IC|CI
CNR Istituto di Cristallografia

EXPO&more Workshop

30 September – 3 October 2019

RootProf

from theory to practice

RootProf website
http://www.ba.ic.cnr.it/softwareic/rootprof/

Belviso Benny Danilo
Bari, 3rd October 2019

Outlines

- Command file and GUI
- Unsupervised quantitative analysis (XAS data)
- Qualitative analysis (SAXS data)
- Supervised quantitative analysis (FT-IR data)

RootProf website
http://www.ba.ic.cnr.it/softwareic/rootprof/

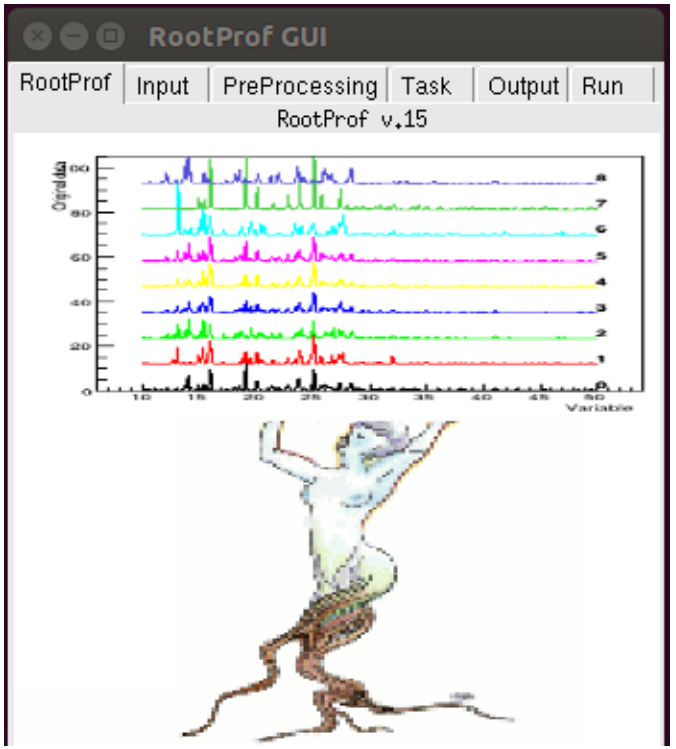
Command file and GUI

Command file

Task
Output
Preproc
Input file

```
whichanalysis 0
figpaper 1
preprocess 1 2 0 1
range 5000 6500
file kinetic_01.extract
file kinetic_02.extract
file kinetic_03.extract
file kinetic04.extract
```

GUI



http://www.ba.ic.cnr.it/softwareic/rootprof/

Command file

Keywords

Task
Output
Preproc
Input file

```
whichanalysis 0
figpaper 1
preprocess 1 2 0 1
range 5000 6500
file kinetic_01.extract
file kinetic_02.extract
file kinetic_03.extract
file kinetic04.extract
```

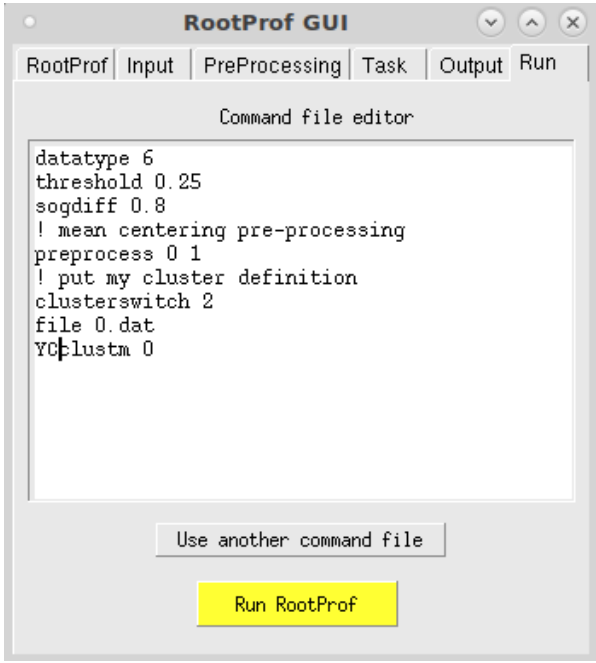
Keywords (general)			
whichanalysis	verbose	preprocess	skiplines
kolmog	savefig	backscale	file
	figpaper	shiftfactor	range
	startfromone	varbin	datatype
	writefiles	wgen	Skipdata

Keywords (X-ray data)			
	writehkl		standardrx
			ignoresigma
			nophase

http://www.ba.ic.cnr.it/softwareic/rootprof/

Command file
Command editing

- Commands must be edited in an input file
- Free format
- Commands are case insensitive
- Comments start with "!"
- Each keyword is followed by a number related to the specific operation to perform



Command file

Examples

whichanalysis

This command is the core of RootProf. It defines which type of analysis has to be performed.

0 First Sight analysis
1 Qualitative analysis
2 Correlation analysis
3 Quantitative analysis
4 Supervised Quantitative
5 Covariance analysis
6 Size analysis
7 MED analysis
8 Only clustering
9 Generation of profiles
10 Crystallinity analysis

= **whichanalysis 1**

range

It defines the range of the independent variable

Range/ranges

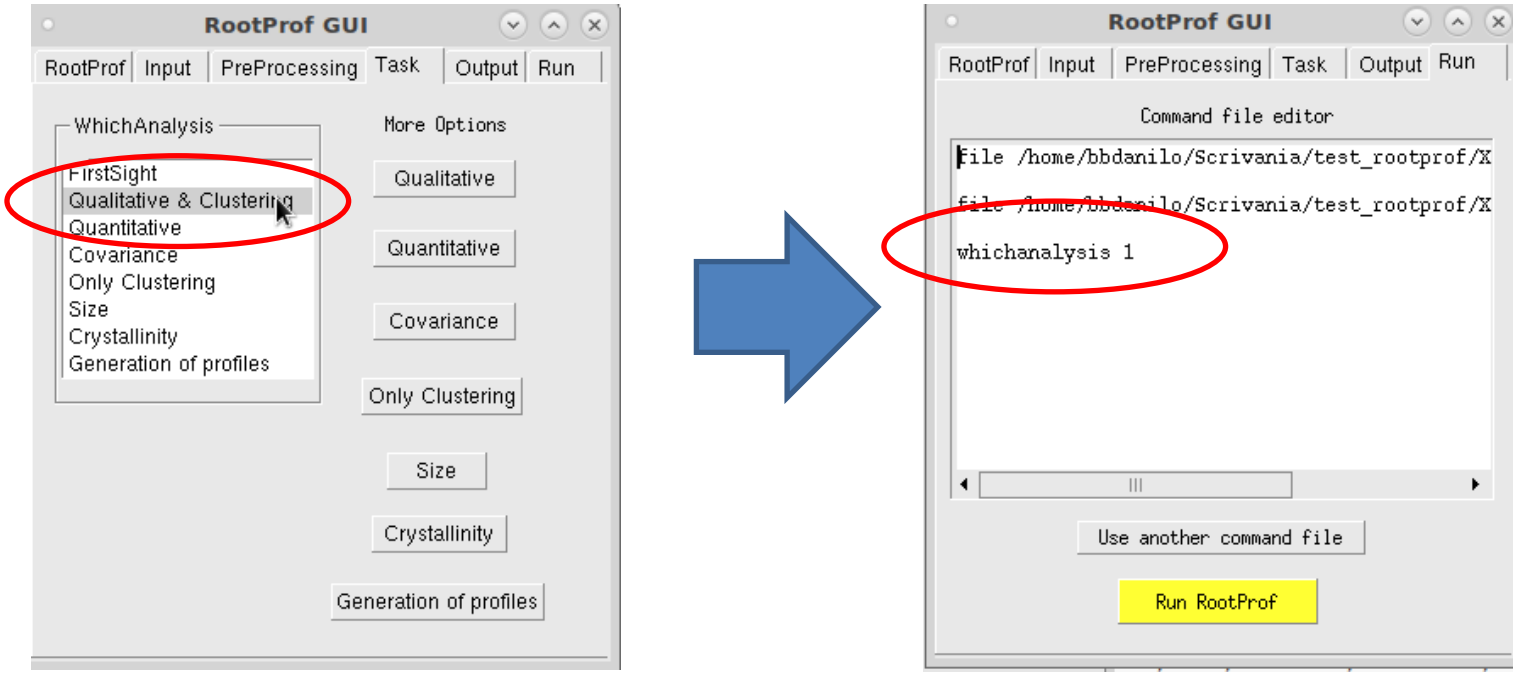
= **range 0 10 50 70**

http://www.ba.ic.cnr.it/softwareic/rootprof/

Command file

Command log file by GUI

While you are learning RootProf, you may want to keep a record of all the commands you give in a plain text log-file that you can read and edit.

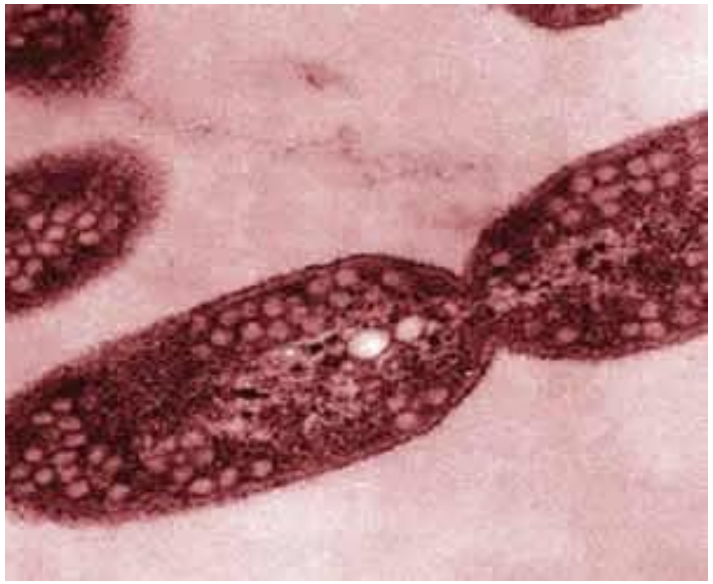


All your commands, typed or clicked in the GUI, will be automatically recorded in the input command file prepared by RootProf.

http://www.ba.ic.cnr.it/softwareic/rootprof/

XAS data analysis by RootProf

Cr reduction by *R. sphaeroides*



Rhodobacter sphaeroides is an **incredibly efficient photosynthetic bacterium** that grows in deep lakes and stagnant waters.

***R. sphaeroides* is able to adapt to a wide range of environmental conditions and can be exploited in bioremediation**

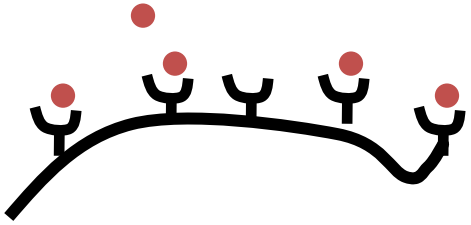
Which is the kinetic of chromium reduction by *R. sphaeroides*?

Tutorial 6 at
https://users.ba.cnr.it//ic/crisrc25/RootProf/TutorialPage.htm

XAS data analysis by RootProf

Sample preparation and XAS measurements

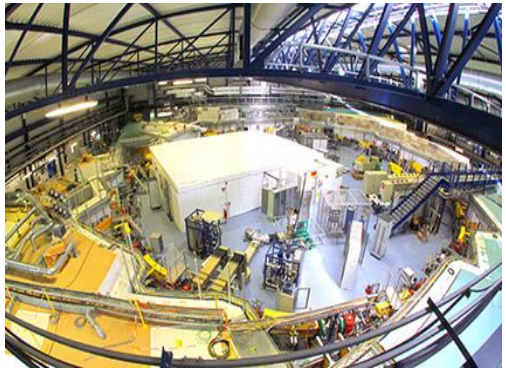
Whole cell



Cell envelope fraction

Soluble fraction

- Dataset: unidimensional XAS spectra (energy, I/I0)
- Data collected by using flow cell
- Measurements at different time to monitor chromate reduction kinetic (Cr(VI) -> Cr(III)) by bacteria cells.



