4th Winter School on Soft X-rays in Macromolecular Crystallography

A General Introduction to the Use of Softer X-rays in MX

S-SAD and X-SAD Approaches for Phase Determination

Bi-Cheng Wang

Biochemistry and Molecular Biology and SER-CAT
University of Georgia

ESRF, Grenoble, France
February 6, 2012

Introduction

The use of S-SAD for phase determination started more than 30 years ago.

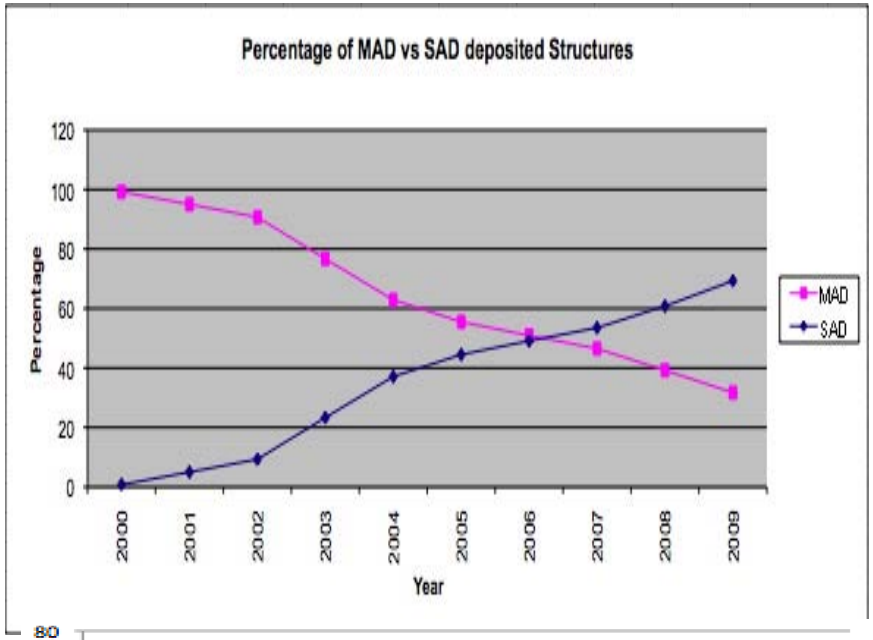
In the early 1980s, one important consideration of using Single-wavelength Anomalous Scattering data (called SAS at the time) was on how to resolve phase ambiguity in SAD data.

The structure determination of Crambin (4.7 kDa, 6 sulfurs) in 1981 (Hendrickson and Teeter, *Nature*, 289, 366, 1981) from the S-SAD data inspired great interest in SAD phasing at the time from many crystallographers, including our research group.

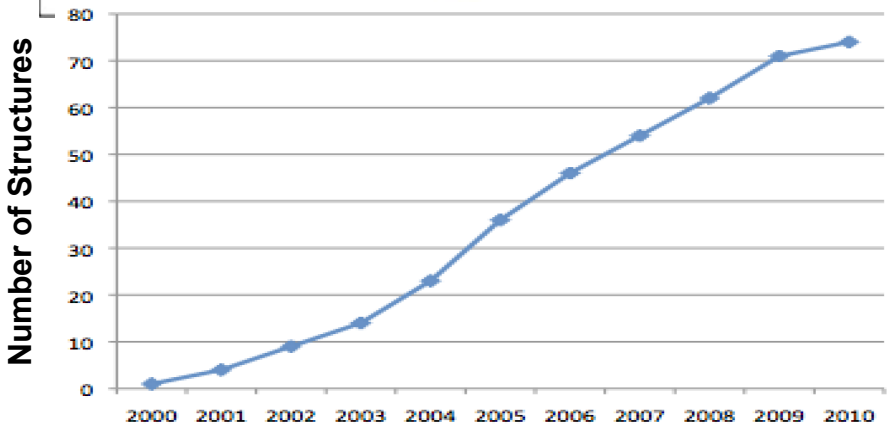
However, the general interest in using any SAD data for structure determination quickly diminished and replaced by MAD approach in the late-80s.

It would be interesting to take a quick look on the growth of *de novo* PDB structures based on the phasing methods used and focus our attention on some key aspects of the evolution of phasing methods from MAD to SAD in this lecture.

De Novo PDB Structures



The Use of **MAD Data** and **SAD (SAS) Data** in the 10 year period from 2000 to 2009



The Use of **S-SAD (S-SAS) Data** from 2000 to 2010

There are theoretical differences between the MAD and SAD methods. There are also theoretical differences between the first S-SAD structure (crambin, 1981) and those determined 19 years later (starting with obelin, Liu, et al., 2000). Let's first focus our attention to the differences in the basic theory behind the SAD phasing as an introduction to **Any-SAD** phasing.

Outline

- *The SAD Methods - 1940s - 1980s*
  - *Resolving the phase ambiguity in SAD data*
  - *Determining the handedness*
- *The Use of S-SAD and X-SAD Methods - 1980s - 2012*
- *Continued Efforts for S-SAD and X-SAD Phasing - 2000 -*
  - *Various approaches for improving I/σ (I)*
  - *The use of softer (extended wavelength) X-rays*
- *A Shared "Resource" at the APS to Advance Extended Wavelength MX - 2010 -*
- *Database SSAD_DB at UGA - 2011 -*

Selected Slides Presented at Workshops/Meetings, including:

IUCr2011, Madrid, Spain, 2011
APS Extended Wavelength MX Workshop, 2011
ACA One-Day S-SAD Workshop, Chicago, 2010
Pittsburgh Diffraction Conference One-Day S-SAD Workshop, 2009
EMBO World Lecture on MX, Pune, India, 2008
Winter School on Softer X-rays in MX, Seefeld, 2006; Berlin 2009
Cold Spring Harbor Laboratory X-ray Course at IBP, Beijing 4/29-5/15, 2008
MSC SAD Phasing Workshop Houston, April 3-4, 2006
ISAS Workshop, Tsinghua University, Beijing, China, June 6, 2002
ACA Summer School in Crystallography, Pitt & UGA, 1993-2002
ACA ISAS Workshop, Los Angeles, July 21, 2001
ISAS Workshop, University of Georgia, Athens, GA, April 6-7, 2000

UGA, Athens, Georgia, USA April 6, 2000

Basic Theory Behind the SAD Phasing

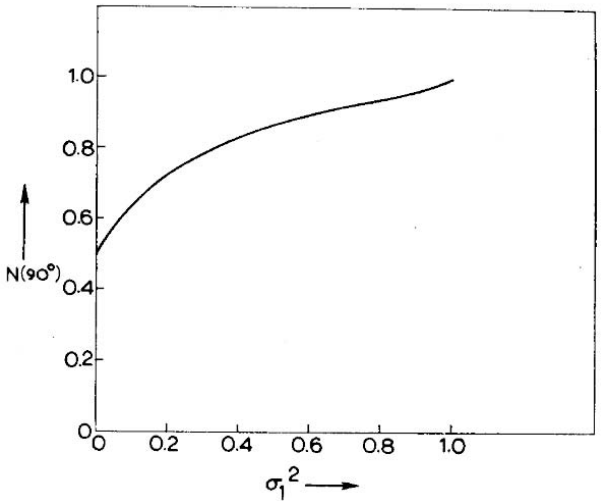
Two different S-SAD approaches proposed in early 1980s

The First Approach

1981: **Crambin (4.7 kDa, 6 sulfurs)** (Hendrickson and Teeter, *Nature*, 289, 366, 1981) proposed **A single-step process** called **R**esolved **A**nomalous (**RA**) phasing.

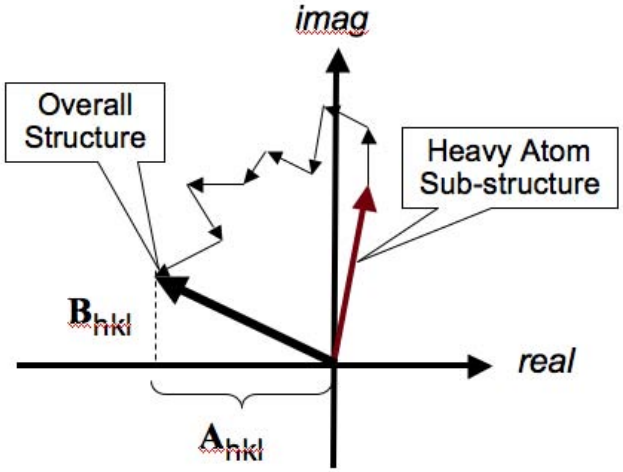
This approach is very similar to the **quasi-anomalous phasing** * method of Ramachandran and Srinivasan (1970).

* Pages 174-176, G.N. Ramachandran and R. Srinivasan, "Fourier Methods in Crystallography." Wiley (Interscience), New York, 1970



S. Parthasarathy, Acta Cryst. 18, 1028 (1968).

These approaches used the phases of the sub-structure of the anomalous scatters to resolve the phase ambiguity. They would work for small molecules, but not for the average macromolecules.



The Second Approach

1982: **Rhe (12.5 kDa, 2 sulfurs)** (Wang, *Method Enzymol*. 115, 90-122, 1985)
A multi-step process using "filters", Fourier transform and iteration, called **I**terative **S**ingle-wavelength **A**nomalous **S**cattering (**ISAS**) phasing.

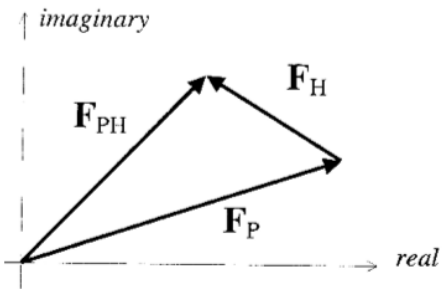
Simulation results showed that **each sulfur atom** can phase at least **57** amino acid residues.

This approach uses protein image itself, not sulfur atom sub-structure, to "resolve" protein phase ambiguity

At the time our group was developing a general method for resolving the phase ambiguity in single isomorphous replacement (SIR) data and realized that a similar approach could be applied to SAS data as well with great effectiveness.

The Phase Ambiguity Problem and how it was treated from the 1940s to 1970s

P. 153, Protein Crystallography, Blundell and Johnson, 1976



"…Let us assume that we have measured F_{PH} and F_P and that we know the arrangement of heavy atoms in the crystal unit cell, i.e. we can calculate the vector F_H. What can we then derive about the phase? …using the cosine law:

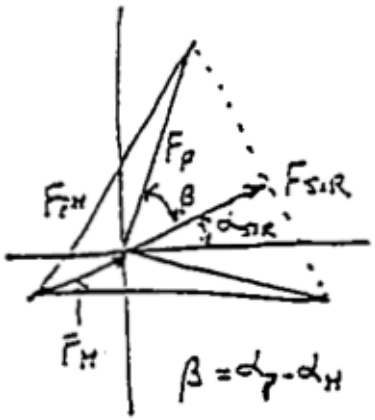
$$\alpha_P = \alpha_H + \cos^{-1}\left(\frac{F_{PH}^2 - F_P^2 - F_H^2}{2F_PF_H}\right) = \alpha_H \pm \alpha'$$

The equation shows that there are two possible values for α_P which cannot be distinguished with one isomorphous derivative…"

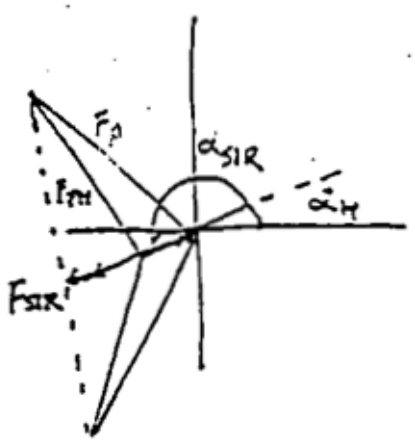
P. 157-178, Protein Crystallography, Blundell and Johnson, 1976

"… there are two closely similar ways of including the phase information from single isomorphous replacement in a Fourier synthesis. The mean of the two phases can be used with suitable weighting (Blow and Rossmann, 1961)… Alternatively, both phases can be used in the Fourier calculation. This is known as a "double-phased synthesis" (Bijvoet, 1949; Bokhoven *et al*., 1951). It can be shown that syntheses using single isomorphous replacement are actually equivalent to the protein structure plus the inverse structure convolved with the phase-squared structure of the H atoms (Ramachandran and Srinivasen, 1970) (if the heavy atoms also have a noncentrosymmetric array)…The method can then give a protein electron density map as was shown by Blow and Rossmann (1961) but the high level of background has meant that it has been rarely used by protein crystallographers…"

"…the protein structure plus the inverse structure convolved with the phase-squared structure of the H atoms…(Ramachandran and Srinivasen, 1970)… "



$\alpha_{SIR} = \alpha_H$



$\alpha_{SIR} = \alpha_H + \pi$

The SIR vector is defined as the projection of the F_P vector onto the F_H vector. It follows that

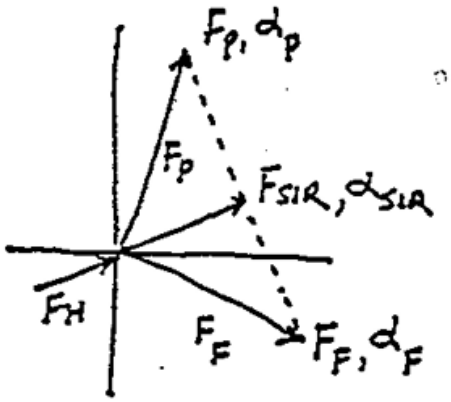
$$|F_{SIR}| = |F_P| \cos\beta = m|F_P|$$

where m is the figure-of-merit,

and $\alpha_{SIR} = \alpha_H$ when $|F_{PH}| - |F_P| > 0$

and $\alpha_{SIR} = \alpha_H + \pi$ when $|F_{PH}| - |F_P| < 0$.

