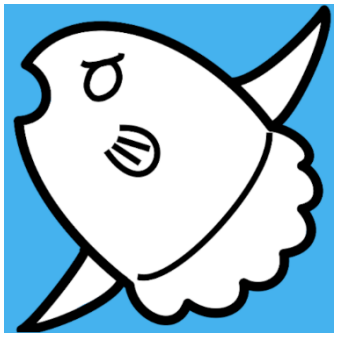
MOLASS User's Guide

for ver. 1.0.12



Mola mola

2023-06-28

Masatsuyo TAKAHASHI, Kento YONEZAWA and Nobutaka SHIMIZU

Photon Factory

Institute of Materials Structure Science

High Energy Accelerator Research Organization, KEK

Contents

0 Introduction

a This document describes MOLASS[1] — 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 infinite dilution

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: https://doi.org/10.2142/biophysico.bppb-v20.0001.

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 7.

b X-ray scattering data in the input folder 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.[2]

d To follow the steps below, users can use an example set of data from the download site.

e If you are to use it in the easiest way, proceed as follows.

1. Run the "molass.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.

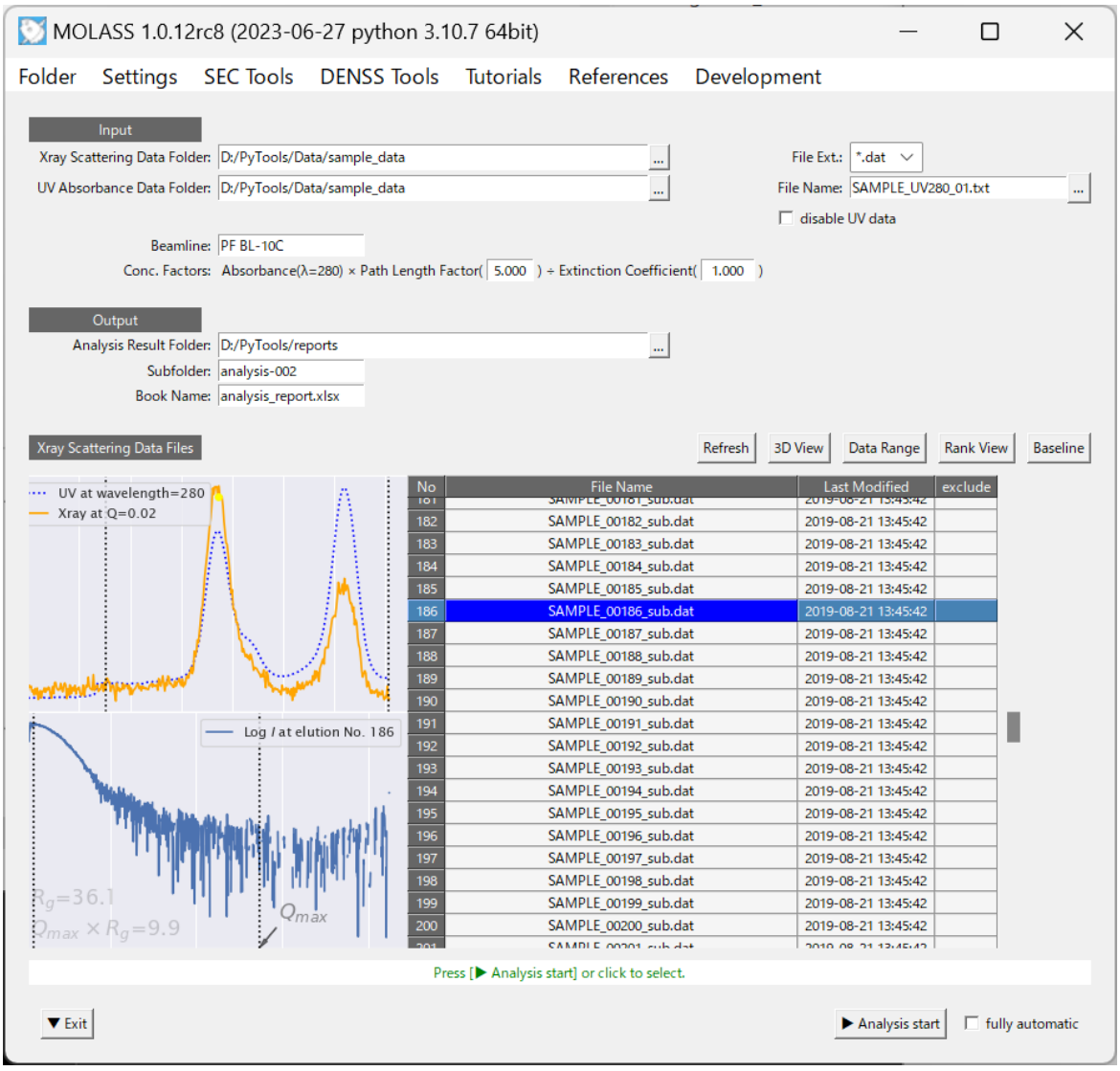
g That is all for the easiest usage.

---

[1] We renamed the program from SerialAnalyzer to MOLASS — Matrix-based Optimization with Low rank factorization for Automated analysis of SEC SAXS — in order to better imply its character after it has been changed significantly over years of groping started in January, 2016.

[2] Handling of those X-ray data not coupled with UV data is possible and described separately in Chapter 8.

Fig. 2-1　Main Dialog



3　Standard Procedure

3.1　Basic Steps

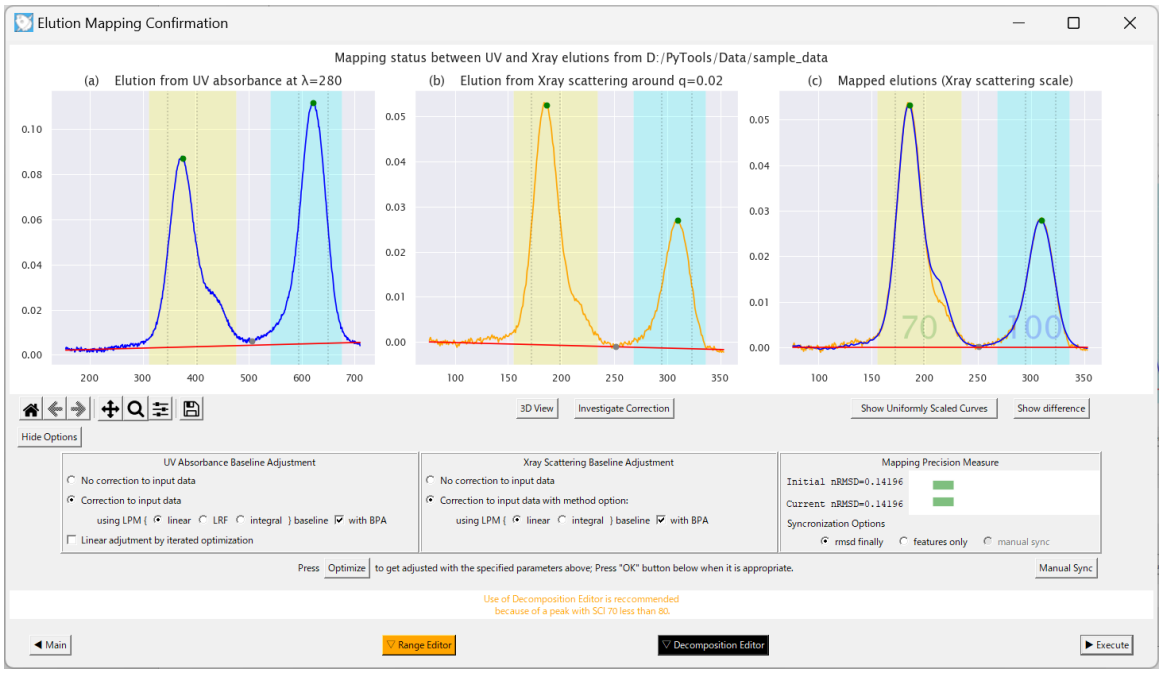
a　The following are the basic steps automatically performed when you press "▶ Analysis start" button with "fully automatic" checked in the above main dialog (Fig. 2-1).

b　You should follow these steps manually unless you choose "fully automatic". (Here is described only how to manipulate. See Chapter 7 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[3] of those two.

[3] UV elution curve in this figure has been locally scaled to make it possible to compare. See Section 7.2 for details.

Fig. 3-1　Mapping Dialog



3. At this stage, if needed[4], you can further use “Decomposition Editor” to better control the extrapolation (LRF) process. In that case, the “Decomposition Editor” button instead will be colored green (Fig. 3-2) after you have applied decomposition, suggesting that the decomposition result be used thereafter. See the next section 3.2 for such control.
4. In the “fully automatic” mode, a decomposition step using “Decomposition Editor” is performed depending on the data condition. It is performed, currently, when some of the peaks are suspected of multi-component, judging from the following conditions.
   ① There exists a peak with SCI[5] less than 80.
   ② Or, there exists a significant[6] overlapping between adjacent peaks.
   This decomposition is not necessary when these conditions are not met.
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 “▶ Execute” button, which executes the following steps.
   ① Baseline correction of X-ray data
   ② Guinier Analysis of all scattering curves

---

[4] This decomposition step is required when the peaks seem to include multiple components.
[5] Single Component Indicator. See Section 7.2.2 for its definition.
[6] It is judged significant if the peaks have some overlapping in their feet above 5% in height against their peak tops, measure by modeled elution curves.

③ Extrapolation to infinite dilution

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

Fig. 3-2 Mapping Dialog after the decomposition

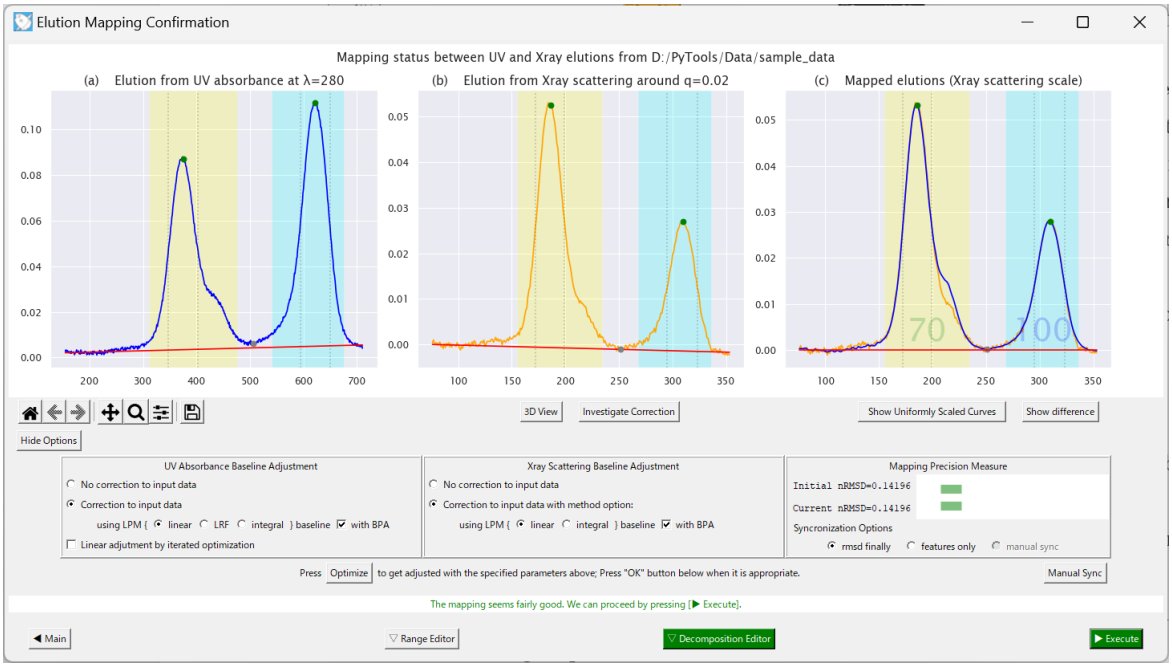
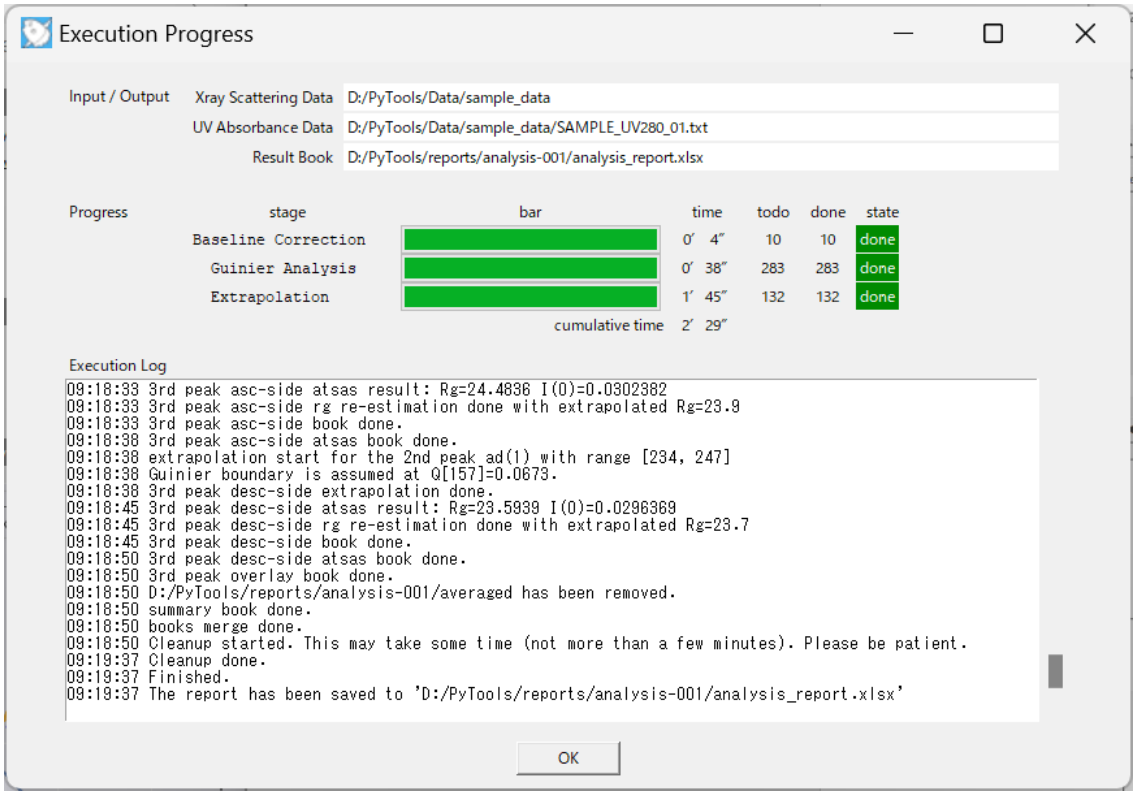


