

RootProf

TUTORIAL 7

Qualitative analysis of SAXS profiles

Contents

Chapter 1: The data set.....pag.2

Chapter 2: Qualitative analysis of SAXS profiles.....pag.3

Chapter 1

The data set

Unidimensional profiles from small angle X-ray scattering (SAXS) measurements on protein solutions compose our dataset. All profiles have been acquired at beamline BM29 of ESRF. They allow determining the aggregation properties of ubiquitin in presence of metal ions. The composition of the dataset is reported in Table 1. The corresponding files are included as demo files. They are formed by two columns, the first containing the q values, the second the corresponding values of intensity.

Table 1: Samples used for SAXS analysis.

Nsample	Code	Description
0	Ub_24h_S4mgml.dat	Pure ubiquitin sample
1	Ub_Cu_24h_S4mgml.dat	Ubiquitin with Cu ions
2	Ub_Zn25_24h_S4mgml.dat	Ubiquitin with Zn ions, 25 mM concentration
3	Ub_Zn200_24h_S4mgml.dat	Ubiquitin with Cu ions, 200 mM concentration

Chapter 2

Qualitative analysis of SAXS profiles

Motivation

Classify SAXS profiles by PCA.

The command file

The list of commands is the following.

```
whichanalysis 1
figpaper 1
range 0.5 1.7
preprocess 3 2 0 0
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
```

The commands have been included in the demo file named *fileInputSAXS*. See the user guide for an explanation of their meaning. Preprocess 3 2 0 0, for example, means that the Log_{10} of the intensity is considered, rescaled by standard normal variate.

Running RootProf

Start ROOT by clicking on his icon, or by typing “root” on a terminal window. Then write the root command:

```
Root> .x RootProf.C("fileInputSAXS")
```

Or

```
Root> .> outputSAXS
```

```
.x RootProf.C("fileInputSAXS")
```

```
.>
```

After some seconds, graphic windows will start appearing on your screen, while text output will appear on the terminal window, or redirected in the file named *outputSAXS*. When the run ends, the root prompt will appear again on the ROOT terminal, and you will be able to edit each single graphic window and read the output file by your text editor.

The graphic output

Input profiles are shown in Figs. 1 and 2. In this latter figure, in particular, one can appreciate the different slope of the profiles.

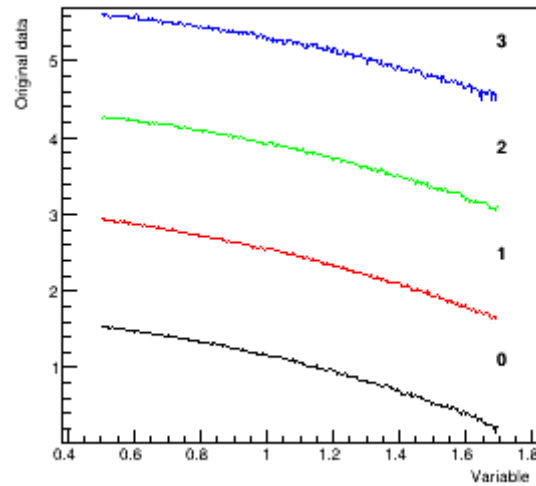


Fig.1 Original data shifted

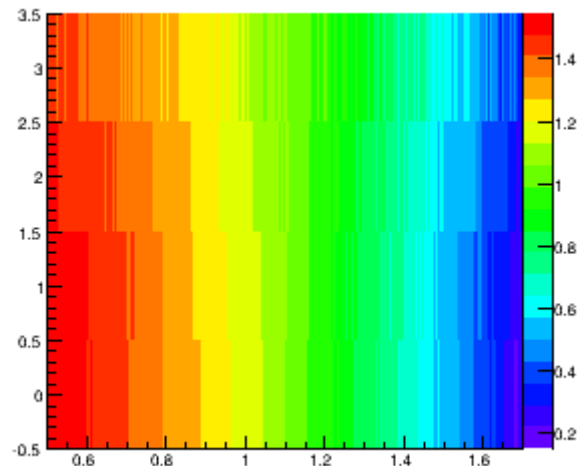


Fig.2 Data Matrix

The data matrix undergoes PCA. The corresponding Scree plot is shown in Fig.3, where one can see that almost all the sample variance is hold by the first principal component.

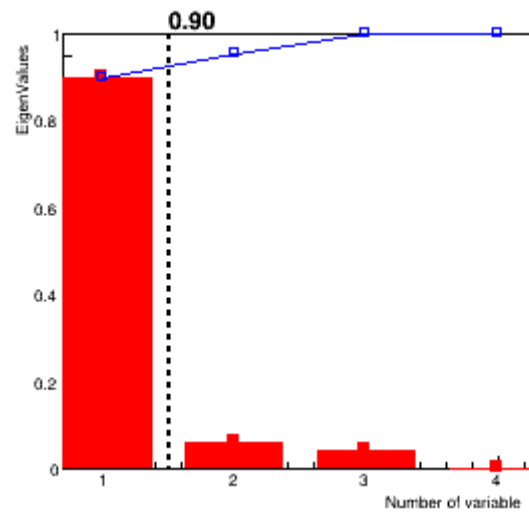


Fig.3 Scree plot

Original profiles are superimposed in Fig.4. Comparison of this figure with Fig.5, where the profiles reconstructed by using uniquely the first principal component are superimposed, confirms that PC1 capture most of the features of the dataset.