ANKA

KIT Karlsruhe Institute of Technology

Belviso B. D. et al. (2013) Biometals, 26, 693–703

XAS data analysis by RootProf

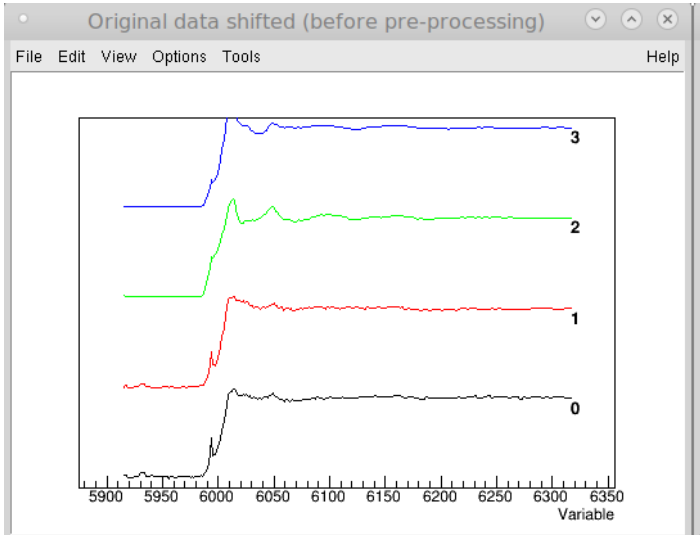
First sight analysis

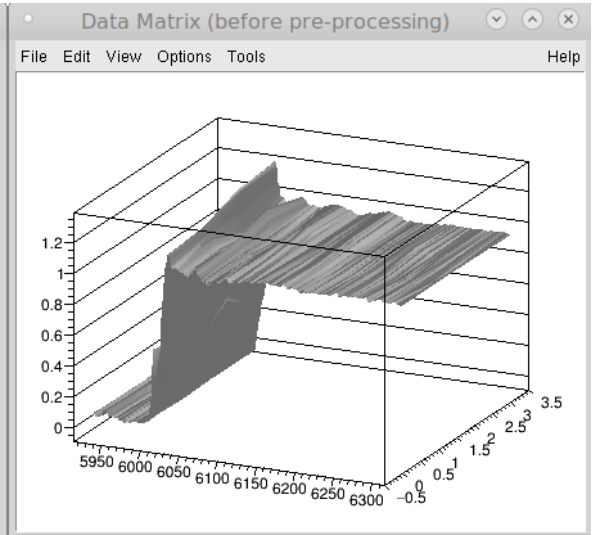
AIM OF THE ANALYSIS

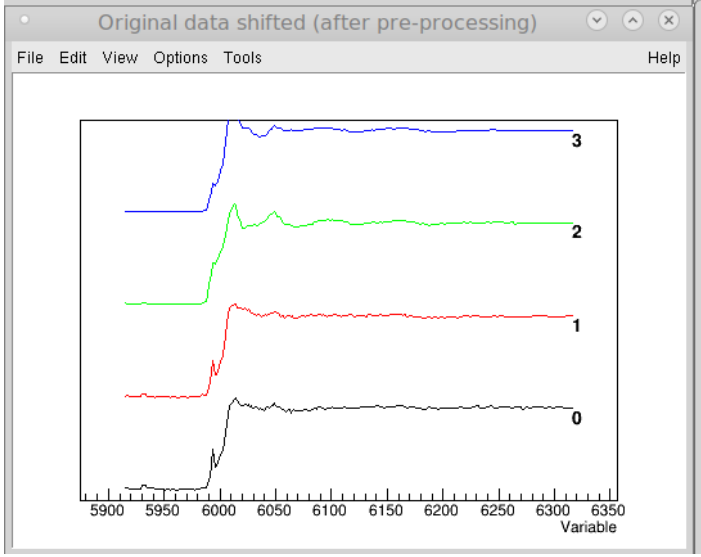
Viewing and comparison of data

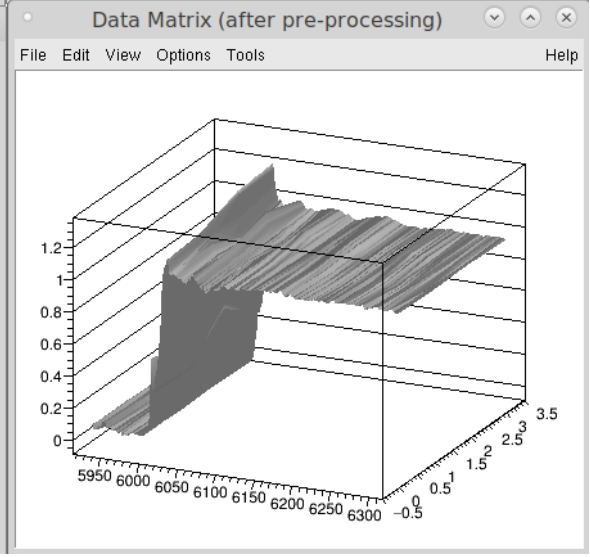
COMMAND FILE

whichanalysis 0
figpaper 1
range 5000 6500
file kinetic_01.extract
file kinetic_02.extract
file kinetic_03.extract
file kinetic04.extract









XAS data analysis by RootProf

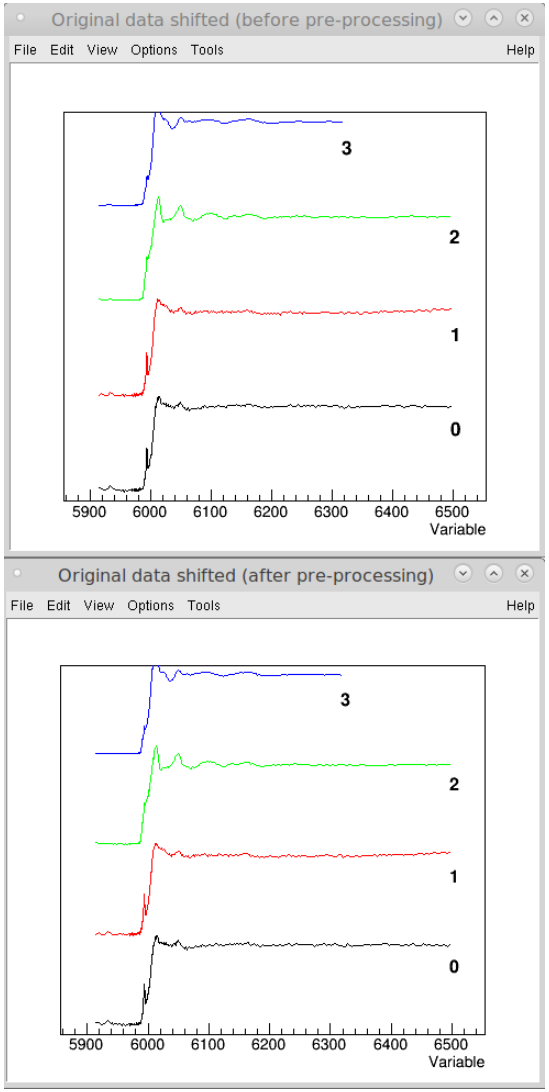
First sight analysis

AIM OF THE ANALYSIS

Viewing and comparison of data

COMMAND FILE

whichanalysis 0
figpaper 1
varbin 1
range 5000 6500
file kinetic_01.extract
file kinetic_02.extract
file kinetic_03.extract
file kinetic04.extract



XAS data analysis by RootProf

First sight analysis

Command: varbin 1

RootProf treats the spectra with variable binning. Spectra keep their original ranges.

varbin 1

```
 Reading input files:
 ---------------------
Sample 0 -> file kinetic_01.extract
        Found 340  points
Sample 1 -> file kinetic_02.extract
        Found 340  points
….
```

Profiles have different binning

varbin 0

```
 Reading input files:
 ---------------------
Sample 0 -> file kinetic_01.extract
        Found 340  points
Sample 1 -> file kinetic_02.extract
        Found 340  points
….
```

Profiles have different binning

Profiles interpolated to have equal binning

Transforming input profiles to equal binning
Chosen range: [5914.76 6317.77]
Number of points: 295

XAS data analysis by RootProf

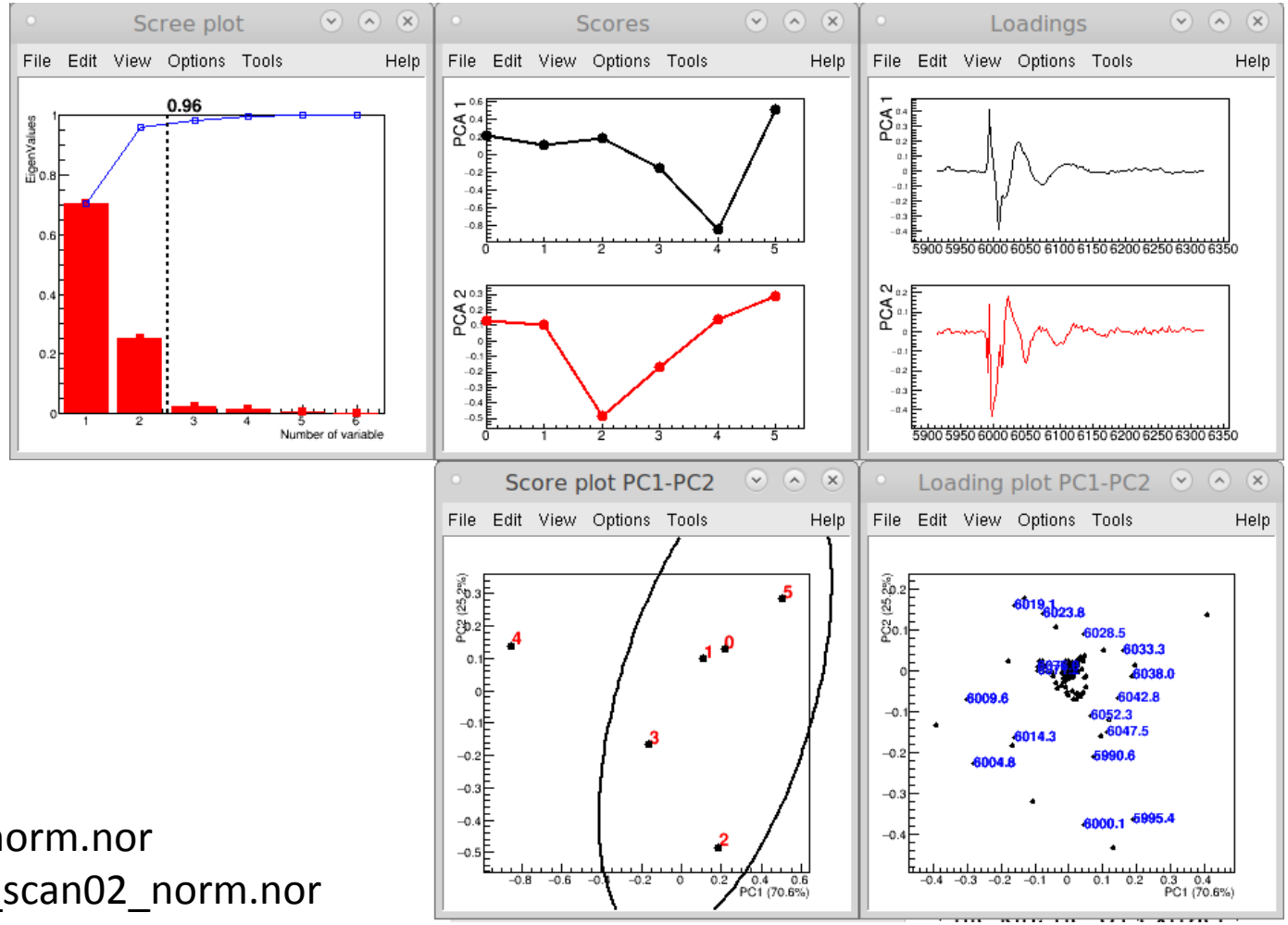
Clustering by PCA

AIM OF THE ANALYSIS

Grouping data by PCA

COMMAND FILE

whichanalysis 1
figpaper 1
threshold 0.96
range 5000 6500
file kinetic_01.extract
file kinetic_02.extract
file kinetic_03.extract
file kinetic04.extract
file cr3_2_exafs_s02_norm.nor
file Cr6_exafs_2_flow_scan02_norm.nor



