EXCURVE

Wolfram Meyer-Klaucke

DESY

Hamburg, Germany

(presentation is based on experiences with BioXAS courses held at EMBL Hamburg, thanks to Ian Harvey and Loretta Murphy)

BioXAS on metalloproteins and organism tissue

Why Excurve ?

- Pros:
  - Well suited for users from biology (non-experts)
  - Comes with ligand data-base (amino acids)
  - Allows for automation of analysis / scripts (towards end of presentation)
  - Small number of parameters to adjust prior start of refinement (defaults work well)
  - k-space refinement (no filtering, considers the noise in your EXAFS)
- Cons:
  - GUI not very helpful (reduces functionality, not covered today)
  - To beginners the commands and options are sometimes not obvious
  - In case defaults are not working well no instructive comment is given on how to proceed

How to obtain it? Send an email to **s.tomic@dl.ac.uk**

New Tools for the Analysis of EXAFS: The DL EXCURV Package

S. Tomic a;¤, B.G. Searle a, A. Wander a, N.M. Harrison a;b,

A.J. Dent c, J.F.W. Mosselmans d, and J.E. Inglesfield e

a *Computational Science and Engineering Department, CCLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, UK*

b *Department of Chemistry, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK*

c *Diamond Light Source Ltd., Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK*

d *Synchrotron Radiation Department, CCLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, UK*

e *Department of Physics and Astronomy, University of Wales Cardiff, Cardiff CF24 3YB, UK*

CCLRC Technical Report DL-TR-2005-001, **ISSN 1362-0207**

¤ Corresponding author.

Email addresses: s.tomic@dl.ac.uk (S. Tomic), b.searle@dl.ac.uk (B.G.Searle).

Excurve – where to find more information?

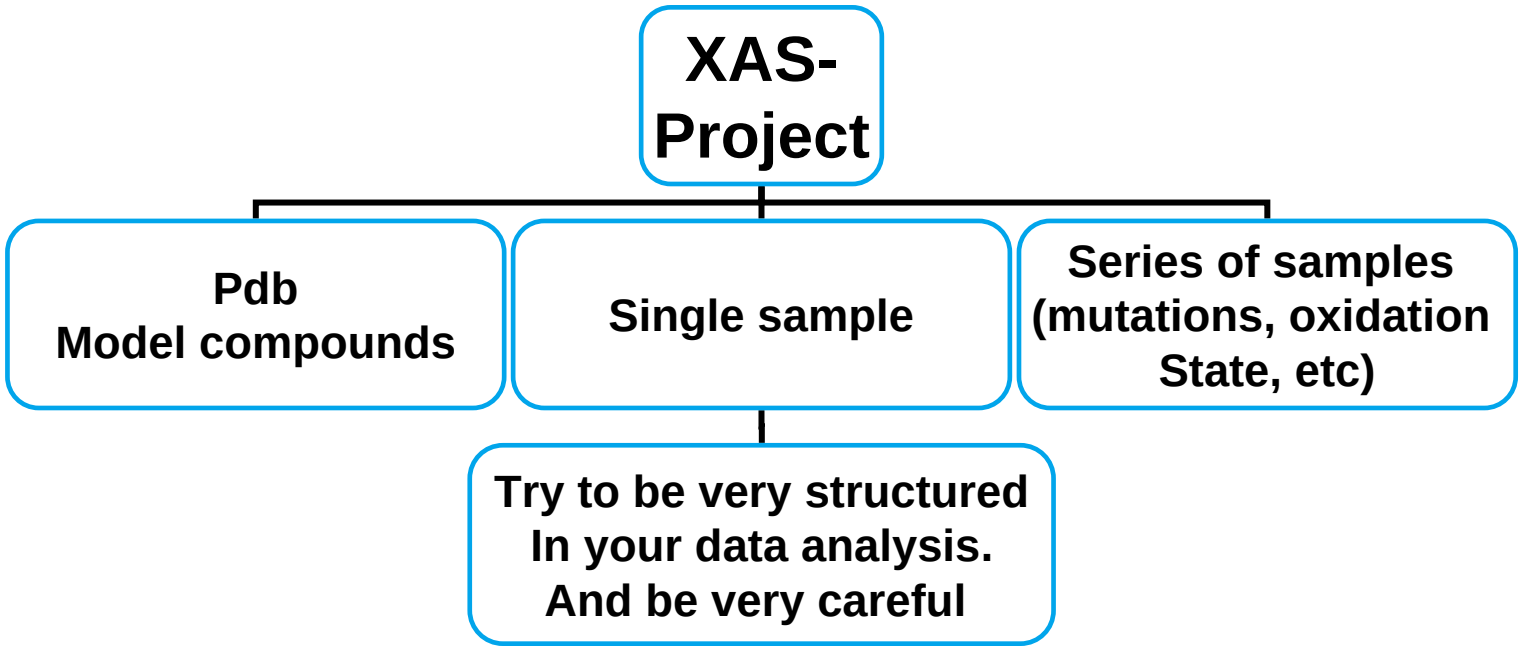
- Official site for DL-Excurv: http://www.cse.scitech.ac.uk/cmg/EXCURV/
- Further information by Fred Mosselmans and colleagues: http://www.diamond.ac.uk/Home/Beamlines/I18/meetings/xafs_course.html
- http://www.diamond.ac.uk/dms/Beamlines/I18/I18_Mar07/basicexcurve.pdf
- On the web: Search for Excurve, Excurv98 or DL-Excurv, or ABRA
- http://iopscience.iop.org/1742-6596/190/1/012033/pdf/1742-6596_190_1_012033.pdf
- http://scripts.iucr.org/cgi-bin/paper?S0909049509040576
- **What about automation** (the advantage of having no GUI)?

How to run Excurve ?

