Long wavelength data collection – why & how?

Gordon Leonard

ESRF Macromolecular Crystallography Group

The Electron Density Equation

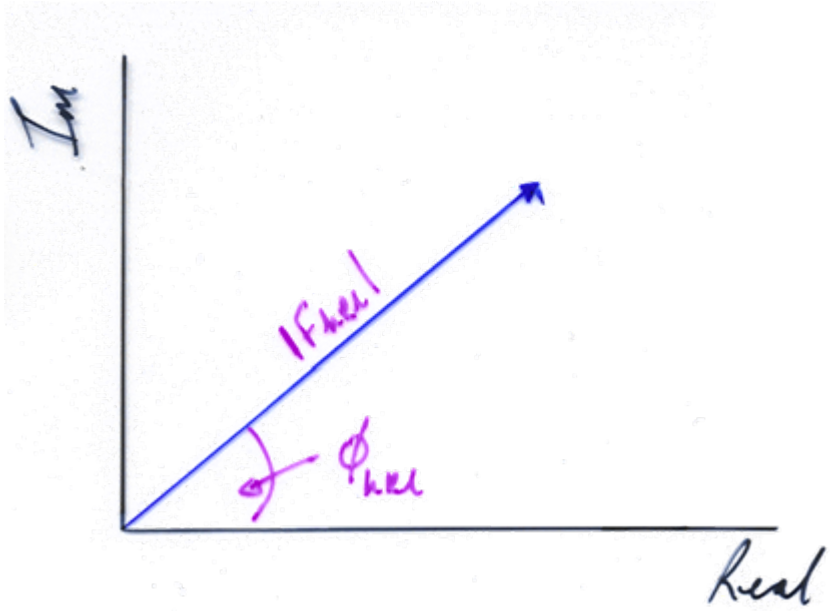
$$\rho_{(x,y,z)} = \frac{1}{V_c} \sum_h \sum_k \sum_l F_{hkl} \exp(-2\pi i(hx + ky + lz))$$

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

$$F_{hkl} = \sum_j |F_{hkl}| \exp -2\pi i(\phi)$$

Amplitude

Phase



From an x-ray diffraction experiment we can ‘measure’ the amplitude but get *no information* about the phase

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

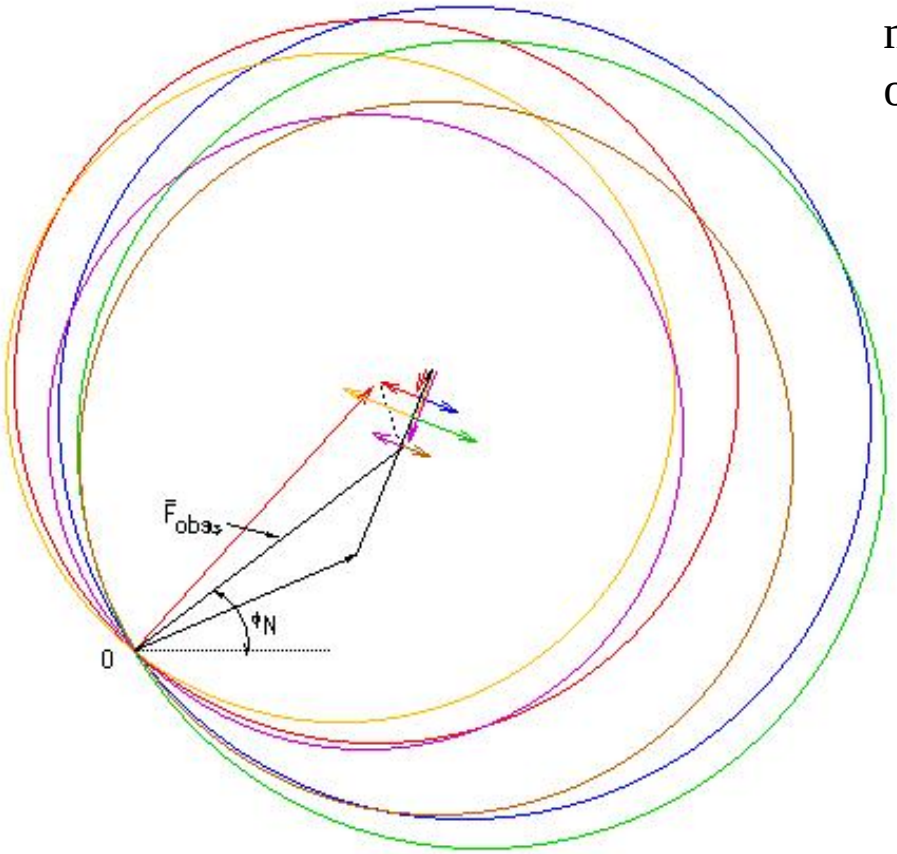
$$f = f_o + f'(\lambda) + if''(\lambda)$$

Wavelength-dependent and vary rapidly close to an absorption edge

When *f*”is large Friedel’s Law is no longer obeyed:

$$F_{hkl} \neq F_{\bar{h}\bar{k}\bar{l}}$$

MAD as Quasi-MIR

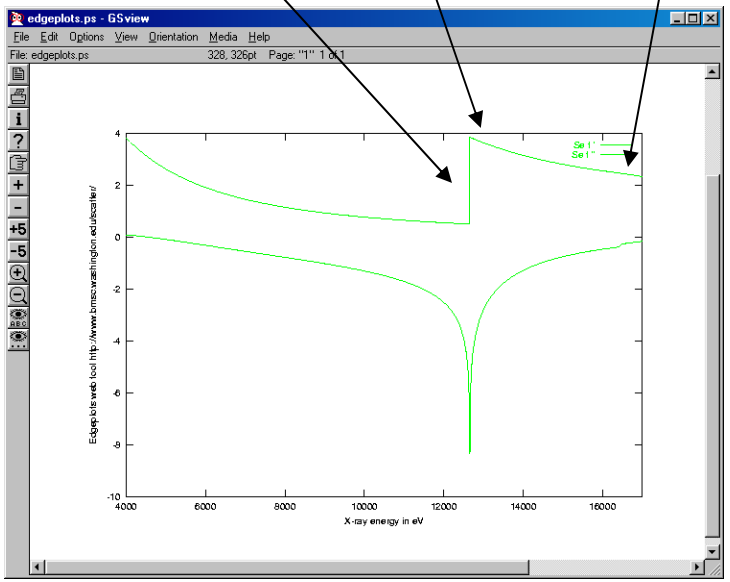


f'at minimum, moderate value of f"

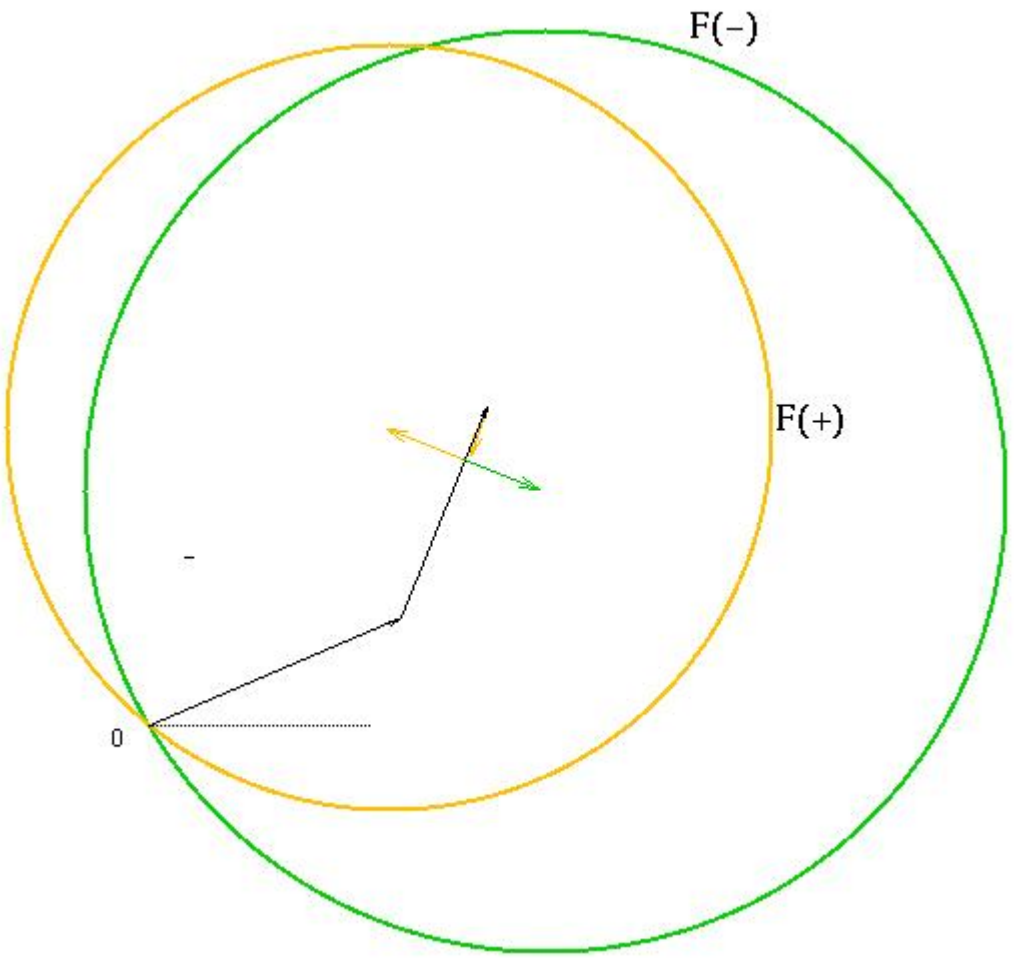
f' moderate, f" moderate

f' moderate,

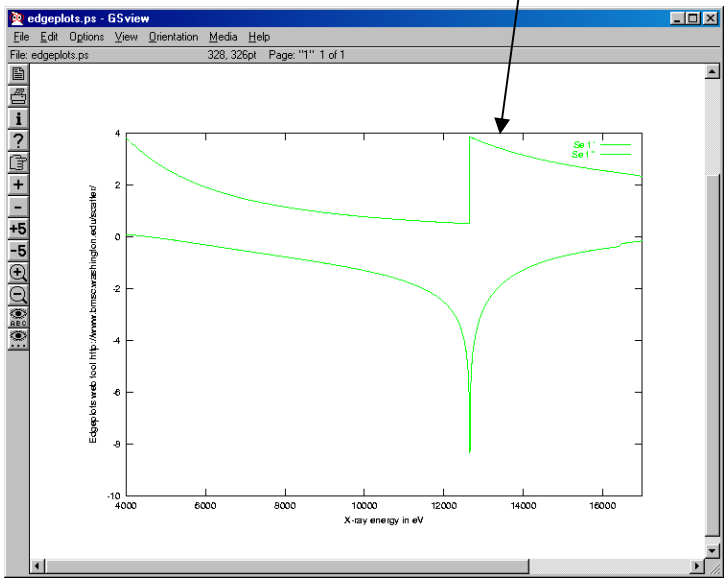
f" large



Deriving phase information using SAD



One can also phase using single wavelength data. Use density modification to break phase ambiguity