"…the protein structure plus the inverse structure convolved with the phase-squared structure of the H atoms …(Ramachandran and Srinivasen, 1970)…"



$$F_{SIR} = (1/2)(F_P + F_F)$$

$$\alpha_{SIR} = (1/2)(\alpha_P + \alpha_F)$$

$$\alpha_F = 2\,\alpha_{SIR} - \alpha_P$$

$$= 2\,\alpha_H - \alpha_P$$

$$F_{SIR} = |F_{SIR}|e^{i\alpha_{SIR}} = m|F_P|e^{i\alpha_{SIR}}$$

but $$F_{SIR} = (1/2)[|F_P|e^{i\alpha_P} + |F_F|e^{i\alpha_F}]$$

$$= (1/2)[|F_P|e^{i\alpha_P} + |F_P|e^{i(2\alpha_H - \alpha_P)}]$$

Thus, $$m|F_P|e^{i\alpha_{SIR}} = (1/2)[|F_P|e^{i\alpha_P} + |F_P|e^{i(2\alpha_H - \alpha_P)}]$$

$$\rho_{SIR} = 0.5\,[\rho_{Protein} + \rho_{Noise}]$$

or

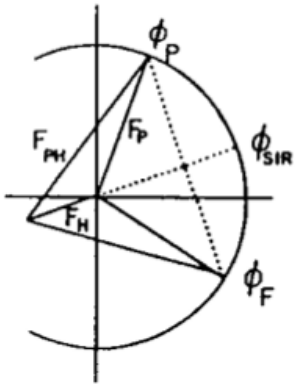
$$\rho_{Protein} = 2\rho_{SIR} - \rho_{Noise}$$

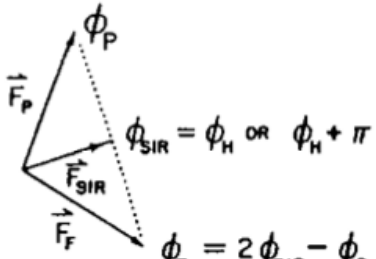
"…it has been rarely used by protein crystallographers…"

Our attention then shifted: ***How about building a <u>filter</u> to remove ρ_{Noise}?***

How can we treat SAD in a similar way as we treat SIR?

The Nature of SIR and SAS (SAD) Maps



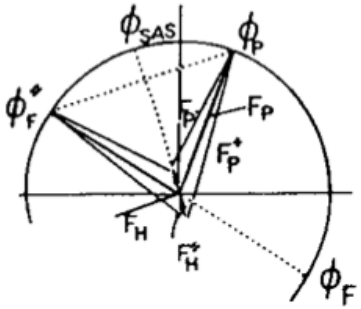


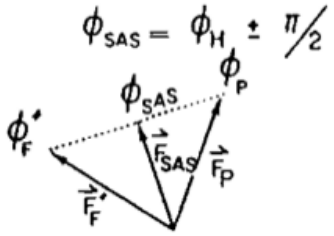
$$\phi_{SIR} = \frac{1}{2}(\phi_P + \phi_F)$$

$$2 \times \vec{F}_{SIR} = \vec{F}_P + \vec{F}_F$$

$$2 \times m F_P e^{i\phi_{SIR}} = F_P e^{i\phi_P} + F_P e^{i2\phi_H} e^{-i\phi_P}$$

$$2 \times \int_{SIR} = \int_{protein} + (Noise)_{SIR}$$





$$\phi_{SAS} = \frac{1}{2}(\phi_P + \phi_F')$$

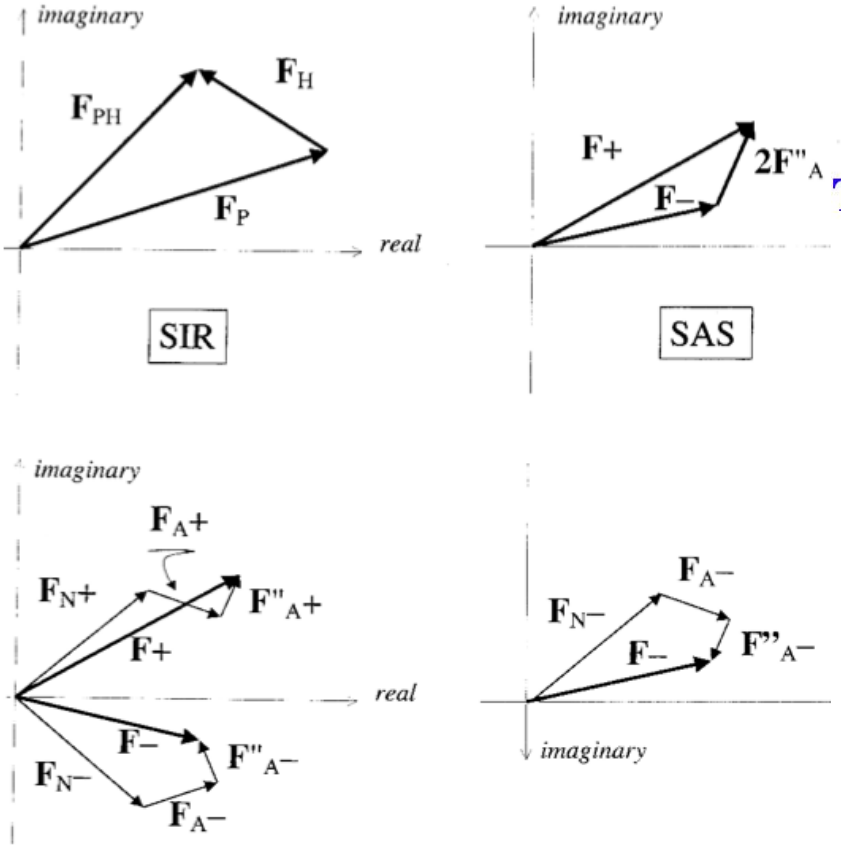
$$2 \times \vec{F}_{SAS} = \vec{F}_P + \vec{F}_F'$$

$$2 \times m F_P e^{i\phi_{SAS}} = F_P e^{i\phi_P} - F_P e^{i2\phi_H} e^{-i\phi_P}$$

$$2 \times \int_{SAS} = \int_{protein} + (Noise)_{SAS}$$

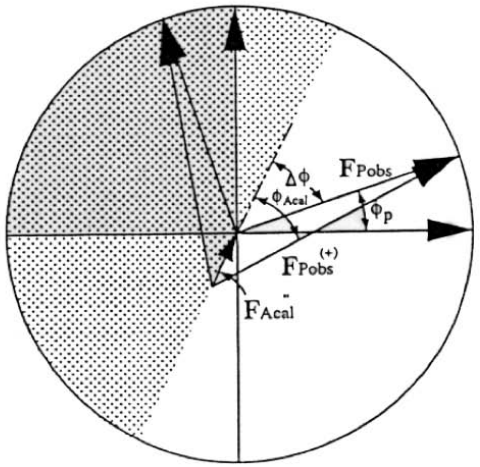
Wang (1985), Methods Enzym, 115, 90-112

Phase Triangle Relationships in SIR and SAS



Basic Principle of SIR and SAS Methods

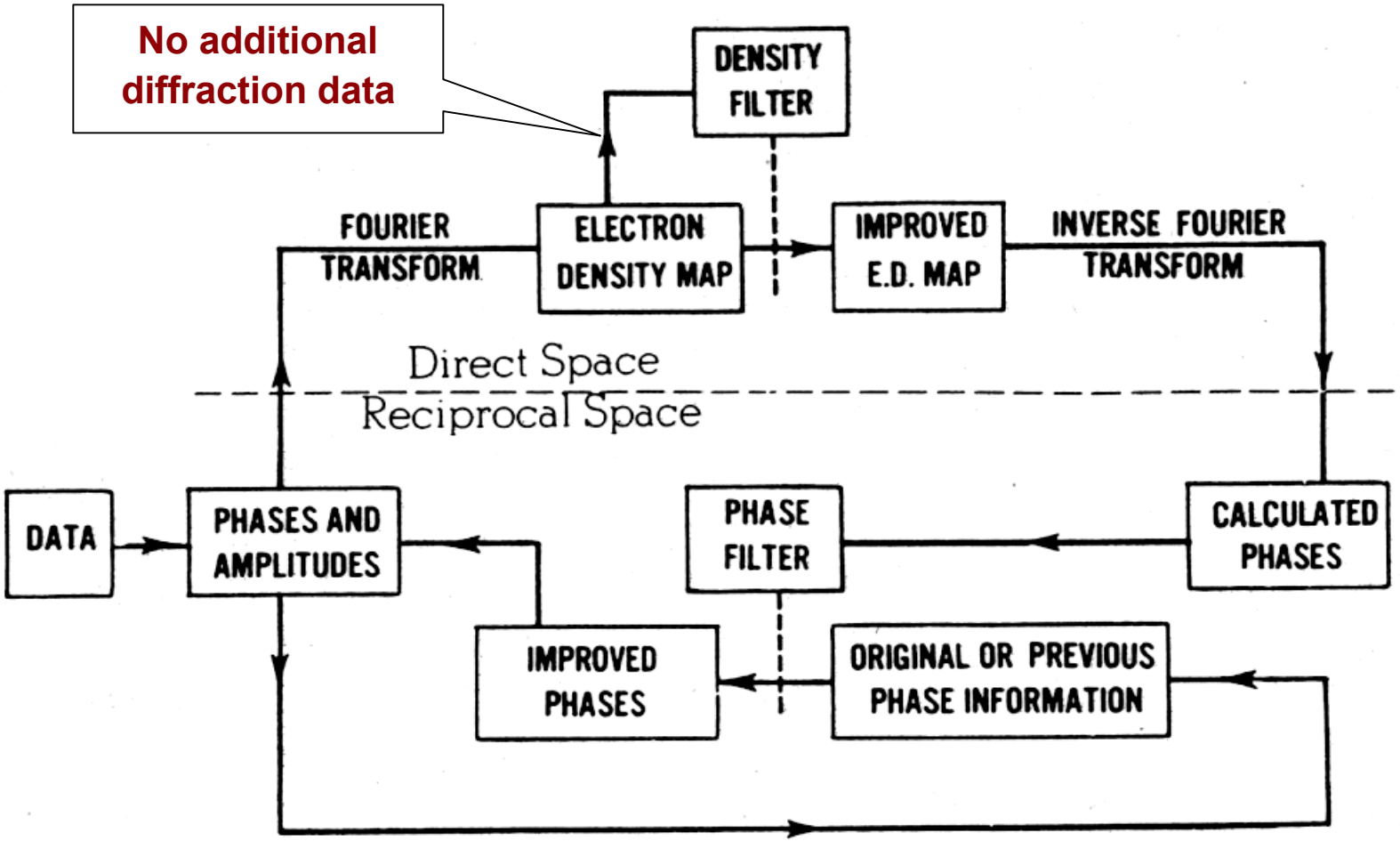
To find the orientation of a phase triangle from the orientation of one of its sides – two possibilities



Wang's lecture notes, ACA Summer School in Crystallography (1993–2002)
UGA ISAS Workshop (2000), ACA ISAS Workshop (2001), and others

So, we may treat the SAS data in a manner similar to how we treated the SIR data

The Proposed "Noise Filtering" Process for "Resolving" Phase Ambiguity



Wang (1985), Methods Enzym, 115, 90-112

The Phase Ambiguity Problem

Phase ambiguity is **not an inherent property** of single isomorphous replacement (SIR) or single anomalous scattering (SAD) data.

This view is analogous to that of "**Phase Problem**" in crystallography.

For example, phase-loss occurs for an **individual reflection** when only intensity is recorded. Phase-loss is not an inherent property when the **intensities of a complete assembly (complete data set) of reflections** are known.

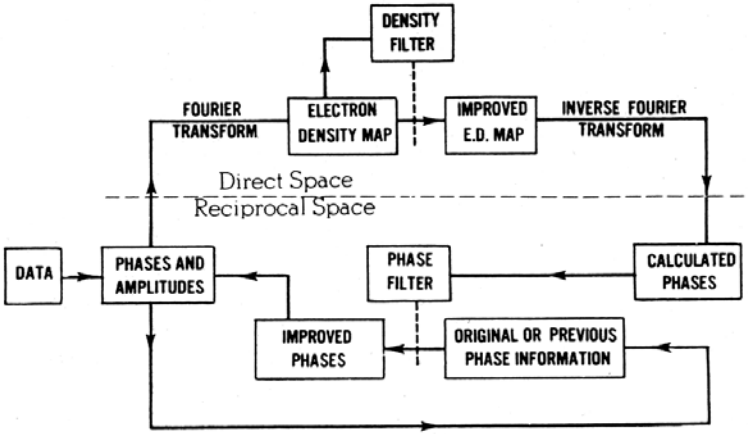
Thus, phase-ambiguity occurs for an **individual reflection pair** when only intensities are measured. Phase-ambiguity is not an inherent property when the **intensities of a complete assembly (complete SAD or SIR data set) of reflection pairs** are known.

If we measure a set of SIR or SAD data accurately enough, the phase ambiguity problem can be solved without "additional" data.

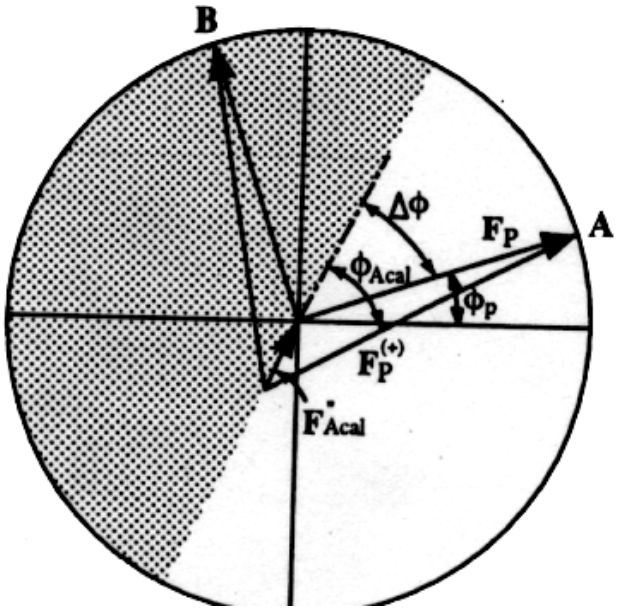
How Can Noise Filtering (Solvent Flattening) Remove the Phase Ambiguity?

1. Conceptual Answer

- **Solvent flattening can locate and enhance the protein image – e.g. whatever is not solvent must be protein!**
- **From the protein image, the phases of the structure factors of the protein can be calculated**



- These calculated phases are then used to **select** the true phases from sets of true and false phases
- Thus, in essence, the phase ambiguity is **resolved by the protein image itself!**



ISAS map from experimental Au-SAS data of RHE Wang, 1985

How Can Noise Filtering Remove the Phase Ambiguity?