Fig. 3-3 Progress Dialog when it is finished

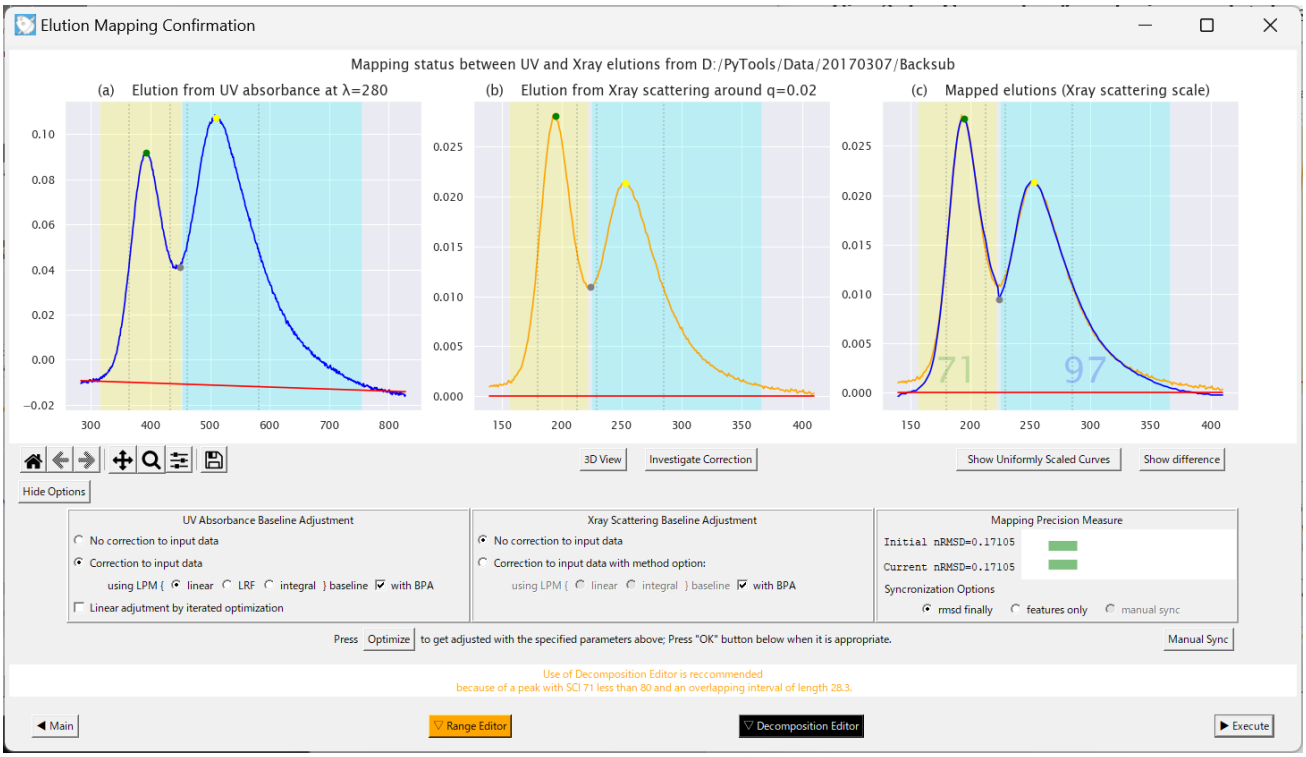


3.2 Decomposition Procedure

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 peaks incompletely separated



c In such situations, decomposition of the total elution curve is required to make available elementary elution curves each of which separately represents the concentration of its 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[7] or EGH[8]) as shown in Fig. 3-5 and Fig. 3-6 below.

---

7 Exponentially Modified Gaussian model.

8 Exponential-Gaussian Hybrid model

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

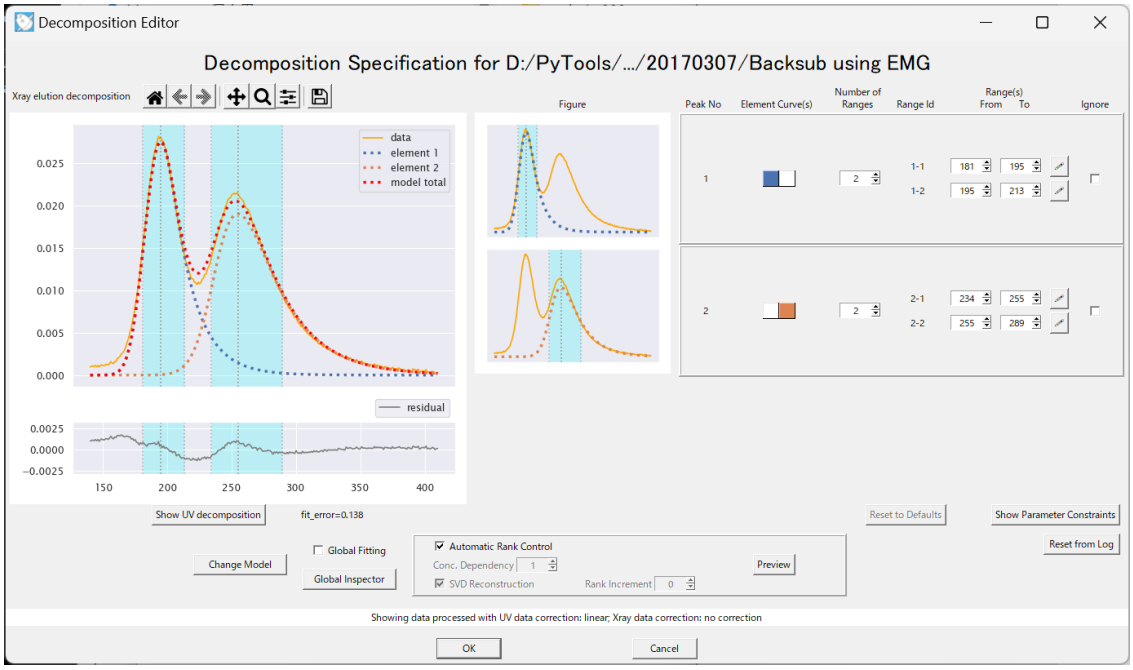
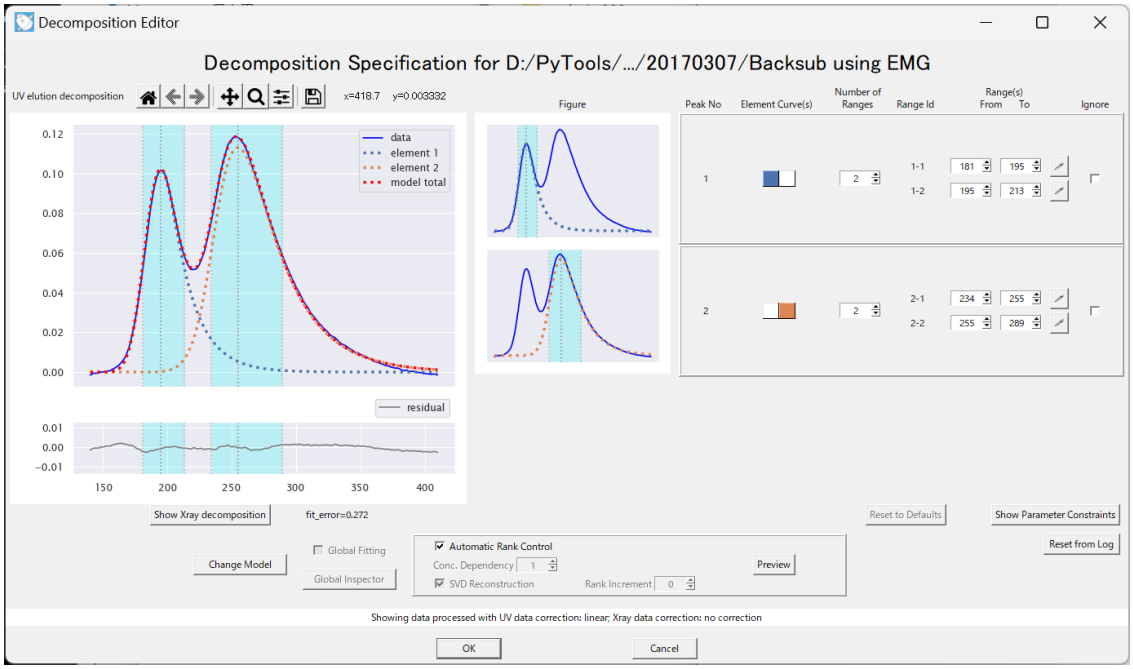


Fig. 3-6 Decomposition Editor showing decomposition of UV elution curve



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 7.4 for, non-default, advanced (or manual) usage of this editor.

4 Input Data

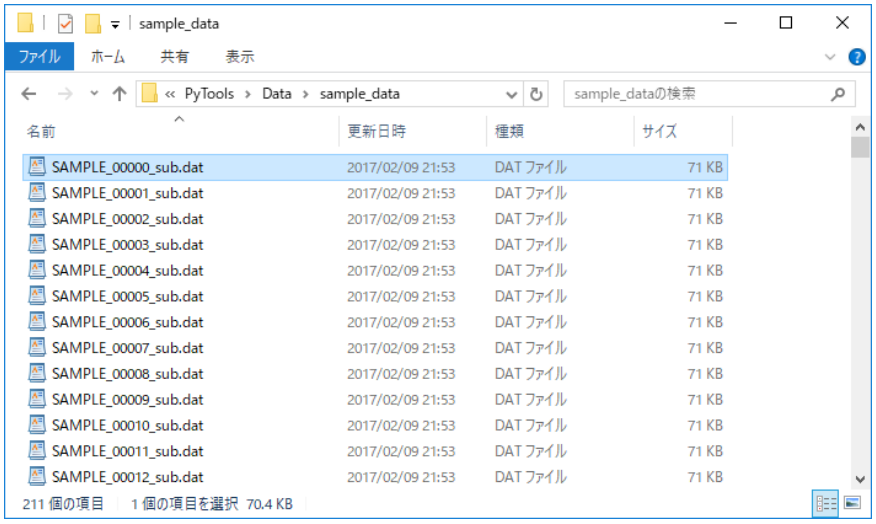
4.1 X-ray Scattering Data

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.

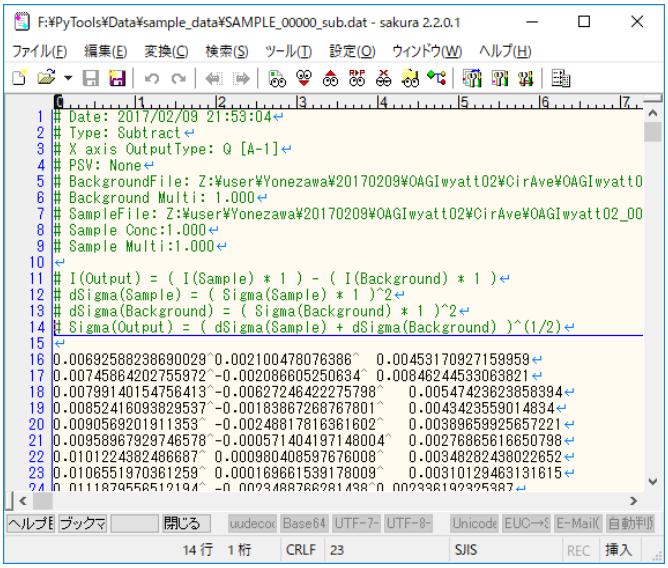
Fig. 4-1 Input folder containing X-ray scattering data files



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

a UV absorbance data are used to evaluate the concentration of each elution points.

b They should be given as text files with ".txt" extension[9], either in one file or in one folder containing a set of multiple files.

c In the one-file style (which is recommended), the file is automatically recognized by the program if the file is placed in the same folder.

d The file format in one-file style is shown in Fig. 4-3, which include comment lines followed by data lines with multiple columns.

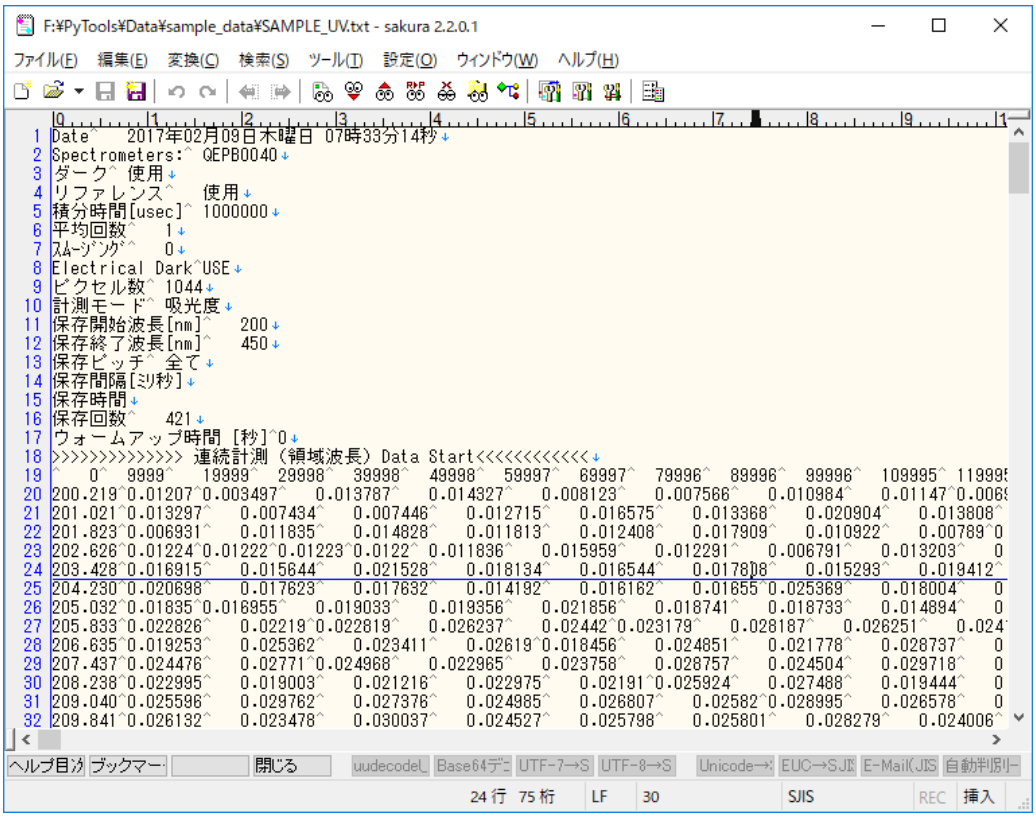
e The first line of data lines is a header line with numbered column names, which identify elution points.

f 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.

g The first column includes wavelengths, which usually vary about from 200 nm to 450 nm.

h 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)



i 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.