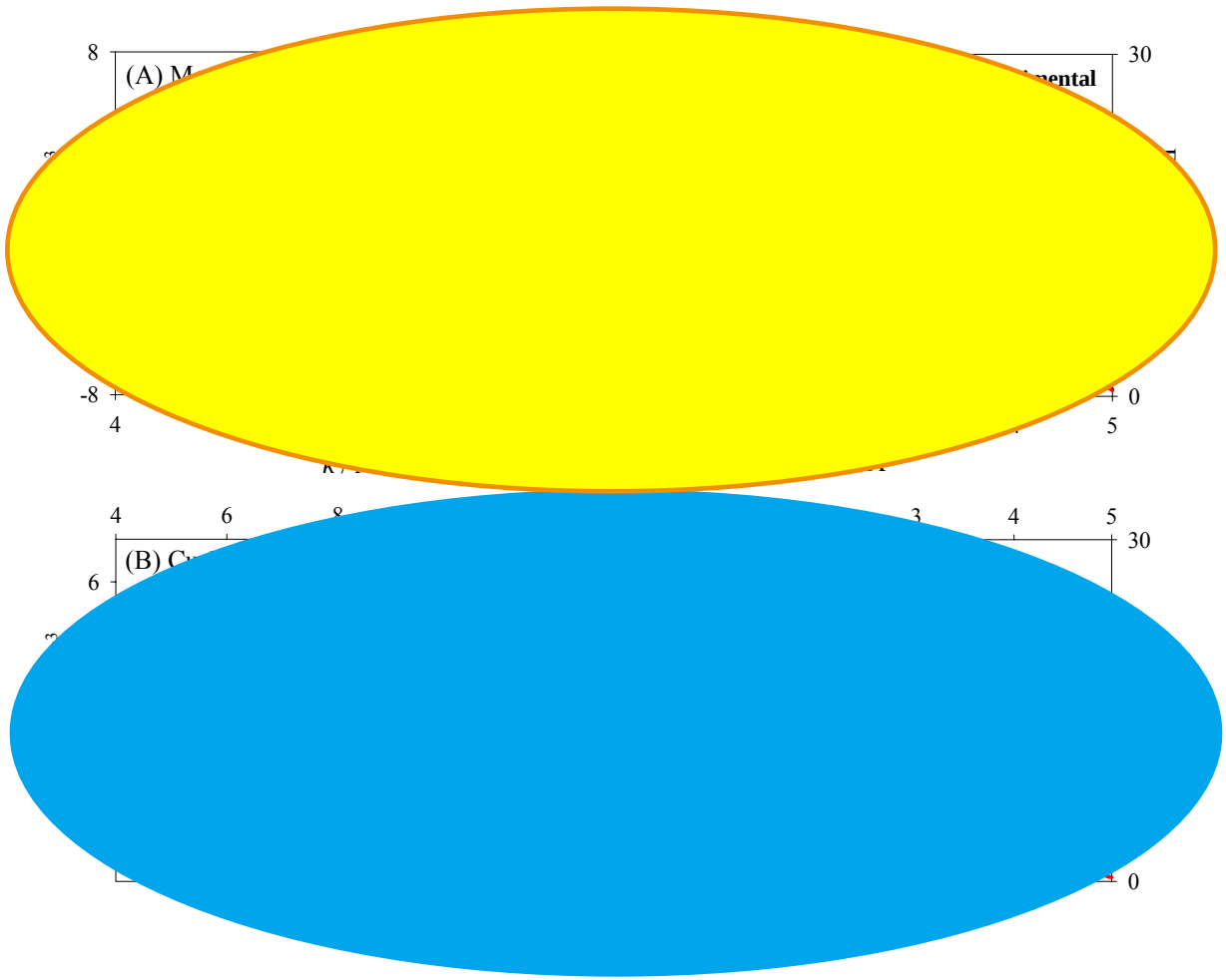
- Installation (here for Windows, linux works similar):
  - Install under C:\Excurve
- Start:
  - Start command prompt under Windows
  - Go for directory with your data and start the program by
  - C:\excurve\excurve
  - (In fact there are different versions depending on the number of data points you collected and the complexity of your analysis, check by

L dim (list dimension)

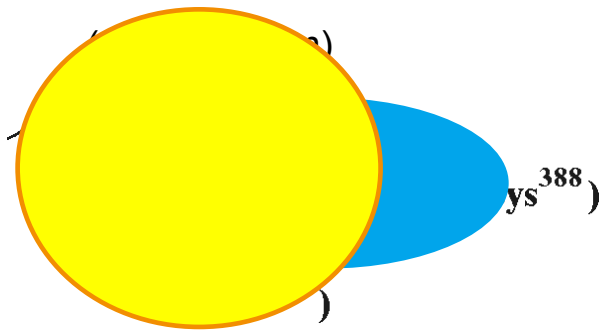
Basic concept:



[S] cluster in CODH



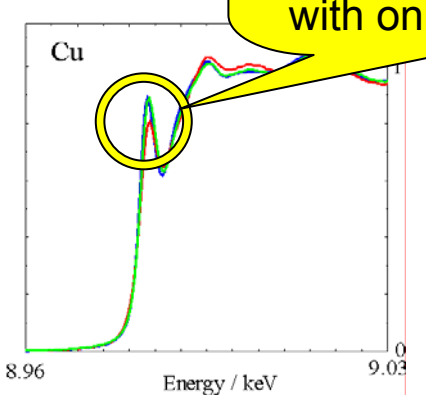
Structural model:

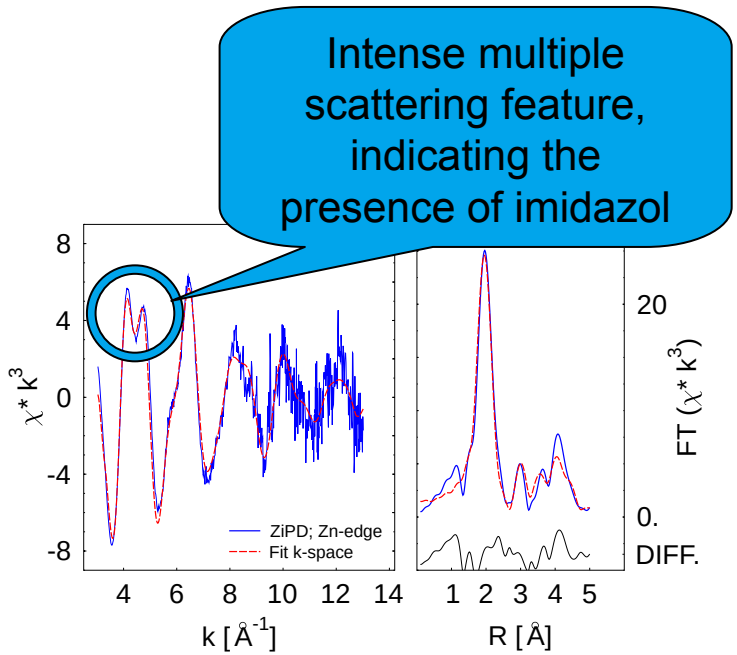


Binuclear cluster
Gnida et al. Biochemistry, 2003

see also:
Dobbek *et al*., PNAS , 15971 (2002)

Available information:





> Raw XANES

> Raw EXAFS

> Reference XANES

> Reference EXAFS

> Perform EXAFS refinement

> Compare results with element specific bond length

- Harding
- Bond valence sum analysis

1. Look at your spectra

> Raw XANES

> Raw EXAFS

> Reference XANES

> Reference EXAFS

> Perform EXAFS refinement

> Compare results with element specific bond length

> ….

Check literature

Measure references for oxidation state with similar ligand set

Edge features (e.g. pre-edge-peak intensity)

1. Look at your spectra

> Raw XANES

> Raw EXAFS

> Reference XANES

> Reference EXAFS

> Perform EXAFS refinement

> Compare results with element specific bond length

> ….

Typical features occur for:

Imidazol

Heme group

Beat nodes

Amplitude of spectra

Apply a reasonable weighting to obtain a rather constant amplitude up to k= 12.5Å^{-1} or higher (for biological sample typically k^3)

If you have **no good data**, you might waste your time with an ab-initio analysis!