2. Numerical Solution by **Iteration**

Start with a SAS map: $\rho_{SAD} = 0.5\,\rho_{Protein} + N_{SAD}$

If molecular boundary is known and N_{SAS} can be separated into four parts and

First SAS map: $\rho_{SAD} = 0.5\,\rho_{Protein} + N_P(+) + N_P(-) + N_S(+) + N_S(-)$

$N_P(-) + N_S(+) + N_S(-)$ **are filterable**

First filtered map: $\rho_{SAD} = 0.5\,\rho_{Protein} + N_P(+)$

After Filter 1 and Cycle 1 (1 direct and 1 reciprocal space filterings)

$$\rho_{SAD} = M\,\rho_{Protein} + N_P(+) + N_P(-) + N_S+) + N_S(-) \text{ where } 0.5 < M < 1.0$$

The filterable terms, $N_P(-) + N_S(+) + N_S(-)$, are getting **smaller and smaller!**

After n cycles of filtering $\rho_{SAD} = \rho_{Protein} + N_P(+) + N_P(-) + N_S(+) + N_S(-)$

We originally named this process Iterative Single-wavelength Anomalous Scattering (ISAS) method (Wang 1985)

Basis for most density modification algorithms in using X-SAD data

Locating the Molecular Boundary for Making the Density Filter

Electron density map, ρ_i, represents a **population of electrons**. It is useful for locating atoms but not for locating molecules.

A suitable algorithm is needed to represent the **population of atoms** (atomic density?) to define the molecule in the same way that population of houses defines a city.

A new map, ρ_j , is constructed such that the density at each grid point is the sum of the positive density with a sphere of radius R from that grid point in the original map

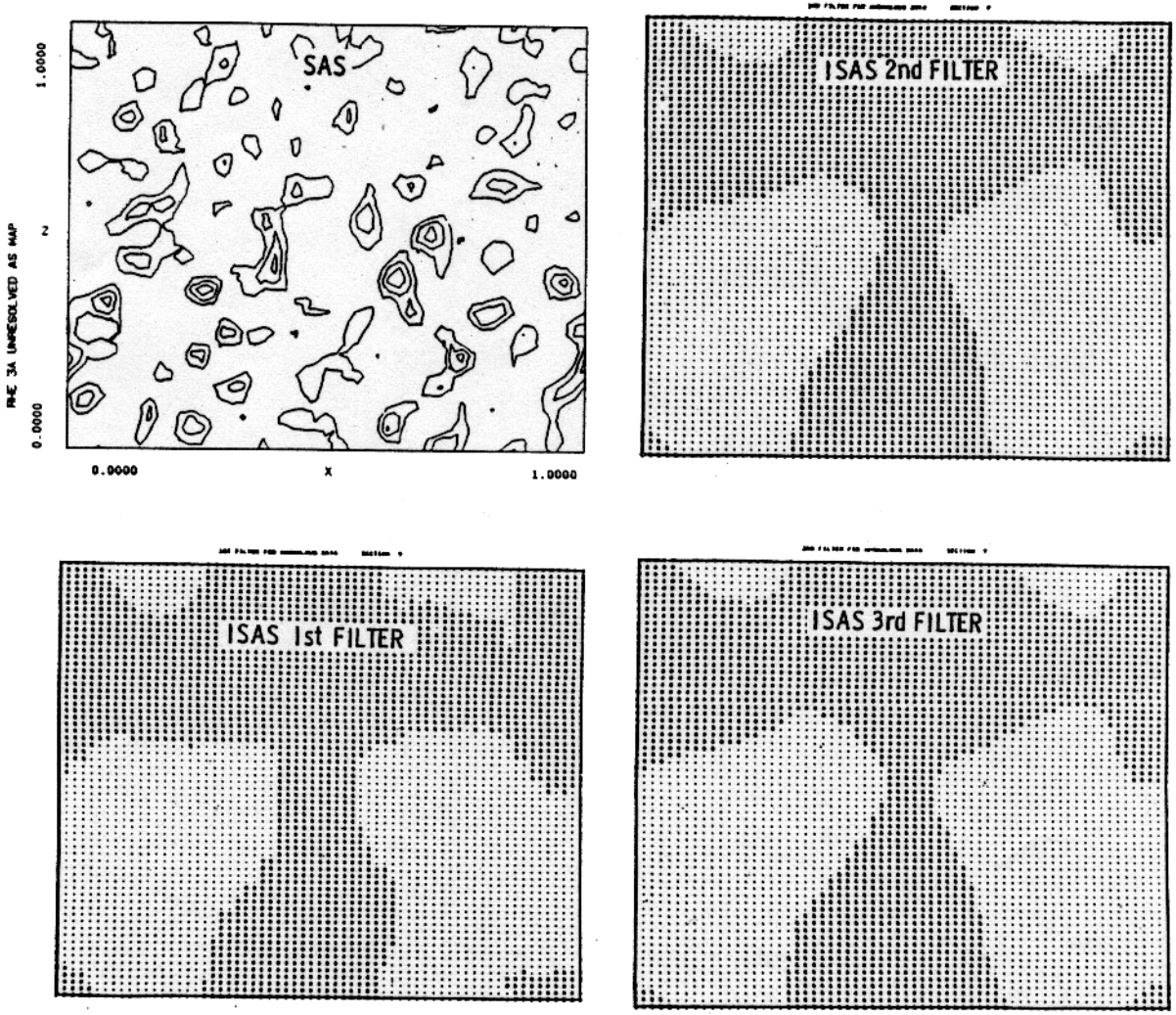
$$\rho_j^B = \sum_i^R w_i \rho_i$$

$$w_i = 1 - r_{ij}/R \quad \text{for } r_{ij} < R \text{ and } \rho_i > 0$$

$$w_i = 0 \quad \text{for } r_{ij} > R \text{ or } \rho_i < 0$$

This map tells us the probability of finding the boundary between the molecule (city) and solvent (outside the city).

An Example of Density Filters



Computer Simulation on Rhe (113 residues, 2 sulfurs) by Sulfur-ISAS Method in 1982

*(Maps Produced **With** and **Without** Iteration)*

Before Iteration

After Filter 1 cycle 1

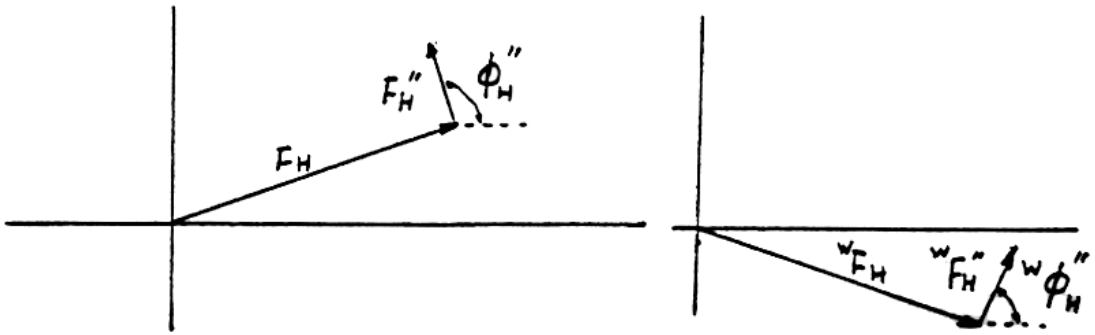
After Filter 3 cycle 8

Map by refined phases

Wang (1985), Methods Enzym, **115**, 90-112

The Handedness of SAD Data and Structure

What Happens if the Handedness of Anomalous Scatterer is Wrong?



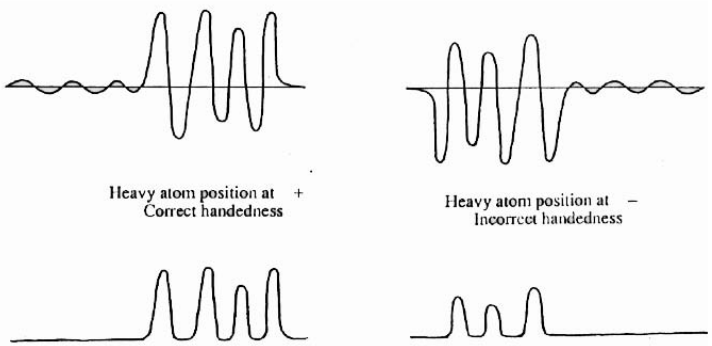
After Fourier transformation, we will have

$$ {}^w\rho_{SAS} = -{}^m\rho_{SAS} $$

where ${}^m\rho_{SAS}$ is the mirror image of ρ_{SAS} , the correct SAS density.

Note the negative sign in front of ${}^m\rho_{SAS}$.

ρ_{SAS} and $-{}^m\rho_{SAS}$ Maps



Anomalous Scattering Workshop, Birmingham, AL (1989)
Montreal ACA Annual Meeting (1995),
UGA ISAS Workshop (2000), ACA ISAS Workshop (2001), and others

Handedness Can be Determined by Solvent Flattening*

The ISAS process is carried twice, once with heavy atom site(s) at refined locations (+++), and once in their inverted locations (---).

- Heavy Atom Handedness and Protein Structure Determination using Single-wavelength Anomalous Scattering Data. Wang, ACA Annual Meeting, Montreal, July 25, 1995.

Handedness and space group enantiomorph

Examples

Data	FOM[1]	Handedness	FOM[2]	R-Factor	Corr. Coef
RHE	**0.54**	**Correct**	**0.82**	**0.26**	**0.958**
	0.54	Incorrect	0.80	0.30	0.940
NP With I[3]	**0.54**	**Correct**	**0.80**	**0.27**	**0.955**
	0.54	Incorrect	0.76	0.36	0.919
NP With I & S[4]	**0.56**	**Correct**	**0.82**	**0.24**	**0.964**
	0.56	Incorrect	0.78	0.35	0.926

[1]: Figure of merit before solvent flattening
[2]: Figure of merit after one filter and four cycles of solvent flattening
[3]: Four Iodine were used for phasing (L. Chen, et el, *PNAS*, 88, 4240-4244 (1991))
[4]: Four Iodine and 56 Sulfur atoms were used for phasing

Heavy Atom Handedness and Protein Structure Determination using Single-Wavelength Anomalous Scattering Data, ACA Annual Meeting, Montreal, July 25, 1995.

Are those small statistical differences significant? Yes!

Handedness and space group enantimorph

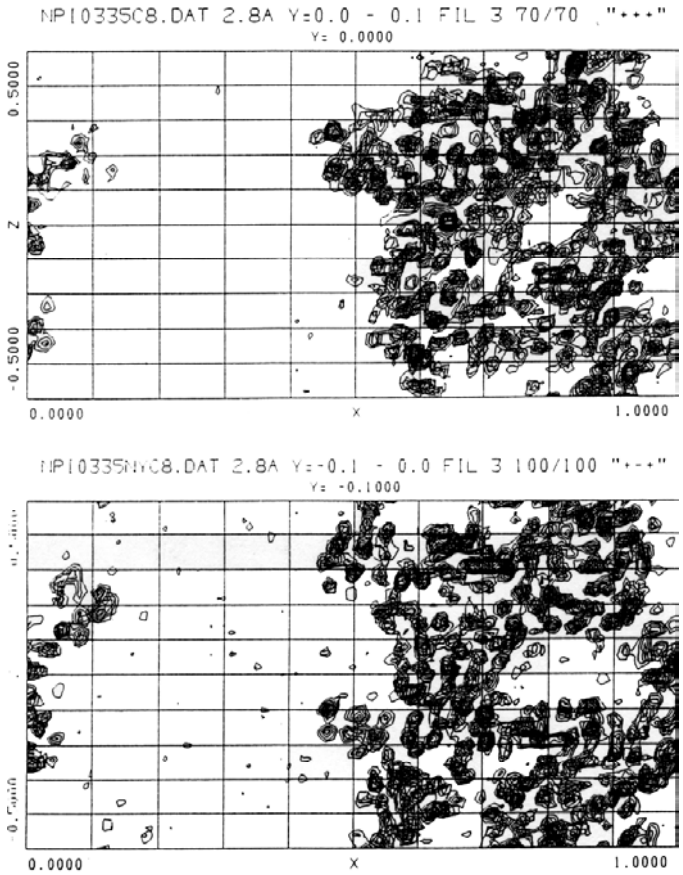


Figure 1.6 The ISAS Electron Density Maps of the "+++" and "+-+" Enantiomers Projected down the Y Axis.
Top: Iodines at (+++), Correct Handedness.
Bottom: Iodines at (+-+), Incorrect Handedness.

Wang, Lecture note, ISAS Workshop, University of Georgia (2000)

Outline

- *The SAD Methods - 1940s - 1980s*
  - *Resolving the phase ambiguity in SAD data*
  - *Determining the handedness*
- ***The Use of S-SAD and X-SAD Methods - 1980 - 2012***
- *Continued Efforts for S-SAD and X-SAD Phasing - 2000 -*
  - *Various approaches for improving I/σ (I)*
  - *The use of softer (extended wavelength) X-rays*
- *A Shared "Resource" at the APS to Advance Extended Wavelength MX - 2010 -*
- *Database SSAD_DB at UGA - 2011 -*

Our Work on the X-SAD Phasing Approach

1989: I-SAS λ = 1.54 Å on a home X-ray source

Crystal structure of a bovine neurophysin II dipeptide complex at 2.8 A determined from the single-wavelength anomalous scattering signal of an incorporated iodine atom. *Proc Natl Acad Sci U S A,* 88, 4240-4244 (1991).
L.Q. Chen, J.P. Rose, E. Breslow, D. Yang, W.R. Chang, W.F. Furey, M. Sax, B.C. Wang

1998: Fe-SAS λ = 1.54 Å on a home X-ray source

The 2.0 Angstrom Structure of Human Ferrochelatase, The Terminal Enzyme of Heme Biosysthesis. *Nature Structural Biology,* 8, 156-160 (2001).
C.K. Wu, H.A. Dailey, J.P. Rose, A. Burden, M. Sellers, B.C. Wang.

