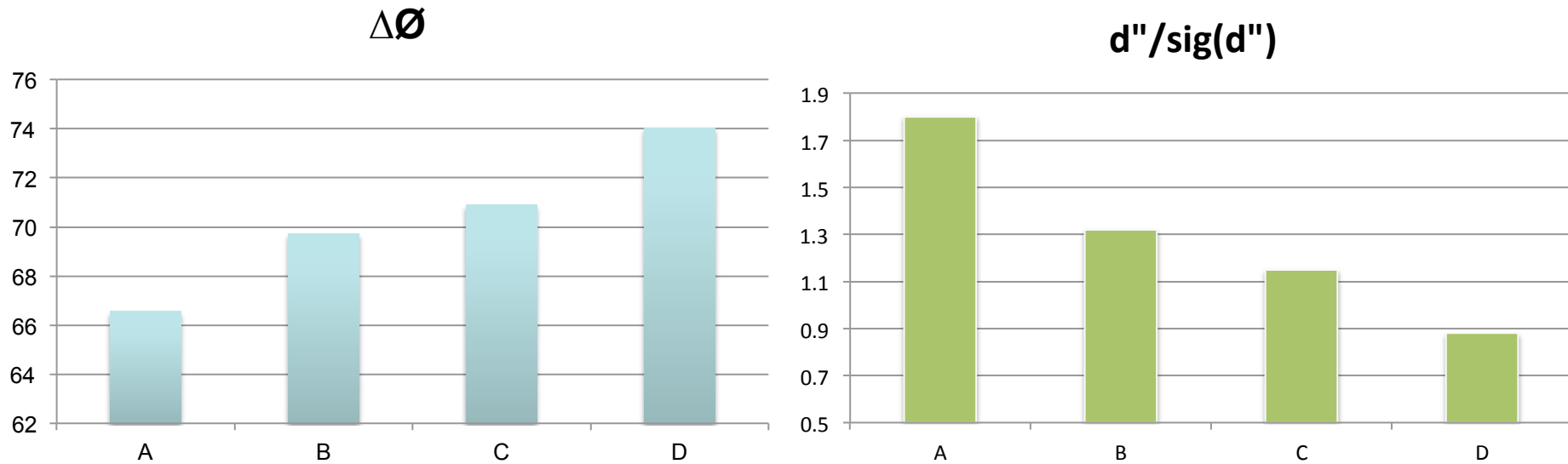


- CC (all/weak): Clear gap between “solved” and “unsolved” data.
- CC (all/weak): Inconsistent with $\Delta\emptyset$ for and weak anomalous data.

Evaluate the anomalous data quality



Comparison of indicators:



➤ $d''/\text{sig}(d'')$: Consistent with $\Delta\emptyset$ for both strong and weak anomalous data.



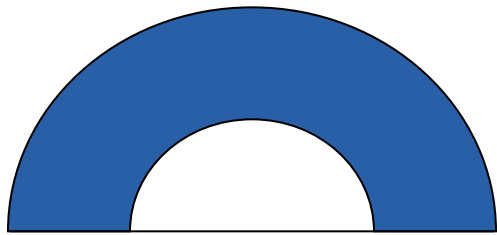


A better way to collect high quality anomalous data

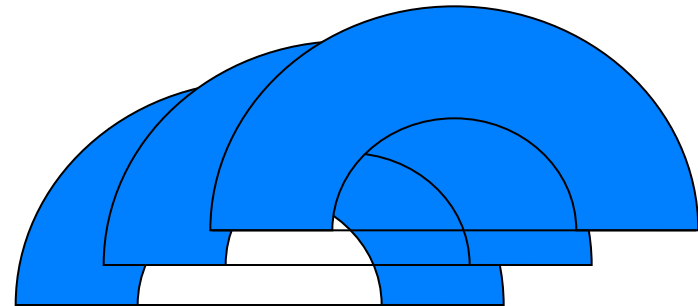
Multi-Data-Set (MDS) data collection strategy

- Traditional data collection strategy:
 - Exposure X seconds per frame scanning a total of Y degrees
- Multi-Data-Set (MDS) data collection strategy:
 - Exposure X/N seconds per frame scanning a total of Y degrees, but repeat the data collection N times

Total X-ray dosage is the same!