[9] No filename rules other than ".txt" extension are required for later versions ≧ MOLASS 1.0.8, while those like "…UV.txt" or "…spectra.txt" had been expected in earlier versions.

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)/Concentration
5. Rg

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. (See also the note below.)

g ATSAS AUTORG quality evaluation, if available, is contrasted with a red line.

h Other analysis results (namely, I(0), I(0)/Conc., Rg) 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 ATSAS AUTORG are shown in a line with orange markers and the values have been set to zeros there if the program does not produce results.

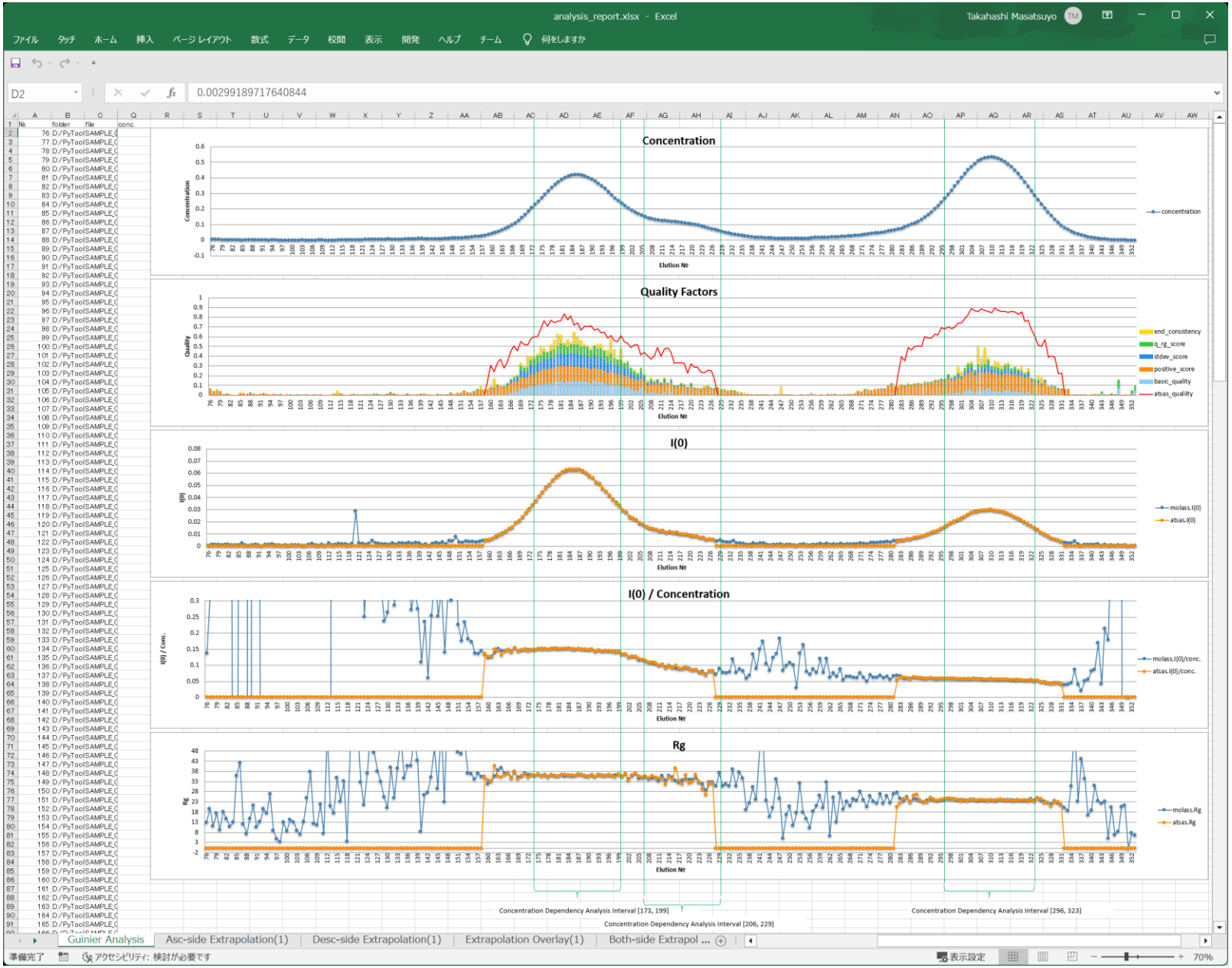
Note on the quality of Guinier Analysis and the qRg < 1.3 convention

The program strictly keeps the well-established qRg < 1.3 restriction for well-conditioned data. However, when the "basic quality" is less than 0.5, it tries to obtain possibly better Rg values by loosening the restriction to qRg < 1.6, 1.8 or 2.3 depending on the worsening of the quality. Users should be aware of this violation and be warned not to use the resulted values separately ignoring the attached qualities.

Tab. 5-1 Data Quality Factors

No	Factor Name	Value Range	Weight	Formula	Description
1	basic quality	0 – 1	0.2	$\exp\left(0.5 \times (1 - \frac{\sum abs(gradients)}{height})\right)$	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 qRg score × basic quality where raw qRg score = 1 if $q_{left} \cdot R_g \leqq 0.2$ 0 if $q_{left} \cdot R_g \geqq 1.2$ $(1.2 - q_{left} \cdot R_g)$ 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 Extrapolation (LRF) Result

a Although it is called "Extrapolation" to infinite dilution from historical reasons, the computation is implemented as a kind of LRF - Low Rank Factorization - explained briefly below.

b Result is provided in several sheets for each (ascending or descending) side of the peaks.

c See Fig. 5-2 and Fig. 5-3 for examples of those sheets.

d 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.

e 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.

f The process of extrapolation to infinite dilution is summarized below. See also Section 7.6.3 for the preview of this result.

1. Suppose you have a measured data set of X-ray intensity and a corresponding elution curve as follows.

$$\text{Measured data set: } M = \begin{bmatrix} \mathrm{M}_{1,1} & \mathrm{M}_{1,2} & \cdots & \mathrm{M}_{1,300} \\ \mathrm{M}_{2,1} & \mathrm{M}_{2,2} & \cdots & \mathrm{M}_{2,300} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{M}_{800,1} & \mathrm{M}_{800,2} & \cdots & \mathrm{M}_{800,300} \end{bmatrix} \tag{5-1}$$

$$\text{Elution curve: } C = [\mathrm{C}_1 \quad \mathrm{C}_2 \quad \cdots \quad \mathrm{C}_{300}] \tag{5-2}$$

2. Also suppose you have selected an elution range 51-70 for the extrapolation purpose, which means that you are to solve the following equation (rank 1 formulation).

$$\begin{bmatrix} \mathrm{M}_{1,51} & \mathrm{M}_{1,52} & \cdots & \mathrm{M}_{1,70} \\ \mathrm{M}_{2,51} & \mathrm{M}_{2,52} & \cdots & \mathrm{M}_{2,70} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{M}_{800,51} & \mathrm{M}_{800,52} & \cdots & \mathrm{M}_{800,70} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{800} \end{bmatrix} [\mathrm{C}_{51} \quad \mathrm{C}_{52} \quad \cdots \quad \mathrm{C}_{70}] \equiv P \cdot C \tag{5-3}$$

3. In cases of unignorable interparticle effects, the equation can be modified as follows (rank 2 formulation) by approximating the effects.[10]

$$\begin{bmatrix} \mathrm{M}_{1,51} & \mathrm{M}_{1,52} & \cdots & \mathrm{M}_{1,70} \\ \mathrm{M}_{2,51} & \mathrm{M}_{2,52} & \cdots & \mathrm{M}_{2,70} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{M}_{800,51} & \mathrm{M}_{800,52} & \cdots & \mathrm{M}_{800,70} \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \\ \vdots & \vdots \\ A_{800} & B_{800} \end{bmatrix} \begin{bmatrix} \mathrm{C}_{51} & \mathrm{C}_{52} & \cdots & \mathrm{C}_{70} \\ {\mathrm{C}_{51}}^2 & {\mathrm{C}_{52}}^2 & \cdots & {\mathrm{C}_{50}}^2 \end{bmatrix} \equiv P \cdot C \tag{5-4}$$

4. Represent either of the above equations using matrices as follows.

$$M = P \cdot C \tag{5-5}$$

5. Or, more realistically using Frobenius norm,

$$\text{Find } P \text{ which minimizes } \|P \cdot C - M\| \tag{5-6}$$

[10] In the formulation, the quadratic relation between I(q) and concentration is based on the formula (2.13) in the classical book[2] by Feigin and Svergun, 1987, ignoring the concentration dependence of $\mathrm{P}(r, v_1)$ which is considered hard to handle properly.

6. It is well known that you can get an approximate solution of the above problem as follows using Moore-Penrose inverse.[11]

$$P = M \cdot C^{+} \tag{5-7}$$

g The following description includes both cases where interparticle effects are ignorable or unignorable. Note that B[i] or $B(Q_i)$ are not computed and do not exist in ignorable cases.

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.

$$\begin{aligned} A(Q_i) &\equiv A[i] \\ B(Q_i) &\equiv B[i] \end{aligned} \tag{5-8}$$

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, ATSAS ALMERGE produces only an A(q) curve for each side, which is shown, scaled[12] 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 For results in unignorable inter-particle effects cases, see Fig. 5-5, Fig. 5-6 and Fig. 5-7 which correspond respectively Fig. 5-2, Fig. 5-3 and Fig. 5-4 in ignorable cases.

[11] See R. Penrose's paper[1] for a general-purpose solution.

[12] The extrapolated curve from ALMERGE is not normalized by concentration. It is scaled to match our result which is concentration-normalized.