1999: S-SAS λ = 1.75 Å at a synchrotron

Structure of the Ca^{2+}-Regulated Photoprotein Obelin at 1.7 Å Resolution Determined Directly from its Sulfur Substructure.
Z.-J. Liu, E. S. Vysotski, C.-J. Chen, J. P. Rose, J. Lee, and B. C. Wang, *Protein Science*, 9, 2085-2093 (2000).

Based on these results a proposal was submitted to NIH in 2000 for the development of Direct Crystallography

Direct Crystallography

(Direct Approach to Protein Crystallography)

(Proposed to NIH in 2000)

Use Unlabeled Native Crystals and Single-wavelength X-rays

Metal atoms: **Fe, Co, Zn, Mn, Ca…., naturally present in metalloproteins, ~ 30% of all proteins contain metals**

Sulfur atoms: **Nearly all proteins have sulfur**

If sulfur phasing is successful then virtually any other anomalous scatterers become available for phasing!

- **Can we use single wavelength X-rays?**
- **Can we use sulfur atoms routinely as phasing probe?**

B.C. Wang, GRC 2004

X-SAD

Native-SAD and Derivative-SAD

Use Unlabeled Native Crystals and Single-wavelength X-rays

Metal atoms: **Fe, Co, Zn, Mn, Ca…., naturally present in metalloproteins, ~ 30% of all proteins contain metals**

Sulfur atoms: **Nearly all proteins have sulfur**

If sulfur phasing is successful then virtually any other anomalous scatterers, X, become available for phasing!

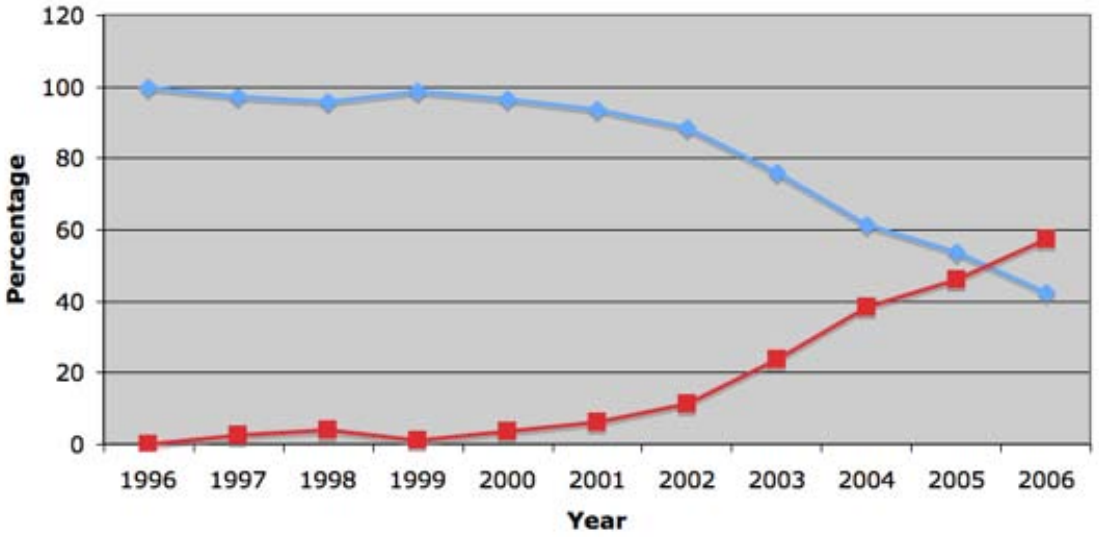
- **Can we use single wavelength X-rays?**
- **Can we use sulfur atoms routinely as phasing probe?**

About the use of single wavelength?

The answer now is "Yes".

The Use of MAD Data and SAD (SAS) Data

PDB Deposits



Our Results
131 structures from 2001-2007

SAD	73	12	by S-SAS
		38	by Se-SAS
		23	by M-SAS
MAD	7		
MR	51		

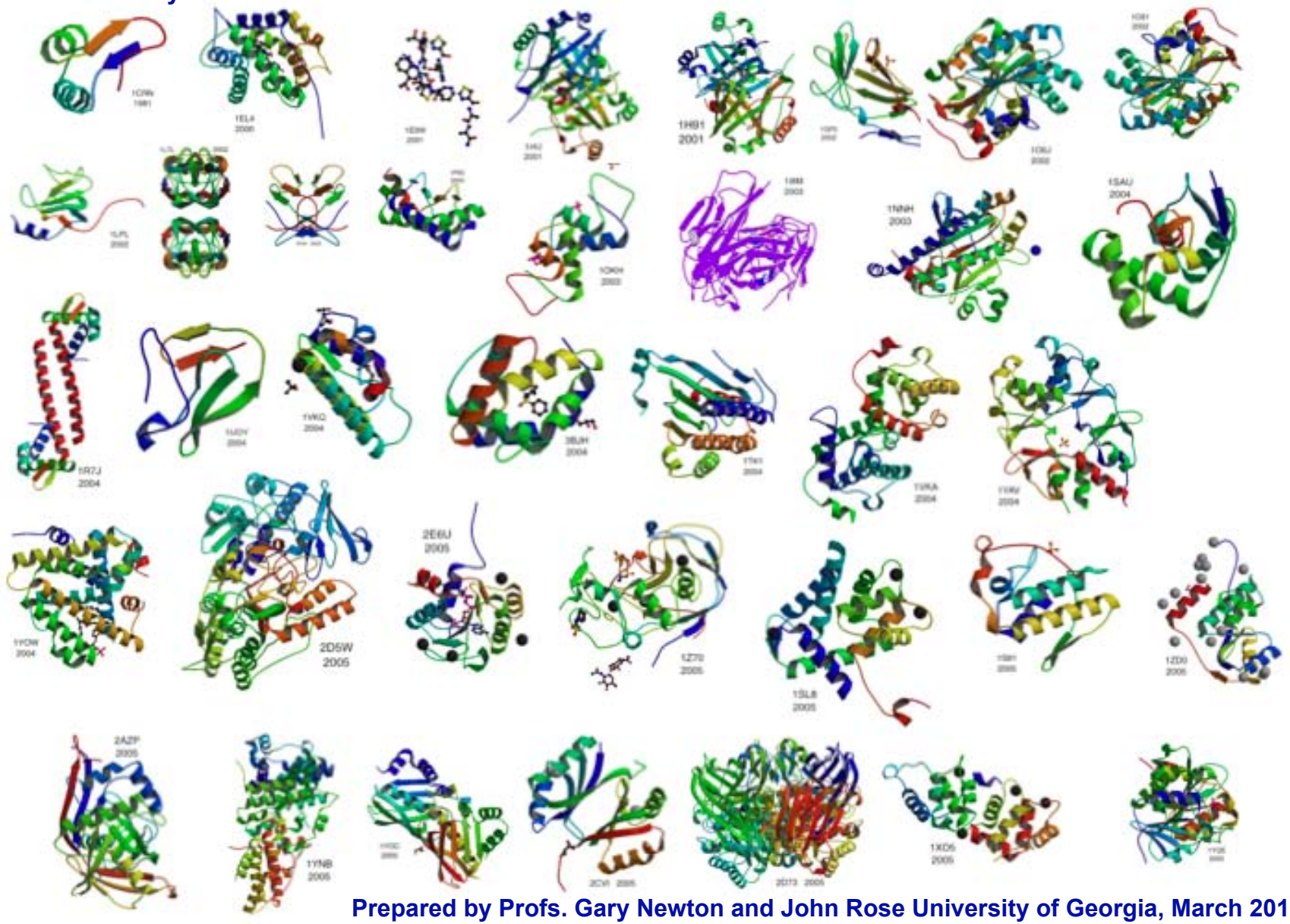
About the use of longer wavelength?

With the increased interest in single wavelength anomalous scattering phasing, data collection using ~1Å X-rays (Se absorption edge) has become an option, instead of a requirement.

37 S-SAD Structures Deposited in PDB from 1981 to 2005

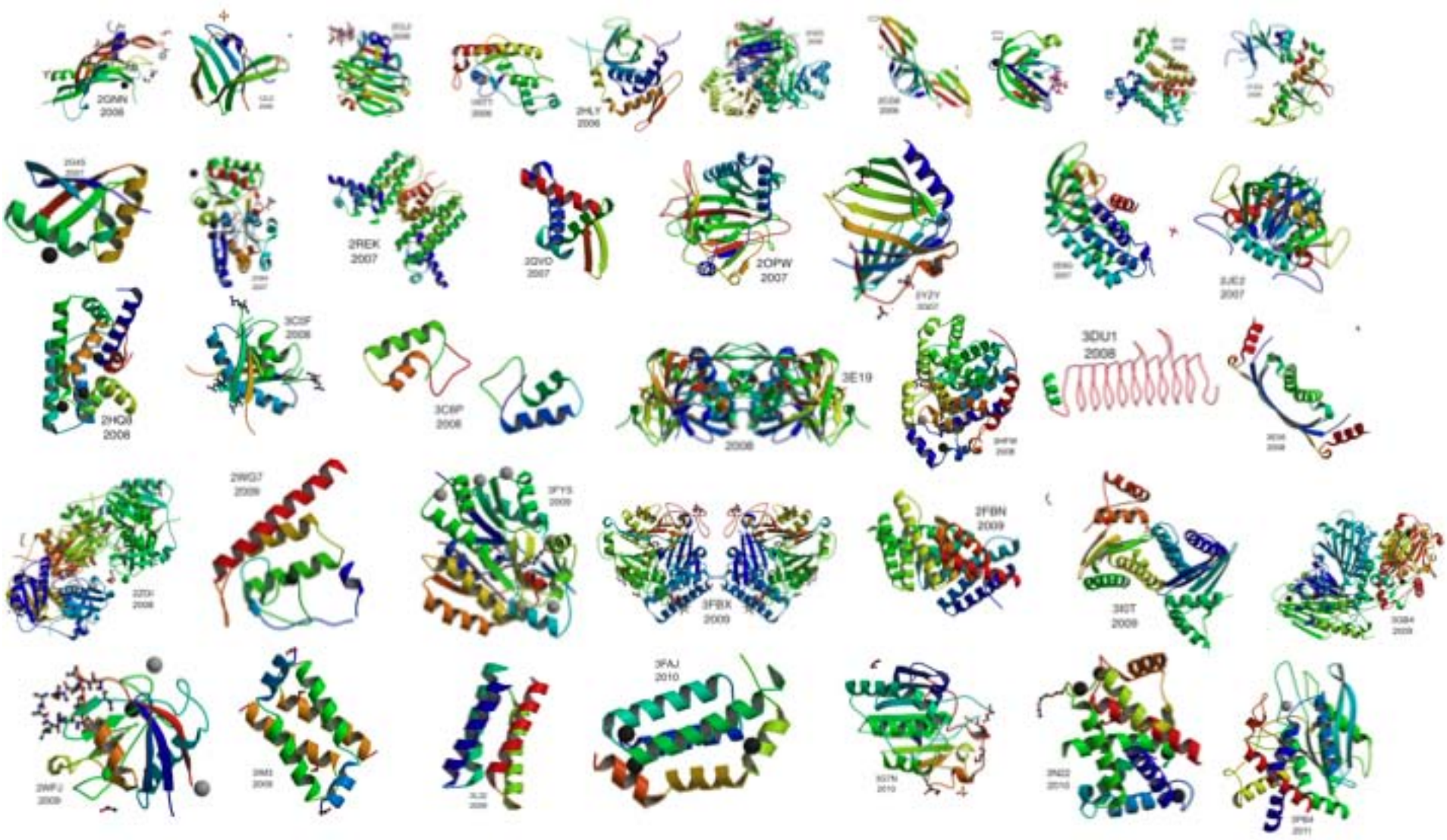
Crambin (1981) **Obelin (2000)**

19 years



Prepared by Profs. Gary Newton and John Rose University of Georgia, March 2011

39 S-SAD Structures Deposited in PDB from 2006 to 2011



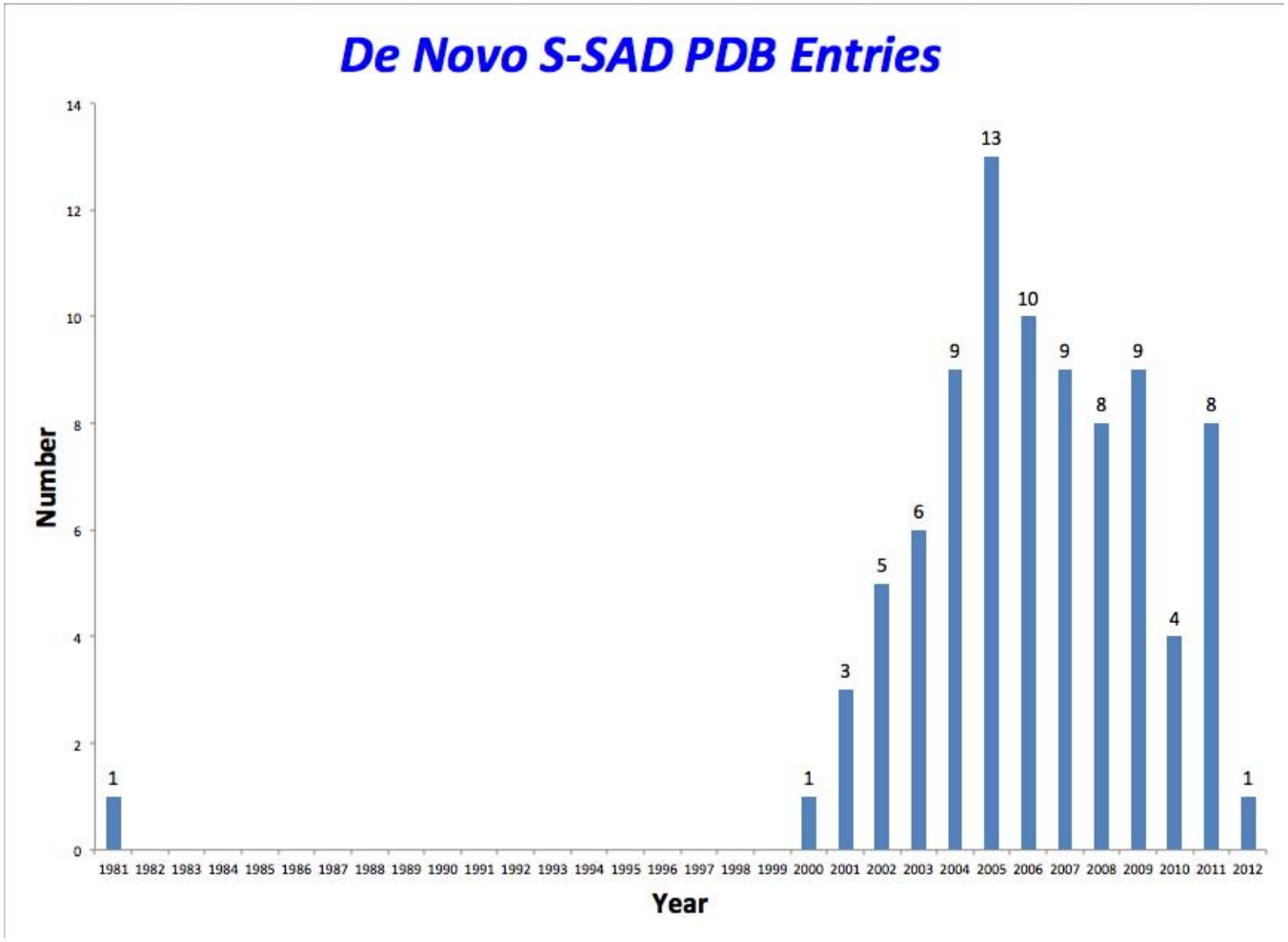
Prepared by Profs. Gary Newton and John Rose University of Georgia, March 2011