Exposure Time: X seconds



Exposure Time: X/N seconds
Repeat N times



Summary

Single data set approach:

$$\sigma_{total}^2 = G[I_s + I_{bg} + (m/n)I_{bg}] + m(K/A)^2 I_s^2$$

Multiple Data set (MDS) approach:

$$\sigma_{total}^2 = G[I_s + I_{bg} + (m/n)I_{bg}] + \frac{m(K/A)^2 I_s^2}{N}$$

This illustrates the theoretical advantage of applying the MDS approach over the traditional approach.

Test of the MDS strategy



Crystals: Zn free cubic insulin crystals

X-ray source: Rigaku MicroMax-007 Cu, SER-CAT 22-ID

Detector: Saturn 944⁺, Mar 225 CCD

Exposures: X seconds, X/3 seconds

Scan range: 50.0°, 50.0° x 3

Oscillation: 1.0°

Test of the MDS strategy



Table 1a Data collection and refinement statistics

	Crystal 1		Crystal 2		Crystal 3	
Data collection						
X-ray source	Rigaku MicroMax-007					
X-ray Optics	VariMax HR					
Detector	Rigaku Saturn 944+					
Wavelength (Å)	1.5419					
Space group	I2 ₁ 3					
Cell dimensions:						
<i>a</i> = <i>b</i> = <i>c</i> (Å)	77.96		77.59		78.42	
Exposure (s) [§]	45	15	45	15	45	15
Scan range (°)	50	3 X 50	50	3 X 50	50	3 X 50
Resolution(Å)*	50.00-2.00 (2.07-2.00)	50.00-2.00 (2.07-2.00)	50.00-1.95 (2.02-1.95)	50.00-2.10 (2.18-2.10)	50.00-1.95 (2.02-1.95)	50.00-2.10 (2.18-2.10)
R _{sym} (%)	5.3(22.7)	5.5(44.5)	4.8(33.7)	6.9(38.8)	3.9(23.5)	5.8(48.8)
I/σI	47.84(6.4)	66.21(6.18)	39.60(4.71)	53.16(10.06)	42.07(5.58)	51.52(5.17)
Completeness (%)	99.6(99.8)	99.8(100.0)	93.5(61.1)	98.4(90.9)	99.8(100.0)	98.4(90.9)
Redundancy	5.3	16.0	5.5	16.9	5.2	15.5
Refinement						
Resolution (Å)	50.0-2.5		50.0-2.5		50.0-2.5	
R/Rfree (%)	25.94/23.69	25.33/22.35	26.13/26.92	26.05/23.72	24.87/26.52	24.58/26.21

[§]: long exposure data was collected before short wavelength to avoid the radiation damage artifact.

*: Numbers in parentheses are statistics for the highest resolution shell.

Test of the MDS strategy



Table 1b Data collection and refinement statistics

	Crystal4		Crystal5		Crystal6	
Data collection						
X-ray source			SER-CAT 22-ID			
X-ray Optics			Monochromator			
Detector			Mar 225 CCD			
Wavelength (Å)			2.0			
Space group			I2 ₁ 3			
Cell dimensions:						
<i>a</i> = <i>b</i> = <i>c</i> (Å)		77.84		78.58		77.76
Exposure (seconds)	9.0	3.0	9.0	3.0	9.0	3.0
Scan range (°)	90.0	3 X 90.0	90.0	3 X 90.0	90.0	3 X 90.0
Resolution(Å)*	50.00-2.30 (2.38-2.30)	50.00-2.30 (2.38-2.30)	50.00-2.30 (2.38-2.30)	50.00-2.30 (2.38-2.30)	50.00-2.30 (2.38-2.30)	50.00-2.30 (2.38-2.30)
R _{svm} (%)	5.2(8.9)	6.5(12.1)	5.3(7.7)	5.8(10.0)	5.1(11.1)	6.7(17.6)
I/σI	62.3(45.3)	89.4(58.0)	69.8(55.1)	106.5(69.8)	58.2(37.5)	105.7(96.0)
Completeness (%)	99.24(99.14)	99.38(99.19)	99.14(99.05)	99.36(99.19)	99.46(99.30)	99.41(99.29)
Redundancy	10.3	30.8	10.2	30.3	10.3	30.4

