Chooch
Calculation of Anomalous Scattering Factors from X-ray fluorescence data

Gwyndaf Evans

October 27, 1998

1 Introduction

The effects of anomalous scattering which manifest themselves most when the incident X-ray energy approaches the absorption edge energies of an atom are described mathematically by two correction terms which are applied to the normal atomic form factor or Thompson scattering factor f_o. The modified scattering factor is given by $f = f_o + f' + if''$ where f' is the real part and f'' the imaginary part of the anomalous scattering correction term.

When thought of in classical terms anomalous scattering is essentially analogous to any resonance effect such as an electrical LC circuit. The optical theorem [1] demonstrates that the imaginary term f'' is directly related to the atomic absorption coefficient for an atom by

$$f'' = mc\epsilon_o E\mu_a/e^2\hbar \tag{1}$$

where μ_a is the atomic absorption coefficient, E the X-ray energy and all other symbols take there usual meaning. As in other resonance phenomena such as dielectric susceptibility, the real part of the dispersive term is related to the imaginary part by a Kramers-Kronig (K-K) transformation. In the case of X-ray scattering the K-K transform takes the following form

$$f'(Eo) = \frac{2}{\pi}\oint_o^\infty \frac{(Ef''(E))}{(E_o^2 - E^2)}dE \tag{2}$$

Why do we need to know f'' and f'? When performing Multiple wavelength Anomalous Diffraction (MAD) experiments a crucial prerequisite is knowing at which wavelengths to measure diffraction data. This can only be determined at the time of the experiment due to two main reasons

1. For a particular heavy atom element the X-ray energies to be measured are largely dependent on the environment of that element within the protein sample and its orientation with respect to the polarization vector of the incident X-ray beam.

2. The calibration of the incident X-ray energy at different X-ray beam lines will rarely be the same and as yet no calibration standards have been established which are common to all crystallographic facilities.

In addition the calibration of each beam line may vary over time. As previously stated the f'' value is directly related to the atomic absorption coefficient for an atom. For a discussion of the difficulties and solutions associated with X-ray energy calibration for MAD see [2].

The absorption is directly proportional to the X-ray fluorescence emitted from the atom as a result of absorption of the incident X-rays. This provides the experimenter with a means of determining the dependence of f'' on the X-ray energy. f' may then be determined computationally using the K-K relationship. This provides the necessary information with which to make a rational choice of which wavelengths to measure for the experiment. Clearly we also establish the magnitudes of the anomalous scattering factors as a function of X-ray energy. These values are potentially useful as starting points for heavy atom refinement during the latter stages of data analysis. However the most important aspect of the evaluation of anomalous scattering factors is the correct determination of the energy dependence of f' and f''.

2 Determination of f'' and f'

Obtaining f'' from fluorescence data

Fluorescence spectra are generally measured directly from the same frozen protein crystal sample from which the diffraction data is to be measured. The spectra are typically recorded using a photo-multiplier (e.g. Bicron tube) or an energy resolving photo-diode type detector (e.g. Amptek). In both cases the fluorescence signal is recorded on an arbitrary scale. Determination of the corresponding f'' spectra is done via two stages.

Firstly the raw fluorescence spectrum must be background subtracted and corrected to subtract out any additional scattering effects which may be energy dependent. This procedure is typically very straight forward for data measured using a good energy resolving detector such as the Amptek since the measured signal is essentially dominated by fluorescent X-ray counts. However photo-multiplier tubes which have poor energy resolution will typically measure the elastic scattering components of the X-rays as well as the fluorescence signal and will therefore usually require a more careful background correction.