9 S-SAD Structures Deposited in PDB since May, 2011

8 more in 2011:
3AII
2Y7P
4A7D
3TBD
2YIL
2YIO
2XRH
3UI5

1 in 2012
3QB8

A total of 85 S-SAD Structures deposited in PDB



Outline

- *The SAD Methods - 1940s - 1980s*
  - *Resolving the phase ambiguity in SAD data*
  - *Determining the handedness*
- *The Use of S-SAD and X-SAD Methods - 1980s - 2012*
- ***Continued Efforts on S-SAD and X-SAD Phasing - 2000 -***
  - ***Various approaches for improving I/σ (I) in data***
  - ***The use of softer (extended wavelength) X-rays***
- *A Shared "Resource" at the APS to Advance Extended Wavelength MX - 2010 -*
- *Database SSAD_DB at UGA - 2011 -*

Various Approaches for Improving I/σ (I)

1. *Signal-Based Data Collection from One or multiple Crystals*

2. *Data Processing by a "Good Translator"*

3. *The MDS Data Collection Approach Using a Single Crystal - radiation slicing*

4. *Data Collection Using Highly Sensitive Detectors*

1. Signal-Based Data Collection (SBDC)

An automated data collection platform aimed at producing a data set having a "preset" anomalous scattering signal known to be sufficient for solving a macromolecular structure.

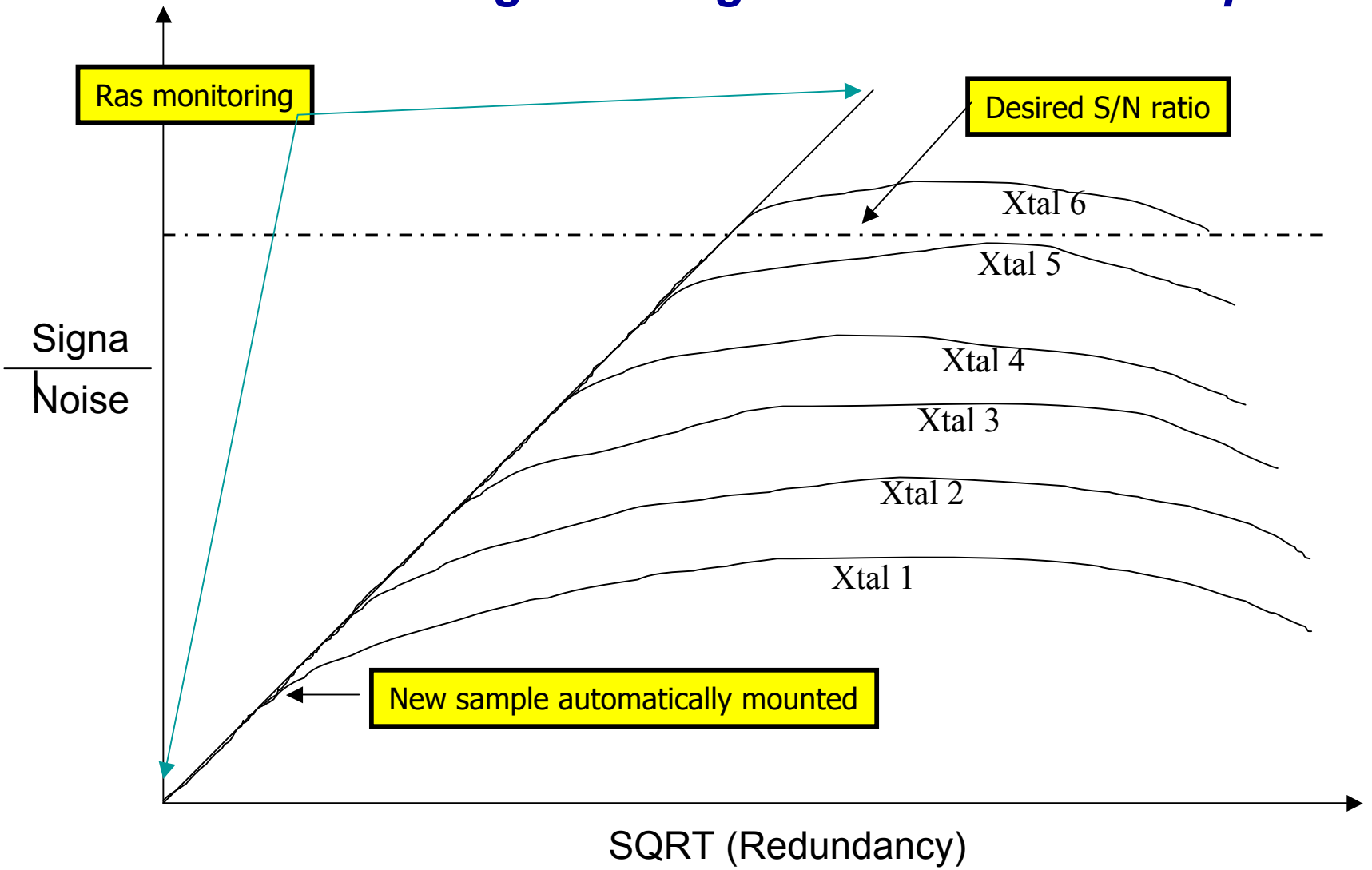
Ultimate Goal

By combining technologies already available, one may achieve total automation. This includes routine sulfur phasing at SER-CAT's beamlines.

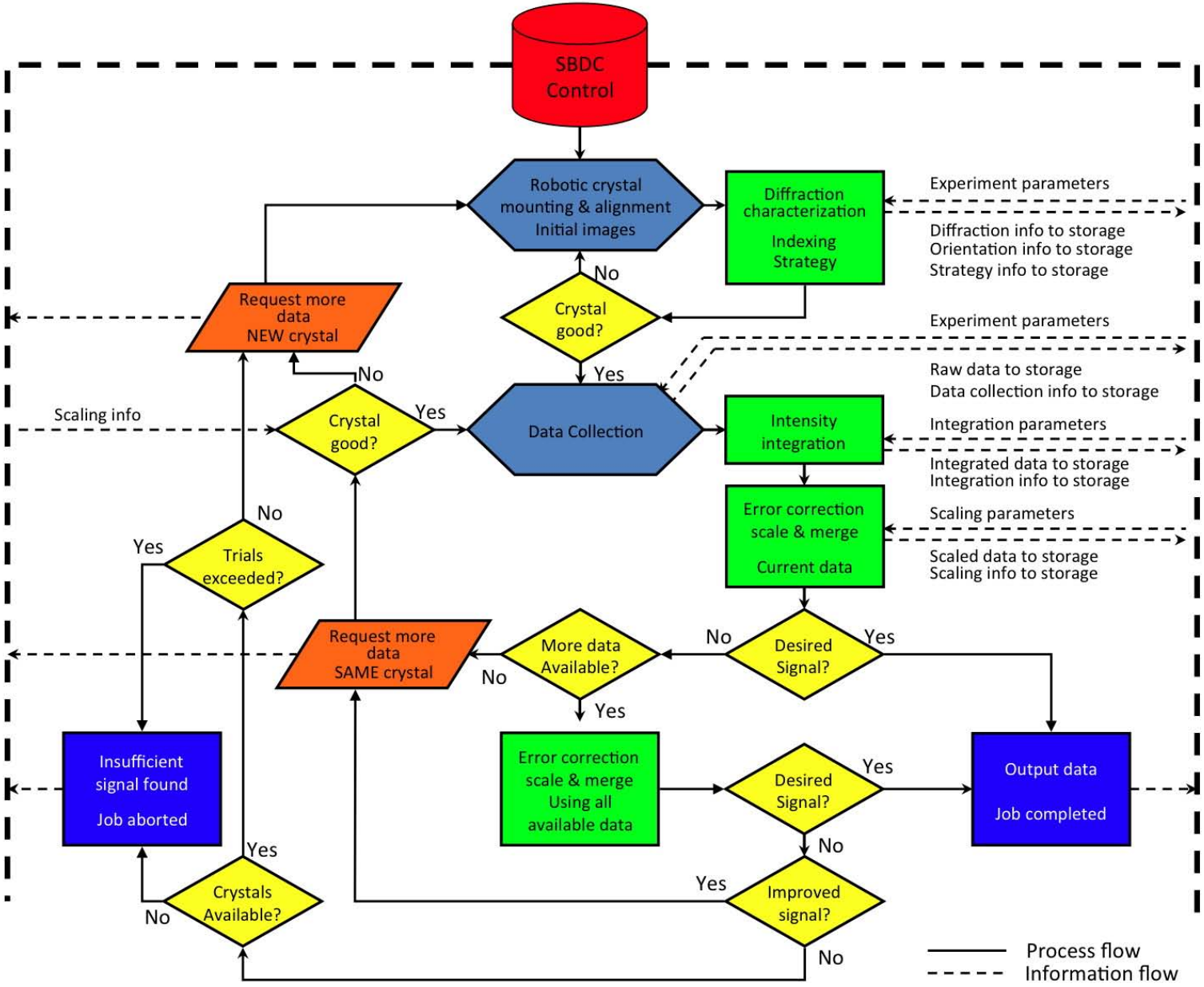
The SBDC platform is illustrated on the next slide

2005 – Signal Based Data Collection

Increasing S/N using robotics and multiple crystals
To meet the remaining challenges of routine sulfur phasing.



2005 – Signal Based Data Collection



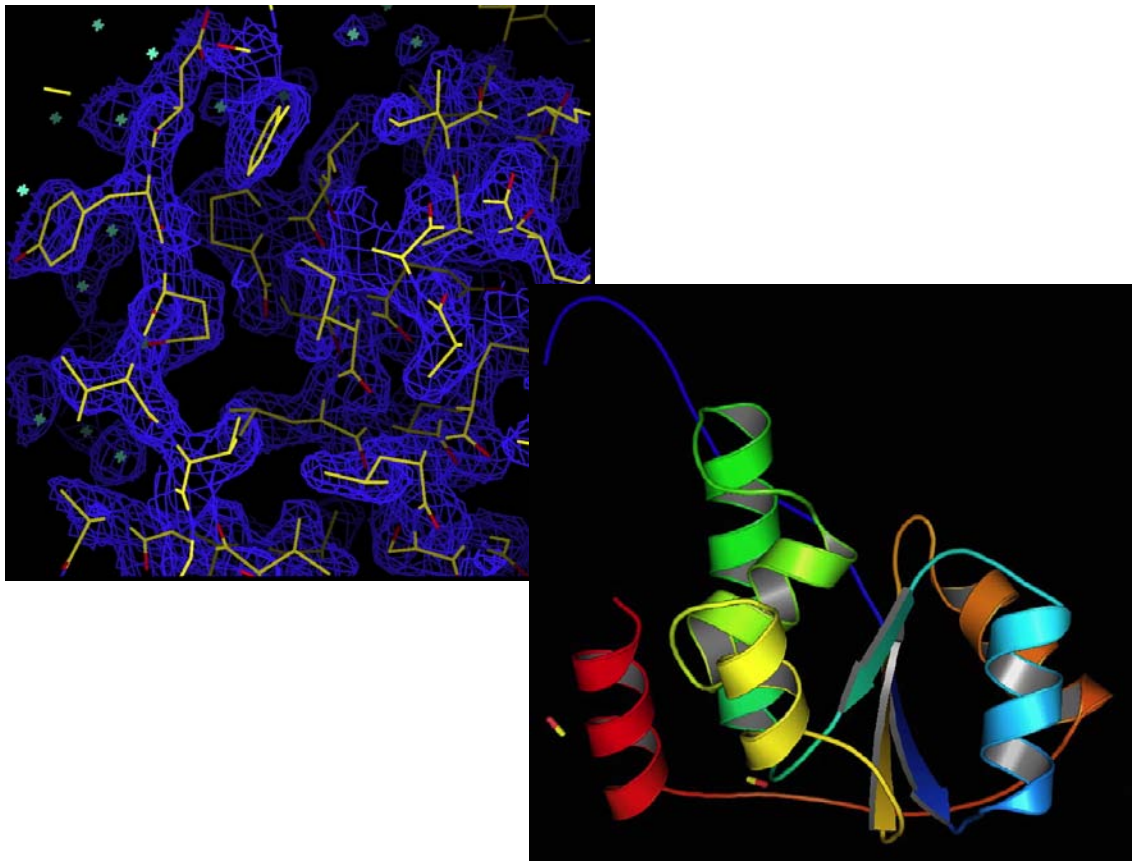
PPCW2005, NIH, Feb. 2-3, 2005

1. Multi-Crystal Averaging

2005 – PF0523 Solved by Merging 2 Data Sets

Protein	*PF0523*
# residue / sulfurs	**150 / 4**
Sequence	Poly ALA
X-ray Source	Cr-RAG
Wavelength(Å)	2.2909
Orientation	Random
Oscillation step (°)	1.0
# images	**360 X 2**
Exposure time (s)	600
Cell Constants (Å)	53.7, 86.6
Space Group	$P3_121$
Resolution(Å)	2.4
Completeness(%)	100
Redundancy	37.9
Rmerg (%)	5.5
Rmerg (high shell)	17.2
$\langle I/s_I \rangle$	137.2
PDB Entry	1ZD0