Data collected by UGA team

Test of the MDS strategy



Table 2a Anomalous signal calculation

	Crystal 1		Crystal 2		Crystal 3	
Resolution (Å)	50.0-2.5		50.0-2.5		50.0-2.5	
RPH [§]	1.66	2.46	2.96	3.19	2.92	3.19
Map CC ^{&}	0.37	0.53	0.58	0.61	0.52	0.66
R cc [%]	1.43		1.05		1.27	

[§] RPH: Relative peak height is the ratio of average peak height of peaks 1, 2 and 3 divided by the average peak height of peaks 7, 8 and 9 in anomalous difference map calculated at 50.0 – 2.5Å resolution.

[&] Map CC: Map correlation coefficient between sulfur SAD phased map and model phased map at 50.0 – 2.5Å resolution.

[%] Rcc: Ratio of map CC between MDS data and the long exposed data of the same crystal.

Test of the MDS strategy



Table 2b Anomalous signal calculation

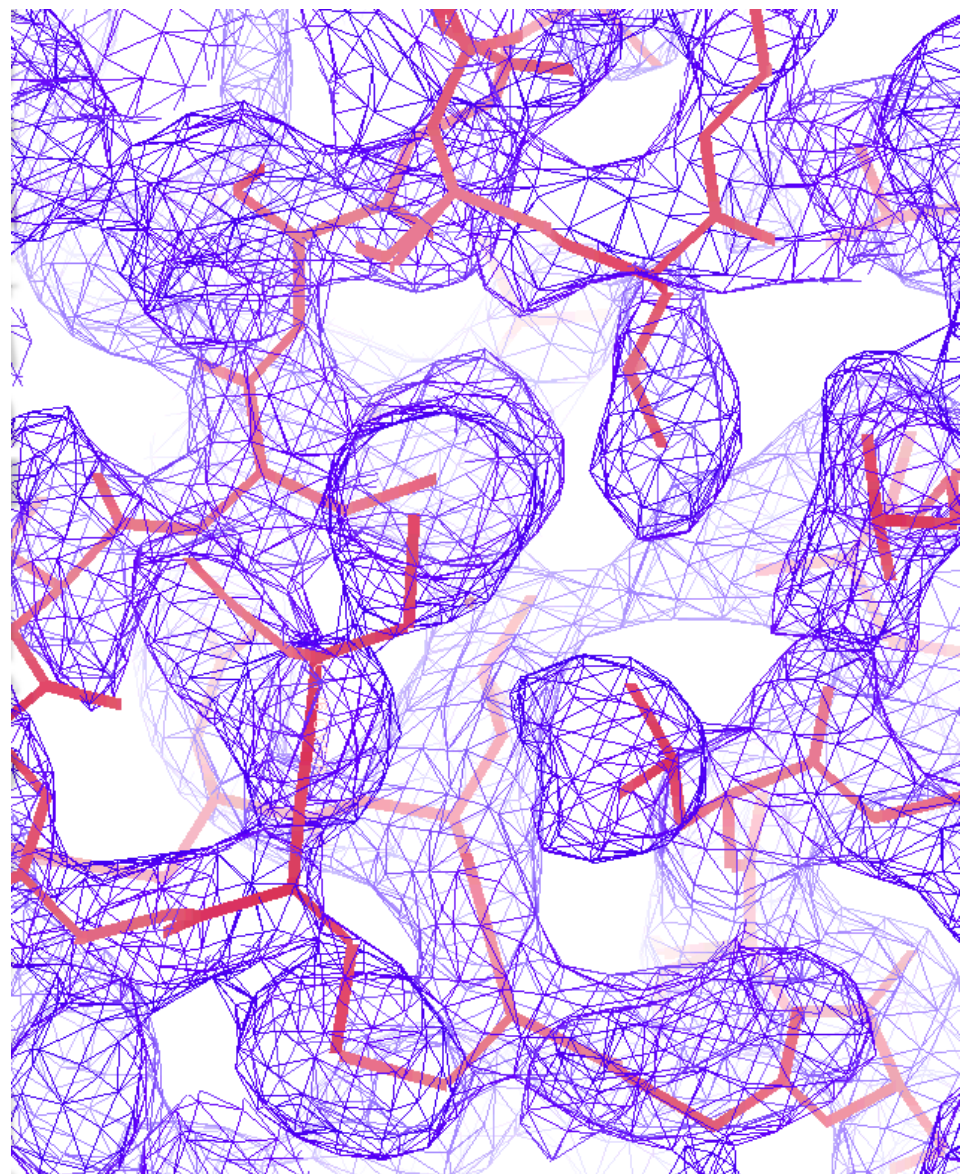
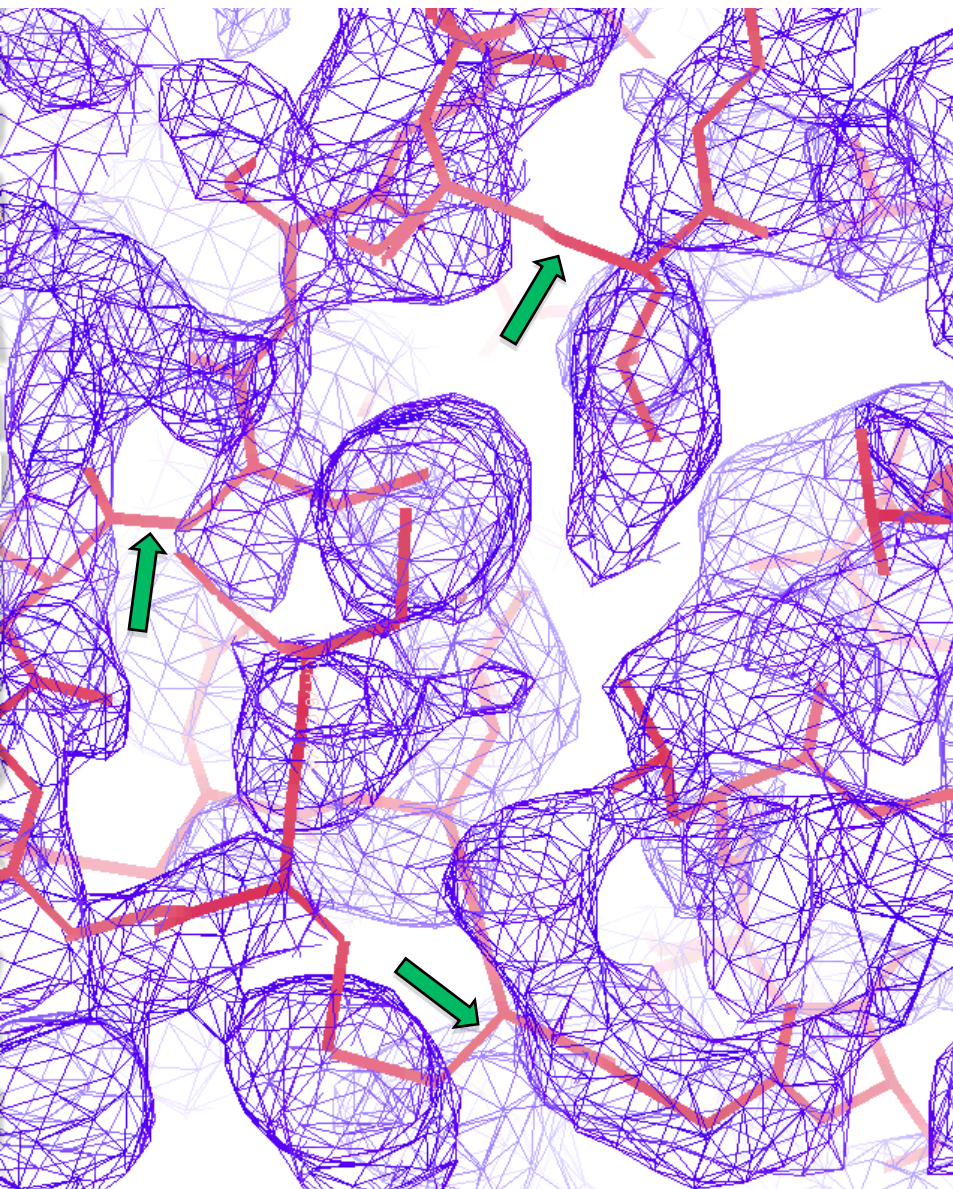
	Crystal 4		Crystal 5		Crystal 6	
Resolution (Å)	50.0-2.5		50.0-2.5		50.0-2.5	
RPH [§]	2.43	2.64	2.42	2.54	2.33	2.55
Map CC ^{&}	0.767	0.804	0.726	0.757	0.787	0.839
R cc [%]	1.05		1.05		1.27	