➔ Limit analysis to comparisons

2. Consider available information

> Raw XANES

> Raw EXAFS

> Reference XANES

> Reference EXAFS

> Perform EXAFS refinement

> Compare results with element specific bond length

> ….

Define starting model for first shell:

Typical element types in biological samples: O,N,S

Total number of ligands: 2,3,4,5,6

Ratio of ligands: e.g. $FeO_nN_mS_k$

Correlate Debye-Waller parameter for different shells (e.g. a[1,3] links Debye-Waller factor for 1.shell and 3shell)

Introduce central atom and potential ligands:

Set Atom1 Cu; Set Atom2 O; Set Atom3 N; Set Atom4 C

2. Consider available information

> 1. shell analysis

Introduce central atom and potential ligands:

Change Atom1 Cu; C Atom2 O; C Atom3 N; C Atom4 C

Define a starting model $ZnO_nN_mS_0$ (or better check a variety of models):

Change ns 2 (ns: Number of shells)

Change n1 2; C t1 O

C n2 2; c t2 N (t: element type)

Starting values for distance and disorder (Debye-Waller factor):

C r1 2.00; c a1 0.008

C r2 2.00; c a2 0.008

Define starting model for first shell:

Typical element types in biological samples: O,N,S

Total number of ligands: 2,3,4,5,6

Ratio of ligands: e.g. $ZnO_nN_mS_k$

Correlate Debye-Waller parameter for different shells (e.g. a[1,3] links Debye-Waller factor for 1.shell and 3shell)

NOTE on Debye-Waller factor:

Excurve uses 2 σ^2

Feff uses σ^2

2. Consider available information

> Raw XANES

> Raw EXAFS

> Reference XANES

> Reference EXAFS

> Perform EXAFS refinement

> Compare results with element specific bond length

> ….

You might want to include now further shells and multiple scattering (his).

Focus on constrained refinement.

Correlate Debye-Waller parameter (a[1,3])

Correlate distances e.g. of His and O (r[1,3])

Correlate data from different absorption edges

2. Consider available information

We can typically not distinguish O and N, but we can detect multiple scattering from imidazol rings!

Thus:

C t1 o; c r1 1.99

C r2 2.00; c t2 his

Sort

C ns 6 (O plus imidazole ring)
(ns: number od shells)

Set cor on (treat ring as rigid)

Ref

R[1,2];a[1,2];a[3,4];a[5,6];ef;=

;;;;;;;

You might want to include now further shells and multiple scattering (his).

Focus on constrained refinement.

Correlate Debye-Waller parameter (a[1,3])

Correlate distances e.g. of His and O (r[1,3])

Correlate data from different absorption edges

XAS Data Analysis Overview – an enzyme example

Step 1 - Collect all available information

studies on other enzymes with similar functionality

biochemical studies on enzyme

Step 2 – XANES analysis

possible coordination/geometry of the Fe site

Step 3 – EXAFS analysis

visual inspection of the spectrum and its FT

test different structural (physically reasonable!) models

validate the best refined model by checking its consistency with XANES and available knowledge

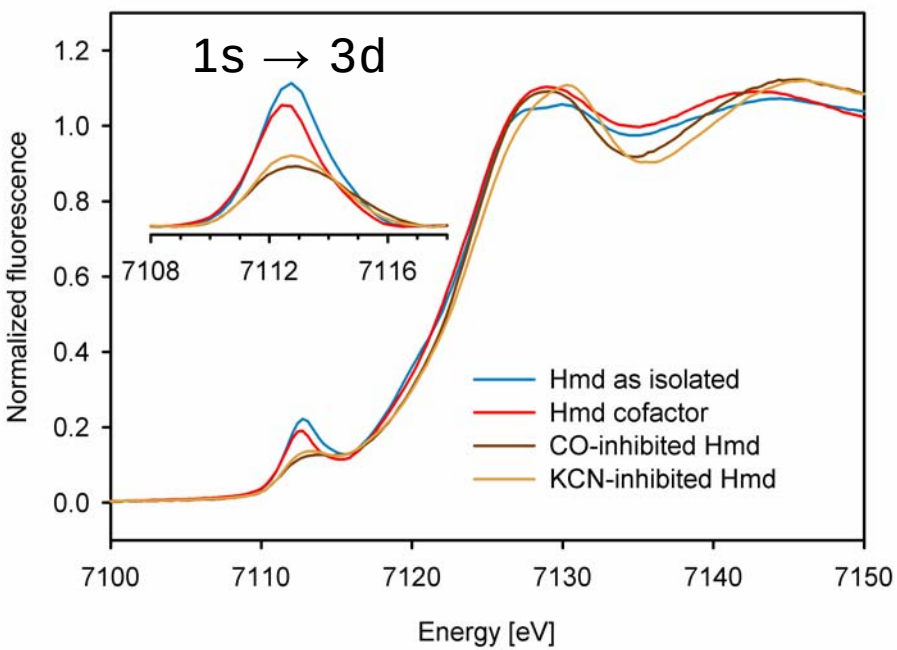
Step 1 – Background knowledge

> Here we leave out step 1, to keep the example more general

Step 2 – XANES analysis

Fe K-edge XAS spectra have a pre-edge feature originating from 1s → 3d transition

Generally pre-edge intensity increases with a decrease in coordination number and in metal site symmetry