Structure determined by
SCA2Structure
72/150 A.A. fit by Arp/wARP

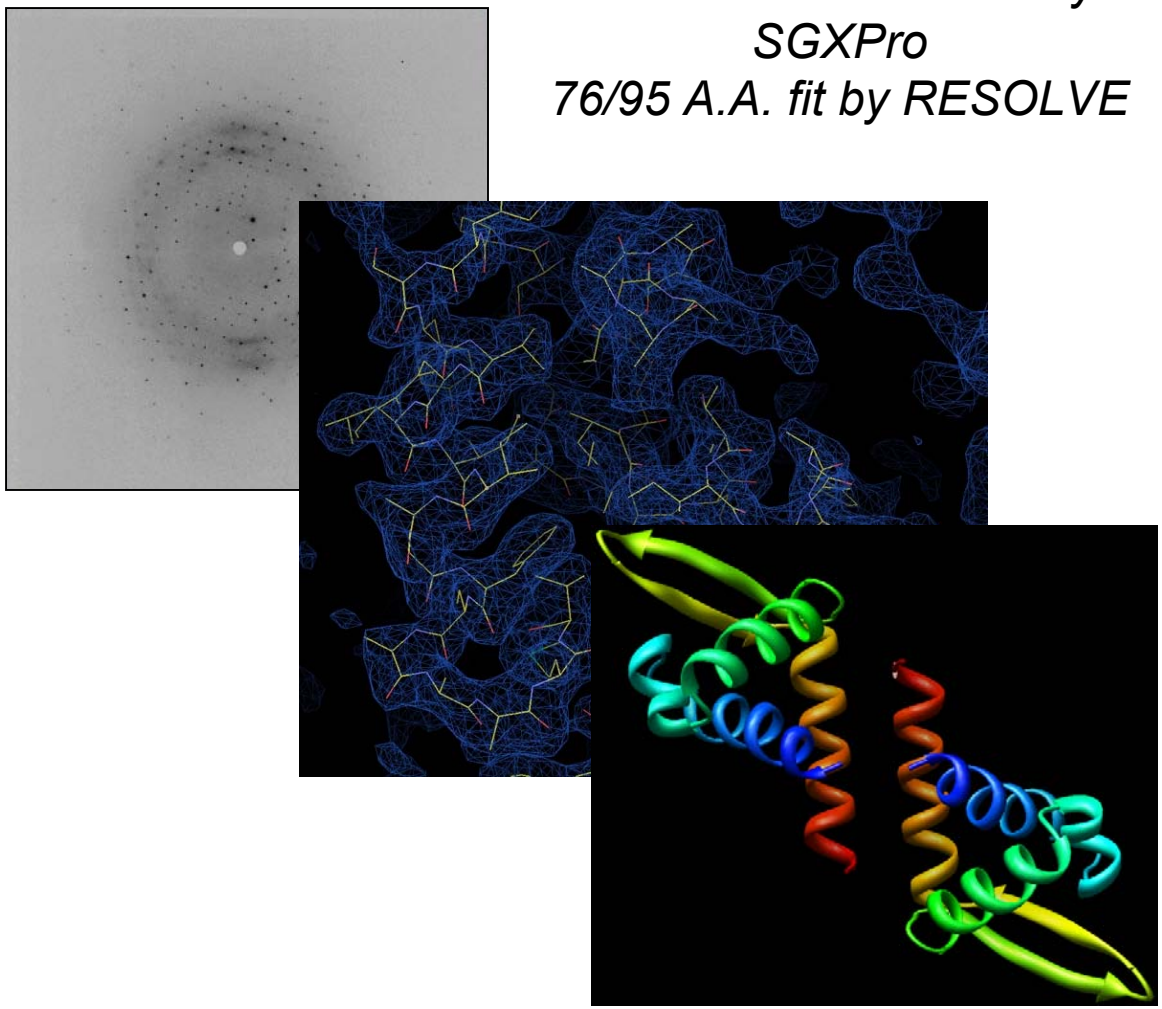


J. Habel, University of Georgia, 2005

1. Multi-Crystal Averaging
2007 - AF1382 Solved by Merging 2 Data Sets

Protein	AF1382
# residue / sulfurs	**95 / 5**
Sequence	Poly ALA
X-ray Source	22-ID
Wavelength(Å)	1.9
Orientation	Random
Oscillation step (°)	1.0
# images	**360 X 2**
Exposure time (s)	3
Cell Constants (Å)	53.5, 41.3
Space Group	$P4_2$
Resolution (Å)	2.3
Completeness(%)	100
Redundancy	25.3
Rsym(%)	4.8
Rsym (high shell)	26.0
$\langle I/s_I \rangle$	85.7
PDB Entry	303K

Structure determined by SGXPro
76/95 A.A. fit by RESOLVE



Two data sets processed with HKL2000

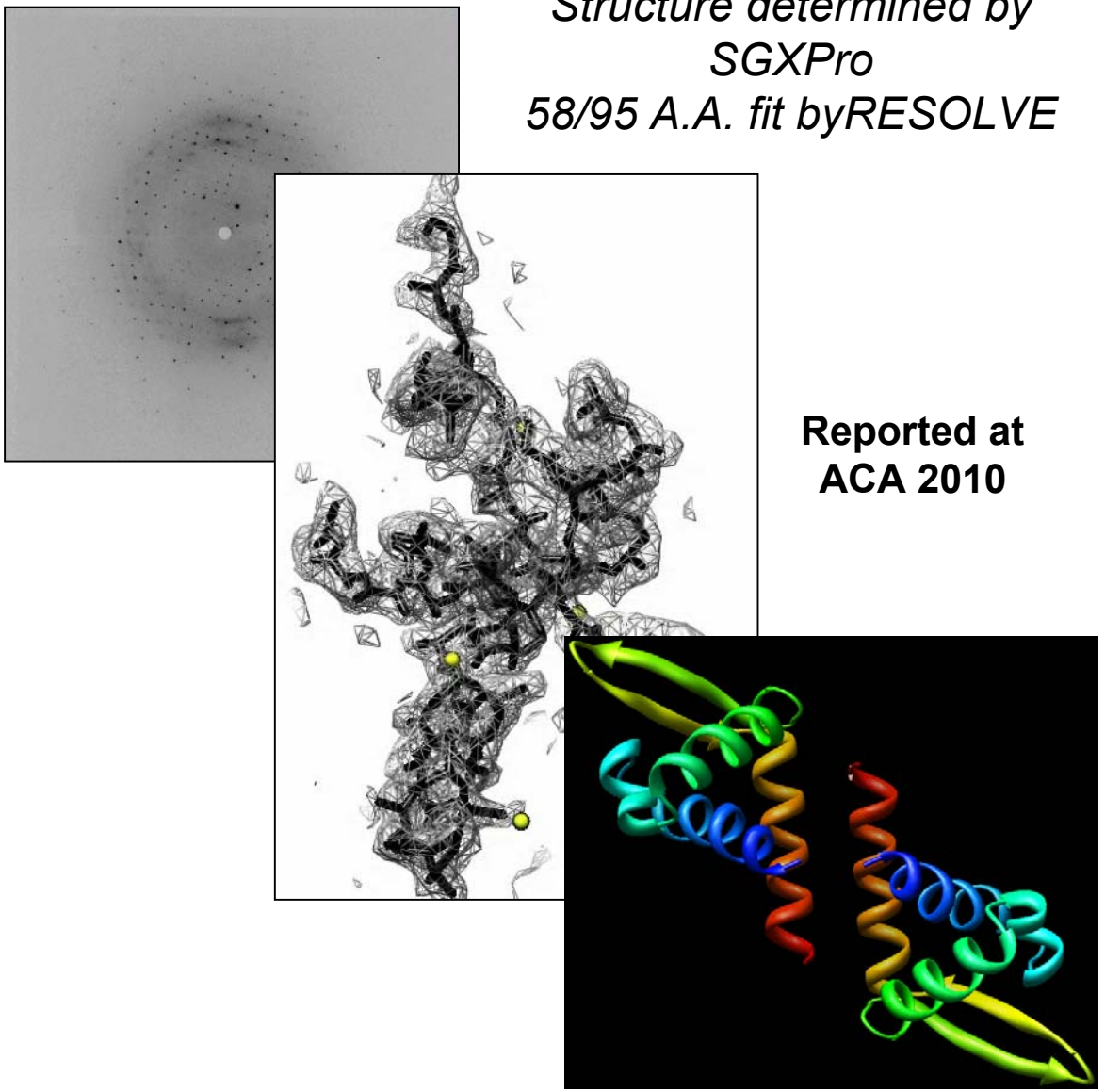
J. Zhu, University of Georgia, 2008

2. Data Processing by A Different "Translator"

2010 – AF1382

Protein	AF1382
# residue / sulfurs	**95 / 5**
Sequence	Poly ALA
X-ray Source	22-ID
Wavelength(Å)	1.9
Orientation	Random
Oscillation step (°)	1.0
# images	**360 X 1**
Exposure time (s)	3
Cell Constants (Å)	53.5, 41.3
Space Group	$P4_2$
Resolution (Å)	2.3
Completeness(%)	100
Redundancy	25.3
Rsym(%)	4.8
Rsym (high shell)	26.0
$\langle I/s_I \rangle$	85.7

Structure determined by SGXPro
58/95 A.A. fit byRESOLVE



Reported at ACA 2010

Single data set processed with Proteum 2

J. Swindell, University of Georgia, 2010

3. MDS Data Collection Using a Single Crystal

The MDS * (*M****ultiple* ***D****ata* ***S****ets / Crystal) Approach*
- Each data set with reduced radiation dosage

This is an idea of taking what might be called "slices" of radiation dosage" to produce multiple scans.

* ***Liu et al., Acta Cryst., (2011), A67, 544-549***

Basis of the MDS Approach

Look at the Equation for Sigma

During the 1960's when single counting diffractometers for small molecules were developed, reflection Sigma values were calculated by the following formula:

$\sigma_{total} = (\sigma_{Is}^2 + \sigma_{Ins}^2)^{1/2}$ *(counting statistics + Instrumental error)*

$= k\,(Sc_{peak} + Sc_{bg} + \varepsilon^2 \times Sc^2)^{1/2}$

where σ - Sigma

Sc - photon scan counts

ε - Experimental (ignorance) factor, generally $0.02 < \varepsilon < 0.10$

When the area detectors were developed in the 1980's for macromolecular data collection, the Sigma values of individual reflections from the 2-D detectors were also modeled by the two types of above errors. For example,

$$\sigma^2_{total} = \sigma_{Is}^2 + m\sigma_{Ins}^2$$

$$= G[I_s + I_{bg} + (m/n)I_{bg}] + m(K/A)^2 I_s^2 \qquad (11.2.5.17)^*$$

* A.G.W. Leslie, Integration of Macromolecular Diffraction Data, *International Table for Crystallography, Volume. F,* page 214 (2001).

Basis of the MDS Approach

Looking at the 2nd Term of the Equation:

$$\sigma^2_{total} = \sigma_{Is}^2 + \boldsymbol{m\sigma_{Ins}^2}$$

$$= G[I_s + I_{bg} + (m/n)I_{bg}] + \boldsymbol{m(K/A)^2 I_s^2} \qquad (11.2.5.17)$$

σ^2_{total} increases quickly with increasing in $\boldsymbol{I_s}$.

How about if we reduce the exposure time by a factor of N, such that $\boldsymbol{I_j} = I_s/N$?

$$\sigma^2_j = G[\boldsymbol{I_{sj}} + \boldsymbol{I_{bgj}} + (m/n)\boldsymbol{I_{bgj}}] + m(K/A)^2 \boldsymbol{I_{sj}}^2$$

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

We then compensate the weaker data by:

1. **repeating data collection N times and**
2. **add the intensities of all the equivalent reflections together.**

Then,

$$I_s = I_{j1} + I_{j2} + I_{j3} + I_{j4} + \ldots\ldots + I_{jN} = N\, I_j$$

$$\sigma^2_{total} = \sigma^2_1 + \sigma^2_2 + \sigma^2_3 + \sigma^2_4 + \ldots\ldots + \sigma^2_N = N\,\sigma^2_j$$

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

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

The reflection's INTENSITY is recovered by the MDS approach AND the reflection's SIGMA value is improved!

Basis of the MDS Approach

Summary

Single data set approach:

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

Multiple Data set (MDS) approach:

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

This illustrates the theoretical advantage of applying the MDS approach over the traditional approach, which is summarized in the statement below:

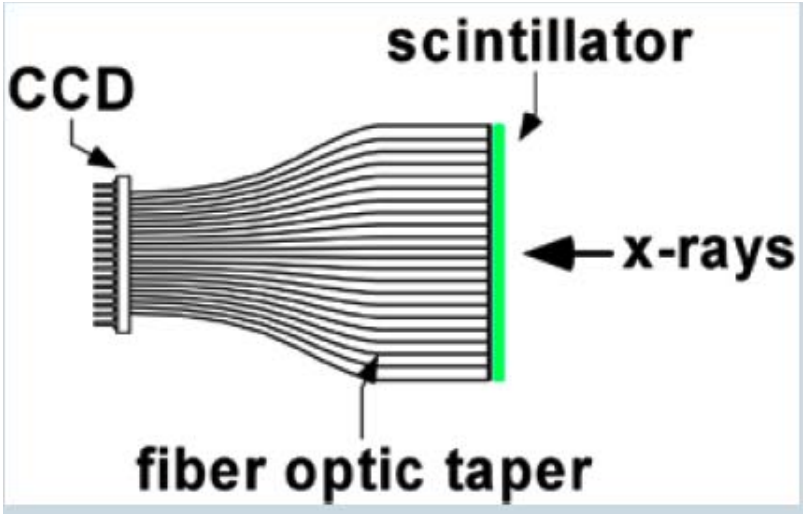
For a fixed X-ray dose, collecting Multiple Data Sets using short (low dose) exposures can produce better data than collecting a single set of data using long (high dose) exposures.