XAS data analysis by RootProf

Clustering by PCA

========== Dendrogram ==========

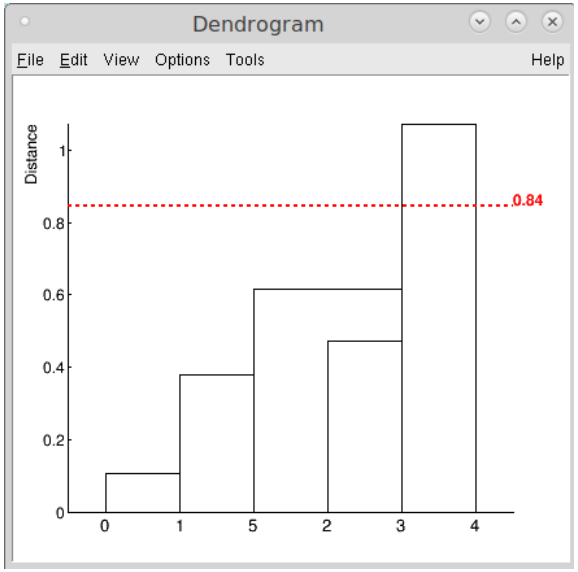
Step	Dist	Sample 1	Sample 2
5	1.0710	0	4
4	0.6171	0	2
3	0.4718	2	3
2	0.3805	0	5
1	0.1094	0	1

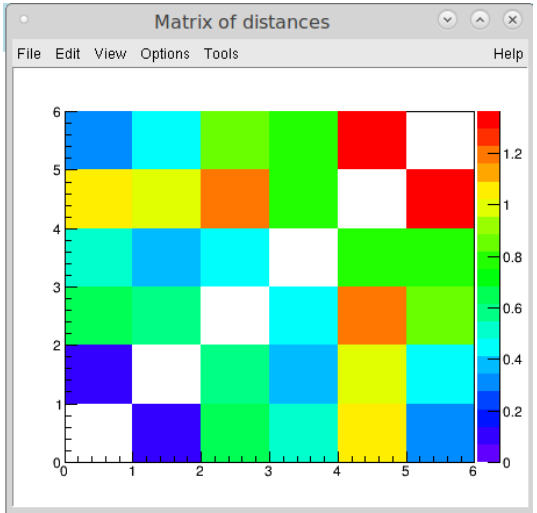
==========

Normalized Cluster threshold: 0.200000 (0.763999)

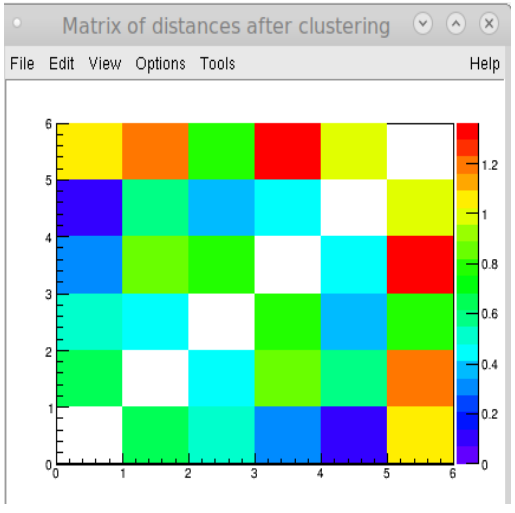
Normalized Cluster threshold redefined: (0.200000) 0.763999

Cluster Threshold 0.844





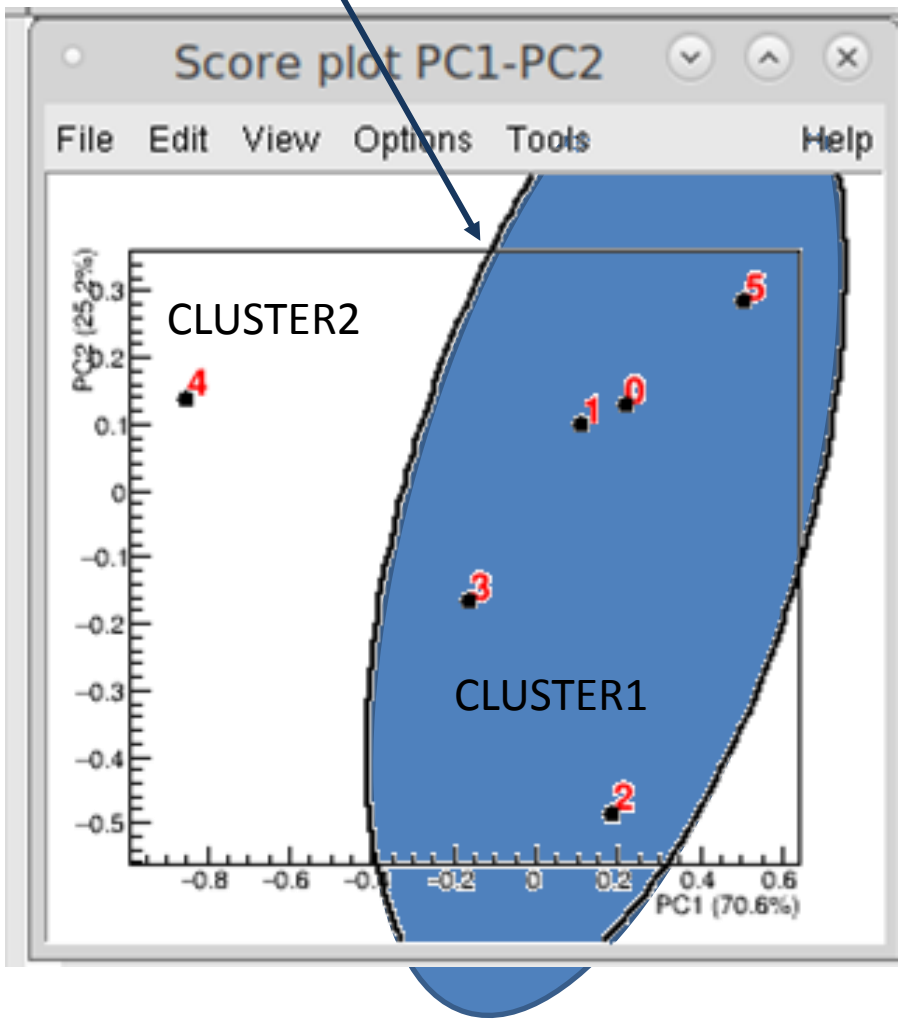
clustering

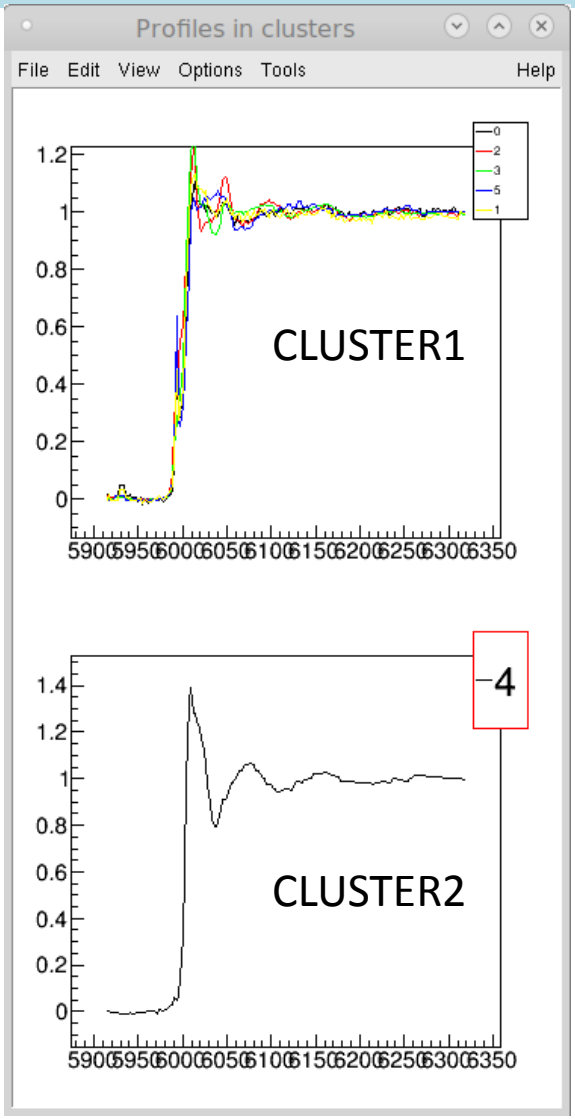


XAS data analysis by RootProf

Clustering by PCA

95% confidence level ellipse





XAS data analysis by RootProf

Quantitative analysis

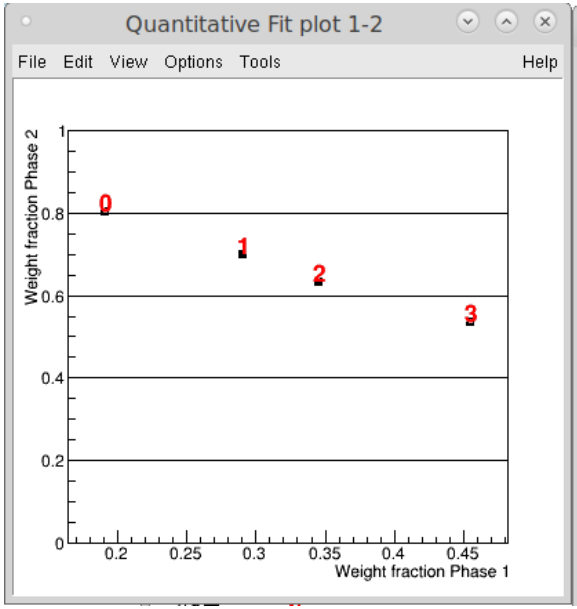
AIM OF THE ANALYSIS

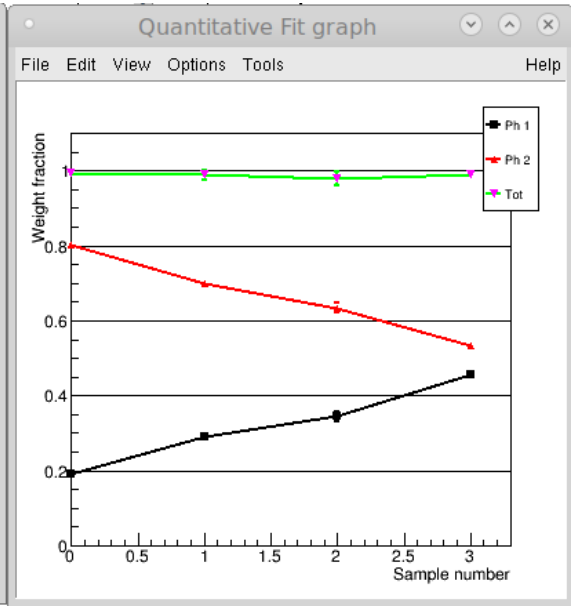
Quantitative analysis by the MultiFit approach.