The procedure involves applying a low order polynomial fit separately to the below edge region of the spectrum and the above edge region of the spectrum. The fits should be generated away from the absorption edge where the XANES effects are smallest. These two polynomials can then be applied to the raw spectrum such that it is normalized to be zero fare below the absorption edge and unity above the edge. The normalized signal $N(E)$ is obtained by

$$N(E) = R(E)\left\{f''_{above}(E) - f''_{below}(E)\right\} + f''_{below}(E) \tag{3}$$

where R(E) is the raw data, f''_{below} is the polynomial fit in the below edge region and f''_{above} the fit for the above edge region. All are functions of the X- ray energy E. Theoretical values of f'' have been calculated by Cromer & Libermann [3]. The calculations however take no account of the effects of coordination of anomalous scattering atoms to other atoms. The effects of coordination are most visible in the near edge region which also happens to be the region of interest for MAD. Therefore the Cromer & Libermann tables are not applicable in the near edge region. However away from the absorption edge above and below in energy the tables provide a good estimate of the true f'' values and therefore provide a means by which the normalized fluorescence data can be converted to a f'' spectrum. The theoretical spectrum is essentially multiplied into the experimentally determined spectrum to produce an experimentally determined f'' spectrum.

Obtaining f' from f''

Given a f'' spectrum the K-K transformation may be used to directly obtain a f' spectrum. An algorithm has been described [4] which allows this to be carried out computationally. Complications arise in the calculation because of the singularity in the integrand of Equation 2 arising when E is equal to Eo and also because of the impractical limits of integration. The singularity is dealt with conveniently by the above algorithm and the integration limits are chosen such that the calculation remains possible but does not become inaccurate. Integration limits which extend only a few keV above and below the absorption edge will usually provide an accurate estimate of the X-ray energy corresponding to the minimum value of f' but the magnitude of the f' curve will in general be incorrect. To obtain highly accurate magnitudes integration limits are chosen which extend up to $50\times$ absorption edge energy and to very low energies of say 1 keV. These calculations however are time consuming and not totally necessary given the experimental requirements. Therefore modest integration limits may be chosen such that the duration of the calculation is tolerable as well as the accuracy of the f' curve. In the case of the Se K edge recommended integration limits for a full calculation are 1.2 keV and 630 keV. This calculation takes 120 secs. on a 150 MHz Pentium MMX for a 101 point spectrum. Using integration limits of 1.2 keV and 30 keV takes 8 secs. on the same CPU and introduces only a $+0.3$ e error into the resulting f' curve. Such errors are acceptable for the majority of cases.

3 Organization of the program

Calculating anomalous scattering factors from raw fluorescence data with CHOOCH requires the use of two programs, BENNY and CHOOCH. The programs are both called from a shell script, `Chooch.sh`, which is excecuted by the user. The input and output files used and generated by the programs are described in detail below.

BENNY

BENNY reads in raw fluorescence data from a measurement perfomed on a MAD crystal sample and performs a number of manipulative tasks. Firstly it performs background correction by fitting a polynomial [5] of degree 0 to 3 to the below edge and above edge regions of the spectrum and normalizing such that the fluorescence is zero far below the edge and unity far above the edge. The background fitting step is far from being automatic. The reason for this is that fluorescence spetra are measured in many different ways, using different detectors and over differeing energy ranges. This can give the spectra unusual background properties and make the detection of the true background level difficult.

The ideal fluorecence spectrum is one where

1. the background scatter is low and varies slowly and smoothly with energy.
2. the data is measured from well below the absorption edge ($< (E_{edge} - 200)$ eV), to allow the background level below the edge to be easily established, to well above the

edge ($> E_{edge} + 200$ eV), to establish the level of signal + background above the edge and allow good normalisation to be performed.

The use of a good energy discriminating detector will often help satisfy the first criteria. However only the user and beamline staff can satisfy the second suggestion.

If the spectra has been measured well there will be enough data either side of the absorption edge to allow a good fit of the backround levels to be made. The program requires the user to input values of the X-ray energy between which the background fits will be made. This is done graphically with the cursor. First click on the low energy point and the on the high energy point. The user can also choose the type of fit to be generated by choosing the polynomial order. This will be either a straight line, a quadratic or a cubic. Choosing the [0] option requires that the user select the actual background level with zero slope (a completely manual option reserved for poorly measured data where it is not possible to determine the background by fitting).