4. Data Collection Using A Highly Sensitive Detector

Further increase the $I/\sigma(I)$
using a CCD detector with a smaller taper ratio
(those retaining a higher ratio of
electrons per x-ray photon)

	Large CCD taper ratio detectors	Small taper ratio detectors
1.00 Å	~8.5 e/photon	
1.54 Å	~6.5 e/photon	up to 80 e/photon

A CCD Detector

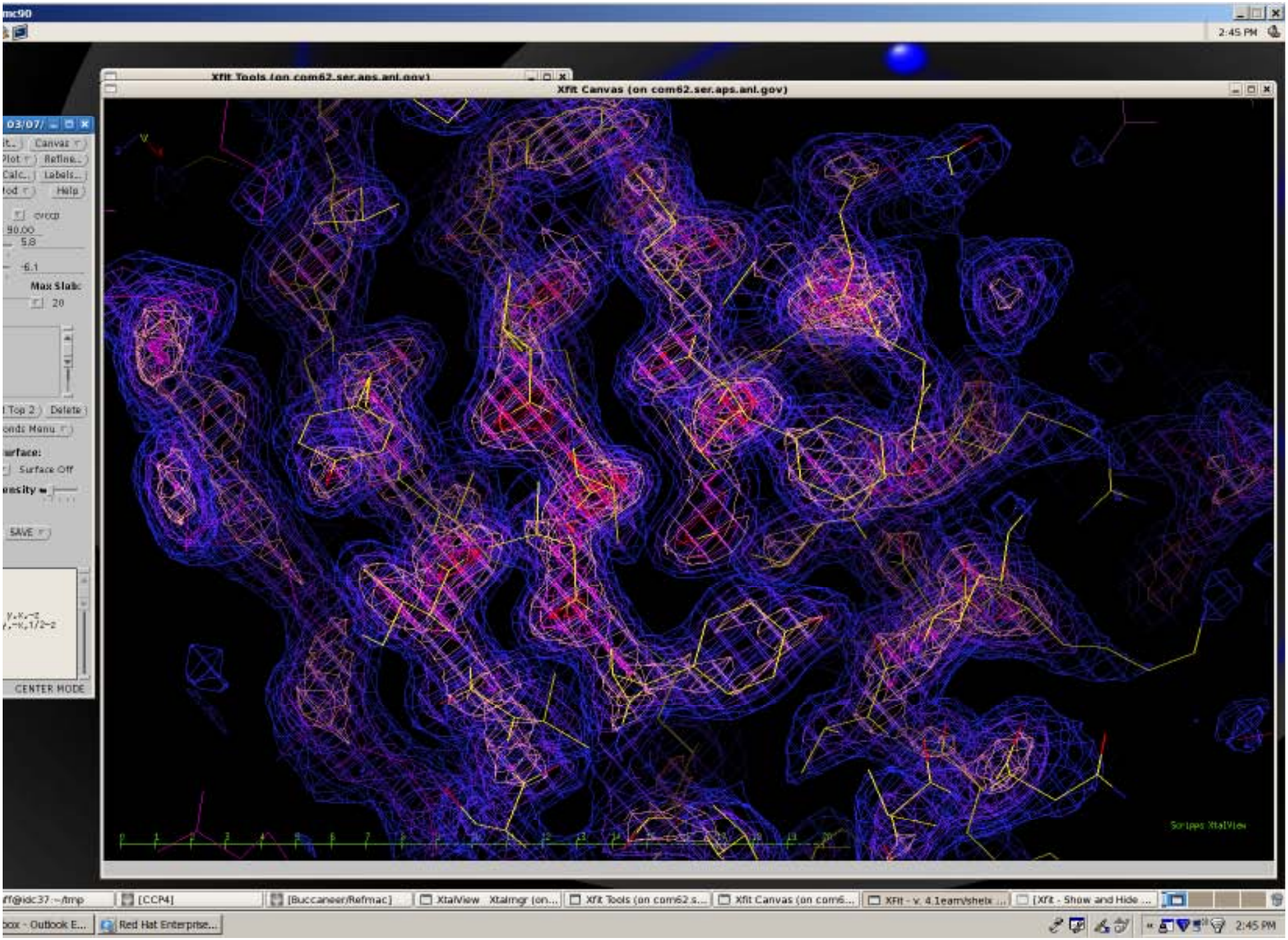


The X-rays excite a scintillator screen to produce visible photons. These photons are focused onto a CCD imager where they produce electrons.

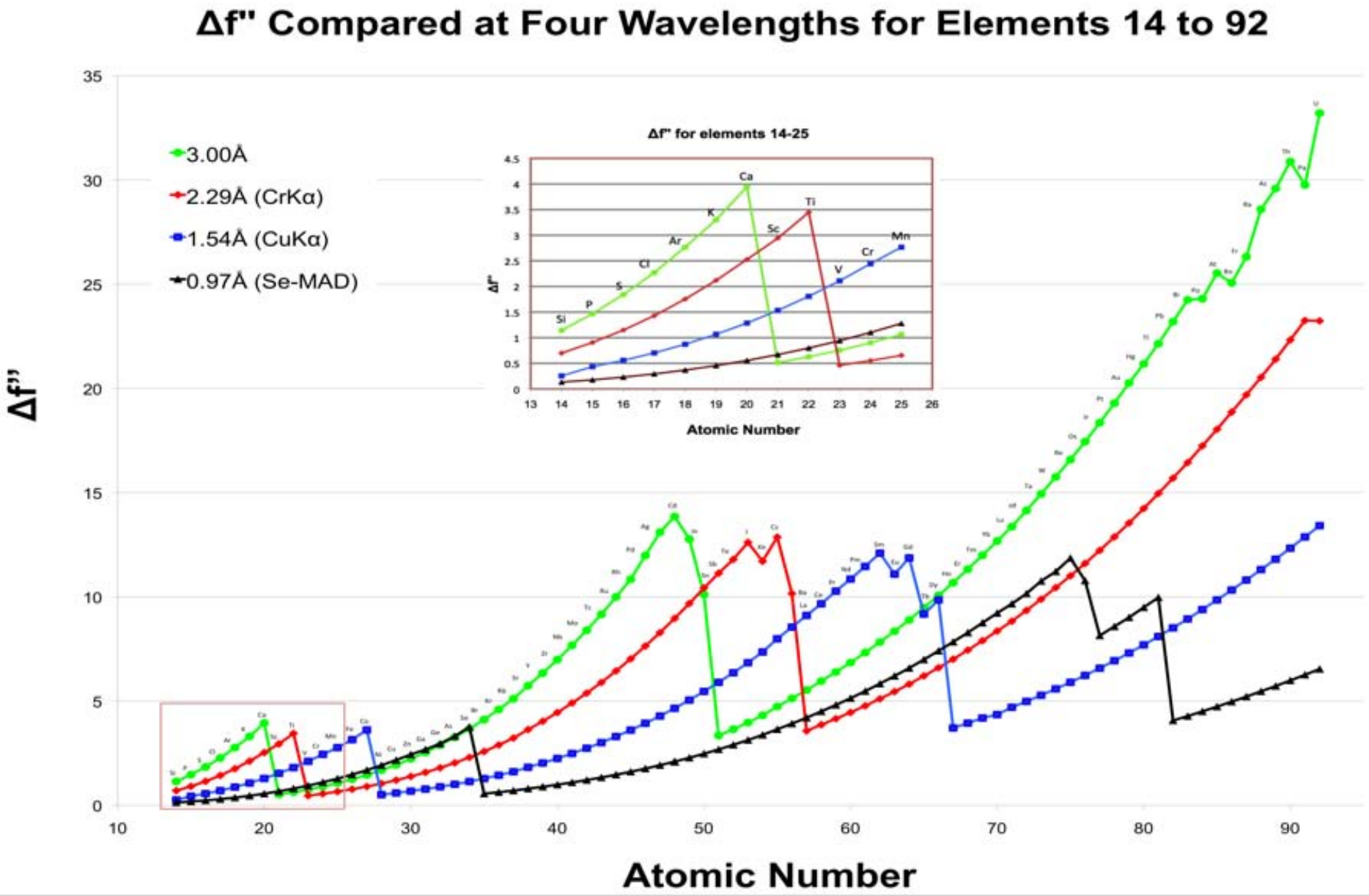
The transmission of fiber optics decreases with the square of the taper ratio. The larger the taper ratio the higher the light loss, and the less sensitive of the detector in terms of the # of light photons arriving at the CCD per X-ray.

Test results of using a detector with a smaller taper ratio

1. Initial crystal screening on UGA's CuKα home RAG X-ray source, this test crystal was observed to diffract to 2.8 Å.
2. Eight data sets with a total of 8,600 images (0.2° scan/image) from one crystal were collected in 30 hours on a detector with a sensitivity of 40 e^-/photon (Bruker's Platinum 135 detector on a home Cu X-ray Source).
3. The structure was originally determined by Se-SAD and refined to 2.7 Å (cutoff-sigma = 2) in 2006. It is now re-determined by S-SAD and refined to 2.2 Å.

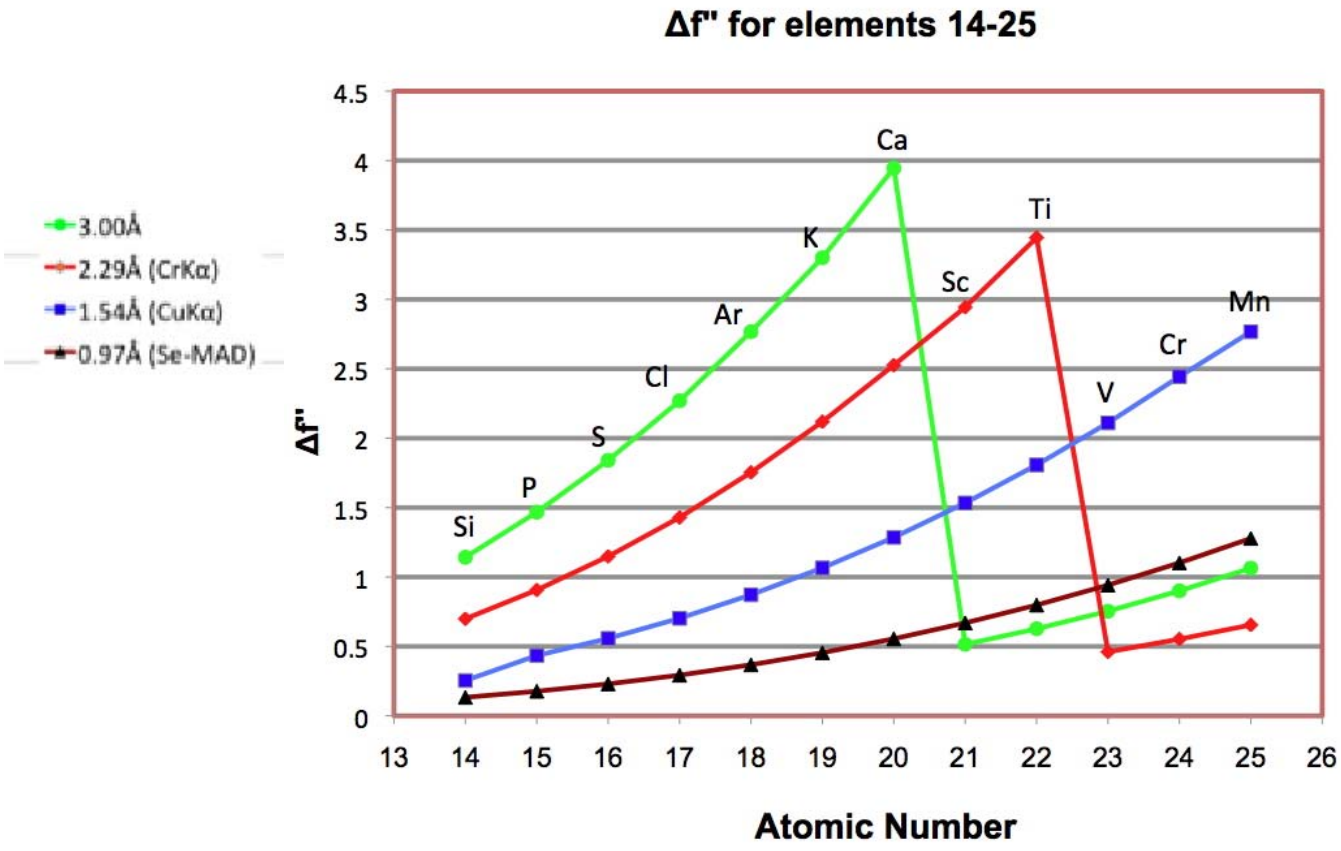

Xfit Canvas (on com62.ser.aps.anl.gov)
Canvas
Plot
Refine...
Calc...
Labels...
Help
Max Slab:
20
Top 2
Delete
Surface Off
SAVE
CENTER MODE
Scripps XtalView
2:45 PM
[CCP4]
[Buccaneer/Refmac]
XtalView
Xfit Tools (on com62.s...
Xfit Canvas (on com6...
[Xfit - Show and Hide ...
Red Hat Enterprise...
2:45 PM

The Use of softer (Extended Wavelength) X-rays



Advantages of Using Longer Wavelength X-rays

a. The increased anomalous scattering signal of some important light atoms would allow the use of unlabeled native crystals for routine structure determination.



b. Thirty seven (37) elements having large Δf" in this wavelength region (1.48 to 3.10 Å) can be used as X-SAD phasing probes, if needed.

