

# SerialAnalyzer User's Guide

for ver. 1.3.1

2020-04-08

Masatsuyo TAKAHASHI, Kento YONEZAWA and Nobutaka SHIMIZU

Photon Factory

Institute of Materials Structure Science

High Energy Accelerator Research Organization, KEK

# Contents

0	Introduction .....	1
1	Installation.....	1
2	Introductory Procedure .....	1
3	Standard Procedure.....	2
3.1	Most Basic Steps .....	2
3.2	Decomposition Procedure .....	5
4	Input Data.....	7
4.1	X-ray Scattering Data .....	7
4.2	UV Absorbance Data .....	8
5	Output Results.....	9
5.1	Analysis Result Book .....	9
5.1.1	Guinier Analysis Result .....	9
5.1.2	Zero-Concentration Analysis Result.....	12
5.1.3	Analysis Summary .....	16
5.2	Other Output Files.....	16
6	Detailed Description of the Dialogs.....	20
6.1	Main Dialog .....	20
6.1.1	Purpose and Required Inputs.....	20
6.1.2	Quick Inspection of Data.....	21
6.1.3	Data Range Restriction .....	21
6.1.4	Preliminary Synchronization View .....	24
6.1.5	Removing Abnormal Points .....	24
6.1.6	Toggling the Mode concerning UV Data Usage.....	25
6.2	Mapping Dialog .....	26
6.2.1	Purpose and Manipulation Flow .....	26
6.2.2	Checking the Mapping Adequacy .....	27
6.2.3	Data Investigation.....	29
6.2.4	Attracting User's Attention .....	31
6.2.5	Baseline Correction Hierarchy .....	32
6.2.6	Note on the differences of use ranges of UV and X-ray data .....	33
6.2.7	UV Absorbance Baseline Correction Options .....	34
6.2.8	X-ray Scattering Baseline Correction Options .....	34
6.2.9	Low Percentile Method .....	35
6.2.10	Matrix Factorization Adjustment.....	36
6.2.11	X-ray Scattering Baseline Investigation.....	37
6.2.12	Synchronization Control .....	37

6.3	Range Editor.....	39
6.3.1	Purpose and Manipulation Flow .....	39
6.3.2	How to Change Analysis Ranges .....	39
6.3.3	Concentration Options .....	40
6.4	Decomposition Editor.....	41
6.4.1	Purpose and Manipulation Flow .....	41
6.4.2	Two Types of Data by Two Types of Models.....	42
6.4.3	Components, Peaks and Elution Elements .....	43
6.4.4	Distinction of Peak Significance and Number of Ranges .....	43
6.4.5	Editing Combination of Elution Elements .....	44
6.4.6	Elution Model Parameter Control.....	46
6.5	Analyzer Dialog.....	47
6.5.1	Purpose and Manipulation Flow .....	47
6.5.2	Description .....	47
6.6	Progress Dialog.....	49
6.6.1	Purpose and Manipulation Flow .....	49
6.6.2	Description .....	49
6.7	Extrapolation Preview .....	51
6.7.1	Preview Button Frame .....	51
6.7.2	Purpose and Manipulation Flow .....	51
6.7.3	Description .....	52
6.7.4	Optimization Options .....	55
6.7.5	Saving the Preview Results .....	56
6.7.6	Giving Known Scattering Curves.....	56
6.7.7	Background Execution of DENSS jobs.....	57
6.7.8	Re-optimization and Accessory Tools .....	58
6.8	Settings Dialog and Setting Menu.....	60
6.8.1	Purpose and Manipulation Flow .....	60
6.8.2	ATSAS AUTORG Location.....	60
6.8.3	Standard Mapping Plane .....	60
6.8.4	Angle Range Restriction Start.....	61
6.8.5	Extrapolation Regularization Weights.....	61
6.8.6	Other Options .....	61
6.8.7	Restoring Settings from a Result Folder .....	62
6.8.8	Showing Environment Information .....	62
6.9	SEC Tools.....	63
6.9.1	Abnormal Data Handling.....	63
6.9.2	Data Range Restriction.....	64

6.9.3	Elution Curve Picker .....	64
6.9.4	Scattering Curve Plotter .....	65
6.9.5	SVD Viewer .....	66
6.10	MCT-SAXS Support .....	67
6.11	Data Requirements .....	67
6.12	Manipulation Outline .....	67
6.13	MCT Tools .....	69
6.13.1	Average Subtractor .....	69
6.13.2	Individual Subtractor .....	69
6.13.3	Concentration Normalizer .....	70
6.13.4	Elution Curve Picker .....	70
6.13.5	Preprocess (Individual Subtractor with 3D Visualization) .....	70
6.14	DENSS Tools.....	71
6.14.1	DENSS GUI .....	71
6.14.2	DENSS Manager .....	72
6.14.3	Electron Density Viewer .....	73
6.14.4	SAXS Simulator.....	73
6.15	AutoGuinier Interface.....	75
7	X-ray Data only Mode .....	78
7.1	Design Philosophy.....	78
7.2	Implementation .....	78
7.3	Different Manipulation Points from the Normal Usage .....	78
7.3.1	Main Dialog.....	78
7.3.2	Mapping Dialog.....	79
8	Implementation Details .....	81
8.1	AutoGuinier .....	81
8.1.1	Outline.....	81
8.1.2	Command Line Interface .....	81
8.1.3	Preliminary Process .....	81
8.1.4	Scattering Curve Smoothing.....	81
8.1.5	Guinier-Porod Model Fitting .....	81
8.1.6	Guinier Interval Determination.....	81
8.1.7	Estimation of $R_g$ and $I(0)$ .....	81
8.1.8	Evaluation of Quality Factors .....	81
8.2	SerialAnalyzer .....	82
8.2.1	Outline.....	82
8.2.2	Used Data Range Restriction .....	82
8.2.3	Standard Mapping Plane .....	82



8.2.4	Peak Recognition of the Elution Curves .....	83
8.2.5	Flow Change Point Detection in UV absorbance .....	83
8.2.6	Optimization of the Mapping between UV absorbance and X-ray scattering .....	84
8.2.7	Range Determination for Zero-Concentration Analysis .....	84
8.2.8	Extrapolation to Zero-Concentration .....	84
8.2.9	UV absorbance baseline correction .....	84
8.2.10	X-ray scattering baseline correction .....	84
8.2.11	Intermediate files .....	84
	Appendix.....	85
A	Note on the Meaning of Error on Estimated $R_g$ .....	85
B	Intermediate files .....	86
C	KEK-staff Options .....	87
C.1	Specialist Options Dialog .....	87
C.1.1	Absorbance Data Treatment .....	88
C.1.2	Data Correction Options.....	88
C.1.3	Elution Mapping Options .....	88
C.1.4	Elution Curve Modeling .....	88
C.1.5	Guinier Analysis Options .....	88
C.1.6	Extrapolation Options .....	88
C.1.7	Deprecated Extrapolation Options .....	88
C.1.8	Maintenance Mode (for Guinier Analysis) .....	88
D	Developer Tools .....	89
D.1	Developer Options Dialog.....	89
D.2	Tester Dialog .....	90
E	Tutorials .....	91
E.1	Linear Transformation .....	91
E.2	SVD (2D).....	92
E.3	SVD (3D).....	92
E.4	Matrix Factorization.....	93
E.5	Conjugate Gradient Algorithm .....	93
E.6	Affine Transformation of Baseline Adjustment.....	94
E.7	Extrapolation with Spherical Particles.....	94
E.8	Noise and Sigma-width Dependency of Base Percentile Offset .....	95
	References .....	96
	Index .....	97

## 0 Introduction

- a This document describes SerialAnalyzer — a program intended to make easier the following basic analysis of data from SEC-SAXS (HPLC) experiments at KEK-PF.
  1. Guinier Analysis
  2. Extrapolation to Zero-Concentration
- b Supported platform is Windows 64bit only.
- c If the ATSAS program suite is available, it also uses AUTORG, ALMERGE and DATGNOM in the analysis in order to enable users to contrast and validate the results.
- d If you use results of this program, please cite: [Software for serial data analysis measured by SEC-SAXS/UV-Vis spectroscopy](#).

## 1 Installation

- a Download and unzip a latest zipped distribution file from [KEK download site](#).

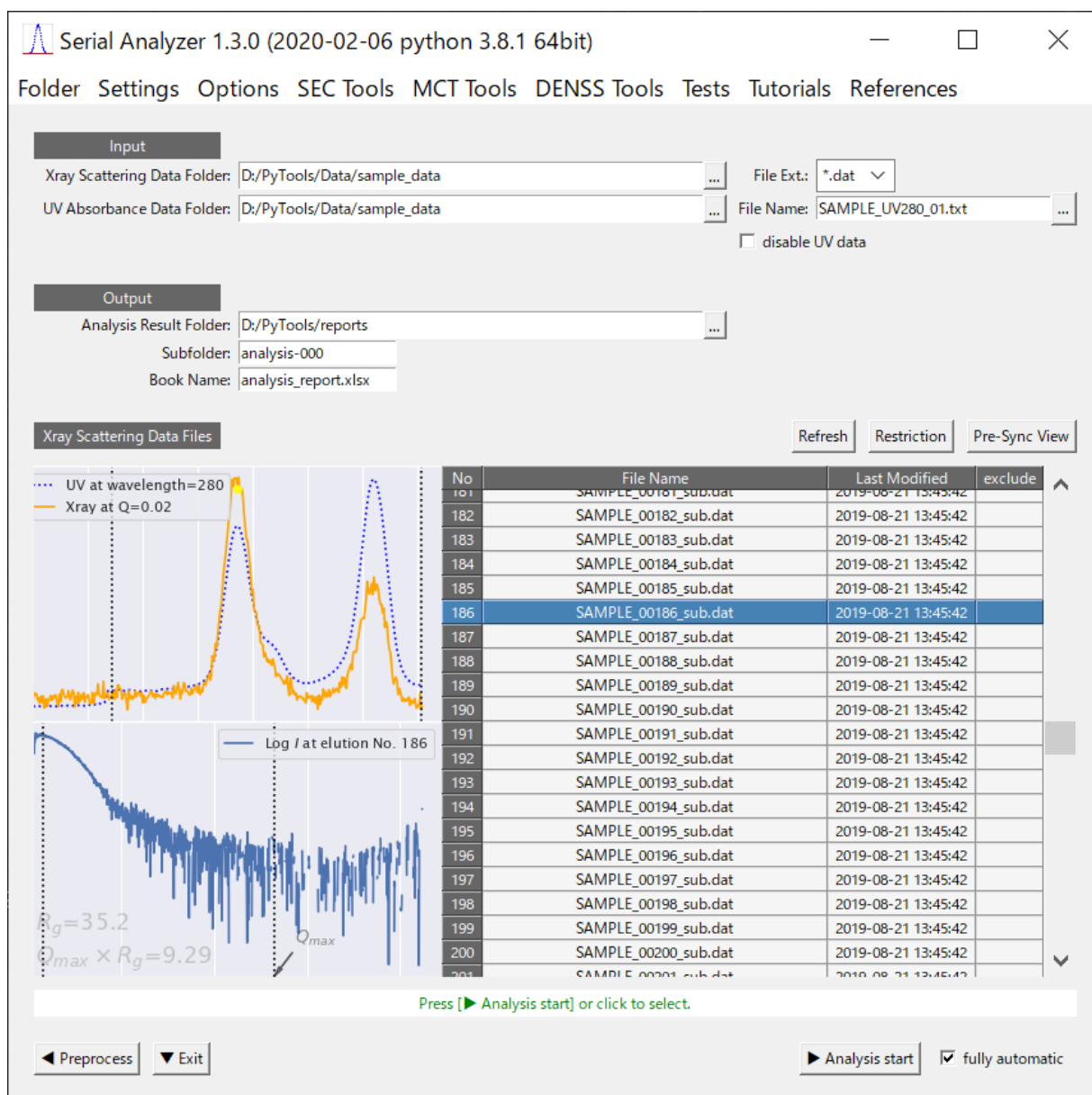
## 2 Introductory Procedure

- a Easiest usage is summarized here to give a quick idea of what it is. If you need any deviation from these basics or would like to know the reasons behind, please consult detailed description in Chapter 6.
- b Input X-ray scattering data are assumed to be pre-processed using SAngler (or other equivalent programs) in the following two steps. See Chapter 4 for their detailed requirements.
  1. Circular Averaging
  2. Background Subtraction
- c The name of a UV absorbance data file, named like “...UV.txt” or “...spectra.txt”, placed in the same folder is automatically recognized and filled in the entry.<sup>1</sup>
- d To follow the steps below, users can use an example set of data from [the download site](#).
- e The following are the easiest steps.
  1. Run the “serial\_analyzer.exe” in the installed folder.
  2. Specify the input/output folders in the Main Dialog. (Fig. 2-1)
  3. Press “► Analysis start” button to proceed further.
  4. If you do this with the “fully automatic” button checked, the standard steps described in the next Chapter 3 will be performed automatically with default parameters or options.
  5. Analysis results will be available in an excel book in the analysis result folder.
- f See Chapter 5 to understand information in the result book.
- g That is all for the easiest usage.

---

<sup>1</sup> Handling of those X-ray data not coupled with UV data is possible and described separately in Chapter 7.

Fig. 2-1 Main Dialog



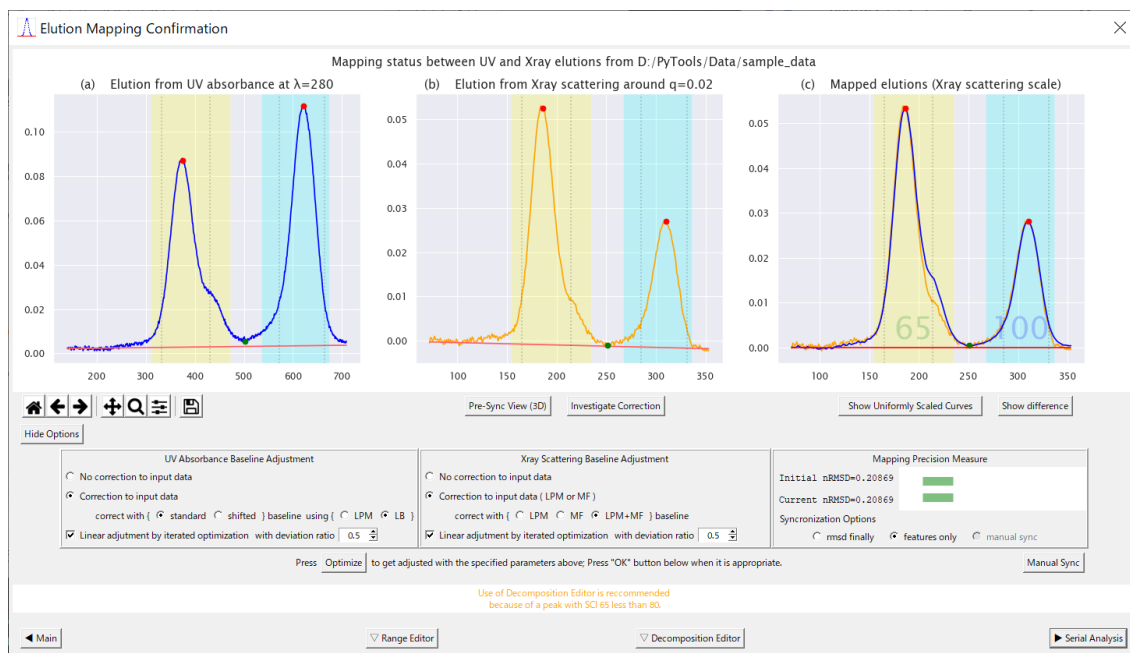
### 3 Standard Procedure

#### 3.1 Most Basic Steps

- a The following are the most basic steps automatically performed when you press “▶ Analysis start” button with “fully automatic” checked in the above main dialog (Fig. 2-1).
- b You should follow these steps manually unless you choose “fully automatic”. (Here is described only how to manipulate. See Chapter 6 for details and reasons.)
  1. Just after “▶ Analysis start” button press, the Mapping Dialog (Fig. 3-1) will appear.
  2. In the dialog, there are three figures which show the elution curves of, from left to right, UV absorbance data, X-ray scattering data and the overlay<sup>2</sup> of those two.

<sup>2</sup> UV elution curve in this figure has been locally scaled to make it possible to compare. See Section 6.2 for details.

**Fig. 3-1 Mapping Dialog**



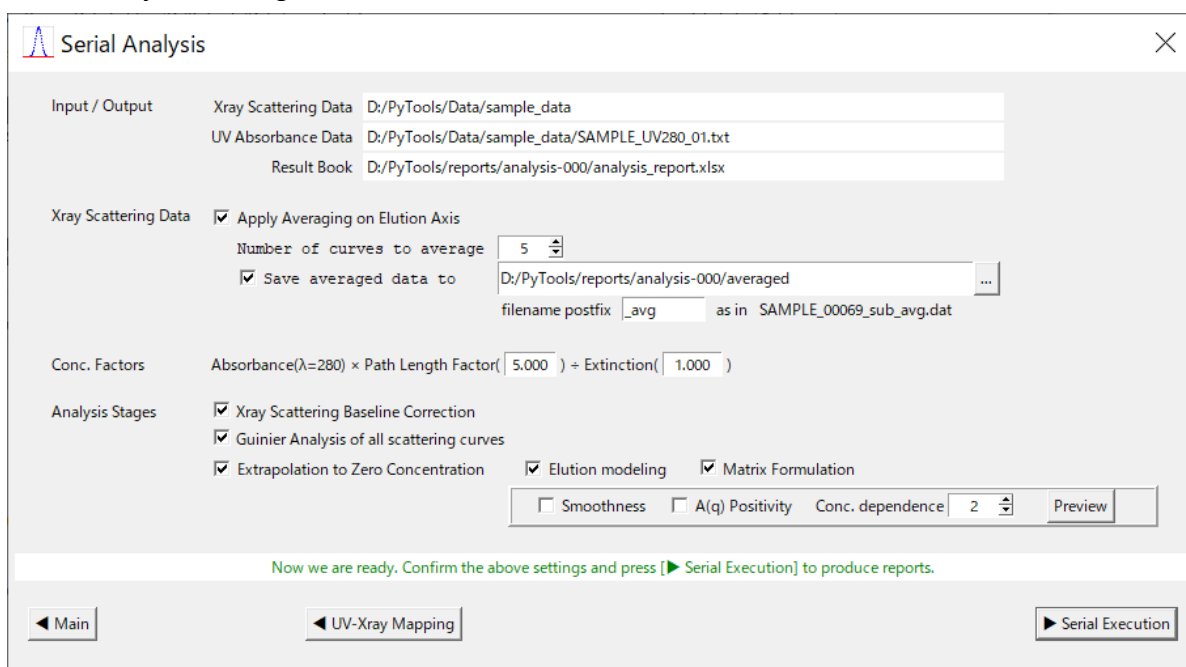
3. At this stage, if needed<sup>3</sup>, you can further use “Decomposition Editor” to better control the extrapolation process. See the next section 3.2 for such control.
4. In the “fully automatic” mode, use of “Decomposition Editor” is decided depending on data. It is used, currently, when some of the peaks are suspected of multi-component, judging by the following conditions.
  - ① There exists a peak with SCI<sup>4</sup> less than 80.
  - ② Or, there exists a significant<sup>5</sup> overlapping between adjacent peaks.
5. Looking at the figures in this dialog, confirm the elution mapping adequacy between the UV absorbance data and the X-ray scattering data.
6. For reference, the program shows its rough judgement of the adequacy in the guide message box at the lower part.
7. If the mapping is acceptable, proceed by pressing “► Serial Analysis” button, which shows the next Analyzer Dialog (Fig. 3-2).
8. Press “► Serial Analysis” button in the dialog, which executes the following steps.
  - ① Baseline correction of X-ray data
  - ② Guinier Analysis of all scattering curves
  - ③ Extrapolation to Zero-Concentration

<sup>3</sup> This decomposition step is required when the peaks seem to include multiple components.

<sup>4</sup> Single Component Indicator. See Section 6.2.2 for its definition.

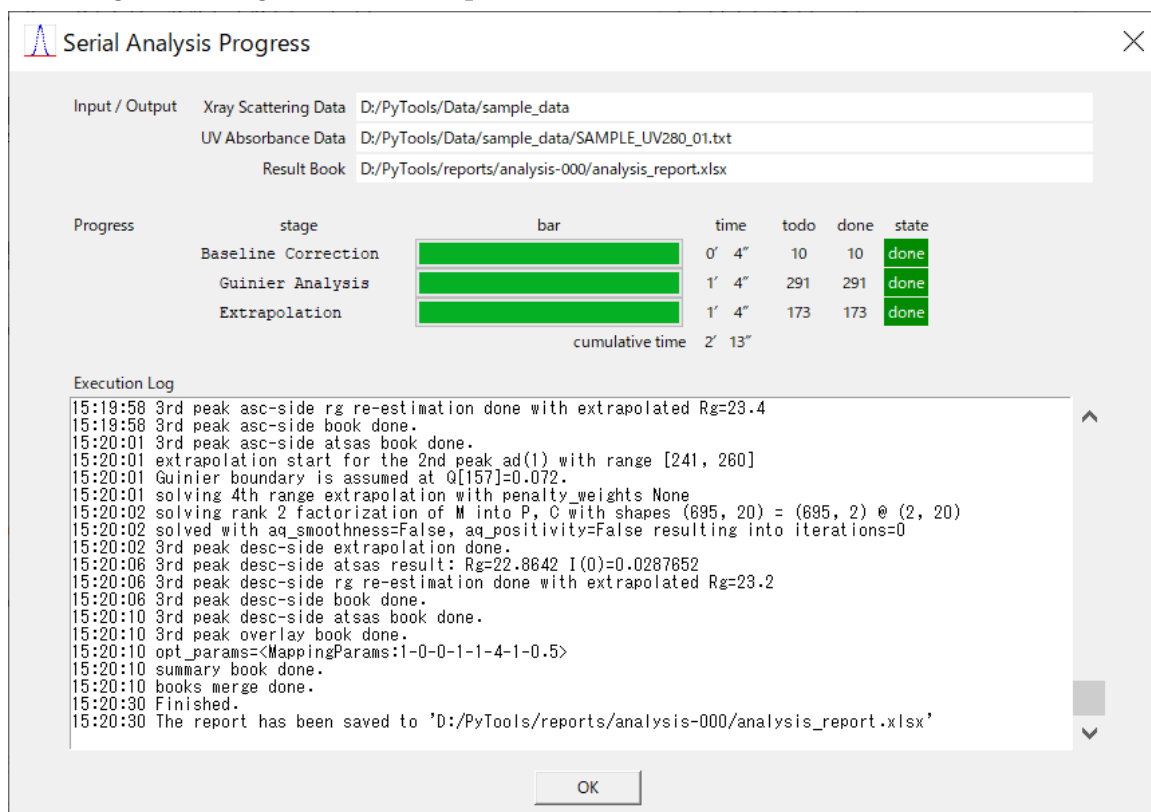
<sup>5</sup> It is judged significant if the peaks have some overlapping in their feet above 10% in height against their peak tops, measure by modeled elution curves.

**Fig. 3-2 Analyzer Dialog**



9. Wait several minutes until the states in the Progress Dialog (Fig. 3-3) are all “done”.
10. When finished, results are available in the output folder as stated in the progress log.

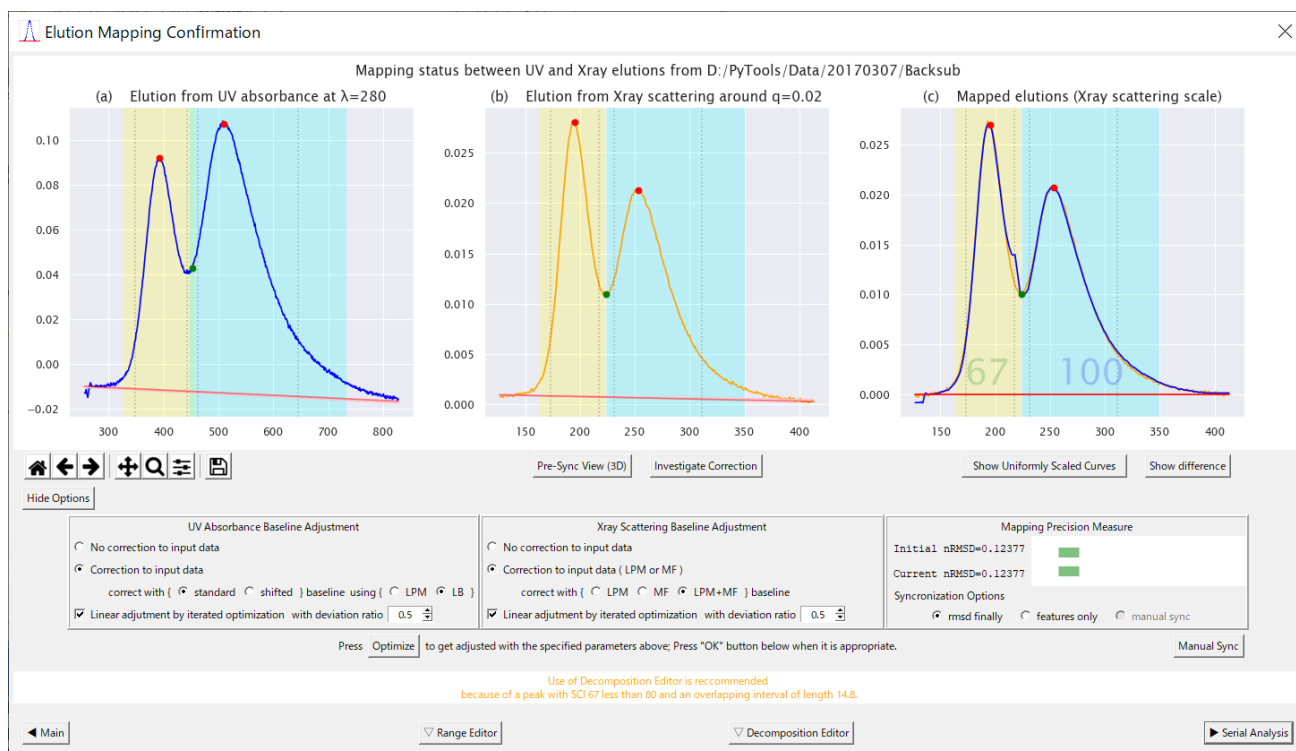
**Fig. 3-3 Progress Dialog when it is complete**



### 3.2 Decomposition Procedure

- a You can do without this procedure if you are confident that you have acquired data consisting only of single component peaks.
- b However, in general, such lucky situations are not always expected, and you may get peaks unclearly separated as shown in the next figure.

**Fig. 3-4 Example of a peak with another different component**

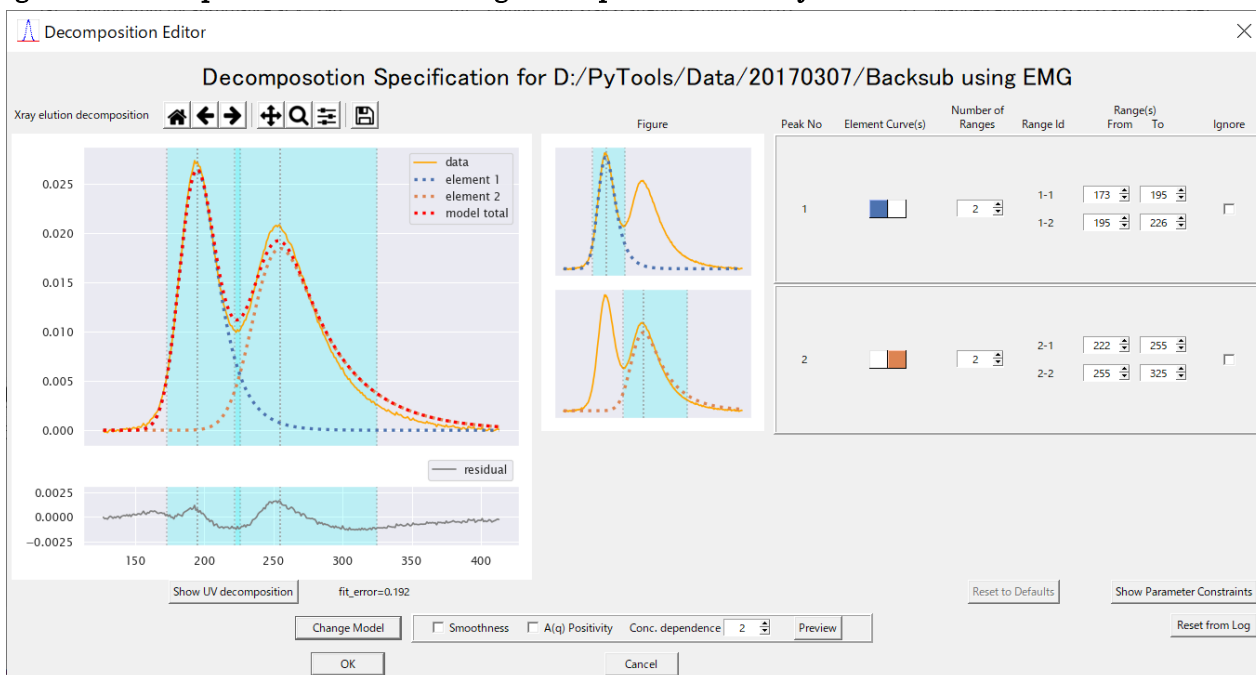


- c In such situations, decomposition of the total elution curve is required to make available elementary elution curves which separately represent the concentration of each single component.
- d Proceed to “Decomposition Editor” by pressing the button to confirm (and modify) the decomposition attained by the program’s algorithm.
- e You can decompose the elution curves, extracted from both of X-ray and UV data, using elution curve models (EMG<sup>6</sup> or EGH<sup>7</sup>) as shown in Fig. 3-5 and Fig. 3-6 below.

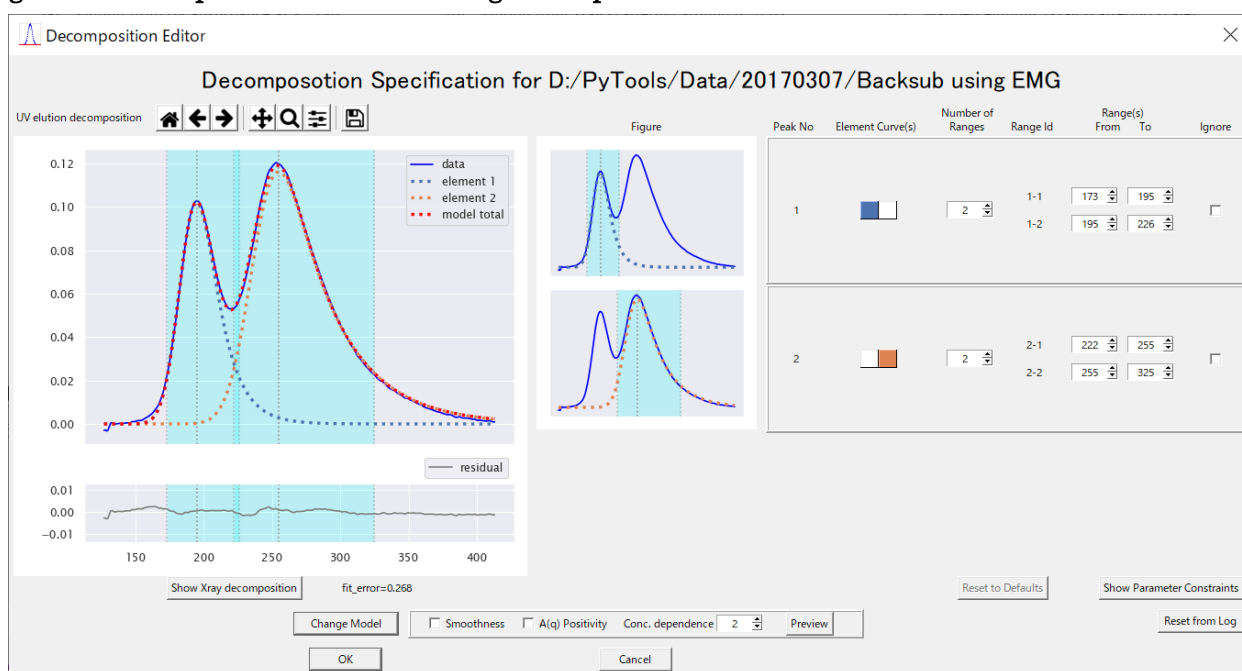
<sup>6</sup> Exponentially Modified Gaussian model.

<sup>7</sup> Exponential-Gaussian Hybrid model

**Fig. 3-5 Decomposition Editor showing decomposition of X-ray elution curve**



**Fig. 3-6 Decomposition Editor showing decomposition of UV elution curve**



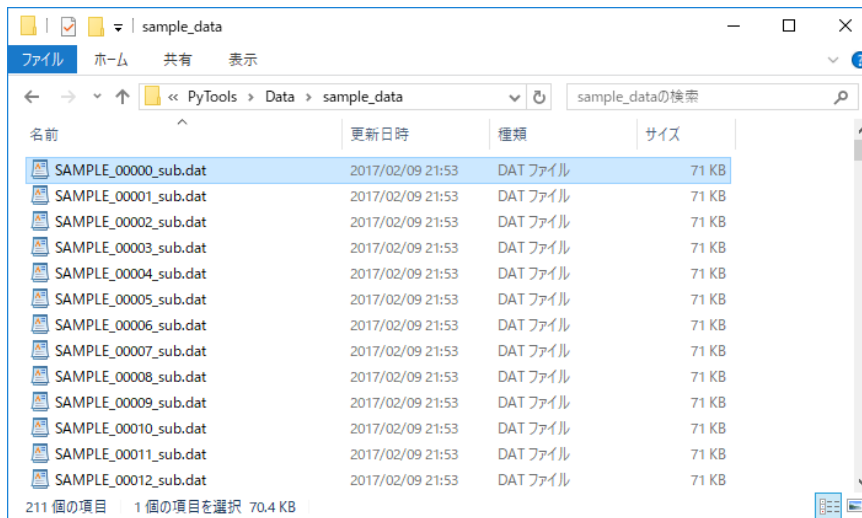
- f Press “OK” if it is acceptable, which leads you back to the Mapping Dialog (Fig. 3-1) introduced in the previous section.
- g This procedure described here above is the default course of action when you have chosen “fully automatic” in the main dialog.
- h See Section 6.4 for, non-default, advanced (or manual) usage of this editor.

## 4 Input Data

### 4.1 X-ray Scattering Data

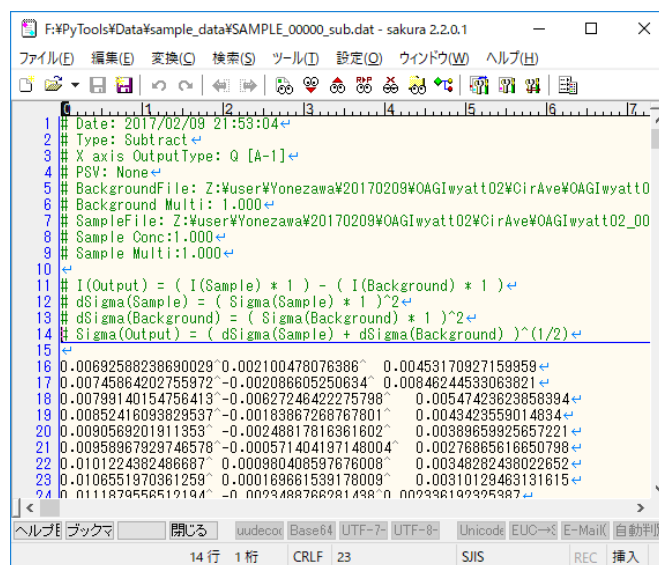
- a X-ray scattering data are assumed to be pre-processed using SAngler (or other equivalent programs) in the following steps.
  1. Circular Averaging
  2. Background Subtraction
- b As a data set from a respective experiment, they should be placed in one folder with file names suffixed by consecutive numbers as shown below.

**Fig. 4-1 Input folder containing X-ray scattering data files**



- c The suffix numbers identify the corresponding elution points.
- d The file format of an X-ray scattering data file is shown below in Fig. 4-2, which includes comment lines followed by data lines with tab-separated three columns of scattering vector length, intensity and estimated error.

**Fig. 4-2 File format of an X-ray scattering data file**





## 4.2 UV Absorbance Data

- UV absorbance data are used to evaluate the concentration of each elution points.
- They should be given either in one file or in one folder containing a set of multiple files.
- In the one-file style (which is recommended), it is convenient if the file is placed in the same folder as contains X-ray scattering files. If such is the case, the file is automatically recognized by the program on the condition that it is named like "...UV.txt" or "...spectra.txt".
- The file format in one-file style is shown in Fig. 4-3, which include comment lines followed by data lines with multiple columns.
- The first line of data lines is a header line with numbered column names, which identify elution points.
- The number of elution points, or the density of measurements, while it is usually larger than that of the corresponding X-ray scattering data, is not required to match to the latter.
- The first column includes wavelengths, which usually vary about from 200 nm to 450 nm.
- In data lines, columns other than the first column include absorbance values.

**Fig. 4-3 File format of a UV absorbance data file (one-file style)**

```

1 Date 2017年02月09日木曜日 07時33分14秒
2 Spectrometers: QEPB0040
3 ダーク使用
4 リファレンス 使用
5 積分時間[usec] 1000000
6 平均回数 1
7 スムージング 0
8 Electrical Dark USE
9 ピクセル数 1044
10 計測モード 吸光度
11 保存開始波長[nm] 200
12 保存終了波長[nm] 450
13 保存ピッチ 全て
14 保存間隔[ミ秒]
15 保存時間
16 保存回数 421
17 ウォームアップ時間[秒]
18 >>>>>>>>>>> 連続計測 (領域波長) Data Start<<<<<<<<<<<<<<<
19 0 9999 19999 29998 39998 49998 59997 69997 79996 89996 99996 109995 119995
20 200.219 0.01207 0.003497 0.013787 0.014327 0.008123 0.007566 0.010984 0.011147 0.006816
21 201.021 0.013297 0.007434 0.007446 0.012715 0.016575 0.013368 0.020904 0.013808
22 201.823 0.006931 0.011835 0.014828 0.011813 0.012408 0.017909 0.010822 0.00789
23 202.626 0.01224 0.01222 0.01223 0.0122 0.011836 0.015959 0.012291 0.006791 0.013203 0
24 203.428 0.016915 0.015644 0.021528 0.018134 0.016544 0.017808 0.015293 0.019412
25 204.230 0.020698 0.017623 0.017632 0.014192 0.016162 0.01655 0.025369 0.018004 0
26 205.032 0.01835 0.016955 0.019033 0.019356 0.021856 0.018741 0.018733 0.014894 0
27 205.833 0.022826 0.02219 0.022819 0.026237 0.02442 0.023179 0.028187 0.026251 0.024
28 206.635 0.019253 0.025362 0.023411 0.02619 0.018456 0.024851 0.021778 0.028737 0
29 207.437 0.024476 0.02771 0.024968 0.022965 0.023758 0.028757 0.024504 0.029718 0
30 208.238 0.022995 0.019003 0.021216 0.022975 0.02191 0.025924 0.027468 0.019444 0
31 209.040 0.025536 0.029762 0.027376 0.024985 0.026807 0.02582 0.028995 0.026578 0
32 209.841 0.026132 0.023478 0.030037 0.024527 0.025798 0.025801 0.028279 0.024006
  
```

- In the one-folder (multiple files) style (which is deprecated), the data are separated into multiple files in the same format, and the filenames are number-suffixed to represent the order of elution points.