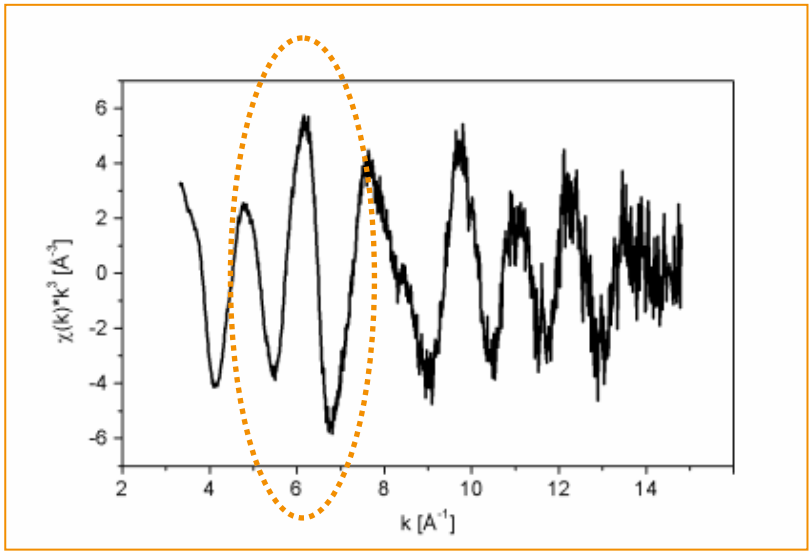


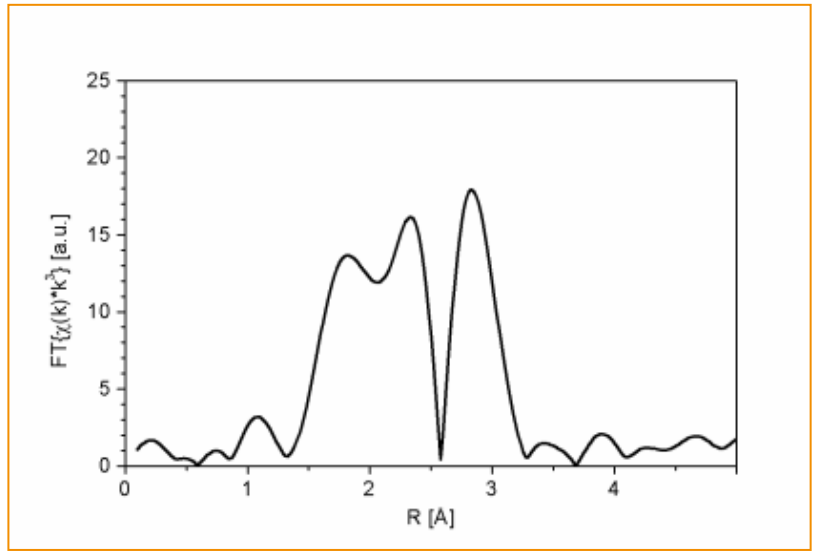
- Calculate the area under pre-edge peak

- Compare its value with those of Fe model complexes
- (no perfect models are available

➔ Either the number of ligands is small or / and heterogeneous

Step 3 – EXAFS analysis





EXAFS inspection – at ~6 Å^{-1} frequency of oscillations changes
➔ Lighter and heavier elements are present (combination of oxygen/nitrogen and sulfur?

FT inspection – three main contributions
~1.8 Å and ~2.9 Å peaks indicate the presence of CO ligands
~2.3 Å distance is the typical iron – sulfur bond length

How to run Excurve ?

- How to do the refinement?
  - ***R data*** (Read in data)
  - If existing read parameter file for starting values (***r par***)
  - You might want to use phases, calculated before (I never do)

    R phases
  - Alternative: just type ***r*** (read starts above sequence)

Run Excurve

Let's do it. Here comes the start-up script for a simple example Zn in a protein

Now we compare one structural model with the data:

```
> r e
> f17200.dat
> 1                 each data point
> 12                column combination (energy, extracted EXAFS signal)
> Zn K              element / absorption edge
> r par
> gcm_s3_his1.par
> calc pot
> c
> n
> 1
> ;;;;;;;;;         replaces returns
> calc phase;;;;;;;;;;;;;;          use defaults
> set ms units      multiple scattering defined in units (here one amino acid: his)
> set w 4           k3-weighting
> set cor on        move units as rigid bodies (constrained refinement)
> c dwmax 0.02      limit parameter space to reasonable values
> c dwmin 0.002     dw: Debye-Waller parameter
```

Run Excurve

> *c r1 2.00* change distance for first shell, here
> *c r2 2.29* change distance for second shell

> *ref* [iii] iii: indicator for step width [50 – 200] leave blank
> *ef;r1;r2;a[1-2];a[3-4];a[5-6];=*
> *c;p*
> correlation_matrix.cor
> *pr par*
> parameter_file.par
> *pr spec*
> exafs_exp_theory.spc
> 2;3;4;0
> *pr spec*
> FT_exp_theory.ft
> 3;5;6;0

Details:

The usual syntax for entering commands is:- **command keyword option** e.g. Change N1 2 or Set COR ON, where the commands are change or set, respective keywords are N1 or COR, and the options are 2 or ON.

The **command %<filename>** will execute a list of commands within the file <filename> and is very useful if you create startup files for reading in phaseshifts and setting up options for graphics, etc).

The character = is used as an escape character, to terminate or skip to the next part of a command.

The semicolon ; is used to enter several commands on one line e.g. C ATOM1 CU;C ATOM2 C;C ATOM3 N

The asterisk * followed by a Unix command performs that command (N.B. you cannot use wildcards in Unix commands sent from within Excurv98).

Parameters are changed using the command **Change** and viewed using **List** e.g. Change ATOM1 AG (or C ATOM1 AG), List Unit 1 (or L U 1) etc.

Part of the Script used to generate ASCII Output

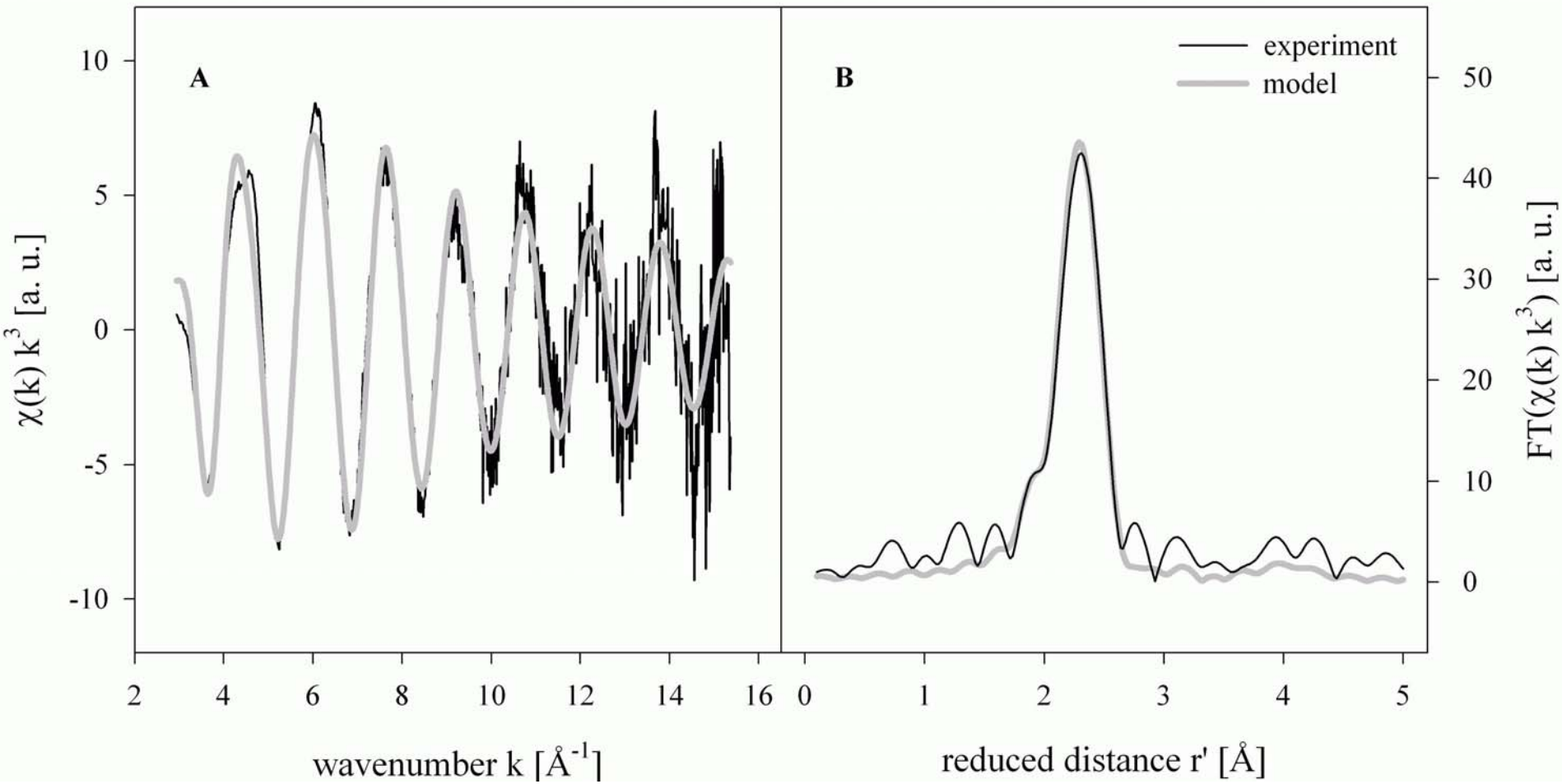
```
> ref
> ef;r1;r2;a[1-2];a[3-4];a[5-6];=
> c;p;c
> correlation.cor                 use names linked to sample and fit model e.g. Zn_gcm_model_s3_his1.cor

> pr par
> parameter.par                   use names linked to sample and fit model e.g. Zn_gcm_model_s3_his1.par

> pr spec
> exafs.spc                       use names linked to sample and fit model e.g. Zn_gcm_model_s3_his1.spc
> 2
> 3
> 4
> 0

> pr spec
> FT.ft                           use names linked to sample and fit model e.g. Zn_gcm_model_s3_his1.ft
> 3
> 5
> 6
> 0
```

