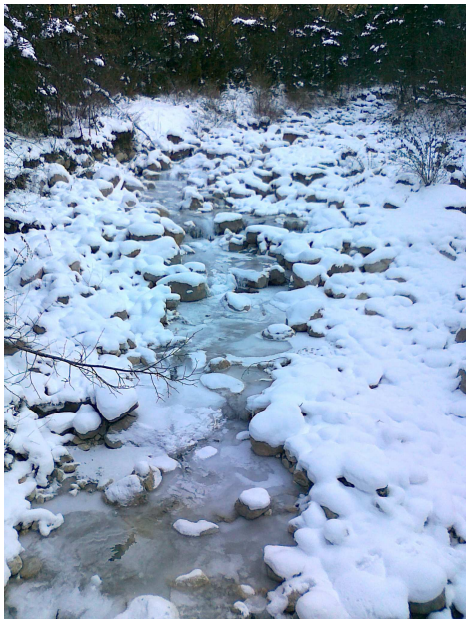
Why we should make more use of softer X-rays in macromolecular crystallography



Gordon Leonard

ESRF Structural Biology Group

4th Winter School on Soft X-rays in MX

What are Soft X-rays?

X-rays are at the short wavelength, high energy end of the electromagnetic spectrum. Only gamma rays carry more energy. It is convenient to describe x-rays in terms of the energy they carry, in units of thousands of electron volts (keV). X-rays have energies ranging from less than 1 keV to greater than 100 keV.

Hard x-rays are the highest energy x-rays, while the lower energy x-rays are referred to as soft x-rays. The distinction between hard and soft x-rays is not well defined. Hard x-rays are typically those with energies greater than around 10 keV.

$$\lambda(\text{Å}) = 12.398 / E(keV)$$

The lower (softer) the energy, the longer the wavelength

http://hesperia.gsfc.nasa.gov/sftheory/xray.htm

How can we use Soft X-rays in MX?

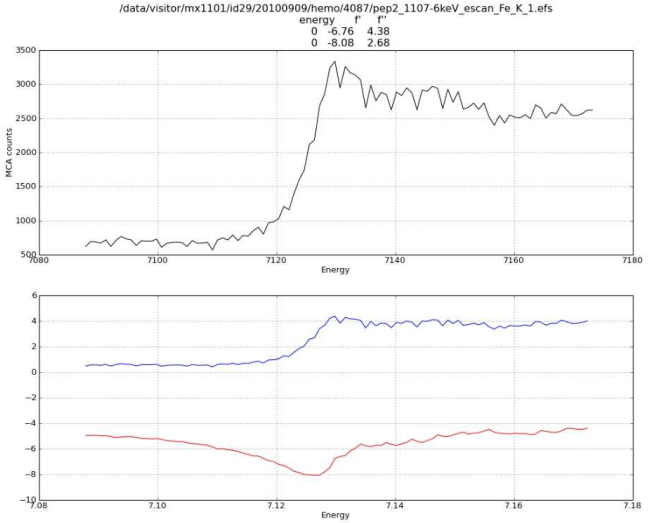
1. Producing *de novo* phasing information for structure determination
    1. Crystals of native proteins/nucleic acids
    2. Heavy atom derivatives
2. Improving phase information from molecular replacement protocols
3. Checking the correctness of chain tracing [both proteins and nucleic acids]
4. Improving models of macromolecule-'solvent' interactions

For all of this we take advantage of (enhanced) anomalous scattering properties of some elements at softer X-ray energies

$$F_{hkl} = \sum_j f_j \exp^{-2\pi i (hx_j + ky_j + lz_j)}$$

$$f = f_o + f'(\lambda) + if''(\lambda)$$

Wavelength-dependent; vary rapidly around an absorption edge



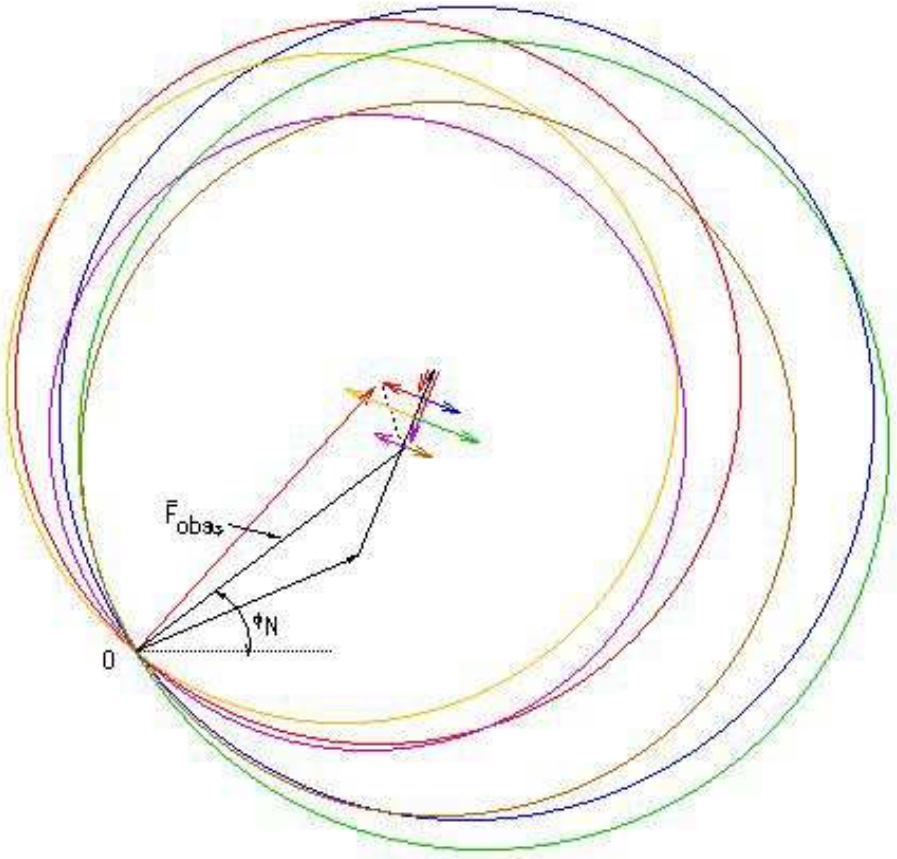
Anomalous differences: $F_{hkl}^{\lambda_i} \neq F_{\bar{h}\bar{k}\bar{l}}^{\lambda_i}$

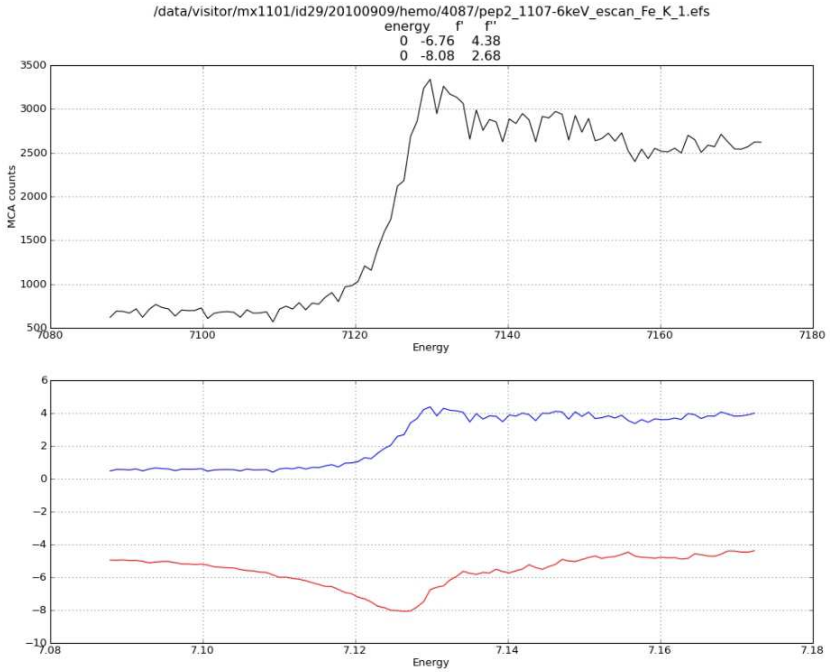
Dispersive differences: $F_{hkl}^{\lambda_i} \neq F_{\bar{h}\bar{k}\bar{l}}^{\lambda j}$

Can use these differences to:

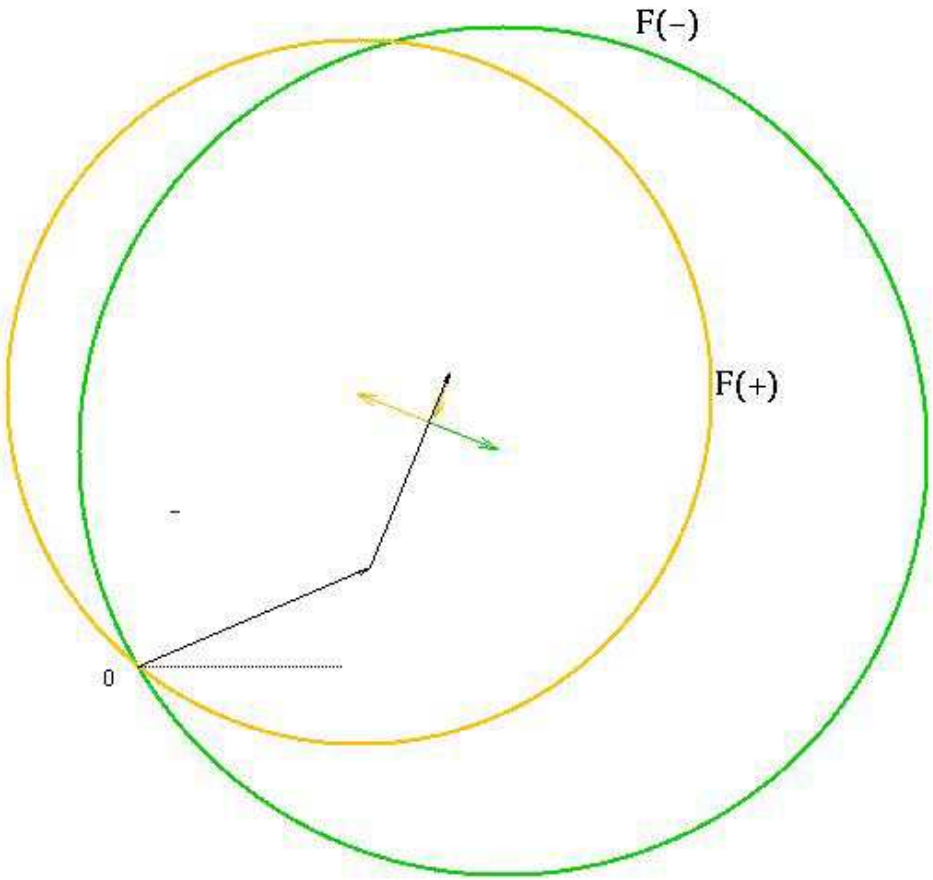
1. Derive phase information
2. Calculate 'difference' fouriers

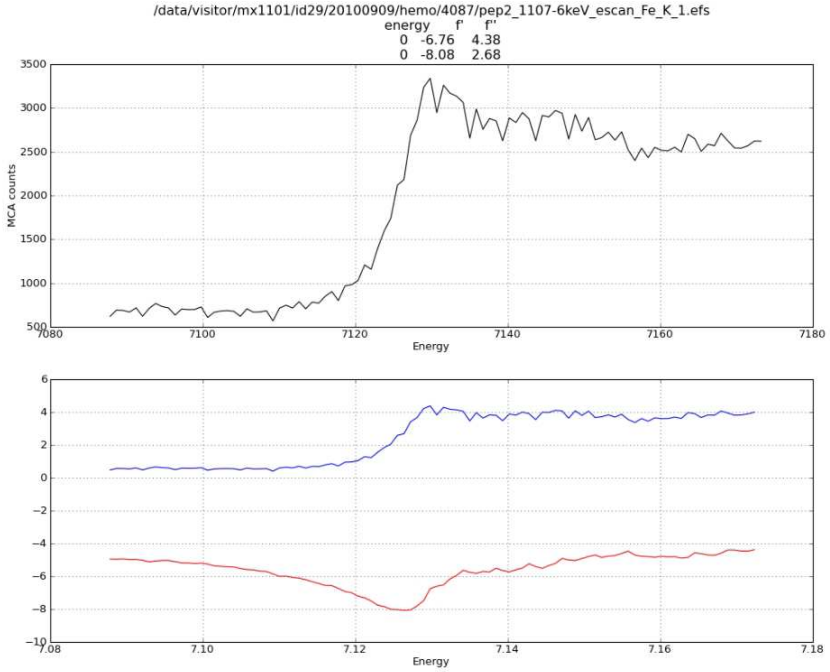
Deriving de novo phase information: MAD (as Quasi-MIR)





Deriving de novo phase information: SAD





$\Phi_p = \Phi_h \pm \Phi'$: Phase ambiguity but can get around this

"From a practical point of view a larger f" will give more accurate values for the possible solutions [of the phase]…."

Woolfson & Fan, "Solving Crystal Structures" (1995) Cambridge University Press

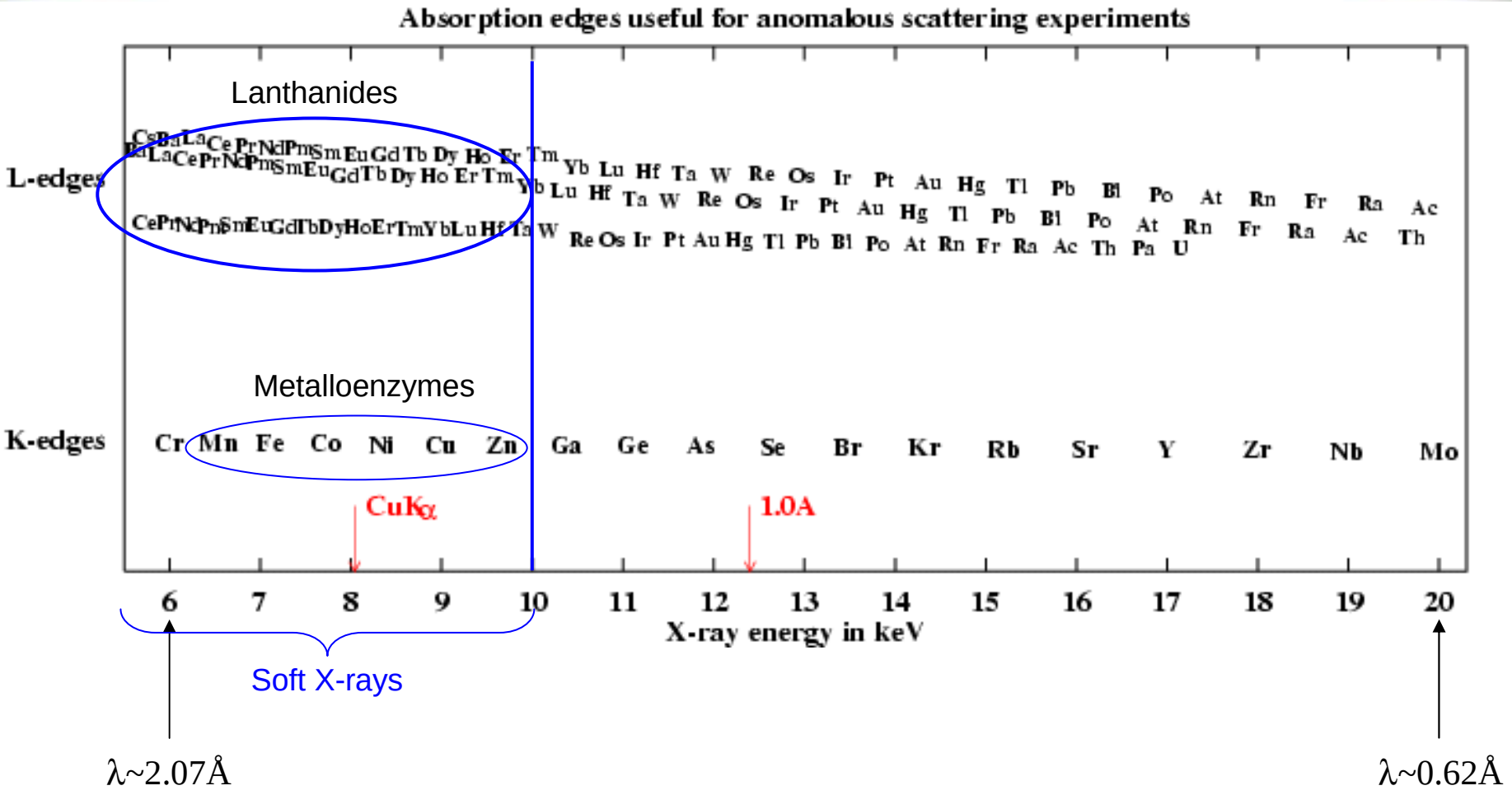
$$'signal\ '=<\Delta F\ /\ F>\approx \frac{1}{\sqrt{2}}\frac{\sqrt{N_A}\,2\,f''}{<|F_T|>}$$

K- absorption edges: f" usually has a maximum around **4e⁻**; can rise to **6-7e⁻** with 'white line'.

L- absorption edges: f" usually has a maximum around **12e⁻**; can rise to **~30e⁻** with 'white line'.

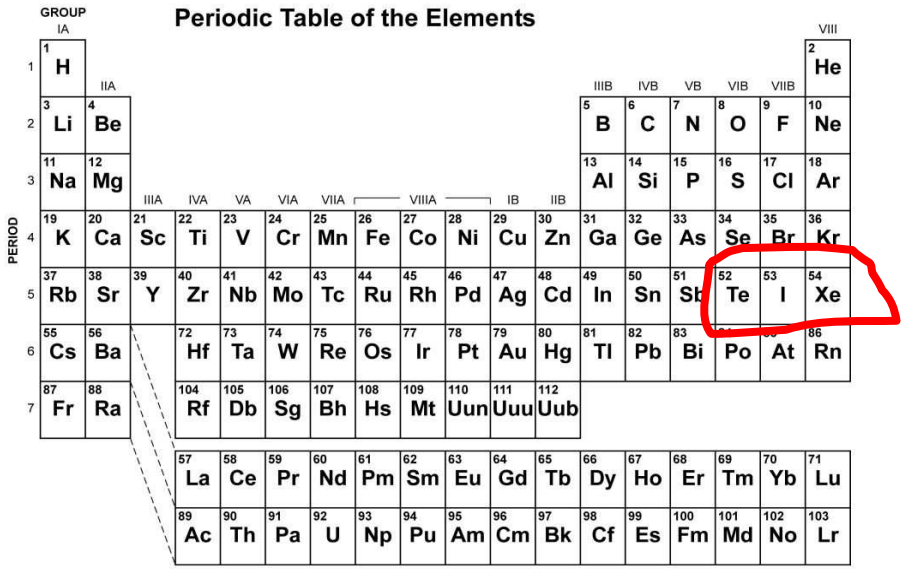
So (if you have a choice) L- absorption edges: will always give bigger signals

Routinely accessible energies at MX Beamlines



1. We already make a lot of use of soft x-rays – Metalloenzymes represent ~30% of all proteins
2. Are there potential benefits in extending this range?