## 5 Output Results

- a Analysis results are produced to an Excel book and other text files in the specified folder, which are described below in this chapter.

### 5.1 Analysis Result Book

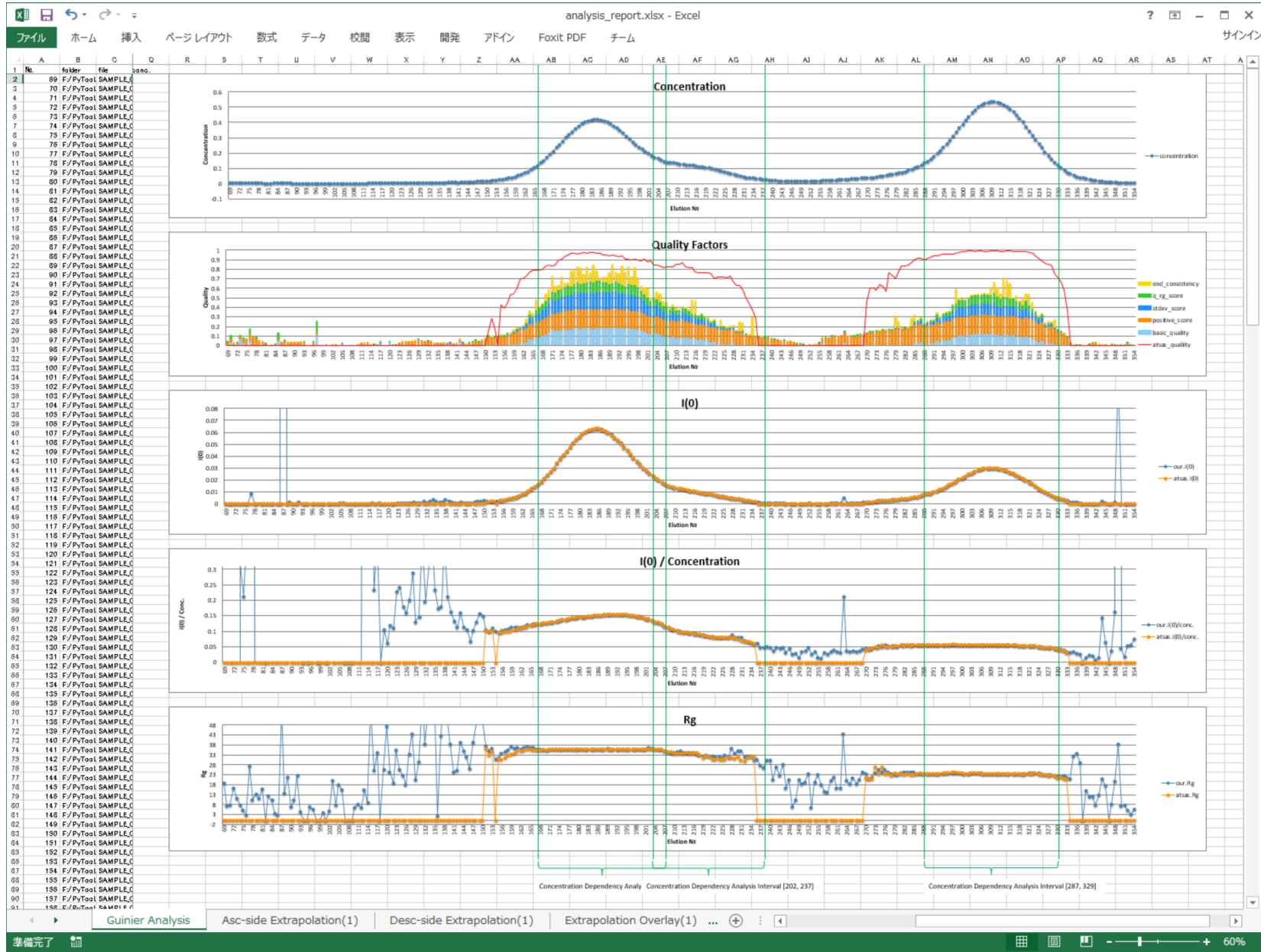
#### 5.1.1 Guinier Analysis Result

- a See Fig. 5-1 for overview of an example.
- b It is provided in a sheet containing five charts, which show each variation during the experiment of the following quantities.
  1. Concentration
  2. Quality
  3.  $I(0)$
  4.  $I(0)/\text{Concentration}$
  5.  $R_g$
- c Green lines indicating the intervals for further analyses are overlaid on the charts.
- d Fig. 5-1 shows two peaks with overlaid interval lines, which will be the input of extrapolation to zero-concentration analysis described later.
- e Sequential numbers starting from zero on the horizontal axes identify each elution point.
- f Data quality evaluation shown with vertical bars is computed as a weighted sum of the five factors listed in Tab. 5-1.
- g ATASAS AUTORG quality evaluation, if available, is contrasted with a red line.
- h Other analysis results (namely,  $I(0)$ ,  $I(0)/\text{Conc.}$ ,  $R_g$ ) from AutoGuinier are shown in a line with steel blue markers, which are only reliable when the corresponding quality score is sufficiently high ( $> 0.5$ ).
- i Corresponding results from ATASAS AUTORG are shown in a line with orange markers and the values have been set to zeros there if the program do not produce results.

**Tab. 5-1 Data Quality Factors**

No	Factor Name	Value Range	Weight	Formula	Description
1	basic quality	0 – 1	0.2	$\exp\left(0.5 \times \left(1 - \frac{\sum abs(gradient)}{height}\right)\right)$	Indicates smoothness or noiselessness of the scattering curve. Computed, as shown by the formula, from the monotonicity of the small-angle region (roughly determined as the 1/8-th of the whole region). Expected greater if the data seem less noisy. The scaling, 0.5 in the formula, has been chosen to better demonstrate the quality differences that we are interested in.
2	positive score	0 – 1	0.2	$2 \times \left(\max\left(0.5, \sum_{intensity[i]>0} weights[i]\right) - 0.5\right)$	While intensity values should be positive, data may contain negative values, which is not desirable. This score indicates the desirability of the data in this sense. Weights are chosen linearly to satisfy $\sum weights = 1$ , being greater for the smaller-angle region, using the slope from the 95 and 5 percentile values. It is 1 if the data contain no negative value and zero if about half of the data are negative, considering weights.
3	$R_g$ stdev score	0 – 1	0.2	$\exp\left(-10 \times \frac{R_g \text{ stdev}}{R_g}\right)$	Indicates the conditional reliability of estimated value of $R_g$ , assuming the appropriate determination of the Guinier interval. Computed, as shown by the formula, from the $R_g$ stdev propagated from the slope error of the linear regression in the Guinier plot. It is greater if the $R_g$ stdev is smaller. The scaling, 10 in the formula, has been adjusted from hundreds of profiles, including bad quality ones, from more than 50 experiments.
4	left $qR_g$ score	0 – 1	0.2	raw $qR_g$ score $\times$ basic quality where raw $qR_g$ score = 1 if $q_{left} \cdot R_g \leq 0.2$ 0 if $q_{left} \cdot R_g \geq 1.2$ ( $1.2 - q_{left} \cdot R_g$ ) otherwise	The right end of Guinier interval is bounded by the conventional constraint that $q \cdot R_g < 1.3$ , which comes from ensuring the accuracy of the Taylor expansion, while the left end selection depends on the linearity of the scattering curve in the lowest Q region. This score is intended to indicate the latter difference. It is multiplied by the basic quality to avoid accidentally getting unfairly high scores in low quality cases.
5	end consistency	0 – 1	0.2	$\exp\left(-20 \times abs\left(1 - \frac{left R_g}{right R_g}\right)\right)$	This score evaluates the linearity of the curve in the Guinier interval, in another way, from the ratio between $R_g$ 's at both ends of the interval as shown in the formula. The scaling, 20 in the formula, has been adjusted in the same way as in $R_g$ stdev score. It is greater if the $R_g$ 's at both ends of the Guinier interval is closer.

Fig. 5-1 Guinier Analysis Result



### 5.1.2 Zero-Concentration Analysis Result

- a Result is provided in several sheets for each (ascending or descending) side of the peaks.
- b See Fig. 5-2 and Fig. 5-3 for examples of those sheets.
- c In each sheet, three charts at top are intended to show the variations of concentration, estimated Rg's and I(0)/c during the extrapolation process.
- d In the two charts at center and right, namely, Rg extrapolated and I(0)/c extrapolated charts, green lines show the estimated values from input data, while orange lines show those from re-constructed (or extrapolated) data.
- e Extrapolation to zero-concentration is summarized in Section 6.7.3 for the matrix norm formulation, which is the default course of action. Here follows a deprecated non-matrix iterative description.

1. Suppose we have selected one side of a peak and extracted a data set as follows. (Where, data sizes 800, 20 have been chosen arbitrarily only for description convenience)

a vector of 800 scattering vector length values  $Q_i$

$$[Q_0 \quad Q_1 \quad \cdots \quad Q_{799}]$$

a vector of 20 concentration values  $C_j$

$$[C_0 \quad C_1 \quad \cdots \quad C_{19}]$$

a matrix of  $800 \times 20$  intensity values  $I_{i,j}$  for each  $(Q_i, C_j)$  pair

$$\begin{bmatrix} I_{0,0} & I_{0,1} & \cdots & I_{0,19} \\ I_{1,0} & I_{1,1} & \cdots & I_{1,19} \\ \vdots & \vdots & \ddots & \vdots \\ I_{799,0} & I_{799,1} & \cdots & I_{799,19} \end{bmatrix}$$

2. For each fixed row (or  $i$ ) in the matrix, solve the following linear regression problem (linear on those coefficient parameters) and determine the parameters  $A[i]$  and  $B[i]$ .<sup>8</sup>

$$I_{i,j} = C_j A[i] + C_j^2 B[i] \text{ for each } C_j \text{ in a selected side of a peak}$$

or, to express visually,

$$I_{i,0} = C_0 A[i] + C_0^2 B[i]$$

$$I_{i,1} = C_1 A[i] + C_1^2 B[i]$$

...

$$I_{i,19} = C_{19} A[i] + C_{19}^2 B[i].$$

3. Since we now have parameters  $A[i]$  and  $B[i]$  for all  $Q_i$ , we can re-construct scattering intensity  $I_c(Q_i)$  for any desired concentration  $c$  from the quadratic relation

$$I_c(Q_i) = A[i]c + B[i]c^2. \tag{5-1}$$

---

<sup>8</sup> In the formulation, the quadratic relation between I(q) and concentration is based on the formula (2.13) in the classical book[1] by Feigin and Svergun, 1987. See also Section 6.7 for the formulation using matrices.

Fig. 5-2 Zero-concentration Analysis result for an ascending side of a peak

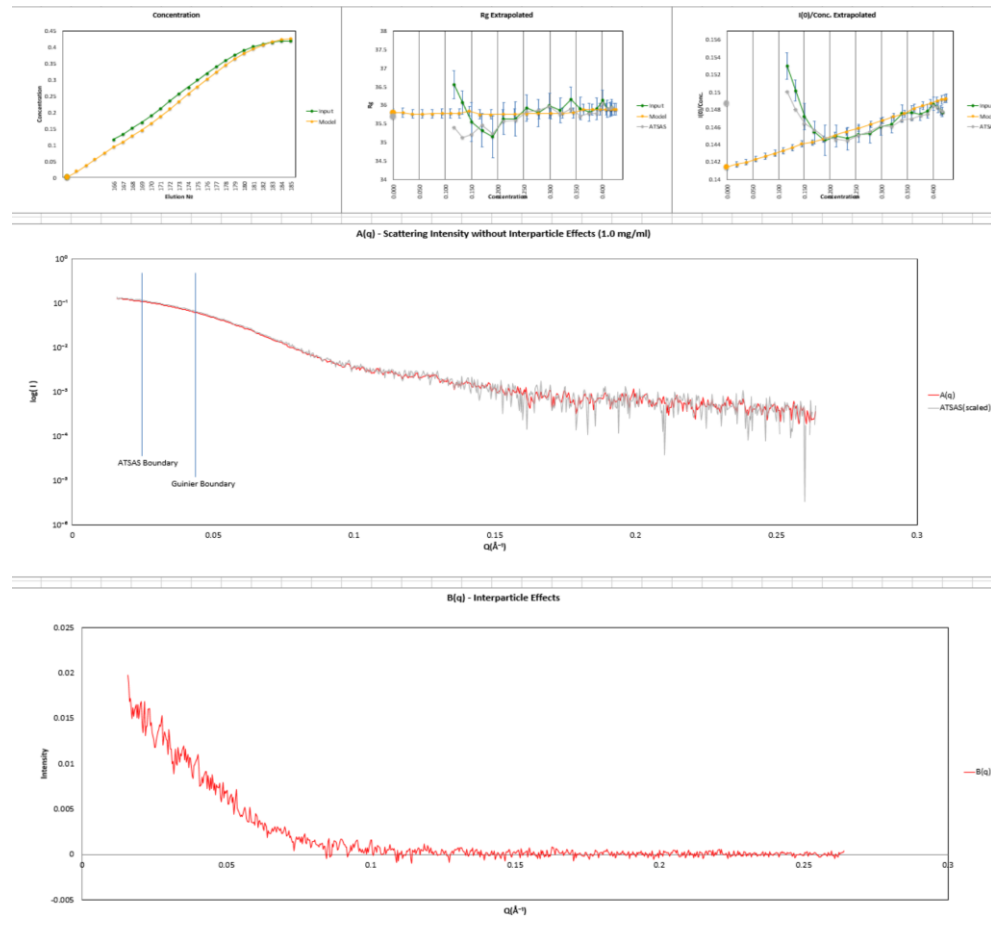
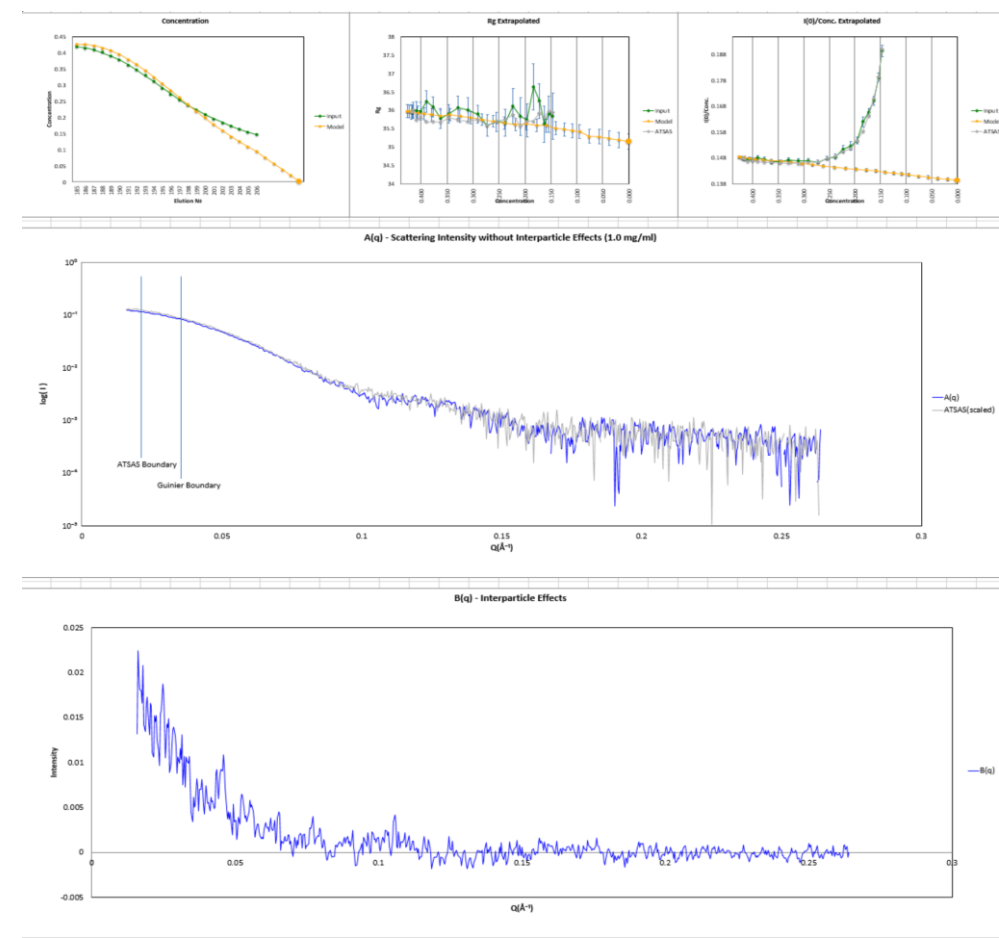


Fig. 5-3 Zero-concentration Analysis result for a descending side of a peak



f For zero-concentration analysis, we can consider the following equivalent formula instead, because their difference is only a parallel shift of amount  $\log(c)$  in the logarithmic scaling of Guinier plot.

$$I_c(Q_i)/c = A[i] + B[i]c. \quad (5-2)$$

g Formula (5-2) can be interpreted as follows for  $c=0$ .

$$I_0(Q_i)/0 \equiv A[i]. \quad (5-3)$$

h Parameters  $A[i]$  and  $B[i]$ , interpreted as functions of the scattering vector length  $Q$ , are shown in charts at the center and bottom of the sheet respectively.

$$\begin{aligned} A(Q_i) &\equiv A[i] \\ B(Q_i) &\equiv B[i] \end{aligned} \quad (5-4)$$

i As the chart titles say, the meaning of the functions  $A(q)$  and  $B(q)$  are, respectively,

1. scattering intensity without inter-particle effects,
2. inter-particle effects.

j The colors of these (in a sense, separated) curves are shown in red and blue for the ascending and descending sides respectively so that they can be easily identified in the overlaid charts, which are provided in another sheet. (Fig. 5-4)

k As for the separated curves, ATAS ALMERGE produces only an  $A(q)$  curve for each side, which is shown, scaled<sup>9</sup> to match the above  $A(q)$ , if available, in gray at the center of its respective sheet. (Fig. 5-2, Fig. 5-3)

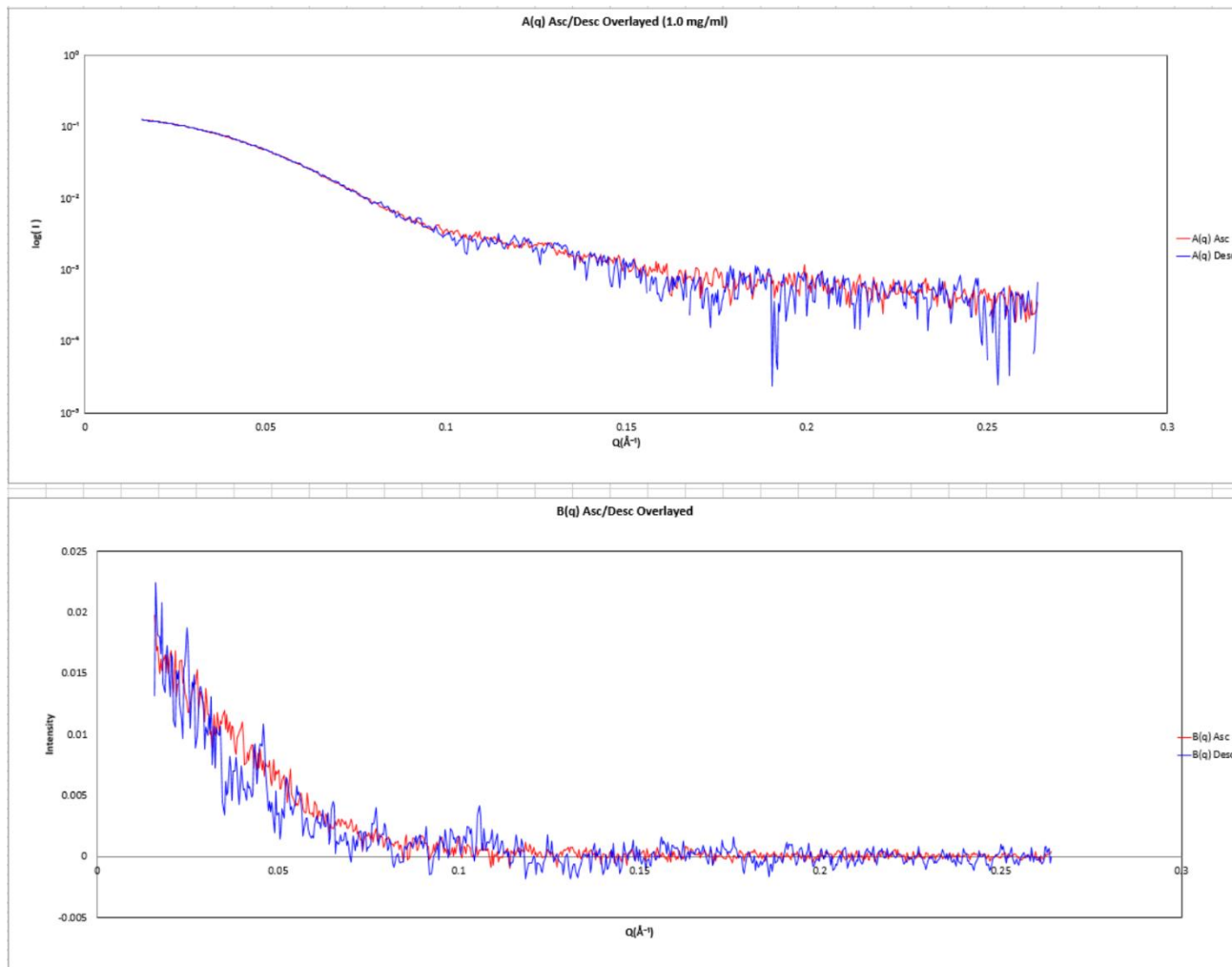
l In the chart of curve  $A(q)$ , vertical lines indicate boundaries where the curve construction policy changes between the small and wide-angle regions.

m Those policies are different depending on the programs. See section 8.2.7 for details.

---

<sup>9</sup> The extrapolated curve from ALMERGE is not normalized by concentration. It is scaled to match our result which is concentration-normalized.

Fig. 5-4 Overlaid charts of both sides of a peak





### 5.1.3 Analysis Summary

- a It is provided in three sheets named as follows.
  1. Extrapolation Summary (Fig. 5-5)
  2. Entire Summary (Fig. 5-6)
  3. Summary for Publication (Fig. 5-7)
- b The extrapolation summary contains estimated values and quality information for each ascending or descending side of peaks. The quality scores are available both in amounts and as a chart as shown in Fig. 5-5.
- c The entire summary contains enough information to identify and reproduce the analysis, as well as analysis evaluation information.
- d The summary for publication contains some of the items available at this stage, which are recommended in the SAXS publication guideline [4].

### 5.2 Other Output Files

- a Other output files are listed in Tab. 5-2.
- b Files postfixed as “\_asc.dat-” or “\_dsc.dat-” are numbered with prefixes, like “pk1\_”, in the order of elution for both cases with a single peak or multiple peaks.
- c “\_cn” postfix is added to suggest that the file contains concentration-normalized result.

**Tab. 5-2 Other output files**

No	File name	Description
1	atsas/pk1_asc.dat	Extrapolated curve data for the ascending side by ALMERGE
2	atsas/pk1_dsc.dat	Extrapolated curve data for the descending side by ALMERGE
3	pk1_asc_A_cn.dat	Extrapolated A(q) data for the ascending side by this program
4	pk1_asc_B_cn.dat	Extrapolated B(q) data for the ascending side by this program
5	pk1_dsc_A_cn.dat	Extrapolated A(q) data for the descending side by this program
6	pk1_dsc_B_cn.dat	Extrapolated B(q) data for the descending side by this program
7	serial_analyzer.log	Analysis log from this program

Fig. 5-5 Extrapolation Summary Sheet

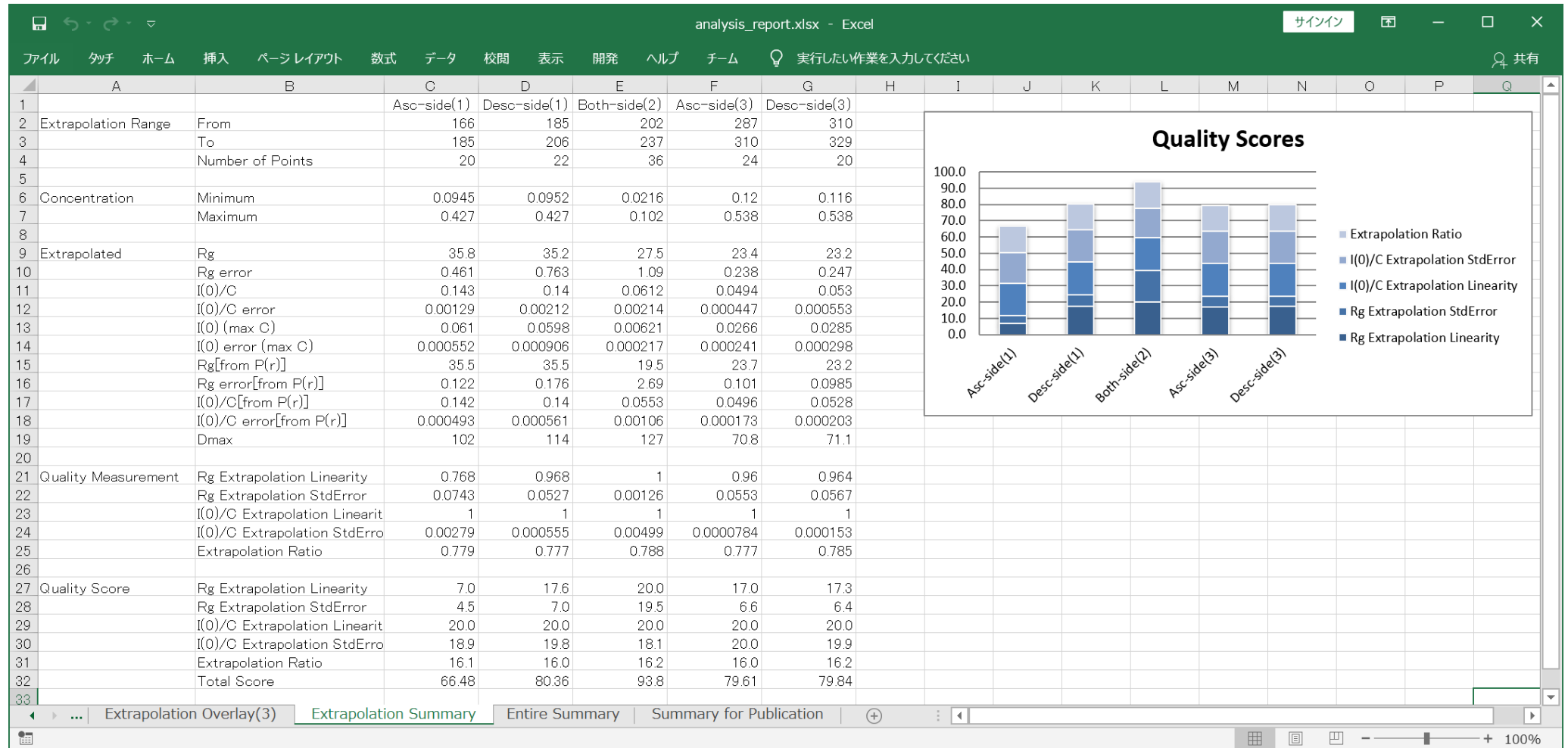


Fig. 5-6 Entire Summary Sheet

	A	B	C
1	Category	Item Name	Item Value
2	Input Data	Xray Scattering Data Folder	D:/PyTools/Data/sample_data
3		Xray Scattering Data Shape	(286, 695, 3)
4		UV Absorbance Data Folder/File	D:/PyTools/Data/sample_data/SAMPLE_UV280_01.txt
5		UV Absorbance Data Shape	(318, 571)
6		Output Data	Analysis Result Folder/Book
7	Averaged Data Folder		D:/PyTools/reports/analysis-000/averaged
8	Extrapolated Data Folder		
9	Input Data Pre-processing	Number of Elution Points for Averaging	5
10	Concentration Factors	Path Length Factor	5.0000
11		Extinction	1.0000
12		UV/Xr-ay Mapping Settings	UV Absorbance Baseline Correction
13		Xray Scattering Baseline Correction	LPM(linear)+MF with adjustment(0.5)
14	Zero-Concentration Analysis Settings	Optimizer Formulation	Matrix formulation with elution models
15		Range Determination Policy	Decomposed Elution Range
16		Analysis Ranges	[[166, 185], [185, 206]], [[202, 237]], [[287, 310], [310, 329]]
17	Serial Analysis Quality	UV/Xray Mapping Adequacy	
18		Xray Scattering Baseline Fitness	
19	Time Info	Date/Time Created	2019-10-18 15:20:10
20		Basesurface Correction	00:02
21		Guinier Analysis Execution	01:06
22		Extrapolation Execution	00:45
23	Other Info	Program Version	Serial Analyzer 1.2.1 (2019-10-21 python 3.8.0 64bit)
24		Excel Version	16.0
25		Logical Number of CPUs	24
26		CUDA Tools Version	(cupy 6.3.0, cuda 10010)
27			

Fig. 5-7 Summary for Publication Sheet

	A	B	C	D	E	F
1			Component-1	Component-2	Component-3	
2	Data-collection parameters					
3	Beamline					
4	Beam geometry ( $\mu\text{m}$ )					
5	Wavelength ( $\text{\AA}$ )					
6	q range ( $\text{\AA}^{-1}$ )	0.0158 - 0.264				
7	Exposure time (s)					
8	Concentration range ( $\text{mg ml}^{-1}$ )		0.0952 - 0.427	0.0216 - 0.102	0.116 - 0.538	
9	Temperature (K)					
10						
11	Structural parameters					
12	$I(0)$ ( $\text{cm}^{-1}$ ) [from $P(r)$ ]		$0.14 \pm 0.000561$	$0.0553 \pm 0.00106$	$0.0528 \pm 0.000203$	
13	$R_g$ ( $\text{\AA}$ ) [from $P(r)$ ]		$35.5 \pm 0.176$	$19.5 \pm 2.69$	$23.2 \pm 0.0985$	
14	$I(0)$ ( $\text{cm}^{-1}$ ) (from Guinier)		$0.14 \pm 0.00212$	$0.0612 \pm 0.00214$	$0.053 \pm 0.000553$	
15	$R_g$ ( $\text{\AA}$ ) (from Guinier)		$35.2 \pm 0.763$	$27.5 \pm 1.09$	$23.2 \pm 0.247$	
16	$D_{\text{max}}$ ( $\text{\AA}$ )		114	127	71.1	
17	Porod volume estimate ( $\text{\AA}^3$ )					
18	Dry volume calculated from sequence ( $\text{\AA}^3$ )					
19						
20	Molecular-mass determination					
21	Partial specific volume ( $\text{cm}^3 \text{g}^{-1}$ )					
22	Contrast ( $\Delta \rho \times 10^{10} \text{cm}^{-2}$ )					
23	Molecular mass $M_r$ [from $I(0)$ ]					
24	Calculated monomeric $M_r$ from sequence					
25						
26	Software employed					
27	Guinier Analysis	AutoGuinier				
28	Zero-concentration Extrapolation	SerialAnalyzer				

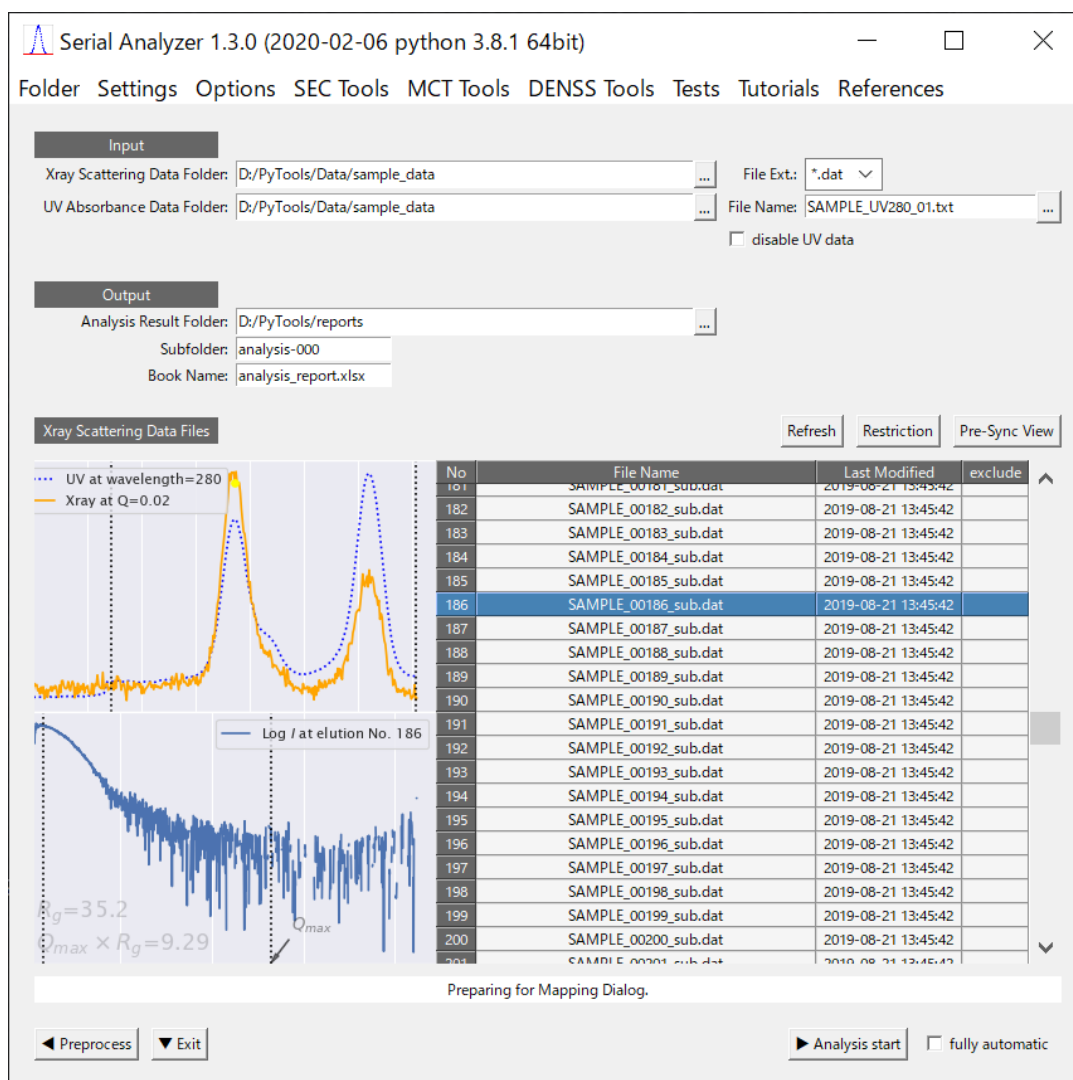
## 6 Detailed Description of the Dialogs

### 6.1 Main Dialog

#### 6.1.1 Purpose and Required Inputs

- a The Main Dialog is for entries of input / output folders and files as shown in Fig. 6-1.

**Fig. 6-1 Main Dialog**



- b For every folder or file entry, users can use any of the following three input methods, namely,
1. direct typing (including copy & paste),
  2. invoking folder/file-selection dialog by “...” button,
  3. drag and drop mouse manipulation
- c On entry in the “X-ray Scattering Data Folder”, the “UV Absorbance Data Folder” is automatically filled in if the UV absorbance file is in the same folder and properly named, and the file list below is filled.

- d The expected naming for UV absorbance file is like "...UV.txt" or "...spectra.txt". When the file naming is not an expected one, it fails to recognize and users must specify the folder or the file in the subsequent entries.
- e The "File Ext." entry is for specifying the file extension of X-ray scattering data files, which is usually supposed to be ".dat", and, otherwise, it can be changed to ".csv".
- f For the usage of "disable UV data" checkbox, see Section 6.1.6 below.
- g The output folders are to be specified hierarchically in "Analysis Result Folder" and "Subfolder" entries.
- h The "Subfolder" entry must include numbers as in the figure for automatic numbering, which increments the number if there already exists a file with the same name.
- i The "► Analysis start" button invokes a preliminary process and shows the "Mapping Dialog" described in the next section.<sup>10</sup>
- j The "Refresh" button is used when the file list must be updated in only such cases as there occurs no other trigger events to update the list.

#### 6.1.2 Quick Inspection of Data

- a At the lower left of the Main Dialog, two figures, called "Outline Figures", are shown to give a quick view of the X-ray data.
- b The upper figure shows the elution curve with a yellow selected point, and the lower figure shows the scattering curve with its R<sub>g</sub> value which corresponds to the selected point in the elution curve.
- c The row selection in file list at the lower right of the Main Dialog synchronizes with the point selection on the elution curve, and their synchronization is bi-directional.
- d The file list can be used as an interface to AutoGuinier program. See Section 6.15 for such usage.