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

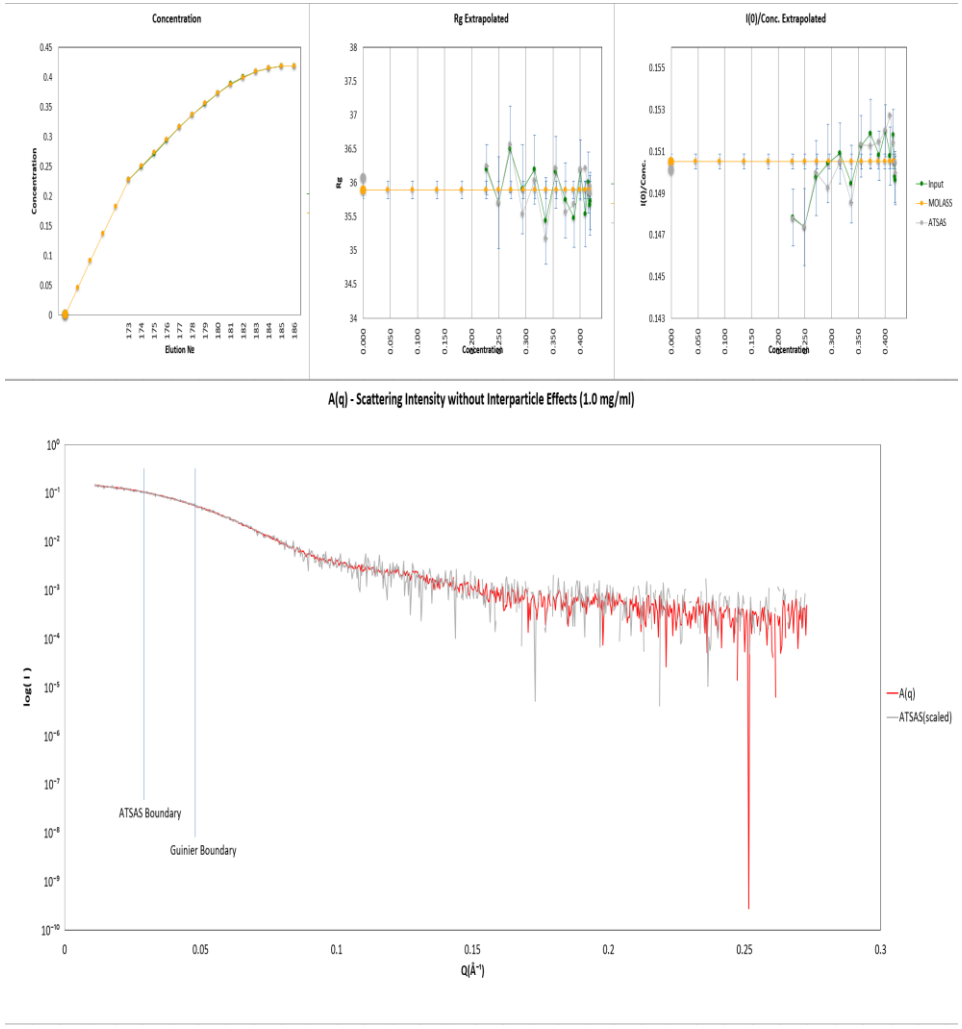


Fig. 5-3 Zero-concentration Analysis result for a descending side of a peak

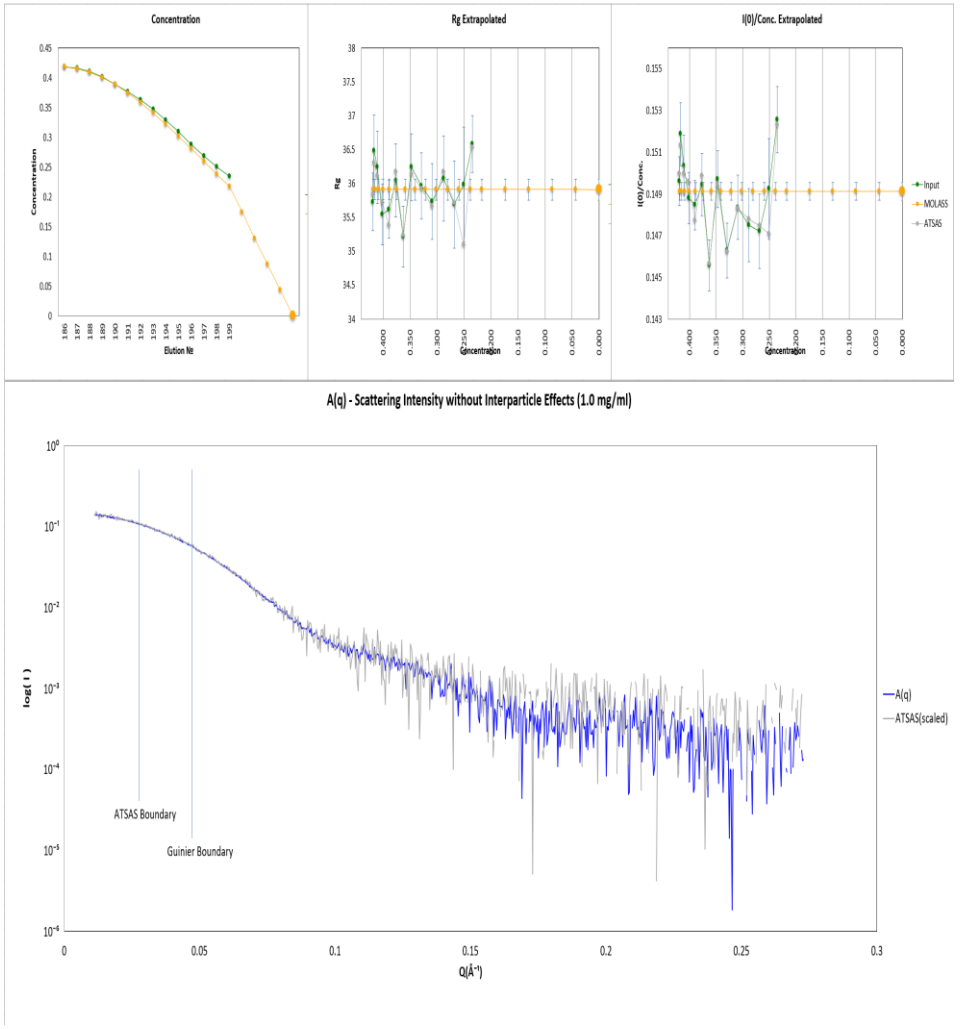


Fig. 5-4　Overlaid charts of both sides of a peak

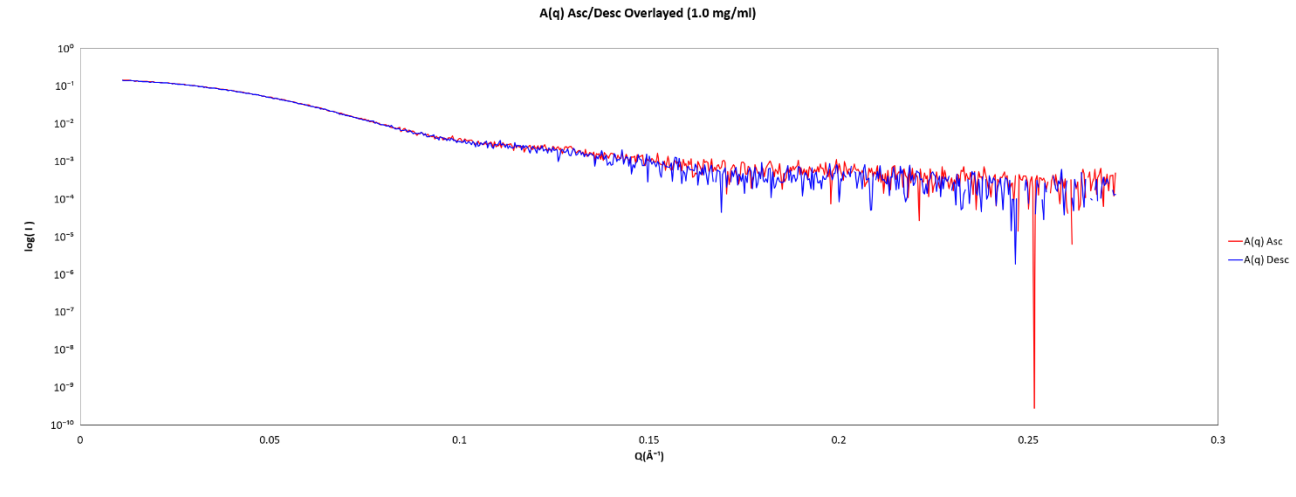


Fig. 5-5 Zero-concentration Analysis result for an ascending side of a peak

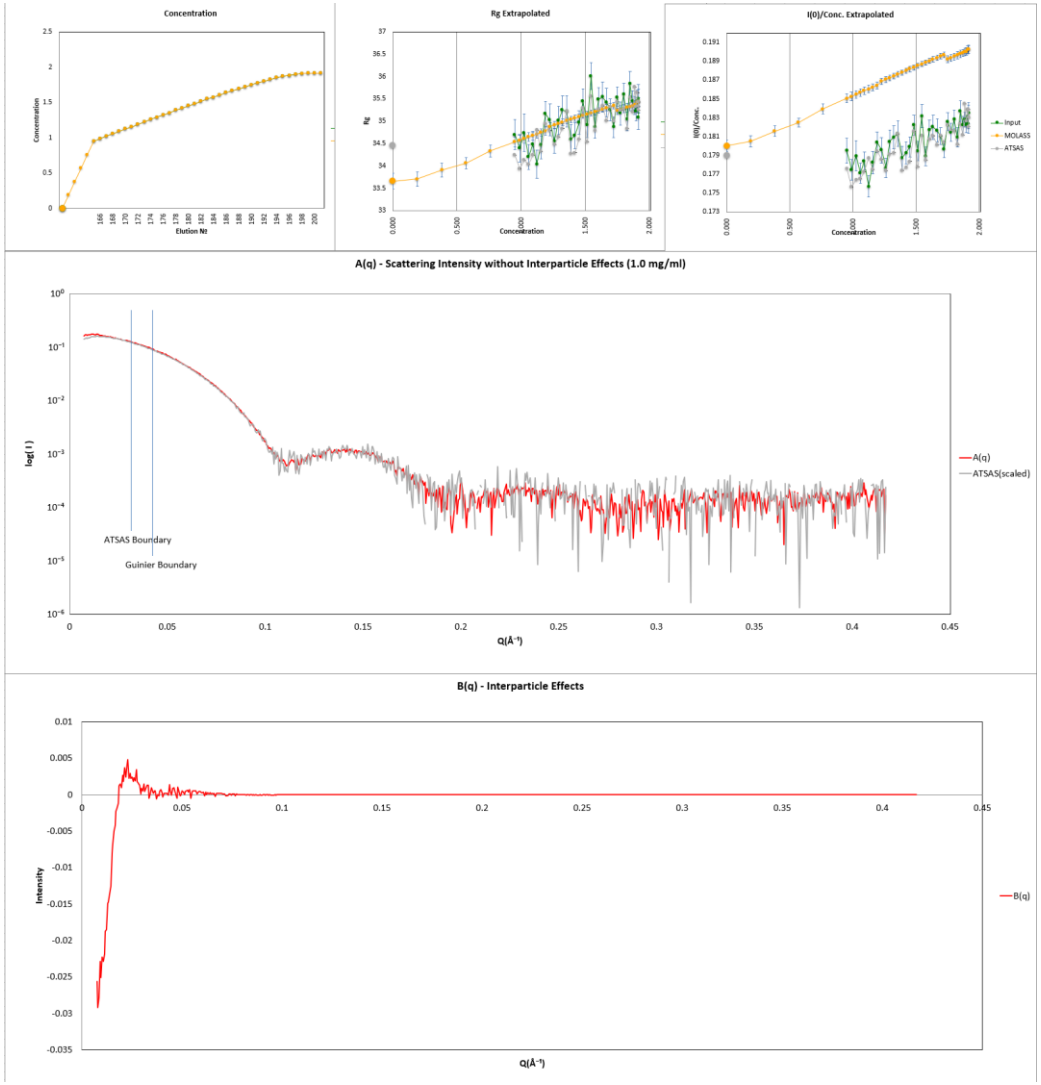


Fig. 5-6 Zero-concentration Analysis result for a descending side of a peak

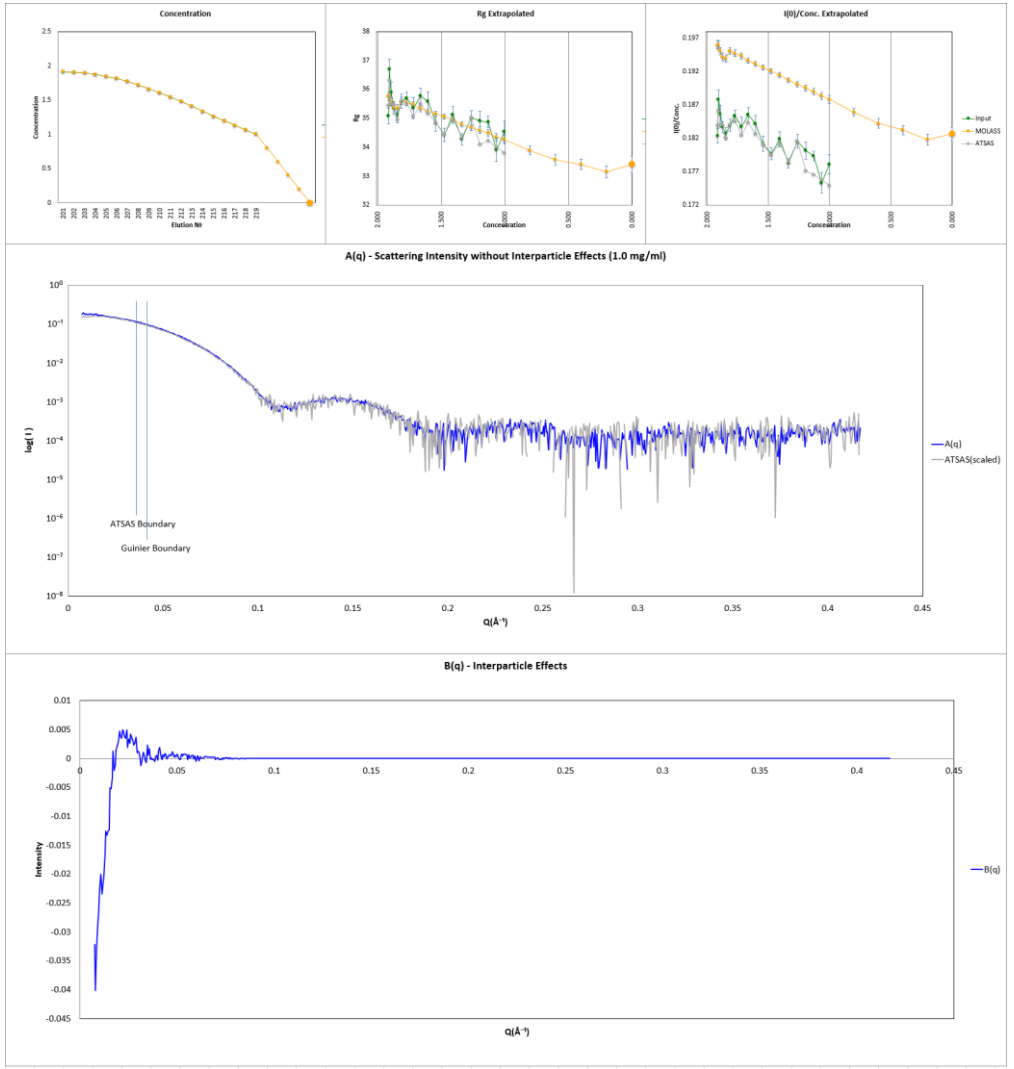
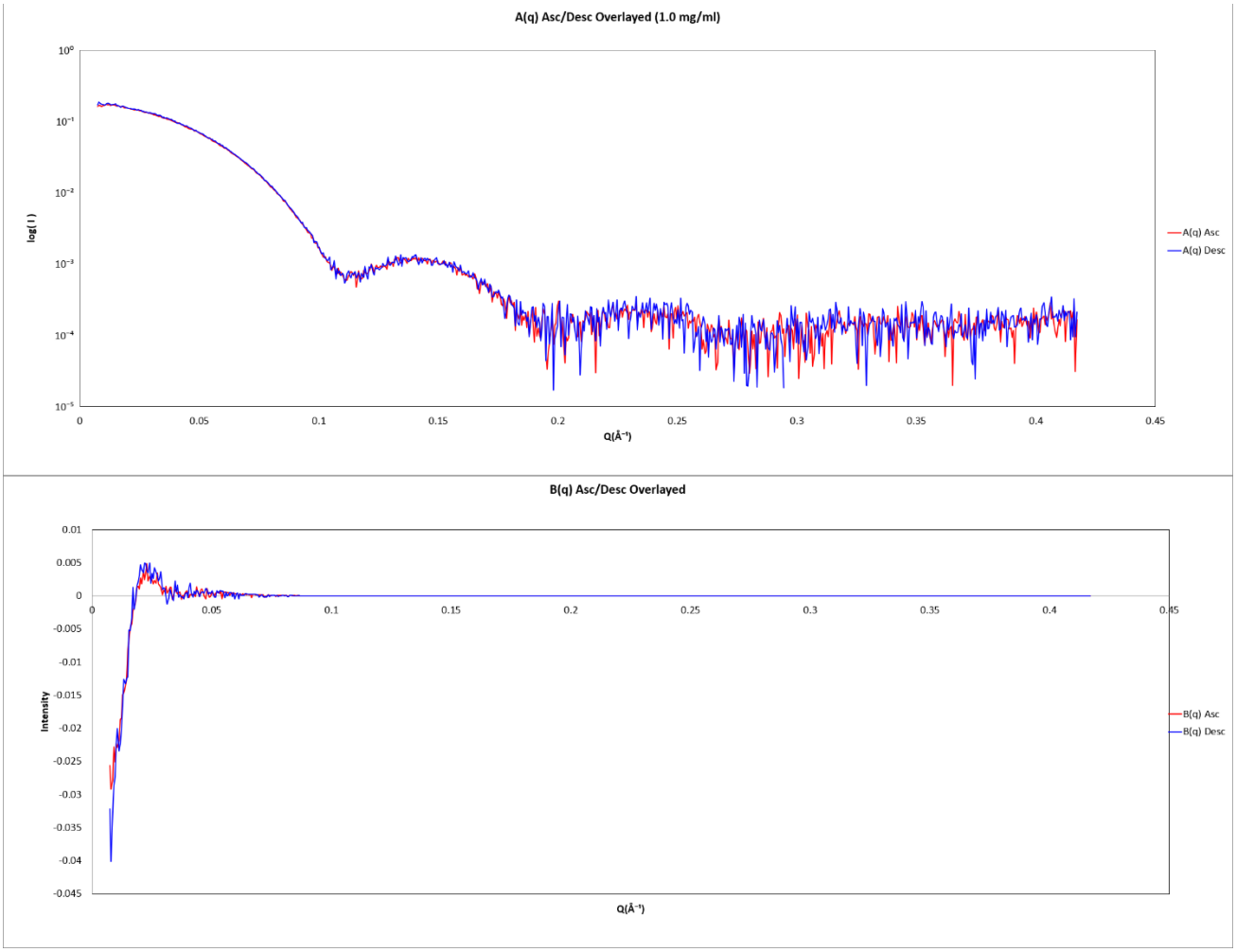


Fig. 5-7　Overlaid charts of both sides of a peak



5.1.3 Analysis Summary

a It is provided in three sheets named as follows.
   1. Extrapolation Summary (Fig. 5-8)
   2. Entire Summary (Fig. 5-9)
   3. Summary for Publication (Fig. 5-10)

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. 5-8.