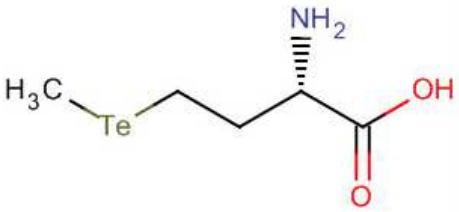
Useful absorption edges accessible with an extended wavelength range



K- edges:
Te: 31.8 keV (λ = 0.39 Å)
I: 33.2 keV (λ = 0.37 Å)
Xe: 34.6 keV (λ = 0.36 Å)

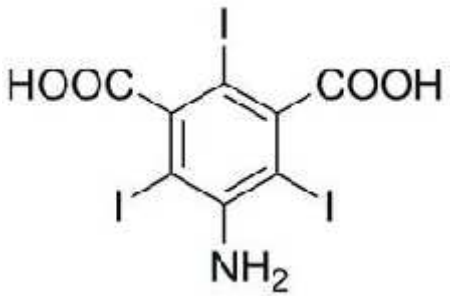
L- edges:
Te: 4.9; 4.6; 4.3 keV (λ = 2.5; 2.7; 2.9 Å)
I: 5.2; 4.8; 4.6 keV (λ = 2.4; 2.6; 2.7 Å)
Xe: 5.4; 5.1; 4.8 keV (λ = 2.3; 2.4; 2.6 Å)

Telluromethionine



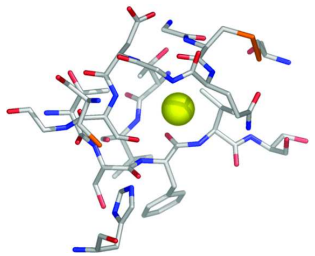
Budisa, N. *et al.*, (1997) *J. Mol. Biol.* **270**, 616-623.

'Magic Triangle'



Beck, T. *et al.*, (2008) *Acta Cryst.* **D64**, 1179-1182.

Xe derivative of elastase



Evans & Bricogne(2002) *Acta Cryst*. **D58**, 976-991.

html pdf cited in buy

J. Appl. Cryst. (2004). **37**, 925-933 [doi:10.1107/S0021889804023076]

Multi-wavelength anomalous diffraction method for I and Xe atoms using ultra-high-energy X-rays from SPring-8

K. Takeda, H. Miyatake, S.-Y. Park, M. Kawamoto, N. Kamiya and K. Miki

Abstract: The first successful multi-wavelength anomalous diffraction (MAD) experiments using ultra-high-energy X-rays (~35 keV) were performed for iodine and xenon derivatives of hen egg-white lysozyme crystals. The beamline BL41XU of SPring-8 enabled the collection of high-quality MAD data, which led to the calculation of anomalous or dispersive difference Patterson maps that determined the positions of iodine and xenon atoms. The electron density maps obtained by the density modification method for both cases proved to be of sufficient quality for building molecular models. I-MAD and Xe-MAD phasing are now available at SPring-8, and the utilization of ultra-high-energy X-rays will make a significant contribution to the solution of the phase problem in protein crystallography.

Keywords: multi-wavelength anomalous diffraction; macromolecular crystallography; protein crystallography; ultra-high-energy X-rays; electron density distribution.

html pdf buy

J. Appl. Cryst. (2006). **39**, 831-841 [doi:10.1107/S0021889806036387]

Anomalous diffraction at ultra-high energy for protein crystallography

J. Jakoncic, M. Di Michiel, Z. Zhong, V. Honkimaki, Y. Jouanneau and V. Stojanoff

Abstract: Single-wavelength anomalous diffraction (SAD), multiwavelength anomalous diffraction (MAD) and single isomorphous replacement with anomalous scattering (SIRAS) phasing at ultra-high X-ray energy, 55 keV, are used successfully to determine a high-quality and high-resolution experimental electronic density map of hen egg-white lysozyme, a model protein. Several combinations, between single- and three-wavelength, with native data were exploited to demonstrate that standard phasing procedures with standard equipment and software can successfully be applied to three-dimensional crystal structure determination of a macromolecule, even at these very short wavelengths. For the first time, a high-quality three-dimensional molecular structure is reported from SAD phasing with ultra-high-energy X-rays. The quality of the crystallographic data and the experimental electron density maps meet current standards. The 2.7% anomalous signal from three Ho atoms, at the Ho *K* edge, was sufficient to obtain a remarkable electron density and build the first lanthanide structure for HEWL in its entirety.

Keywords: ultra-high energy; phasing; MAD; SAD; SIRAS; HEWL; holmium.

1 fully occupied Xe or I or Te atom; 300 a.a. protein; no white line

$$'signal\ '=<\Delta F\ /\ F>\approx \frac{1}{\sqrt{2}}\frac{\sqrt{N_A}\,2\,f''}{<|\,F_T\,|>}$$