$$y_{mod} = \sum_{j=1}^{n} v_j P_j(i + e_j) + y_o$$

COMMAND FILE

whichanalysis 3
varbin 1
figpaper 1
fitmodel 1
range 5000 6500
file kinetic_01.extract
file kinetic_02.extract
file kinetic_03.extract
file kinetic04.extract
file cr3_2_exafs_s02_norm.nor
purephase
file Cr6_exafs_2_flow_scan02_norm.nor
purephase





Phase 1 = CrO_4^{2-}
Phase 2 = Cr^{3+}

XAS data analysis by RootProf

Quantitative analysis

$$y_{mod} = \sum_{j=1}^{n} v_j P_j(i + e_j) + y_o$$

FIT RESULTS:

Profile 0 : kinetic_01.extract

Chi-Square=1.16e-01, Reduced Chi-Square=3.47e-04, NDF=334

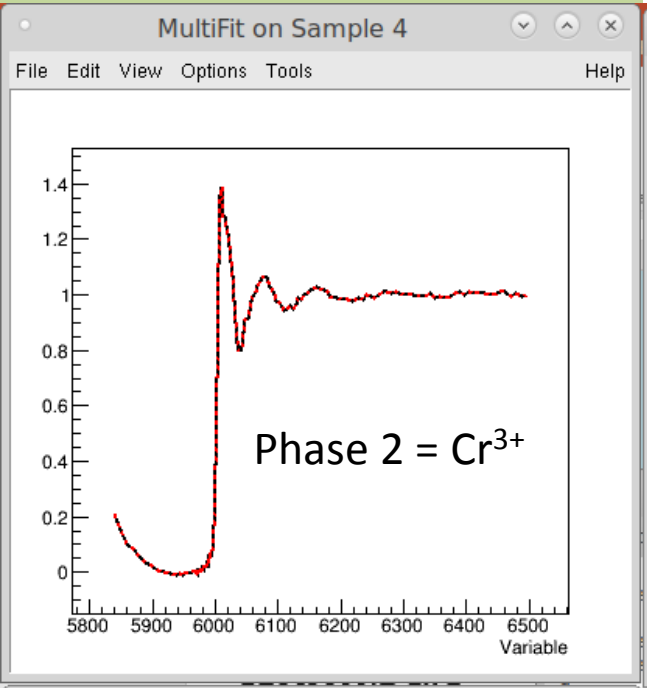
Weight fraction Phase 1 0.191 +- 0.006

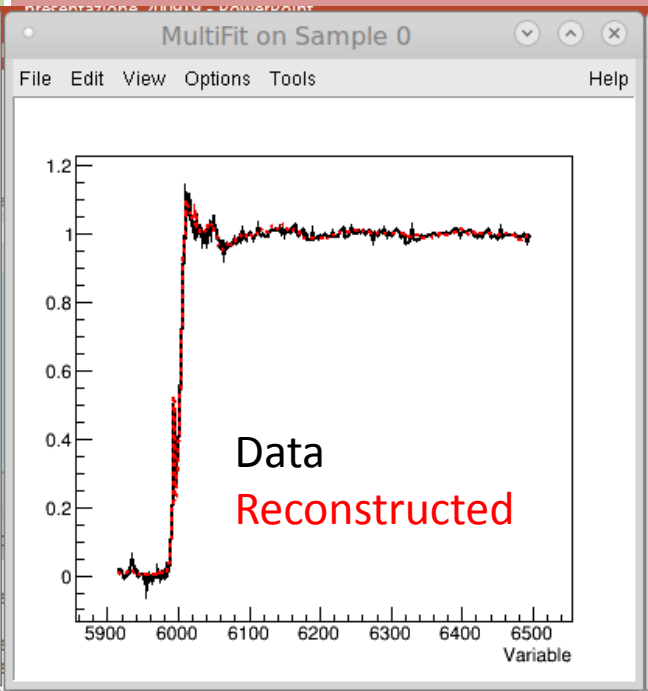
Weight fraction Phase 2 0.802 +- 0.006

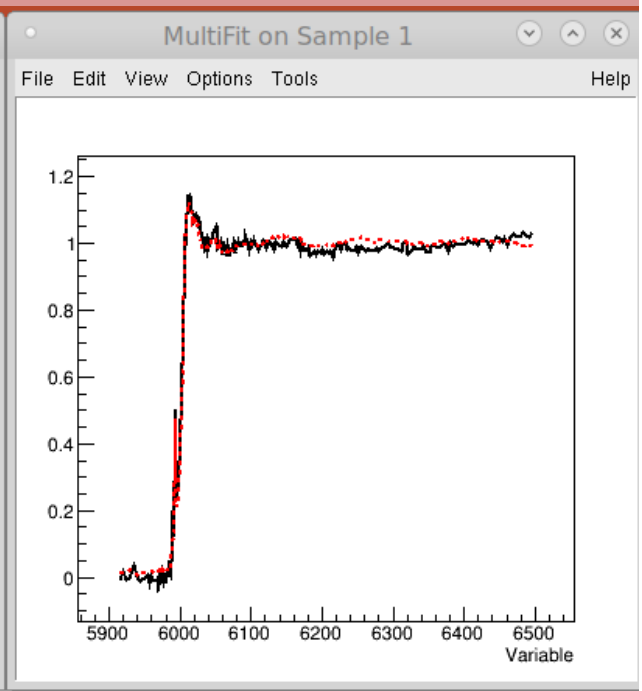
Total weight fraction 0.993 +- 0.008

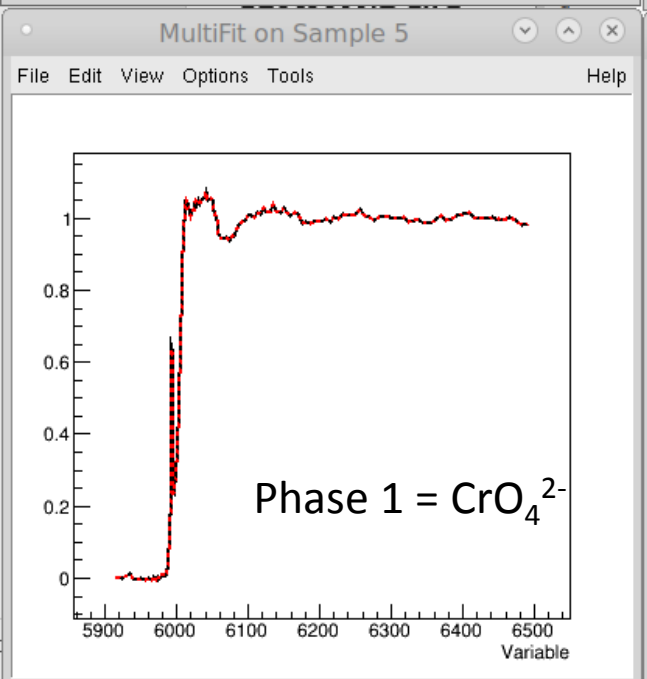
Pure phases

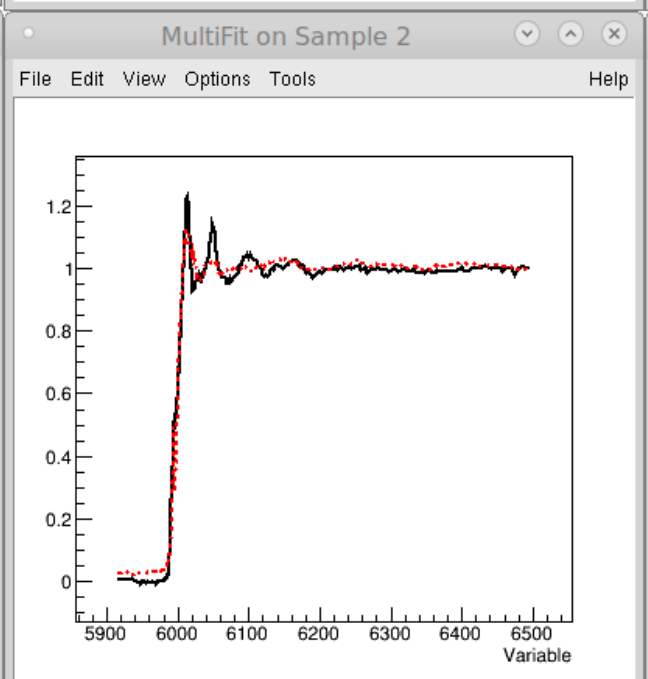
Samples

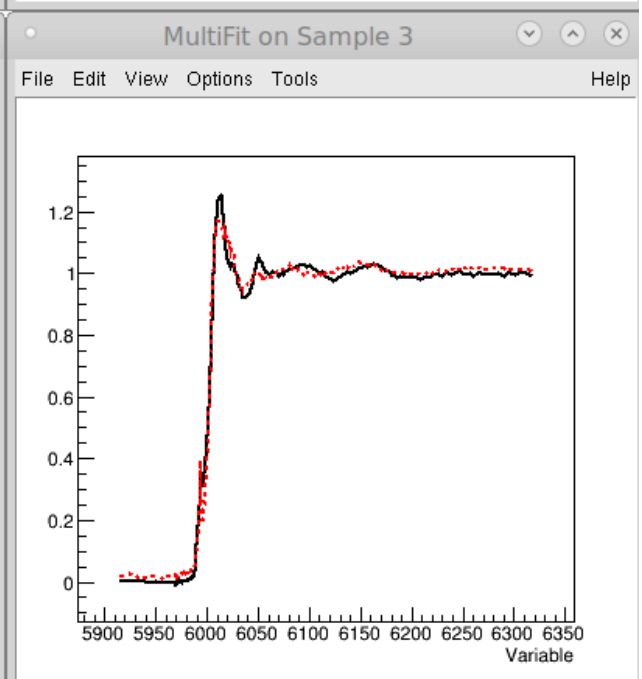






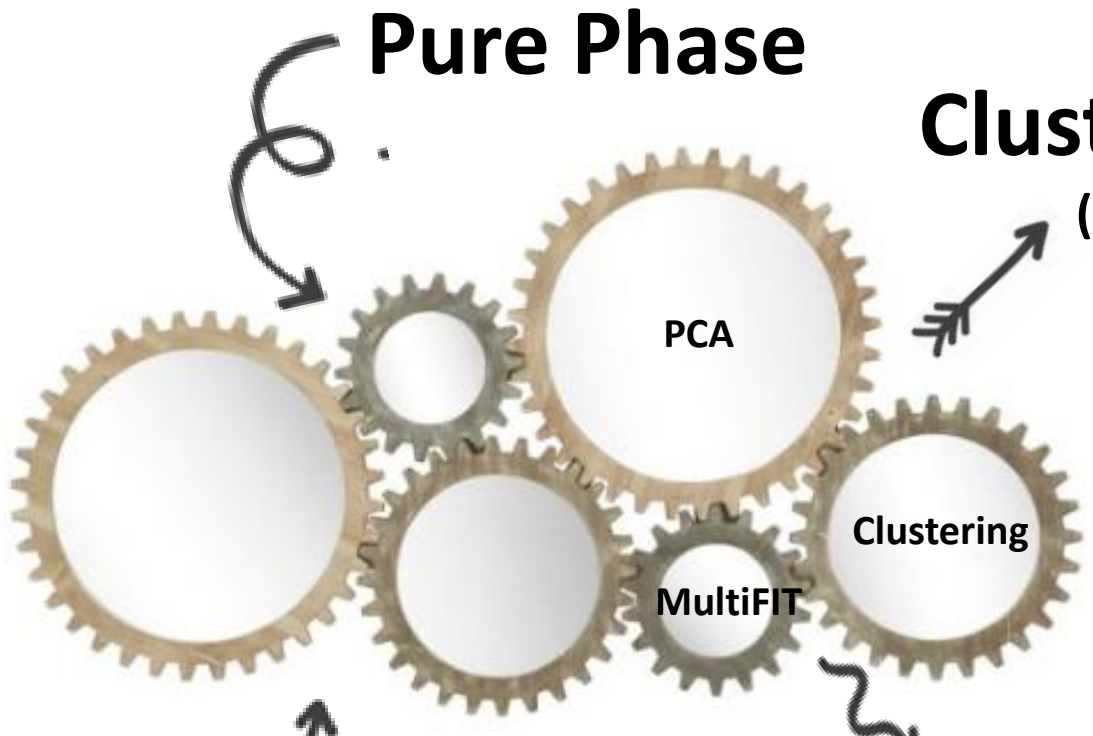






XAS data analysis by RootProf

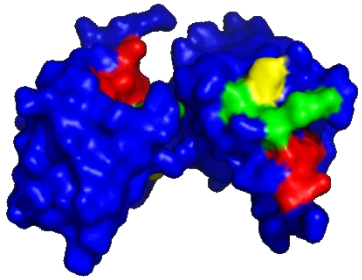
Which is the kinetic of chromium reduction by *R. sphaeroides*?