Single data set shown in example



Cohen, S.X., Moulin, M., Schilling, O., Meyer-Klaucke, W., Schreiber, J., Wegner, M., and Müller, C.W., *The GCM domain is a Zn-coordinating DNA-binding domain.* FEBS Lett, 2002. 528(1-3): p. 95-100.

What happens if we use other structural models?

- We could compare the fit index for each model to do a meta-analysis
- This strategy is found in ABRA

Table 1
Comparison of published Zn-EXAFS data with *ABRA*'s results (1σ errors on the last digit extracted by meta-analysis are given in parentheses).

S, His and O are the zinc donor groups and their numbers are given for the different samples and structural models. The sum of His and O donor groups given by the meta-analysis is abbreviated as Low-*Z* ligands.

	Number of ligands							
	Published results			Best model by *ABRA*			Average of good models (meta-analysis)	
	S	His	O	S	His	O	S	Low-*Z*
Zinc finger proteins								
HPV E7	4.0			4.0			4.0 (4)	0.2 (2)
GCM	3.0	1.0		3.0	1.0		3.0 (0)	1.2 (2)
ZnF-UBP	2.67	1.33		2.67	1.00	0.33	2.8 (3)	1.2 (3)
Catalytic active sites								
Ec ZiPD		2.0	2.5		3.0	1.5	0.0 (0)	4.2 (2)
At Glx2-2		2.5	2.5		4.0	0.5	0.0 (0)	4.2 (2)
Bc bla	0.5	3.0	1.0		3.0	1.0	0.1 (2)	4.0 (4)
Model compounds								
Bis(acetato)bis(imidazol)zinc(II)		2.0	2.0		2.0	2.0	0.0 (0)	4.2 (2)
Tetrakis(imidazole)zinc(II)-perchlorate		4.0			3.5	0.5	0.0 (0)	4.3 (2)

What can we learn from the analysis of other structural models?

> Meta-analysis

	Average of good models (meta-analysis)	
	S	Low-Z
Zinc finger proteins		
HPV E7	4.0 (4)	0.2 (2)
GCM	3.0 (0)	1.2 (2)
ZnF-UBP	2.8 (3)	1.2 (3)
Catalytic active sites		
Ec ZiPD	0.0 (0)	4.2 (2)
At Glx2-2	0.0 (0)	4.2 (2)
Bc bla	0.1 (2)	4.0 (4)
Model compounds		
Bis(acetato)bis(imidazol)zinc(II)	0.0 (0)	4.2 (2)
Tetrakis(imidazole)zinc(II)-perchlorate	0.0 (0)	4.3 (2)

Wellenreuther, G., Parthasarathy, V., and Meyer-Klaucke, W.,
Towards a Black-Box for Biological EXAFS Data Analysis – II. Automatic BioXAS Refinement & Analysis (ABRA).
J Synchrotron Radiat, 2010. 17(1): p. 25-35.

EXAFS data analysis: The standard way

BioXAS data

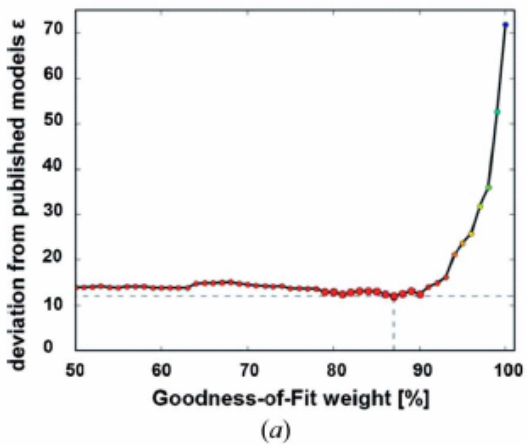
User supplied info

Pool of models

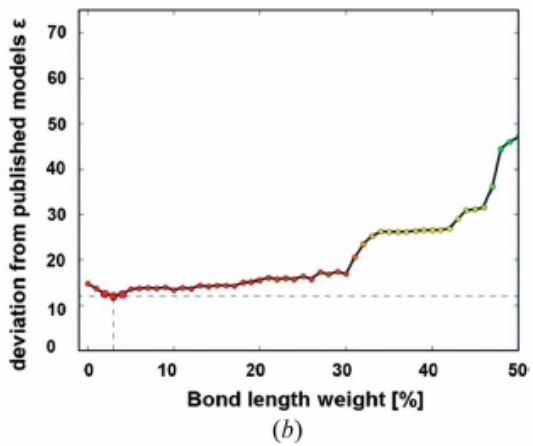
D L E X C U R V

Fit-Index
R-Factor
χ^2

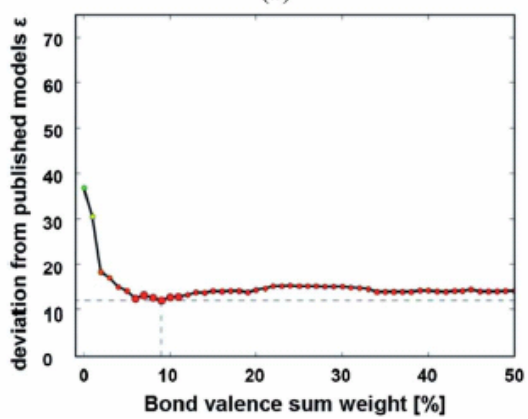
«unreasonable» model

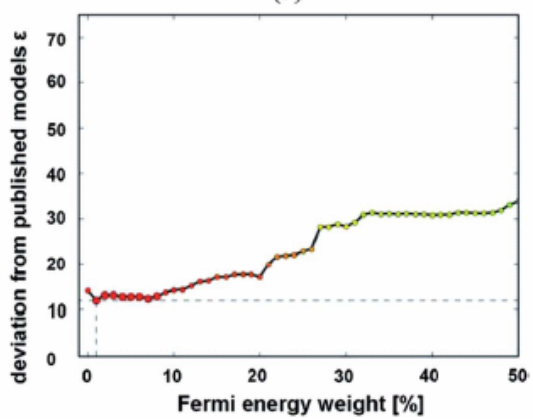


(a)