SAD Signals

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

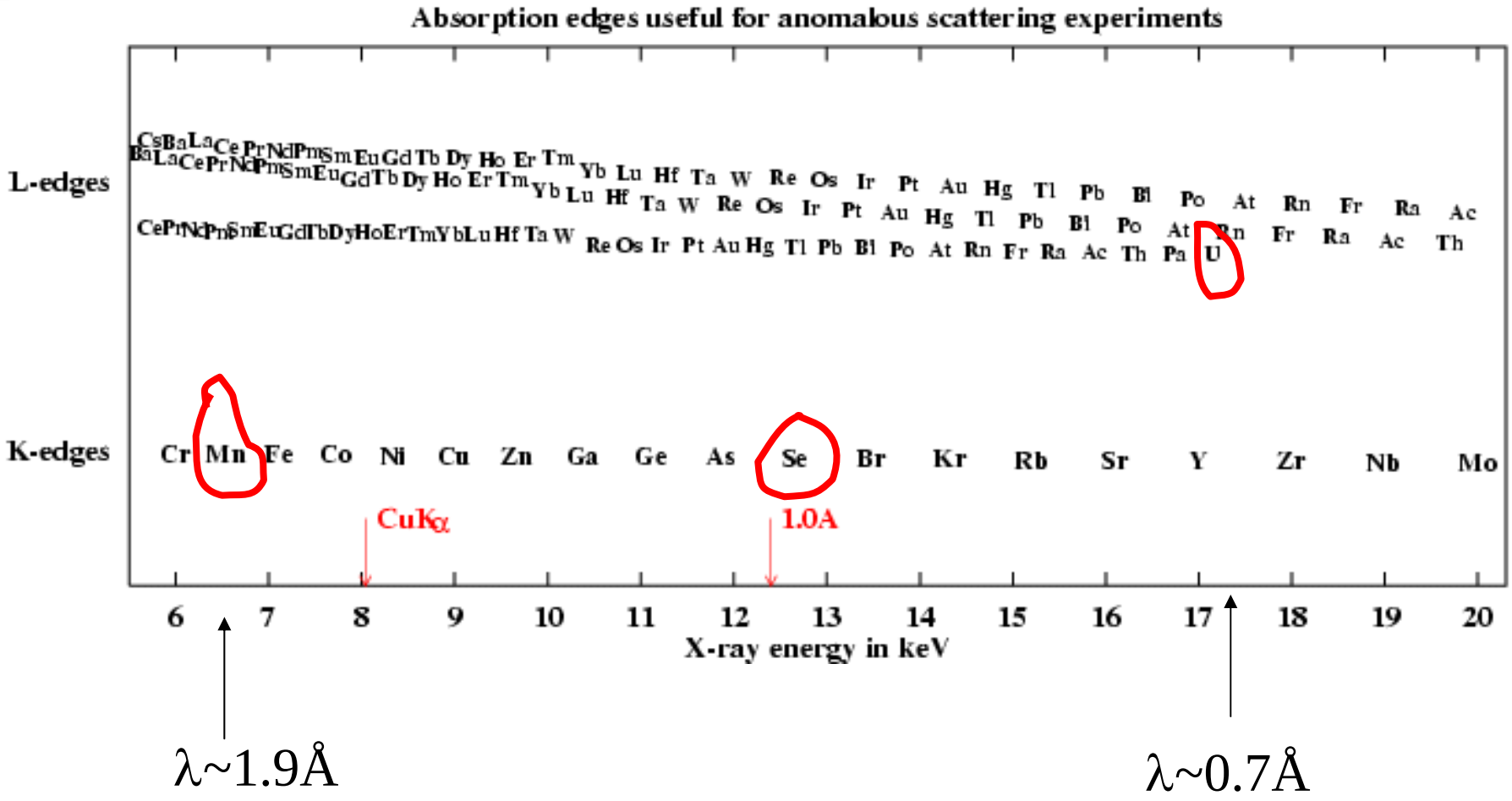
L absorption edges: f" usually has a maximum around **12e⁻**
can rise to **~30e⁻** with 'white line'.

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

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

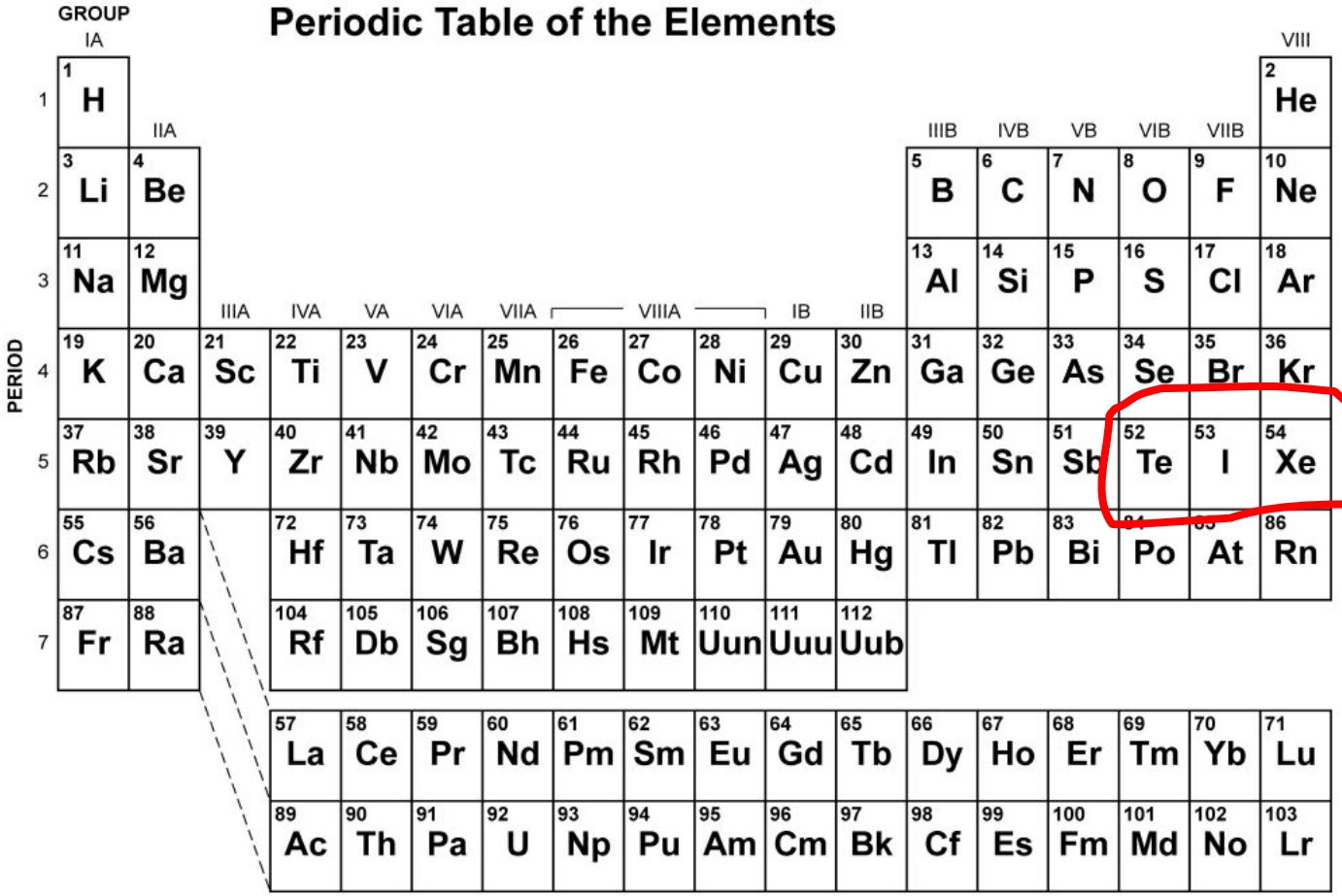
Woolfson & Fan, "Solving Crystal Structures" (1995) Cambridge University Press

Energy of absorption edges commonly used in MAD/SAD phasing procedures



Are there potential benefits in going beyond this range?