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.

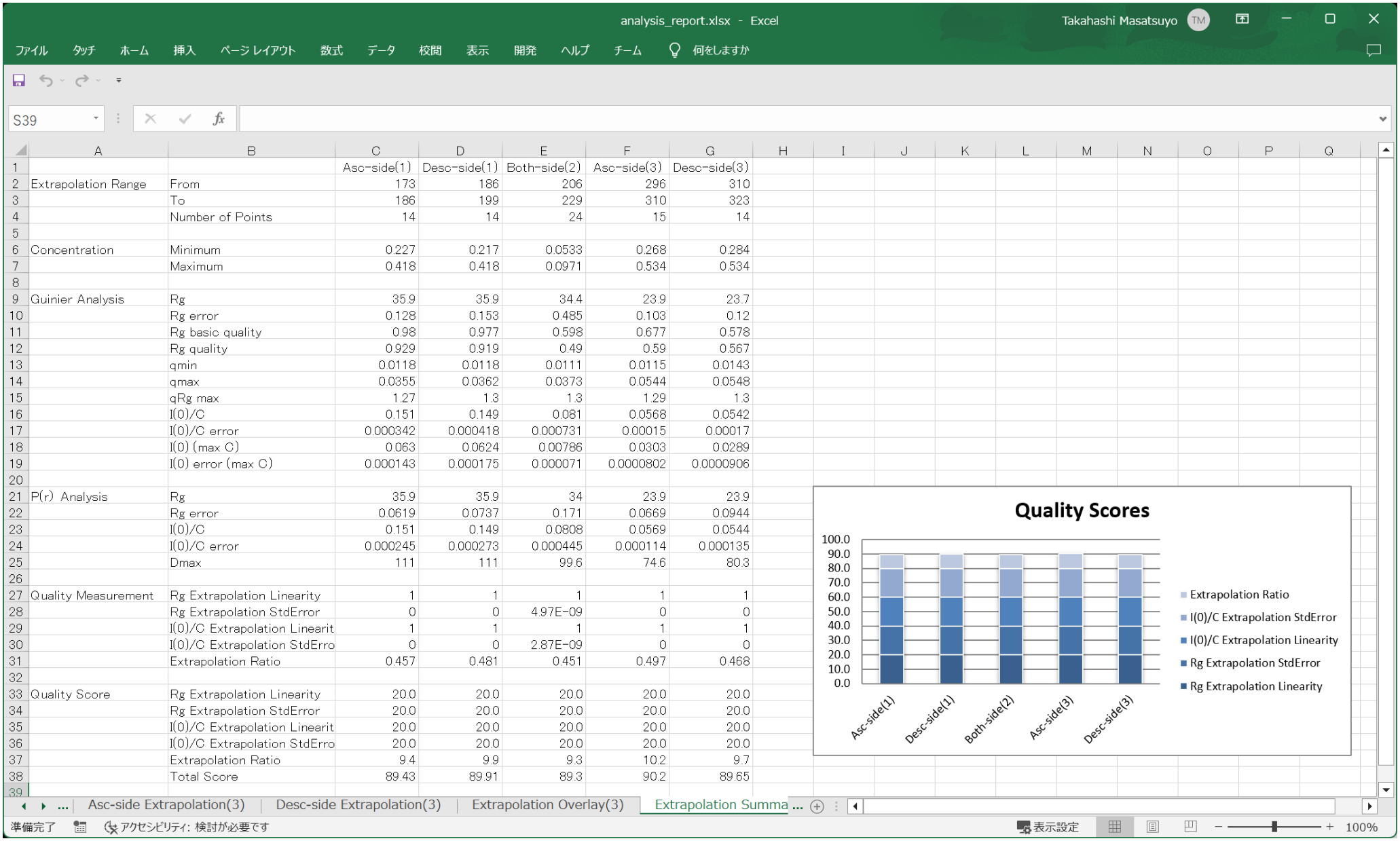
b Files postfixed as "_asc.dat-" or "_dsc.dat-" are numbered with prefixes, like "pk1_", in the order of elution for both cases with a single peak or multiple peaks.

c "_cn" postfix is added to suggest that the file contains concentration-normalized result.

Tab. 5-2 Other output files

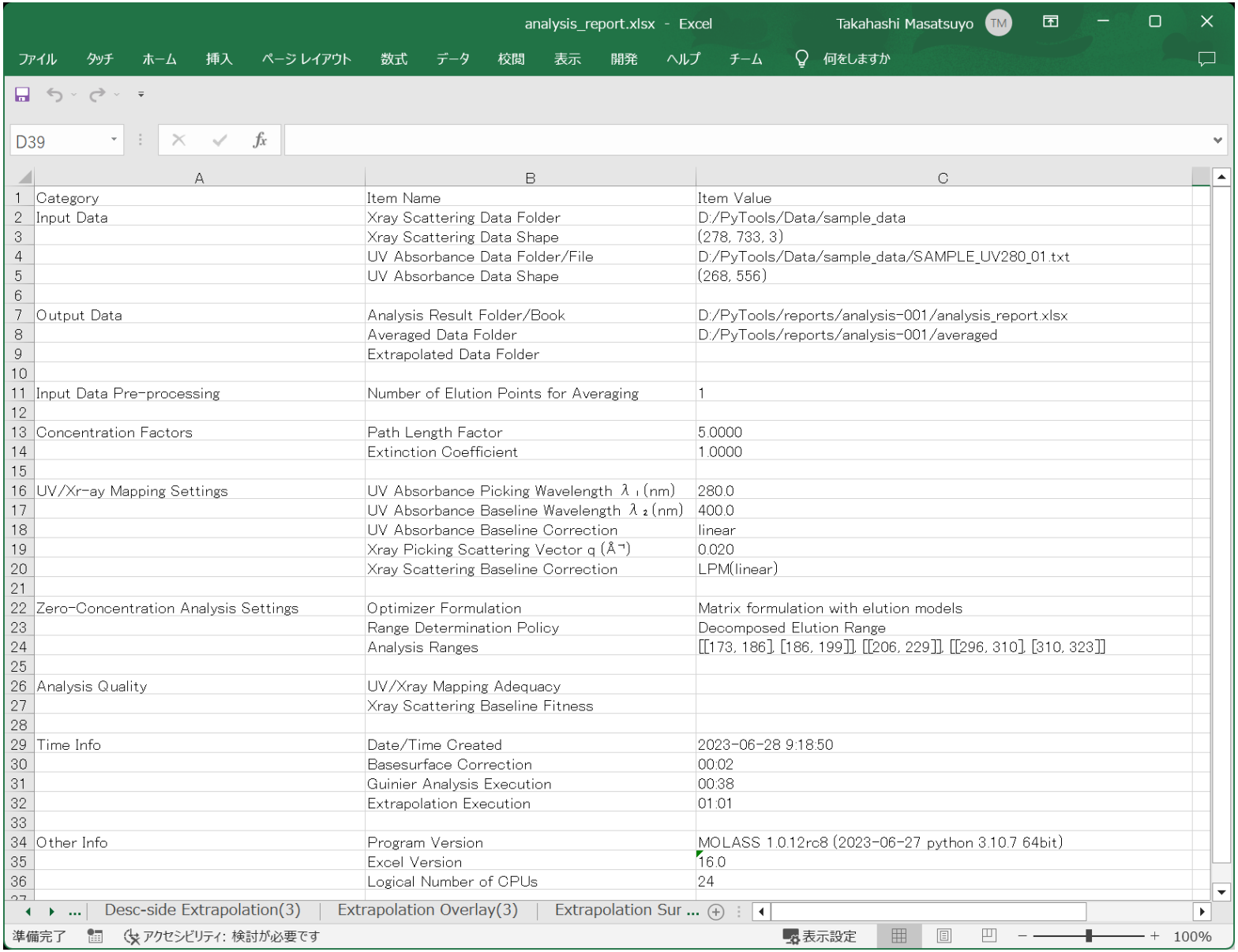
No	File name	Description
1	atsas/pk1_asc.dat	Extrapolated curve data for the ascending side by ALMERGE
2	atsas/pk1_dsc.dat	Extrapolated curve data for the descending side by ALMERGE
3	pk1_asc_A_cn.dat	Extrapolated A(q) data for the ascending side by this program
4	pk1_asc_B_cn.dat	Extrapolated B(q) data 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. 5-8　Extrapolation Summary Sheet



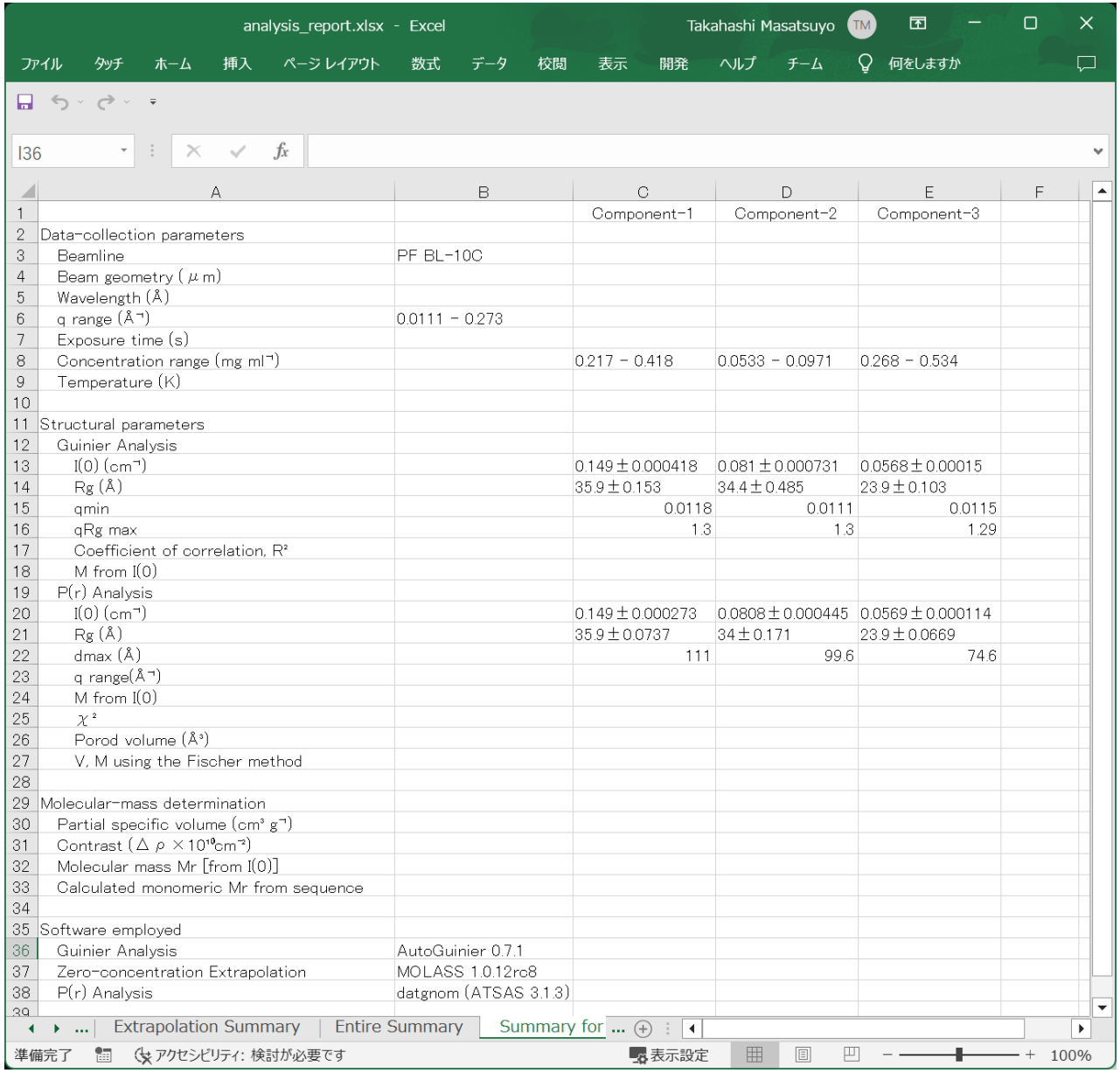
	A	B	C	D	E	F	G
1			Asc-side(1)	Desc-side(1)	Both-side(2)	Asc-side(3)	Desc-side(3)
2	Extrapolation Range	From	173	186	206	296	310
3		To	186	199	229	310	323
4		Number of Points	14	14	24	15	14
5							
6	Concentration	Minimum	0.227	0.217	0.0533	0.268	0.284
7		Maximum	0.418	0.418	0.0971	0.534	0.534
8							
9	Guinier Analysis	Rg	35.9	35.9	34.4	23.9	23.7
10		Rg error	0.128	0.153	0.485	0.103	0.12
11		Rg basic quality	0.98	0.977	0.598	0.677	0.578
12		Rg quality	0.929	0.919	0.49	0.59	0.567
13		qmin	0.0118	0.0118	0.0111	0.0115	0.0143
14		qmax	0.0355	0.0362	0.0373	0.0544	0.0548
15		qRg max	1.27	1.3	1.3	1.29	1.3
16		I(0)/C	0.151	0.149	0.081	0.0568	0.0542
17		I(0)/C error	0.000342	0.000418	0.000731	0.00015	0.00017
18		I(0) (max C)	0.063	0.0624	0.00786	0.0303	0.0289
19		I(0) error (max C)	0.000143	0.000175	0.000071	0.0000802	0.0000906
20							
21	P(r) Analysis	Rg	35.9	35.9	34	23.9	23.9
22		Rg error	0.0619	0.0737	0.171	0.0669	0.0944
23		I(0)/C	0.151	0.149	0.0808	0.0569	0.0544
24		I(0)/C error	0.000245	0.000273	0.000445	0.000114	0.000135
25		Dmax	111	111	99.6	74.6	80.3
26							
27	Quality Measurement	Rg Extrapolation Linearity	1	1	1	1	1
28		Rg Extrapolation StdError	0	0	4.97E-09	0	0
29		I(0)/C Extrapolation Linearit	1	1	1	1	1
30		I(0)/C Extrapolation StdErro	0	0	2.87E-09	0	0
31		Extrapolation Ratio	0.457	0.481	0.451	0.497	0.468
32							
33	Quality Score	Rg Extrapolation Linearity	20.0	20.0	20.0	20.0	20.0
34		Rg Extrapolation StdError	20.0	20.0	20.0	20.0	20.0
35		I(0)/C Extrapolation Linearit	20.0	20.0	20.0	20.0	20.0
36		I(0)/C Extrapolation StdErro	20.0	20.0	20.0	20.0	20.0
37		Extrapolation Ratio	9.4	9.9	9.3	10.2	9.7
38		Total Score	89.43	89.91	89.3	90.2	89.65
39							