K- absorption edge: <ΔF/F> = 1.5%

L_I- absorption edge: <ΔF/F> = 5.8%

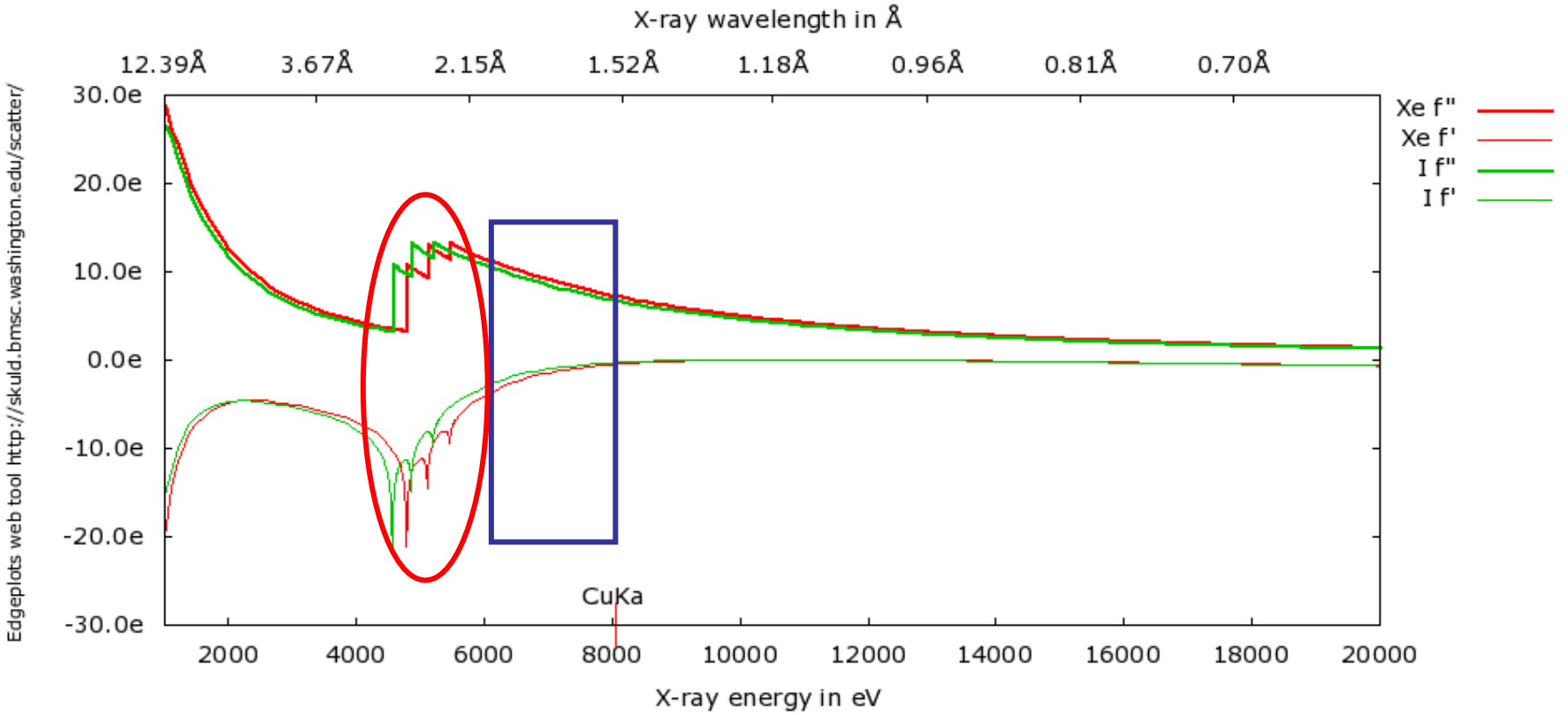
6 keV: <ΔF/F> = 5.0%

7 keV: <ΔF/F> = 3.9%

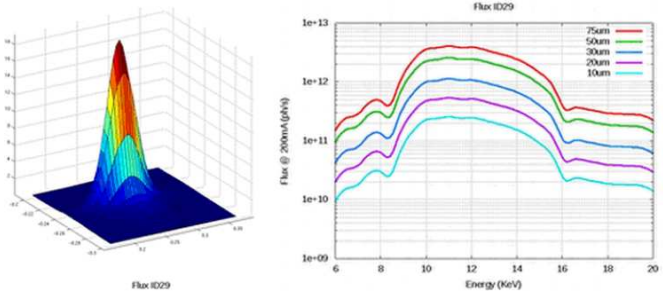
8 keV: <ΔF/F> = 3.2%

http://www.ruppweb.org/new_comp/anomalous_scattering.htm

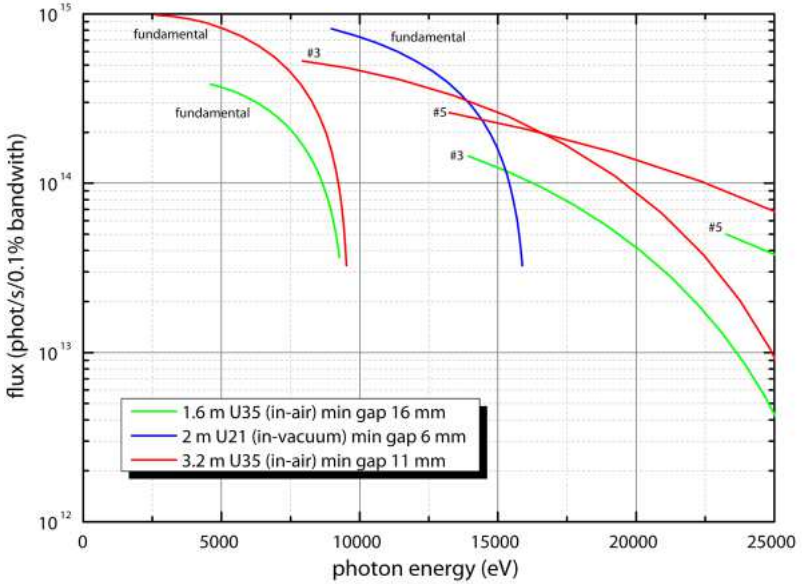
Experiments at or close to L-absorption edges give much bigger signals than those around K-edges



Disadvantage (?) – SAD not MAD



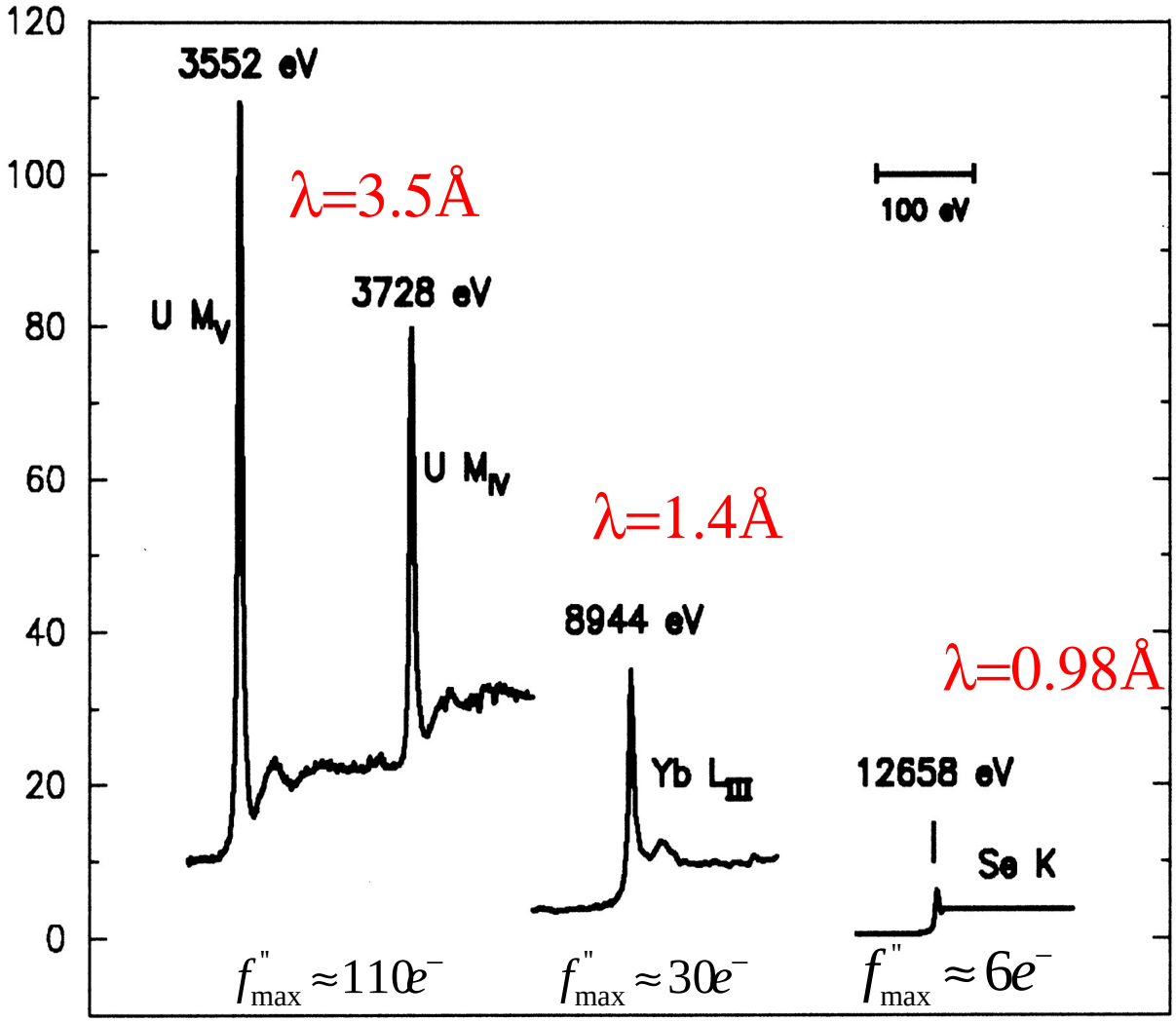
MAD experiments around Xe and I L-absorption edges at ESRF in the near future?



Beam size	10 x 10 μm²		5 x 5μm2		
	Collimated Current	Collimated Current	Collimated + New Udulators	+ UHV-He	+ μfocus
6.0 keV	3.2×10^{7}	3.2×10^{7}	6.4×10^{7}	1.9×10^{8}	1.5×10^{10}
8.0 keV	6.4×10^{7}	6.4×10^{7}	1.3×10^{8}	2.1×10^{8}	2.0×10^{10}
10.0 keV	6.4×10^{8}	6.4×10^{8}	6.4×10^{8}	8.0×10^{8}	4.8×10^{10}
12.0 keV	6.4×10^{8}	6.4×10^{8}	6.4×10^{8}	7.3×10^{8}	4.3×10^{10}
14.0 keV	4.7×10^{8}	4.7×10^{8}	7.0×10^{8}	7.6×10^{8}	5.2×10^{10}
16.0 keV	6.4×10^{7}	6.4×10^{7}	1.3×10^{8}	1.4×10^{8}	9.4×10^{9}
20.0 keV	4.7×10^{7}	4.7×10^{7}	1.4×10^{8}	1.5×10^{8}	9.9×10^{9}

Also other BLs at other SR sources

Signals can be even bigger at VERY long wavelength



Liu, Ogata, & Hendrickson (2001). Proc. Natl. Acad. Sci. USA, Vol. 98, 10648-10653

What could we do at VERY long wavelengths?

$$< \frac{\Delta F}{F} > \approx \frac{1}{\sqrt{2}} \frac{\sqrt{N_A}\, 2\, f''}{<|F_T|>}$$

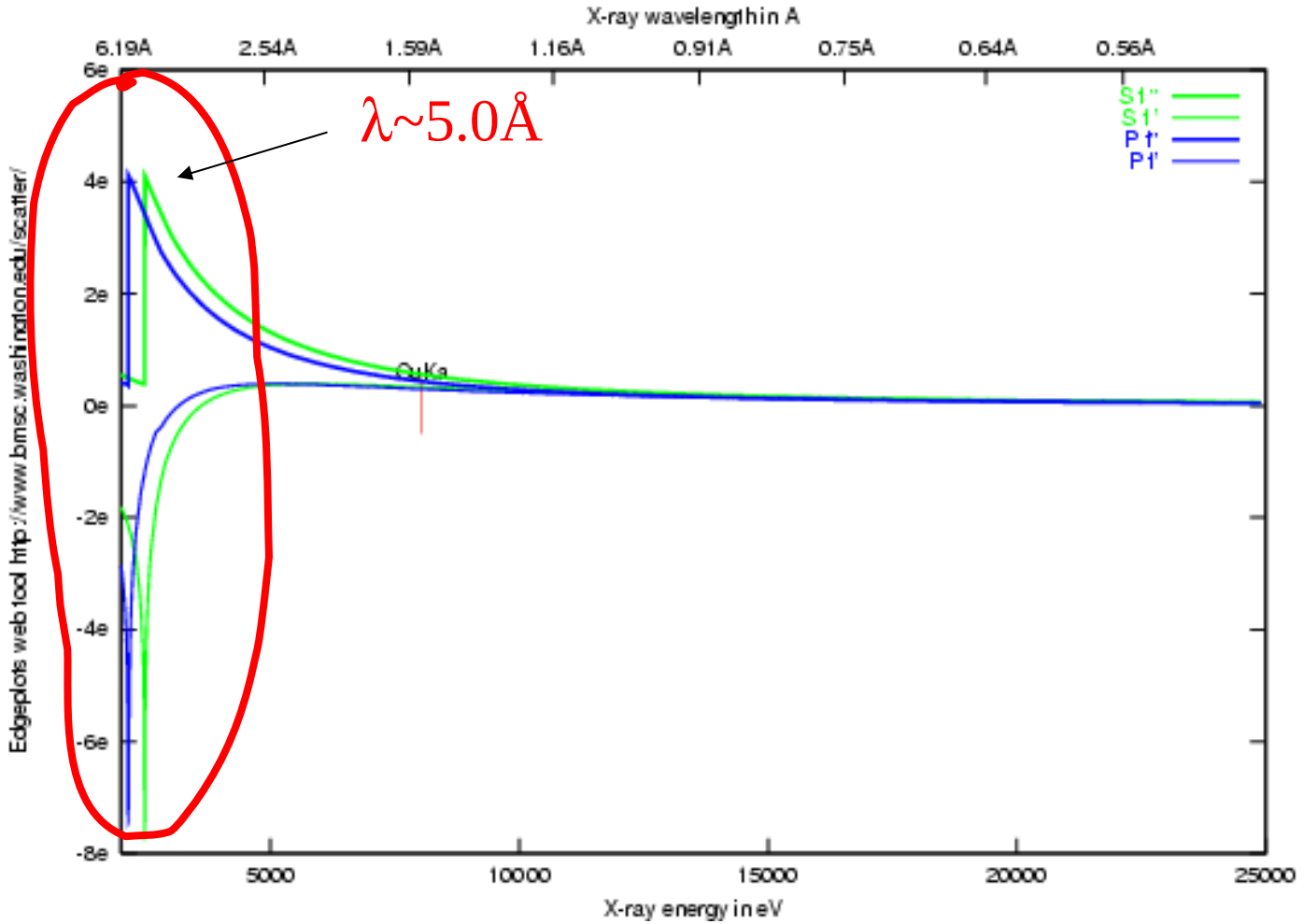
For Proteins: $<|F_T|> \sim 6.70 \bullet [\#\ \text{Atoms}]^{1/2}$