#### 6.1.3 Data Range Restriction

- a In the outline figures mentioned above (Fig. 6-1), there may appear two vertical dotted lines for each figure, which indicate respective data ranges suitable for analysis.
- b The elution range is determined, by default, from the flow change points when they are observed.
- c The angle range start is determined, by default, to the extendable limit, in the small angle side, beyond the Guinier region start. The limit can be controlled using "R<sub>g</sub>-consistency" in the Settings Dialog. See section 6.8.4 for details.
- d The angle range end is determined, by default, to avoid the unusable high-Q part which comes from the flange, when it is observed.
- e When they seem to be badly determined by the program, you can change those ranges by

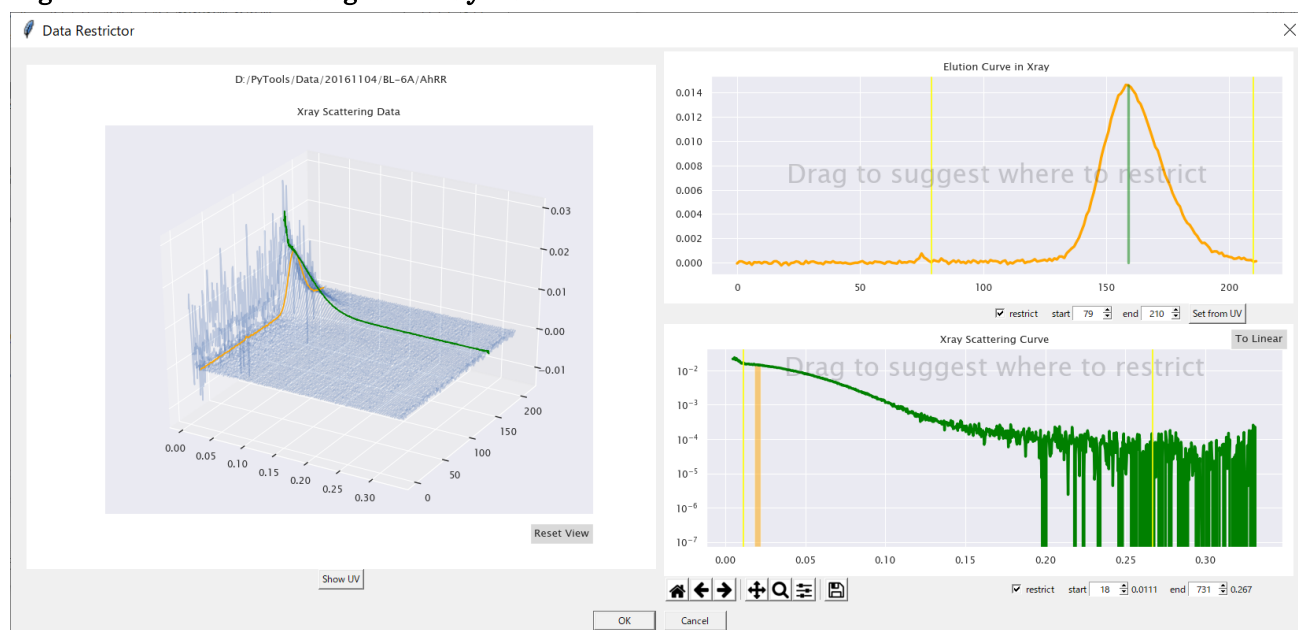
---

<sup>10</sup> You can change this auto-navigation in the Settings Dialog. See Section 6.8.

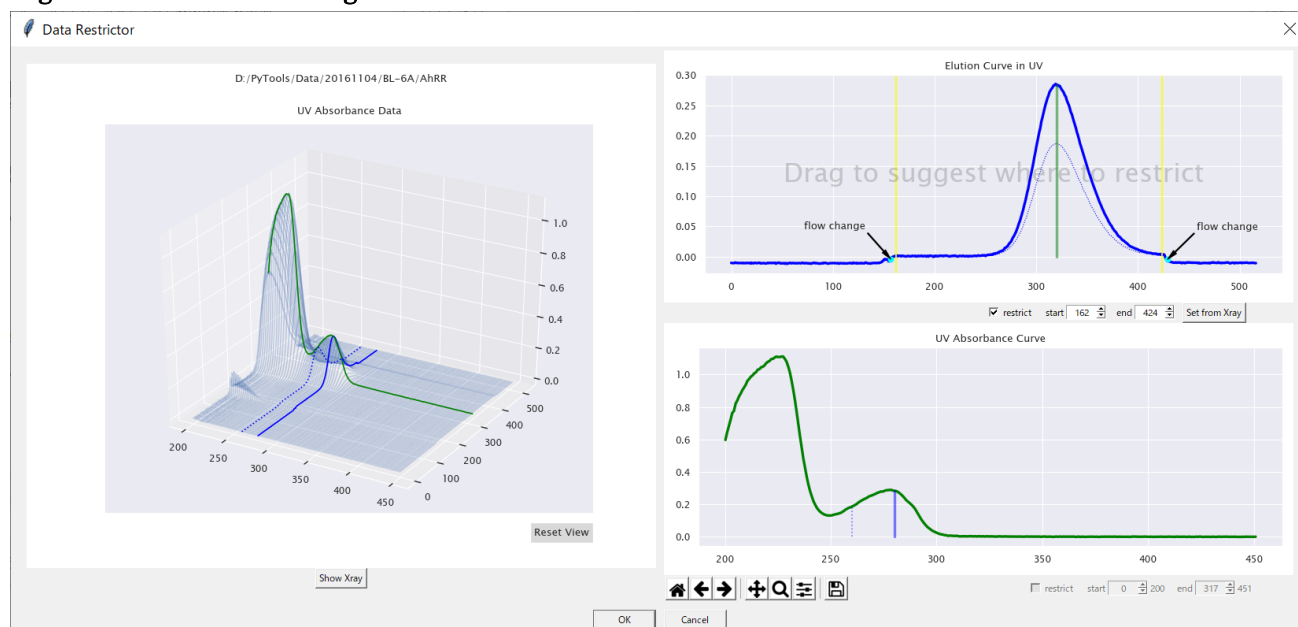
pressing the “Restrict” button, which shows the restrictor dialogs shown below. (Fig. 6-2, Fig. 6-3)

- f These paired dialogs take care of range restriction for X-ray and UV data respectively, and they can be toggled by the buttons, namely “Show UV” or “Show X-ray”, at the bottom of each 3D view figure.

**Fig. 6-2 Restrictor Dialog for X-ray Data**



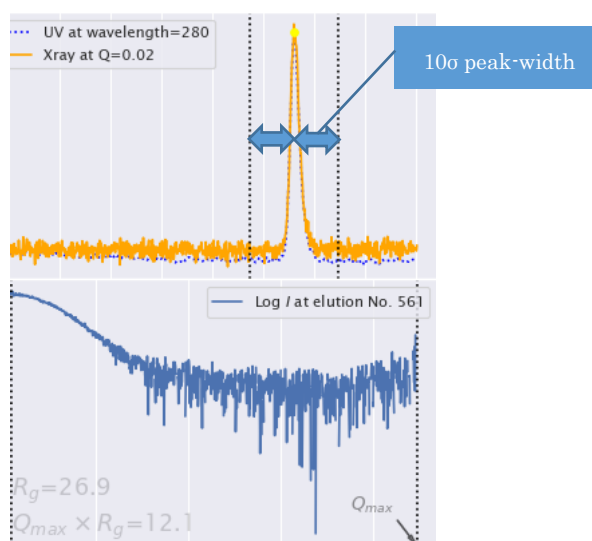
**Fig. 6-3 Restrictor Dialog for UV Data**



- g To change the restriction, drag your mouse roughly to set initial start-end values, and then use each spin button to modify.

- h When you have changed the range of one type of data, e.g., Xray (or UV) data, the range of the other type of data, i.e., UV (or Xray) data, can be easily changed using the “Set from Xray (UV)” button, which automatically calculates and sets the corresponding range from preliminary rough synchronization.
- i The range restriction for wave length values of UV data is not supported in this version.
- j As an additional note for elution restriction, non-peak part of the either side of the elution will be excluded when it extends beyond  $10\sigma$  peak-width, where “ $\sigma$ ” here should be understood to correspond to that of a gaussian elution.

Fig. 6-4 Restriction by  $10\sigma$  peak-width

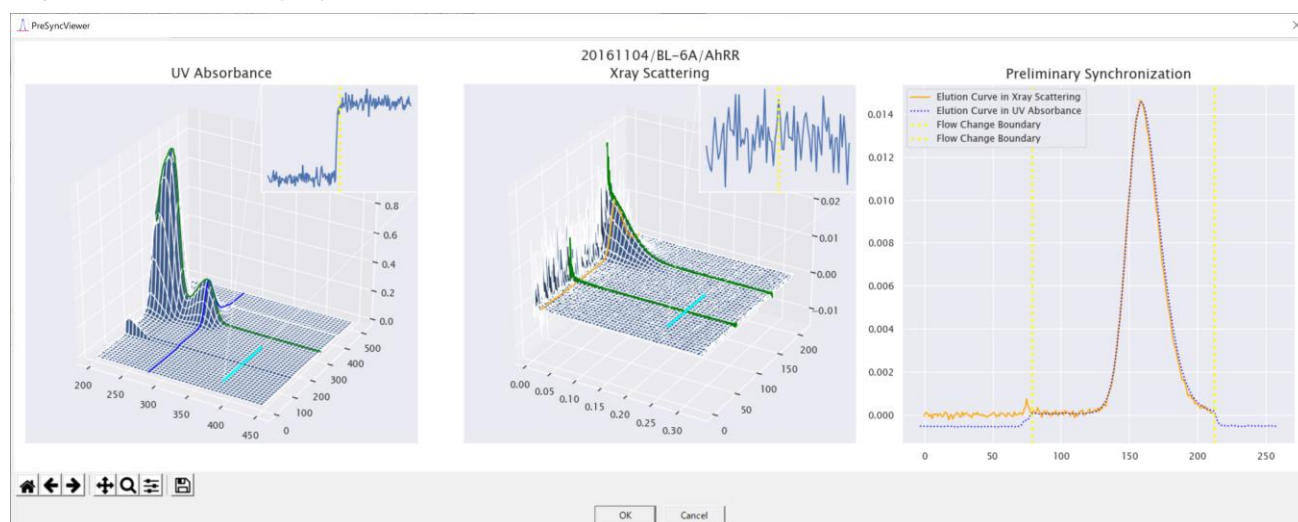




#### 6.1.4 Preliminary Synchronization View

- a This view, shown from “Pre-Sync View” button, is provided for users to observe both sets of data, namely UV absorbance and X-ray scattering, at a glance.
- b When flow change points are recognized, it also gives a closer view near the points in the inset box.
- c At the right is shown an overlaid figure of the elution curves, which is a result of preliminary synchronization process of both sets of data.

**Fig. 6-5 Preliminary Synchronization View**



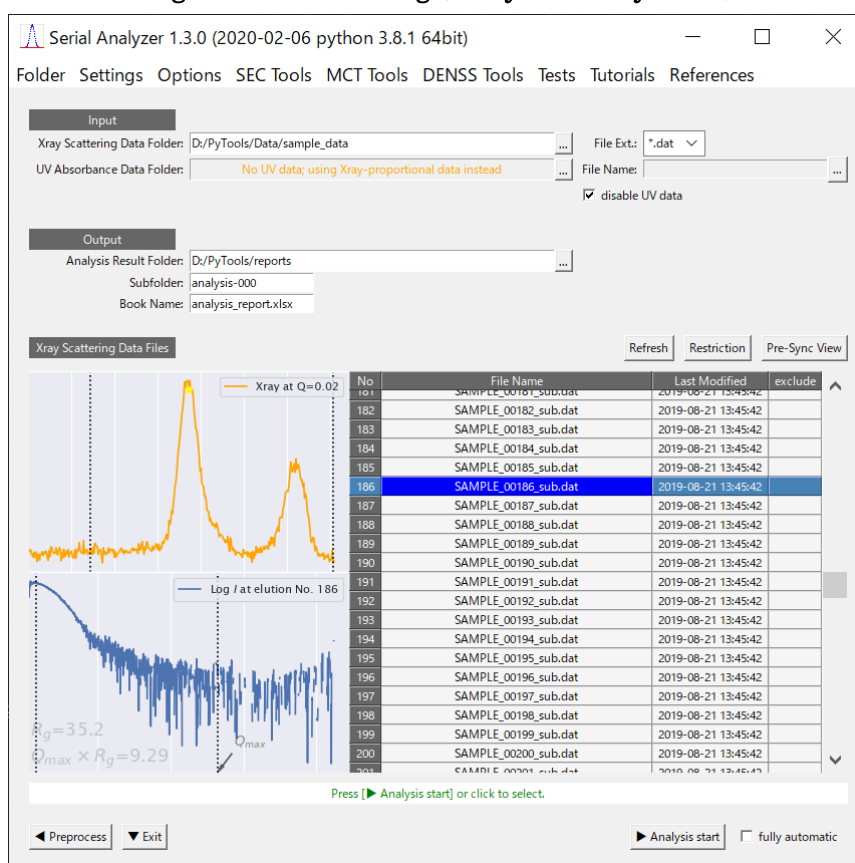
#### 6.1.5 Removing Abnormal Points

- a The “exclude” column in the data file table is used to exclude abnormal points. See Section 6.9.1 for details.

### 6.1.6 Toggling the Mode concerning UV Data Usage

- a In the standard usage, the program is supposed to analyze UV data paired with X-ray data.
- b Such standard mode is useful not only for the attachment of the absolute concentration, but also for the detection of different components from the deviation of the pair of elution curves.
- c However, there can be cases where the program fails to establish a precise mapping between those elution curves.
- d If such is the case, or the other merits of UV data usage are not considered important, users can choose the X-ray data only mode by checking the “disable UV data” button, which changes the Main Dialog as shown in Fig. 6-6 below.
- e It is toggled between the two modes by checking and unchecking the button.
- f See Chapter 7 for details of the X-ray data only mode.

**Fig. 6-6 Main Dialog (X-ray data only mode)**

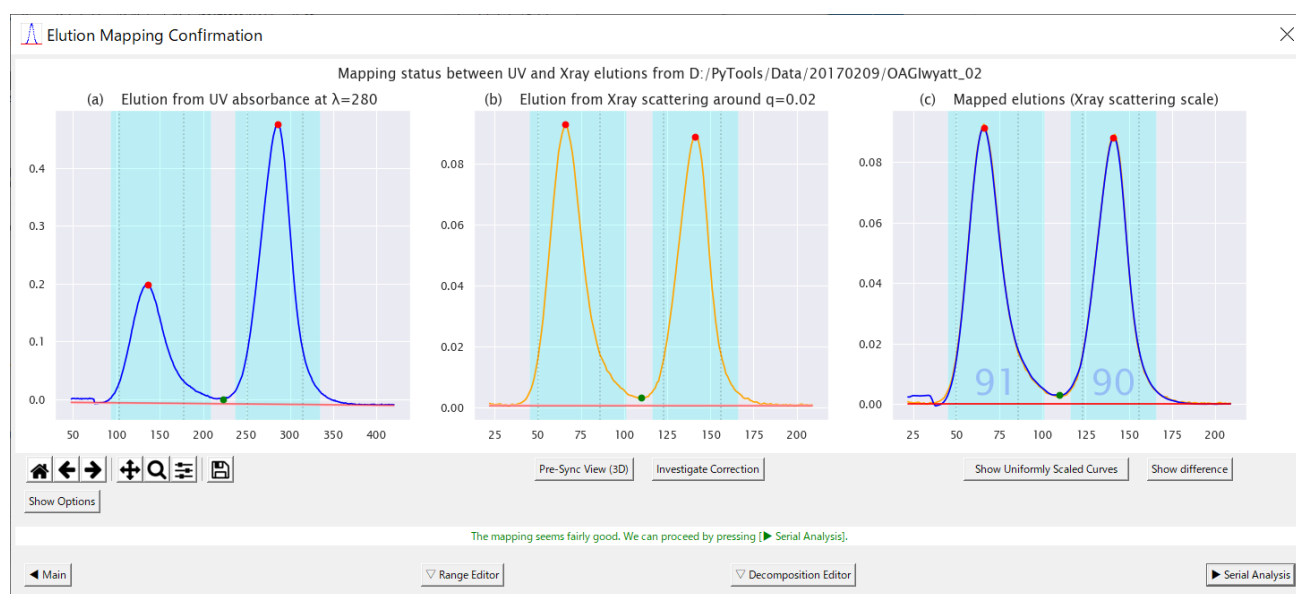


## 6.2 Mapping Dialog

### 6.2.1 Purpose and Manipulation Flow

- g This dialog, as shown in Fig. 6-7, is for checking and adjusting the mapping between the UV absorbance data and the X-ray scattering data.
- h It is invoked either by the “► Analysis start” button in Main Dialog or by the “◄ UV-Xray Mapping” button in Analyzer Dialog.
- i The figures in the upper part of the dialog show the elution curves of both data in the standard mapping plane, which is usually set at the points where wavelength  $\lambda=280\text{nm}$  for UV absorbance and scattering vector length  $q=0.02\text{\AA}^{-1}$  for X-ray scattering.

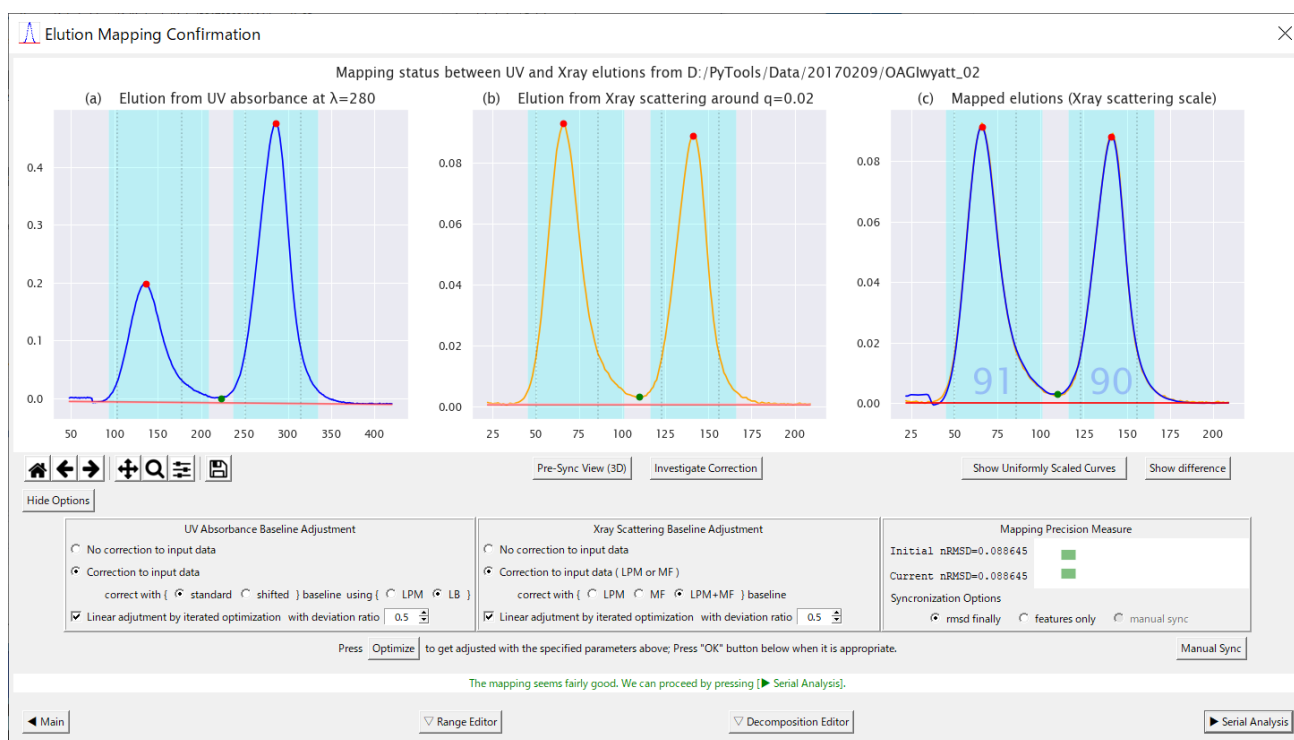
**Fig. 6-7 Mapping Dialog when options are hidden**



- j The lower part of the dialog is hidden if the mapping seems adequate and there seems no need of baseline correction of the X-ray scattering data.
- k It is shown otherwise or by clicking the “Show Options” toggle button. (See Fig. 6-8)
- l For usage of the toolbar (🏠◀▶🔍📏📄), please refer to the following page in matplotlib document.

Interactive navigation [https://matplotlib.org/users/navigation\\_toolbar.html](https://matplotlib.org/users/navigation_toolbar.html).

**Fig. 6-8 Mapping Dialog when options are shown**



### 6.2.2 Checking the Mapping Adequacy

- a Users can check the adequacy of mapping by observing the overlaying state of curves in the right figure or by reading either of the following information indicating the mapping quality.
  1. SCI — Single Component Indicator — displayed in percent values in the figure,
  2. nRMSD — Normalized Root Mean Square Deviation — shown in the lower right box.
- b SCI, in percent values, indicates the ratio of the length of interval where both lines fit well against the whole length of each analysis range. (See another example in Fig. 6-9 on the next page)
- c For nRMSD, there are two indicators for the initial value and the latest (or current) value which varies as the user changes the mapping options and parameters.
- d The color of each indicator bar changes among cyan, yellow and pink depending on the adequacy, as shown in Tab. 6-1 below.

**Tab. 6-1 Colored indication of mapping adequacy**

Color	Adequacy indication
Cyan	Ok
Yellow	Better be improved
Pink	Bad (need adjustment)

**Fig. 6-9 Another Mapping Example where the Deviation is significant between the mapped Curves**



### 6.2.3 Data Investigation

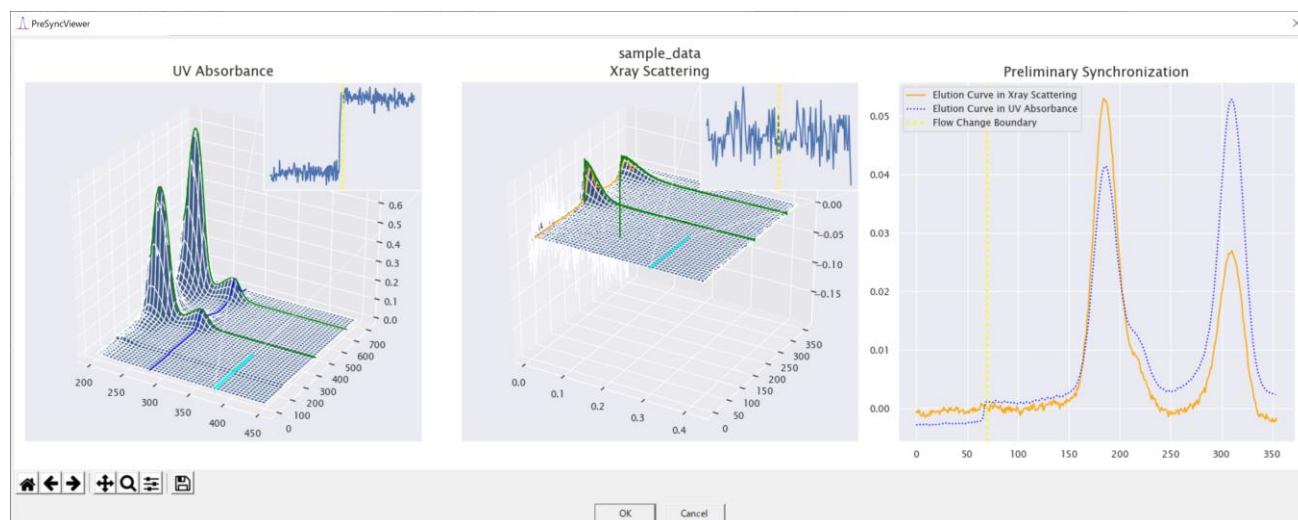
- a There are four buttons for data investigation and range adjustment summarized in the table below.

**Tab. 6-2 Buttons for data investigation and range selection (from left to right)**

Name	Description	Figure
Pre-Sync View (3D)	Shows both sets of data in 3D plot	Fig. 6-10
Investigate Correction	Shows X-ray scattering data at each Q	Fig. 6-20
Show uniformly scaled curves	Shows a uniformly scaled UV curve overlaid	Fig. 6-11
Show difference	Shows difference of the mapped elution curves	Fig. 6-12

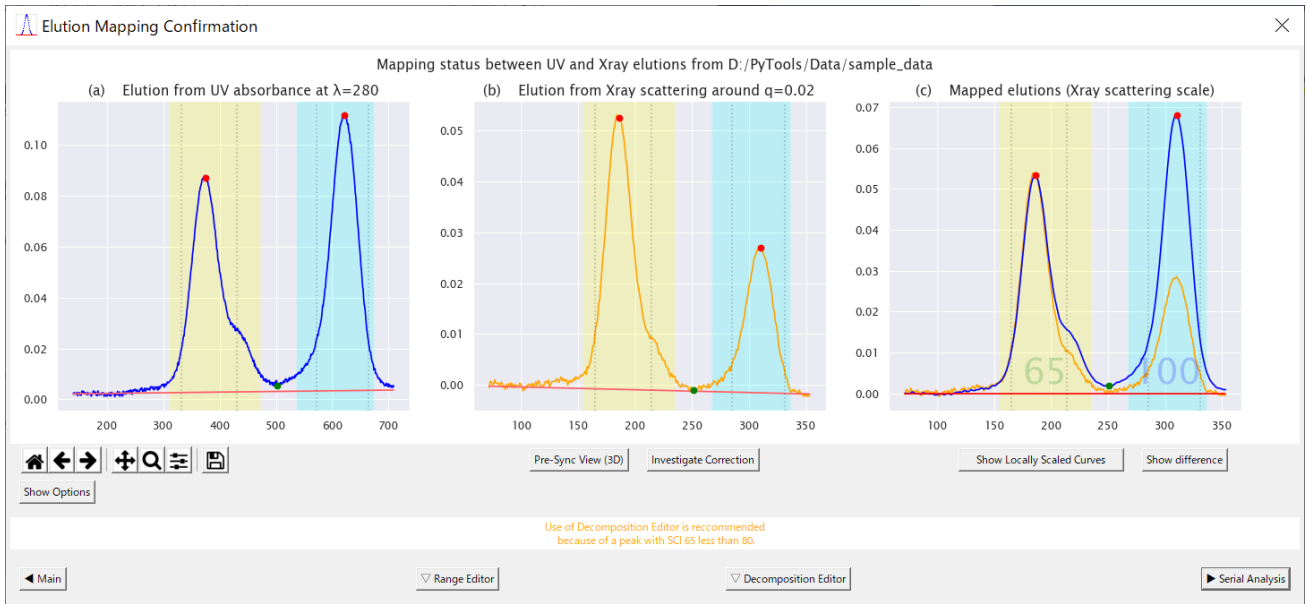
- b The 3D-plots are sometimes required to observe the situations in other parts of the data, not apparent in the 2D figures within the standard mapping plane.
- c Another purpose of the “Pre-Sync View” is to observe the entire sets of data before restriction. (See also Section 6.1.4)

**Fig. 6-10 Preliminary Synchronization View**

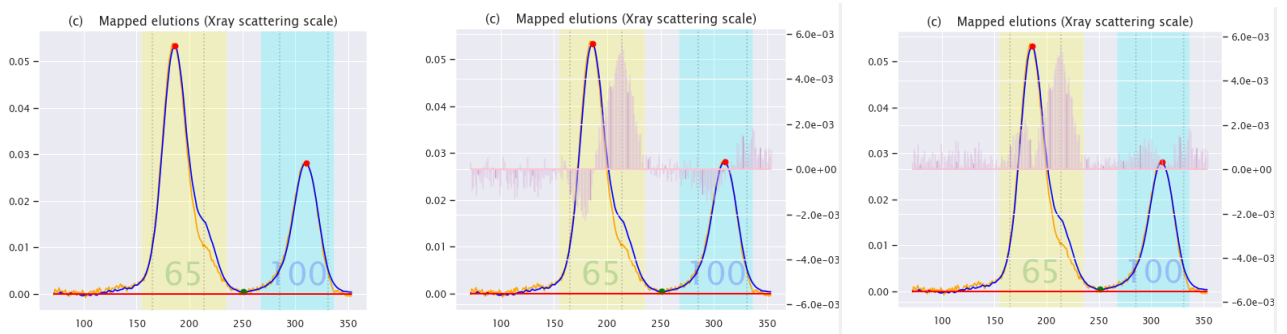


- d As for the “Investigate Correction” button, see Section 6.2.11.
- e The uniformly scaled curves, shown in Fig. 6-11, make clearer the fact that they need locally adjusted scaling for each peak in order to overlap.
- f The differences between the overlapped curves can be shown using the “Show difference” button, which shows them in two ways, namely the differences and their absolute values. (Fig. 6-12)

**Fig. 6-11 Mapping Dialog with Uniformly Scaled Curves**



**Fig. 6-12 Three states toggled by “Show difference” button**



## 6.2.4 Attracting User's Attention

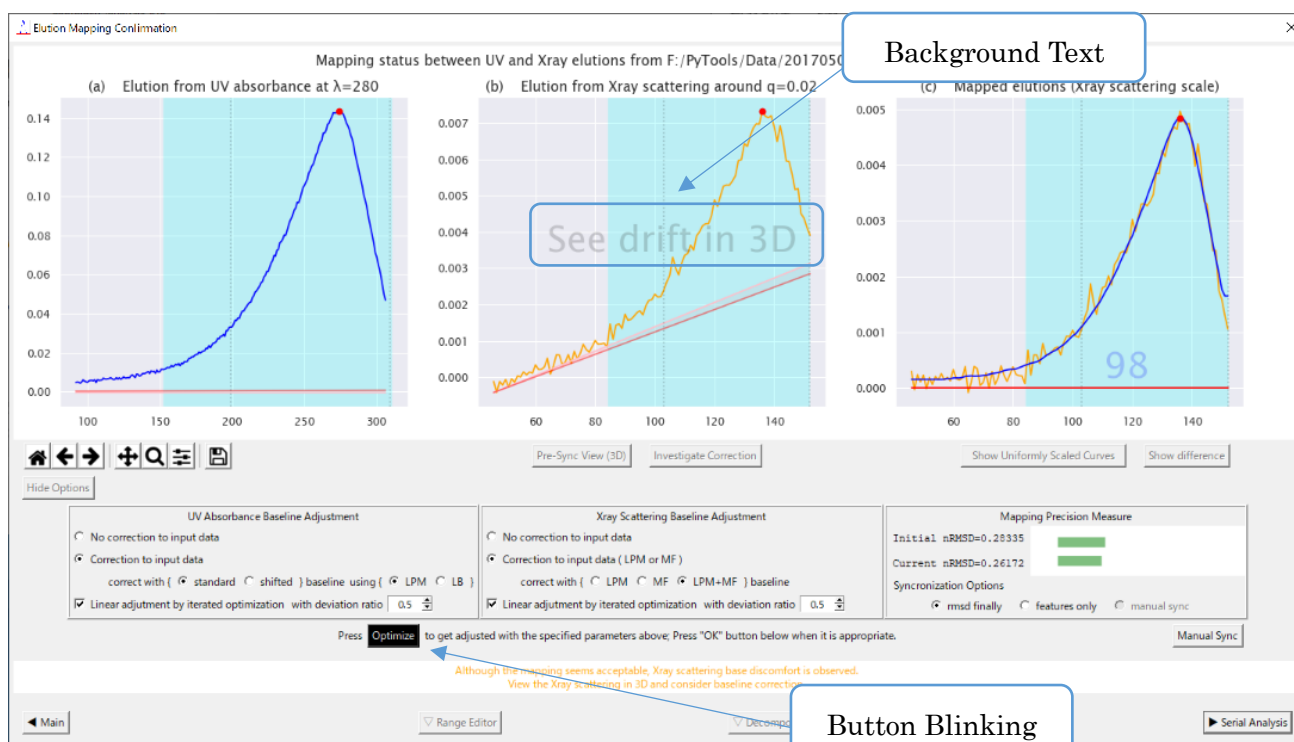
- a The dialog tries to attract user's attention by button blinking or background text messages in cases listed in Tab. 6-3. See Fig. 6-13 for such an example.

**Tab. 6-3 Situations requiring user's attention**

Attention Sign	Situation
"See Drift in 3D" Message	When the program has detected significant baseline drift in X-ray scattering data and wants users to confirm in 3D plot.
Optimize Button Blinking	When the figures and the mapping precision indicators are inconsistent with the option/parameter settings after they have been changed, until they are updated by the "Optimize" button.

- b Users may ignore such suggestions if they are confident that there is no need to address to them.
- c While the "Optimize" button is blinking, pressing "► Serial Analysis" button is judged as an illegal action in order to avoid allowing possibly unintended setting without confirmation, and guided to go back to "◀ Main" instead of "► Serial Analysis" in order to quit.

**Fig. 6-13 Mapping Dialog with Button-Blinking and Background Text**





### 6.2.5 Baseline Correction Hierarchy

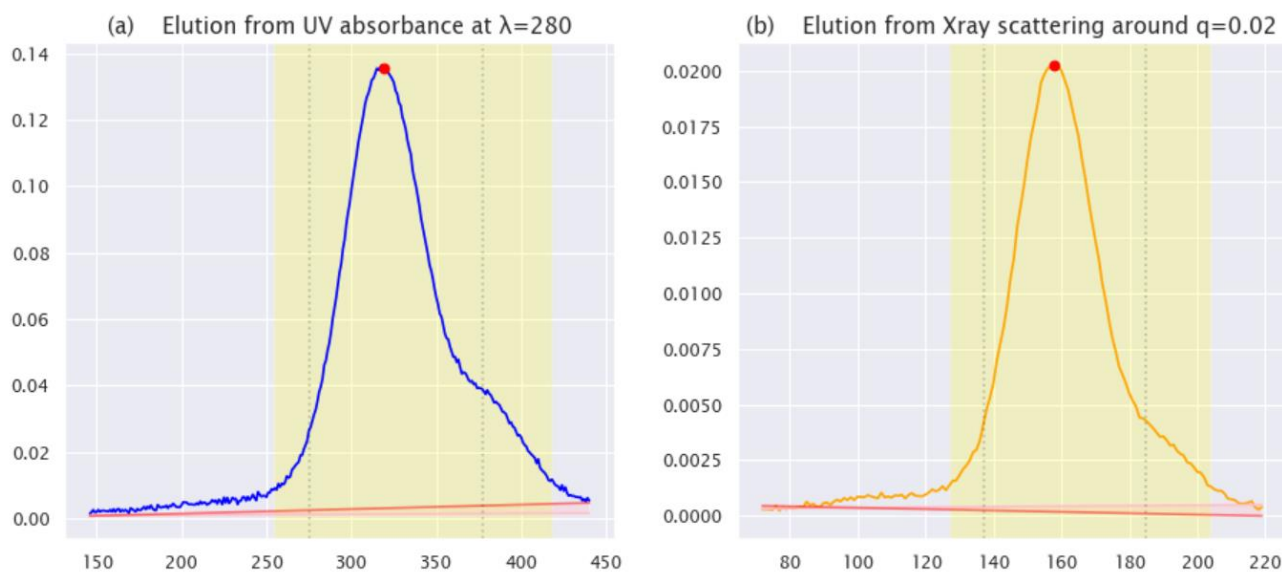
- a For both of UV and X-ray data, baseline correction is implemented similarly in the following two hierarchical ways.
  1. Correction without adjustment
  2. Correction with adjustment
- b The former uses only one side of data, namely, UV only or X-ray only.
- c The latter calculates additional adjustment amounts, comparing both of the former's correction results.
- d Users can choose any combination of these variations as shown in the figure below.

**Fig. 6-14 Baseline Correction Options (in default selection)**

UV Absorbance Baseline Adjustment	Xray Scattering Baseline Adjustment
<input type="radio"/> No correction to input data <input checked="" type="radio"/> Correction to input data correct with { <input checked="" type="radio"/> standard <input type="radio"/> shifted } baseline using { <input type="radio"/> LPM <input checked="" type="radio"/> LB }	<input type="radio"/> No correction to input data <input checked="" type="radio"/> Correction to input data ( LPM or MF ) correct with { <input type="radio"/> LPM <input type="radio"/> MF <input checked="" type="radio"/> LPM+MF } baseline
<input checked="" type="checkbox"/> Linear adjustment by iterated optimization with deviation ratio <input style="width: 50px;" type="text" value="0.5"/>	<input checked="" type="checkbox"/> Linear adjustment by iterated optimization with deviation ratio <input style="width: 50px;" type="text" value="0.5"/>

- e Therefore, there are, in all,  $9 = 3 \times 3$  patterns of combination if “No correction” options for both data are counted.
- f When adjustment options for both data are selected, users can specify the “deviation ratios” to control the balance of deviations from the “corrections without adjustment”.
- g The deviation ratios are automatically controlled to sum up to 1. In other words, if you change one of the ratios from 0.5 to 0.7 for example, the other is automatically updated to 0.3.
- h Corrected baselines and adjustment amounts are shown in red lines and pink belts in the figures (See Fig. 6-15).
- i Detailed descriptions of this section are provided in the following three sections.

**Fig. 6-15 Corrected Baselines and Adjustment Amounts shown in the figures**



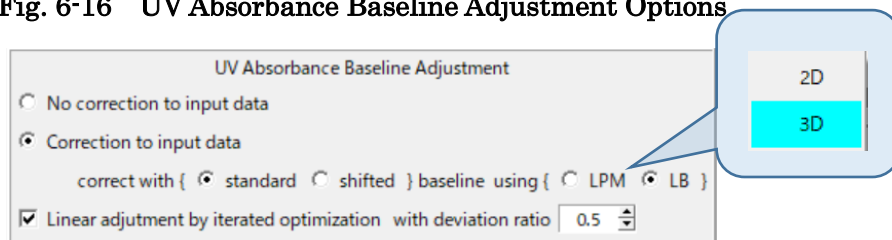
#### 6.2.6 Note on the differences of use ranges of UV and X-ray data

- a UV and X-ray data have similar 3D structures as we have seen in the 3D-plots.
- b However, we note here that their use ranges in this program are different.
- c This fact is important to understand the differences in handling of both data and application of corrections described below.
- d For the subsequent processing, the use range of UV absorbance data is restricted to the mapping plane to attach concentration data to corresponding scattering profiles. Therefore, the correction is applied only to the elution curve in the plane.
- e On the other hand, X-ray data are, of course, fully used. Therefore, the baseline correction is applied to all elution curves.
- f Since the mapping between UV and X-ray data is established only on the standard mapping plane, the adjustment amount is conveyed to other planes using affine transformation. See Chapter 8 for details.

### 6.2.7 UV Absorbance Baseline Correction Options

- a With “No correction” option, input data are used without any correction.
- b “Correction to input data” option has two kinds of variations, namely, “standard” and “shifted”. See Chapter 8 for details.
- c Under the default setting, this correction method (LB or LPM) is automatically selected to give a better mapping. Select the other and re-optimize if it seems inappropriate. The difference between these methods is similar to that of LPM and MF in the next section. See sections 6.2.9 and 6.2.10 for explanation.
- d For UV data only, there are two options for LPM, namely 2D or 3D, and you can choose between them by right-clicking of the LPM button as shown in the bubble in the figure below. As for the difference of the two options, the baseline is determined using the UV elution curve data for 2D, whereas for 3D, data from a wider range of UV wavelengths are used.
- e Linear adjustment is determined to minimize the RMSD between the mapped elution curves, using the deviation ratios to divide the difference proportionately.
- f This deviation ratio is paired with that of X-ray scattering, and on change of one, the other is automatically updated so that they sum up to 1.