Fig. 5-9　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	(278, 733, 3)
4		UV Absorbance Data Folder/File	D:/PyTools/Data/sample_data/SAMPLE_UV280_01.txt
5		UV Absorbance Data Shape	(268, 556)
6			
7	Output Data	Analysis Result Folder/Book	D:/PyTools/reports/analysis-001/analysis_report.xlsx
8		Averaged Data Folder	D:/PyTools/reports/analysis-001/averaged
9		Extrapolated Data Folder	
10			
11	Input Data Pre-processing	Number of Elution Points for Averaging	1
12			
13	Concentration Factors	Path Length Factor	5.0000
14		Extinction Coefficient	1.0000
15			
16	UV/Xr-ay Mapping Settings	UV Absorbance Picking Wavelength λ_1 (nm)	280.0
17		UV Absorbance Baseline Wavelength λ_2 (nm)	400.0
18		UV Absorbance Baseline Correction	linear
19		Xray Picking Scattering Vector q (Å⁻¹)	0.020
20		Xray Scattering Baseline Correction	LPM(linear)
21			
22	Zero-Concentration Analysis Settings	Optimizer Formulation	Matrix formulation with elution models
23		Range Determination Policy	Decomposed Elution Range
24		Analysis Ranges	[[173, 186], [186, 199]], [[206, 229]], [[296, 310], [310, 323]]
25			
26	Analysis Quality	UV/Xray Mapping Adequacy	
27		Xray Scattering Baseline Fitness	
28			
29	Time Info	Date/Time Created	2023-06-28 9:18:50
30		Basesurface Correction	00:02
31		Guinier Analysis Execution	00:38
32		Extrapolation Execution	01:01
33			
34	Other Info	Program Version	MOLASS 1.0.12rc8 (2023-06-27 python 3.10.7 64bit)
35		Excel Version	16.0
36		Logical Number of CPUs	24

Fig. 5-10　Summary for Publication Sheet



analysis_report.xlsx - Excel　　Takahashi Masatsuyo

ファイル　タッチ　ホーム　挿入　ページ レイアウト　数式　データ　校閲　表示　開発　ヘルプ　チーム　何をしますか

I36

	A	B	C	D	E	F
1			Component-1	Component-2	Component-3	
2	Data-collection parameters					
3	Beamline	PF BL-10C				
4	Beam geometry (μm)					
5	Wavelength (Å)					
6	q range (Å¯)	0.0111 - 0.273				
7	Exposure time (s)					
8	Concentration range (mg ml¯)		0.217 - 0.418	0.0533 - 0.0971	0.268 - 0.534	
9	Temperature (K)					
10						
11	Structural parameters					
12	Guinier Analysis					
13	I(0) (cm¯)		0.149±0.000418	0.081±0.000731	0.0568±0.00015	
14	Rg (Å)		35.9±0.153	34.4±0.485	23.9±0.103	
15	qmin		0.0118	0.0111	0.0115	
16	qRg max		1.3	1.3	1.29	
17	Coefficient of correlation, R²					
18	M from I(0)					
19	P(r) Analysis					
20	I(0) (cm¯)		0.149±0.000273	0.0808±0.000445	0.0569±0.000114	
21	Rg (Å)		35.9±0.0737	34±0.171	23.9±0.0669	
22	dmax (Å)		111	99.6	74.6	
23	q range(Å¯)					
24	M from I(0)					
25	χ^2					
26	Porod volume (Å³)					
27	V, M using the Fischer method					
28						
29	Molecular-mass determination					
30	Partial specific volume (cm³ g¯)					
31	Contrast (Δρ ×10¹⁰cm¯²)					
32	Molecular mass Mr [from I(0)]					
33	Calculated monomeric Mr from sequence					
34						
35	Software employed					
36	Guinier Analysis	AutoGuinier 0.7.1				
37	Zero-concentration Extrapolation	MOLASS 1.0.12rc8				
38	P(r) Analysis	datgnom (ATSAS 3.1.3)				
39						

Extrapolation Summary | Entire Summary | Summary for ...

準備完了　アクセシビリティ: 検討が必要です　表示設定　100%

6 Miscellaneous Notes

6.1 Manipulation Logging

a Normal manipulation log is saved to a file in the analysis result folder specified in the main dialog, named for example “…¥reports¥analysis000¥molass.log”.

b Error messages are saved either to the above mentioned log file, or to files in the log folder in the installed folder, named like “…¥molass-1_0_0¥log¥final_error.log”.

c Please attach these log files to your trouble report to get support in trouble shooting.

6.2 Persistent Memory Storage

a Some of user dependent information are stored persistently in files in a folder usually located in “C:¥Users¥username¥.KekTools¥molass”.

b This folder may safely be deleted when any of the files got damaged in troubles, which may occur, for example, when run in multiple instances mentioned below.

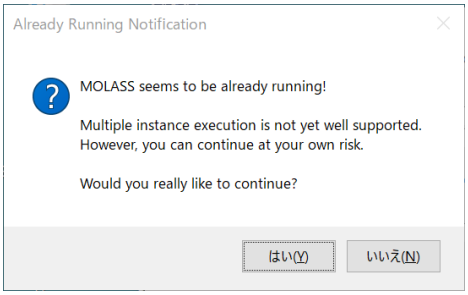
c The program is supposed to start up in default states in such abnormal cases.

6.3 Multiple Instance Execution

a It is not recommended to run multiple instances of the program, because the storage files mentioned above are only designed for single instance use.

b Therefore, the program warns you with a message shown below when you try to start up multiple instances unintendedly.

Fig. 6-1 Already Running Notification



6.4 Batch Mode Execution

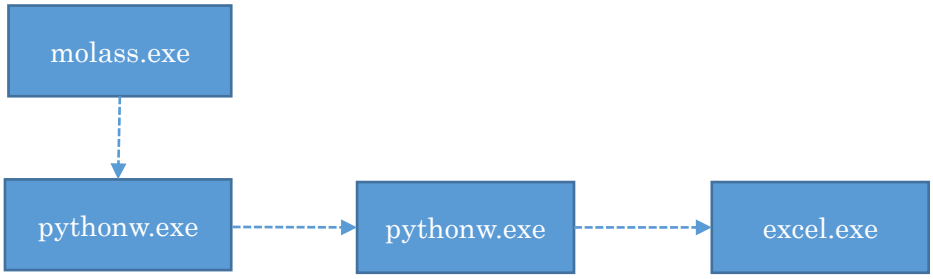
a The program is usually run in the GUI mode, which can be inconvenient in trouble shooting especially when it terminates before the first dialog appears.

b In such cases, you can double click “molass-debug.bat” file to get it started in the batch mode to get information about the situation from the command prompt window.

6.5 Multiprocessing and Cleanup

a During execution, the MOLASS program usually runs with multiple processes as shown in the following figure.

Fig. 6-2 Multiprocessing of MOLASS



b The “molass.exe” is provided just to simplify the starting procedure, which actually invokes “pythonw.exe” with the main script “molass.py” located in the “build” folder.

c The first invoked “pythonw.exe” main process performs the apparent actions, while the secondary “pythonw.exe” process executes Excel book handling in parallel in order to make faster the entire execution.

d Those child processes are coded to try to terminate themselves automatically when the parent “molass.exe” has been terminated, even in such abnormal cases when you terminate one using Task Manager in some due situation.

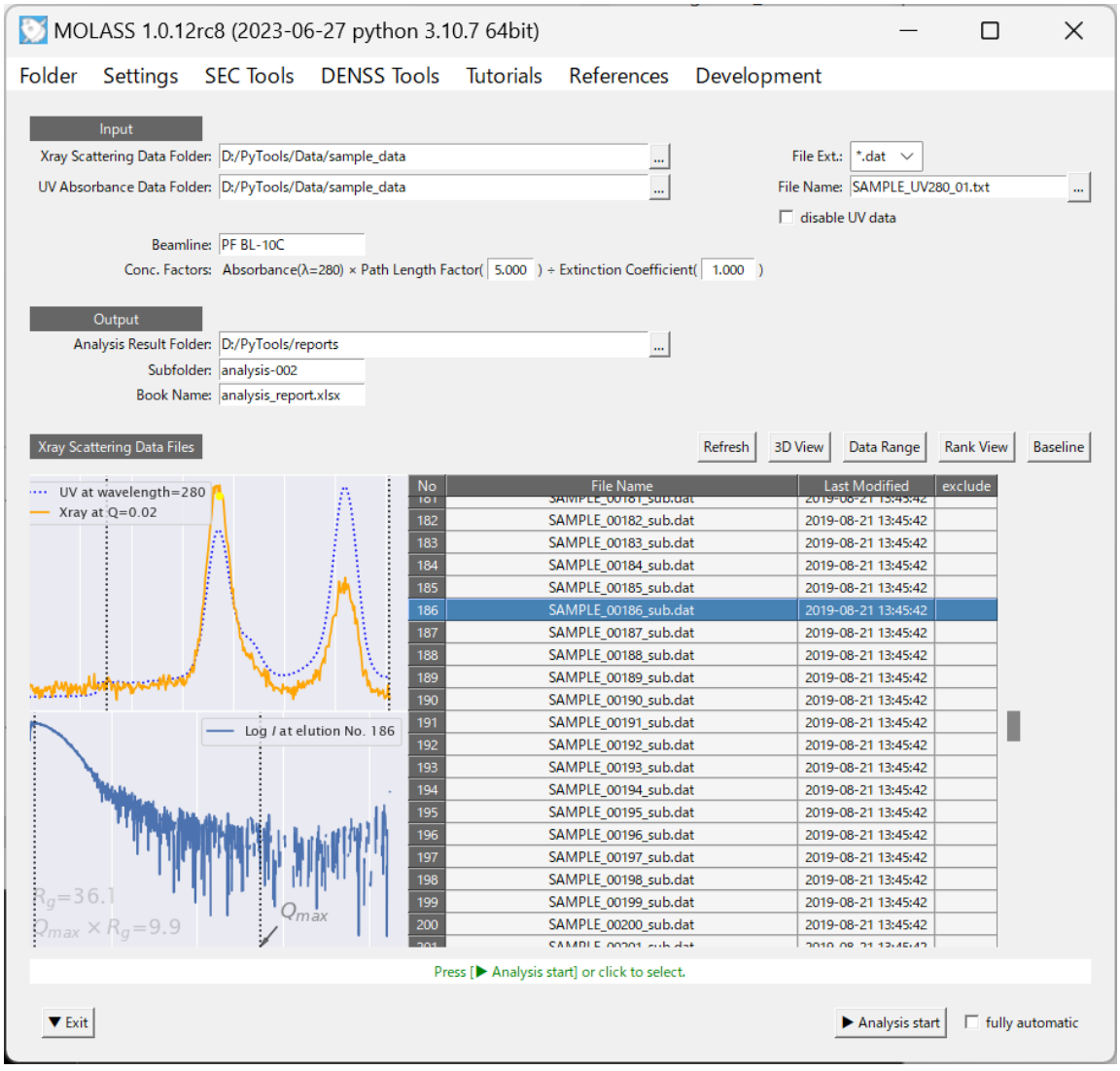
7 Detailed Description of the Dialogs

7.1 Main Dialog

7.1.1 Purpose and Required Inputs

a The Main Dialog is for entries of input / output folders and files as shown in Fig. 7-1.

Fig. 7-1 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 7.1.7 below.

g 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. 7-1 Default Values for Concentration Factors

Factor Name	if measured before 2019-9-30	if measured after 2019-10-01		
		PF	SPring-8	Others
Path Length Factor	5.0	8.437	8.155	7.071 (= $10/\sqrt{2}$)
Extinction Coefficient, Absorbance (OD) @ 1mg/ml	1.0			

h These factors are not essentially relevant to the computation of Rg and I(0) or extrapolation to zero-concentration, however, they do affect the values in the results on the summary sheets mentioned in Section 0. Therefore, don't forget to change them if they are different from default values.

i Beamline name is identified from "Spectrometers" field in the UV data file.

j The output folders are to be specified hierarchically in "Analysis Result Folder" and "Subfolder" entries.

k The "Subfolder" entry must include numbers for automatic numbering, as shown in the figure. The number will be incremented if there already exists a file with the same name.

l The "▶ Analysis start" button invokes a preliminary process and shows the "Mapping Dialog" described in the next section.

m 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.

7.1.2 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 input data.

b The upper figure shows the elution curve with a yellow selected point, and the lower figure shows the scattering curve with its estimated Rg value which corresponds to the selected point in the elution curve.

c The row selection in file list will be synchronize 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 7.10 for such usage.

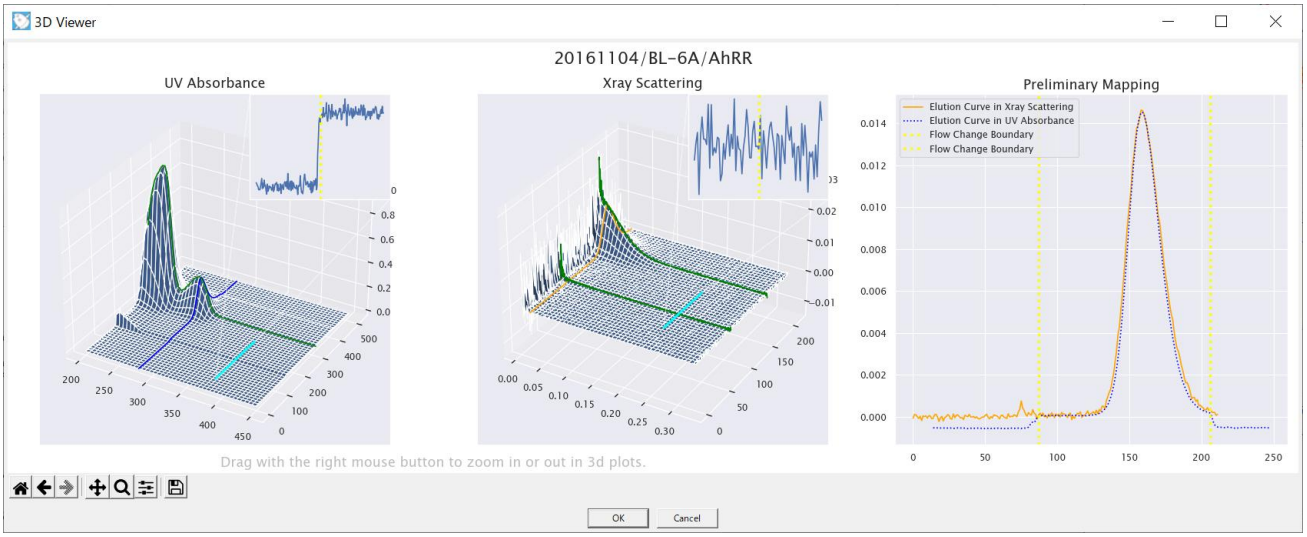
7.1.3 3D View

a This view, shown from “3D 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 mapping process of both sets of data.