$\sim (3.14 \bullet M_r)^{1/2}$

At M_V edge ($f'' \sim 110e^-$) $<\Delta F/F> \sim 6.9\%$ for one U atom in 1.6MDa
(i.e. one fully occupied U atom could 'phase' the ribosome!)

Liu, Ogata, & Hendrickson (2001). Proc. Natl. Acad. Sci. USA, Vol. 98, 10648-10653

Accessing very, very long l would allow MAD Experiments at S & P K-absorption edges



This would provide enough signal to phase the vast majority of native macromolecular crystal structures

Phasing with Sulphur atoms at K-absorption edge: Possibilities

Genome	% S-a.a[#]	<ΔF/F> (f'' = 4e$^-$) (%)
H. sapiens	4.4	6.7
A. thaliana	4.3	6.7
C. elegans	4.7	6.9
D. melanogaster	4.2	6.2
E. coli K12	4.0	6.0
S. cerevisiae	3.4	5.8
C. pneumoniae	3.5	5.8

[#]from http://www.ebi.ac.uk/proteome/

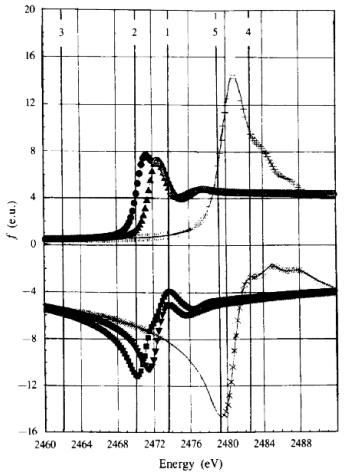


Figure 1
Anomalous dispersion of sulfur near the *K*-absorption edge. ▽, △ are f', f'' of methionine (cysteine); ■, ● are f', f'' of cystine; × and + are f', f'' of sulfate ions. The scattering factors are given in electron units (e.u. = scattering length per electron). The numbers inside the figure denote the indices of the chosen energies E_1–E_5.

White lines will almost double signal! S. Stuhrmann *et al., J. Synchrotron Rad.* (1997). 4, 298-310.

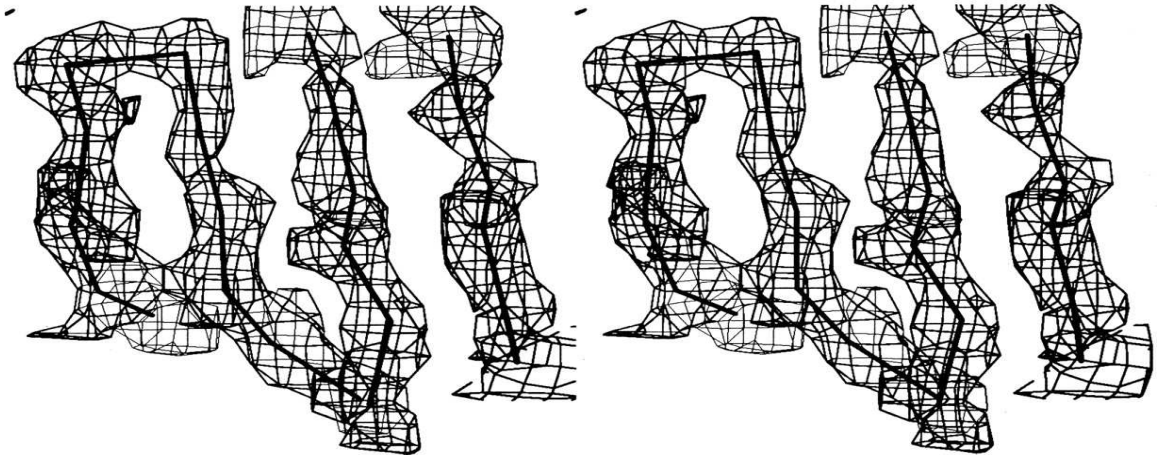
Data Collection at very long λ - I

It is feasible:

Lehmann, Müller & Stuhrmann (1993). *Acta Cryst*. **D49**, 308-310

Kahn, et al. & Stuhrmann (2000). *J. Synchrotron Rad.* **7**, 131-138.

Liu, Ogata, & Hendrickson (2001). *Proc. Natl. Acad. Sci*. USA, **98**, 10648-10653.



MAD-phased electron density from uranyl derivative of elastase. Data collected around U M_{IV} edge

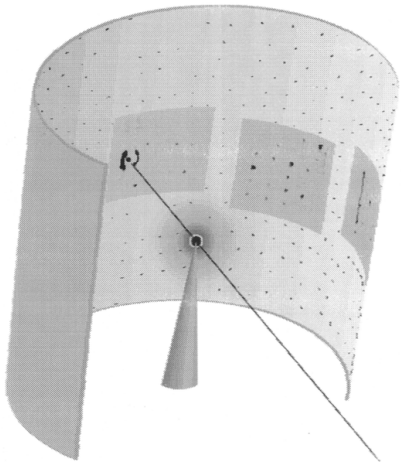
BUT……

Data Collection at very long λ - II

It is experimentally <u>very</u> difficult:

- Absorption (both from air & sample) a real problem. Would need
  - small samples – to reduce absorption
  - evacuated/He-filled 'experiment' – reduced air scatter, attenuation of diffracted X-rays
  - specialized beam-lines (no absorbing material between source & sample) – improve intensity at sample position

- Diffraction angles very high
  - $2\theta \sim 112.8^{o}$ at $d_{min} = 3.0Å$ & $\lambda = 5.0Å$
  - specially shaped detectors



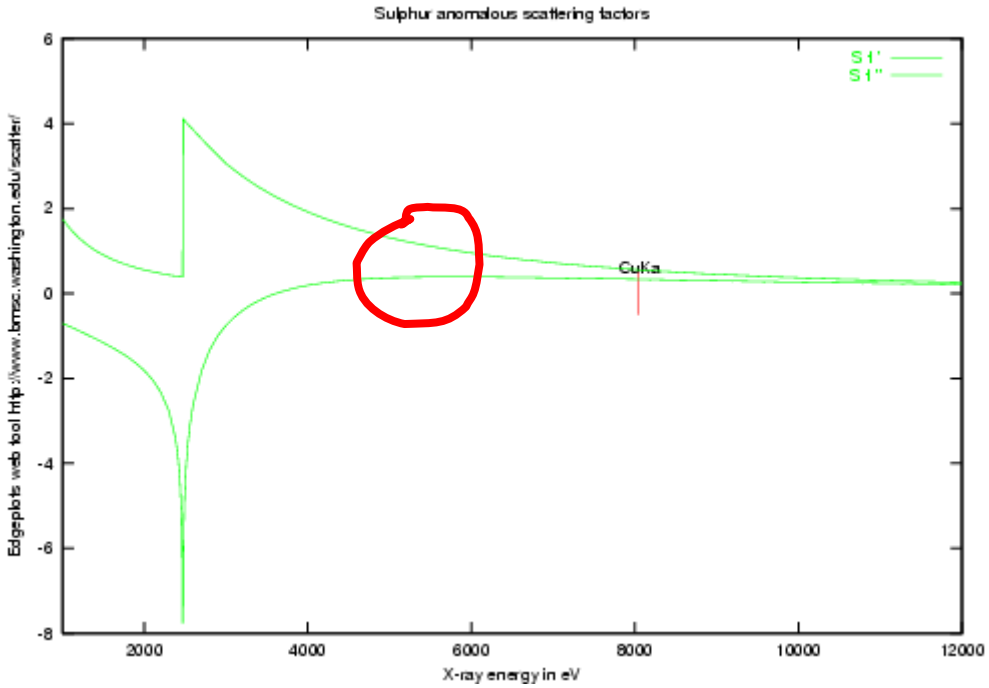
Could one do these experiments on a routine basis?

Yes!!

- S is naturally present in significant amounts in nearly all proteins.
- No need for heavy atom derivatives (including SeMet)
    - crystal quality not compromised
    - no non-isomorphism
    - less time spent in biochemistry laboratory
    - could truly be a magic bullet

SAD Phasing with S at more reasonable wavelength

Problems are smaller but so is signal….