[§] RPH: Relative peak height is the ratio of average peak height of peaks 1, 2 and 3 divided by the average peak height of peaks 7, 8 and 9 in anomalous difference map calculated at 50.0 – 2.5Å resolution.

[&] Map CC: Map correlation coefficient between sulfur SAD phased map and model phased map at 50.0 – 2.5Å resolution.

[%] Rcc: Ratio of map CC between the MPC data and the long exposed data of the same crystal.

Test of the MDS strategy



“Parameter Space Screening” method based phasing pipeline:

Pipeline architecture

Dictionary driven Web-based user interface

- user authentication
- builds web page based on dictionary specifications
- upload require data
- dictionary provides for
 - On-line validation
 - easy pipeline updates and modifications

BioperlPipeline based workflow manager

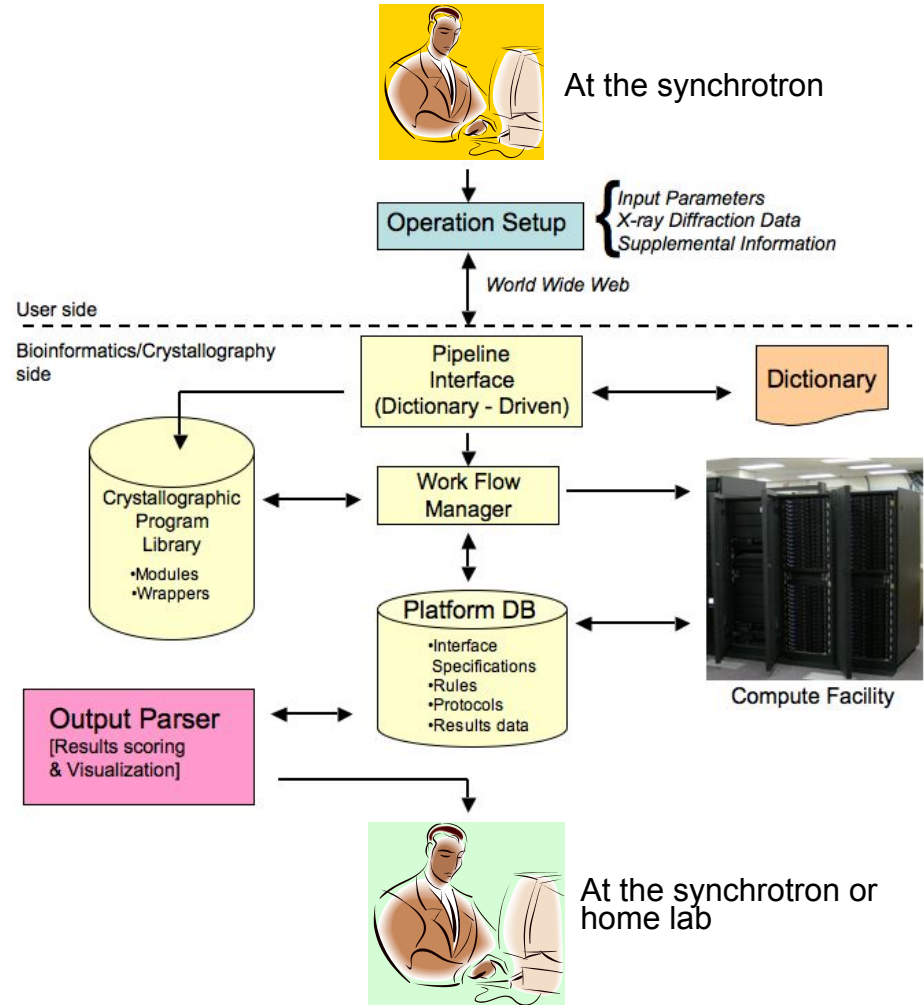
- manages interaction of the software components
- build the pipeline based on dictionary specifications
 - using a predefined configuration file
 - using input/out software wrappers based on Bioperl modules
- submit jobs to the cluster
- monitor jobs to ensure efficient use of computing resources