After normalisation the program goes away and applies a spline fit [6] to the normalized data creating a smooth curve. It additionally calculates 1st, 2nd and 3rd derivatives of smoothed data for input into CHOOCH. This step is fully automatic. Parameters varying the type of fit performed are hard coded into the program and have so far served well on all cases that I have tested the program on so if it should fail please let me know.

Files

***file*.raw** The raw input fluorescence data to BENNY. The first line should contain the number of data points (integer). The second line in usually blank but is interpreted as text. It can be used as a comment line for the data. Each subsequent line should contain three values referring to one data point - The data point number (integer), the X-ray energy in eV (real) and the fluorescence signal on an arbitrary scale (real).

e.g.

```
Fluor. spectrum for element Qu  ; Title (a80)
801                             ; No. data points (free format)
12300.0  2002                   ; Energy (eV), Flu. Signal (free format)
12300.5  2030
12301.0  2035
12301.5  2450
.
.
.
12700.0  6700
```

splinor Output by BENNY and input to CHOOCH containing the X-ray energy, smoothed normalized fluorescence data 1st, 2nd and 3rd derivatives (format(5f13.3)).

e.g.

```
Fluor. spectrum for element Qu  ; Title (a80)
     101                        ; No. of data points (i8)
```

```
      12654.043      0.013     -0.005      0.052     -0.083
      12654.438      0.015      0.009      0.019     -0.081
      12654.831      0.019      0.011     -0.008     -0.051
      12655.226      0.022      0.006     -0.015      0.018
   .
   .
   .
      12693.076      1.004      0.007      0.031      0.047
      12693.471      1.009      0.023      0.050      0.047
```

CHOOCH

Takes output from Benny and reads input data about the element and absorption edge in question from a command file. It then calculates f'' and f' from the smoothed normalized fluorescence data and displays the resulting curve. The program automatically selects the peak f'' energy and the minimum f' energy and outputs them. An important requirement of the program is that all the data be measured on a strictly increasing X-ray energy scale of constant energy increments this requirement is satisfied however by the smoothing procedure performed by BENNY.

Files

splinor Output from BENNY as described above.

element.**dat** Data files for each element at a specific absorption egde containing information pertinent to the calculation of the anomalous scattering factors. The files are contained in the `Chooch-xxx/data` directory. The files contain the following information for each element. At present all files are set up foe absorption edge most commonly measured for an element, e.g. K-edge for Se, L3 edge for Pb etc. These files contain information about the integration limits used during the Kramers-Kronig calculation and depending on the CPU speed of the machine performing the sums, it may be necessary to adjust the limits to make the calculation time tolerable and practical. The information in the file is compiled from a number of sources the most prominent being the article by Cromer and Libermann on the calculation of anomalous scattering factors [3] and the McMaster tables [7]

e.g. Selenium K edge

```
78.96                  ; Atomic weight of atom in Atomic units.
34.0                   ; Atomic number of atom.
k                      ; Absorption edge of interest (k or l3 only)
12658.0                ; K-edge energy in eV
1652.0                 ; LI-edge energy in eV
1474.3                 ; LII-edge energy in eV
1433.9                 ; LIII-edge energy in eV
229.6                  ; MI-edge energy in eV
1.160                  ; LI jump factor    from McMasters tables
1.410                  ; LII jump factor   from McMasters tables
4.587                  ; LIII jump factor  from McMasters tables
```

```
0.0000  1.2677e+01  1.4704e+01  1.3075e+01 ; McMasters matrix elements
0.0000 -2.3975e+00 -2.3885e+00  1.8323e-01 ; McMasters matrix elements
0.0000  0.0000     -1.0587e-01 -6.9426e-01 ; McMasters matrix elements
0.0000  0.0000      0.0000      5.0228e-02 ; McMasters matrix elements
-0.215                ; Cromer \& Liberman 5Etot/3mc**2 term.
1652.0                ; Lower energy limit for integration.
30000.0               ; Upper energy limit for integration.
12781.0 3.76 12915.1 3.68 ; Describes a straight line approximation to
12523.7 0.50 12651.5 0.50 ; theoretical f" data above and below edge.
```

file.**efs** Output from CHOOCH containing calculated anomalous scattering factors. The example below is taken from the examples directory in the CHOOCH distribution.