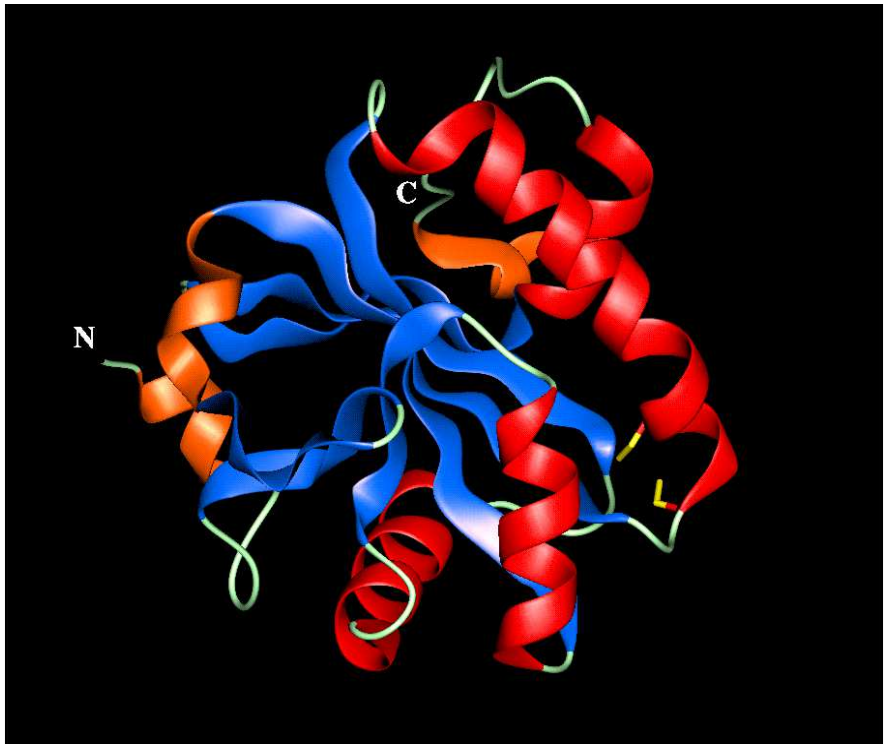
Principle already demonstrated by Hendrickson & Teeter (1981), Wang (1985), Dauter (1999)

Signal from S atoms at λ = 2.07 Å (~6.0 keV)

Genome	%S-a.a[#]	$\langle \Delta F/F \rangle_{6.0\ keV}$	$\langle \Delta F/F \rangle_{5.0\ keV}$
H. Sapiens	4.4	~1.5%	~2.0%
A. Thaliana	4.3	~1.5%	~2.0%
C. Elegans	4.7	~1.5%	~2.1%
D. Melanogaster	4.2	~1.5%	~2.0%
E. coli K12	4.0	~1.4%	~1.9%
S. cerevisiae	3.4	~1.3%	~1.8%
C. pneumoniae	3.4	~1.3%	~1.8%

[#]from http://www.ebi.ac.uk/proteome/

S-SAD Works

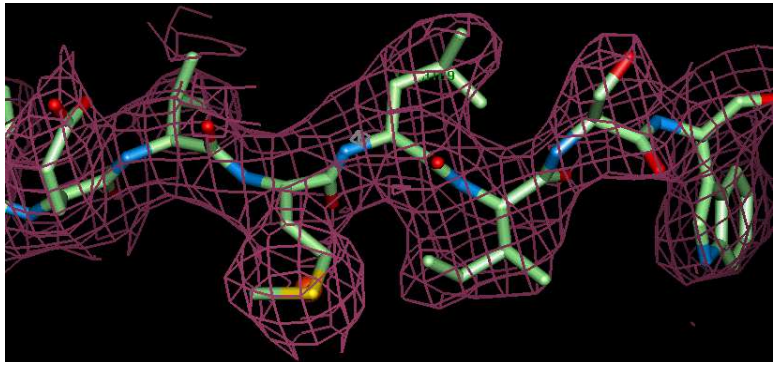


Mol weight	18KDa x 2
Wavelength	1.77Å
Space Group	$P4_32_12$
Oscillation range (°)	1.0
No. of frames	560
No. of sulphurs	8 x 2
Resolution range	40 - 2.7Å
Redundancy	30.0
I/σ(I)	52.0
$I/\sigma(I)_{high}$	9.5
No. S found	14
$\langle \Delta F/F \rangle$	~1.1%*

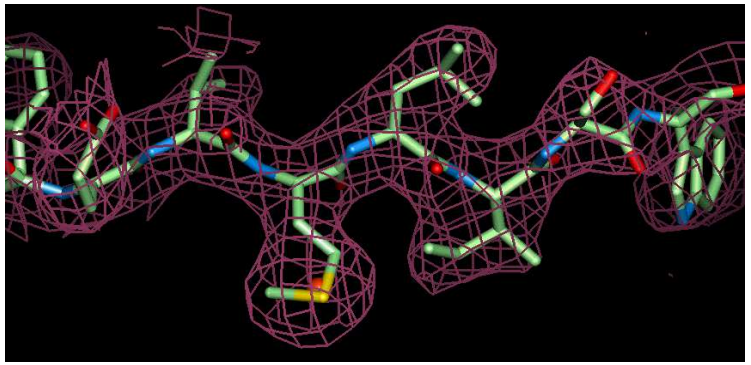
*Tryparedoxin-II, 14S/300 ordered residues $\langle \Delta F_{hkl}/F_{hkl} \rangle$ ~1.1%
% S-containing residues - 4.7%

S-SAD-phased electron density

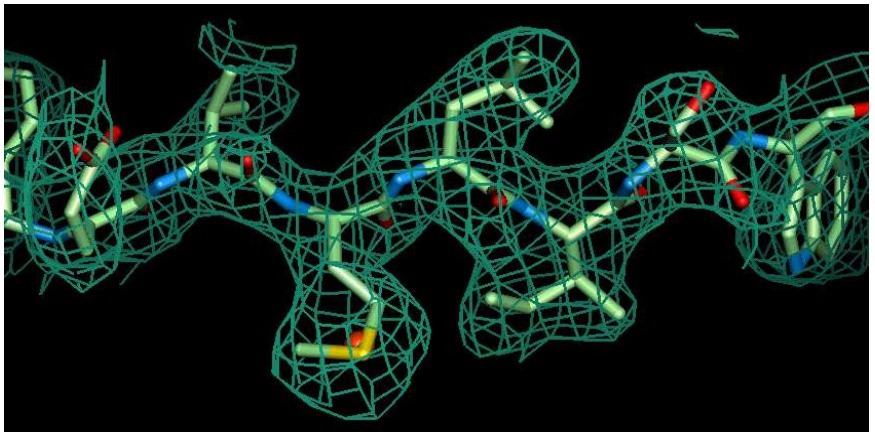
Sites + SHARP & Solomon 2.7Å



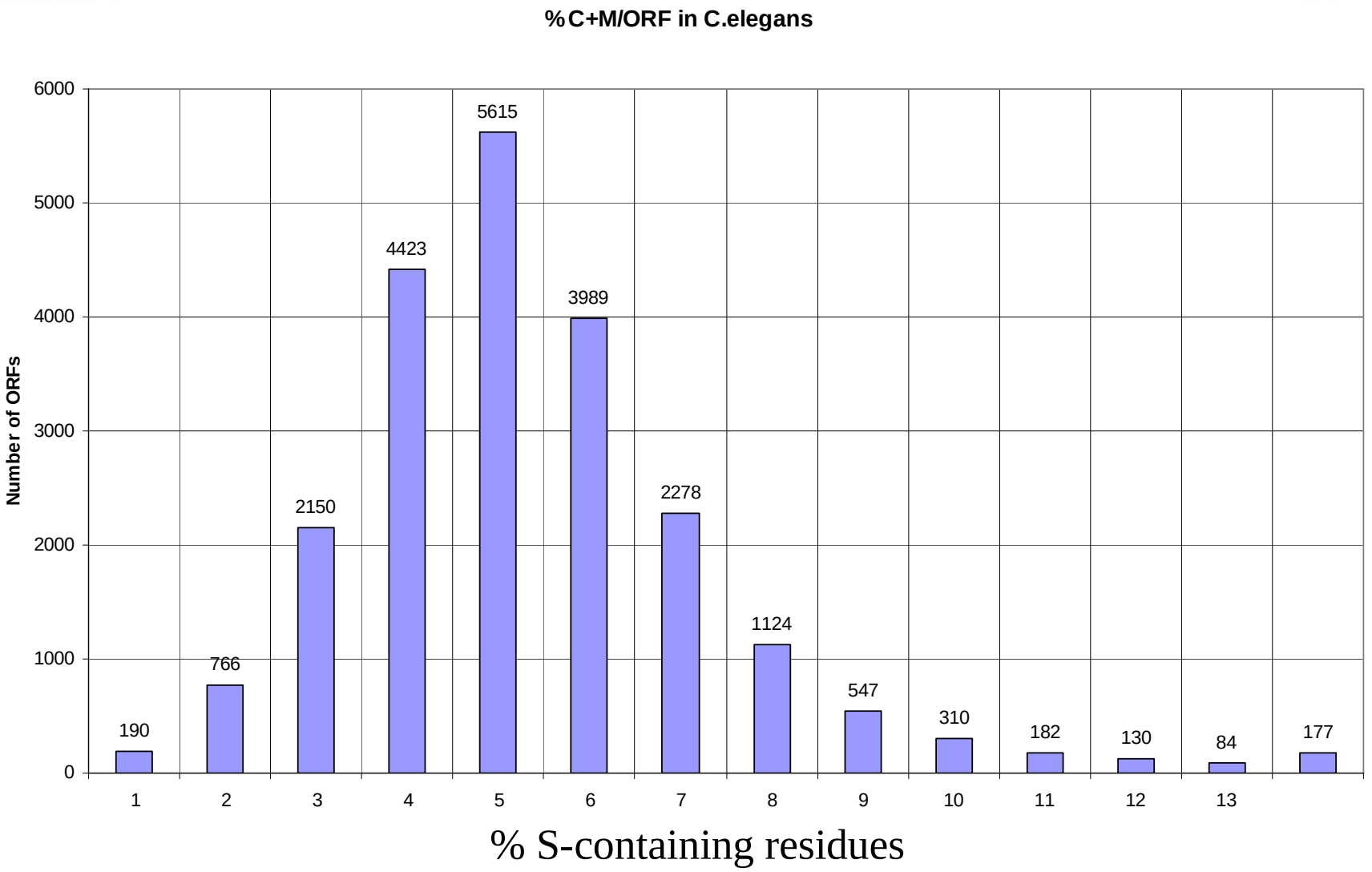
Sites + DM & NCS; 2.7Å



DM & NCS averaging 2.35Å



Micossi *et al*., (2002). *Acta Cryst*. **D58**, 21-28


%C+M/ORF in C.elegans
Number of ORFs
6000
5000
4000
3000
2000
1000
0
190
766
2150
4423
5615
3989
2278
1124
547
310
182
130
84
177
1
2
3
4
5
6
7
8
9
10
11
12
13
% S-containing residues

Acta Cryst. (2006). D**62**, 1475-1483 [doi:10.1107/S0907444906038534]

What can be done with a good crystal and an accurate beamline?