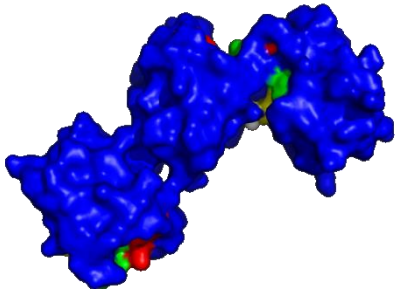
SAXS data analysis by RootProf

Ubiquitin aggregation by metal ions

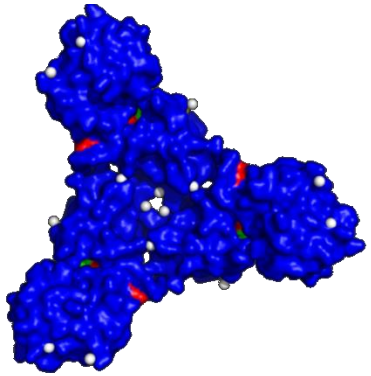
Ubiquitin without zinc



Ubiquitin:zinc=1:20

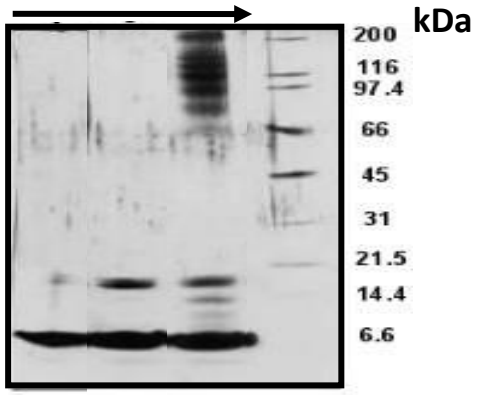


Ubiquitin:zinc=1:70



7 days, T=37°C, [Ub] = 350μM

eq Zn(II)= 0 20 70



In the crystal form, ubiquitin oligomerizes in the presence of metal ions.

What does it happen in solution?

Tutorial 7 at https://users.ba.cnr.it//ic/crisrc25/RootProf/TutorialPage.htm

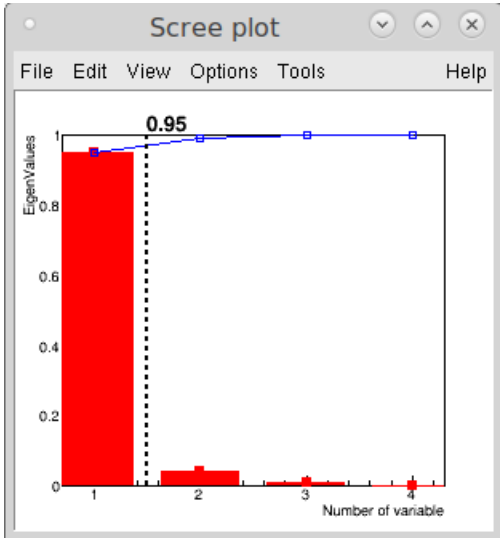
F. Arnesano, B.D. Belviso, R. Caliandro, G. Falini, S. Fermani, G. Natile, D. Siliqi Chemistry, 17,1569-1578

SAXS data analysis by RootProf

Clustering by PCA

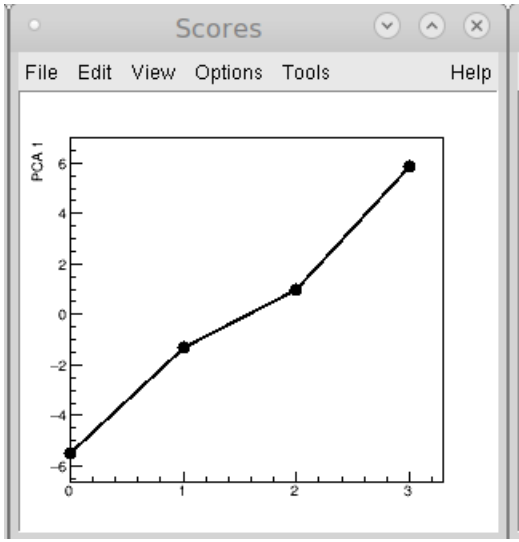
AIM OF THE ANALYSIS

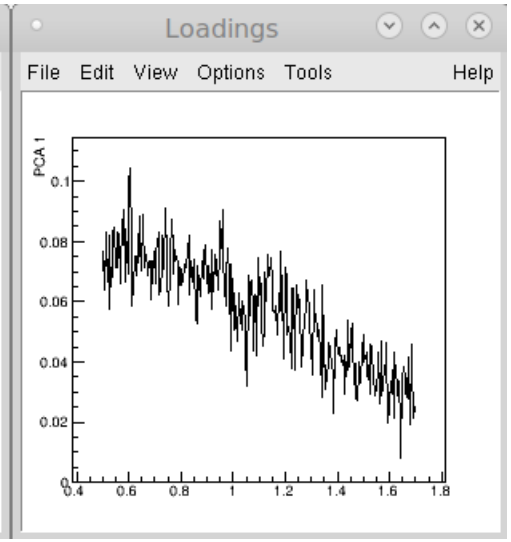
Grouping data by PCA



COMMAND FILE

whichanalysis 1
figpaper 1
range 0.5 1.7
file Ub_24h_S4mgml.dat
file Ub_Cu_24h_S4mgml.dat
file Ub_Zn25_24h_S4mgml.dat
file Ub_Zn200_24h_S4mgml.dat
clusterswitch 0





SAXS data analysis by RootProf

Data reconstruction by PCA

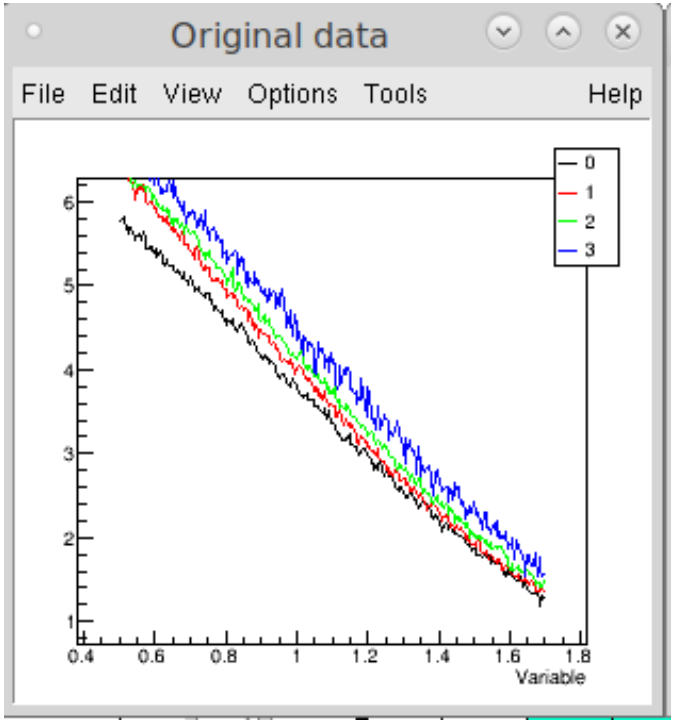
Data matrix | Loadings | Scores

$$X_{s,v} L_{v,n} = S_{s,n}$$

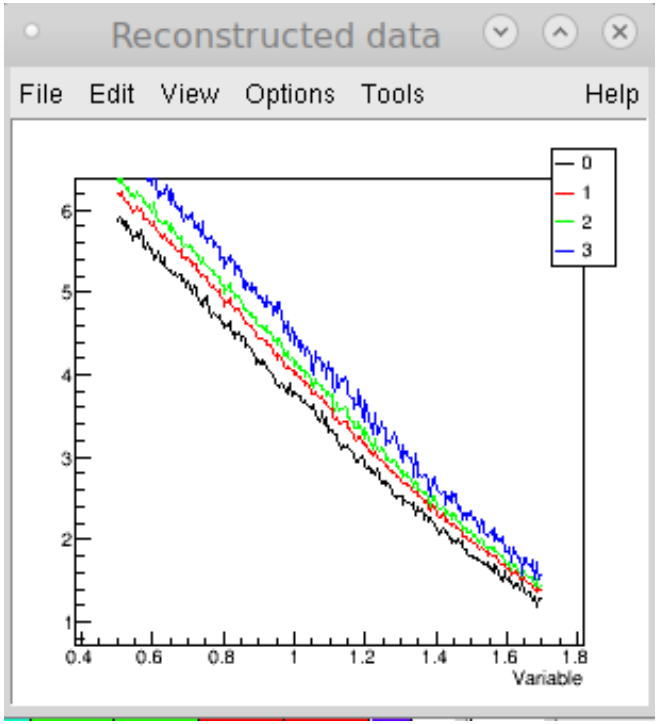
s number of samples
v number of variables
n number of PCs