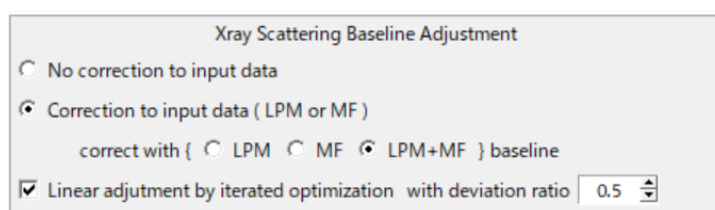
**Fig. 6-16 UV Absorbance Baseline Adjustment Options**



### 6.2.8 X-ray Scattering Baseline Correction Options

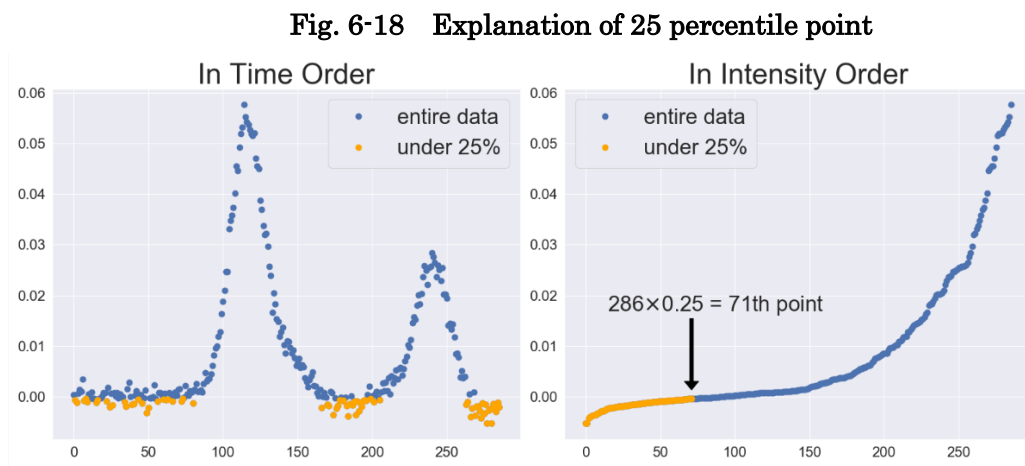
- a With “No correction” option, input data are used without any correction.
- b “Correction to input data” has three options, namely, “LPM”, “MF” and “LPM+MF”. See the next two sections for these meanings.
- c Linear adjustment is determined to minimize the RMSD between the mapped elution curves, using the deviation ratios to divide the difference proportionately.
- d This deviation ratio is paired with that of UV absorbance, and on change of one, the other is automatically updated so that they sum up to 1.

**Fig. 6-17 X-ray Scattering Baseline Adjustment Options**

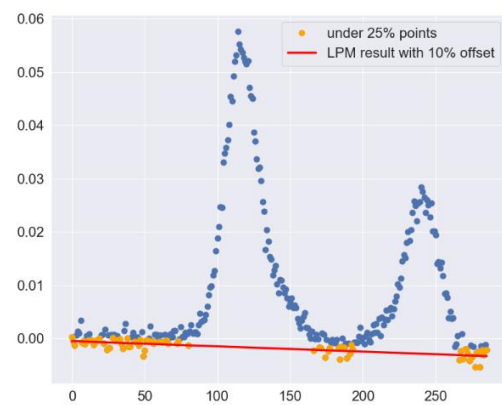


### 6.2.9 Low Percentile Method

- a The first step of baseline correction is performed according to the following algorithm, which we call “Low Percentile Method”.
1. Take a low percentile (e.g., less than 25%) part of the elution curve. (Fig. 6-18)
  2. Using linear regression, determine a line representing that part.
  3. Subtract the determined line, which is considered as an approximate baseline, from the elution curve.
  4. To get a better slope, repeat the above steps until it converges to a certain state.
  5. To get a better vertical position, shift the line so that it pass through an optimal percentile (e.g., 10%) point, which should be determined according to the relative width of the peaks and the noise level of the curve.<sup>11</sup>
  6. Apply the above to all elution curves.



**Fig. 6-19 Baseline determined at 10 percentile point**



<sup>11</sup> The optimal percentile is larger if the curve is noisier or the baseline inclusion rate is smaller. The exact relation can be estimated using Monte Carlo simulation, assuming Gaussian curve and noise.

### 6.2.10 Matrix Factorization Adjustment

- a The Low Percentile Method described above is considered appropriate when the elution curve includes enough baseline data.
- b On the other hand, we have another expectation that the baseline should be determined by the law, such as Lambert-Beer law concerning absorbance or the equivalent (first order) approximation in X-ray scattering.
- c And the law is expected be safely independent of baseline inclusion rate of the data in contract to the harder requirements for LPM.
- d Fortunately, the following minimization formulation gives a stable solution when the base surface can be seen as a plane, where the meanings of symbols are defined in the table below.

$$\operatorname{argmin}_{a,b,c} \left[ \|P_{a,b,c} \cdot C_{a,b,c} - M_{a,b,c}\|_F^2 \right] \text{ or} \quad (6-1)$$

$$\operatorname{argmin}_{a,b,c} \left[ \|M_{a,b,c}[* , j] \cdot M_{a,b,c}[i , *] - M_{a,b,c}\|_F^2 \right] \quad (6-2)$$

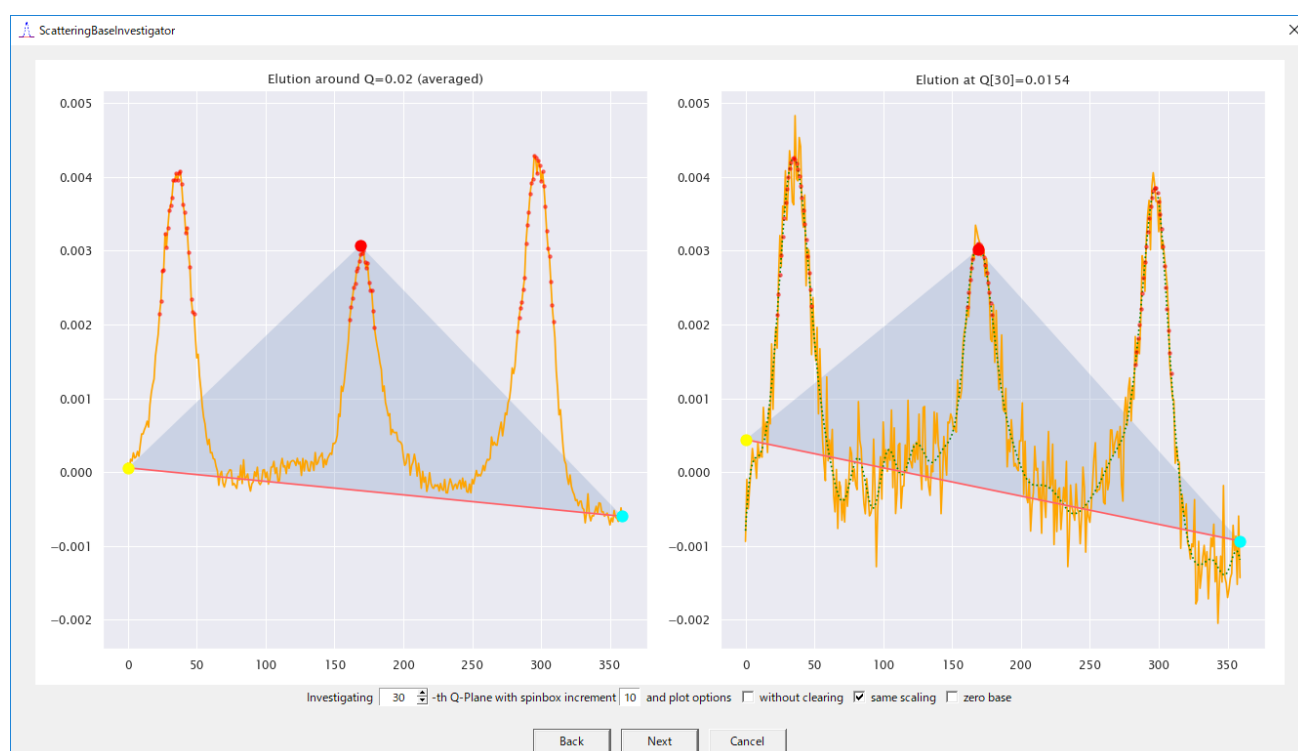
Symbol	Definition
$a, b, c$	Parameters to define a base plane using a formula $z = ax + by + c$ in xyz 3D-space or $B[i, j] = ai + bj + c$ in matrix base data, which also suggest the subscripted symbols below are interpreted as adjusted (subtracted) by the plane $B[i, j]$ .
$P_{a,b,c}$	Matrix consisting of column vectors of scattering (absorbance) curve, which is a part of $M_{a,b,c}$ , e.g., the $j$ -th column $M_{a,b,c}[* , j]$ .
$C_{a,b,c}$	Matrix consisting of row vectors of elution curve, which is a part of $M_{a,b,c}$ , e.g., the $i$ -th row $M_{a,b,c}[i , *]$ .
$M_{a,b,c}$	Matrix of the entire measured data.
$\cdot$	Matrix multiplication.
$\  \quad \ _F^2$	Square of a Frobenius norm of a matrix.

- e Therefore, we can use the plane determined by the above formulation as a base surface to correct the measured data.
- f We can assume that the above requirement — the base should be a plane — can be fulfilled for the set of data, to which we have applied LPM to all of its elution curves respectively.
- g Using an acronym “MF” for the above matrix factorization, this treatment is identified by “LPM+MF” in the correction options.
- h In place of MF, the acronym LB from Lambert-Beer is used for UV absorbance data, meaning the same solution.

### 6.2.11 X-ray Scattering Baseline Investigation

- a As mentioned in Section 6.2.6, baseline correction is applied to all elution curves of X-ray scattering data.
- b Therefore, users are concerned on the states of baseline correction in planes other than the standard mapping plane.
- c To investigate those states, use “Investigate Correction” button to show such figures shown below.

**Fig. 6-20** Affine transformation of adjustment from the standard plane to others



- d In this investigation dialog, the left figure shows the state of standard mapping plane and the right figure shows that of another plane, which can be moved to any desired position using the buttons.
- e The affine transformation of adjusted amount is described in Chapter 8.

### 6.2.12 Synchronization Control

- a Although the program tries its best to deduce the most likely mapping between UV and X-ray elutions, the result may contradict to user's intuition or may be incorrect due to bugs.
- b Here are some recipes to cope with such situations.
- c To understand, note that the synchronization process is roughly made of the following two steps.
  1. Feature point matching

2. Minimization of the RMSD

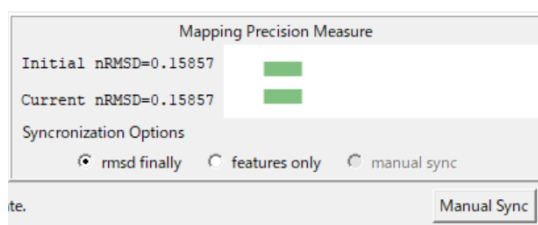
d Keeping the above in mind, users can choose from the three options in the table below.

**Tab. 6-4 Synchronization Options**

Option Name	Step1	Step 2	Description
rmsd finally	<input type="radio"/>	<input type="radio"/>	Adopt the result of step 2 after conducting step 1
features only	<input type="radio"/>		Adopt the result of step 1 only.
manual sync			Ignore the program results and adopt the parameters determined by “Manual Synchronization Dialog”

e Before the first posting of the dialog, the program compares the results of the first two options and shows the result in the “Mapping Precision Measure” box at lower right. (Fig. 6-21)

**Fig. 6-21 Radio Buttons for Synchronization Control**

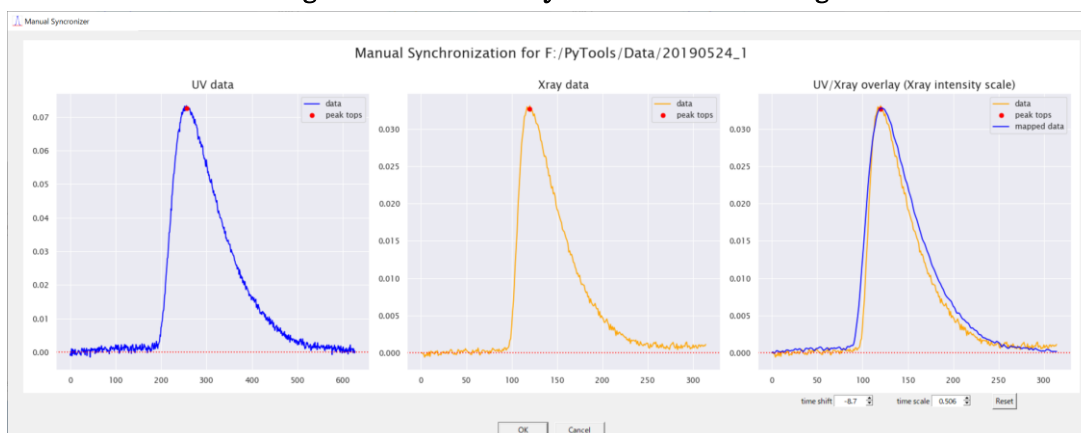


f Users can choose another option with those radio buttons.

g “manual sync” is available only after appropriate parameters are given at “Manual Synchronization Dialog”, which can be invoked by pressing “Manual Sync” button.

h Users are supposed to give two parameters, namely “time shift” and “time scale”, which correspond to the intercept and slope of the temporal linear mapping of elution numbers.

**Fig. 6-22 Manual Synchronization Dialog**

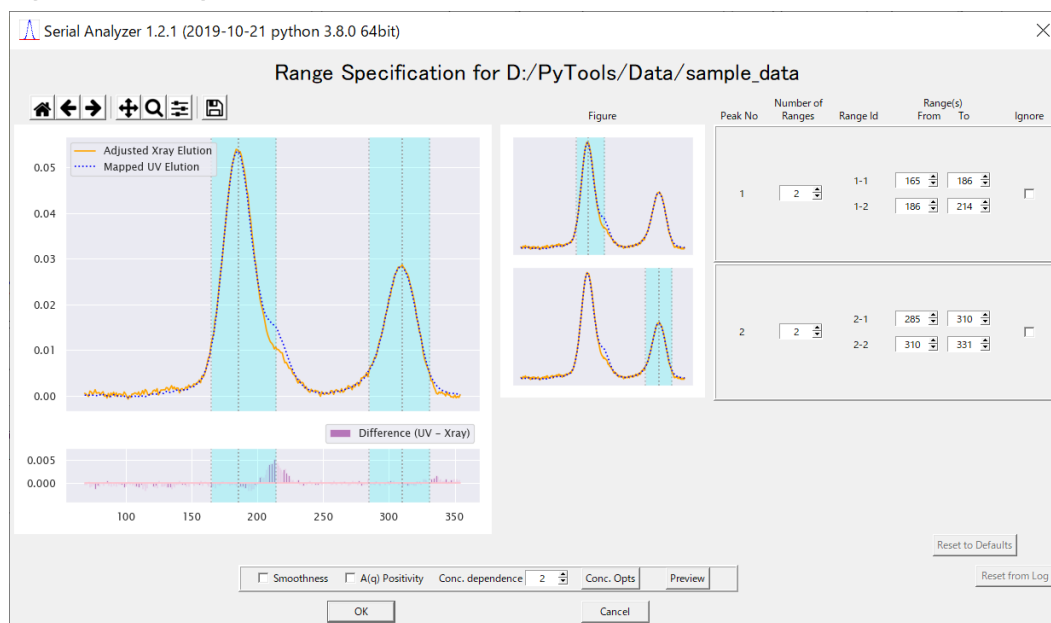


## 6.3 Range Editor

### 6.3.1 Purpose and Manipulation Flow

- The default analysis ranges, depicted with cyan bars in Fig. 6-23 below, are automatically determined by the program, and this dialog is for changing them.
- It is invoked by the “▽ Range Editor” button in the Mapping Dialog.
- As for the “Preview” button with the preceding options, except “Conc. Opts” button described soon below, see Section 6.7.

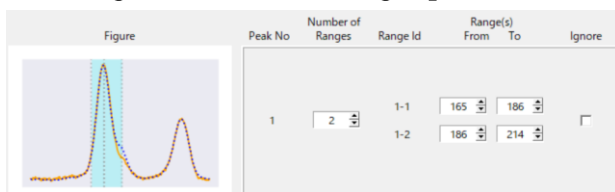
**Fig. 6-23 Range Editor**



### 6.3.2 How to Change Analysis Ranges

- “Number of Ranges” per peak can either be one or two depending on your needs.
- In other words, you should select “double ranges” if you want separate extrapolation results for each side of the peak, otherwise select “single range”.
- Using this spin box, you can select either of these numbers, as shown in the following figures.

**Fig. 6-24 Double Ranges per Peak**



**Fig. 6-25 Single Range per Peak**



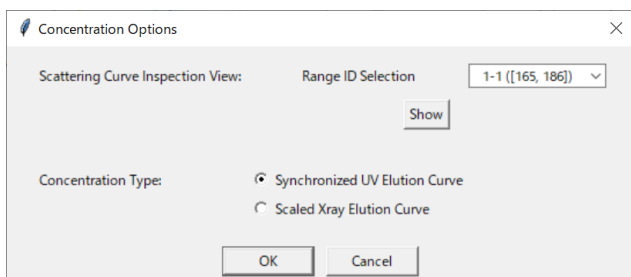
- To change the width of each range, use the spin boxes corresponding to the end points of the range.



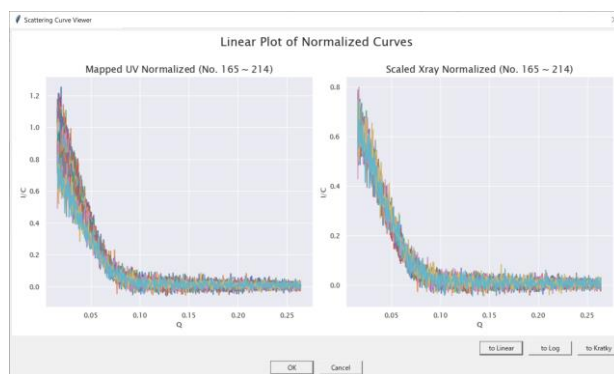
### 6.3.3 Concentration Options

- a In the current version of the program, the elution curve used as concentration is selectable.
- b By default, the synchronized UV elution curve is used for that purpose.
- c However, such default can be inappropriate when it is hard for the program to find good parameters which map the UV and X-ray data resulting in small enough differences.
- d An alternative solution is to use a “scaled X-ray elution curve”, which is made, peak by peak, from an X-ray elution curve multiplied by a peak dependent scale calculated from the height ratio between the UV and X-ray curve peaks.
- e To select the alternative, press “Conc. Opts” button to show “Concentration Options Dialog” shown in Fig. 6-26 below.
- f To observe the difference of these two options, three types of figures, shown in the figures below, are provided

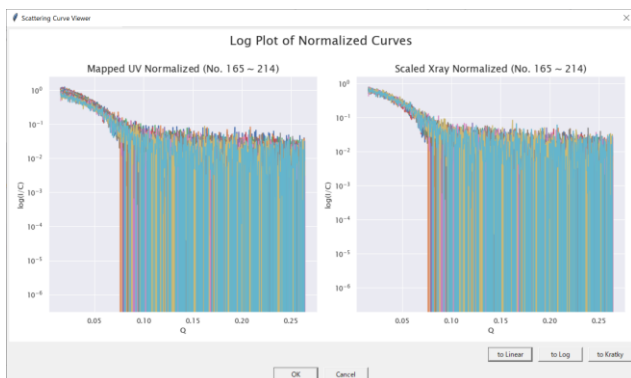
**Fig. 6-26 Concentration Options Dialog**



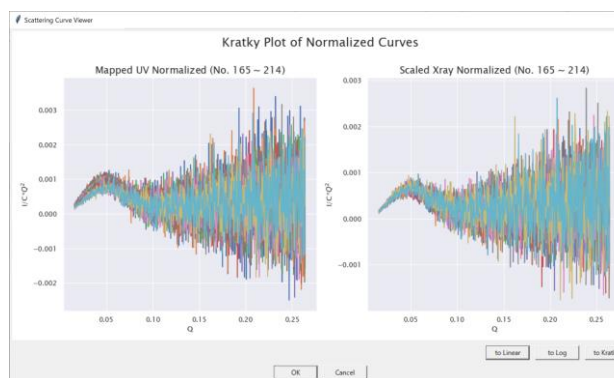
**Fig. 6-27 Linear Plot of the Curves**



**Fig. 6-28 Log Plot of the Curves**



**Fig. 6-29 Kratky Plot of the Curves**

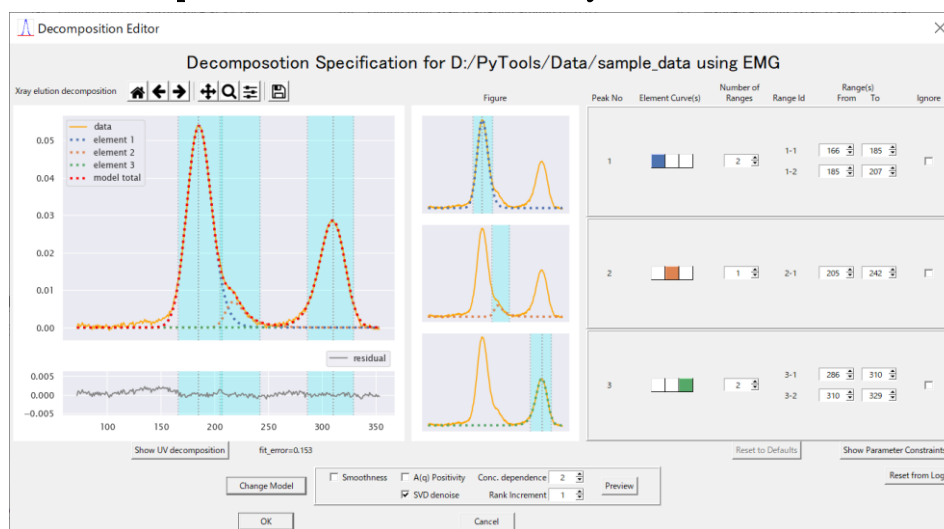


## 6.4 Decomposition Editor

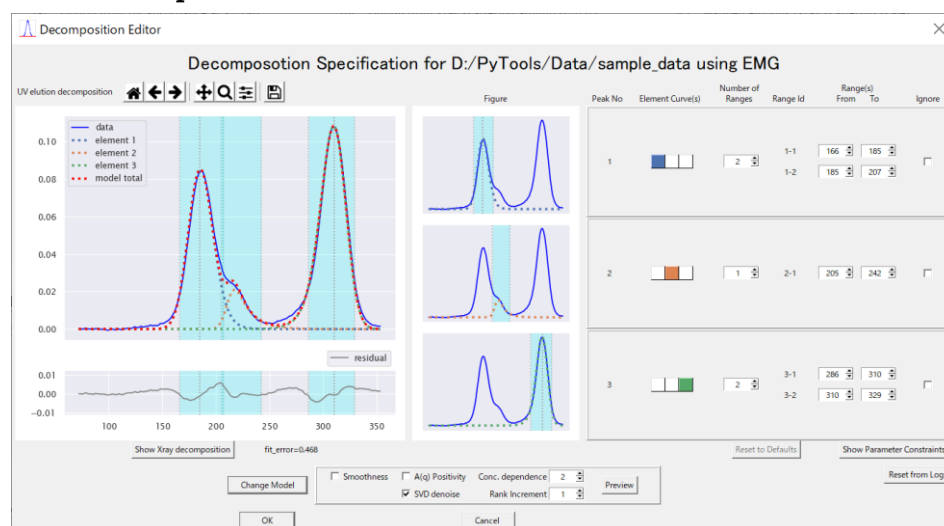
### 6.4.1 Purpose and Manipulation Flow

- This dialog, as shown in Fig. 6-30 and its name suggests, is provided in order to better control the decomposition of the total elution curve into the separate curves of its components.
- For this purpose, elution curve models such as EMG<sup>12</sup> or EGH<sup>13</sup> are employed.
- It is invoked by the “ $\nabla$  Decomposition Editor” button in the Mapping Dialog.
- Alternatively, you can set the “Dialog Navigation Preferences” in the Settings Dialog (see Section 6.8) to let it automatically navigate you along to this editor.
- As for the “Preview” button with the preceding buttons, see Section 6.7.

**Fig. 6-30 Decomposition Editor (EMG with X-ray data)**



**Fig. 6-31 Decomposition Editor (EMG with UV data)**



<sup>12</sup> Exponentially Modified Gaussian

<sup>13</sup> Exponential Gaussian Hybrid

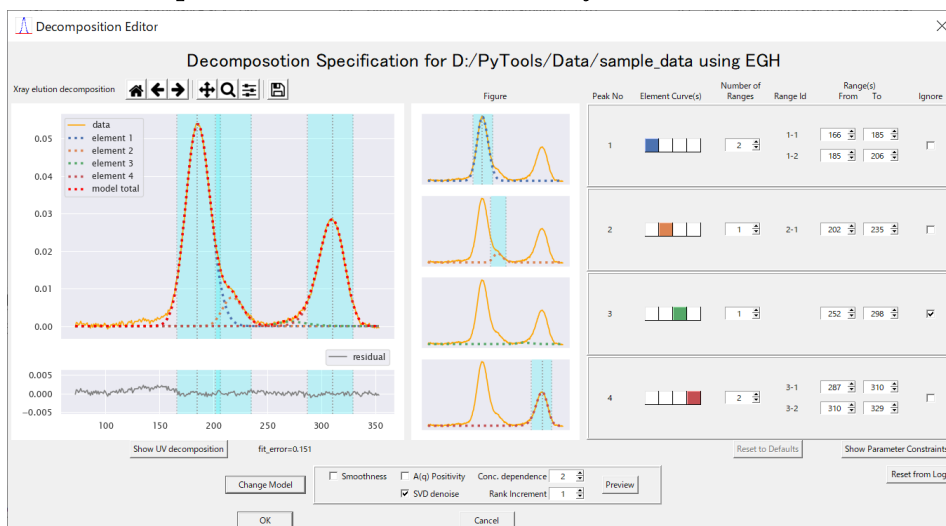
### 6.4.2 Two Types of Data by Two Types of Models

- In the editor, you can confirm the state of decomposition in two views, namely X-ray view and UV view, and, for each view, you can select elution models between EMG and EGH.
- Therefore, in all, you have four types of views to choose which model is better.
- Confirm the meaning of the above statement and observe the slight differences among them by the following table and figures (Fig. 6-30, Fig. 6-31, Fig. 6-32, Fig. 6-33).

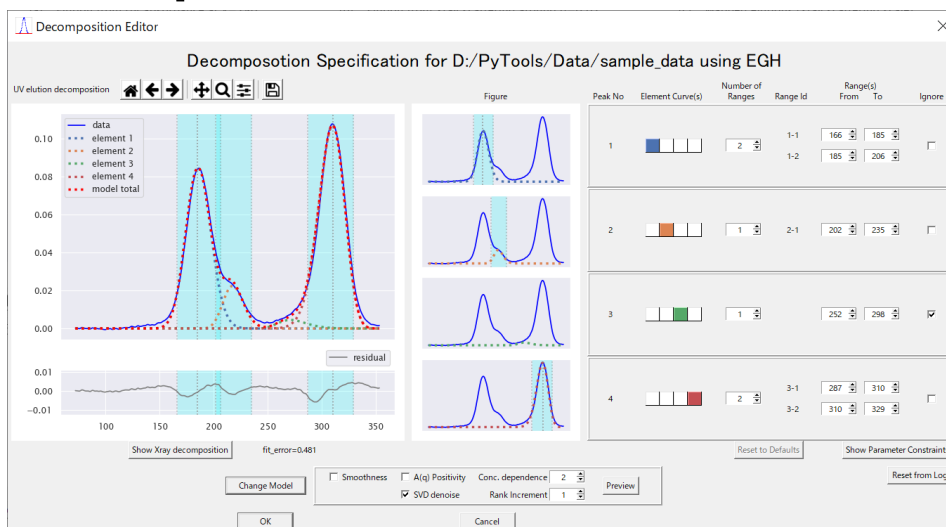
**Tab. 6-5 Four Types Views in Decomposition Editor**

View Type No.	Data	Model
1	X-ray	EMG
2	UV	EMG
3	X-ray	EGH
4	UV	EGH

**Fig. 6-32 Decomposition Editor (EGH with X-ray data)**



**Fig. 6-33 Decomposition Editor (EGH with UV data)**



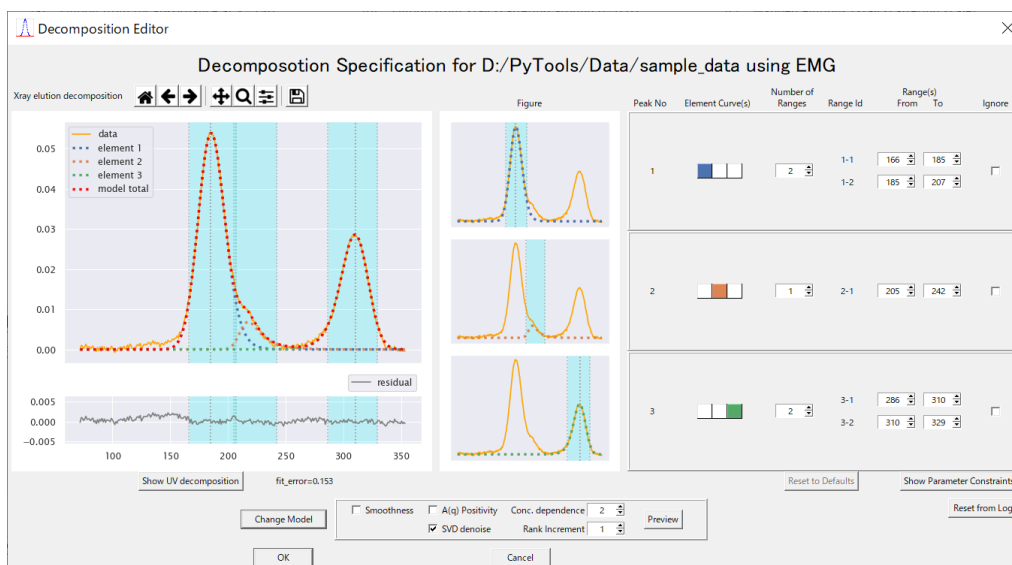
### 6.4.3 Components, Peaks and Elution Elements

- a To explain how to edit, we distinguish confusing concepts using the following definitions.
  1. component: a substance consisting of a single kind of molecules
  2. peak: a peak in an elution curve, often (but not always) expected to be made from a single component in SEC-SAXS
  3. elution element: a set of data modeled by a single elution model
- b Note that one peak with a single component may be modeled using either a single elution element or multiple elution elements.<sup>14</sup>
- c In the example shown in the above figures, there observed are four elution elements identified by colors: ■, ■, ■, ■.

### 6.4.4 Distinction of Peak Significance and Number of Ranges

- a For the moment, forget the above figures and suppose all the elements in the above example make up separate peaks respectively.
- b Such situation is shown in the following figure.

**Fig. 6-34 Separate Peaks with each Single Element**



- c Here are distinguished two kinds of peaks, namely, major peaks (peak No. 1, 3) and a minor peak (peak No. 2).
- d Observe that, in default initial state, the number of ranges is set to two for major peaks and one for minor peaks.
- e Ranges here are supposed to be used as the unit of analysis in the extrapolation process.
- f As for variation, the number of ranges for a minor element can be either zero or one, while that a major element can be either one or two.

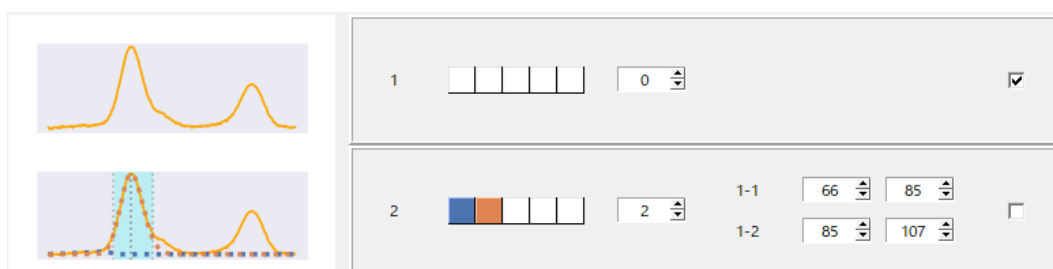
<sup>14</sup> A composite of multiple elution elements is used when a single element is not enough to model a peak.

- g The above rules come from the usual needs in analysis such that we usually want to see separately the ascending side and the descending side of a major peak, while such distinction is not appropriate for a minor peak.
- h Setting the number to zero is supposed to mean, just as a user interface, that the element should not be used alone but be used as a part of another peak.

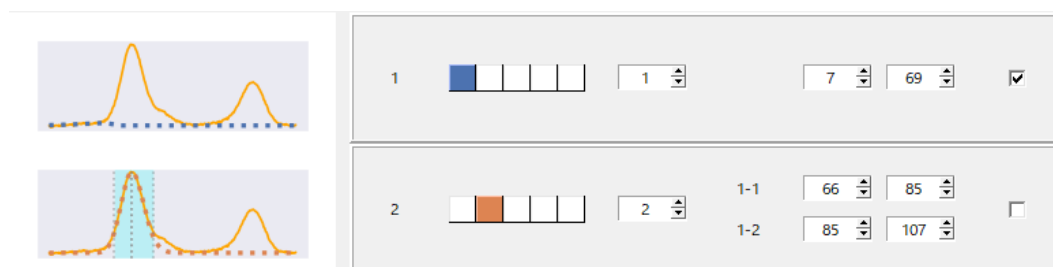
#### 6.4.5 Editing Combination of Elution Elements

- a Now, using the EGH model in Fig. 6-32, let's take up the first two elements (■, ■) and see how the combination can be edited.
- b Compare the following two figures which represent two different interpretations of the experiment data.

**Fig. 6-35 One major peak consisting of two elements**



**Fig. 6-36 Two peaks (minor and major) with each single element**

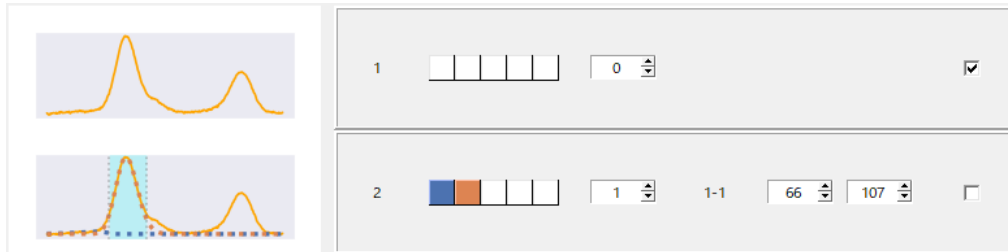


- c With the first interpretation (Fig. 6-35), the left major peak will be analyzed as of a single component and the corresponding elution curve (i.e., the variation of concentration) is calculated as the sum of the two elements.
- d On the other hand, with the second interpretation (Fig. 6-36), the same peak will be analyzed as of double components, i.e. as a compound of one minor peak and another major peak, and the elution elements are used as separate concentration for each extrapolation of the respective peak.
- e To change the interpretation between the above two, we use “number of ranges” of minor peaks (which correspond to the upper frame — peak No. 1 — in this case).<sup>15</sup>

<sup>15</sup> For this toggle purpose, we use the number of ranges for minor peaks because it is either one or zero and easy to

- f If you reduce the number to zero, you will get the combination of Fig. 6-35.<sup>16</sup>
- g If you increase the number to one, you will get the combination of Fig. 6-36.
- h Reducing the number for major peaks to one, as in the following figure, does not affect the combination and only suggests that both (ascending and descending) sides of the peak are treated as one range in the extrapolation process.

**Fig. 6-37 One major peak with one analysis range**



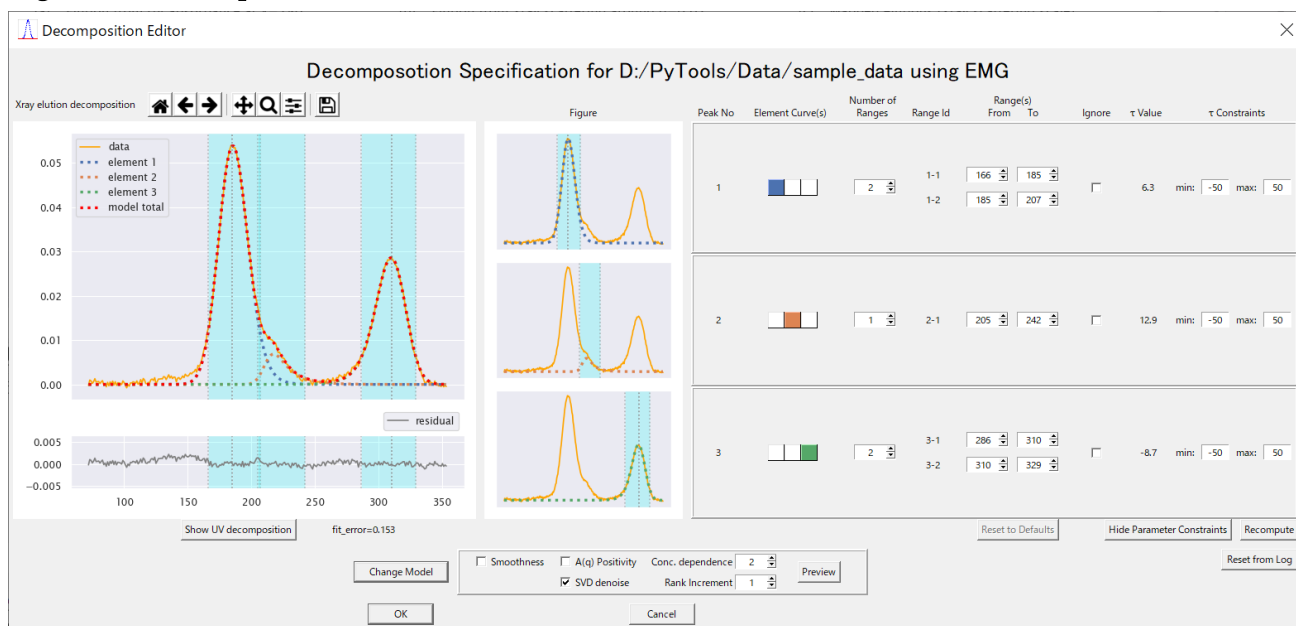
indicate whether the peak exists alone or the element is a part of another.

<sup>16</sup> When the number of ranges is set to zero and the element is indicated to be a part of another, there are, in general, two possibilities whether it belongs to a left or right adjacent major peak. It is determined by the distance between the centers of peaks which side it should belong.