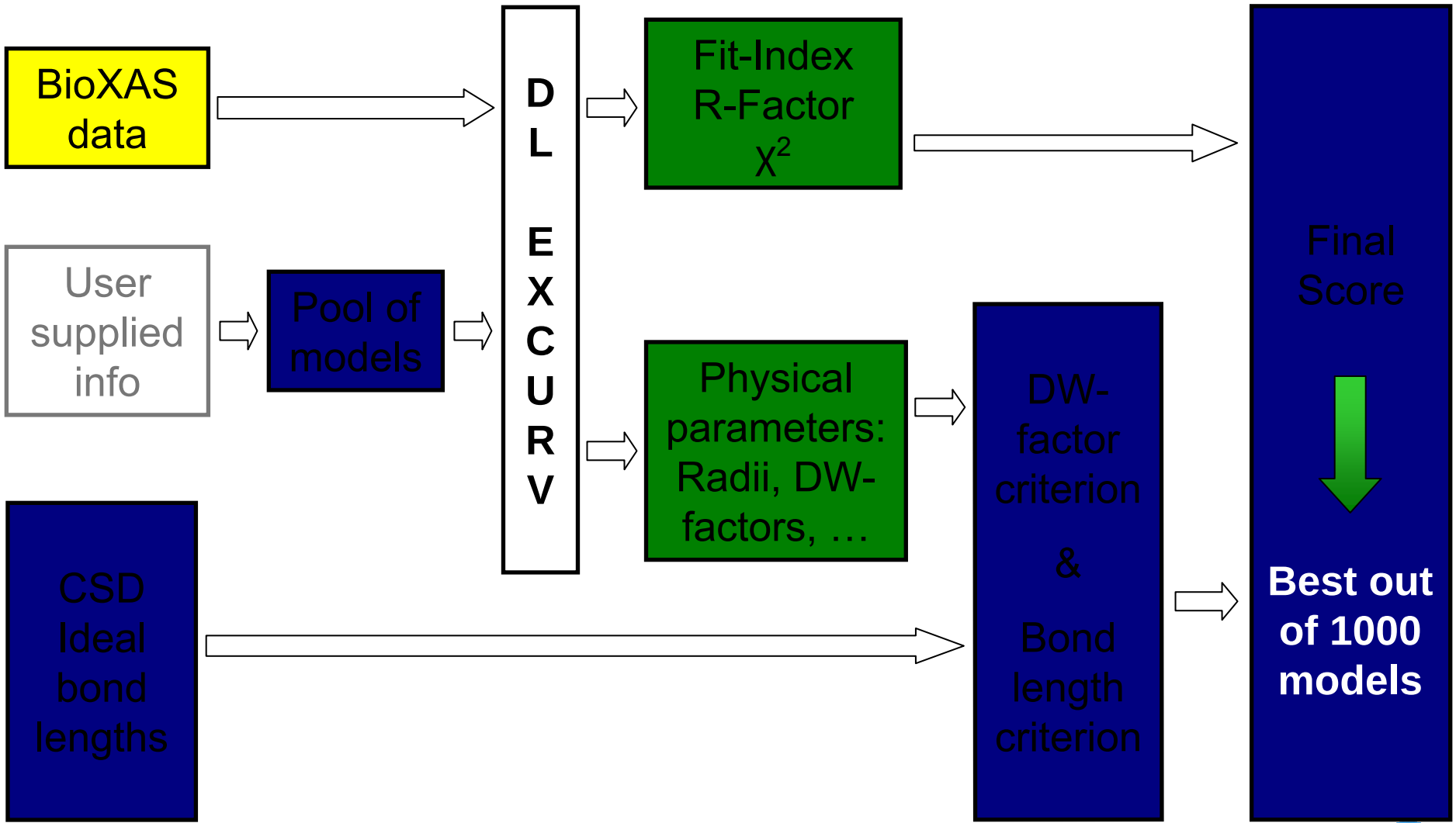


(b)





ABRA: Automated biological EXAFS refinement and analysis



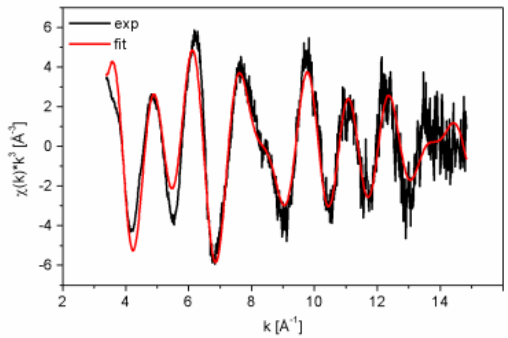
ABRA: G. Wellenreuther & W. Meyer-Klaucke, JSR, 2010

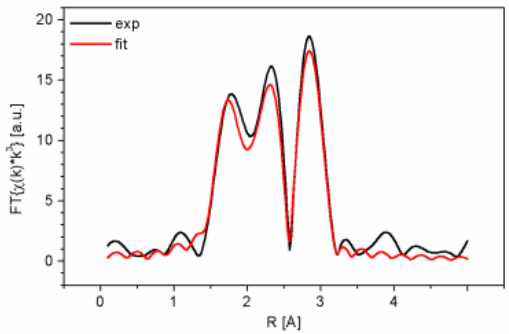
2nd example:

- This example will be part of my presentation **MS.42.2**
- Here, comes the start of our analysis:

Step 3 – Ab initio EXAFS analysis (2nd example)

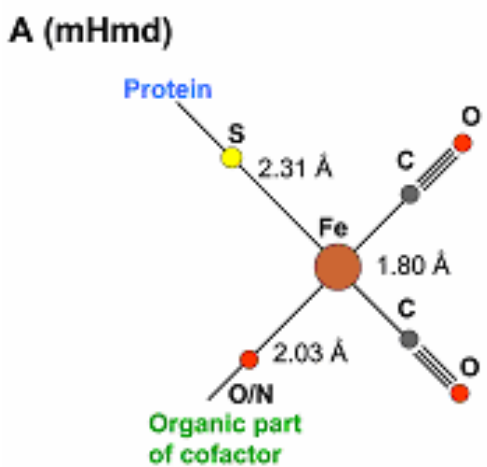
Several tested models: $Fe(CO)_xO_yS_z$, x = 1,2,3 ; y = 0,1,2 ; z = 0,1,2