Absorption edges accessible at shorter wavelengths



K-edges of Te, I, Xe accessible at ~0.3 Å ~(30 KeV)

But, K-edges give f" of (only) ~4 e^-

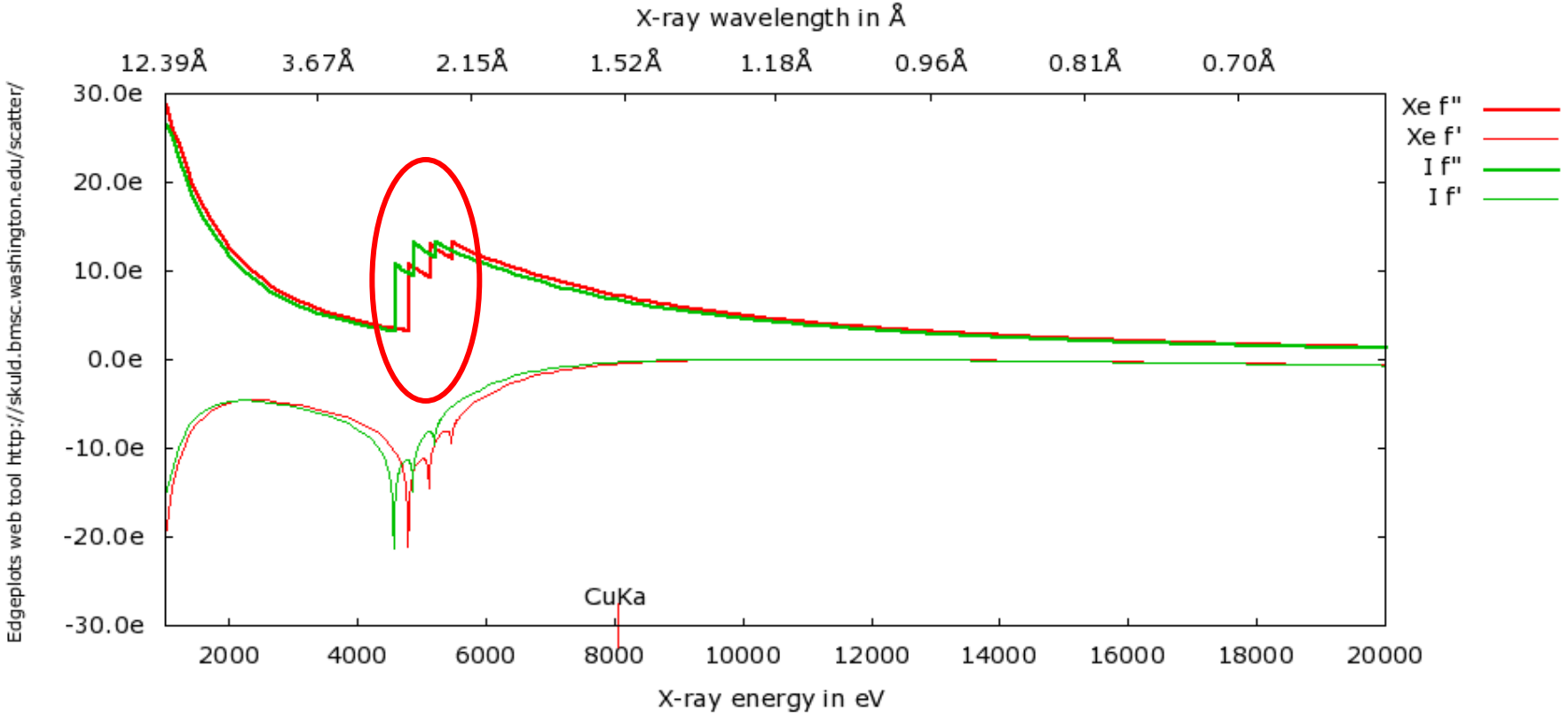
Evans, G. & Bricogne, G. (2002) *Acta Cryst.*, **D58**, 976-991 ; Sauer O., Schmidt A. and Kratky C. (1997) *J. Appl. Crystallogr.*, **30**, 476-486.

f" for I, Xe are (much) higher at longer wavelength

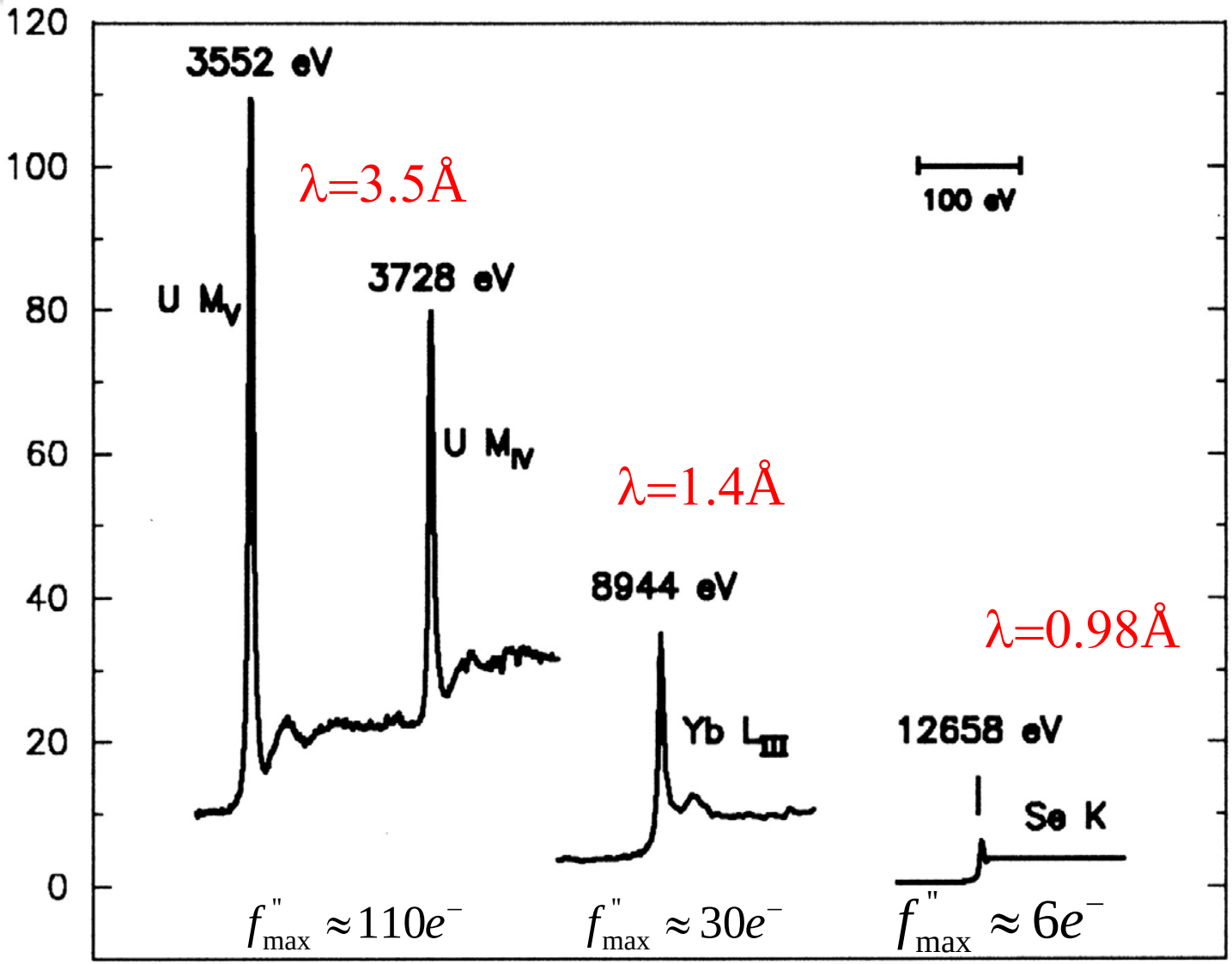
Energy (keV)/Wavelength (Å)	f''_{Xe} (e)	f''_{I} (e)
14.2/0.87	2.8	2.6
12.4/1.00	3.6	3.3
10.0/1.24	5.1	4.8
8.0/1.55	7.4	6.9
7.0/1.78	9.2	8.6
6.5/1.91	10.3	9.7
6.0/2.07	11.6	10.9
5.5/2.25	13.2	12.3

In principle, better to collect data at longer wavelength (even though phasing would be carried out using SAD)

MAD experiments are possible around L_I - edges of I & Xe at $\lambda \sim 2.5$ Å



Signals can be even bigger at VERY long wavelength



What could we do at VERY long wavelengths?

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

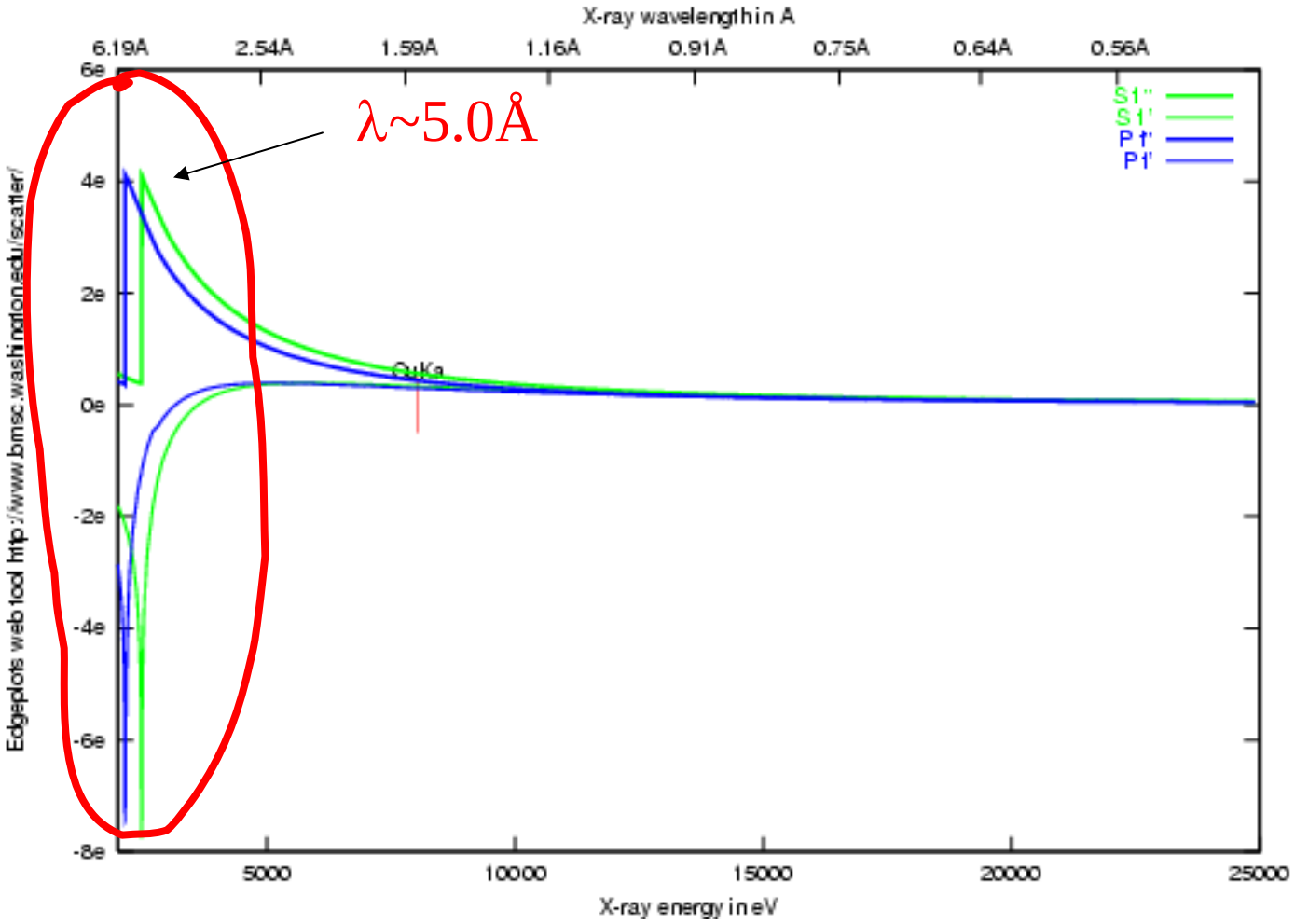
For Proteins: $<|F_T|> \sim 6.70 \bullet [\#\ \text{Atoms}]^{1/2}$