$$X_{s,v} L_{v,g} = S_{s,g}$$

s number of samples
v number of variables
g selected PCs



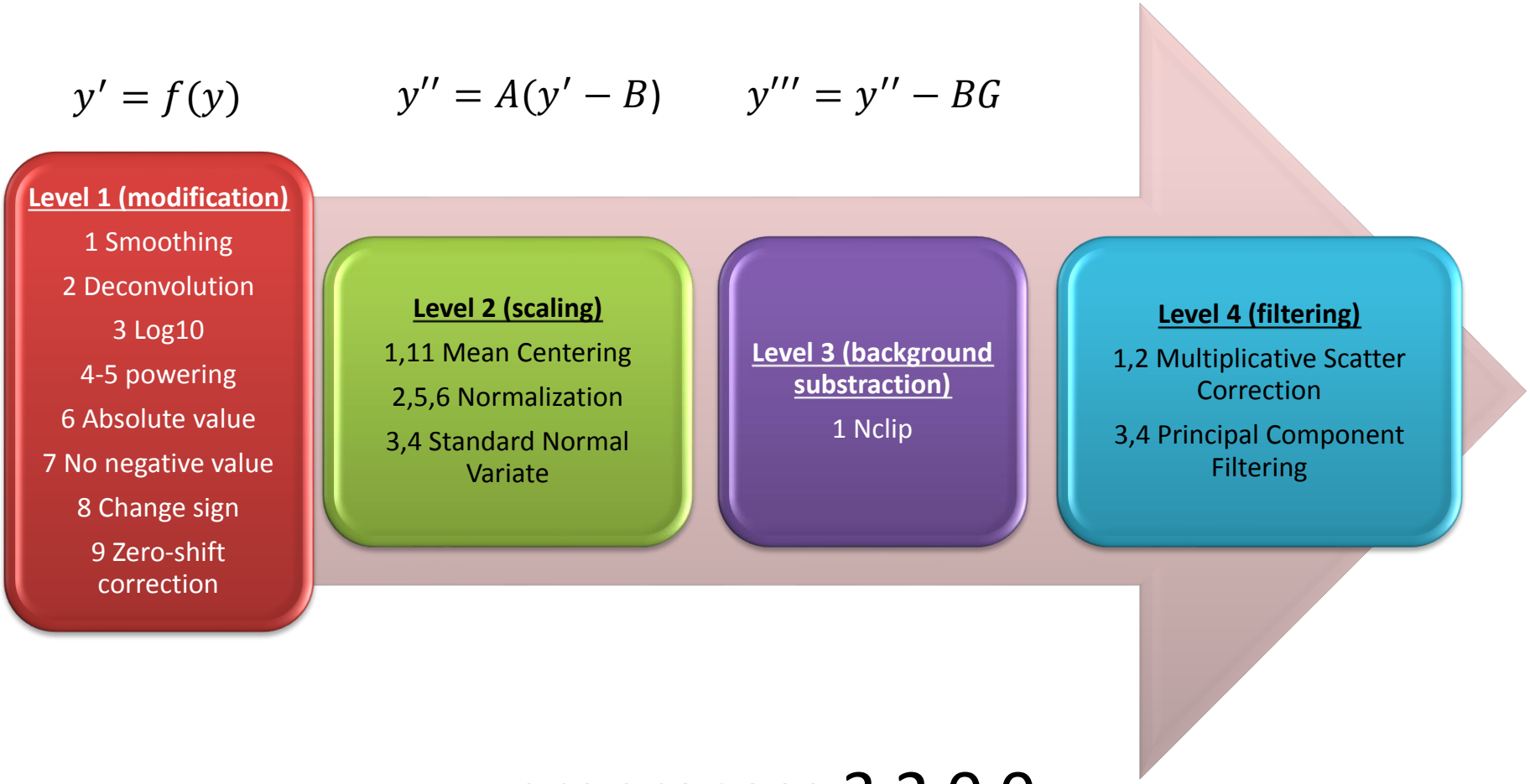
Reconstruction by using PCAs explaining highest variability



SAXS data analysis by RootProf

Preprocessing

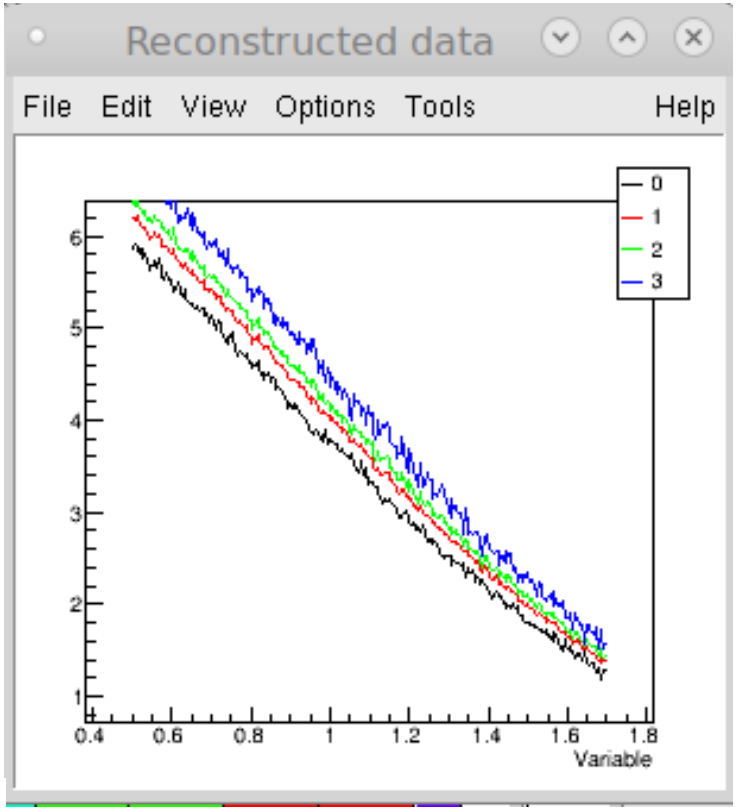
$y' = f(y)$ $y'' = A(y' - B)$ $y''' = y'' - BG$



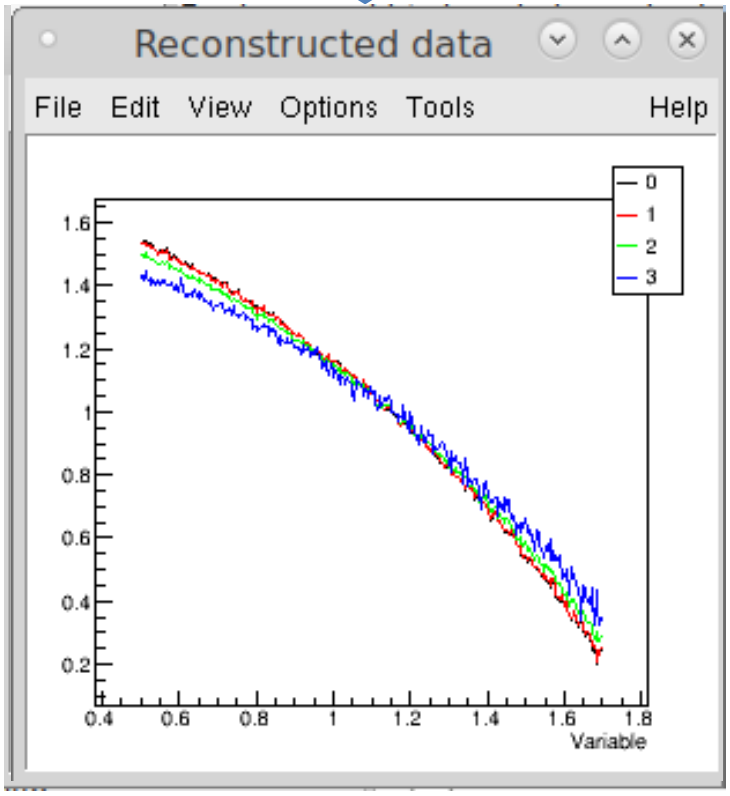
preprocess 3 2 0 0

SAXS data analysis by RootProf

Clustering by PCA: preprocessing



Preprocess 3 2 0 0
Level 1 = log (y)
Level 2 = Normalization

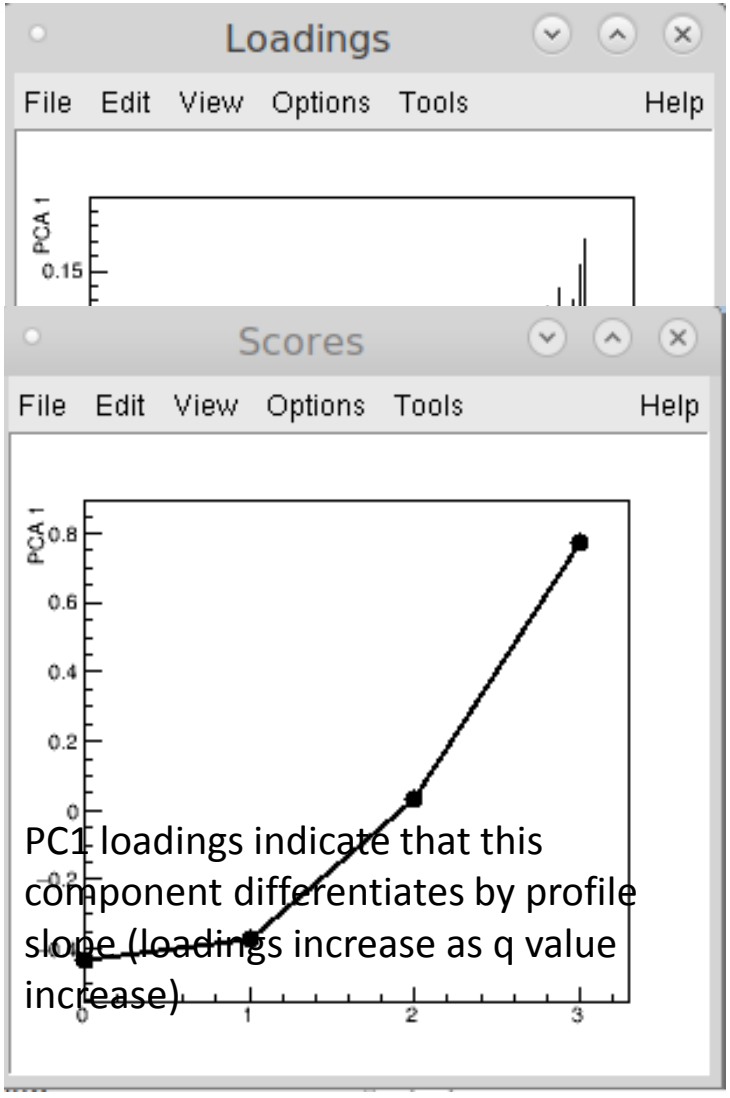


Normalization

Profiles are rescaled so that their area is equal to 1

SAXS data analysis by RootProf

Clustering by PCA: preprocessing



PC1 scores show that

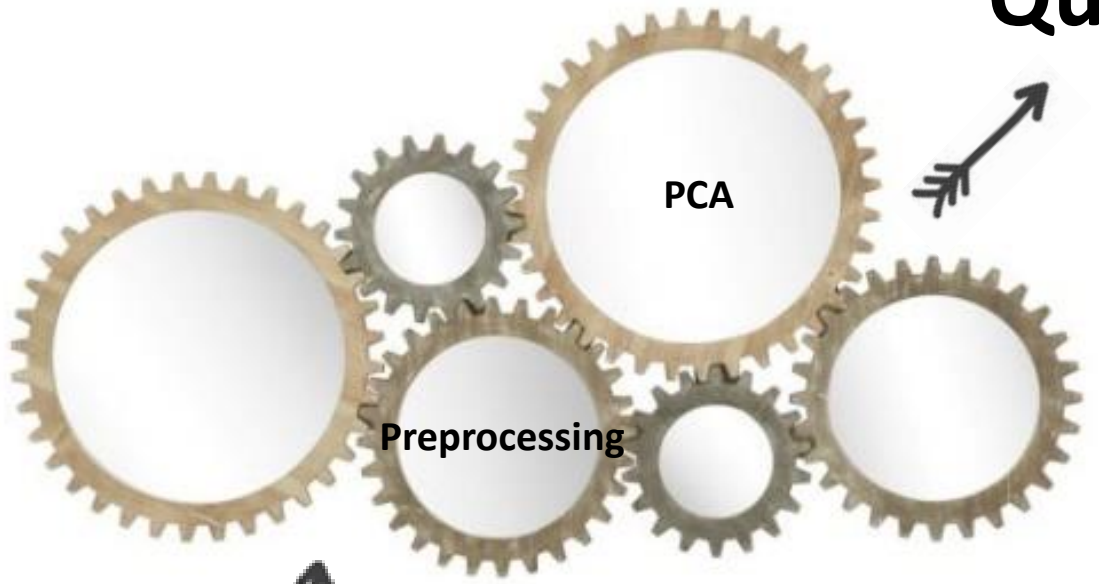
- $[Cu^{2+}]$<25mM has negligible effect on Ub oligomerization
- $[Zn^{2+}]$>25mM affects Ub oligomerization and the effect is proportional to Zn concentration

PC1 loadings indicate that this component differentiates by profile slope (loadings increase as q value increase)

SAXS data analysis by RootProf

Clustering by PCA: preprocessing

What does it happen in solution regarding Ub oligomerization?