J. Wang, M. Dauter and Z. Dauter

Abstract: X-ray single-wavelength anomalous diffraction (SAD) data from a crystal of proteinase K were collected using synchrotron radiation of 0.98 Å wavelength at SER-CAT 22-ID beamline, Advanced Photon Source, Argonne National Laboratory. At this wavelength, the expected Bijvoet ratio resulting from the presence of one calcium, one chloride and ten S atoms in the 279-residue protein is extremely small at ~0.46%. The direct-methods program *SHELXD* located 11 anomalous sites using data truncated to 2 Å resolution. *SHELXE* was used to produce an easily interpretable electron-density map. This study shows that an accurate beamline and a good-quality crystal provide the possibility of successfully using a very weak anomalous signal of sulfur measured at a short wavelength for phasing a protein structure, even if a small degree of radiation damage is present.

PDB reference: 2id8

Keywords: anomalous scattering; SAD phasing.

Acta Cryst. (2003). D**59**, 1020-1027 [doi:10.1107/S0907444903007467]

Phasing on anomalous signal of sulfurs: what is the limit?

U. A. Ramagopal, M. Dauter and Z. Dauter

Abstract: Recent years have witnessed significant advancements in X-ray data-acquisition techniques and phasing algorithms, which have made possible the successful use of a very small anomalous diffraction signal for the solution of crystal structures of macromolecules. Two crystal structures, a 44 kDa glucose isomerase containing nine sulfurs and a 33 kDa xylanase containing five sulfurs, have been solved from single-wavelength anomalous data using widely available methods and programs. These two enzymes contain less sulfur than most proteins in the bacterial or eukaryotic proteomes, providing a Bijvoet ratio of about 0.6%. For glucose isomerase the automatically interpretable electron-density maps could be obtained at high as well as low resolution. The S-SAD approach relies on the anomalous signal of sulfur naturally occurring in proteins and alleviates all need for sample derivatization. It may therefore be applicable to all protein crystals able to provide accurate diffraction data.

Keywords: anomalous scattering; sulfur; SAD; xylanase; glucose isomerase.

For most systems S-SAD should be successful. Why isn't it?

1. Signals are (sort of) small
2. Radiation damage (High multiplicity to extract signal)?
3. Poor experiment planning and/or execution?

How can we make the use of soft X-rays for phasing routine?

Improve our experiments – reduce noise

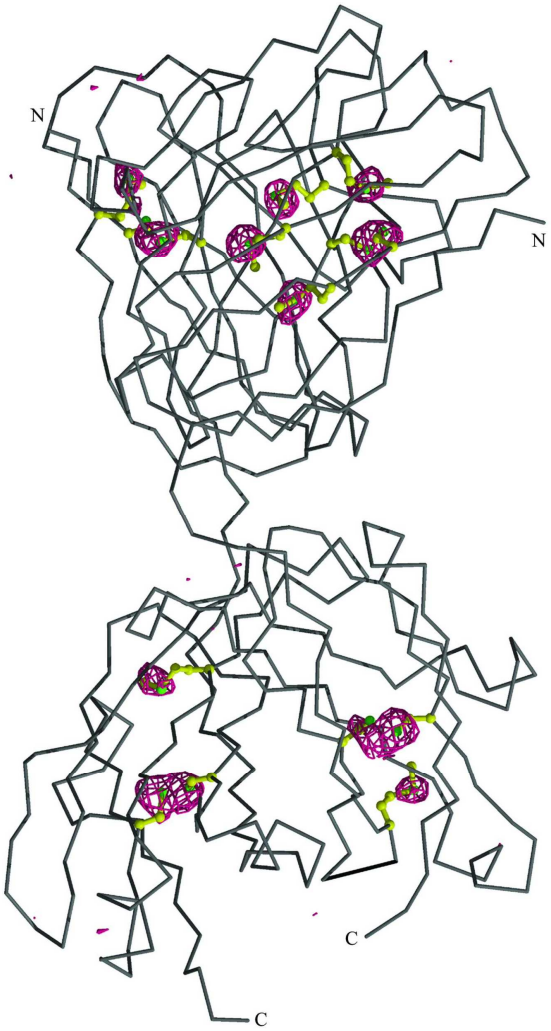
- Reduce air scatter
  - He-filled sample environment? vacuum until end of slit box?
- Eliminate contamination of beam by harmonics from monochromator
  - better use of pushers, use of small mirror before sample
- Routine use of (mini-)kappa
  - Anomalous differences measured on same image, same X-ray dose
  - Collection about more than one axis
- Correcting for/reducing absorption by sample (crystal + surroundings)
  - Analytical absorption corrections?
  - 'loopless' mounting of frozen crystals
  - Scaling long wavelength data against reference data set?
- Improving detector efficiency at wavelengths of interest?
- Better treatment & use of radiation damage
- Use software to help predict strategy, reduce radiation damage, perform multi-position / multi crystal experiments, merge data from most isomorphous crystals

Increase the signal

- Longer wavelengths than currently routinely used?

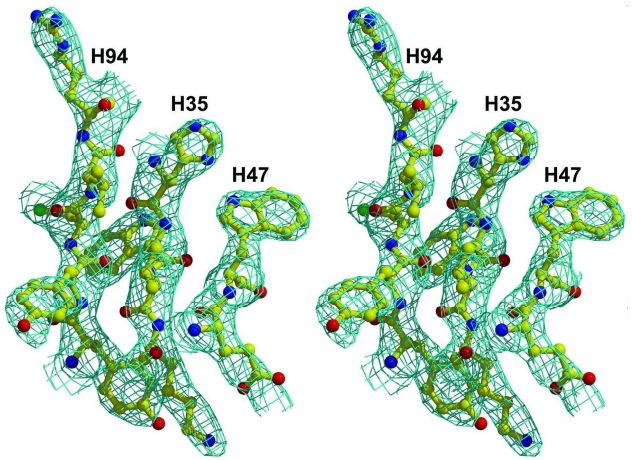
Feedback during experiment

- Improvement in statistics (R_{pim}; CC_{anom}; etc.)
- Automatic substructure determination
- Is structure solved?



<ΔF/F> ~ 0.8%: Anomalous difference fourier (ΔF_{ano}, α – 90º) calculated using phases from a MR solution and anomalous differences measured at λ = 1.54 Å (E = 8.05 keV).

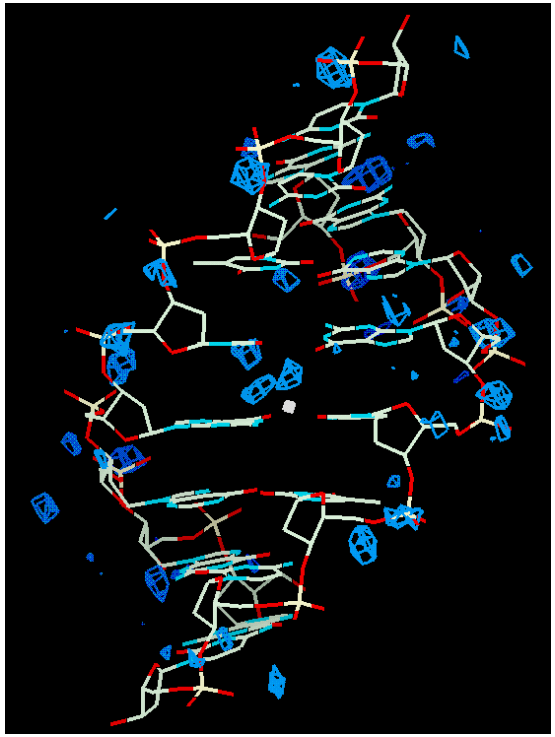
S-SAD phases (d_{min} = 3.5 Å); density modification; NCS averaging



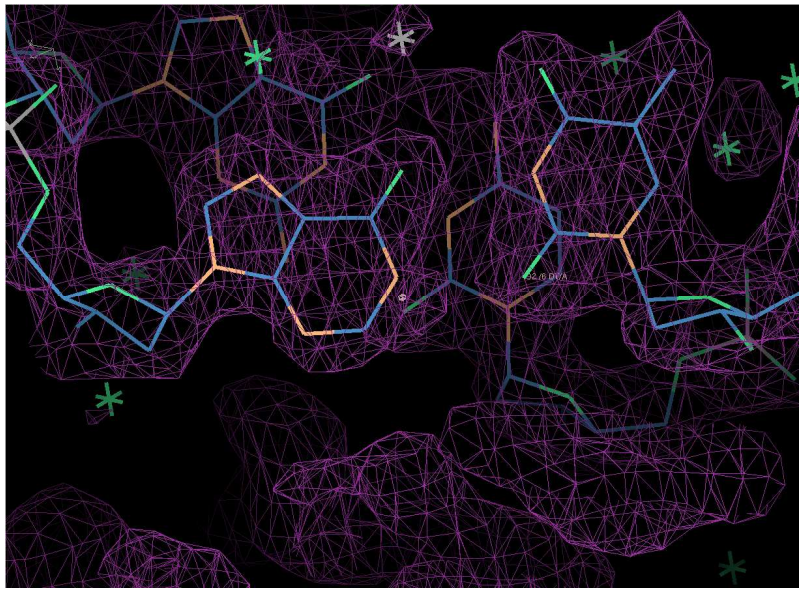
Schuermann & Tanner, (2003). *Acta Cryst*. **D59**, 1731-1736.