e.g.

```
Se test data from a foil Chooch test data ; Title (a80)
     101                                  ; No. of data points (i8)
  12654.04      0.51     -6.47            ; Energy, f'', f' (3f10.2)
  12654.44      0.51     -6.52
  12654.83      0.53     -6.58
.
.
.
  12692.29      3.77     -4.81
  12692.68      3.76     -4.79
  12693.08      3.76     -4.72
  12693.47      3.78     -4.69
```

file.**inf** Output from CHOOCH containing summary of calculation.

e.g.

```
Se test data from a foil Chooch test data
 Total points integrated     :      71890
 Integration limits low/high :    1653.06  30000.25
 First/last data points at   :   12654.04  12693.47
 Energy scale increment      :  0.394

 Inflection point at   12665.48 with  f' of  -9.7
             Peak at   12667.45 with f'' of   6.6
```

file.**ps** Output from CHOOCH. Once the anomalous scattering factors have been calculated and displayed you can select the 'p' option and dump a POSTSCRIPT file of the plot. If you do not press 'p' the file is still generated unfortunately but will contain inly a pre-amble.

N.B. Out of all the above files only file.raw need be created by the user - all other are produced by either BENNY or CHOOCH.

4 Installation

If you are reading this manual the chances are you have already installed the program and have read the installation guide README.Install

5 Running the program

The program should be excecuted in the directory containing the raw fluorescence data file, *file*.**raw** by typing `Chooch.sh` <element> *file* at the unix prompt. For example to run the example fluorescence data named SeFoil.raw type

```
Chooch.sh  Se  SeFoil
```

If all is properly installed a PGPLOT window will appear displaying your fluorescence spectrum. You will then be guided through the procedure. A simple run down of the procedure follows:

1. Fitting the below edge region - in the PGPLOT window just type 0,1,2 or 3 depending on the type of fit you would like. More often that not you will use either 0 or 1. Fluorescence data from proteins is typically measured over a fairly limited range which requires a bit of guess work when determining the background level (hence the need for option 0). Anyway, after entering a choice the cursor will appear and if you chose option 0 click at an appropriate level for the below edge background. If you chose any other option select two energies - low first, then high - which ideally should be $E_{edge} - 100$ and $E_{edge} - 25$ or so. If the data doesn't extend that far below then use option 0 or be very careful!

2. Fitting the above edge region - this is the same as for the below edge region but you should take care that you ignore any near edge effects when selecting the energy range for a fit. You shouldn't bias the fit with a large white line peak. Therefore use caution and select a low energy which is away from the near edge region where the XANES ripples begin to die out ($\sim E_{edge} + 30$).

3. When inspecting the normalisation result you can decide you don't like it and by typing '`n`' you can return to the beginning and refit the backgrounds.

4. If you do like it then just click the mouse to continue and the program will smooth the data for you.

5. Proceed by hitting '`c`' and the PGPLOT window will disappear while CHOOCH calculates the anomalous scattering factors. When it's complete another PGPLOT window will appear with the results. It also prints estimates of the f'' peak energy and the f' minimum energy.

Using the zoom facility

At most stages of the procedure you can zoom in on your spectrum by pressing '`z`' and selecting a low then a high X-ray energy with the mouse cursor. The zoomed region will then appear in the same window. You can redraw the original data range by pressing '`r`'.