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

At M_v edge (λ = 3.49 Åf " ~ 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



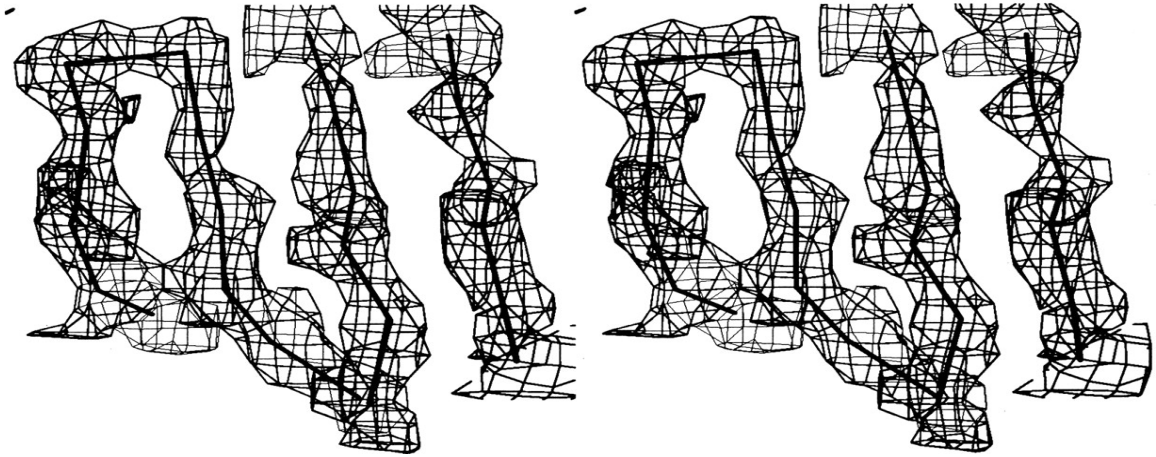
This would provide enough signal to phase the nearly all native macromolecular crystal structures (K-edge of S has a (big) white line)

It is feasible:

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

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

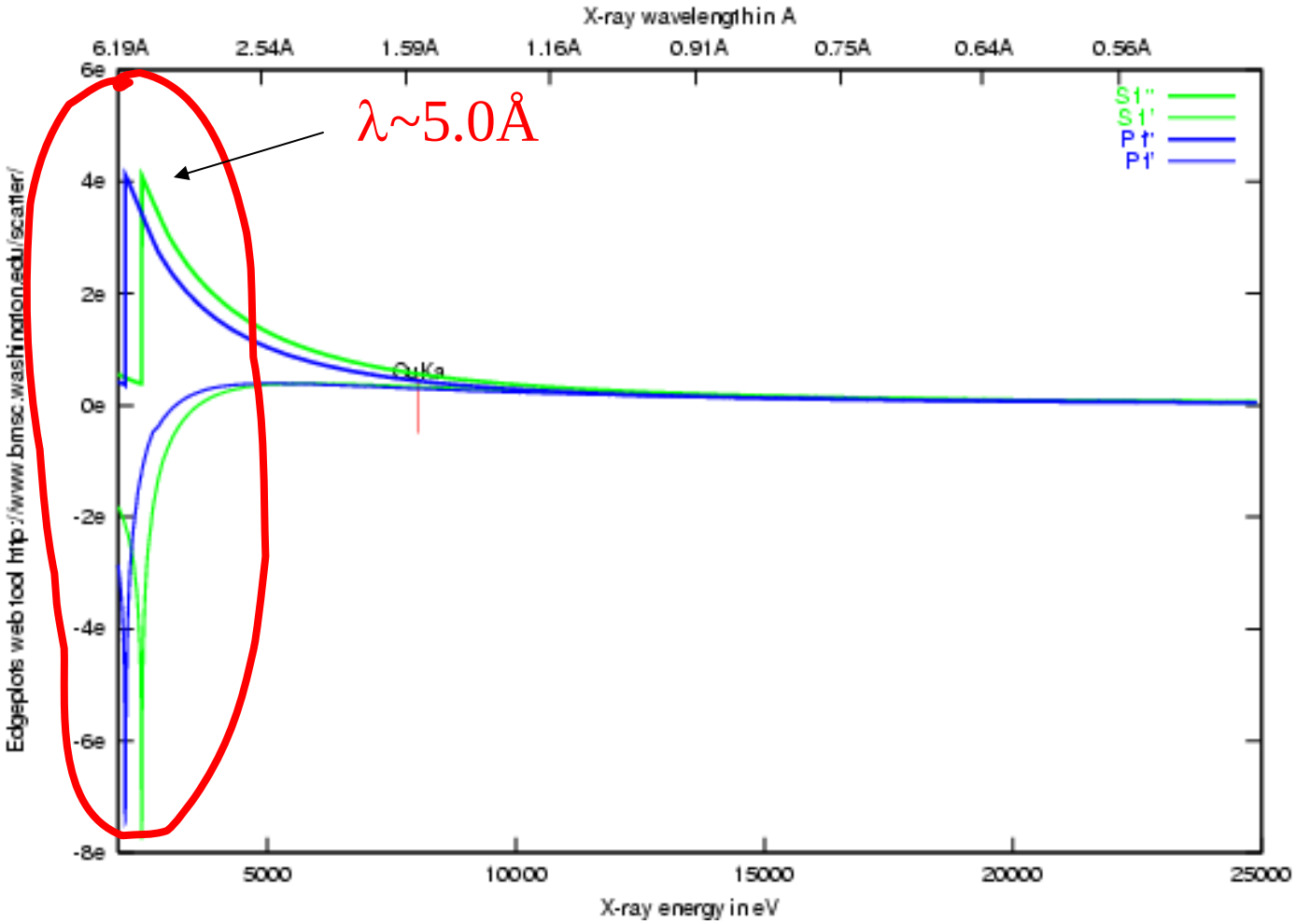
Liu, Ogata, & Hendrickson (2001). *Proc. Natl. Acad. Sci.* USA, **98**, 10648-10653.



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

BUT……

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



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

Phasing with Sulphur atoms at K-absorption edge: Signal Possibilities

Genome	% S-a.a[#]	$\langle \Delta F/F \rangle$ ($f'' = 4e^-$) (%)
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/

In reality signals should be even higher!!

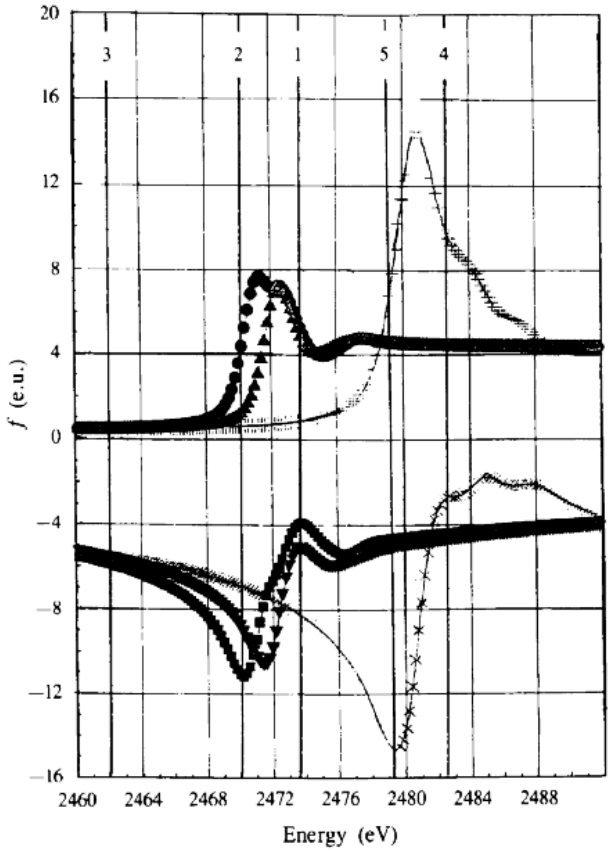


Figure 1
Anomalous dispersion of sulfur near the *K*-absorption edge. ▽, △ are f', f'' of methionine (cysteine); ■, ● are f', f'' of cystine; × and + are f', f'' of sulfate ions. The scattering factors are given in electron units (e.u. = scattering length per electron). The numbers inside the figure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

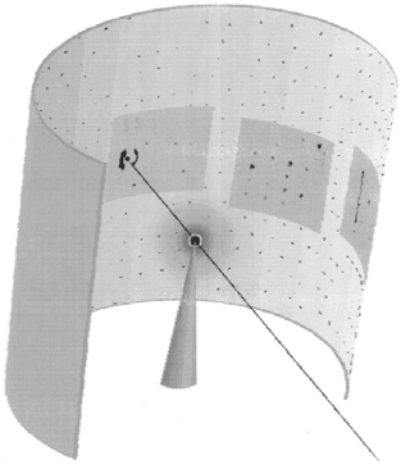
White lines will almost double signal!!