MR-P-SAD @ 6 keV

Anomalous difference fourier
(ΔF_{ano}, α – 90º)



d_{min} = 2.3 Å; Iterative improvement of P positions, phase calculation + density modification in SHARP



Unbiased electron density map

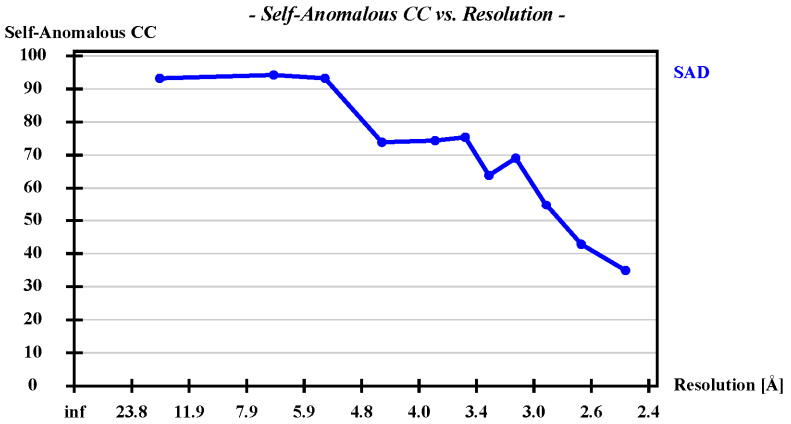
Barbara Durling: d(GGIGCTCC)$_2$; Space Group P6$_5$; d_{min} = 2.3 Å; R_{sym} = 4.9%; multiplicity = 18.4

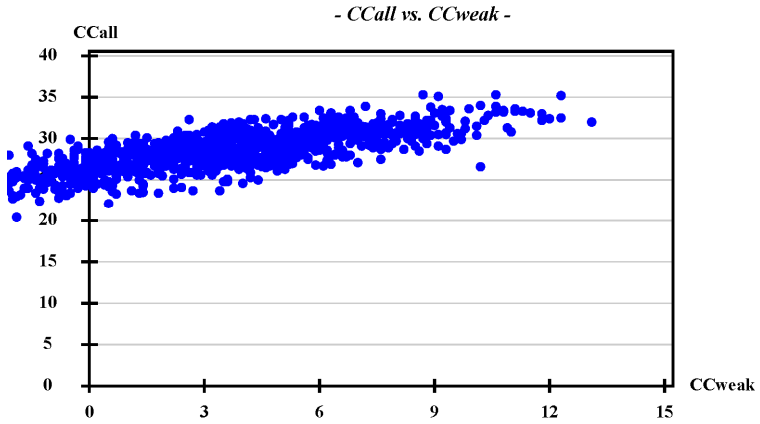
Why did we resort to MR-P-SAD ?

N	1/d^2	Dmin(Å)	Rmrg	Rfull	Rcum	Ranom	Rpim
$$							
1	0.0189	7.27	0.031	0.031	0.031	0.022	0.010
2	0.0378	5.14	0.032	0.032	0.032	0.022	0.011
3	0.0567	4.20	0.031	0.031	0.032	0.017	0.010
4	0.0756	3.64	0.034	0.034	0.033	0.014	0.011
5	0.0945	3.25	0.039	0.039	0.034	0.019	0.013
6	0.1134	2.97	0.047	0.047	0.036	0.019	0.015
7	0.1323	2.75	0.068	0.068	0.039	0.025	0.023
8	0.1512	2.57	0.112	0.112	0.043	0.031	0.038
9	0.1701	2.42	0.136	0.136	0.046	0.041	0.048
10	0.1890	2.30	0.167	0.167	0.049	0.049	0.062
$$							

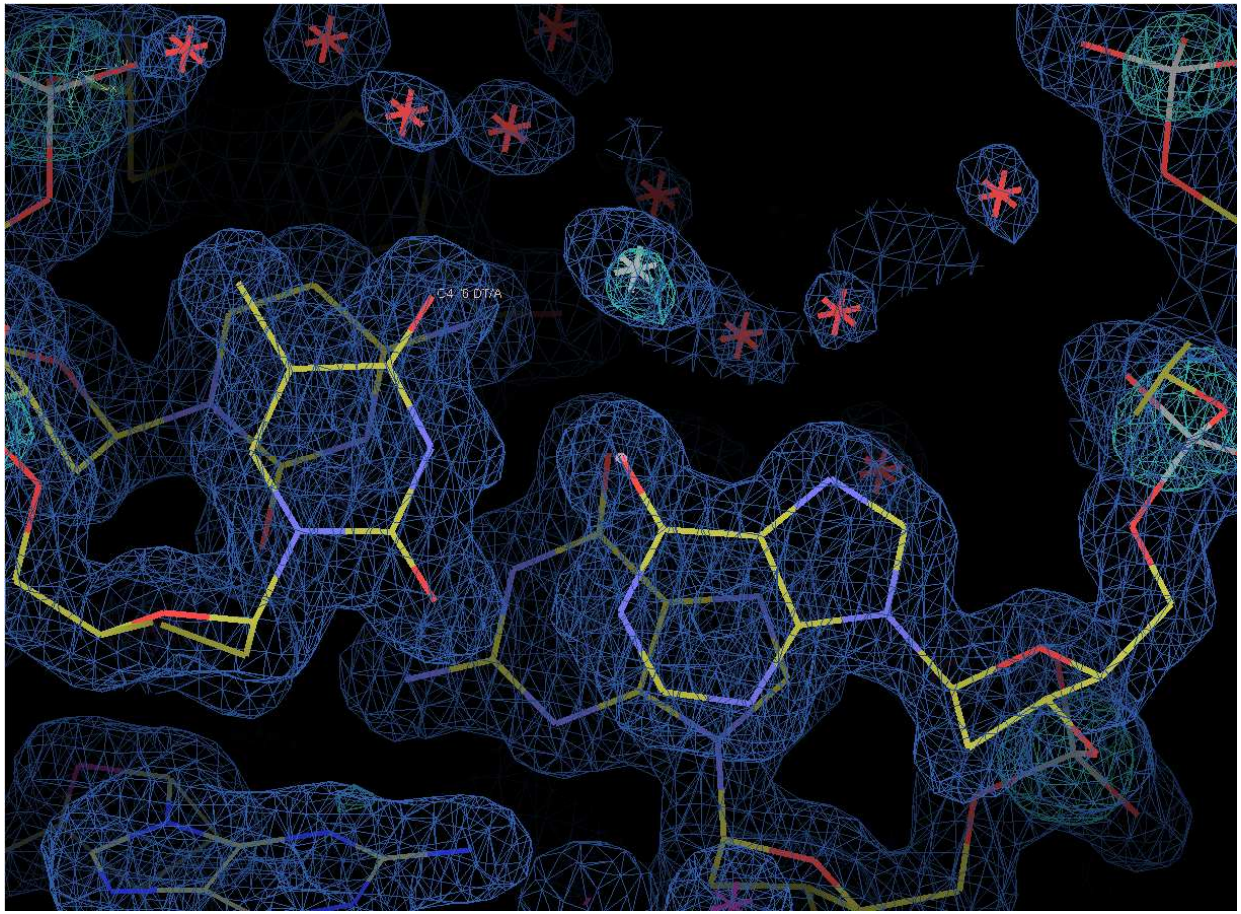
$$R_{anom} / R_{p.i.m} \geq 1.2$$

N	1/resol^2	dmax	CC_anom
$$			
1	0.0189	7.27	0.575
2	0.0378	5.14	0.632
3	0.0567	4.20	0.651
4	0.0756	3.64	0.517
5	0.0945	3.25	0.537
6	0.1134	2.97	0.468
7	0.1323	2.75	0.452
8	0.1512	2.57	0.488
9	0.1701	2.42	0.135
10	0.1890	2.30	0.142
$$			
Overall			0.507





Identifying ions



Cyan: d(GGIGCTCC)$_2$ anomalous difference fourier (ΔF_{ano}, α – 90º) using data collected at 6 keV. Phases are from refined model.

Moiety bridging I•T base-pair not a water molecule?

Summary

1. Softer energies in MX are not just about S-SAD.
   - S-SAD has enormous potential, but we (still) need to learn how to do it properly
     - Multi-position and/or multi-crystal data collections; cluster analysis to choose best bits?
2. Soft X-rays will allow MAD / optimised SAD experiments on Xe, I, U etc derivative crystals.
   - Bigger anomalous signals for larger systems
3. Soft X-rays provide information that can improve the success of:
   - MR protocols
   - Chain Tracing
   - Proper identification of ions that may be important in biological processes

Thanks for your attention.