Fig. 7-2 3D View



7.1.4 Data Range Trimming

a In the outline figures mentioned above (Fig. 7-1), there may appear two vertical dotted lines for each figure, which indicate data range limits suitable for analysis.

b The elution range is determined, by default, from the flow change points when they are observed.

c The small-angle limit is determined, by default, to the extendable limit beyond the Guinier region start. The limit can be controlled using “Rg-consistency” in the Settings Dialog. See section 7.7.4 for details.

d The wide-angle limit is determined, by default, to the maximal position which can 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 pressing the “Data Range” button, which shows the Data Range dialogs shown below. (Fig. 7-3, Fig. 7-4)

f These paired dialogs take care of range trimming 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.

g To change the range, 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 mapping.

Fig. 7-3 Data Range Dialog for X-ray Data

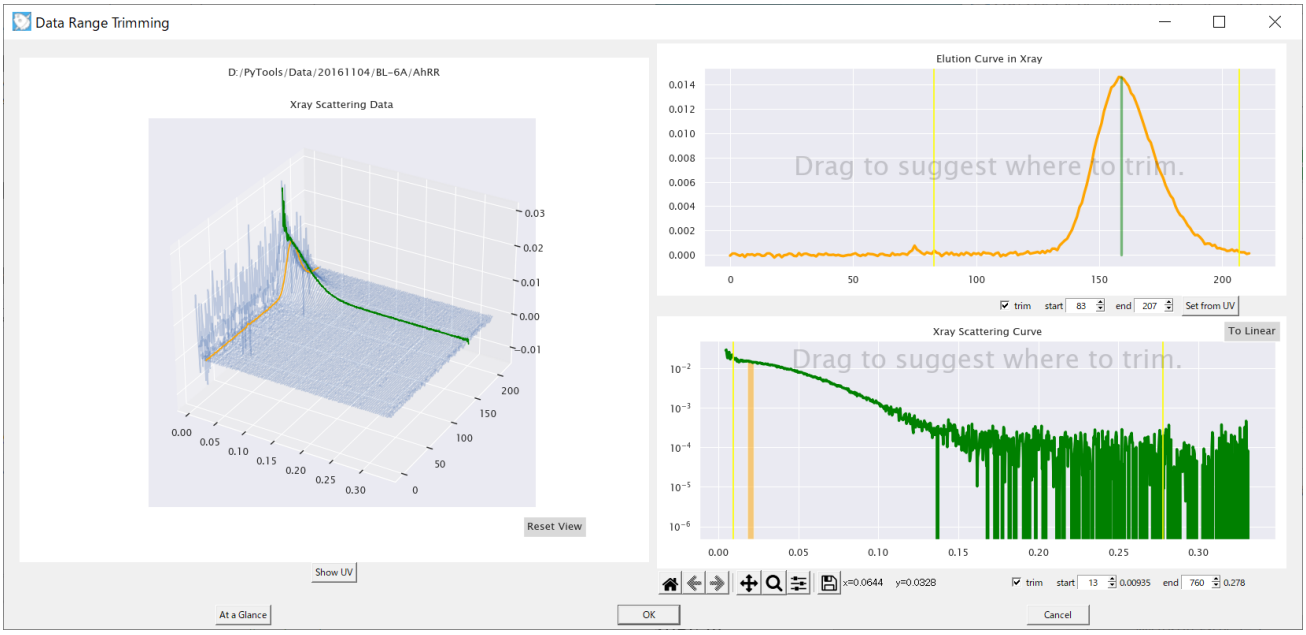
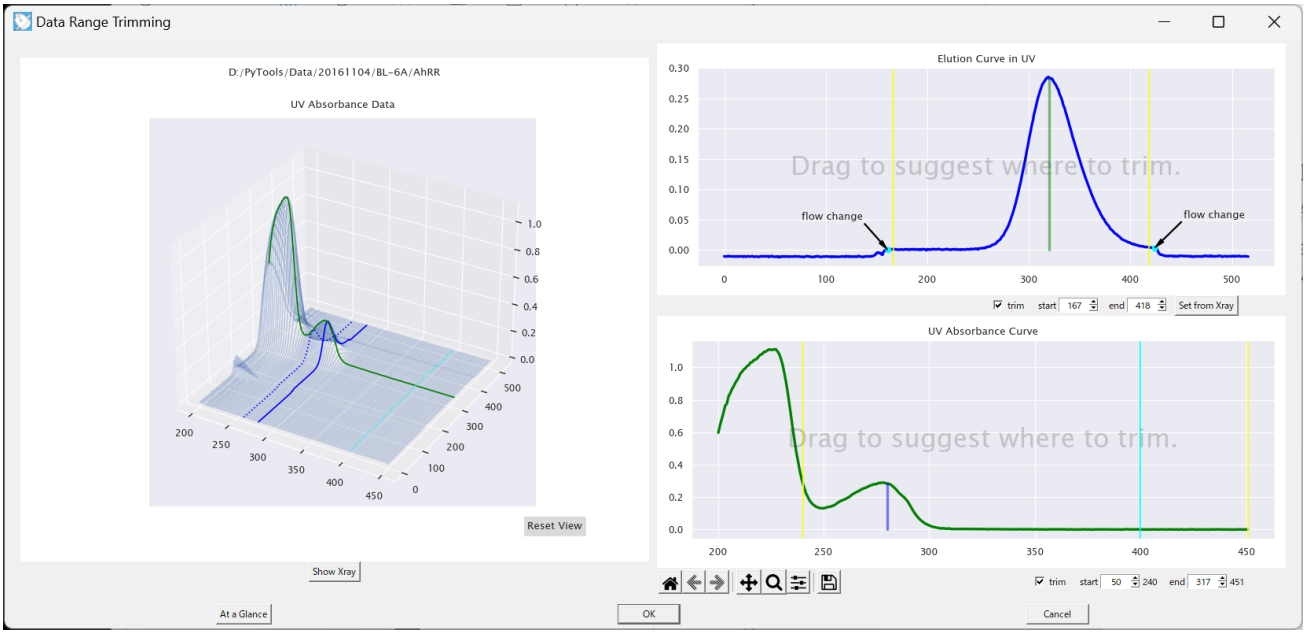


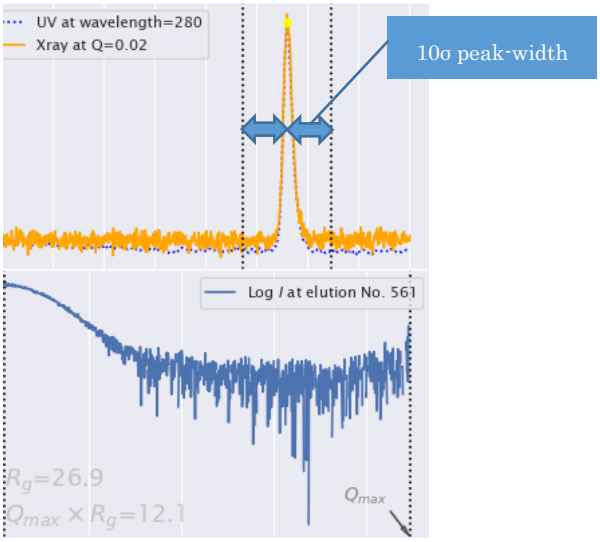
Fig. 7-4　Data Range Dialog for UV Data



i　The cyan lines suggest the wavelength referenced to determine the UV baseline.

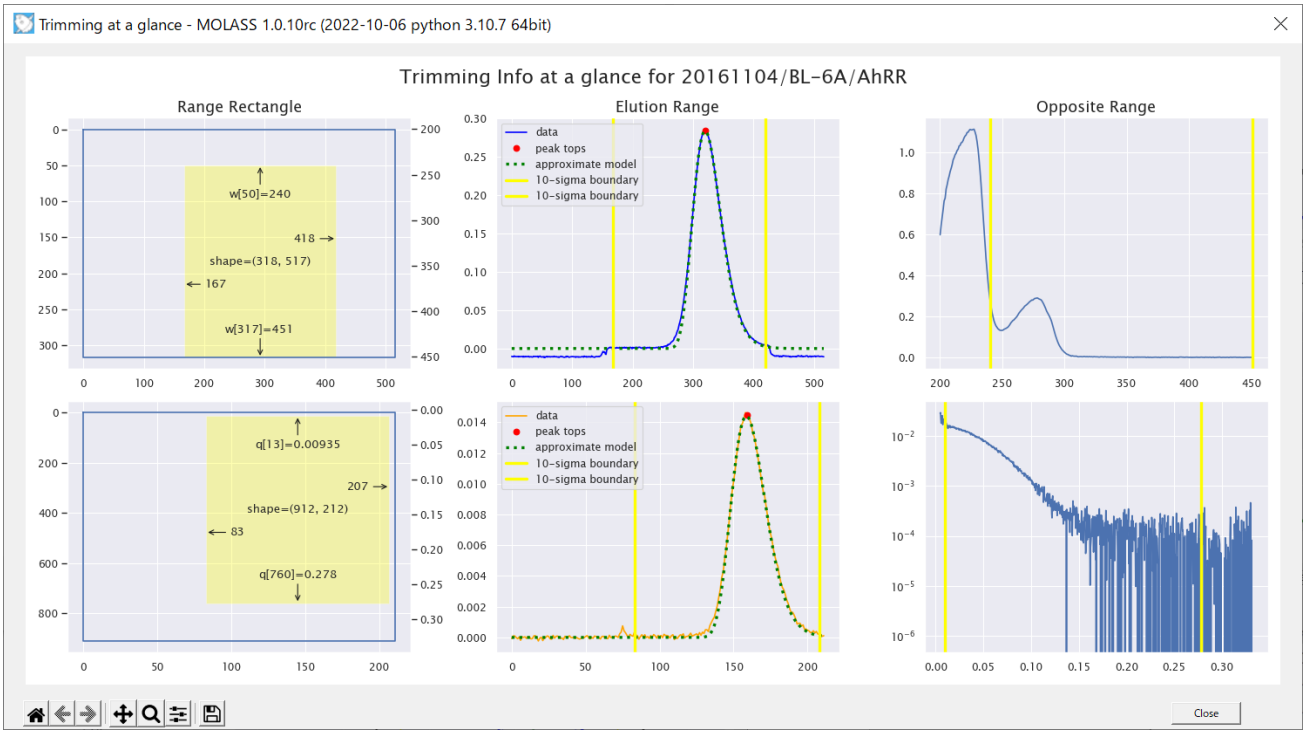
j　As an additional note for elution range trimming, non-peak part of the either side of the elution will be excluded when it extends beyond 10σ peak-width, where "σ" here should be understood to correspond to that of a gaussian elution.

Fig. 7-5　Restriction by 10σ peak-width



k　"At a Glance" button at left bottom shows a compact view of the both data ranges as shown below.

Fig. 7-6 “At a Glance” View of Data Range



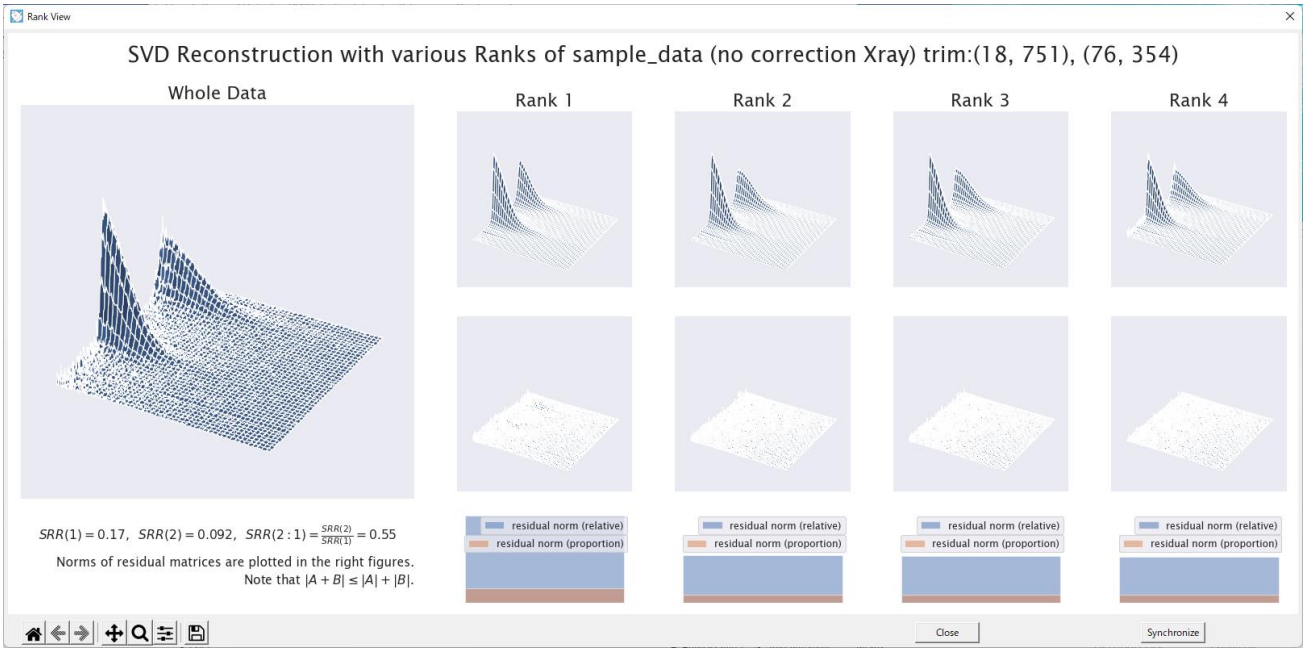
7.1.5 Rank View

a Observing ranks of data matrices is useful for the following purposes.

1. to estimate the number of components contained
2. to judge whether the concentration dependence is ignorable

b For those purposes, “Rank View” provides figures as shown below.