### 6.4.6 Elution Model Parameter Control

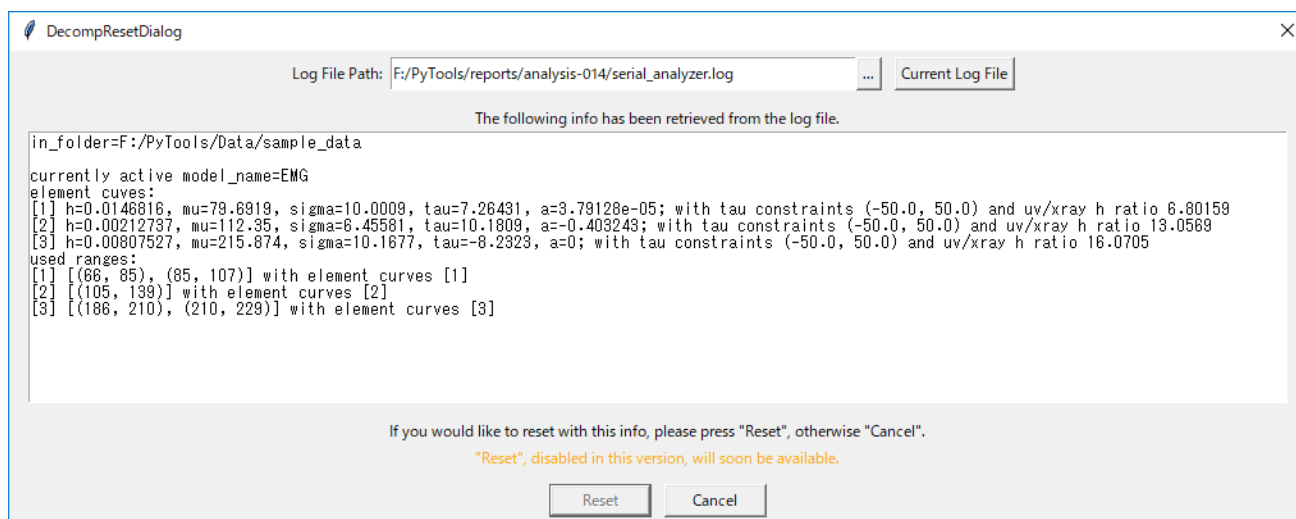
- By pressing the “Show Parameter Constraints” button, you can observe and set constraints for the  $\tau$  parameter of the selected elution model.
- After specifying the minimum or maximum value, press “Recompute” button to get an updated view of the figures and parameter values.

**Fig. 6-38 Decomposition Editor with the Parameter Constraints shown**



- In case you need to see detailed information of the model curves, press “Reset from Log” button, followed by “Current Log File” button press, which shows a dialog as shown below.
- This dialog can be used to restore the set of parameters previously used if you choose the corresponding old log file.

**Fig. 6-39 “Reset from Log” Dialog**



## 6.5 Analyzer Dialog

### 6.5.1 Purpose and Manipulation Flow

- This dialog is for confirmation on parameters or options before the execution of analysis and provides a chance to change some of them.
- It usually appears just after pressing “► Serial Analysis” button in Mapping Dialog.<sup>17</sup>
- As for the “Preview” button with the preceding three check buttons, see Section 6.7.

**Fig. 6-40 Analyzer Dialog**

### 6.5.2 Description

- Input/output information at top is read-only just for confirmation.
- As default, X-ray scattering curves are averaged to get better signal-to-noise (S/N) ratio. You can change those parameters and the folder (and file-naming) to save the resulting curves in.
- Concentration factors input is required to convert from UV absorbance to concentration. The default values are set according to the rule shown in the following table.

**Tab. 6-6 Default Values for Concentration Factors**

Factor Name	if measured before 2019-9-30	if measured after 2019-9-30
Path Length Factor	5.0	7.071 (= $10/\sqrt{2}$ )
Extinction	1.0	1.0

<sup>17</sup> You can change this manipulation flow in the Settings Dialog. See Section 6.8.



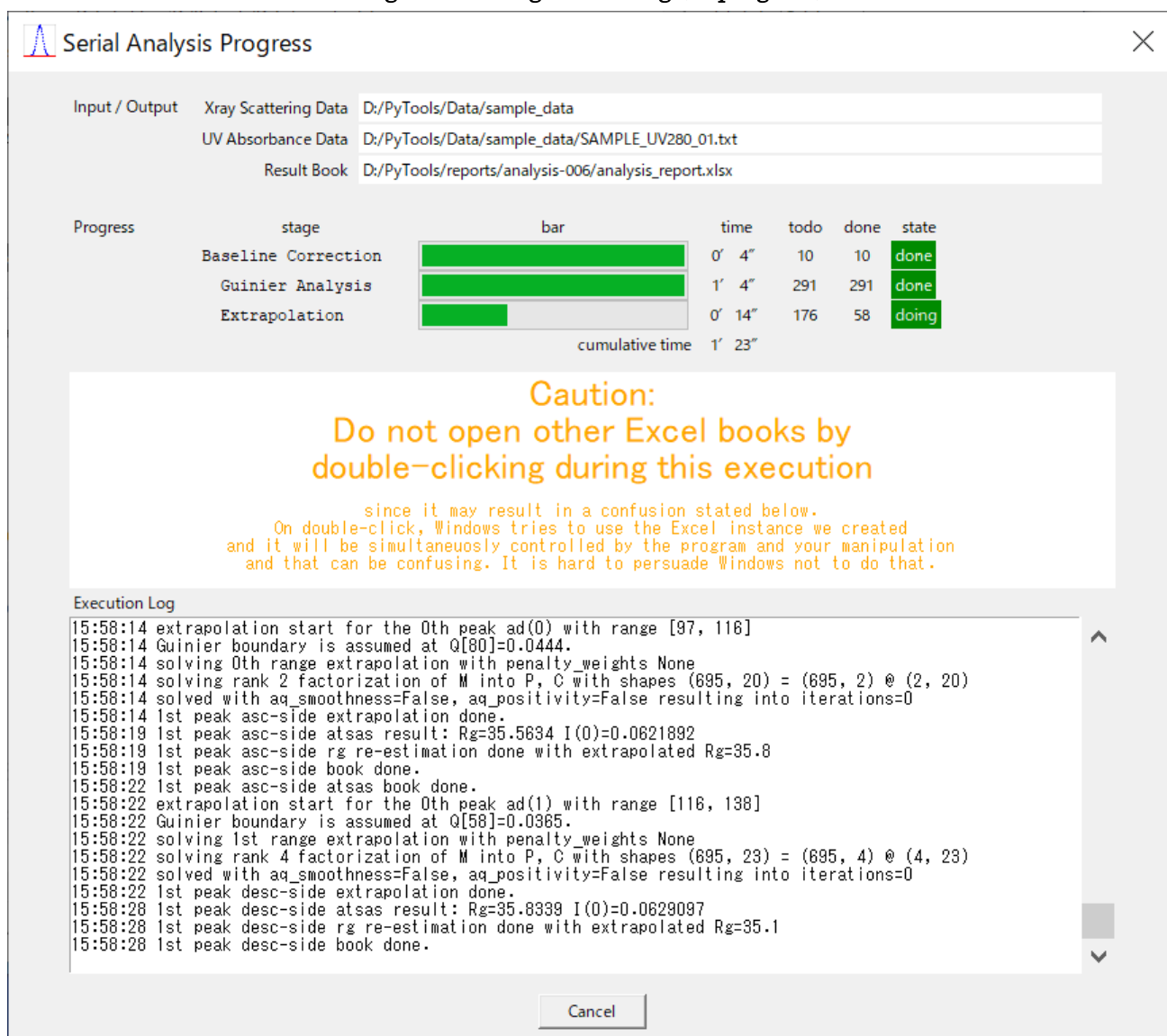
- d These factors are not essentially relevant to the computation of  $R_g$  and  $I(0)$  or extrapolation to zero-concentration, however, they do affect the values in the results on the summary sheets mentioned in Section 5.1.3. Therefore, don't forget to change them if they are different from default values.
- e “◀ UV-Xray Mapping” button shows the Mapping Dialog described in Section 6.2.
- f “▶ Serial Execution” button invokes a serial analysis, which consists of the following three stages.
  1. Baseline correction of all the X-ray scattering data
  2. Guinier analyses of all the X-ray scattering data
  3. Extrapolation to zero-concentration of each ascending or descending side of all the peaks
- g Right after the analysis is invoked normally, “Progress Dialog” appears, which is described in the next section.
- h Although they take the form of check buttons, “Baseline correction” is automatically selected depending on the selected options in Mapping Dialog, and “Guinier analysis” is always selected. “Extrapolation” only is changeable.
- i By selecting “Matrix Formulation”, you can preview the extrapolation result using “Preview” button. (See Section 6.7)
- j If you don't select “Matrix Formulation”, the extrapolation process will be conducted in an iterated way varying  $Q_i$  as described in Section 5.1.2.

## 6.6 Progress Dialog

### 6.6.1 Purpose and Manipulation Flow

- This dialog appears just after pressing “Run” button in Analyzer Dialog.
- It show the progress with progress bars and log messages.
- You can use the “Cancel” button to stop the execution on the way.

**Fig. 6-41 Progress Dialog (in progress)**

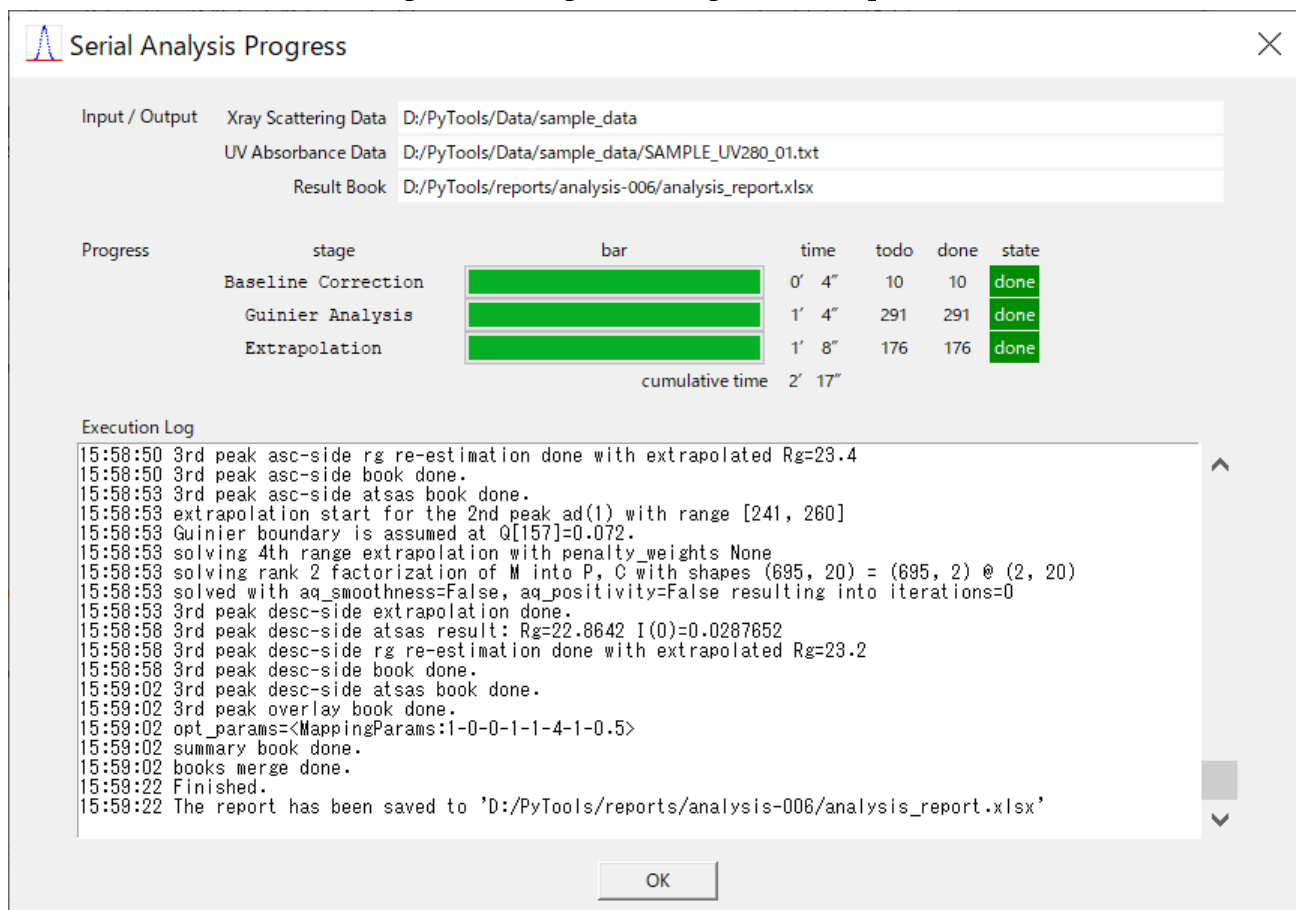


### 6.6.2 Description

- The two figures (Fig. 6-41 above, Fig. 6-42 below) show the dialog in the states of “in progress” and “when complete”.
- Input/output information in the top three rows is read-only just for confirmation.
- The numbers, which appear under the column title “todo” and “done”, show rough counts in internal process units, and should not be interpreted strictly.

- d Other items are self-explanatory.
- e Users are advised not to open other Excel books during the execution of analysis to avoid a simultaneous use of the same Excel instance by human and program.<sup>18</sup>

**Fig. 6-42 Progress Dialog (when complete)**



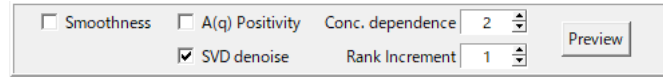
<sup>18</sup> Although the program creates an invisible Excel instance for its exclusive use, Windows usually forces the use of the same instance when requested by humans by double clicking books. It is not easy to persuade the OS so as not to do that.

## 6.7 Extrapolation Preview

### 6.7.1 Preview Button Frame

- a The button which invokes this preview is placed in a frame shown below with its preceding control buttons.

**Fig. 6-43 Preview Button Frame**

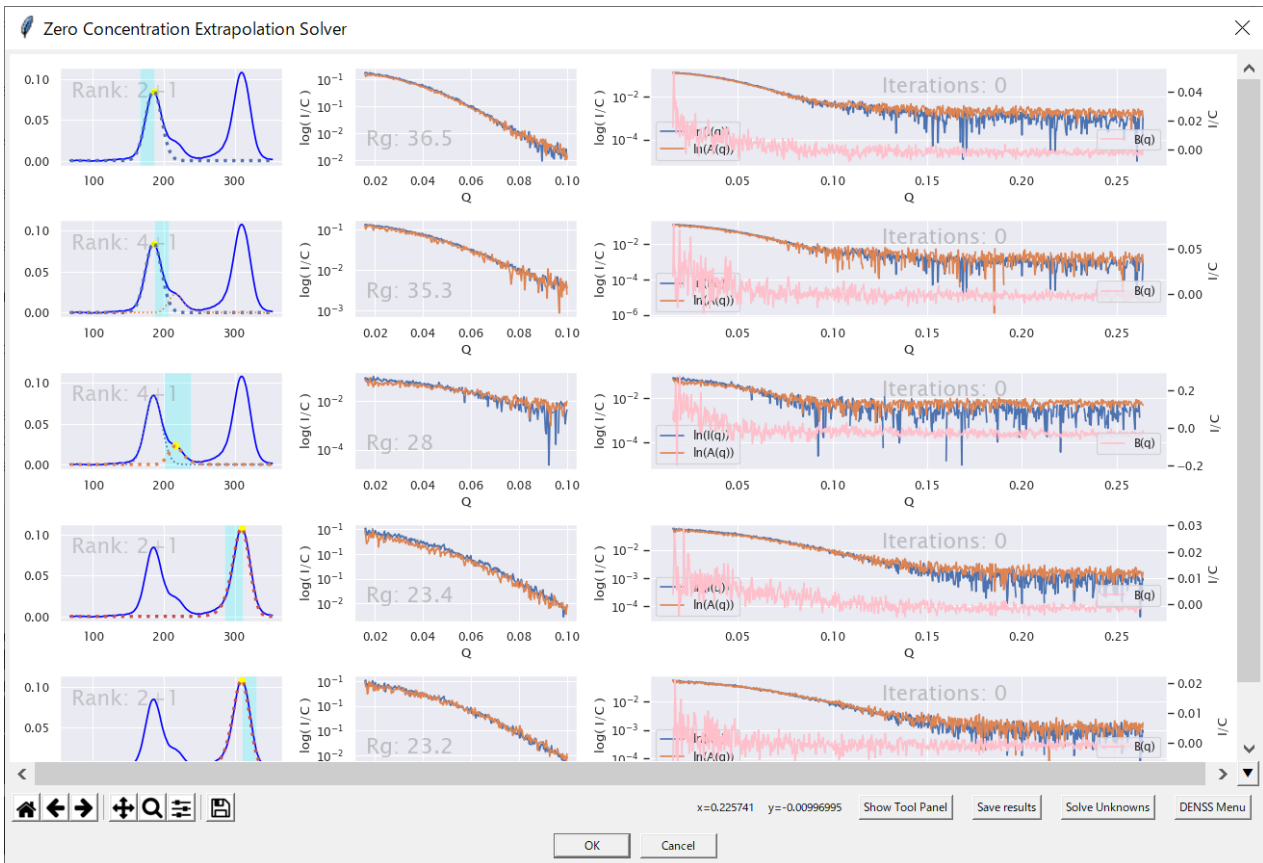


- b This "Preview" button frame is placed in each of the following three dialogs: Decomposition Editor, Range Editor or Analyzer Dialog.
- c The invocation options can be controlled with the preceding three check buttons. See Section 6.7.4 for concrete description.

### 6.7.2 Purpose and Manipulation Flow

- a This preview dialog is for quickly observing the results of extrapolation process and adjusting the penalty weights for its optimization.
- b It is invoked by the "Preview" button mentioned in the previous section.
- c This preview is available only either when you are in the Decomposition Editor or when you have selected "Matrix Formulation" option in the Analyzer Dialog.

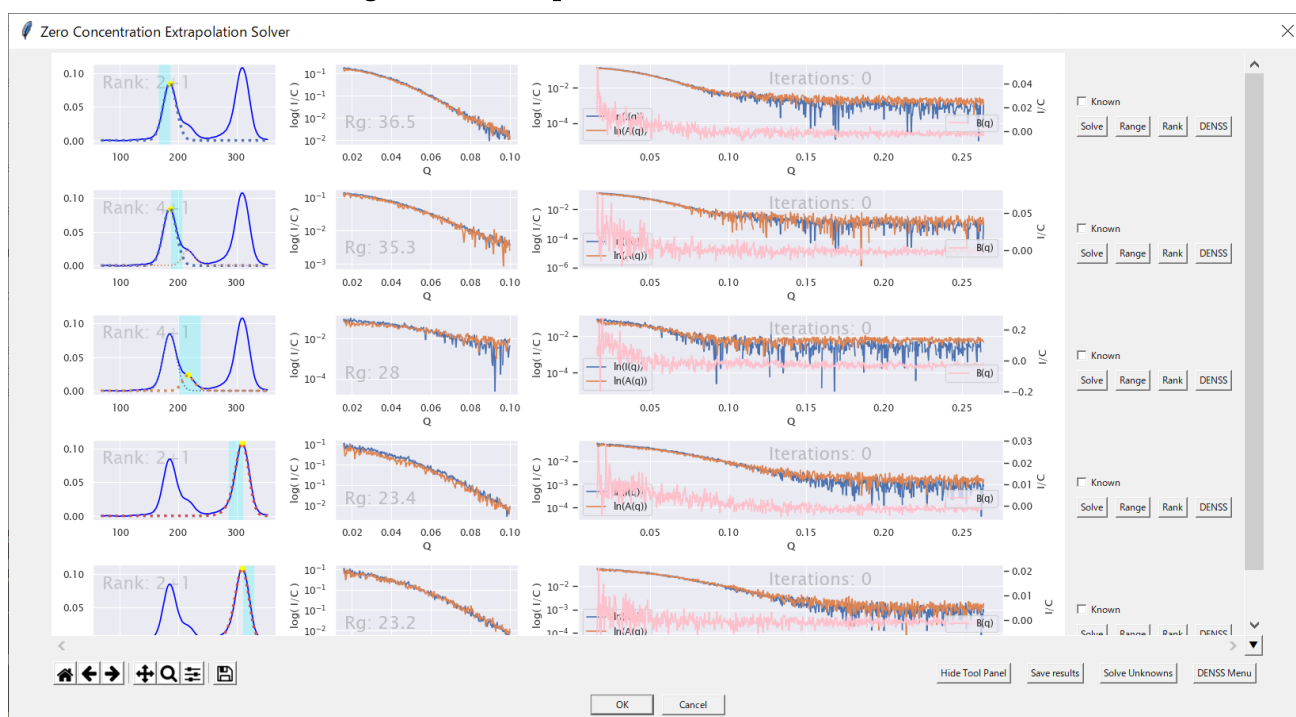
**Fig. 6-44 Extrapolation Preview**



### 6.7.3 Description

- a The preview consists of rows, each of which contains three figures and a hidden panel.
- b Each row corresponds to an analysis range as an optimization unit of extrapolation process.
- c The three figures show, from left to right respectively, the following items:
  1. The analysis range as a pale blue belt in the total elution curve with the rank of matrix used in the optimization,
  2. The original scattering curve in the small angle region and the extrapolated curve (i.e.  $A(q)$  curve) with its  $R_g$  value,
  3. All the three curves including the  $B(q)$  curve in the entire angle region.
- d The hidden panel for regularization weights adjustment can be shown using the “Show Tool Panel” button.(Fig. 6-45)

**Fig. 6-45 Extrapolation Preview with Tool Panel**



- e Tool panels are provided for each analysis range and can be changed separately.
- f For each panel, there are included the three parameters, listed in the following table, with each scale slider.

**Tab. 6-7 Regularization Weights used in the Optimization for Extrapolation Process**

No	Parameter Name	Default Value	Min Value	Max Value
1	A(q) Positivity Penalty	0.1	0	0.2
2	A(q) Smoothness Penalty	0.05	0	0.2
3	B(q) Smoothness Penalty	0.005	0	0.2

- g These parameters correspond, in the listed order respectively, to  $\lambda_1, \lambda_2, \lambda_3$  in the following formula representing the objective function used in the optimization.

$$\operatorname{argmin}_P \left[ \|P \cdot C - \tilde{M}\|_F^2 + \lambda_1 \|A_{\text{pos}}(P)\|_F^2 + \lambda_2 \|S_A(P)\|_F^2 + \lambda_3 \|S_B(P)\|_F^2 \right] \quad (6-3)$$

- h These parameters  $\lambda_1, \lambda_2, \lambda_3$  are called regularization (or overfitting reduction) parameters.
- i You can change default values for the parameters in “Settings Dialog”.
- j Where, the symbols are described in the following table. (See also the next page for those symbol definitions)

**Tab. 6-8 Description of the Symbols used in the Optimization Formula**

No	Symbol	Description
1	M	Data matrix
2	$\tilde{M}$	SVD-denoised data matrix
3	P	Matrix containing extrapolated scattering curves A(q), B(q) as columns
4	C	Matrix containing concentration (elution curves) as rows
5	P · C	Matrix product of matrices P and C
6	$\  \cdot \ _F^2$	Squared Frobenius norm of a matrix
7	A <sub>pos</sub> (P)	Matrix containing only negative elements of A(q) column of P (other elements are set to zeros)
8	S <sub>A</sub> (P)	Smoothness measurement matrix concerning only A(q) curve
9	S <sub>B</sub> (P)	Smoothness measurement matrix concerning only B(q) curve

k Here are the symbol definitions used in formula (6-3). For brevity, let

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} & m_{1,4} \\ m_{2,1} & m_{2,2} & m_{2,3} & m_{2,4} \\ m_{3,1} & m_{3,2} & m_{3,3} & m_{3,4} \\ m_{4,1} & m_{4,2} & m_{4,3} & m_{4,4} \\ m_{5,1} & m_{5,2} & m_{5,3} & m_{5,4} \end{bmatrix}, P = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \\ a_5 & b_5 \end{bmatrix}, C = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \\ c_1^2 & c_2^2 & c_3^2 & c_4^2 \end{bmatrix},$$

then,

$$\|Apos(P)\|_F^2 = \sum_{a_i < 0} a_i^2,$$

$$\|S_A(P)\|_F^2 = \sum_{i=1}^4 (a_i - a_{i+1})^2$$

$$\|S_B(P)\|_F^2 = \sum_{i=1}^4 (b_i - b_{i+1})^2$$

l We used Frobenius norms  $\| \cdot \|_F^2$  here because most of the terms are easily expressed as matrix operations as shown below, to make the expression concise.

$$D = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & 1 & -1 & \\ & & & 1 & -1 \\ & & & & 0 \end{bmatrix}, R_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, R_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$D \cdot P = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & 1 & -1 & \\ & & & 1 & -1 \\ & & & & 0 \end{bmatrix} \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \\ a_5 & b_5 \end{bmatrix} = \begin{bmatrix} a_1 - a_2 & b_1 - b_2 \\ a_2 - a_3 & b_2 - b_3 \\ a_3 - a_4 & b_3 - b_4 \\ a_4 - a_5 & b_4 - b_5 \\ 0 & 0 \end{bmatrix}$$

$$S_A(P) = D \cdot P \cdot R_1 = \begin{bmatrix} a_1 - a_2 & 0 \\ a_2 - a_3 & 0 \\ a_3 - a_4 & 0 \\ a_4 - a_5 & 0 \\ 0 & 0 \end{bmatrix}, S_B(P) = D \cdot P \cdot R_2 = \begin{bmatrix} 0 & b_1 - b_2 \\ 0 & b_2 - b_3 \\ 0 & b_3 - b_4 \\ 0 & b_4 - b_5 \\ 0 & 0 \end{bmatrix}$$

m It is important from a technical point of view that this type of objective functions<sup>19</sup> are differentiable so that we can use gradient descent method to effectively optimize such large number of parameters.

---

<sup>19</sup> Namely, functions which can be expressed using Frobenius norms.

n Detailed description of the optimization will soon be available in the section 8.2.8.

#### 6.7.4 Optimization Options

a The first two buttons in the “Preview Button Frame” show below can be used to change the optimization as shown in the following table.

**Fig. 6-46 Preview Button Frame**



**Tab. 6-9 Check button states and the corresponding formulation**

Smoothness	A(q) Positivity	Formulation
<input type="checkbox"/>	<input type="checkbox"/>	$\operatorname{argmin}_P \left[ \ P \cdot C - \tilde{M}\ _F^2 \right]$
<input checked="" type="checkbox"/>	<input type="checkbox"/>	$\operatorname{argmin}_P \left[ \ P \cdot C - \tilde{M}\ _F^2 + \lambda_2 \ S_A(P)\ _F^2 + \lambda_3 \ S_B(P)\ _F^2 \right]$
<input type="checkbox"/>	<input checked="" type="checkbox"/>	$\operatorname{argmin}_P \left[ \ P \cdot C - \tilde{M}\ _F^2 + \lambda_1 \ Apos(P)\ _F^2 \right]$
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	$\operatorname{argmin}_P \left[ \ P \cdot C - \tilde{M}\ _F^2 + \lambda_1 \ Apos(P)\ _F^2 + \lambda_2 \ S_A(P)\ _F^2 + \lambda_3 \ S_B(P)\ _F^2 \right]$

b If the third option, namely “Conc. dependence”, whose default value is 2, is reduced to 1, the matrices P and C described in the previous section will be constructed as shown below; i.e., there will be no rows or columns concerning B(q); and the rest of the details are modified accordingly.

$$P = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}, C = [c_1 \quad c_2 \quad c_3 \quad c_4],$$

c “SVD denoise” is checked by default. If you uncheck this, denoise process will be skipped so that the original data matrix M be directly used instead of  $\tilde{M}$ .

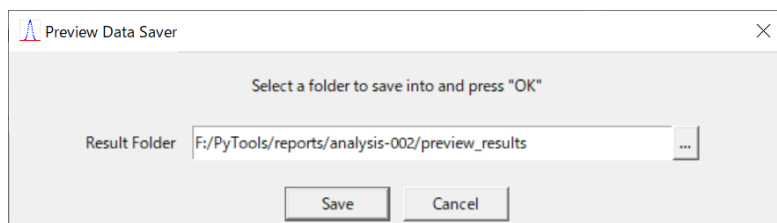
d The rank, used in the default SVD denoise process, is determined to “Conc. dependence” plus “Rank Increment”, where the default increment is set to one to avoid too strict interpretation of the situation.



### 6.7.5 Saving the Preview Results

- a Using the “Save results” button, you can show “Preview Results Saver” shown in the figure below and save your preview results.
- b The results will be saved in a subfolder named “preview\_results” in the analysis result folder.

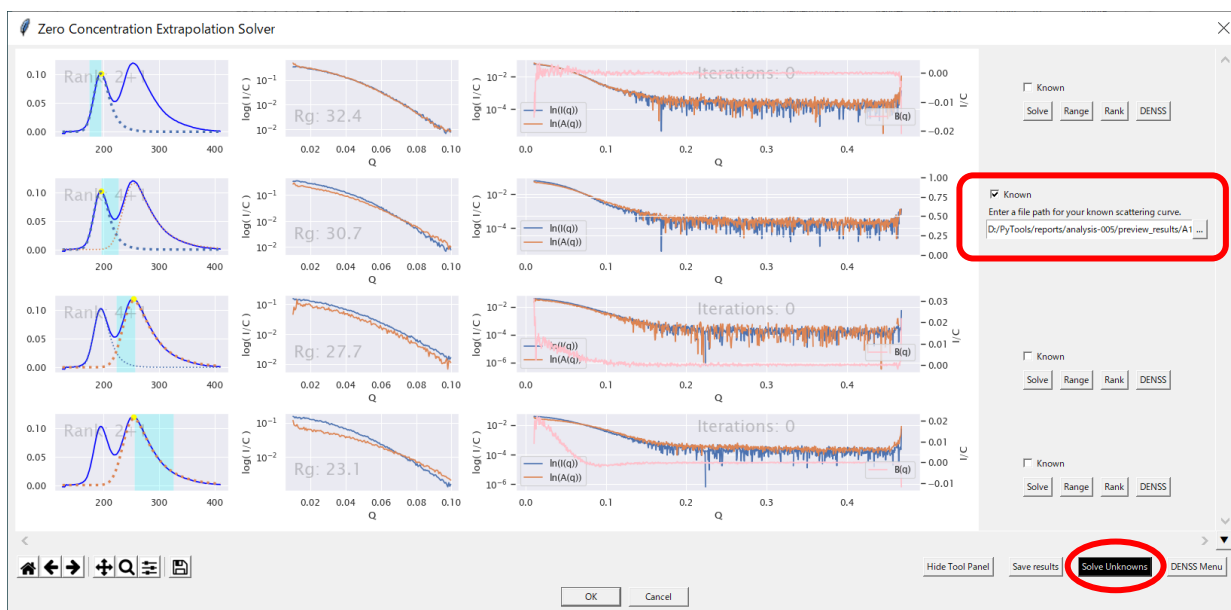
**Fig. 6-47 Preview Results Saver**



### 6.7.6 Giving Known Scattering Curves

- a This feature is experimental.
- b For simplicity, we assume here in this section that the inter-particle effects are negligible.
- c In other words, we consider the case when “Conc. dependence” described above is set to 1.
- d Moreover, suppose that, among all the components involved, the scattering curves from some of them are already known.
- e In such situations, we can use the known results to improve the solution to unknown components.
- f To give a known scattering curve, check the corresponding “Known” check button in the parameter panel, which can be shown by “Show Tool Panel” button.
- g Checking the button will change this part of the panel to the form, where you can enter the path of your known curve data. (Fig. 6-48)

**Fig. 6-48 Giving Known Results in the Preview Parameter Panel**

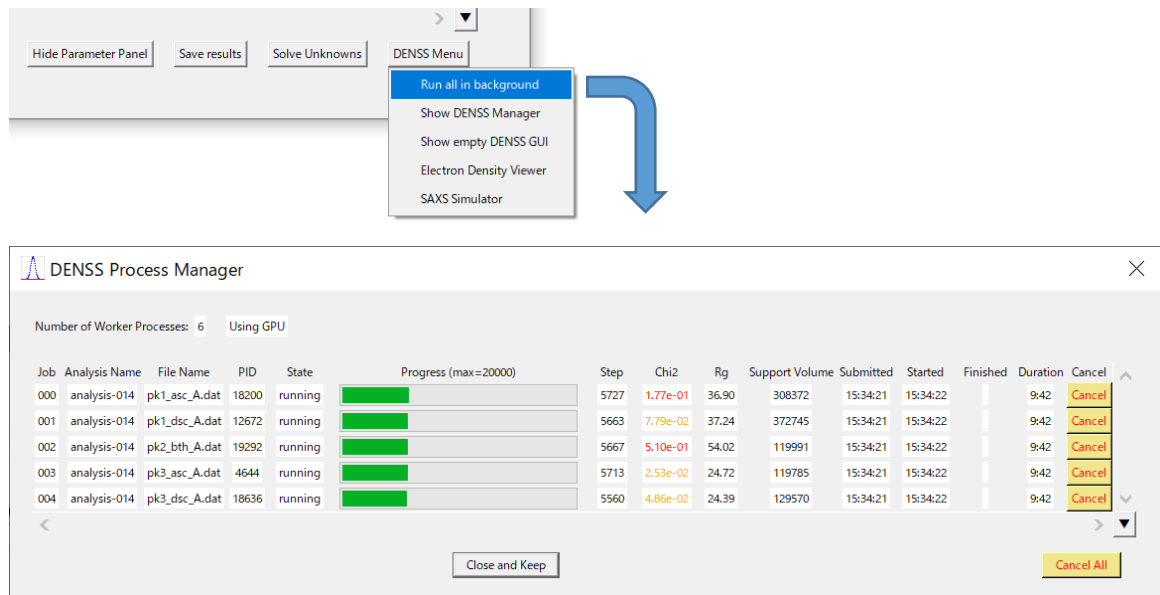


- h On entering the path, the corresponding figures in the same row in the preview will be updated to be consistent with the given known curve.
- i Press “Solve Unknowns” button to let the program to re-compute considering this new condition.
- j To avoid deviating from the matrix formulation described above, the program solves this problem by adding another regularization term like  $\lambda_4 \|P_{*,j} - A_{*,j}\|_F^2$  with large enough parameter  $\lambda_4$ , which forces the corresponding column vector of P as close as possible to the given curve.

### 6.7.7 Background Execution of DENSS jobs

- a We can pass the previewed, or extrapolated, curves to a local [DENSS](#) program and run the jobs in background from “DENSS Menu” located lower bottom in the Preview as shown below.

**Fig. 6-49 Passing the curves to DENSS program**



- b After pressing the “Run all in background” button, “DENSS Process Manager” appears, with which we can monitor the execution of the jobs.
- c Since the jobs are run in background, we can safely close the dialog and re-open it as we like to monitor them.
- d Re-opening can be performed here in this “DENSS Menu” as well as from DENSS Tools menu in the Main Dialog. (See also Section 6.14 DENSS Tools)
- e “Show empty DENSS GUI” button is for running DENSS using a curve file, which may have been made by any other means.
- f Execution using GPUs will be enabled if the machine is equipped with NVIDIA compatible GPUs and you have [CUDA Toolkit 10.2](#) installed.

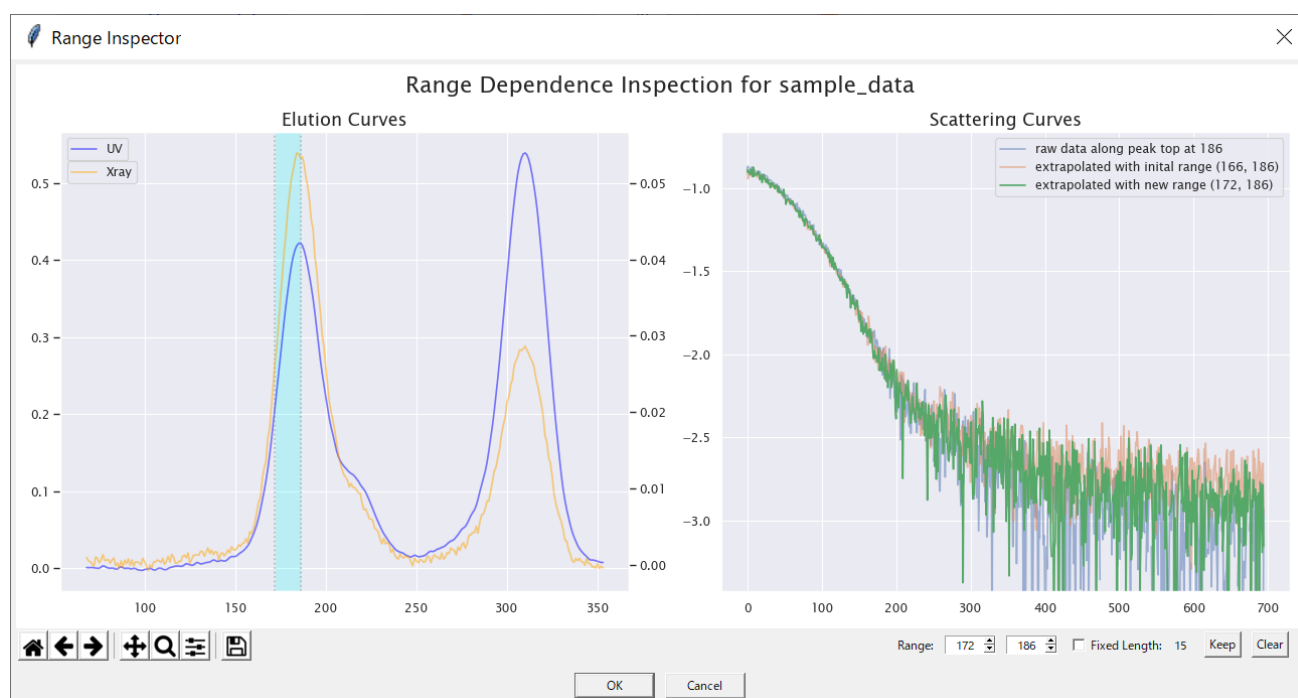
### 6.7.8 Re-optimization and Accessory Tools

- a In the adjustment panel, there are four buttons, which can be used as described in the following table.