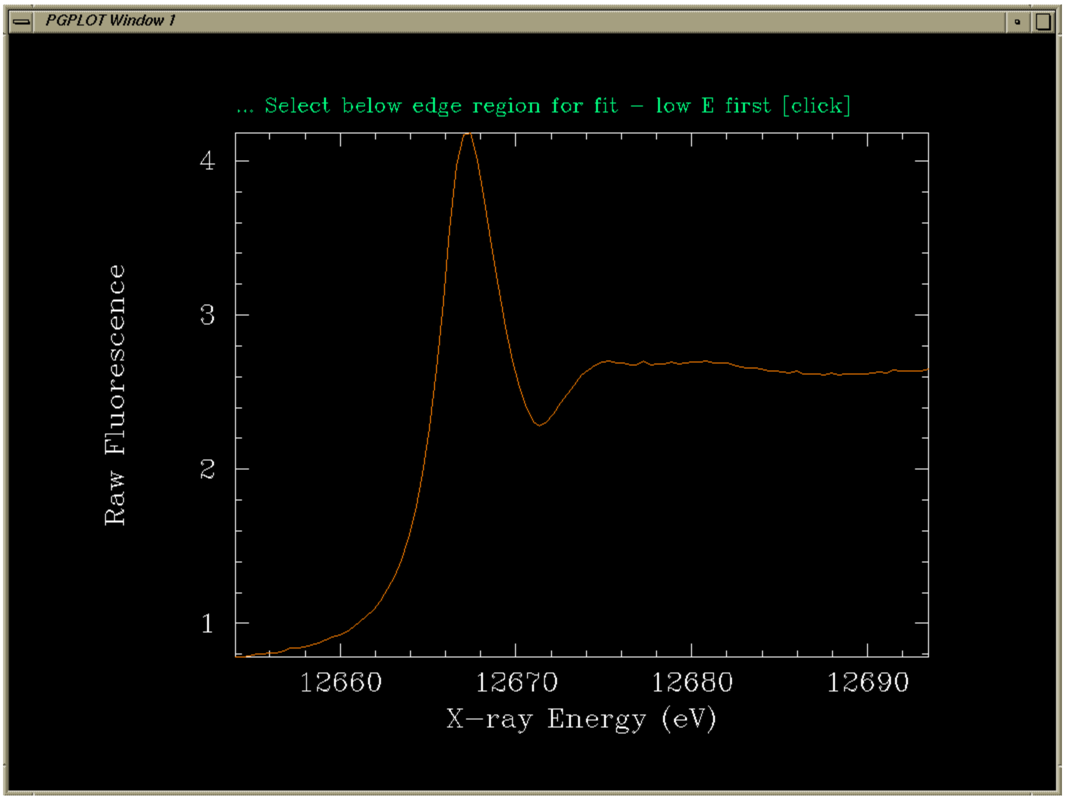

Figure 1: PGPLOT window seen on first excecuting Chooch.sh.

Generating a PostScript plot of the result

Once Chooch has produced the PGPLOT window containing the anomalous scattering curves you can generate a plot of the output by selecting the 'p' option in the PGPLOT window.

References

[1] R.W. James. *The Optical Principles of the Diffraction of X-rays.* G. Bell and sons Ltd, London, 1969.

[2] G. Evans and R. F. Pettifer. Stabilisation and calibration of x-ray wavelengths for anomalous diffraction experiments using synchrotron radiation. *Rev. Sci. Instr.*, 67(10), October 1996.

[3] D.T. Cromer and D. Libermann. Relativistic calculation of anomalous scattering factors for X-rays. *J. Chem. Phys.*, 53:1891–1898, 1970.