Set of analytical tools for harvesting results

- parse out key data
- format into web tables and exported to the user
 - promising solutions are highlighted in red
 - allows sorting of results by key data items

A relational database archive

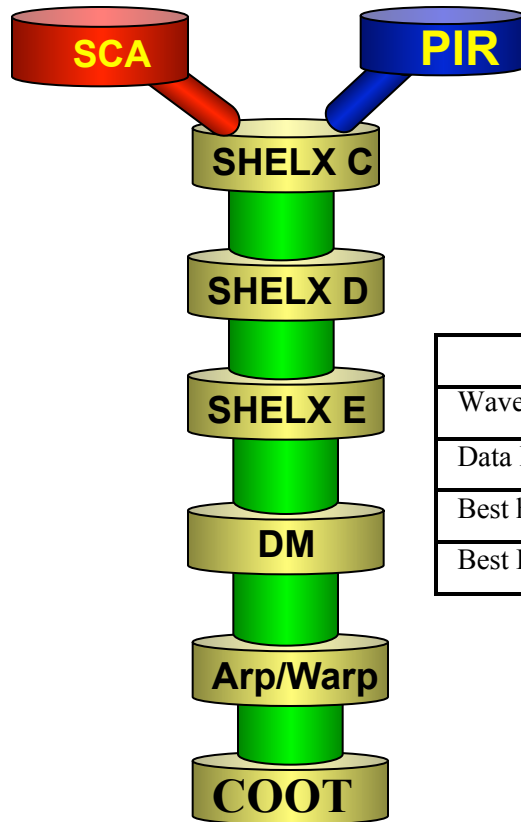
- the dictionary used
- the job process history
- all pipeline input/pipeline output files



A powerful tool for S-SAD phasing



Parameter Space Screening method based pipeline:



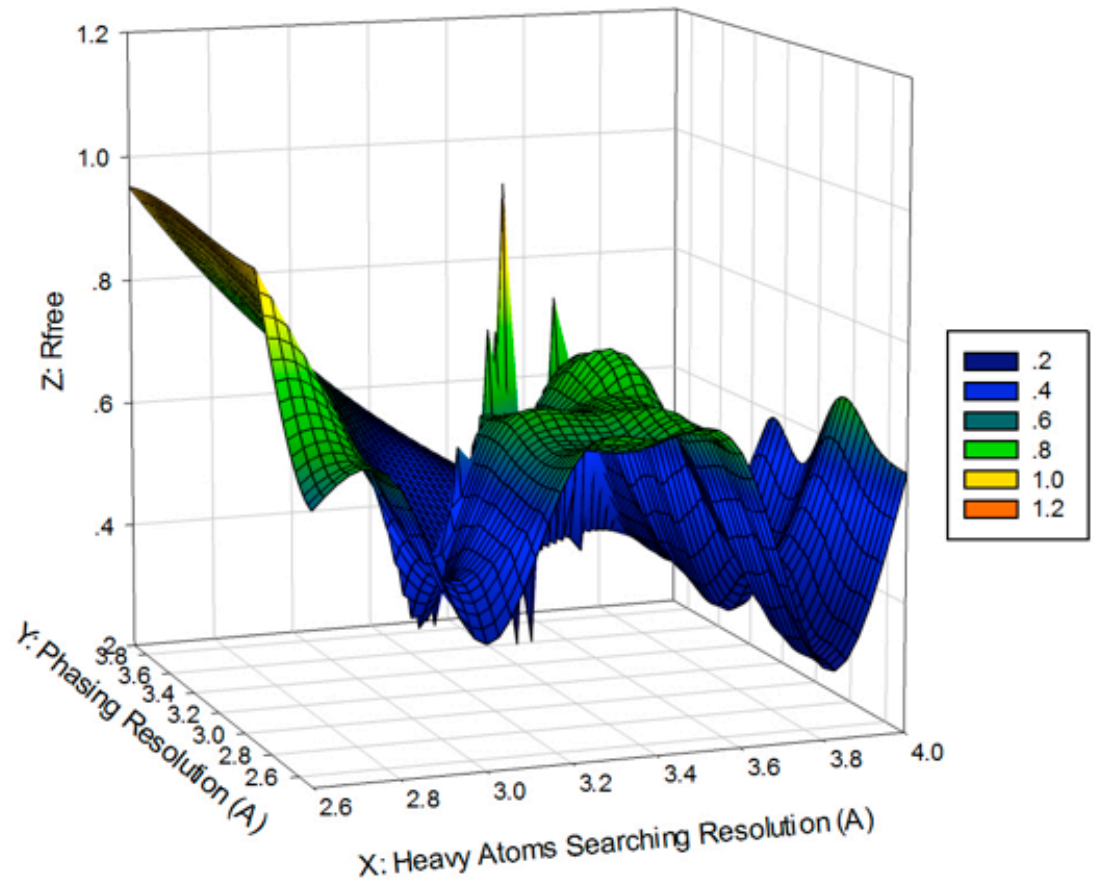
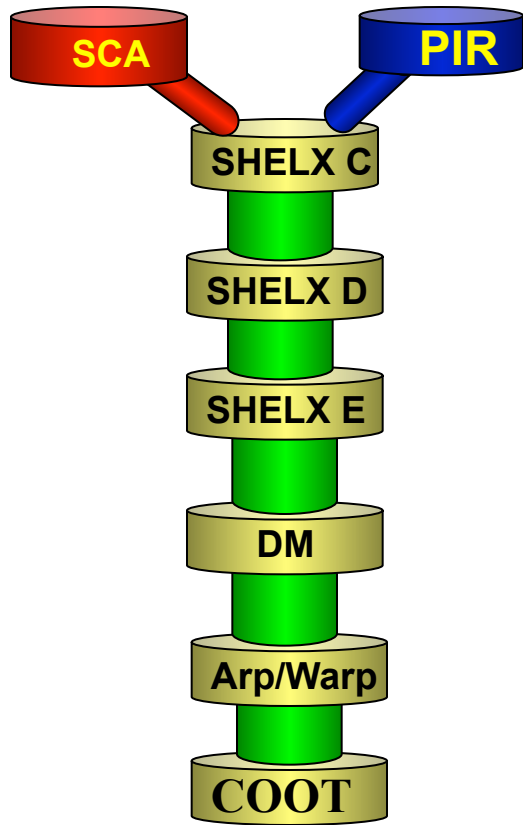
Phasing results from 2D screening of the pipeline

	A2	B1	C1	D1
Wavelength (Å)	2.00	2.00	2.70	2.70
Data Resolution (Å)	50.00-2.70	50.00-2.70	50.00-2.95	50.00-3.20
Best heavy searching resolution (Å)	4.00	3.80	3.60	N/A
Best Phasing Resolution (Å)	3.50	2.70	3.30	N/A

A powerful tool for S-SAD phasing



Parameter Space Screening method based pipeline:



Parameter space screening results for crystal B

Summary



1. Data quality varies drastically with data collection system. It is necessary to check the readiness before data collection.
2. Zn-free insulin crystals are excellent probes for monitoring the facility's status.
3. $R_{\text{ano}}/R_{\text{pim}}$ ratio, $d''/\text{sig}(d'')$ and CC (All/Weak) values are reliable indicators for measuring data with good anomalous signals, while $d''/\text{sig}(d'')$ is more sensitive for grading a wider range of anomalous data qualities in this specific case.
4. MDS strategy is an effective way for improvement of data quality.
5. "Parameter Space Screening" pipeline show potential in S-SAD phasing.



Acknowledgments



Heng Ru

Bi-Cheng Wang

OuYang SongYing

John Rose

Dong Wu

Lirong Chen

Wei Ding

Hua Zhang

Tian Hua

Soichi Wakatsuki

Naohiro Matsugaki

John Chrzas

Albert Fu

Other IBP, PF and SER-CAT staff

NIGMS, UGA, GRA, IBP, CAS, MOST, NSFC