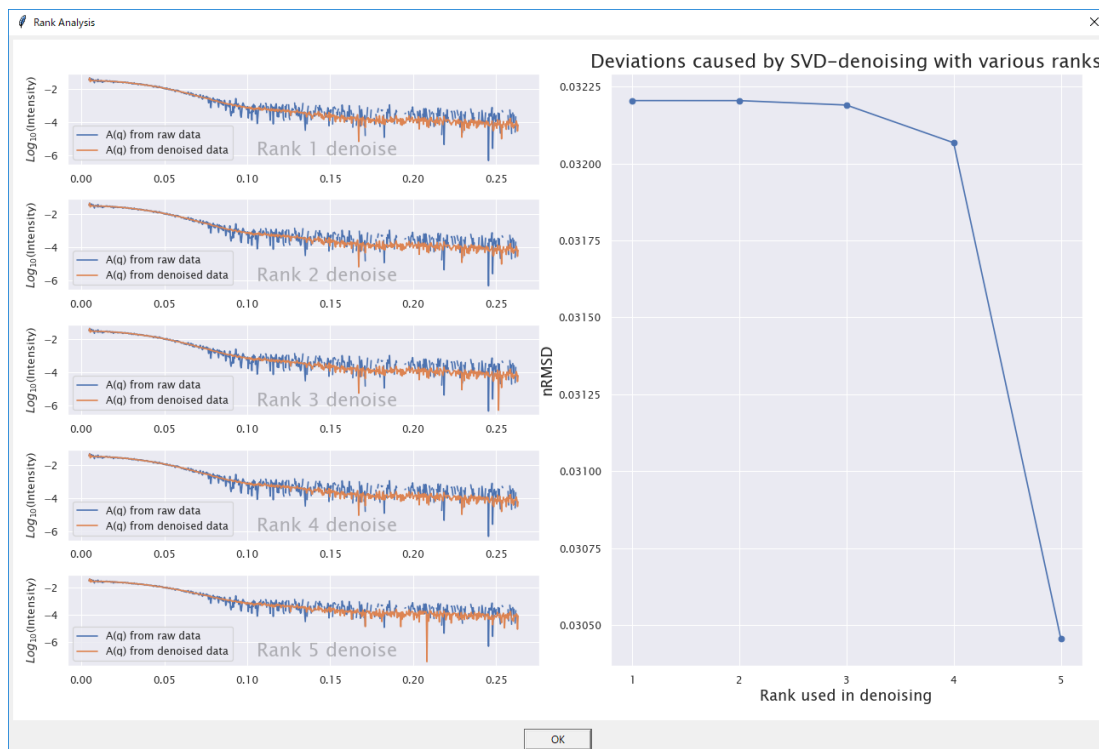
**Tab. 6-10 Re-optimization and Accessory Tools**

Button Name	Action
Solve	Invokes the optimizer with the (changed) penalty weights and re-computes the extrapolation results.
Range	Shows “Range Inspector” (Fig. 6-50) which is used to inspect range dependence of the extrapolation result. After inspecting, the last range will be applied as your analysis range if you press “OK”.
Rank	Shows “Rank Inspection Figure” (Fig. 6-51), which can be used to inspect the rank of the data matrix within the range used in the optimization.
DENSS	Invokes a GUI dialog (Fig. 6-52) to run DENSS program, which is yet an experimental feature in this version.

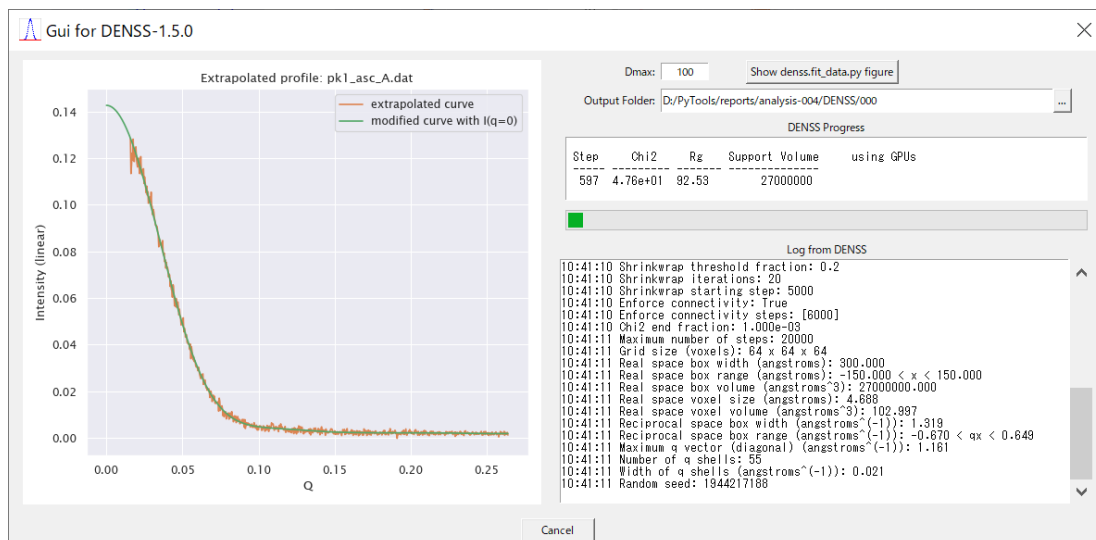
**Fig. 6-50 Range Inspector**



**Fig. 6-51 Rank Inspection Figure**



**Fig. 6-52 DENSS invocation GUI**



## 6.8 Settings Dialog and Setting Menu

### 6.8.1 Purpose and Manipulation Flow

- a This dialog is invoked from the main menu by selecting [Settings]/[Settings Dialog].

**Fig. 6-53 Settings Dialog**

**Settings**

ATSAS AUTORG Location

- C:\Program Files (x86)\ATSAS 3.0.1\bin\autorg.exe
- C:\Program Files (x86)\ATSAS 3.0.0\bin\autorg.exe
- C:\atsas\bin\autorg.exe

Standard Mapping Plane

UV Absorbance Elution Curve: averaging  points around wave length  $\lambda_1 =$   (used in the analysis)  
wave length  $\lambda_2 =$   (for observation only)

X-ray Scattering Elution Curve: averaging  points around scattering vector  $q =$

Angle Range Restriction Start (Initial default setting appearing in the Outline Figure in Main Dialog)

(1) Minimum of the measured Q range  
 (2) Minimum of the Guinier Q range  
 (3) Extended Limit beyond the Guinier Q range

Acceptable Rg-consistency:

Setting 0 here implies the above option (1)  
Setting 1 here implies the above option (2)

Extrapolation Regularization Weights

Negative A(q)	<input type="text" value="0.1"/>
A(q) Smoothness	<input type="text" value="0.05"/>
B(q) Smoothness	<input type="text" value="0.005"/>

Other Options

allow Multiple Instances

### 6.8.2 ATSAS AUTORG Location

- a At top is a list showing the location of ATSAS/AUTORG exe files, which are recognized and to be used by the program. By default, the latest version will be used.
- b In case there exist no expected entries, manually select your right paths.
- c Paths for other ATSAS programs are recognized accordingly using these paths.

### 6.8.3 Standard Mapping Plane

- a You can change the positioning of the Standard Mapping Plane by changing values for wave length or scattering vector length or numbers of points for averaging.
- b The second UV wave length value ( $\lambda_2=260$ ) is just for observation in case of a different kind of protein sample. The elution curve at that wave length can be observed in the usable data

range restriction. (See Fig. 6-3)

#### 6.8.4 Angle Range Restriction Start

- a As for the default angle range, the start setting can be chosen from the three options, while the end of the range is automatically set to a little smaller point than the flange limit if it is significant enough as the program can detect.
- b The third option for the angle range start is determined to the limit (toward the smaller angle) where the estimated  $R_g$  value remains within the acceptable consistency specified. For example, suppose the  $R_g$  in the Guinier region be 50 and you set 0.8 (which is default) to the acceptable  $R_g$ -consistency, then, the limit angle ( $Q$ ) is determined as small as possible while the estimated  $R_g^{20}$  in a small interval beyond the Guinier region remains within a range around from 40 to 60.
- c  $R_g$ -consistency between two  $R_g$ 's, say  $R_{g_1}$  and  $R_{g_2}$  based on  $R_{g_2}$ , is calculated as follows.

$$R_g \text{ consistency} = \exp \left( -\text{abs} \left( 1 - \frac{R_{g_1}}{R_{g_2}} \right) \right).$$

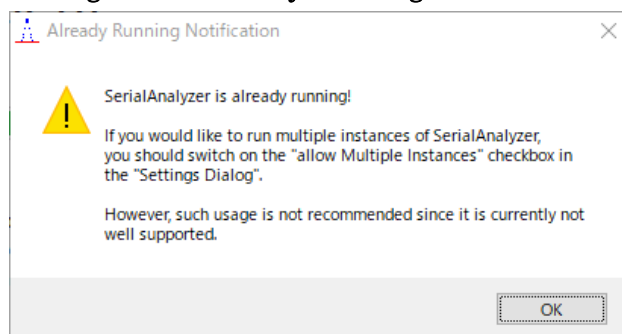
#### 6.8.5 Extrapolation Regularization Weights

- a .Our solver program for zero-concentration extrapolation uses the formulation (6-3) as described in Section 6.7.3.
- b Regularization weights here are the default values to be used by the extrapolation solver to compute extrapolated scattering curves, called  $A(q)$  curves in this document.
- c You can change these default weights temporarily in the

#### 6.8.6 Other Options

- a This application does not support multiple instances. Therefore multiple invocation is prohibited in the default setting. You will be warned about this in a message box as in the following figure.

**Fig. 6-54 Already Running Notification**



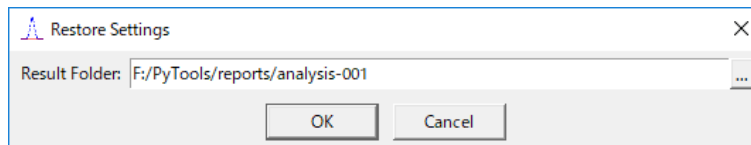
<sup>20</sup> This  $R_g$  value is estimated after simplified extrapolation to zero concentration, using the data near the peak top, to avoid discarding the valid small angle data which apparently look deviated due to inter-particle effects.

- b You can break through this restriction if you check “allow Multiple Instances”.

### 6.8.7 Restoring Settings from a Result Folder

- a The settings used in previous analyses can be restored by selecting [Settings]/[Restore from Result] from the main menu.

**Fig. 6-55 Restoring Settings from a Result Folder**

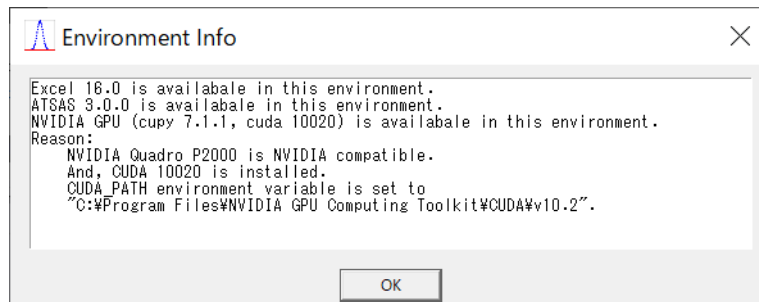


- b Note that this is a risky operation since the settings parameters may change from version to version and this restoring process does not support any conversion.

### 6.8.8 Showing Environment Information

- a From the [Settings]/[Check Environment] menu, you can get information of the environment as shown below.

**Fig. 6-56 Showing Environment Information**



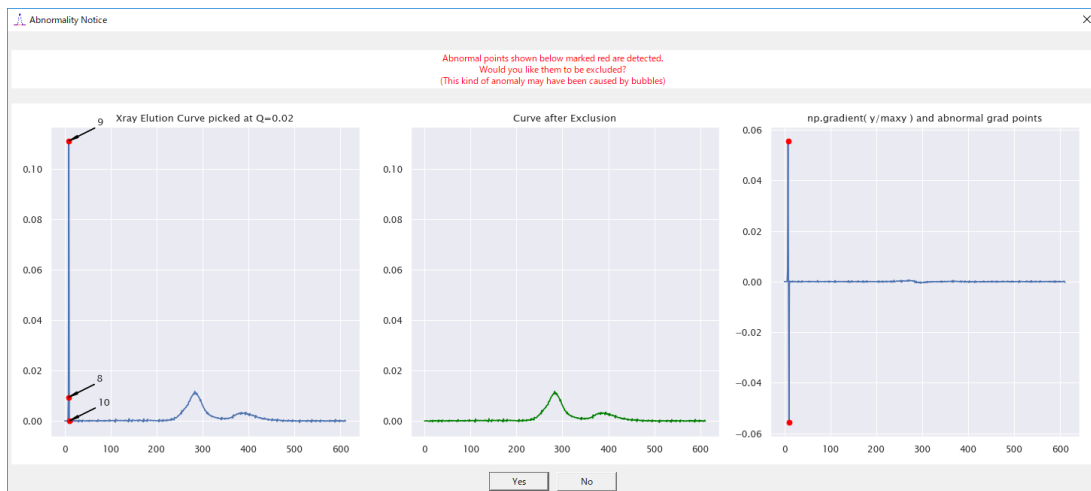
- b Use this information to verify the program’s recognition of the environment.

## 6.9 SEC Tools

### 6.9.1 Abnormal Data Handling

- a This menu item is currently disabled, while abnormal points can be handled in the dialog (Fig. 6-57), which is supposed to be automatically invoked when the program succeeds to detect them properly.

**Fig. 6-57 Detection and removal of abnormal points**



- b Those abnormal points may happen to get mixed into the data in such cases as accidental bubble inclusion.
- c After “Yes” reply to the dialog, they will be removed and replaced by interpolated data, and marked as shown in the following figure.

**Fig. 6-58 Marks in the Data File Table to exclude abnormal points**

No	File Name	Last Modified	exclude
0	RXR01_00000_sub.dat	2017-12-26 21:46:37	
1	RXR01_00001_sub.dat	2017-12-26 21:46:38	
2	RXR01_00002_sub.dat	2017-12-26 21:46:38	
3	RXR01_00003_sub.dat	2017-12-26 21:46:38	
4	RXR01_00004_sub.dat	2017-12-26 21:46:39	
5	RXR01_00005_sub.dat	2017-12-26 21:46:39	
6	RXR01_00006_sub.dat	2017-12-26 21:46:39	
7	RXR01_00007_sub.dat	2017-12-26 21:46:40	
8	RXR01_00008_sub.dat	2017-12-26 21:46:40	X
9	RXR01_00009_sub.dat	2017-12-26 21:46:40	X
10	RXR01_00010_sub.dat	2017-12-26 21:46:41	X
11	RXR01_00011_sub.dat	2017-12-26 21:46:41	
12	RXR01_00012_sub.dat	2017-12-26 21:46:41	
13			
14			
15			
16			
17			
18			
19			

Exclusion Done

Xray data at elution points [8, 9, 10] have been removed and interpolated.

OK

- d The facility for manually setting “exclude” marks to the column is also disabled in this version.



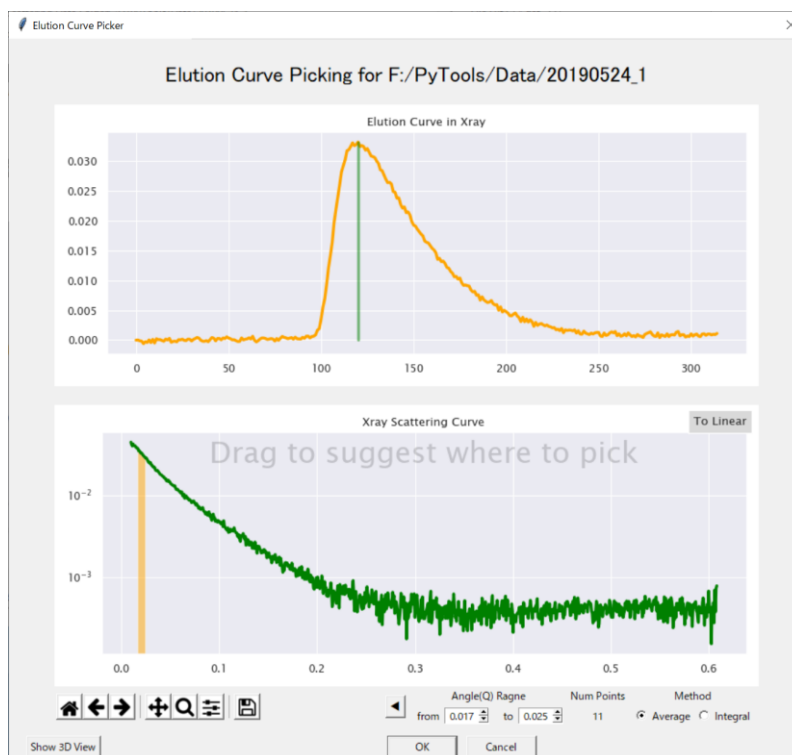
## 6.9.2 Data Range Restriction

- a This menu invokes the same dialog as described in 6.1.3 Data Range Restriction.

## 6.9.3 Elution Curve Picker

- a By default, X-ray Data Range Restriction by averaging 11 curves near the point ( $Q=0.02$ ) on the angular axis.
- b While you can change these default parameters in “Settings Dialog”, you can do it visually by using “Elution Curve Picker” (Fig. 6-59) invoked from the Tools menu.
- c To specify the range for averaging, either drag with your mouse in the lower figure or enter scattering vector ( $Q$ ) values of end points to the corresponding spin boxes.
- d As for “Method” options, “Average” means what it literally means while “Integral” means that the averaged curve values will be multiplied by the angular width, namely “To” value minus “From” value.

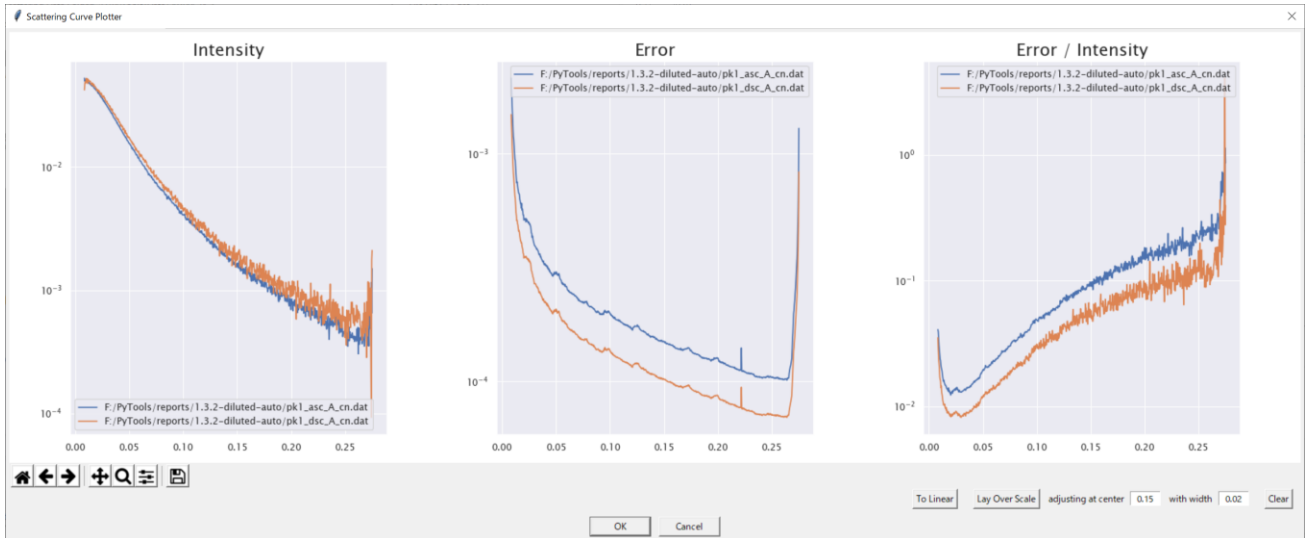
**Fig. 6-59 Elution Curve Picker**



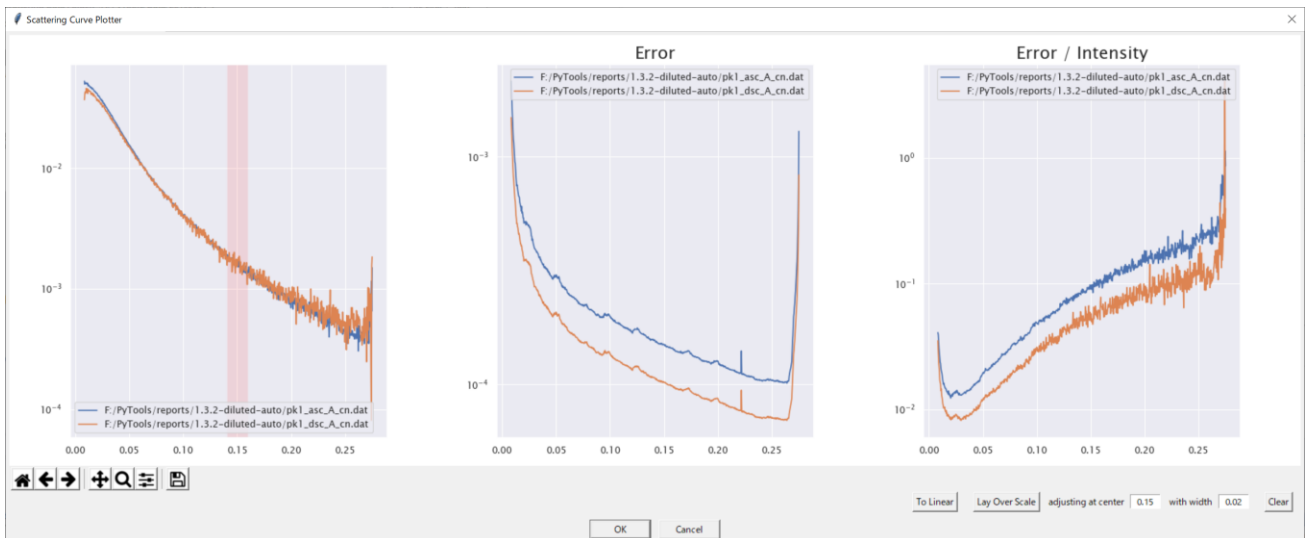
### 6.9.4 Scattering Curve Plotter

- a Using this tool, you can easily plot and compare scattering curves in files conforming to the expected format as described in Section 4.1
- b To plot curves, just drag and drop files into the canvas.
- c To compare, press “Lay Over Scale” button, which scales the other curves to the first curve so that they overlay to each other in the region suggested by the center and width specified.

**Fig. 6-60 Scattering Curve Plotter**



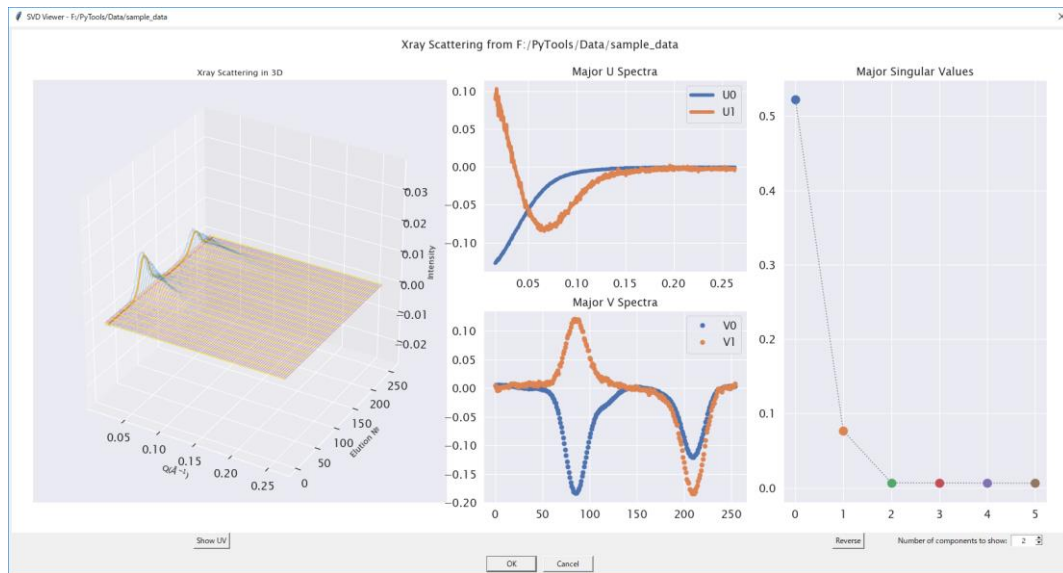
**Fig. 6-61 Scattering Curve Plotter (after “Lay Over Scale” press)**



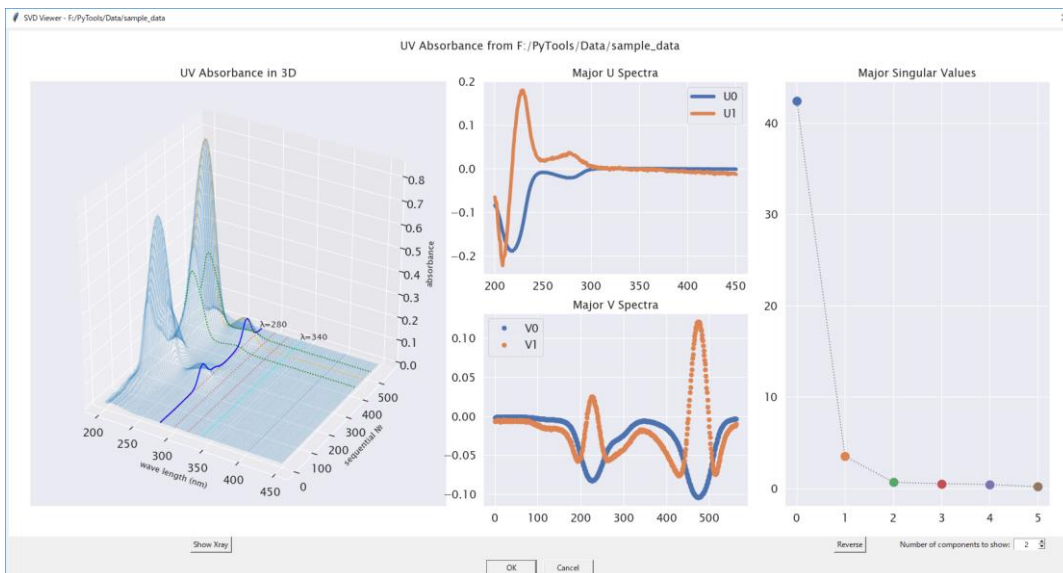
### 6.9.5 SVD Viewer

- a This dialog is for observing the row rank structure of the sets of measured data using SVD, i.e., Singular Value Decomposition.
- b It can toggle, by “Show Xray” or “Show UV” button, between the two sets of figures depicting SVD results for Xray and UV data.

**Fig. 6-62 SVD results for Xray data**



**Fig. 6-63 SVD results for UV data**



- c The number of curves in the central two figures can be changed using the spin box at the lower right corner of the dialog window.
- d “Reverse” button changes the vertical direction of the central two figures.

## 6.10 MCT-SAXS Support

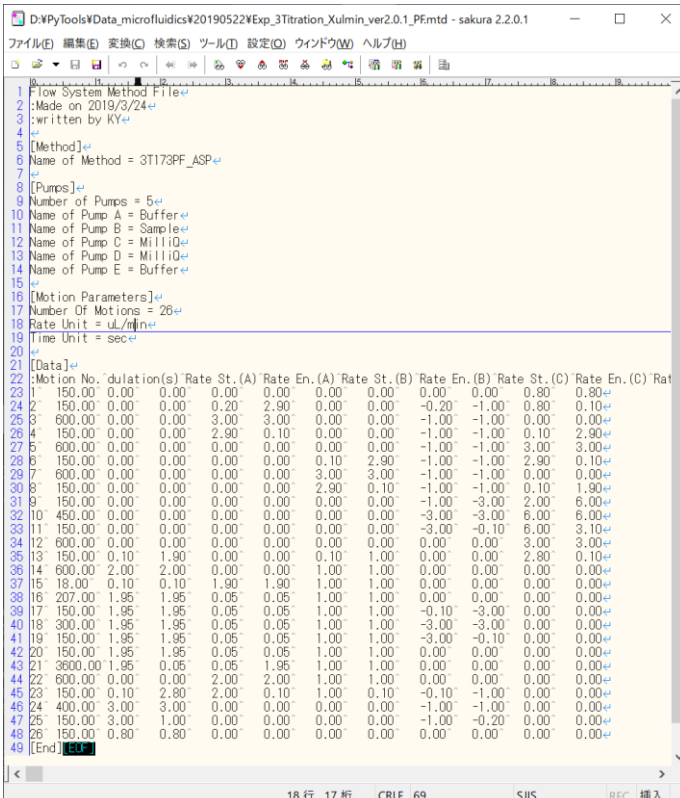
- a MCT is an acronym for Microfluidic Continuous Titration and MCT-SAXS means SAXS using MCT.

Support for MCT-SAXS is experimental and not yet officially supported.

## 6.11 Data Requirements

- a As input to this program, MCT-SAXS data should be given in one folder and be accompanied by a method file with “.mtd” file name extension in the same folder, which was used in the experiment to control the titration devices as shown below.

Fig. 6-64 Method File for MCT-SAXS Experiment



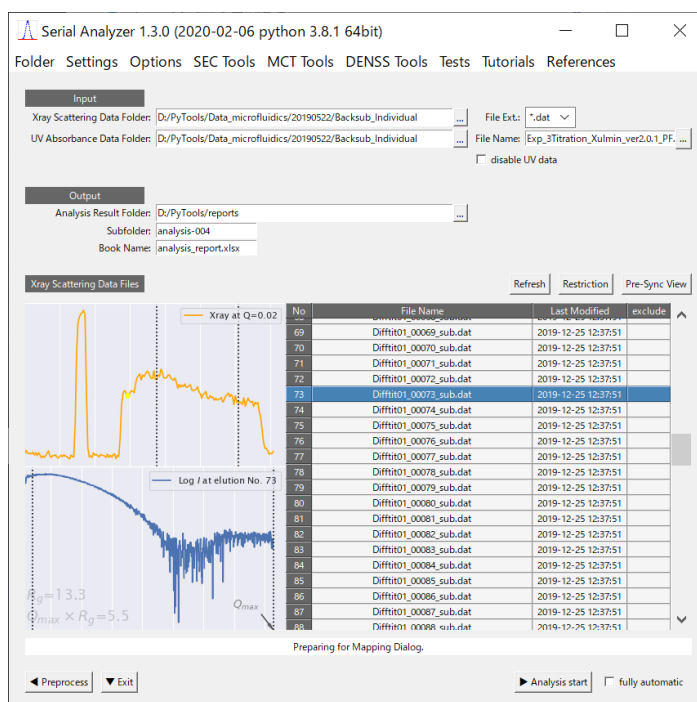
```
D:\VpYTools\Data_microfluidics\20190522\Exp_3Titration_Xulmin_ver2.0.1_Pfemtd - sakura 2.2.0.1
ファイル(E) 編集(C) 変換(O) 検索(S) ツール(T) 設定(O) ウィンドウ(W) ヘルプ(H)
1 Flow System Method File
2 :Made on 2019/3/24
3 :written by KY
4
5 [Method]
6 Name of Method = 3T173PF_ASP
7
8 [Pumps]
9 Number of Pumps = 5
10 Name of Pump A = Buffer
11 Name of Pump B = Sample
12 Name of Pump C = MilliQ
13 Name of Pump D = MilliQ
14 Name of Pump E = Buffer
15
16 [Motion Parameters]
17 Number Of Motions = 26
18 Rate Unit = uL/min
19 Time Unit = sec
20
21 [Data]
22 Motion No. Duration(s) Rate St. (A) Rate En. (A) Rate St. (B) Rate En. (B) Rate St. (C) Rate En. (C) Rate
23 1 150.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.80 0.80
24 2 150.00 0.00 0.00 0.20 2.90 0.00 0.00 -0.20 -1.00 0.80 0.10
25 3 600.00 0.00 0.00 3.00 3.00 0.00 0.00 -1.00 -1.00 0.00 0.00
26 4 150.00 0.00 0.00 2.90 0.10 0.00 0.00 -1.00 -1.00 0.10 2.90
27 5 600.00 0.00 0.00 0.00 0.00 0.00 0.00 -1.00 -1.00 3.00 3.00
28 6 150.00 0.00 0.00 0.00 0.00 0.10 2.90 -1.00 -1.00 2.90 0.10
29 7 600.00 0.00 0.00 0.00 0.00 3.00 3.00 -1.00 -1.00 0.00 0.00
30 8 150.00 0.00 0.00 0.00 0.00 2.90 0.10 -1.00 -1.00 0.10 1.90
31 9 150.00 0.00 0.00 0.00 0.00 0.00 0.00 -1.00 -3.00 2.00 6.00
32 10 450.00 0.00 0.00 0.00 0.00 0.00 0.00 -3.00 -3.00 6.00 6.00
33 11 150.00 0.00 0.00 0.00 0.00 0.00 0.00 -3.00 -0.10 6.00 3.10
34 12 600.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 3.00 3.00
35 13 150.00 0.10 1.90 0.00 0.00 0.10 1.00 0.00 0.00 2.80 0.10
36 14 600.00 2.00 2.00 0.00 0.00 1.00 1.00 0.00 0.00 0.00 0.00
37 15 18.00 0.10 0.10 1.90 1.90 1.00 1.00 0.00 0.00 0.00 0.00
38 16 207.00 1.95 1.95 0.05 0.05 1.00 1.00 0.00 0.00 0.00 0.00
39 17 150.00 1.95 1.95 0.05 0.05 1.00 1.00 -0.10 -3.00 0.00 0.00
40 18 300.00 1.95 1.95 0.05 0.05 1.00 1.00 -3.00 -3.00 0.00 0.00
41 19 150.00 1.95 1.95 0.05 0.05 1.00 1.00 -3.00 -0.10 0.00 0.00
42 20 150.00 1.95 1.95 0.05 0.05 1.00 1.00 0.00 0.00 0.00 0.00
43 21 3600.00 1.95 0.05 0.05 1.95 1.00 1.00 0.00 0.00 0.00 0.00
44 22 600.00 0.00 0.00 2.00 2.00 1.00 1.00 0.00 0.00 0.00 0.00
45 23 150.00 0.10 2.80 2.00 0.10 1.00 0.10 -0.10 -1.00 0.00 0.00
46 24 400.00 3.00 3.00 0.00 0.00 0.00 0.00 -1.00 -1.00 0.00 0.00
47 25 150.00 3.00 1.00 0.00 0.00 0.00 0.00 -1.00 -0.20 0.00 0.00
48 26 150.00 0.80 0.80 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
49 [End]
```

- b The data should have been properly pre-processed using such tools as described later in Section 6.13.

## 6.12 Manipulation Outline

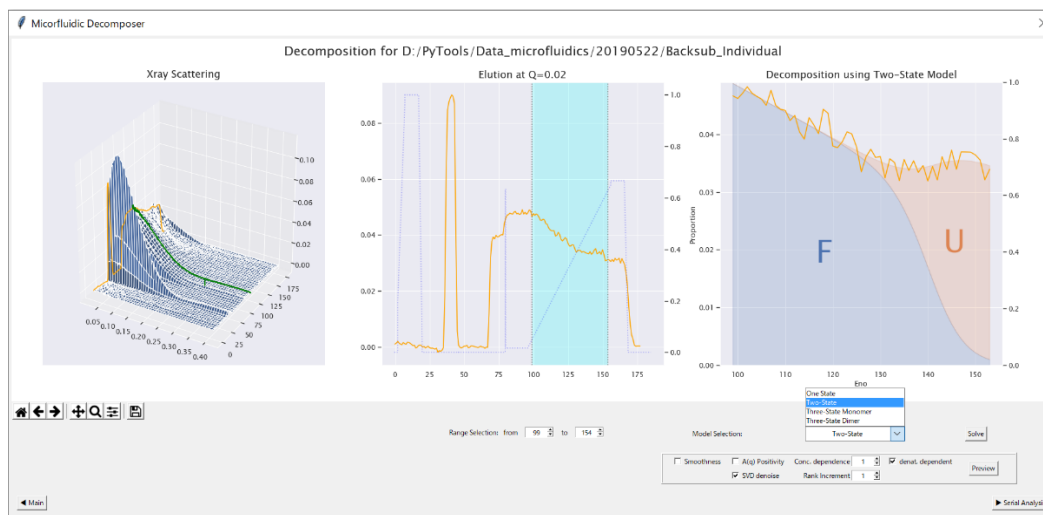
- a Basically, the manipulation for MCT-SAXS data is analogous to that of SEC-SAXS described the previous chapters.
- b Given a properly prepared set of data described above, the program recognizes the type of data and do the job accordingly.
- c The Main Dialog looks almost identical except that the UV file name in SEC-SAXS is replaced with the method file introduced in the previous section.

**Fig. 6-65 Main Dialog with MCT-SAXS Input**



- a There is no Mapping Dialog for MCT data and the next dialog comes as MTC Elution Decomposer shown below, which corresponds to Decomposition Editor for SEC data.

**Fig. 6-66 MCT Elution Decomposer**



- b An appropriate decomposition model should be selected from the model list.<sup>21</sup>
- c Subsequent manipulations are similar and their documentation is omitted in this version.

<sup>21</sup> As of this experimental support, available models are “Two State”, “Three State Monomer” and “Three State Dimer”

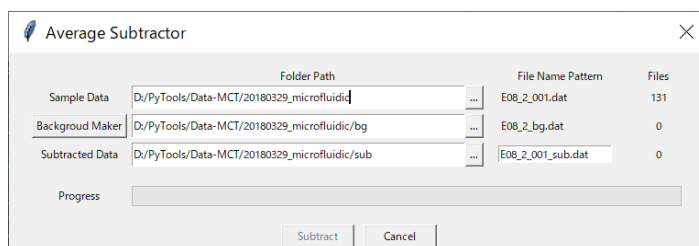
### 6.13 MCT Tools

Tools in this section are experimental and not yet officially supported.

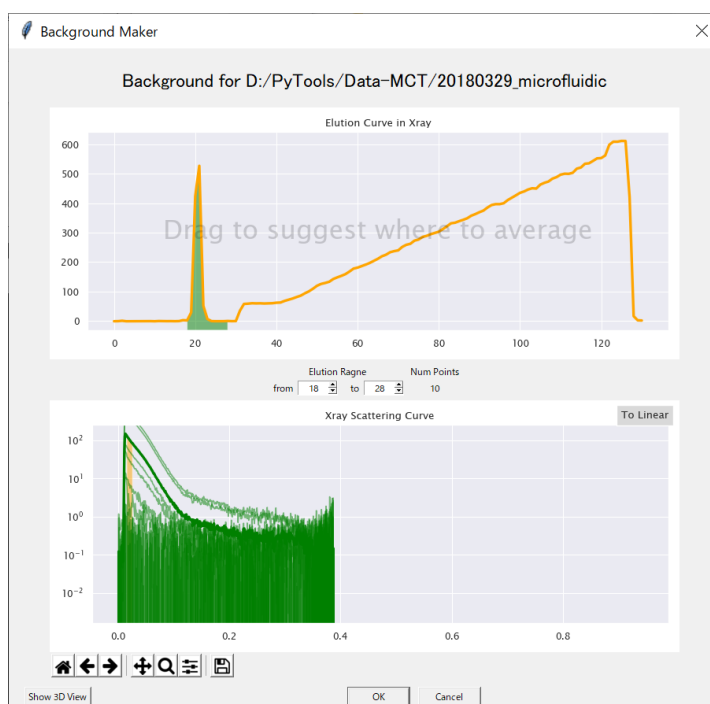
#### 6.13.1 Average Subtractor

- a This is a tool for background subtraction, which makes a background scattering curve averaging a specified elution interval and subtracts it from all the scattering curves. (Fig. 6-67)
- b The interval for the average can be altered in the Background Maker (Fig. 6-68), which is invoked by the button with the same name.