S. Stuhrmann *et al.*, *J. Synchrotron Rad.* (1997). 4, 298-310

Data Collection at very long λ - 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

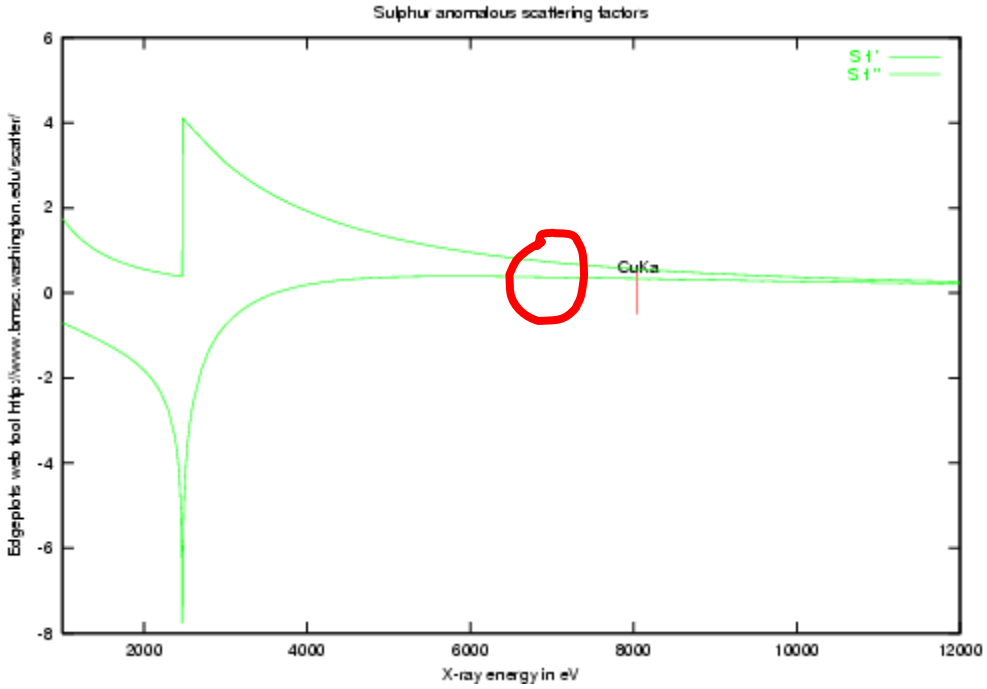
<u>Could probably not do these experiments on a routine basis</u>

Anomalous signal from sulphur/phosphorous in proteins is still significant at λ = 1.8 – 2.2 Å

- Should we try to exploit this in MX?
- 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

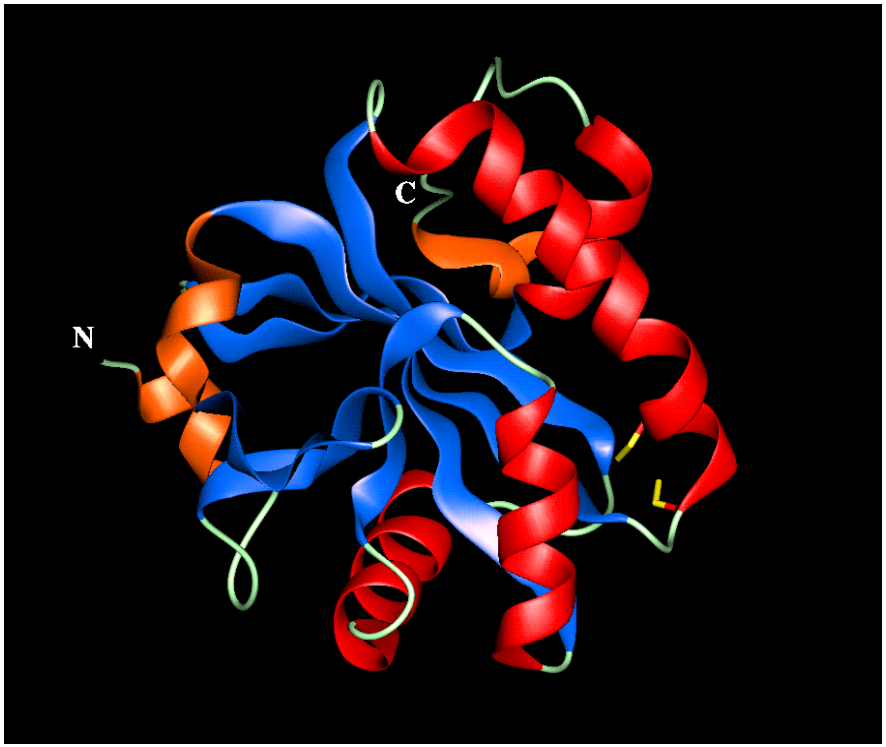
SAD Phasing with S (S-SAD) at more 'reasonable' wavelength

Problems are smaller but so is signal….



At λ = 1.77Å f''=0.72e^{-}; Principle already demonstrated by Hendrickson & Teeter (1981), Wang (1985), Dauter (1999).

S-SAD Works

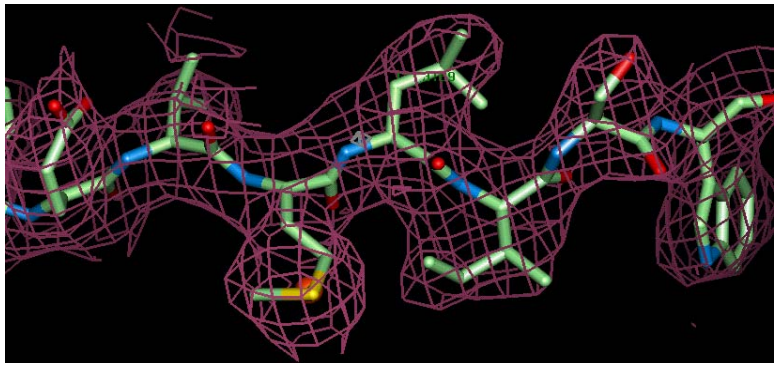


Mol weight	18KDa x 2
Wavelength	1.77Å
Space Group	$P4_32_12$
Oscillation range (°)	1.0
No. of frames	560
No. of sulphurs	8 x 2
Resolution range	40 - 2.7Å
Redundancy	30.0
I/σ(I)	52.0
$I/\sigma(I)_{high}$	9.5
No. S found	14
$\langle \Delta F/F \rangle$	~1.2%*

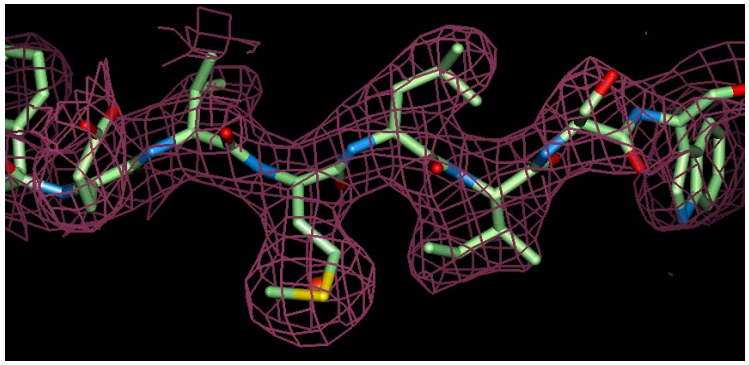
*Tryparedoxin-II, 14S/300 ordered residues $\langle \Delta F_{hkl}/F_{hkl} \rangle$ ~1.2%
% S-containing residues - 4.7%

S-SAD-phased electron density

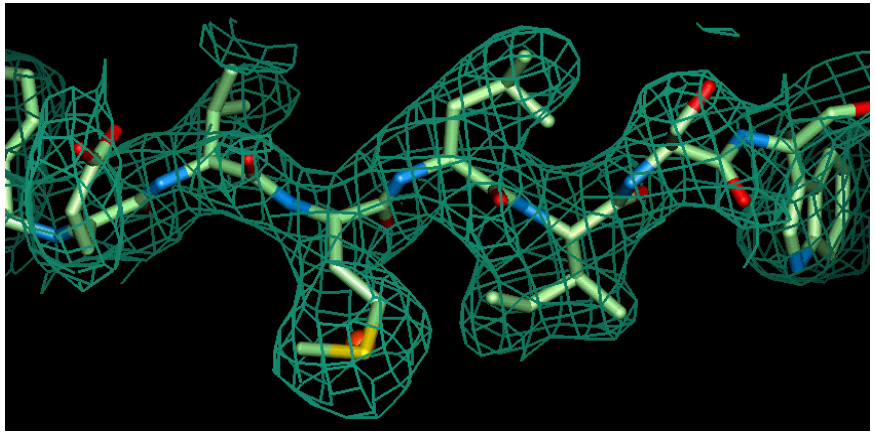
SHARP & Solomon 2.7Å



DM & NCS averaging



DM & NCS averaging 2.35Å



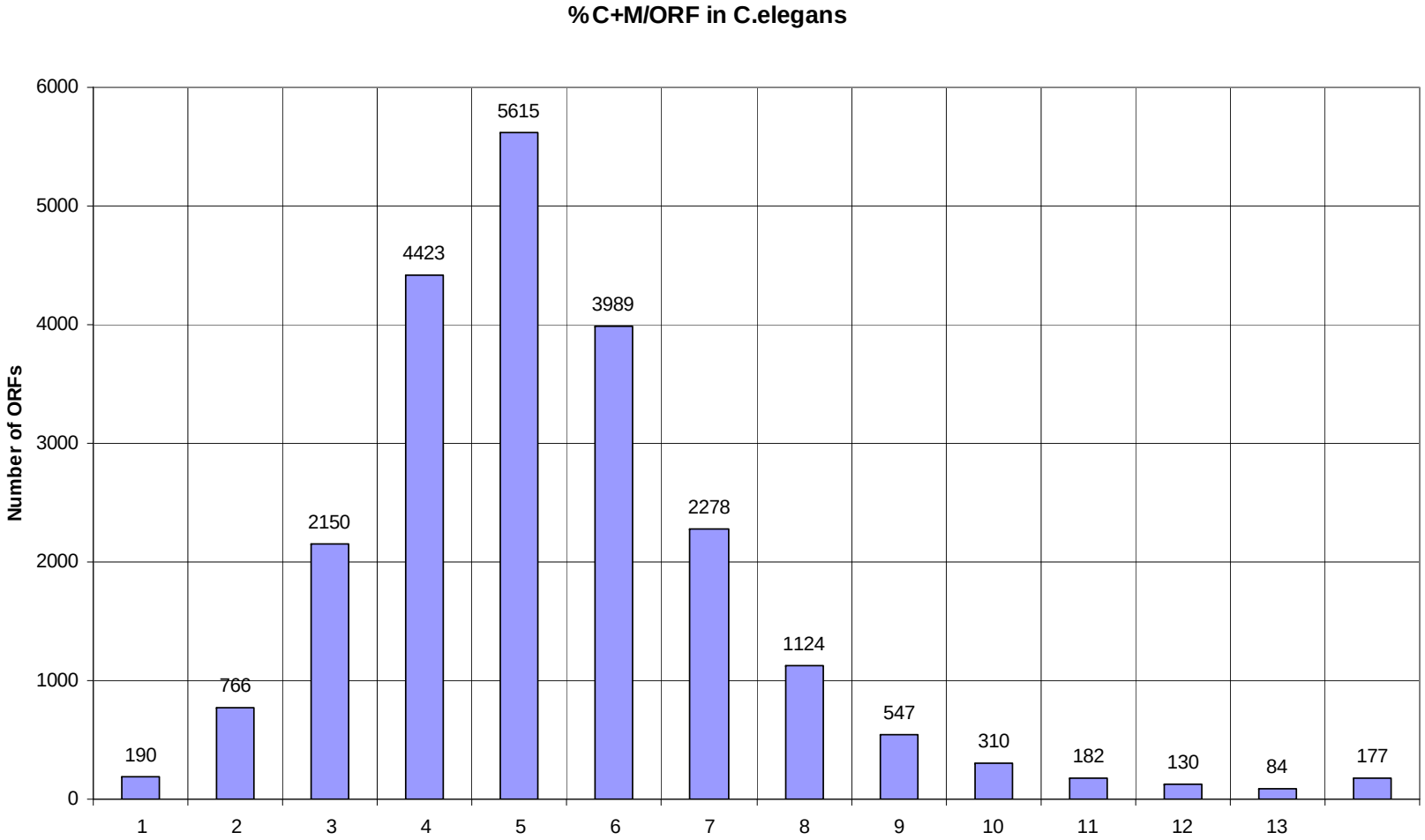
Micossi *et al*., (2002). *Acta Cryst*. **D58**, 21-28