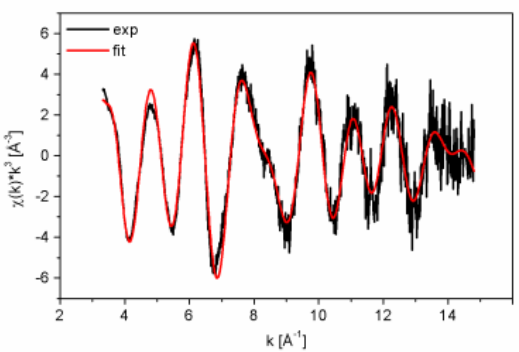


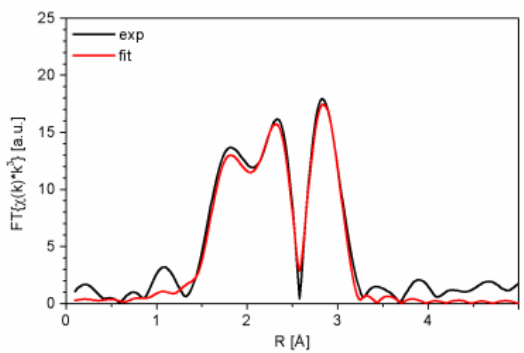


→ $Fe(CO)_2S_1$

- additional light scatterer is required
- multiple scattering within Fe – C – O unit must







→ $Fe(CO)_2O_1S_1$

Reanalysis of MX and EXAFS data for wild-type Hmd assuming C176A binding mode of cofactor

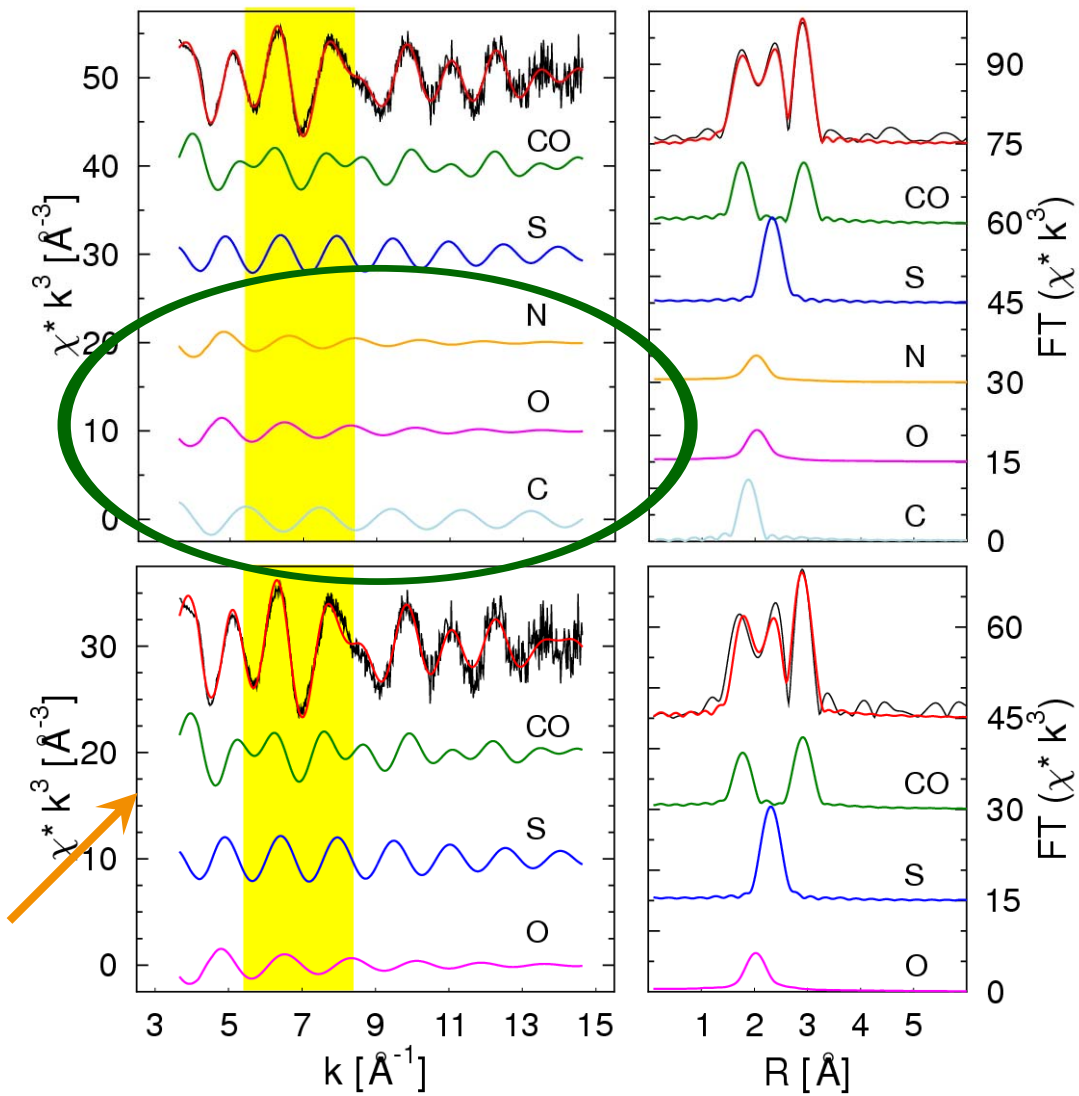
Note destructive interference of backscattering contributions:

C at 1.85 Å and
O/N at 2.04 Å

This model fits the 1.75 Å crystal structure as well.

In the EXAFS we see a slight improvement, but the model has more free parameters.

Initial model



Hiromoto T, Ataka K, Pilak O, Vogt S, Stagni MS, Meyer-Klaucke W, Warkentin E, Thauer RK, Shima S, Ermler U., FEBS Lett. 2009

Salomone-Stagni, M., Vogt, S., Shima, S., and Meyer-Klaucke, W., 2009. 190(012197): p. 1-6.

If you want to learn more about our philosophy in data analysis, read our paper on ABRA

- My Postdoc's motivation was to put my knowledge and habits into software:
- Thus you find several references on criteria helping you to judge the quality of your refinement in the corresponding publications
- Most important:
  - typical metal-ligand distances

    Harding, M. M. (1999). Acta Cryst. D55, 1432–1443.
    Harding, M. M. (2000). Acta Cryst. D56, 857–867.
    Harding, M. M. (2001). Acta Cryst. D57, 401–411.
    Harding, M. M. (2002). Acta Cryst. D58, 872–874.
    Harding, M. M. (2004). Acta Cryst. D60, 849–859.
    Harding, M. M. (2006). Acta Cryst. D62, 678–682.
  - Bond valance sum analysis (keep in mind you obtain information on the oxidation state of the absorber from the edge position)
  - Both concepts do at present not consider the spin-state. Take care!
  - Energy shift (in Excurve called EF) should be similar for similar samples (If not you might have assumed the wrong neighbor atom (turn metals into gold))
- ABRA
  - http://scripts.iucr.org/cgi-bin/paper?S0909049509040576