**Fig. 6-67 Average Subtractor**



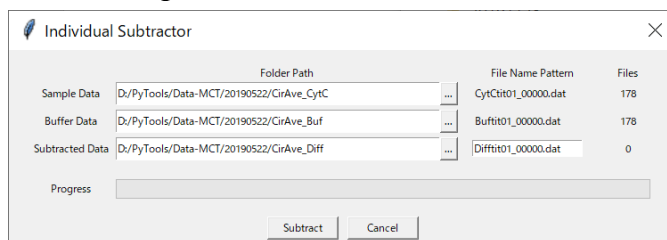
**Fig. 6-68 Background Maker invoked from Average Subtractor**



#### 6.13.2 Individual Subtractor

- a This is a tool for background subtraction, which, given a pair of folders including sample and buffer respectively, for each file in the sample folder, subtracts a corresponding file in the buffer folder. (Fig. 6-69)

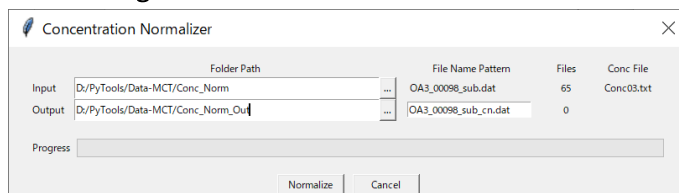
**Fig. 6-69 Individual Subtractor**



6.13.3 Concentration Normalizer

- a This tool divides each file in the input folder with a corresponding absorbance curve in the concentration file with .txt extension in the same folder.

**Fig. 6-70 Concentration Normalizer**



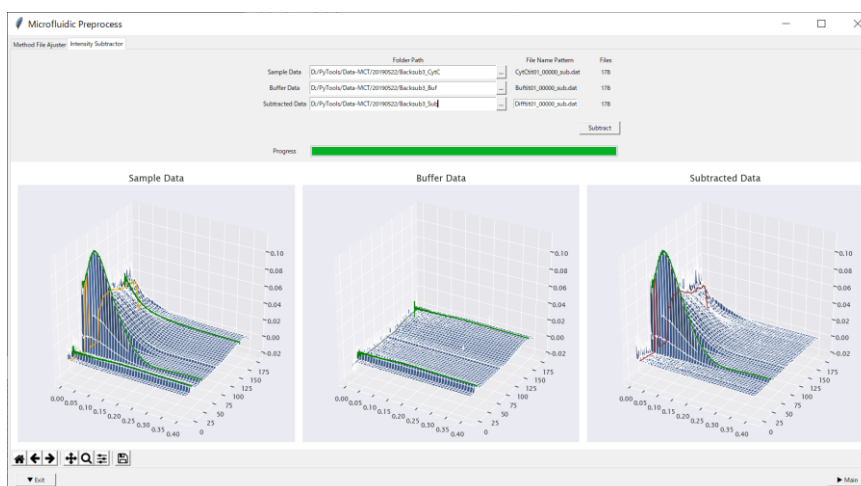
6.13.4 Elution Curve Picker

See Section 6.9.2.

6.13.5 Preprocess (Individual Subtractor with 3D Visualization)

- a This is yet another implementation of an individual subtractor which comes with 3D visualization.

**Fig. 6-71 Individual Subtractor with 3D Visualization**

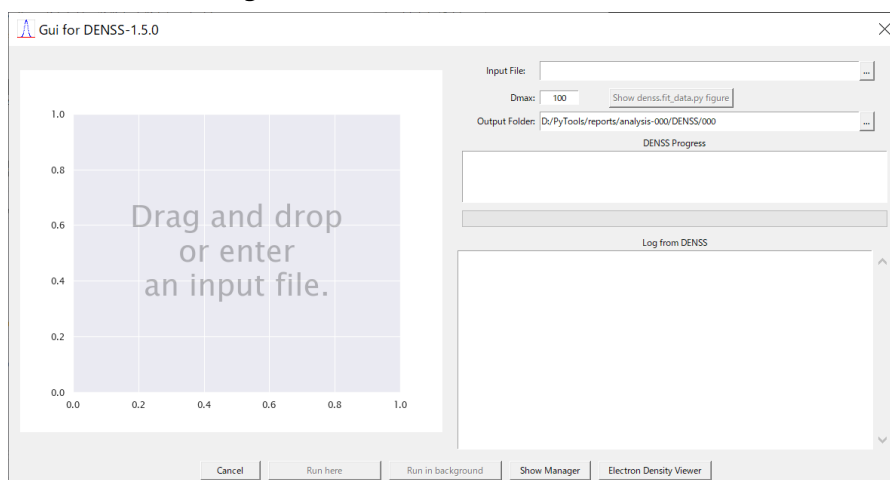


## 6.14 DENSS Tools

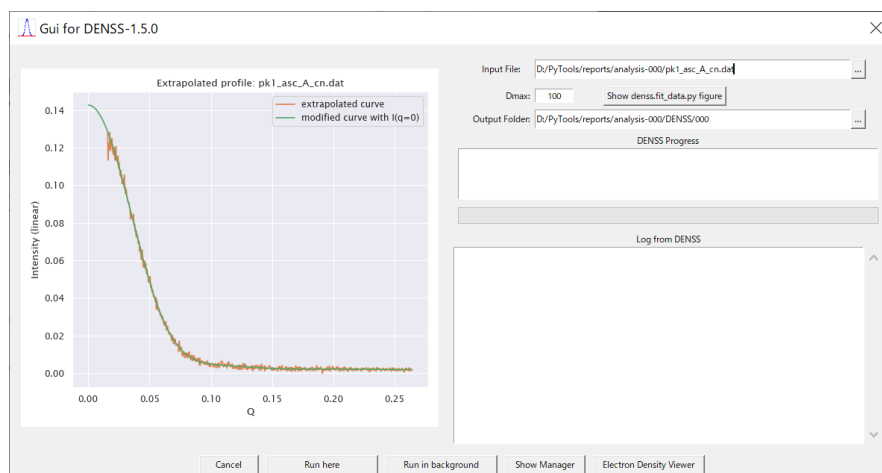
### 6.14.1 DENSS GUI

- This GUI is almost identical to the one introduced in Section 6.7.8, where previewed curves could be passed to a local [DENSS](#) program.
- Difference here is the way that the input curve is given from a file, in contrast to the previous way where it was given from a curve data in memory.
- The input can be specified either by drag and drop to the empty canvas or entry field, or using the selection button beside the entry field.

**Fig. 6-72 DENSS GUI (initial state)**



**Fig. 6-73 DENSS GUI (when given a curve)**

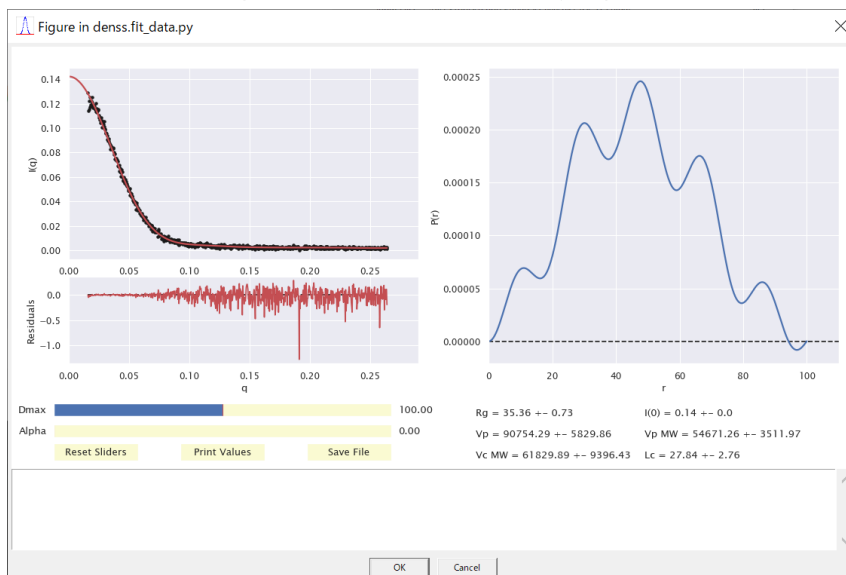


- After the input, the run buttons should be active, and users can run the job either here in the GUI or in background.
- Before running, it is recommended to specify “dmax” after manually determining its proper value using either the dens.fitr\_data.py from the button here (Fig. 6-74), or GNOM in ATLAS program suite. (See also [DENSS tutorial](#))

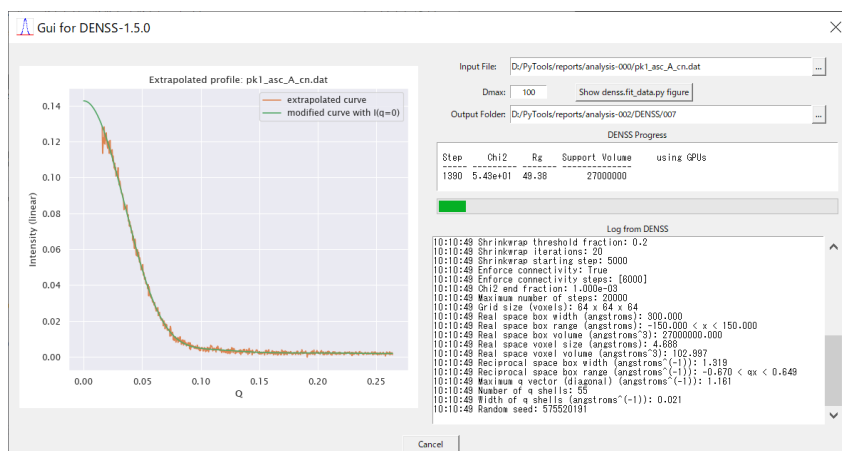


- c When the “Run background” button is pressed, “DENSS Manager” dialog will appear, which is explained in the next section.

**Fig. 6-74 denss.fit\_data.py figure**



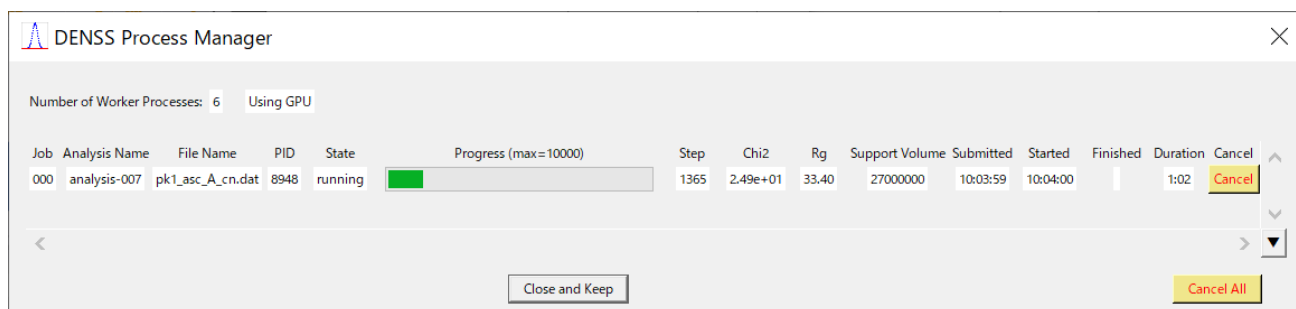
**Fig. 6-75 DENSS GUI (when run here)**



### 6.14.2 DENSS Manager

- a “DENSS Manager” dialog is for monitoring the DENSS jobs, which are initiated by run background buttons either in “DENSS GUI” or in “DENSS Menu” in the Preview Dialog. (See also 6.7.7 Background Execution of DENSS jobs)
- b The jobs are identified by three digit number such as “000”, and listed with progress information.
- c The output files will be placed in subfolders, named according to job number, in a subfolder named “DENSS” in the analysis result folder. For example, the output subfolder will be “analysis-007\DENSS\000” for the job shown below in Fig. 6-76.

**Fig. 6-76 DENSS Manager**

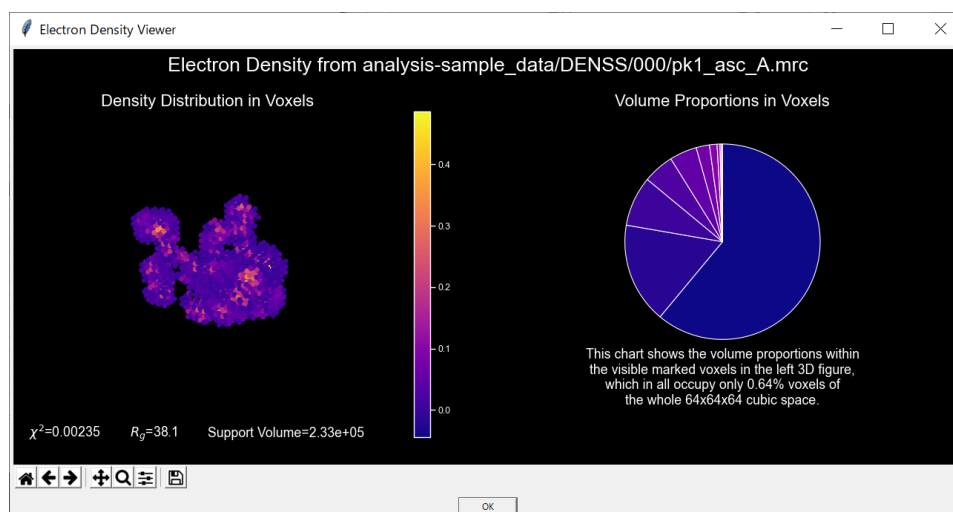


- d Since the jobs are run in background, we can safely close the dialog and re-open it as we like to monitor them.
- e Re-opening can be performed from DENSS Tools menu in the Main Dialog or the Extrapolation Preview.
- f Execution using GPUs will be enabled if the machine is equipped with NVIDIA compatible GPUs and you have [CUDA Toolkit 10.2](#) installed. See Section 6.8.8 to verify GPU availability in your environment.

### 6.14.3 Electron Density Viewer

- a MRC files, which are among the output files from DENSS, can be visualized using the “Electron Density Viewer” as shown in the following figure.

**Fig. 6-77 Electron Density Viewer**

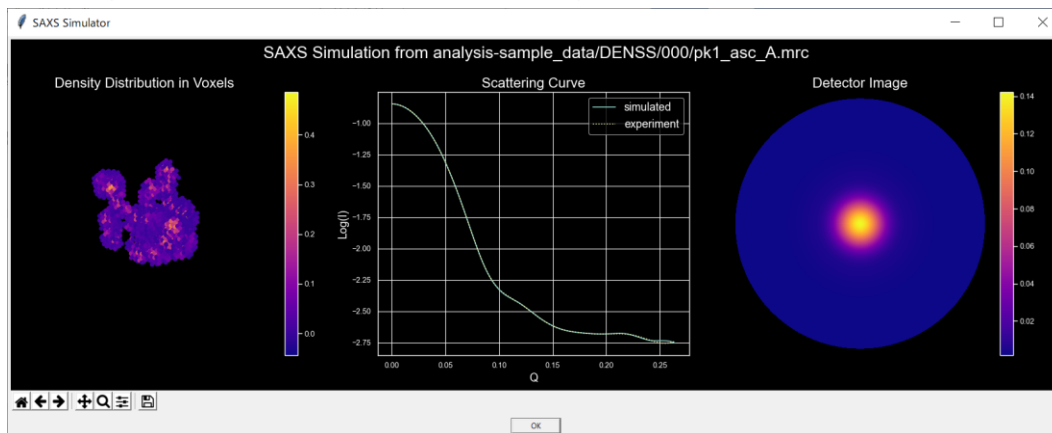


### 6.14.4 SAXS Simulator

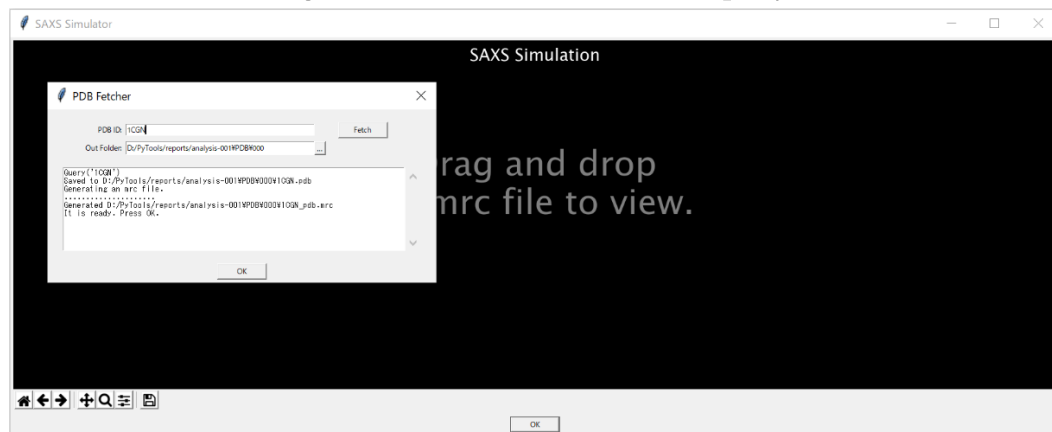
- a Users can also use “SAXS Simulator” to verify the DENSS results as shown in Fig. 6-78.
- b If you are interested in comparing the results with those in PDB, use this simulator by selecting “from PDB” from the right-click menu, which are shown below in Fig. 6-79 and Fig.

6-80.

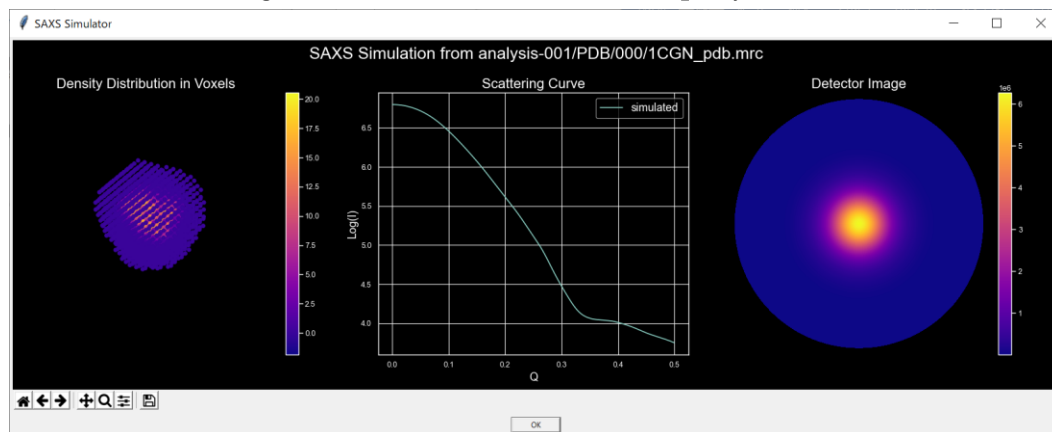
**Fig. 6-78 SAXS Simulator (comparing the simulated and experiment)**



**Fig. 6-79 SAXS Simulator (PDB query)**



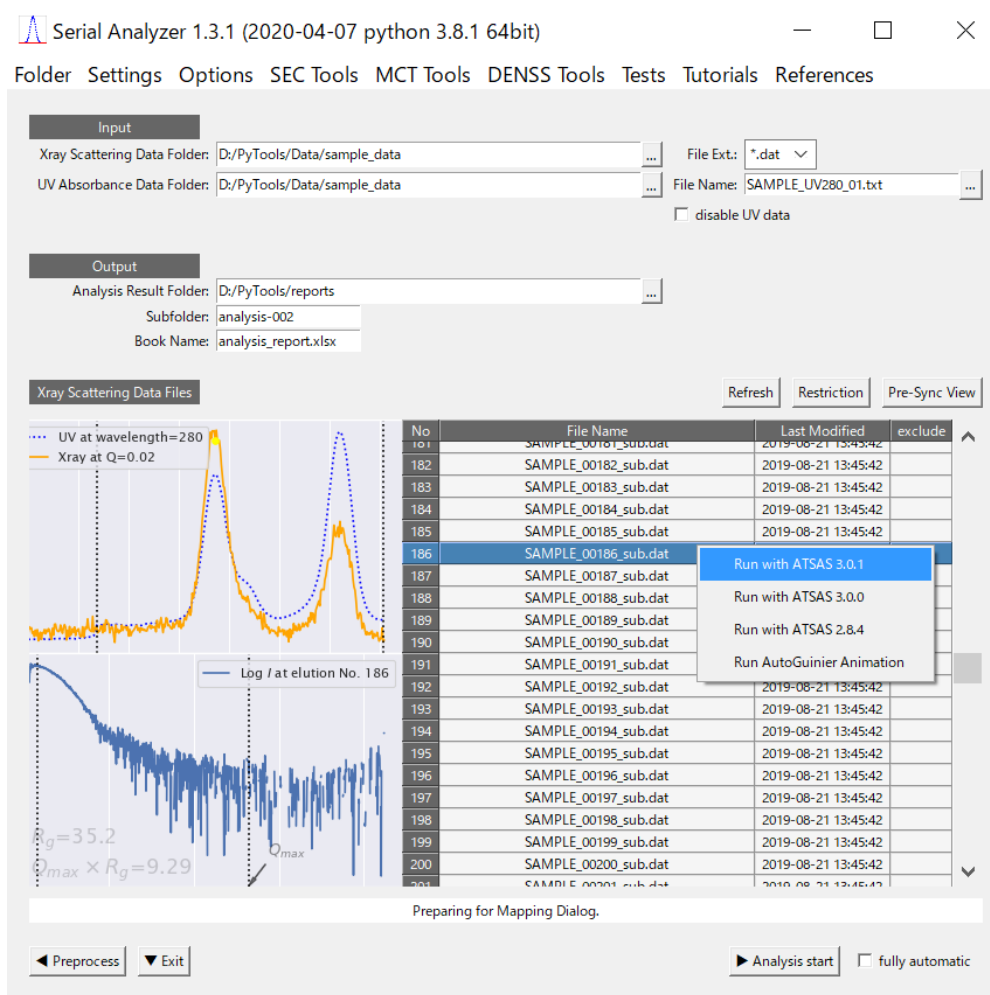
**Fig. 6-80 SAXS Simulator (PDB query result)**



## 6.15 AutoGuinier Interface

- The main dialog (Fig. 6-81) can be used as a GUI to AutoGuinier program.
- For such a usage, double click a desired row to get the result in a dialog as shown in Fig. 6-82.
- Or alternatively, if ATSAS programs are available, select a row and right-click to show the popup menu and select one of the options, which correspond to ATSAS versions specified in the Settings Dialog described in Section 6.8. (Double clicking invokes the first option)

**Fig. 6-81 Single Invocation of AutoGuinier Program**



- The result dialog shows the AutoGuinier result and AUTORG result in a table and three figures as shown in Fig. 6-82.
- The figures at center and right are zoomed ones of the left figure.
- The table shows the estimated  $R_g$ ,  $I(0)$  and related quantities and quality evaluation.
- From the popup menu, you can invoke an animation that explains the algorithm uses in AutoGuinier program (See Fig. 6-83).
- Scoring factors used to determine the appropriate Guinier region are listed in

Fig. 6-82 AutoGuinier Result Dialog

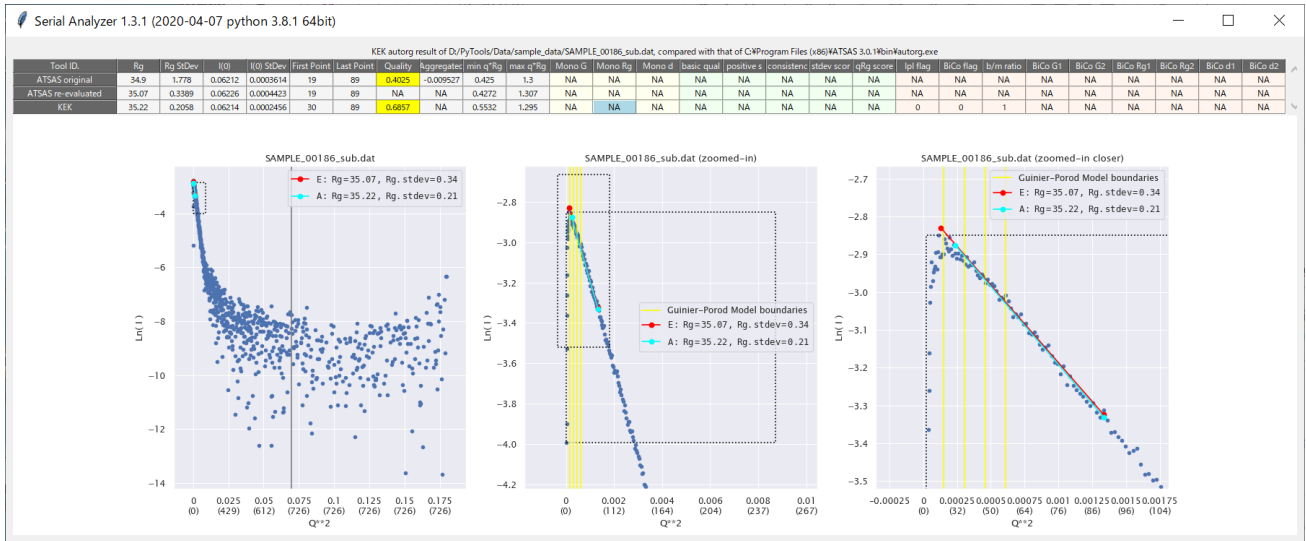
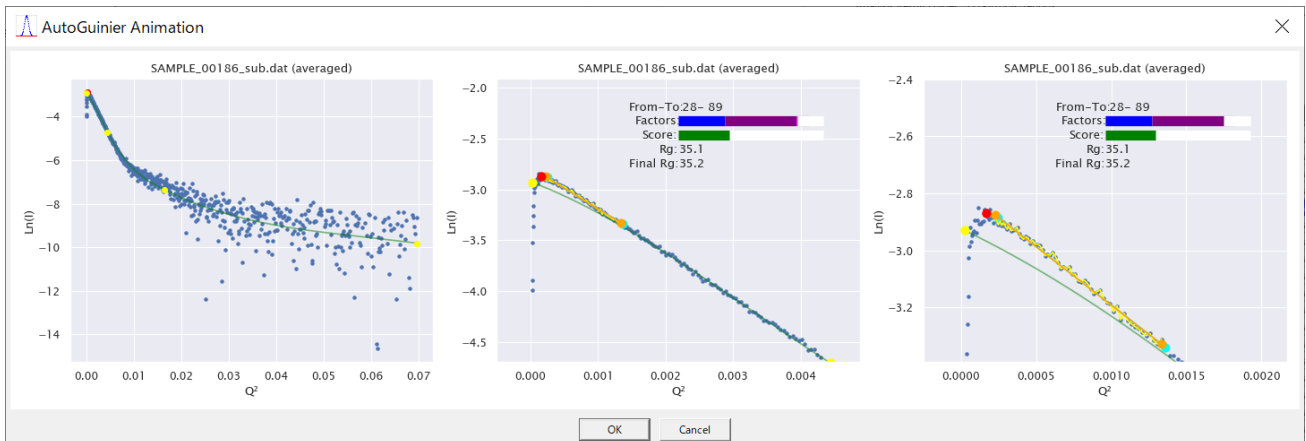


Fig. 6-83 AutoGuinier Animation



**Tab. 6-11 Guinier Region Scoring Factors**

No	Factor Name	Value Range	Weight H	Weight MH	Weight ML	Weight L	Formula	Description
1	size score	0 – 1	0.4	0.6	0.7	0.8	Rg_Consistency(Spline_Rg, Measured_Rg)	Computed from the Rg values of the spline curve and the measured data curve.
2	linearity score	0 – 1	0.5	0.3	0.3	0.2	$\exp(-(r\_value + 1) \times \text{LINEARITY\_SCALE})$	Computed from the r_value of linear regression with. LINEARITY_SCALE = 5
3	end consistency	0 – 1	0.06	0.06	0.0	0.0	Rg_Consistency(SA_End_Rg, LA_End_Rg)	Computed from the Rg values at the small angle side end and the large angle side end.
4	forward consistency	0 – 1	0.04	0.04	0.0	0.0	Rg_Consistency(Extended_SA_End, SA_End)	Computed from the Rg values at extended small angle side end and the small angle side end.

$$\text{Rg\_Consistency}(\text{Rg}_1, \text{Rg}_2) = \exp(-\text{abs}(1 - \text{Rg}_1/\text{Rg}_2) * \text{RG\_CONSISTENCY\_SCALE})$$

$$\text{RG\_CONSISTENCY\_SCALE} = 20$$

## 7 X-ray Data only Mode

### 7.1 Design Philosophy

- a SerialAnalyzer is designed to analyze a series of X-ray data coupled with a set of UV data measured simultaneously in the same experiment.
- b This compound simultaneous measurement strategy has been sought to utilize the merits such as:
  - 1. UV absorbance is free from inter-particle effects and more suitable for the evaluation of concentration especially in the small angle (or low  $Q$ ) regions.
  - 2. With UV data, absolute concentration values can be attached.
  - 3. Difference in the curve forms between X-ray scattering and UV absorbance can be utilized to the decomposition of the elution curves.
- c On the other hand, demerits such as the following may be listed.
  - 1. It is not always easy to reconcile the contradiction between the two different types of measurement data.
  - 2. The analysis programs tend to become more complex and costly.
- d When those merits are considered less important, it is desirable to let the program use X-ray data only (i.e. without UV data).

### 7.2 Implementation

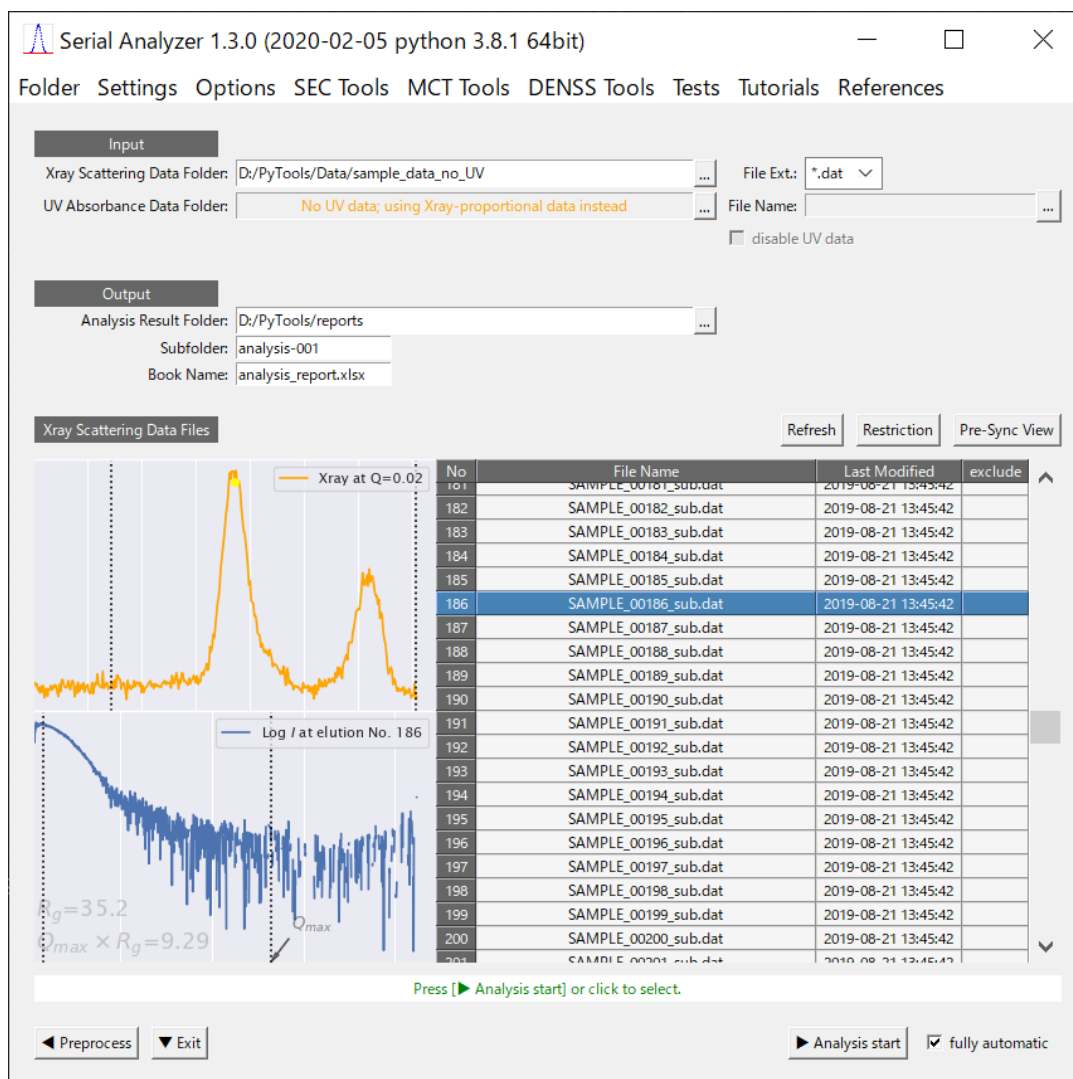
- a It is realized actually by computing the (relative) concentration only from X-ray data and suppressing the mapping functionality.
- b More specifically stated, it is implemented in the following way.
  - 1. Pseudo UV data is generated so that its elution curve be exactly proportional to that of X-ray data in the standard mapping plane.
  - 2. Therefore, the mapping between UV and X-ray is an identity mapping and exact.
  - 3. UV data is so scaled that the maximum value in the standard plane be 1.
- c Different points of manipulation in this case will be described below.

### 7.3 Different Manipulation Points from the Normal Usage

#### 7.3.1 Main Dialog

- a When there exists UV data file in the input folder, then the Main Dialog reflects such situation by disabling the UV-related entries as shown in Fig. 7-1.
- b This X-ray only mode can be activated by checking “disable UV data” button even in those situations where UV data file actually exists. (See also Section 6.1.6)
- c Usage of “Restriction” tool is limited to the X-ray data.

**Fig. 7-1 Main Dialog in X-ray Data only Mode**

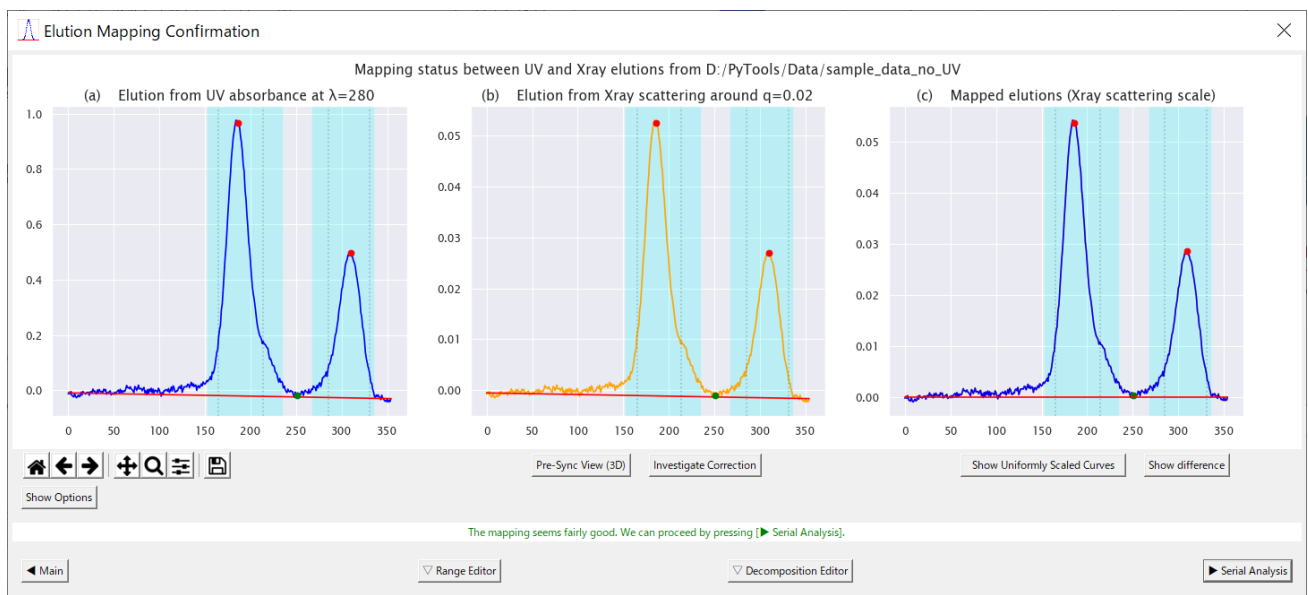


### 7.3.2 Mapping Dialog

- a Mapping is not necessary when you are using X-ray data only.
- b Nevertheless, this dialog will appear in the same manipulation flow as in the normal usage because some optional settings for X-ray data are still valid.
- c In this case, however, optional settings for UV data at lower left are suppressed as shown in Fig. 7-2.
- d Moreover, at the right part, display of SCI values in the “(c) Mapped elutions” figure or nRMSD values in “Mapping Precision Measure” is suppressed because the mapping is always perfect.



Fig. 7-2 Mapping Dialog in the Case of X-ray only Data



## 8 Implementation Details

### 8.1 AutoGuinier

The following sections are in preparation.

#### 8.1.1 Outline

- a The easiest way to understand the outline of implementation of AutoGuinier is to see the animation introduced in Section 6.15, Fig. 6-83.

#### 8.1.2 Command Line Interface

#### 8.1.3 Preliminary Process

#### 8.1.4 Scattering Curve Smoothing

#### 8.1.5 Guinier-Porod Model Fitting

#### 8.1.6 Guinier Interval Determination

#### 8.1.7 Estimation of $R_g$ and $I(0)$

#### 8.1.8 Evaluation of Quality Factors

## 8.2 SerialAnalyzer

The following sections are in preparation.