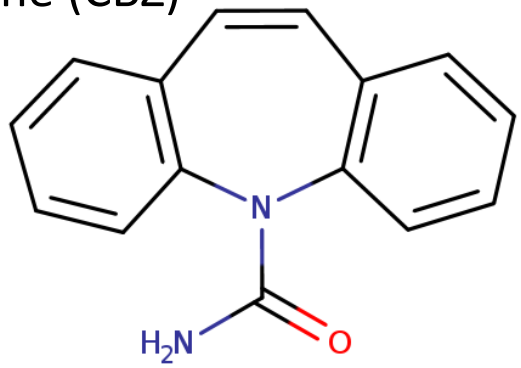
Qualitative Analysis
(<10sec)

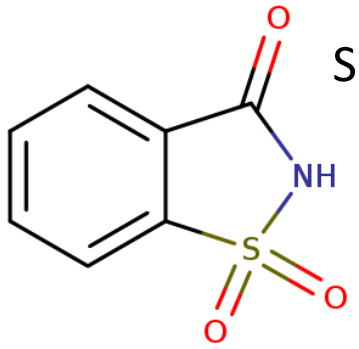
Data

FT-IR data analysis by RootProf

API quantification in drug formulation

Carbamazepine (CBZ)





Saccharine (SAC)

Drugs contain APIs and excipients. Their quantification is important for drug formulation.

Which is the relative amount of CBZ and SAC in a drug formulation?

Tutorial 8 at
https://users.ba.cnr.it//ic/crisrc25/RootProf/TutorialPage.htm

FT-IR data analysis by RootProf

Supervised Quantitative Analysis

AIM OF THE ANALYSIS

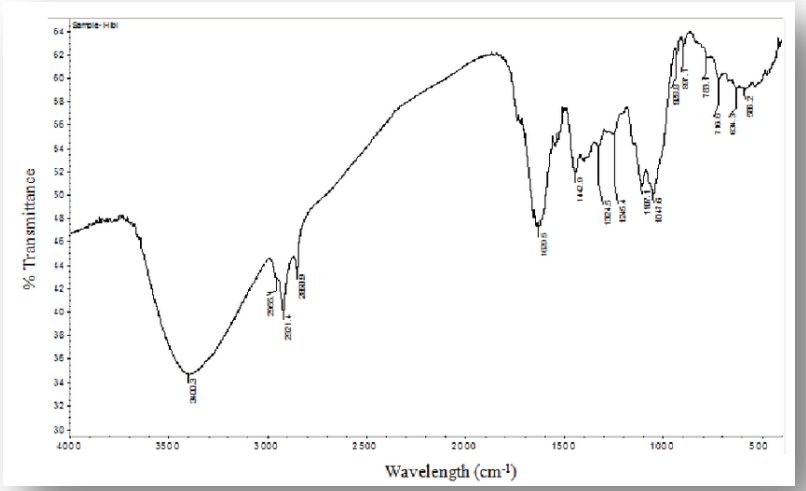
Obtaining a calibration curve by exploiting available data about sample composition

whichanalysis 4

Composition of artificially prepared samples can be used to improve quantitative procedures. (supervised multivariate analysis).

It requires:

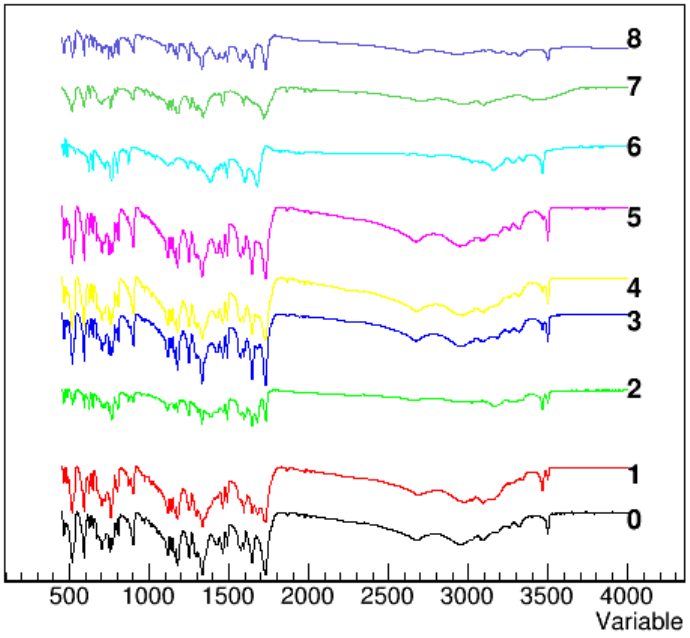
- a calibration set (estimation of model parameters, referw)
- a test set (predictions, test)



datatype

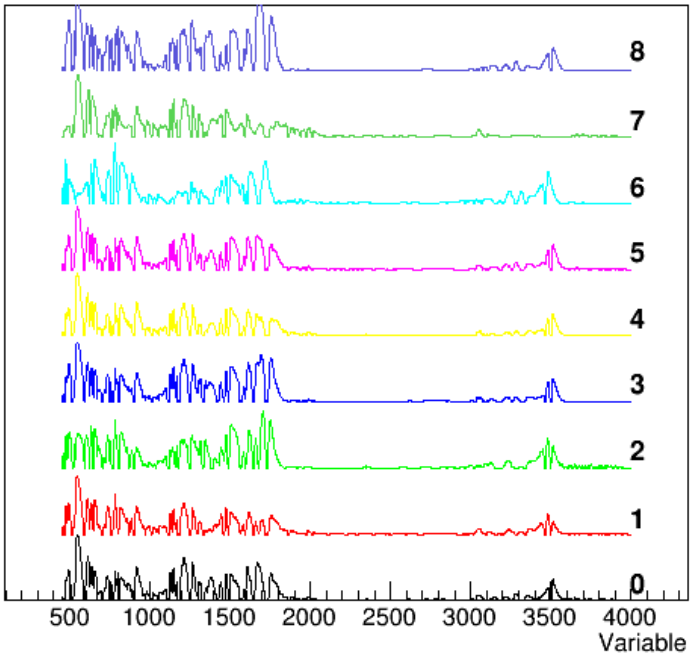
defines the type of input data

1) 2-column file, independent/dependent variable
2) 2-column file, 2theta/Intensity (X-ray data)
3) 4-column file, [HKL]/Intensity (x-ray data single crystal)
4) 2-column file, wavenumber/Intensity (IR data)
5) m,n matrix file, n=variables m=samples
6) m,n matrix file, m=variables n=samples
7) 1-column file, single variable



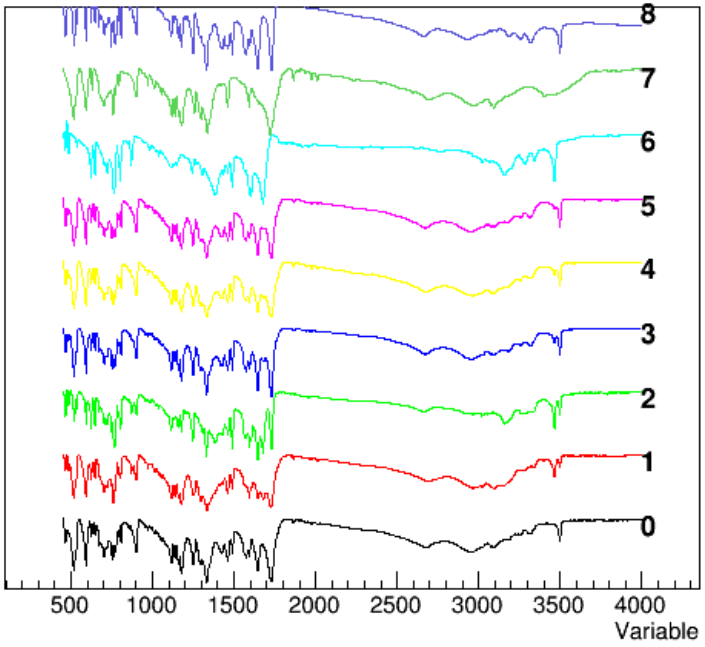
ORIGINAL DATA

Data affected by background (KBr sample holder)



preprocess 0 3 15

Data are rescaled, have the same base line, and peaks are well aligned (SNIP with Nclip=15)



preprocess 0 3

Data are rescaled according to $y'=(y-\langle y\rangle)/\sigma$

This removes background

FT-IR data analysis by RootProf

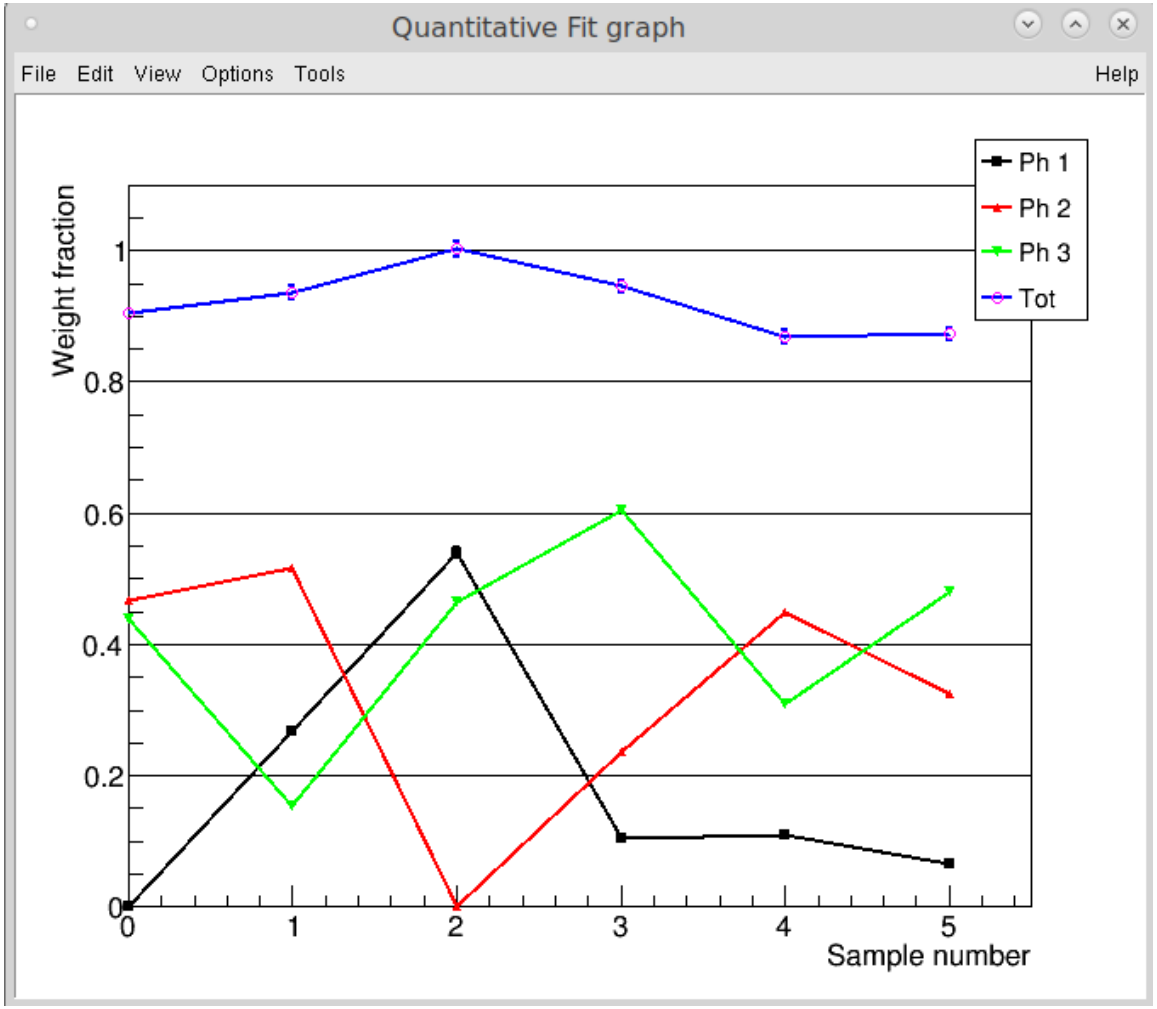
Supervised Quantitative Analysis: preprocessing