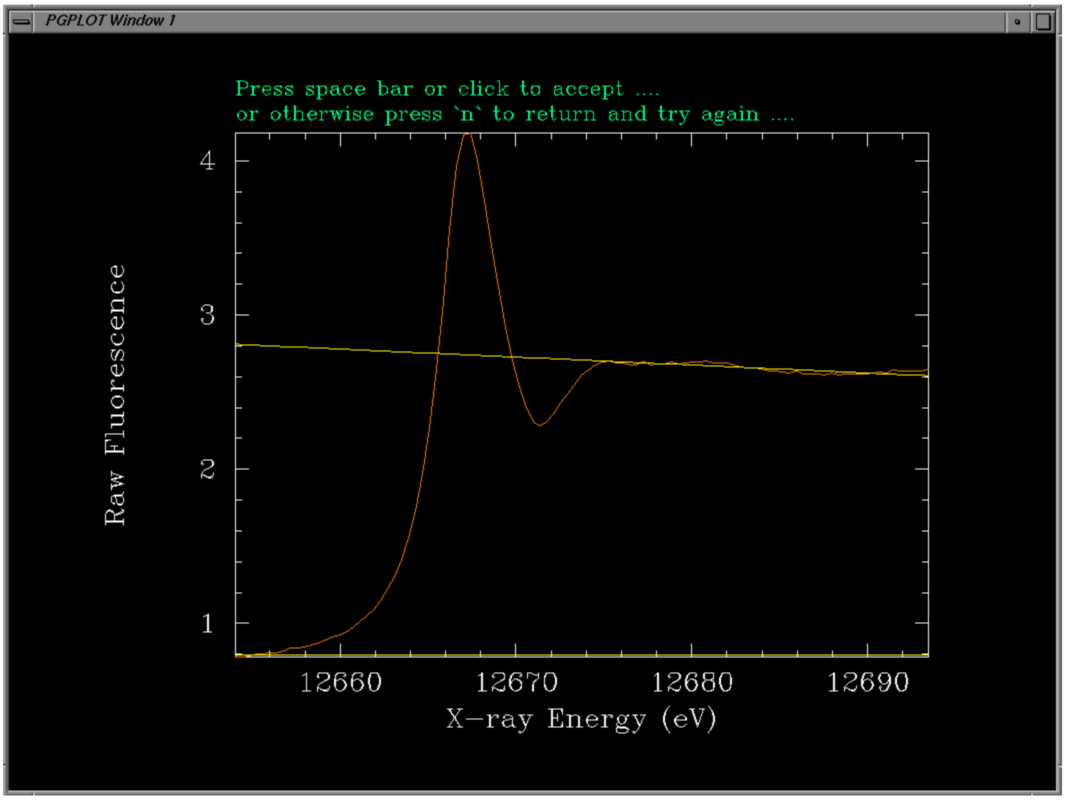

Figure 2: PGPLOT window after fitting the below edge region using option [0] and the above edge region using option [1] fitting between 12675 eV and 12695 eV.