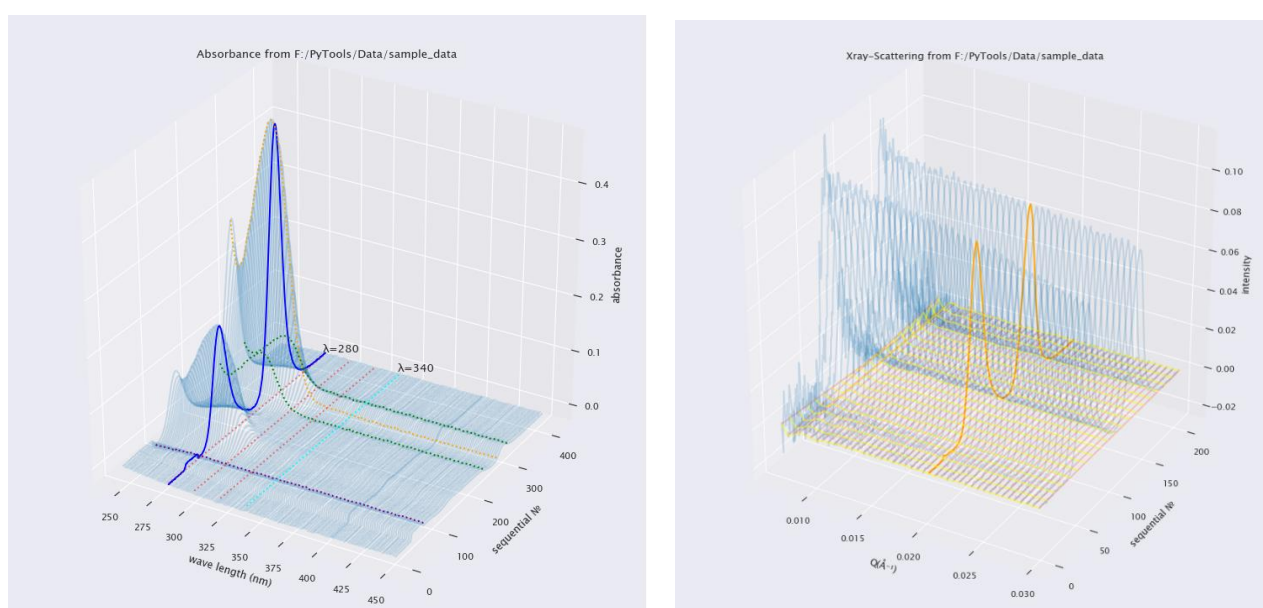
### 8.2.1 Outline

### 8.2.2 Used Data Range Restriction

### 8.2.3 Standard Mapping Plane

- a Both UV absorbance and X-ray scattering data from a serial experiment have similar 3D structures as seen in Fig. 8-1, such that they have, firstly, an elution (or time) axis in common, secondly, another dimension (wave-length or scattering vector length) axis, and finally, the amount (absorbance or intensity) axis.

Fig. 8-1 3D structure of UV absorbance and X-ray scattering data



- b Although they have a common elution (or time) axis, the measurement density (or number of elution points) are usually different.
- c Therefore, we have to make an appropriate mapping in order to get a corresponding concentration value of each scattering intensity.
- d For that purpose, we select a plane for each data set, where the elution curve is observed as clearly as possible, and we call it the standard mapping plane or simply standard plane.
- e In the above figure, the elution curves in the standard mapping plane are highlighted in bold

blue and orange lines.

- f The positions for the standard plane are usually chosen at the wavelength  $\lambda = 280\text{nm}$  for UV absorbance and the scattering vector length  $Q=0.02$  for X-ray scattering.
- g In later descriptions, unless otherwise stated, elution curves in this standard plane are considered.

#### 8.2.4 Peak Recognition of the Elution Curves

- a Similar elution curves are observed in both UV absorbance and X-ray scattering.
- b Therefore, we can recognize their peaks almost in the same way, except that the former have flow-change points, which will be addressed in the next section.
- c

#### 8.2.5 Flow Change Point Detection in UV absorbance

- a Flow change points in UV absorbance data arise from such an operational reason that it is desirable to make shorter the experiment duration time from cost and scarcity.
- b Outside of flow change points, in other words in elution points before and after them, we should ignore the data because they are unreliable there.
- c

- 8.2.6 Optimization of the Mapping between UV absorbance and X-ray scattering
- 8.2.7 Range Determination for Zero-Concentration Analysis
- 8.2.8 Extrapolation to Zero-Concentration
- 8.2.9 UV absorbance baseline correction
- 8.2.10 X-ray scattering baseline correction
- 8.2.11 Intermediate files

## Appendix

### A Note on the Meaning of Error on Estimated Rg

- a We note here that it is just a propagated Gaussian error, on the condition that the Guinier interval is properly determined, and that we do not take into account the other possibilities like improperly determined intervals or cases where multiple Guinier intervals arise from a sample with multiple components.
- b The reason why we don't do so is that we believe such other possibilities should be expressed in other ways, because it could be confusing if different kind of errors from different probability distributions were summed up into one single monotone amount.
- c For instance, suppose a simplified situation depicted in Tab. A-1 where there are only three possibilities.

**Tab. A-1 Simplified probability distribution for discussion**

Candidate Guinier interval	Adoption probability	Estimated Rg	Error
Interval A	1/2	50	5
Interval B	1/4	80	8
Interval C	1/4	40	6

- d The errors in this table are assumed to be usual propagated Gaussian errors.
- e In this situation, suppose further that you have computed the expected Rg and its error as follows.

$$\text{Expected}(R_g) = 50 \cdot \frac{1}{2} + 80 \cdot \frac{1}{4} + 40 \cdot \frac{1}{4} = 55 \quad (\text{A-1})$$

$$\text{Stdev}(R_g) \cong \sqrt{(50 - 55)^2 \cdot \frac{1}{2} + (80 - 55)^2 \cdot \frac{1}{4} + (40 - 55)^2 \cdot \frac{1}{4}} = 15 \quad (\text{A-2})$$

- f In the calculation (A-2), Gaussian errors have been neglected to avoid complexity.
- g Using this example, let us re-state the assertion made at the beginning of this appendix.
- h Even if the situation were like this, our program would choose only one of the possibilities, say interval A, and give its estimated Rg 50 and its error 5 as the result, instead of trying to answer like (A-1) or (A-2).
- i The reason why we would do so is as follows.
  1. Errors computed like (A-2) indicate a different type of variations than Gaussian noises.
  2. The distributions for adoption probability in real world are not so simple as this example, and difficult to model, implement and understand.
  3. It is better to separate different kind of errors, namely Gaussian errors and other systematic errors, because the former are easy to understand and use, while the latter are peculiar and not established and algorithm-dependent.
  4. The other possibilities than choosing one, said as interval A in this example, can better be expressed in other ways.

## B Intermediate files

## C KEK-staff Options

- a Options in this appendix are provided only for KEK-staff for experimental purposes and not intended for general users.

### C.1 Specialist Options Dialog

Fig. C.1-1 Specialist Options Dialog

Specialist Options

Absorbance Data Treatment

Absorbance Curve Picking: averaging  points around wave length  $\lambda =$    
 $\lambda =$   for absorbance bottom line

Data Correction Options

enable X-ray Scattering Baseline Save  ...  
filename postfix  as in SAMPLE\_00000\_sub\_base.dat

Elution Mapping Options

do mapping with peak tops and end points only  
 enable LPM variations: linear, quadratic, spline  
 allow angular slope in MF-baseplane

Elution Curve Modeling

enable Affine Transformation

Guinier Analysis Options

Fix Guinier interval start point to the  -th point; looking at the scattering curve on the  -th peak

Extrapolation Options

enable Concentration Curve Options  
 enable Extended Concentration Dependence ( i.e.,  $I(q) = A(q)*c + B(q)*c^2 + Z(q)*c^3$  )  
 allow Rank Variation ( i.e., no SVD denoise, additional rank allowance )  
 enable Almerge Analyzer

Deprecated Extrapolation Options

Width of Regression in the direction of Q-axis  Five Points  Three Points  One Point  
Regression Boundary  Automatic  Fixed at   No Boundary  
A(q) Curve Construction in B(q)-Extinguished Region  Use max conc. only  Conc-weighted regression

Maintenance

enable Maintenance Mode



C.1.1 Absorbance Data Treatment

C.1.2 Data Correction Options

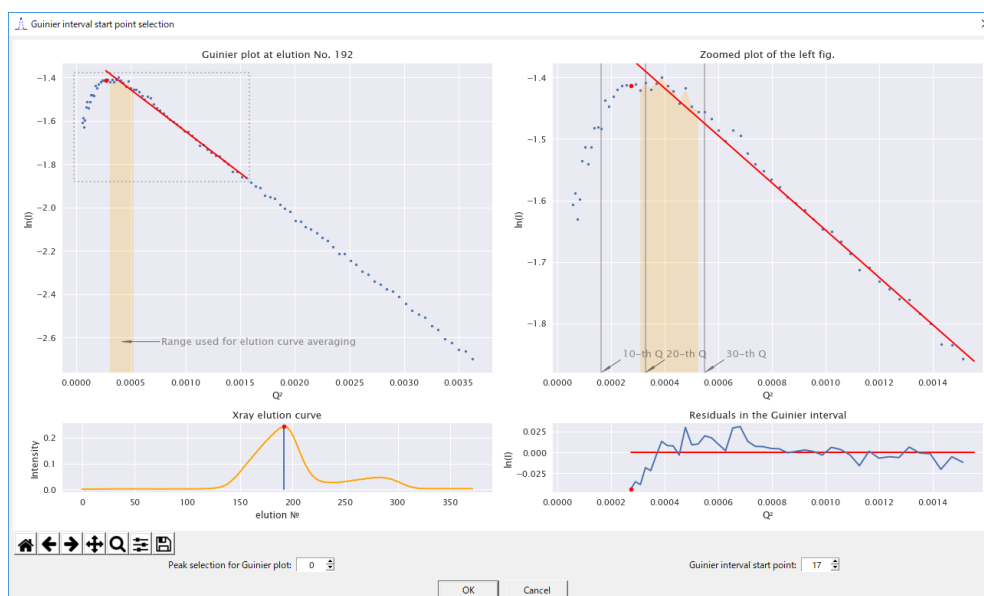
C.1.3 Elution Mapping Options

C.1.4 Elution Curve Modeling

C.1.5 Guinier Analysis Options

- a In Guinier analyses, the program may fail to identify the right Guinier region when the data are ill-conditioned.
- b To cope with such situations, you can directly specify the starting point of the region using a scattering curve at the top of a selected peak as shown in the figure below, which was invoked by the “Guinier plot” button.

**Fig. C.1-2 Fixing a start point of Guinier region**



- c The above figure example shows a fixed starting point to avoid the dropping portion in the very small angle caused by inter-particle effects.

C.1.6 Extrapolation Options

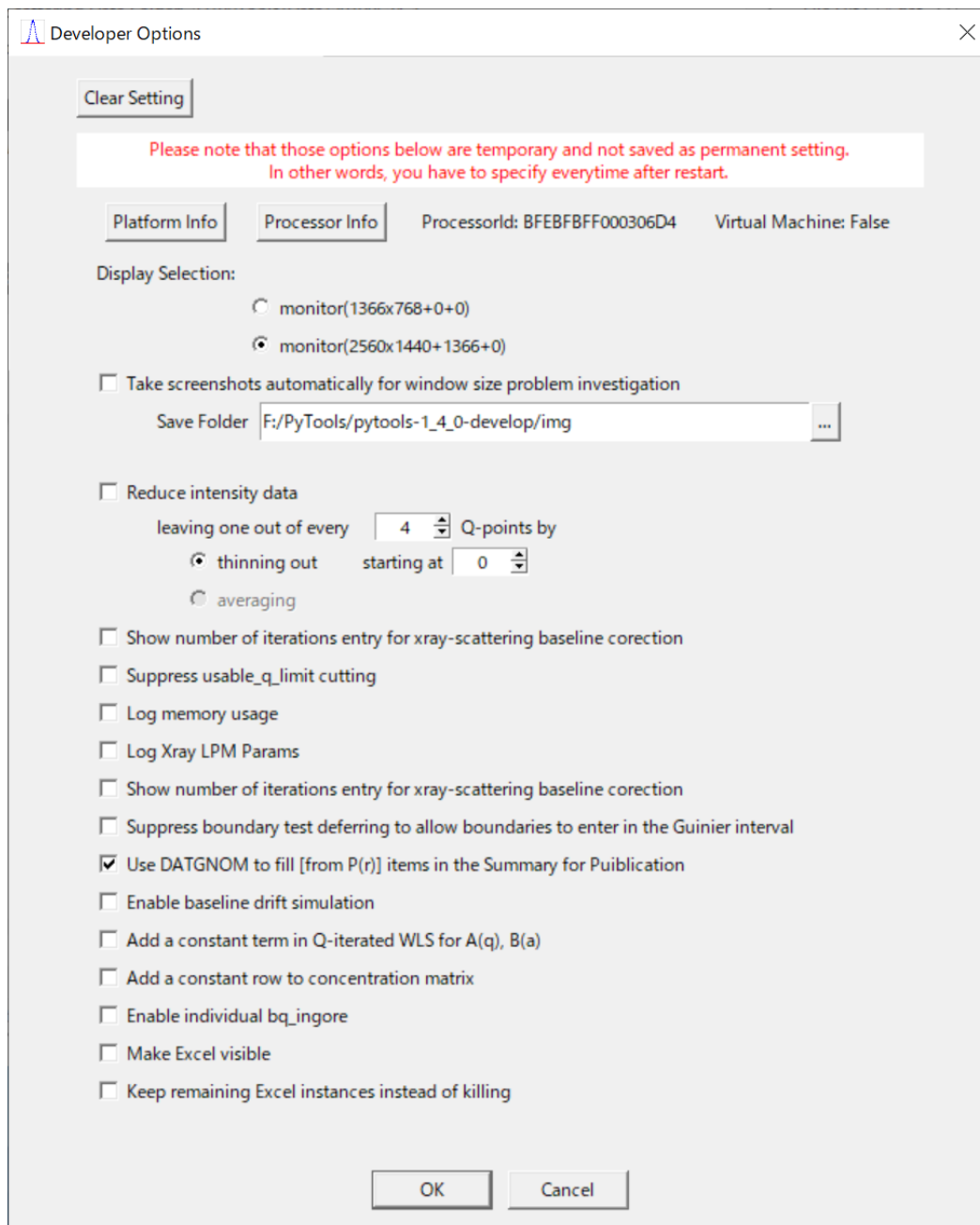
C.1.7 Deprecated Extrapolation Options

C.1.8 Maintenance Mode (for Guinier Analysis)

## D Developer Tools

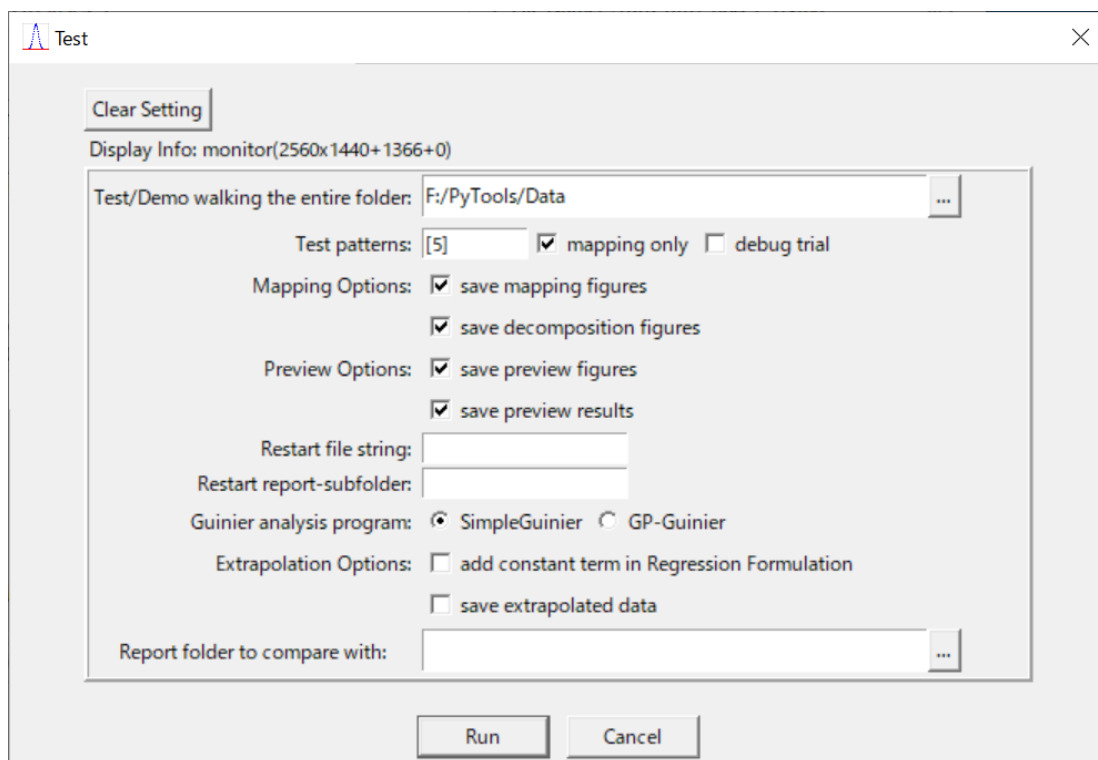
### D.1 Developer Options Dialog

Fig. D.1-1 Developer Options Dialog



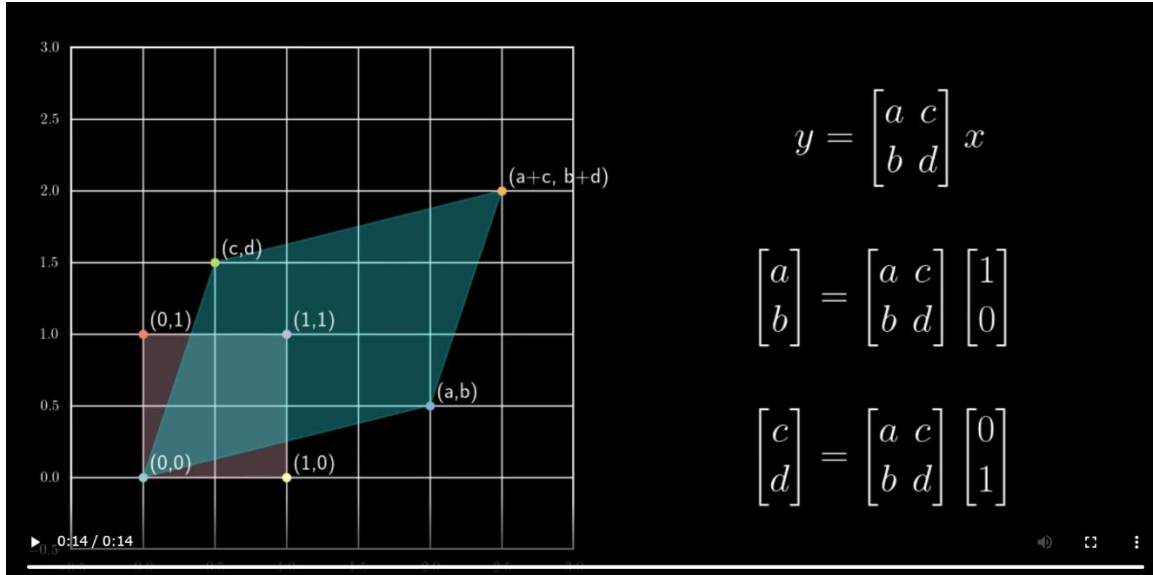
## D.2 Tester Dialog

Fig. D.2-1 Developer Options Dialog

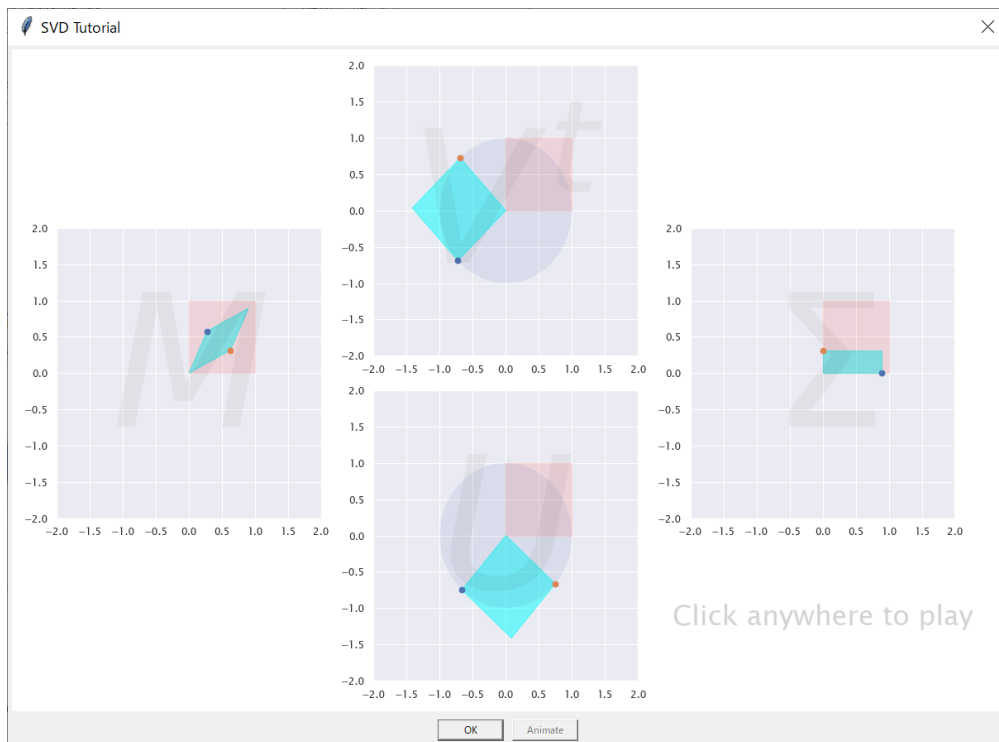


## E Tutorials

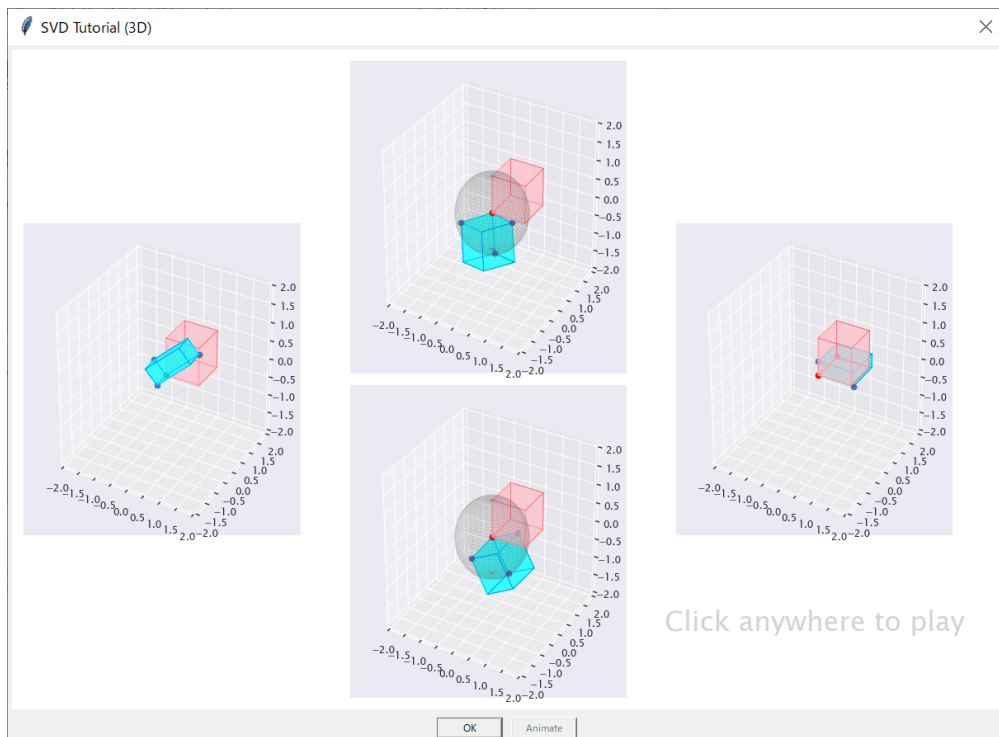
### E.1 Linear Transformation



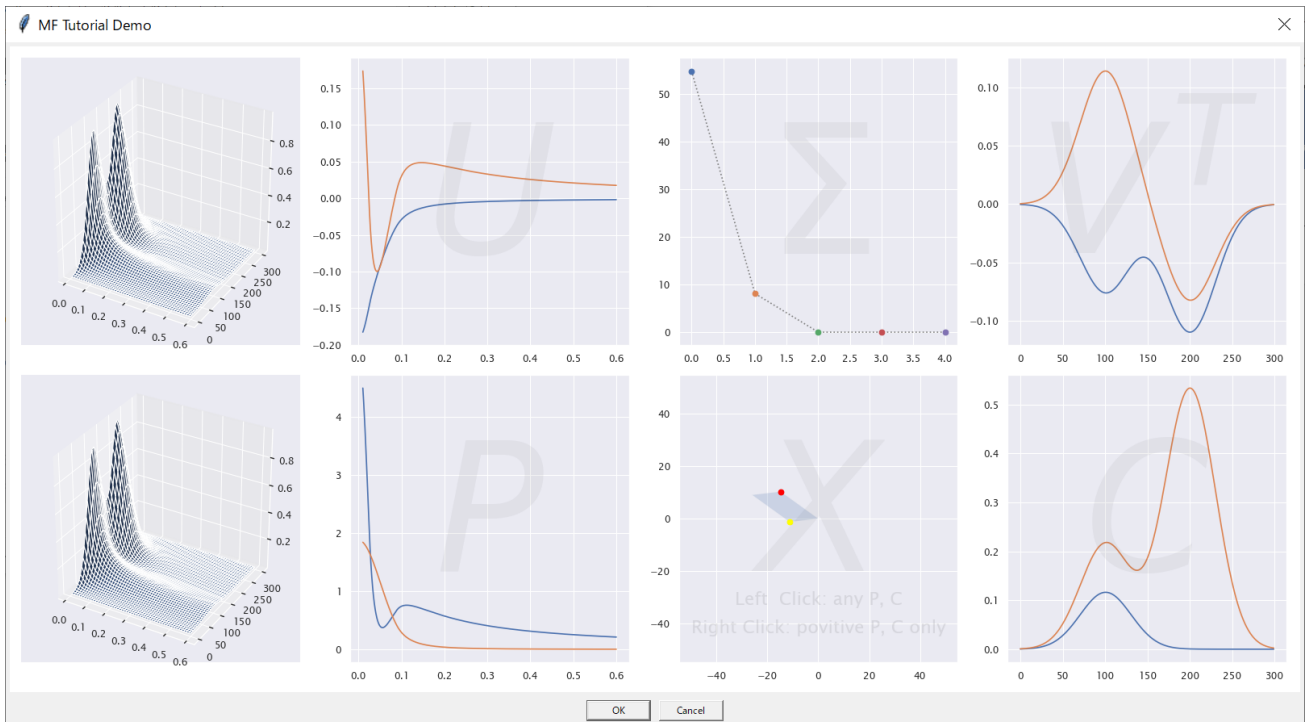
## E.2 SVD (2D)



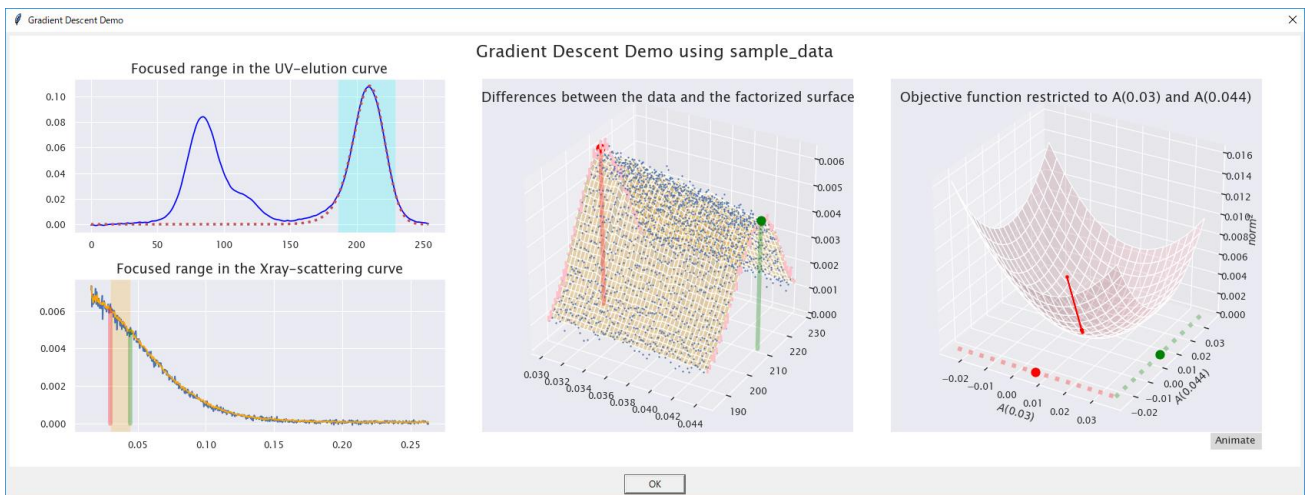
## E.3 SVD (3D)



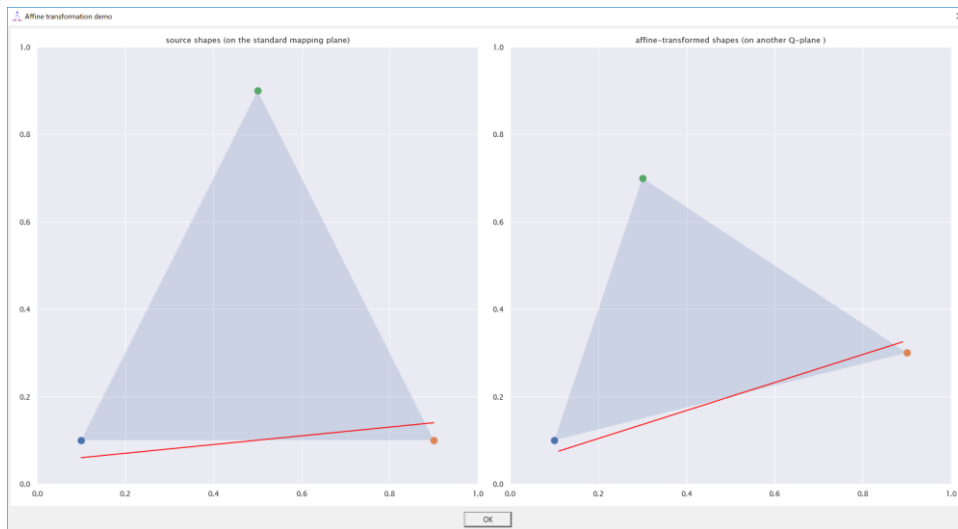
## E.4 Matrix Factorization



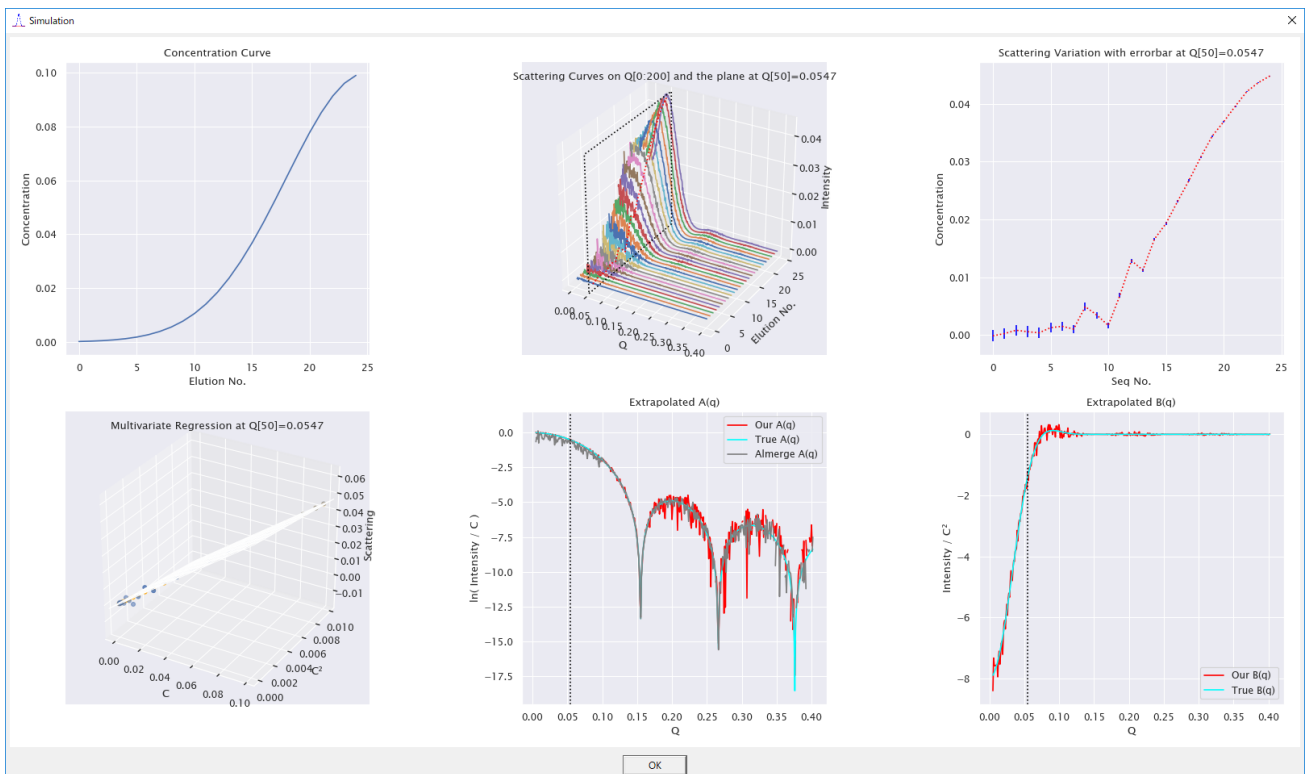
## E.5 Conjugate Gradient Algorithm



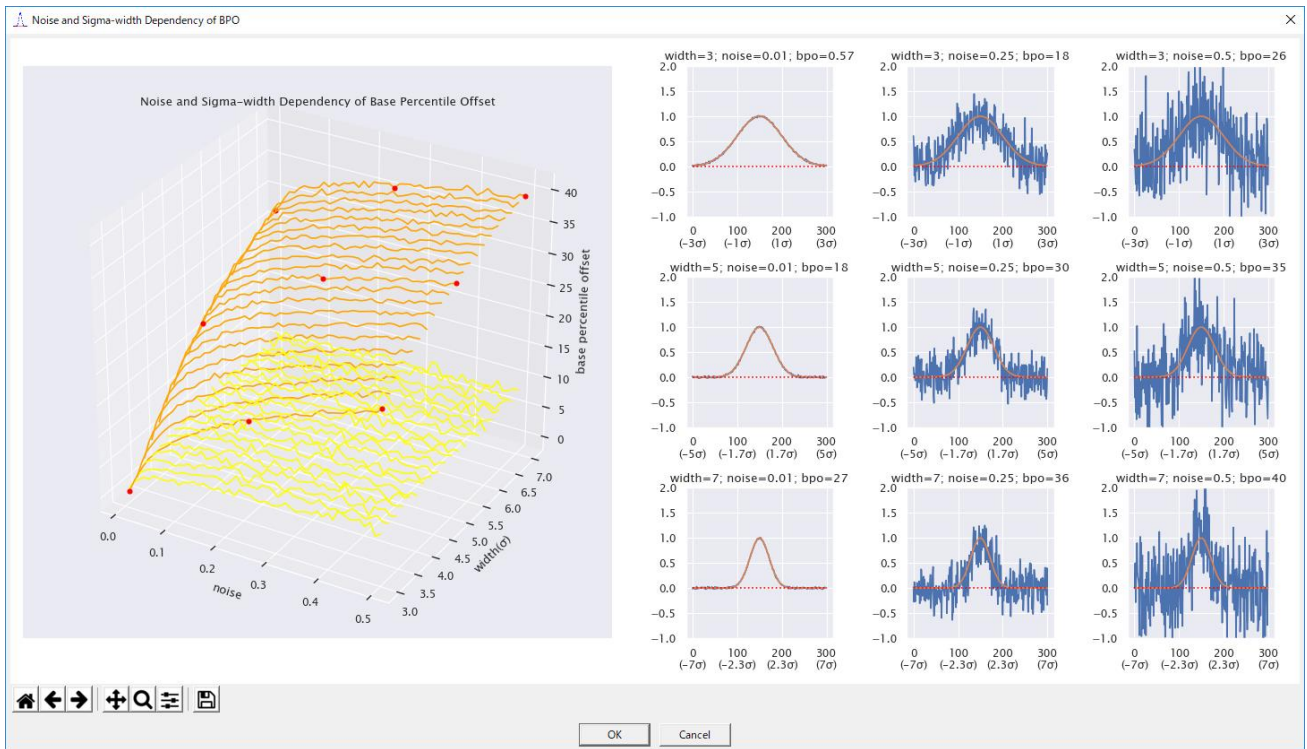
## E.6 Affine Transformation of Baseline Adjustment



## E.7 Extrapolation with Spherical Particles



## E.8 Noise and Sigma-width Dependency of Base Percentile Offset





## References

- [1] R. Penrose (1955). On Best Approximate Solutions of Linear Matrix Equations.
- [2] Feigin, L. A. & Svergun, D. I. (1987). Structure Analysis by Small Angle X-ray and Neutron Scattering. New York: Plenum Press.
- [3] Kevin Lan, James W. Jorgenson (2001). A hybrid of exponential and gaussian functions as a simple model of asymmetric chromatographic peaks.
- [4] David A. Jacques, J. Mitchell Guss, Dmitri I. Svergun and Jill Trehwella (2011). Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution.
- [5] Yuri Kalambet, Yuri Kozmin, Ksenia Mikhailova, Igor Nagaev, Pavel Tikhonov (2011). Reconstruction of chromatographic peaks using the exponentially modified Gaussian function.
- [6] Thomas D. Grant (2018). Ab initio electron density determination directly from solution scattering data.

## Index

### *A*

affine transformation .....33, 37  
analysis results.....9

### *B*

baseline correction .....32

### *C*

concentration .....8  
concentration factor .....47

### *D*

data quality.....9

### *E*

EGH.....41  
elution curve .....26, 82  
elution point.....7, 8  
EMG .....41

### *F*

flange.....21  
flow change point .....21, 83

### *G*

Guinier region .....88

### *I*

inter-particle effects ..... 14, 88

### *O*

outline figures ..... 21  
overfitting reduction..... 53

### *P*

peak-width ..... 23

### *R*

rank of matrix..... 52  
regularization ..... 53  
Rg-consistency ..... 61

### *S*

SCI..... 3, 27  
standard mapping plane ..... 26, 29, 60, 82

### *U*

UV absorbance data ..... 8

### *X*

X-ray scattering data ..... 7