SAD Phasing with Sulfurs at λ = 1.77Å: Signal Possibilities

Genome	Met	Cys	Total
Homo sapiens	2.16	2.23	4.39 %
Arabidopsis thaliana	2.45	1.84	4.29 %
Caenorhabditis elegans	2.61	2.06	4.67 %
Drosophila melanogaster	2.35	1.88	4.23 %
Saccharomyces cerevisiae	2.09	1.30	3.39 %
E.coli K12	2.80	1.17	3.97%
Chlamydia pneumoniae	1.92	1.59	3.51%

Most other Bacteria and Archea have Cys + Met compositions of between 3-3.5%

http://www.ebi.ac.uk/proteome/



N.B. f" for S increases at longer wavelengths (λ = 1.9 Å; f" ~ 1.0 e^-) so we could move to the left of this graph

Anomalous signals – How low can we go?

html | pdf | buy

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.

html | pdf | cited in | buy

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

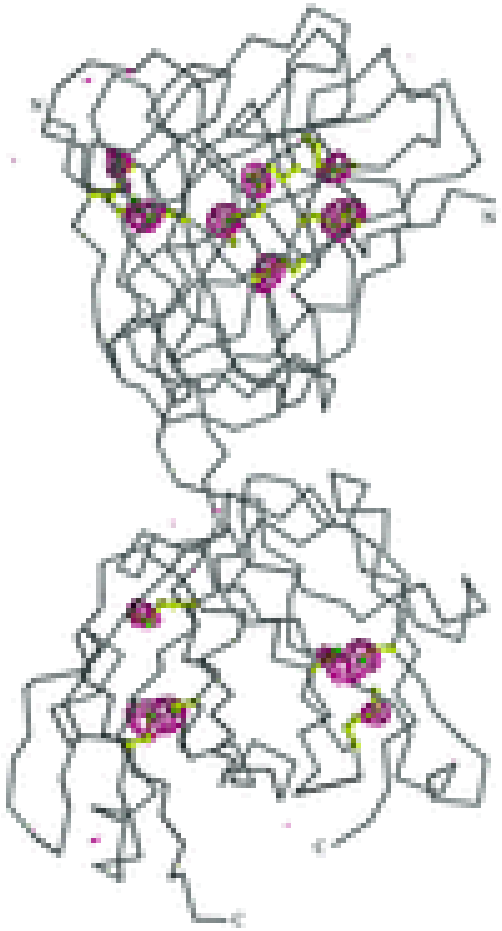
Phasing on anomalous signal of sulfurs: what is the limit?

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 3.6%. 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.

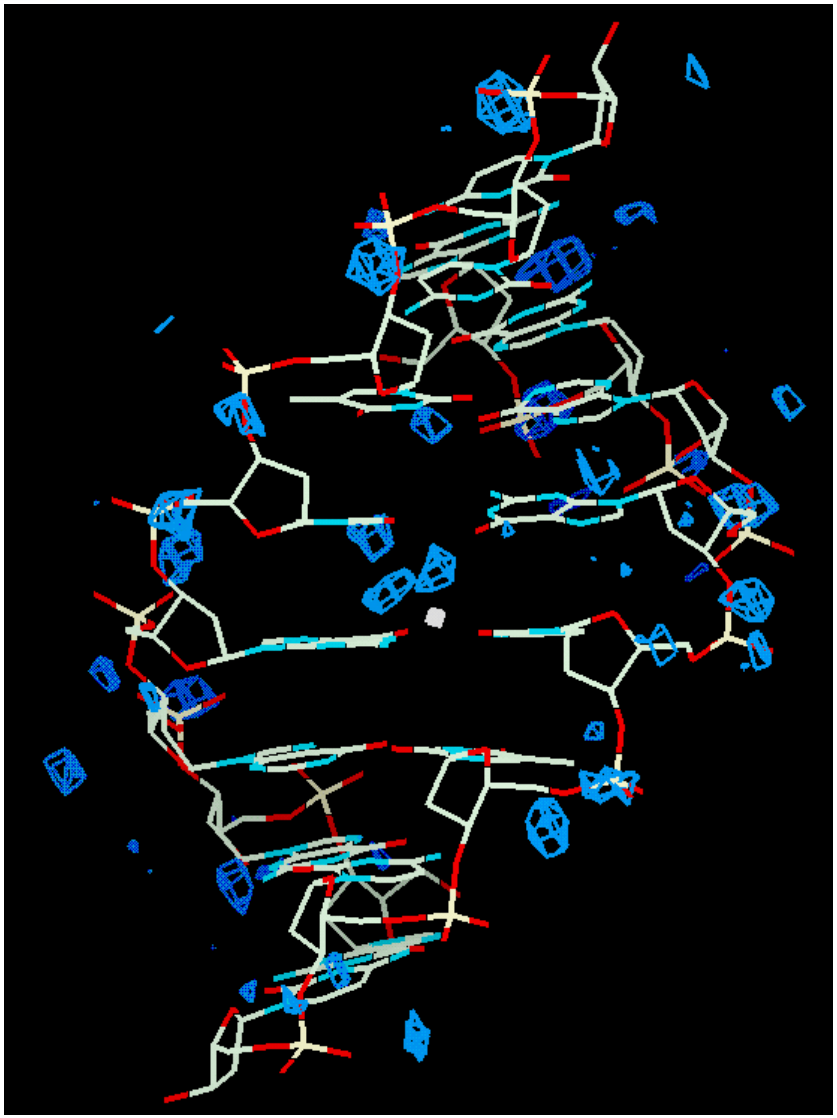
Long λ and Molecular Replacement

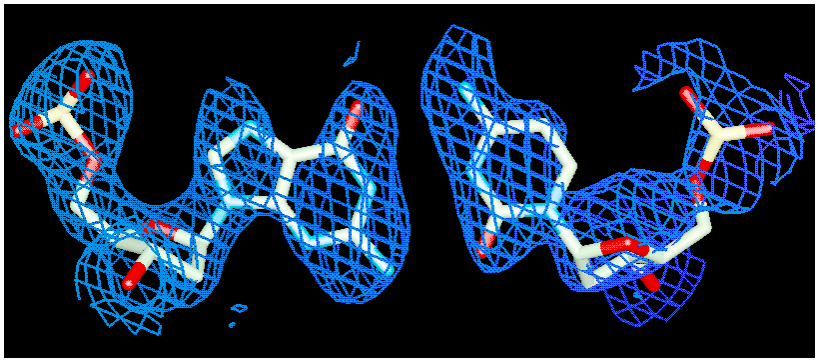


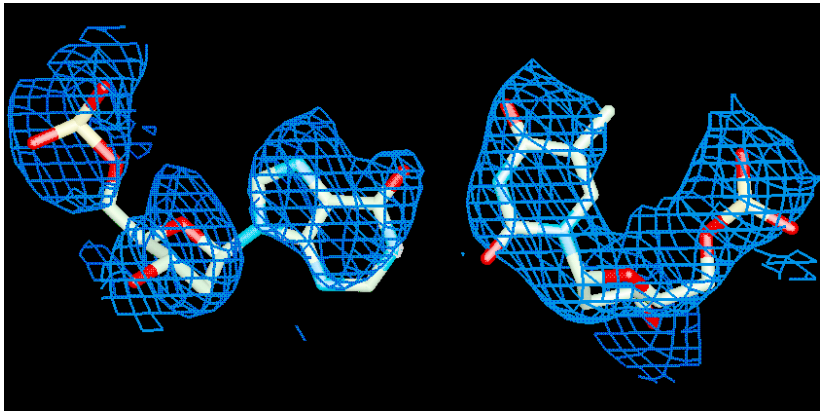
Anomalous difference fourier map (ΔF_{ano}, α – 90º) calculated using phases from a MR solution and anomalous differences measured at λ = 1.54 Å. The peaks in the map represent the 15 S atoms in the asymmetric unit (Schuermann & Tanner, (2003) *Acta Crystallogr*. **D59**, 1731-1736).

Two choices: 1) Use S positions as makers in map interpretation. 2) Calculate phase probability distributions using anomalous differences & S positions. Use these to get an *unbiased* electron density map.