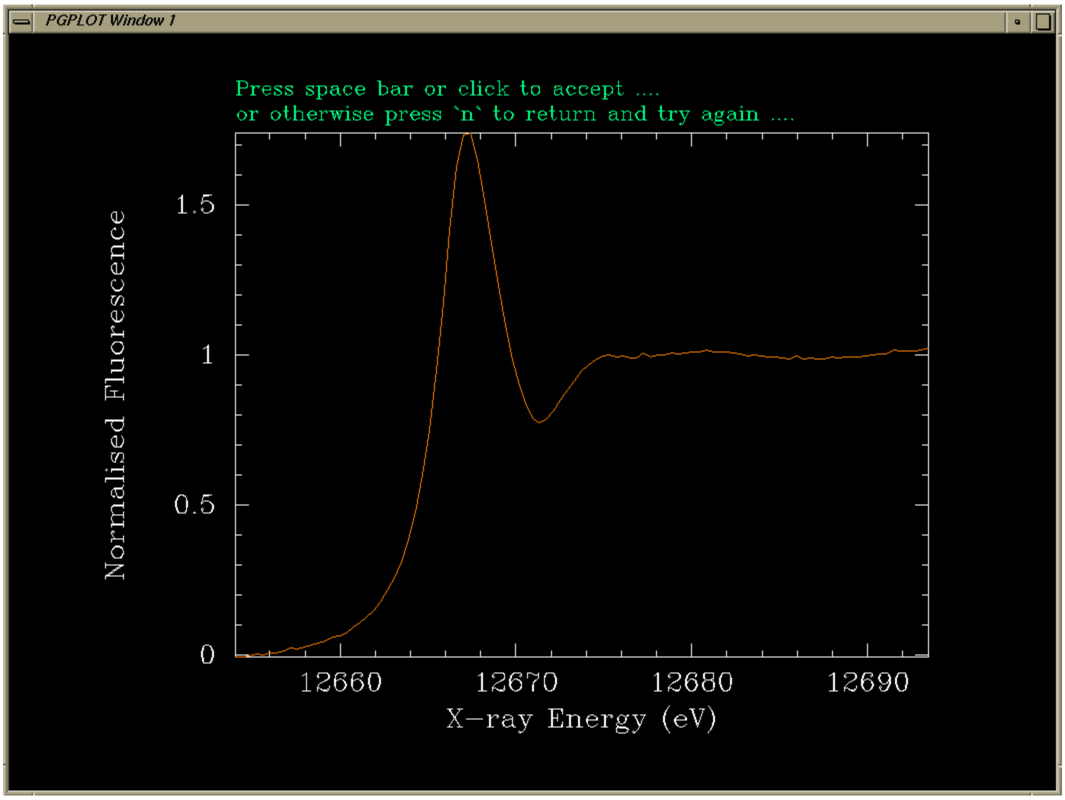

Figure 3: PGPLOT window after normalisation.

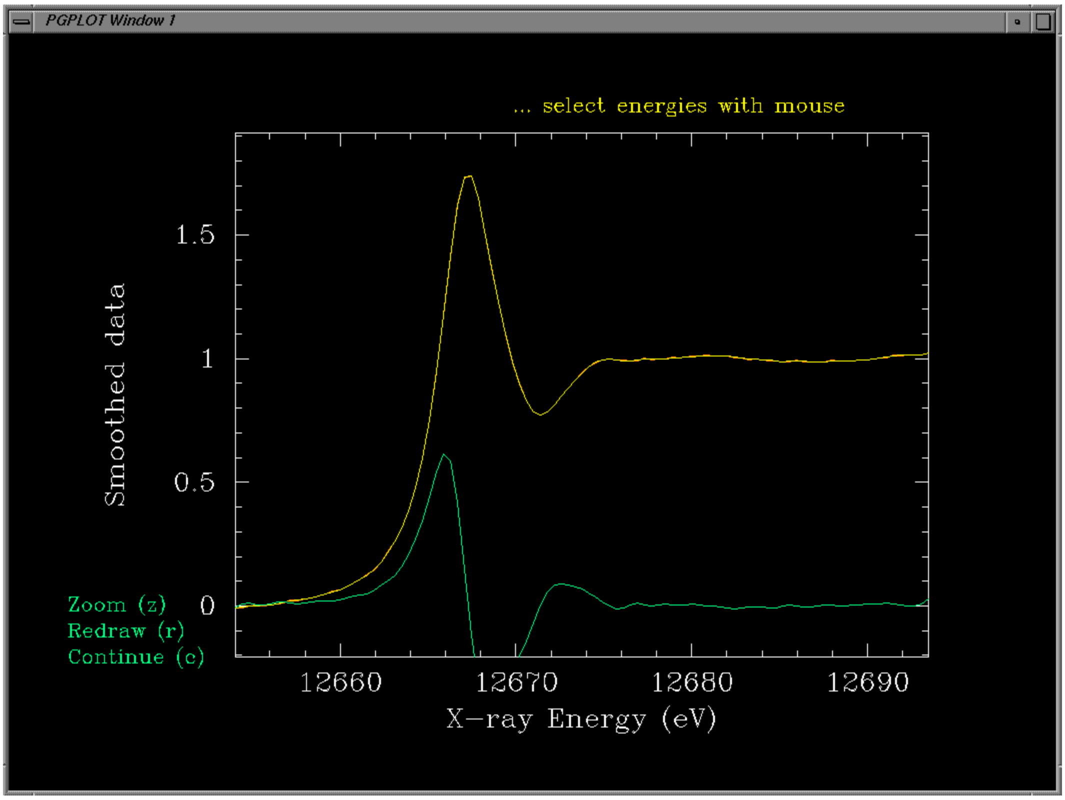

Figure 4: PGPLOT window after smoothing showing the 1st derivative of the smoothed spectrum.