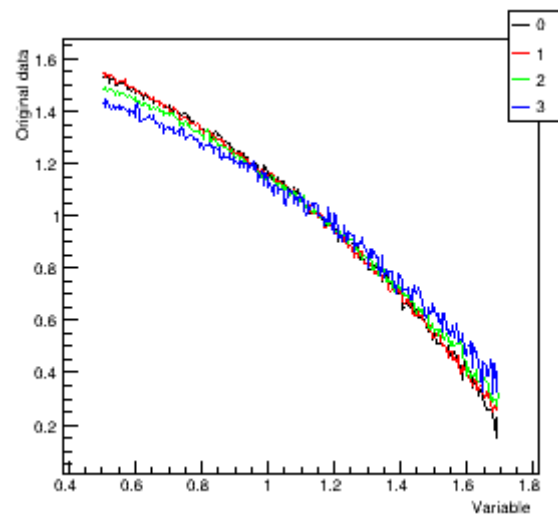


Fig.4 Original data

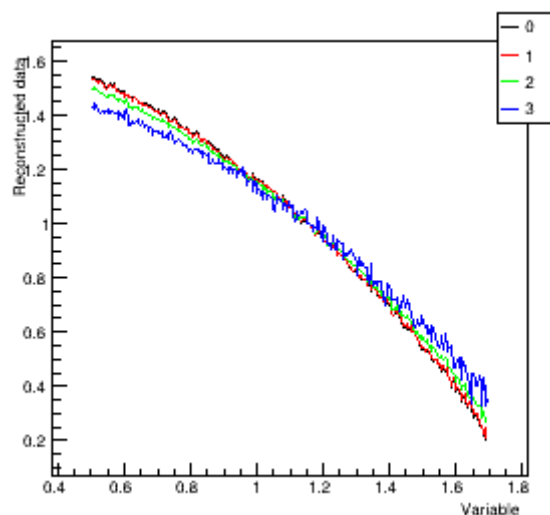


Fig.5 Reconstructed data

PC1 Loadings (Fig.6) indicate that this component has increasing contributions from the q values. This means that profiles are differentiated by PC1 on the basis of their slope.

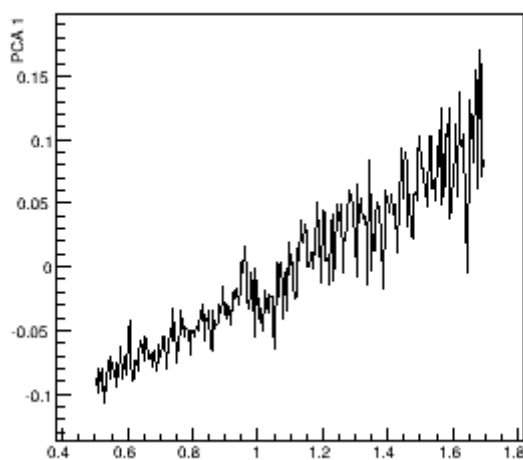


Fig.6 Loadings

PC1 score plot allow classifying profiles according to their slope: addition of metal ions to ubiquitin causes an increase of the slope of their SAXS profiles. This increase is more pronounced in presence of Zn, and proportional to its concentration.

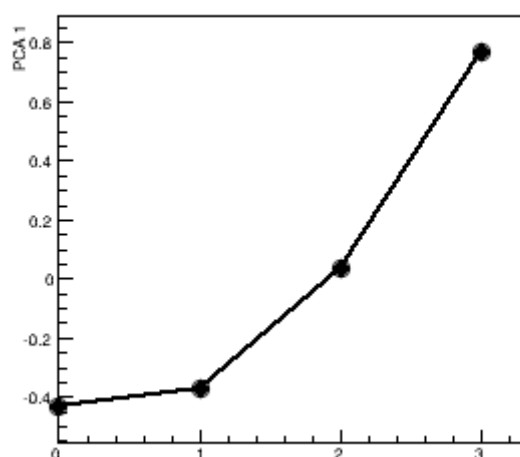


Fig.7 Scores

Output file

The content of the output file named *outputSAXS* is reported below, with comments added.

```
Input from file: fileInputSAXS
```

```
-----
whichanalysis 1
```

```
figpaper 1
```

```
range 0.5 1.7
```

```
preprocess 3 2 0 0
```

```
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
```

The section above shows the commands read from the command file. It should be checked to ensure that they are interpreted correctly.

```
Reading input files:
```

```
-----
Sample 0 -> file Ub_24h_S4mgml.dat
```

```
Found 278 points
```

```
Sample 1 -> file Ub_Cu_24h_S4mgml.dat
```

```
Found 278 points
```

```
Sample 2 -> file Ub_Zn25_24h_S4mgml.dat
```

```
Found 278 points
```

```
Sample 3 -> file Ub_Zn200_24h_S4mgml.dat
```

```
Found 278 points
```

The section above reports the number of data points read within each input file, as determined by the command *range*.

```
Starting Qualitative analysis
```

```
n. points 278  
Eigenvalues: 1 --> 89.63% (89.6%)  
Eigenvalues: 2 --> 5.92% (95.6%)  
Eigenvalues: 3 --> 4.44% (100.0%)  
Eigenvalues: 4 --> 0.00% (100.0%)
```

```
Chosen value of k=1: ratio=0.96 error=0.022
```

The section above shows the results of the PCA analysis. The first eigenvalues are listed as a function of their value, and the number of eigenvalues selected for PCA analysis is reported (k), together with the values of the threshold on the cumulative eigenvalue distribution (ratio), and an estimate of the corresponding error between original and reconstructed data (error).