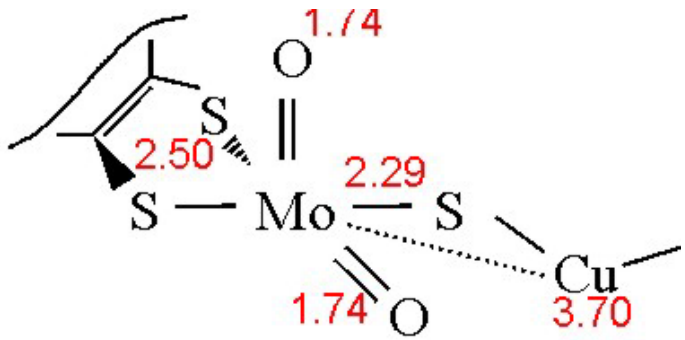
Many thanks for your attention

Take home message:

> Running Excurve is straight forward

Errors in EXAFS

> coordination number for first shell: ~ 20%



> distances: ~ 0.02Å for first shell

> element type: about ±1

Increase significance of results by including all available information

Details:

(ii) To start the program and read in experiment files

At Unix prompt enter excurv98 (N.B. Unix is case sensitive and lower case must be used), and press <return> again when prompted.

ENTER COMMAND:

Read Exp (this reads in the experimental spectrum. The command Read on its own is used to input experiment, a parameter file and phaseshift files. Read PAR is used to input just parameters, and Read PH for phaseshifts only)

Filename for Experiment 1 ?

r8447.exb (this is the name of the file which has had the background removed in exback, ecabs, exbrook, etc.)

Experimental spectrum in r8447.exb

Point frequency [1] ?

1 (this reads in every point of the spectrum. A point frequency of 2 would read in every second point. It is very unlikely you will ever want to use anything but 1)

Details:

Column combination [12] ?

32 (this means that the program will read in energy (in eV) from column 3 of the background subtracted file, and Chi from column 2. Data background subtracted in exback or exspline require the combination 32, ecabs & exbrook output require 12)

Edge ? [CU K]

MO K (The edge that is being analysed. Can be K, L3, L2, L1, e.g. Rb L3 or W L1)

Sequence-number in polarisation set [0]

0 (this is used when reading in a series of spectra of single crystals at different angles. The default 0 should be used for all other cases)

Number of clusters for this experiment [1]

1 (used when defining different structural clusters around the central atom for multiple scattering calculations. If you are using only single scattering, or there is only one type of cluster in the sample use the default of 1)

0.00000 0.09701 42.19988 Background subtracted spectrum

435 Hartrees Absorption Ev Edgenor pre

Number of points read: 435

(iii) Calculating potentials and phaseshifts

ENTER COMMAND:

Change ATOM1 MO; Change ATOM2 CL; C ATOM3 MO (Define an atom type for every element in the sample. You should include your central atom twice as in the potential and scattering calculations an excited atom is not the same as the normal atom). Atom1 must be the central (absorber) atom. The order of the others doesn't matter, but if you save phaseshifts files and parameter files you must read the phaseshift files back into Excurv98 in the same order.

ATOM3 changed from: (0) to: MO (42)

ENTER COMMAND:

CALculate POTentials

(iii) Calculating potentials and phaseshifts

ENTER COMMAND:

Change ATOM1 MO; Change ATOM2 CL; C ATOM3 MO (Define an atom type for every element in the sample. You should include your central atom twice as in the potential and scattering calculations an excited atom is not the same as the normal atom). Atom1 must be the central (absorber) atom. The order of the others doesn't matter, but if you save phaseshifts files and parameter files you must read the phaseshift files back into Excurv98 in the same order.

ATOM3 changed from: (0) to: MO (42)

ENTER COMMAND:

CALculate POTentials

(Hedin-Lunqvist (exchange) Von Barth (GS))

Calculating potentials for all atoms:

Enter : G (Graphics), T (Terminal Output), M (Charge densities) or C (Continue)

http://www.diamond.ac.uk/dms/Beamlines/I18/I18_Mar07/basicexcurve.pdf

CALculate POTentials

(Hedin-Lunqvist (exchange) Von Barth (GS))

Calculating potentials for all atoms:

Enter : G (Graphics), T (Terminal Output), M (Charge densities) or C (Continue)

C

Atom: 1 (MO). Enter neighbouring atom [2 (CL)]

(The general advice is to choose an atom type which most resembles the surrounding of the atom or the lightest neighbouring atom. The program default is shown in square brackets. Press **<return>** to accept the default, or enter the required scatterer.)

Select Code For Exited Atom [1] :

No Correction (0)

1S Core Hole (K-edge) (relaxed approximation) (1)

1S Core Hole (K-edge) (Z+1 approximation) (-1)

1 (Generally use Z+1 for 1st row transition metals and lighter and relaxed for heavier central atoms. This makes very little difference anyway, but try to be consistent.)

http://www.diamond.ac.uk/dms/Beamlines/I18/I18_Mar07/basicexcurve.pdf

1 (Generally use Z+1 for 1st row transition metals and lighter and relaxed for heavier central atoms. This makes very little difference anyway, but try to be consistent.)

MO (CL) Rho0: .4385 Efermi: -3.811 V0: -17.716 Electrons: 41.807 (Z= 42)

The program then asks for atoms surrounding each of the scatterers etc...

The interstitial potential V0 should be the same to within 2 eV for all the elements. If this is not the case a target V0 can be set

ENTER COMMAND:

Set CONstant V (The program takes an average of the potentials calculated, and sets this as a target potential)

Recalculate potentials by using the **CAL POT** routine again The muffin tin radii of the atoms will be refined to give the target potential V0 for all the atoms.

MO (CL) Rho0: .5193 Efermi: -3.751 V0: -19.317 Electrons: 41.807 (Z= 42)

etc…..

The calculated potential can now be used to calculate phaseshifts for the central atom and all the scattering atoms.

http://www.diamond.ac.uk/dms/Beamlines/I18/I18_Mar07/basicexcurve.pdf

CALculate PHaseshifts

Calculating phaseshifts for all atoms:

Do You Require the Atomic Absorption [no] ?

(only required for XANES calculations - generally take the default of NO)

Enter core width (FWHM eV) [5.743]

(again, use default generally)

Rho0: .5193 Efermi: -3.751 V0: -19.317 LMAX: 25 CW: 5.743

Rho0: .4376 Efermi: -5.439 V0: -19.326 LMAX: 25 CW: 5.743

Rho0: .4935 Efermi: -4.275 V0: -19.320 LMAX: 25 CW: 5.743

Having calculated a set of phaseshifts, these can be saved to files which can be read into Excurv98 at a subsequent session. This saves having to recalculate every time and also ensures that the same set of phaseshifts is used in all your analyses.

(iv) Theory, Graphics and FT Parameters Sub Menus

Set Weight 4 (this enables k3-weighting)

Do not play with theory options as a non expert. Keep in mind that any changes made in theory options have to be communicated in publications. **If not, you are cheating**. In case you think changes are required, you need to explain them.