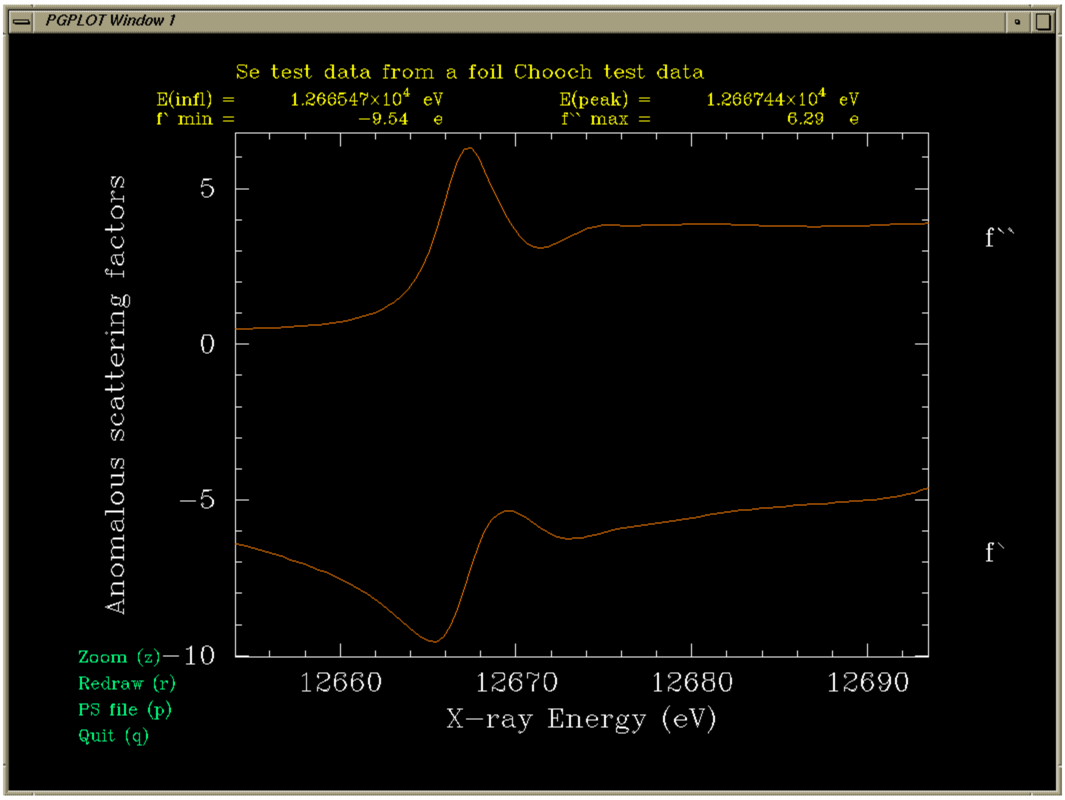

Figure 5: PGPLOT window after running Chooch and calculating the anomalous scattering factors.

[4] J.J. Hoyt, D. de Fontaine, and W.K. Warburton. Determination of the anomalous scattering factors for Cu, Ni and Ti using the dispersion relation. *J. Appl. Cryst.*, 17:344–351, 1984.

[5] L.F. Shampine, S.M. Davenport, and R.E. Huddleston. Fit discrete data in a least squares sense by polynomials in one variable. Fortran Subroutine, 1974.

[6] H.J. Woltring. Test programme for generalized cross-validatory spline smoothing with subroutine gcvspl and function splder using the data of c.l. vaughan, smoothing and differentiation of displacement- time data: an application of splines and digital filtering. Fortran Subroutine, 1986.

[7] W.H. McMasters, N. Kerr, Del Grande, J.H. Mallett, and J.H. Hubbekk. Compilation of X-ray cross sections. Technical Report UCRL-50174, Lawrence Radiation Laboratory (Livermore), 1969.