MRSAD with P at λ=1.89Å

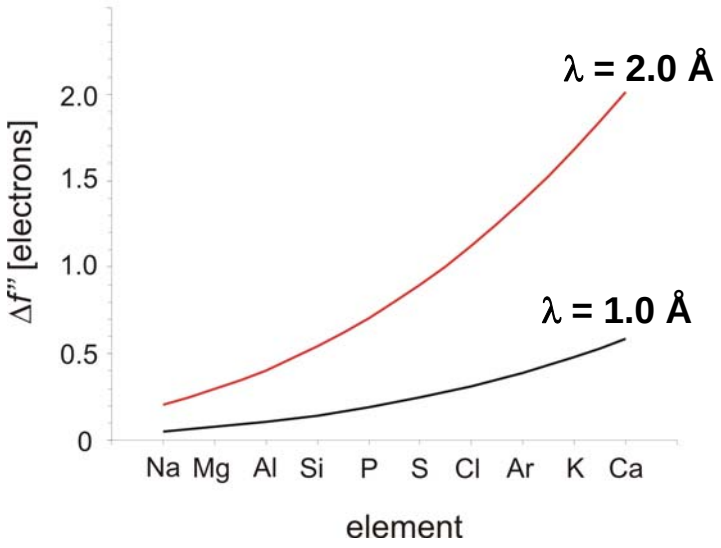






Maps resulting from SHARP phases subject to density modification with DM. d_{min} = 2.4 Å
Model: unrefined MR solution using data to d_{min}=3.0Å

Using long λ to indentify unknown ions



Anomalous difference fourier maps (ΔF_{ano}, α – 90º) will reveal which 'waters' are really ions. Will stand-out better if data collected at longer wavelength. Use XRF analysis to find which ions are likely to be present. If more than one type of ion, peak heights (and coordination sphere) will help discriminate.

Mueller-Dieckmann, C., *et al.*, (2007). *Acta Cryst.,* **D63**, 366-380.

Collecting the best possible data at long λ

Use what you've learned today – it all applies (probably even more) at long λ!!

Other measures you can take include:

- Reduce air scatter/absorption of diffracted beams
  - He-filled sample environment? vacuum until end of slit box?
- Eliminate λ/3 contamination of beam by harmonics from monochromator
  - better use of pushers, use of small mirror before sample
- Routine use of (mini-)kappa (***see talk by A. McCarthy***)
  - 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?
- Better treatment & use of radiation damage (***see talk by M. Weik***)
- Use software to help predict strategy & reduce radiation damage (***see talk by A. Popov***)

Increase the signal

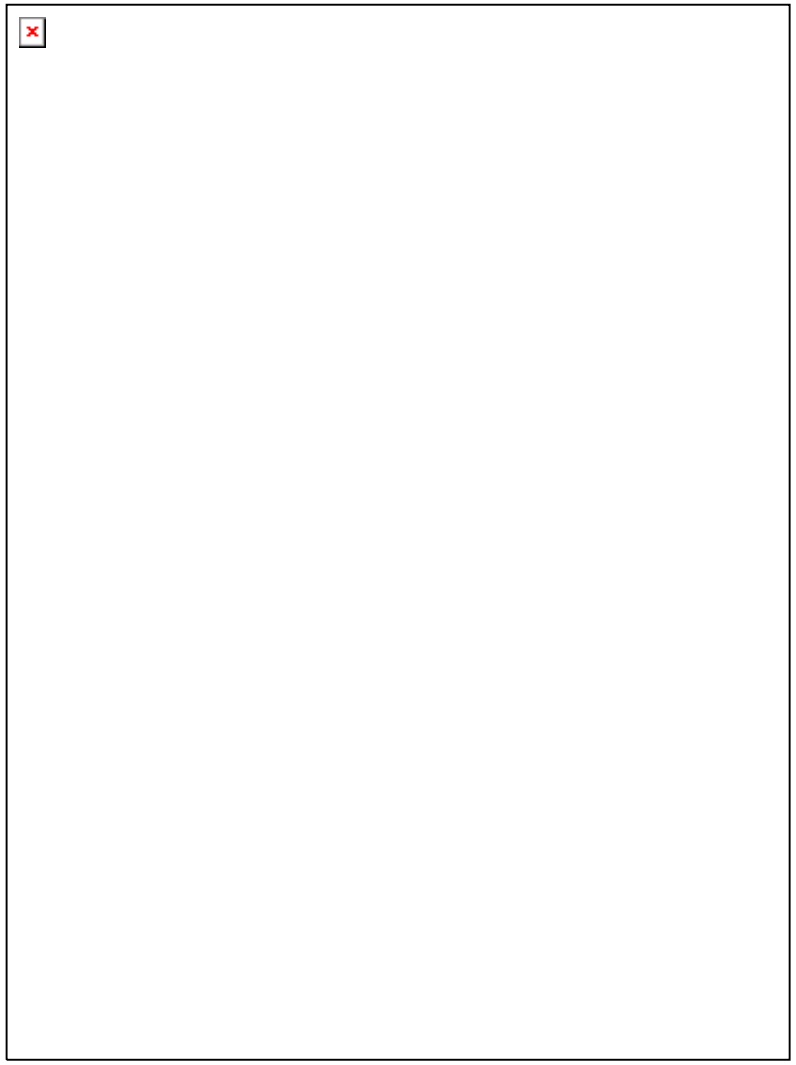
- **Is there an optimum wavelength for collecting our data? Is it λ = 2.1 Å? (S-SAD)

Make sure we collect enough data

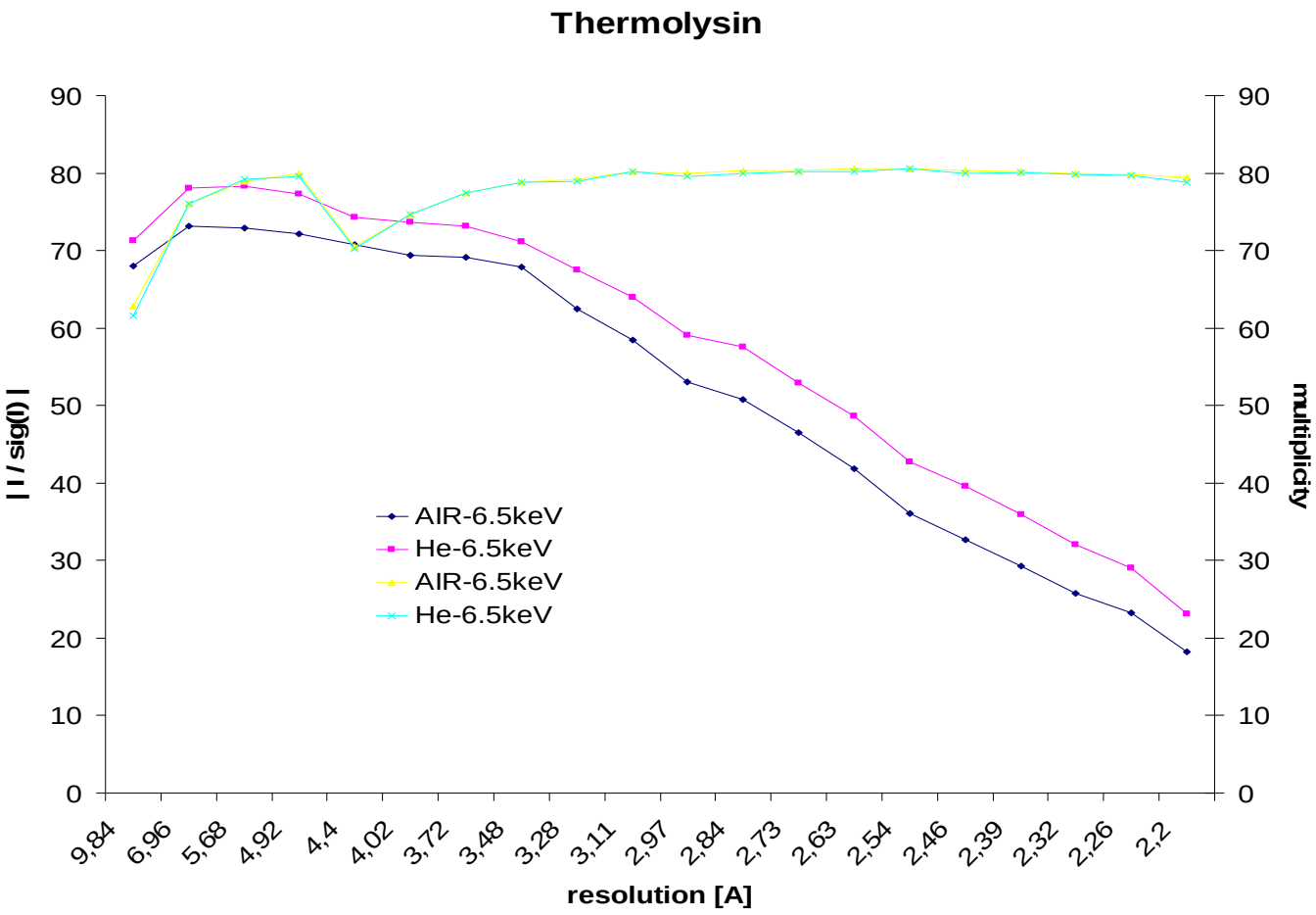
- Automatic substructure determination (and more, ***see talks by T. Schneider & S. Panjikar***)

*Brockhauser *et al.*, (2008). *J. Appl. Cryst.*, **41**, 1057-1066; Leal, *et al.*, (2008) *J. Appl. Cryst.*, **41**, 729-737. **Mueller-Dieckmann et al., (2004) *Acta Crystallogr*. **D60**, 28-38.