Atomic No.	20	21	22	23	24	25	26	27	28	48
Element	**Ca**	**Sc**	**Ti**	**V**	**Cr**	**Mn**	**Fe**	**Co**	**Ni**	**Cd**
K edge, Å	3.0704	2.7596	2.4965	2.2687	2.0701	1.8961	1.7433	1.6083	1.4879	
L1 edge, Å										3.0857

Atomic No.	50	51	52	53	54	55	56	57	58	59
Element	**Sn**	**Sb**	**Te**	**I**	**Xe**	**Cs**	**Ba**	**La**	**Ce**	**Pr**
L1 edge, Å	2.7700	2.6389	2.5102	2.3898	2.2738	2.1697	2.0703	1.9786	1.8932	1.9140
L2 edge, Å	2.9832	2.8304	2.6883	2.5553	2.4293	2.3134	2.2047	2.1048	2.0114	1.9251
L3 edge, Å		3.0040	2.8559	2.7207	2.5926	2.4738	2.3630	2.2614	2.1663	2.0788

Atomic No.	60	61	62	63	64	65	66	67	84	85
Element	**Nd**	**Pm**	**Sm**	**Eu**	**Gd**	**Tb**	**Dy**	**Ho**	**Po**	**At**
L1 edge, Å	1.7399	1.6692	1.6025	1.5398	1.4803					
L2 edge, Å	1.8446	1.7680	1.6957	1.6277	1.5634	1.5025	1.4449			
L3 edge, Å	1.9972	1.9195	1.8460	1.7771	1.7118	1.6500	1.5916	1.5362		
M1 edge, Å									2.9883	2.8720

Atomic No.	86	87	88	89	90	91	92			
Element	**Rn**	**Fr**	**Ra**	**Ac**	**Th**	**Pa**	**U**			
M1 edge, Å	2.7663	2.6652	2.5712	2.4787	2.3926	2.3101	2.2348			
M2 edge, Å							2.3925			
M3 edge, Å							2.8811			

c. Thirty seven (37) elements having large Δf" in this wavelength region (1.48 to 3.10 Å) can be used as X-SAD phasing probes, if needed.

Atomic No.	20	21	22	23	24	25	26	27	28	48
Element	**Ca**	**Sc**	**Ti**	**V**	**Cr**	**Mn**	**Fe**	**Co**	**Ni**	**Cd**
K edge, Å	3.0704	2.7596	2.4965	2.2687	2.0701	1.8961	1.7433	1.6083	1.4879	
L1 edge, Å										3.0857

4.05 e⁻

Atomic No.	50	51	52	53	54	55	56	57	58	59
Element	**Sn**	**Sb**	**Te**	**I**	**Xe**	**Cs**	**Ba**	**La**	**Ce**	**Pr**
L1 edge, Å	2.7700	2.6389	2.5102	2.3898	2.2738	2.1697	2.0703	1.9786	1.8932	1.9140
L2 edge, Å	2.9832	2.8304	2.6883	2.5553	2.4300	2.3134	2.2047	2.1048	2.0114	1.9251
L3 edge, Å		3.0040	2.8559	2.7207	2.5926	2.4738	2.3630	2.2614	2.1663	2.0788

13.5 e⁻

10.9 e⁻

Atomic No.	60	61	62	63	64	65	66	67	84	85
Element	**Nd**	**Pm**	**Sm**	**Eu**	**Gd**	**Tb**	**Dy**	**Ho**	**Po**	**At**
L1 edge, Å	1.7399	1.6692	1.6025	1.5398	1.4803					
L2 edge, Å	1.8446	1.7680	1.6957	1.6277	1.5634	1.5025	1.4449			
L3 edge, Å	1.9972	1.9195	1.8460	1.7771	1.7118	1.6500	1.5916	1.5362		
M1 edge, Å									2.9883	2.8720

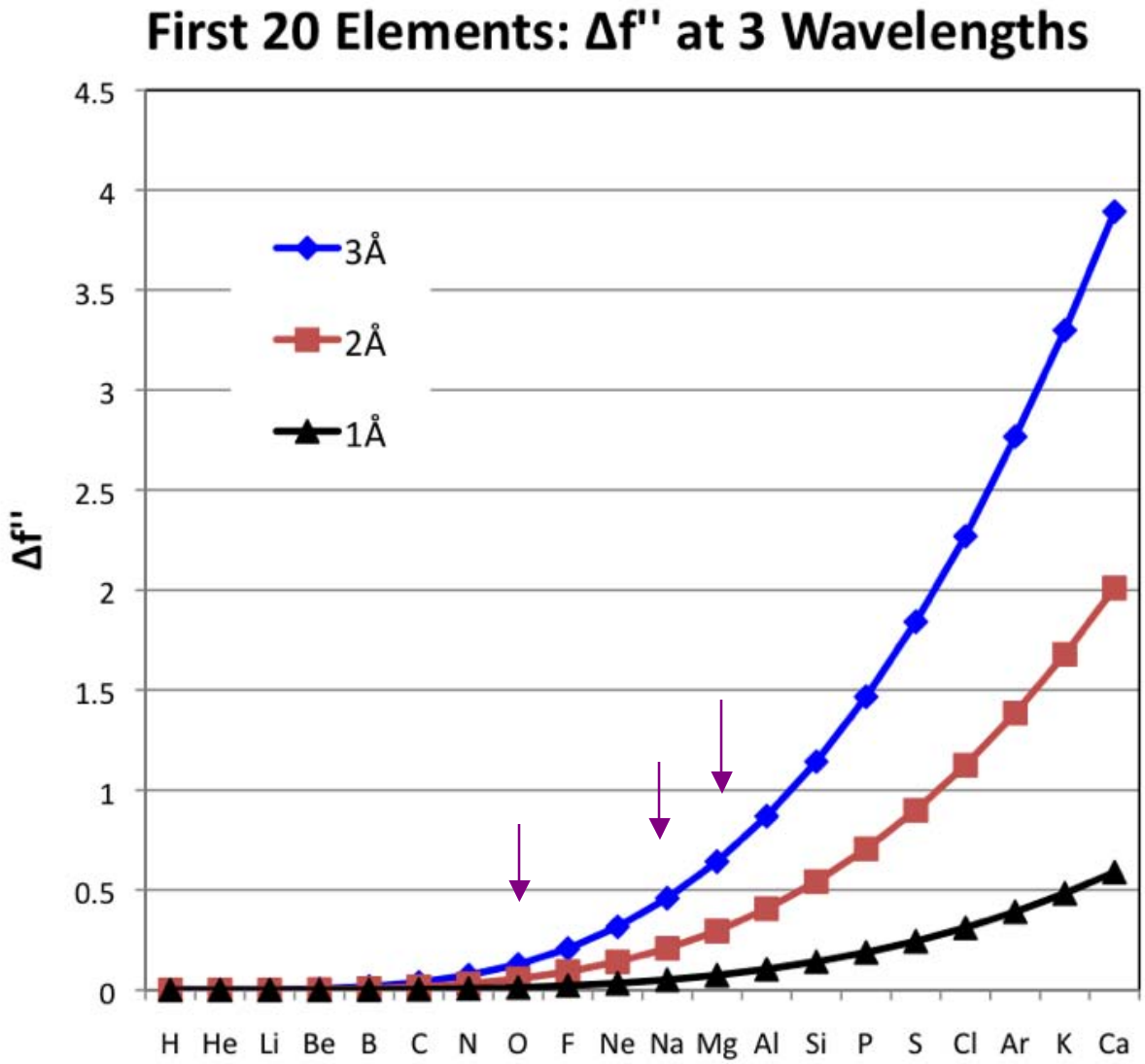
Atomic No.	86	87	88	89	90	91	92			
Element	**Rn**	**Fr**	**Ra**	**Ac**	**Th**	**Pa**	**U**			
M1 edge, Å	2.7663	2.6652	2.5712	2.4787	2.3926	2.3101	2.2378			
M2 edge, Å							2.0925			
M3 edge, Å							2.0611			

23.5 e⁻

35.0 e⁻

***Δf"* for: Ca - 4.05 e⁻(K); Xe - 13.5 e⁻(L1), 10.9 e⁻(L3) and U - 23.5 e⁻ (M1), 35.0 e⁻ (M3)**

C. The long-wavelength X-rays can help to confirm biologically relevant surface-bound or active site ions.

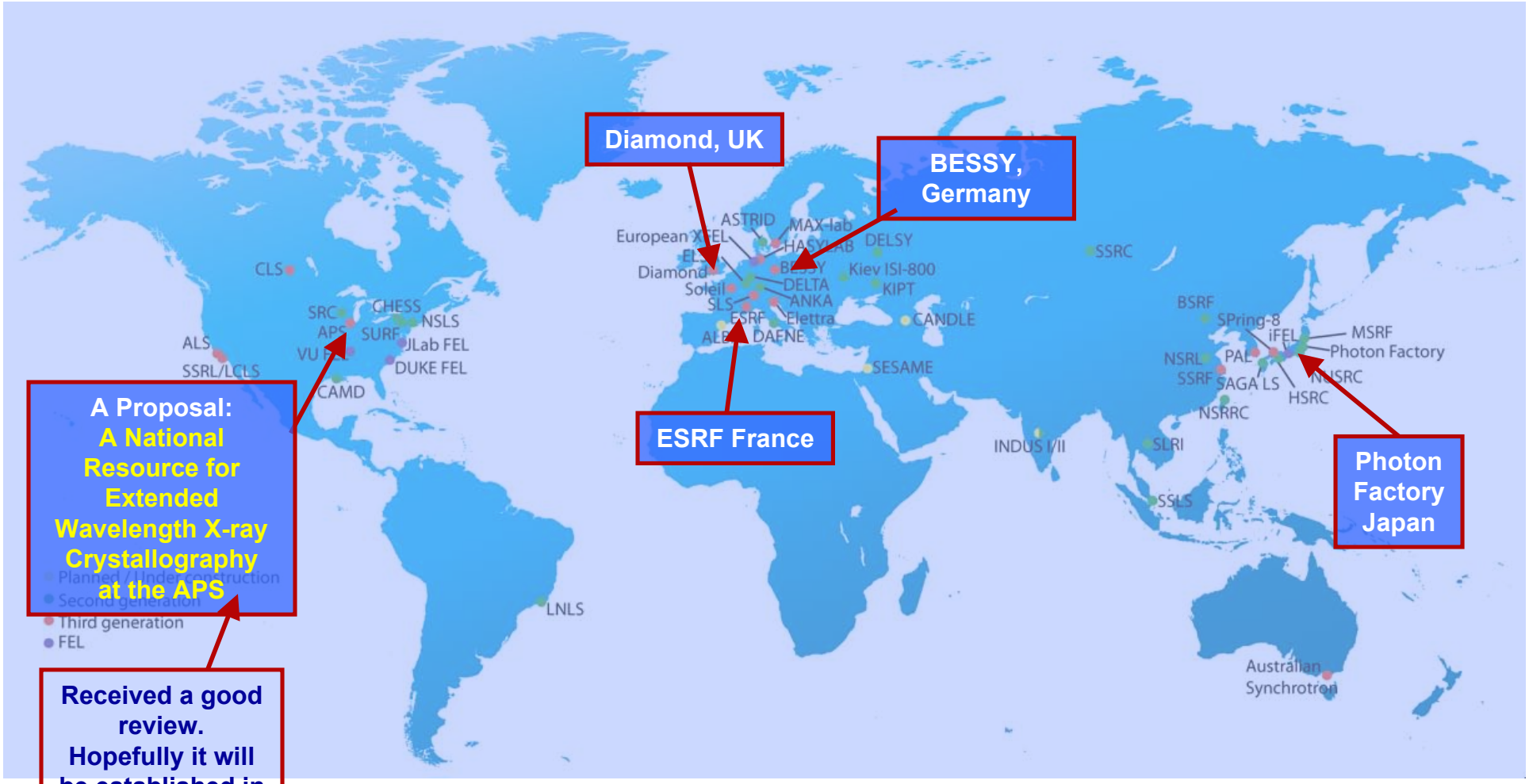


Outline

- *The SAD Methods - 1940s - 1980s*
  - *Resolving the phase ambiguity in SAD data*
  - *Determining the handedness*
- *The Use of S-SAD and X-SAD Methods - 1980s - 2000*
- *Continued Efforts for S-SAD and X-SAD Phasing - 2000 -*
  - *Various approaches for improving I/σ (I)*
  - *The use of softer (extended wavelength) X-rays*
- ***A Shared "Resource" at the APS to Advance Extended Wavelength MX - 2010 -***
- *Database SSAD_DB at UGA - 2011 -*

Map of Synchrotron Facilities in the World

Long-wavelength X-ray beamlines are already in use, or in development, at major synchrotron facilities, such as the Diamond (UK), ESRF (France), BESSY (Germany) and Photon Factory (Japan).



Map and photos were taken from the ESRF website

Photo of Workshop hosts, special supporters, invited speakers and Session Chairs

Group photo of most Workshop participants.

Group photos from APS Users Workshop on Extended Wavelength X-ray Crystallography, May 4, 2011

A Database SSAD_DB at UGA

02.01

SSAD_DB: a database of structures solved by sulfur SAD phasing and related experimental parameters

John Rose[1], Hua Zhang[1], Manfred Weiss[2], Bi-Cheng Wang[1]

[1]Dept. of Biochemistry & Molecular Biology, U. of Georgia, Athens, GA USA, [2]Helmholtz Zentrum Berlin für Materialien und Energie, Berlin, Germany

Rose, et al, 2011 ACA Meeting

Future Plans

(Slide Supplied by John Rose)

SSAD_DB

We have completed the Phase I data harvesting and have identified 87 de novo S-SAD structures (there are probably more).

We have developed a tentative CIF data dictionary.

We are currently harvesting as much information as possible about each structure from the PDB and primary Reference.

We have developed a data sheet that authors can use to update/add data in the SSAD_DB. (web based update is also planned).

We will soon publish an alpha version of the SSAD_DB at the following website:

www.SSAD_DB.uga.edu

How can you help?

If you have a structure solved by S-SAD

- please send the PDB ID to **John Rose (jprose@uga.edu)**

Once the SSAD_DB is on-line we will contact the PDB authors

- please verify the entry and update it if needed

Direct Determination of Macromolecular Structures by Crystallography

…. Although success is still limited, information on what is required for success is accumulating, and has enhanced our belief that routine direct determination of macromolecular structures by crystallography is attainable in the foreseeable future.

Acknowledgements

John Rose
Zheng-Qing (Albert) Fu
Gerd Rosenbaum
Jeff Habel
Liqing Chen
Weihong Zhou
Gary Newton
J. Murray Gibson
Matt Bennig
Joseph Ferrara
Manfred Weiss
Christoph Mueller-Dieckmann
Soichi Wakatsuki
Helen Berman
Zhi-Jie (James) Liu
Edward Wu
John Chrzas
Jinyi (Robin) Zhu
Lirong Chen
Hua Zhang
Robert Fischeti
Denny Mills
Cheng Yang
Nobuhisa Watanabe
Arimin Wagner
Meitian Wang
Samar Hasnain
Wayne Hendrickson

and many other collaborators

NIGMS/NIH, SER-CAT, APS

Georgia Research Alliance, IBM

University of Georgia Research Foundation