FT-IR data analysis by RootProf

Supervised Quantitative Analysis

COMMAND FILE

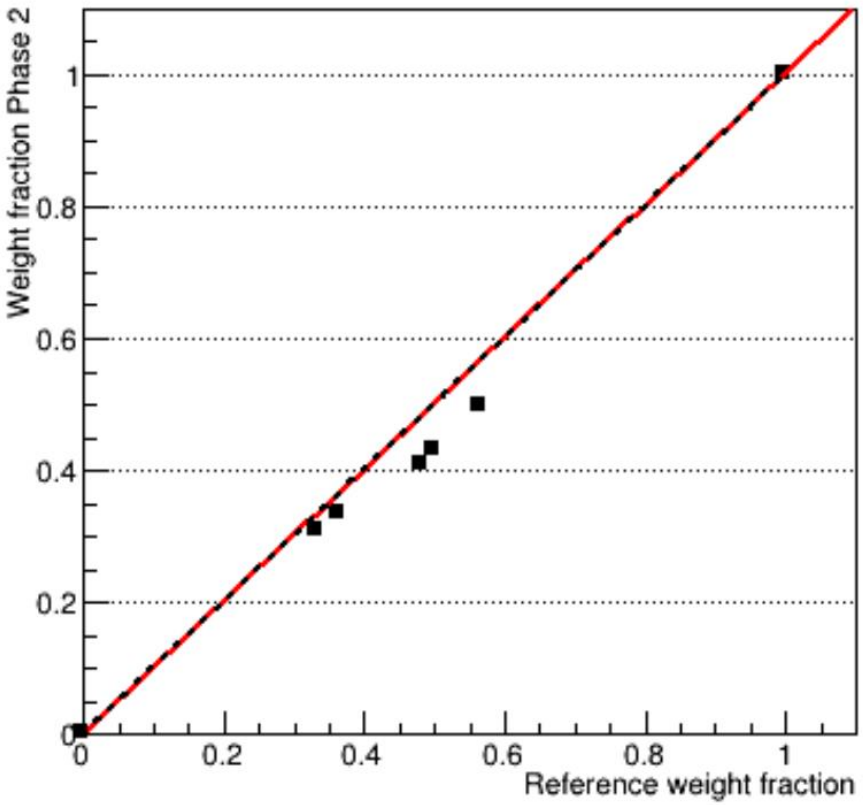
```
whichanalysis 4
calib 2
figpaper 1
dataType 4
range 450 4000
preprocess 0 3 15
file sample_3a.asc.extract
test
file sample_5a.asc.extract
test
file sample_S7a.asc.extract
test
file sample_11a.asc.extract
referw 0.347 0.334 0.319
file sample_21a.asc.extract
referw 0.263 0.482 0.255
file sample_22a.asc.extract
referw 0.238 0.364 0.399
file CBZ_III.asc.extract
referw 1 0 0
purephase
file SAC.asc.extract
referw 0 1 0
purephase
file CBZ-SAC.asc.extract
referw 0 0 1
purephase
```



FT-IR data analysis by RootProf

Supervised Quantitative Analysis: output

referw 0.238 0.364 0.399



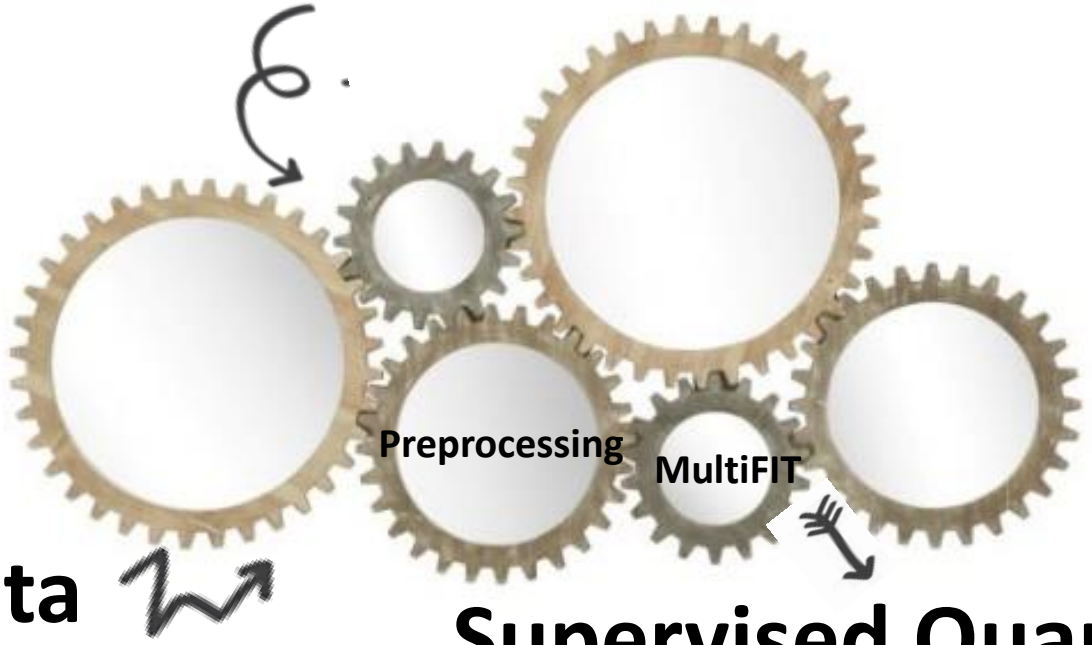
Spectrum 0: sample_3a.asc.extract
Chi-Square=1.09e+01, Reduced Chi-Square=3.08e-03, NDF=3547
Weight fraction Phase 1 0.006 +- 0.034
Weight fraction Phase 2 0.496 +- 0.007
Weight fraction Phase 3 0.378 +- 0.005
Total weight fraction 0.879 +- 0.035

Phase 1: RMSE=0.095 RMSE'=0.117 R2=0.935
R2(NPP)=0.778 AKLD=0.130
Calibration fit: Chi-Square=1.37e+03
intercept=-0.000+-0.000
slope=1.000+-0.000
Limits estimates: LOD=0.03 LOQ=0.08
NPP: Calibration fit: Chi-Square=4.55e+02
intercept=-0.177+-0.011
slope=1.214+-0.029
NPP: Limits estimates: LOD=2.95 LOQ=8.95

FT-IR data analysis by RootProf

Which is the relative amount of CBZ and SAC in our drug formulations?

Pure Phase



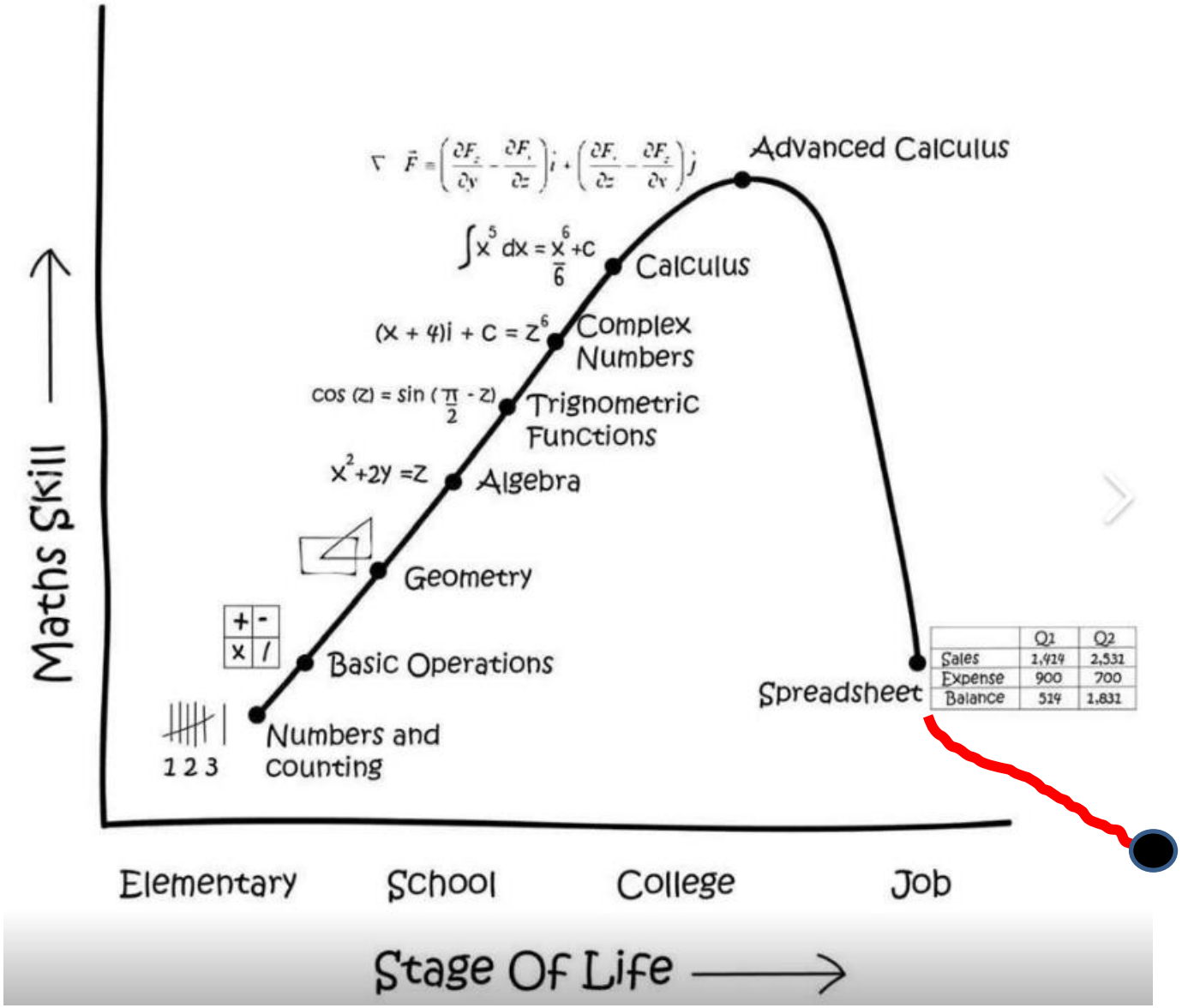
Data

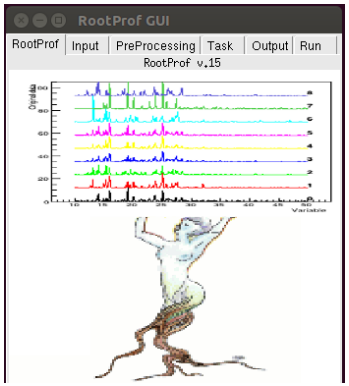
Supervised Quantitative Analysis (<60sec)

Conclusions

- RootProf program allows to perform chemometric analyses on multiple data sets
- It allows to works with several data such as XRPD, FT-IR, XAS, and SAXS spectra
- Qualitative and quantitative phase analysis of mixtures
- Clustering by PCA
- In addition, RootProf allows to combine information from different experimental techniques

Thank you for your attention





RootProf