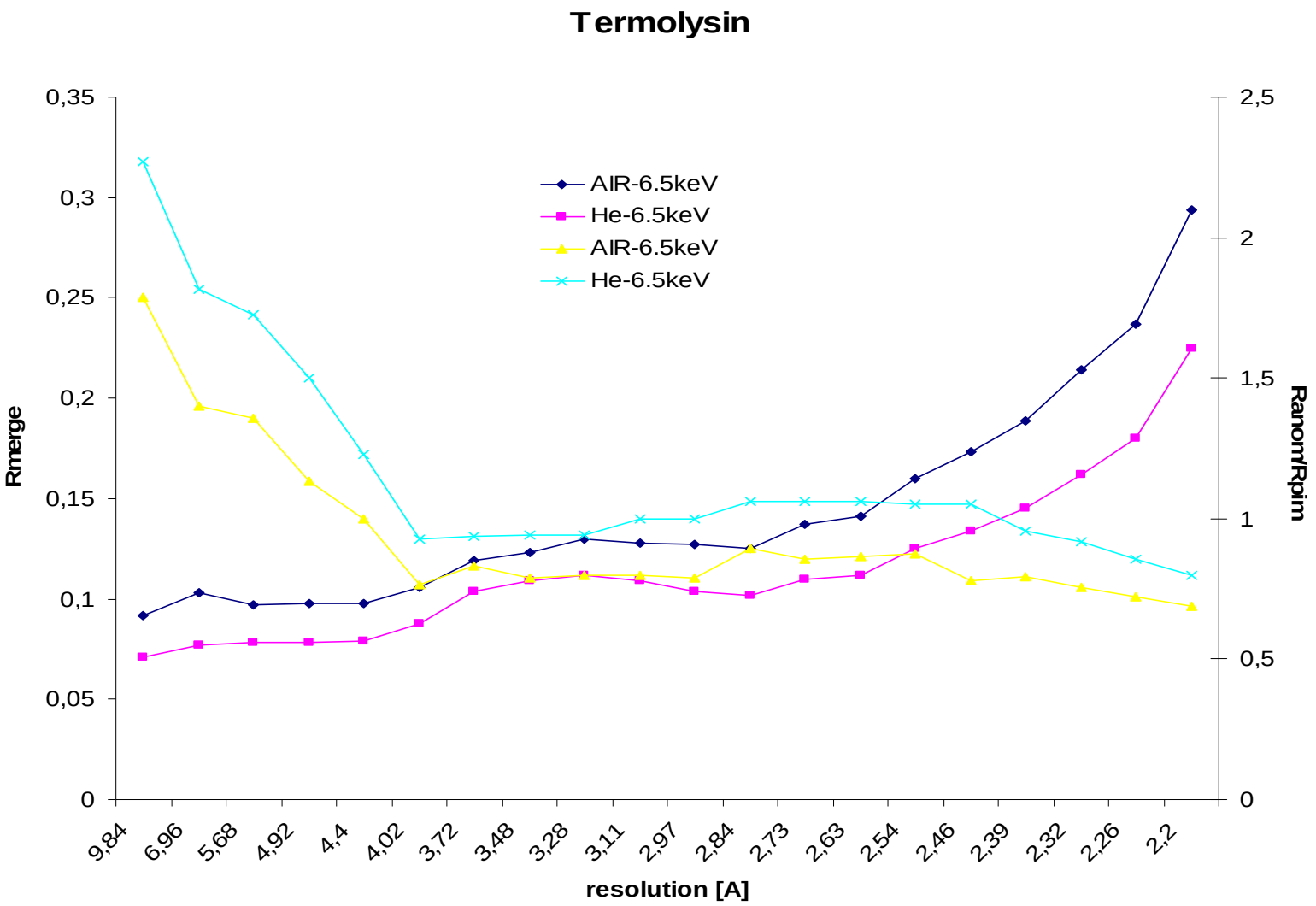
Helium cones - I



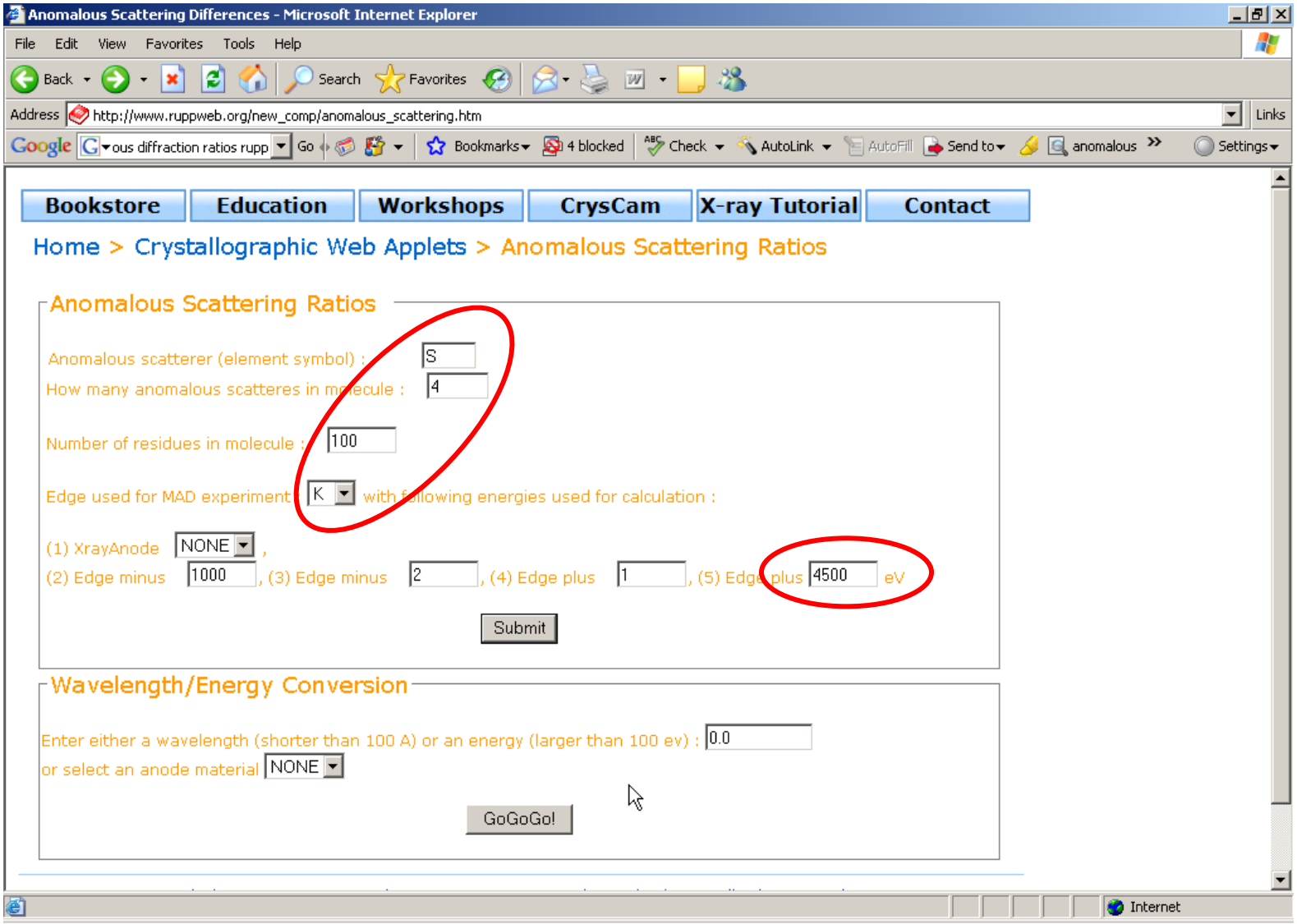
Helium cones - II



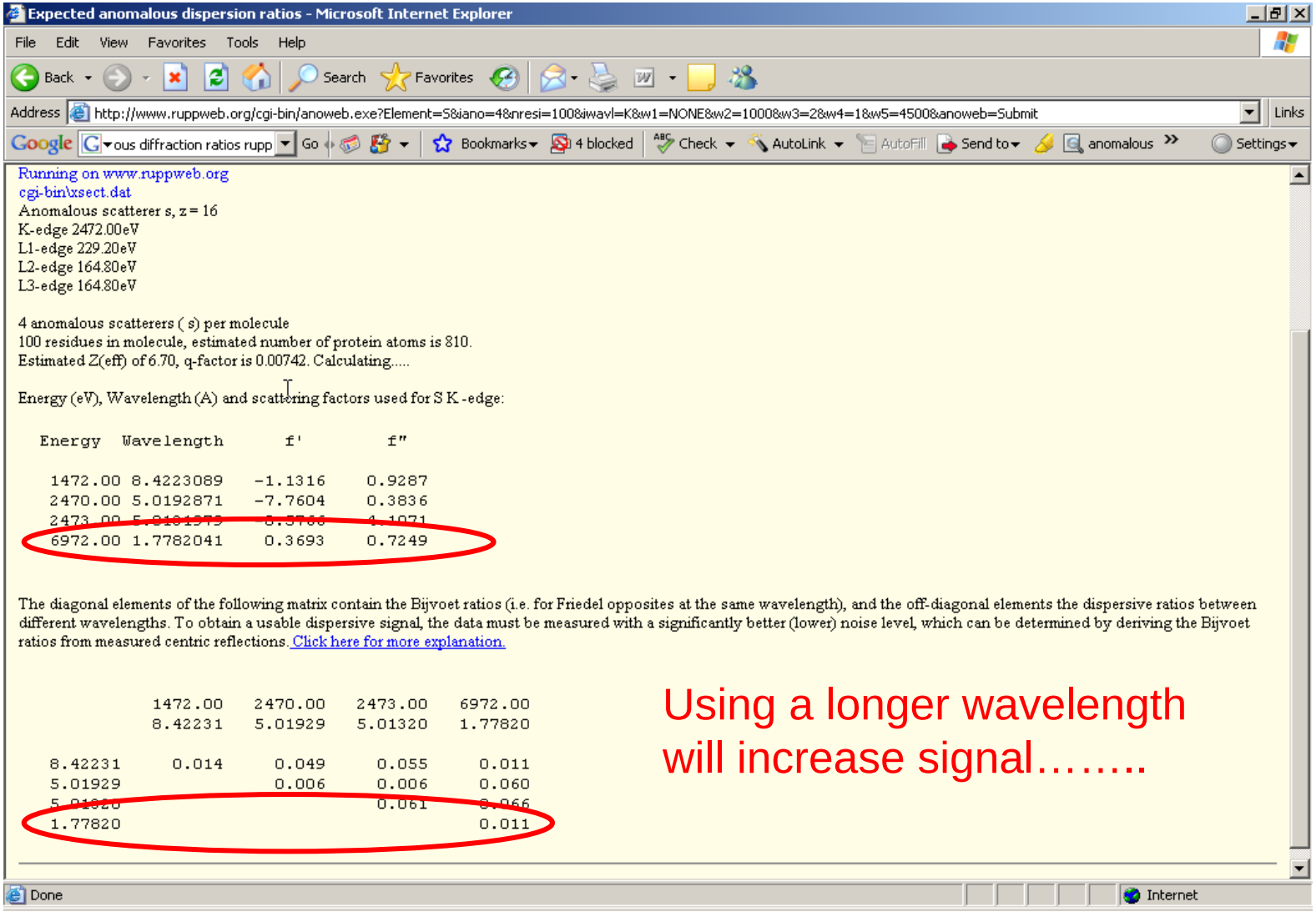
Helium cones - III



Calculating anomalous signals - I



Calculating anomalous signals - II



Using a longer wavelength will increase signal……..

Thanks for your attention!