Fig. 7-7 “Rank View” of Xray Data

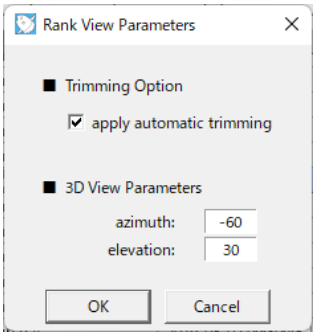


c Rank views are provided in the following variations, which can be selected in the cascade menu at the button.

- UV (no correction)
- Xray (no correction)
- Xray (corrected)

d To get a suitable view, change parameters shown below from the corresponding menu.

Fig. 7-8 Changing View Parameters of “Rank View”



7.1.6 Baseline Correction Inspection

a “Baseline” button shows the following dialog, which is intended to provide information to select the baseline correction method among “None”, “Linear” or “Integral”.

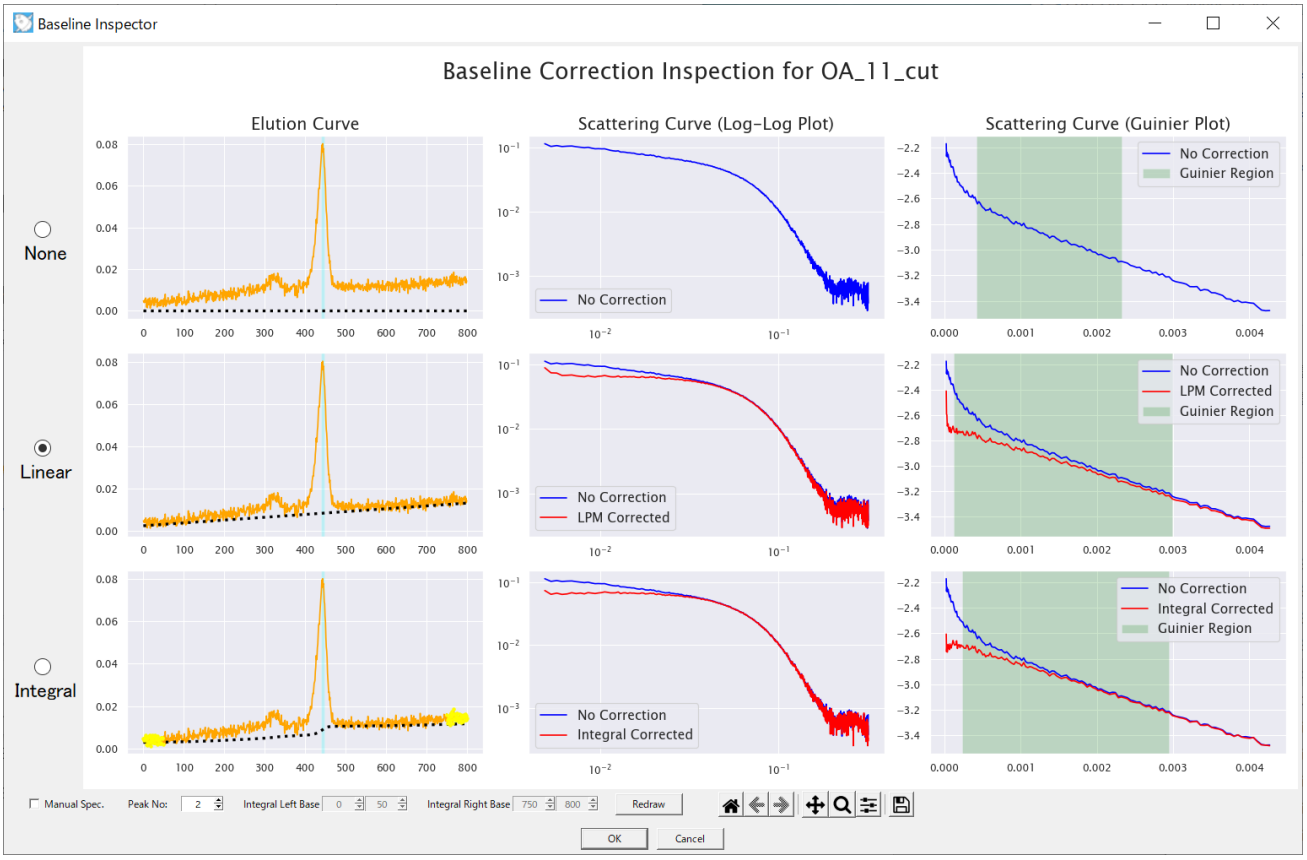
b The top part gives information on no correction case while the other parts give those on “Linear” or “Integral” correction. See Sections 7.2.9, 7.2.10 and 7.2.11 for these meanings and details.

c If you change the default settings in this dialog, automatic method selection will be disabled and those changes will be adopted.

d The ranges colored yellow in the lower left figure, which will be used subtract to get the scattering curve of the accumulated “broken samples”, can be changed either with the spin boxes or dragging the figure with your mouse.

e You can restore the automatic selection mode by unchecking the “Manual Spec.” button at the lower left of the dialog.

Fig. 7-9 Baseline Correction Inspection Dialog



7.1.7 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.

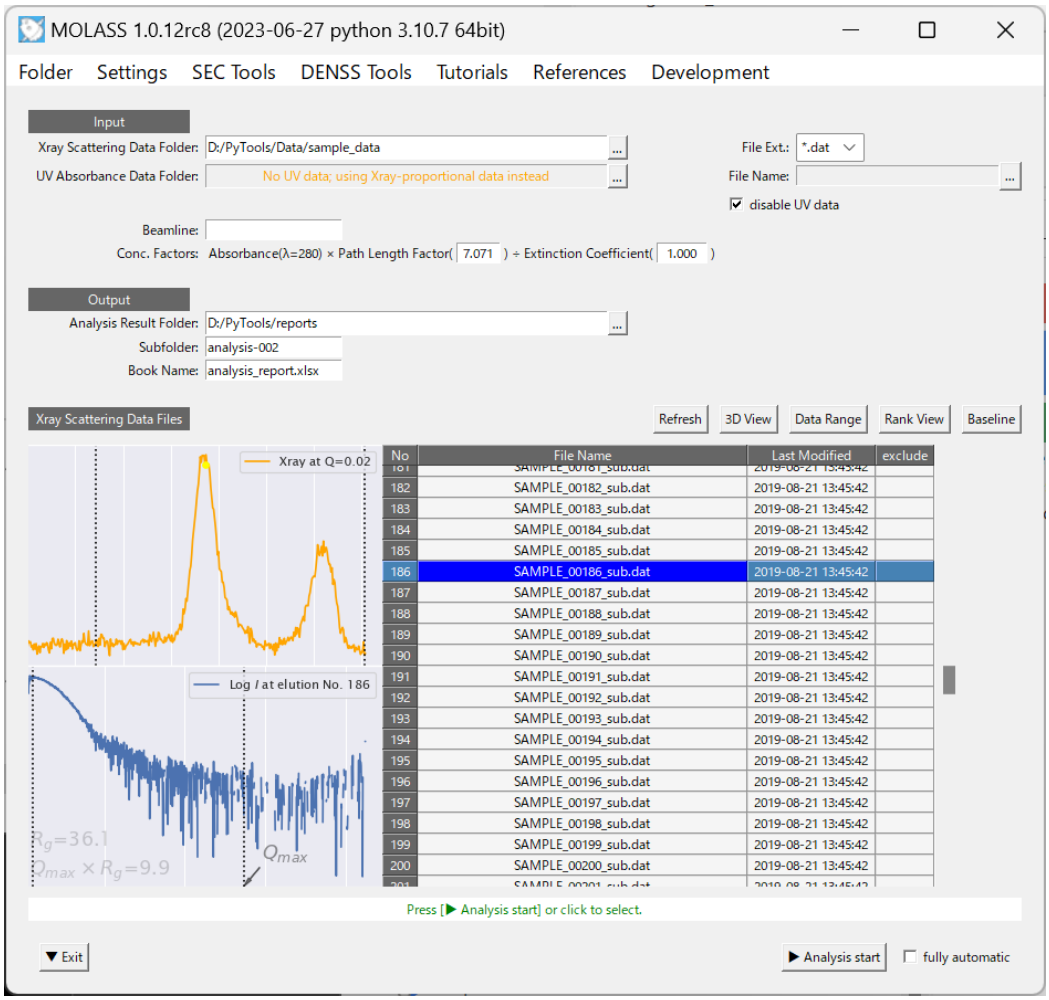
c However, there can be cases where the program fails to establish a precise mapping between those elution curves.

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. 7-10 below.

e It is toggled between the two modes by checking and unchecking the button.

f See Chapter 8 for details of the X-ray data only mode.

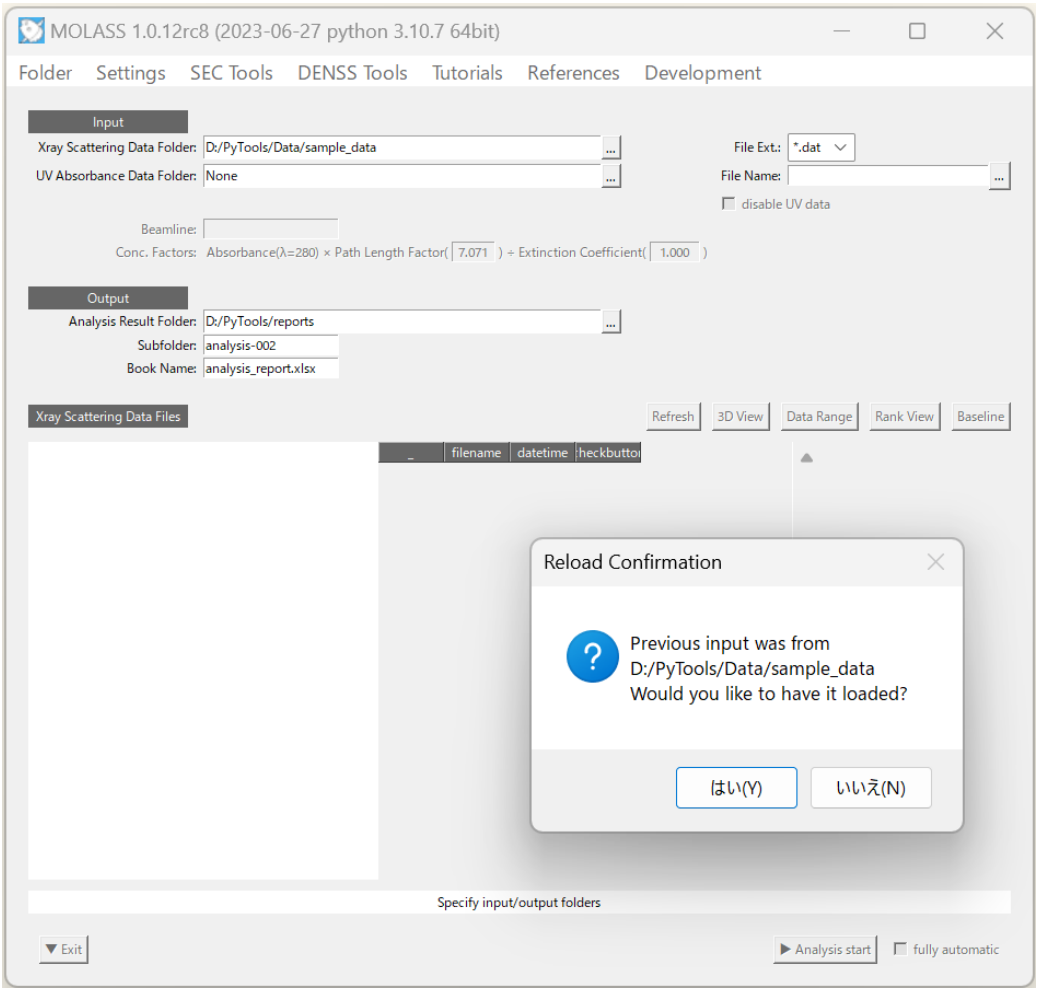
Fig. 7-10 Main Dialog (X-ray data only mode)



7.1.8 Reloading of the Previous Data

a Previously used input folder entry is restored at the startup from the user-dependent persistent storage (see 6.2), and the program asks if you want it to be loaded again in a message box shown below.

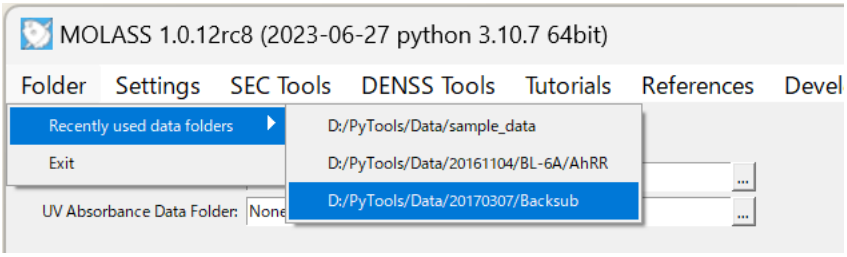
Fig. 7-11 Reloading Confirmation Message



b If that is not what you want, replying “No” will clear the input folder entry to allow you immediately enter your desired new folder path.

c Reloading one of recently used folders is easy if you use the “Recently used folders” submenu cascaded from the “Folder” menu.

Fig. 7-12 Recently used folders Menu



7.1.9 Removed Abnormal Files

a There rarely can be cases where the data include a few abnormal measurements, which happen, for example, from accidentally included bubbles. Such files are automatically removed and replaced with properly interpolated files and indicated with an "X" in the "exclude" column.[13]

[13] In earlier versions of MOLASS, this "exclude" column was also used to enable the user to indicate such abnormal files, which has been abolished for simplicity in later versions.

7.2 Mapping Dialog

7.2.1 Purpose and Manipulation Flow

a This dialog, as shown in Fig. 7-13, is for checking and adjusting the mapping between the UV absorbance data and the X-ray scattering data.

b It is invoked either by the "▶ Analysis start" button in Main Dialog.

c 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 λ=280nm for UV absorbance and scattering vector length q=0.02 $Å^{-1}$ for X-ray scattering.

Fig. 7-13 Mapping Dialog when options are hidden

d 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.

e It is shown otherwise or by clicking the "Show Options" toggle button. (See Fig. 7-14)

Fig. 7-14 Mapping Dialog when options are shown

f For usage of the toolbar (🏠|←|→|✥|🔍|⚙|💾), please refer to the following page in matplotlib document.

Interactive navigation https://matplotlib.org/users/navigation_toolbar.html.

7.2.2 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. 7-15, Fig. 7-16 on the next page)

$$SCI = \frac{\text{well fitting range length}}{\text{whole analysis range length}} \times 100$$

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. 7-2 below.

Tab. 7-2 Colored indication of mapping adequacy

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

Fig. 7-15 Another Mapping Example where the Deviation is significant between the mapped Curves

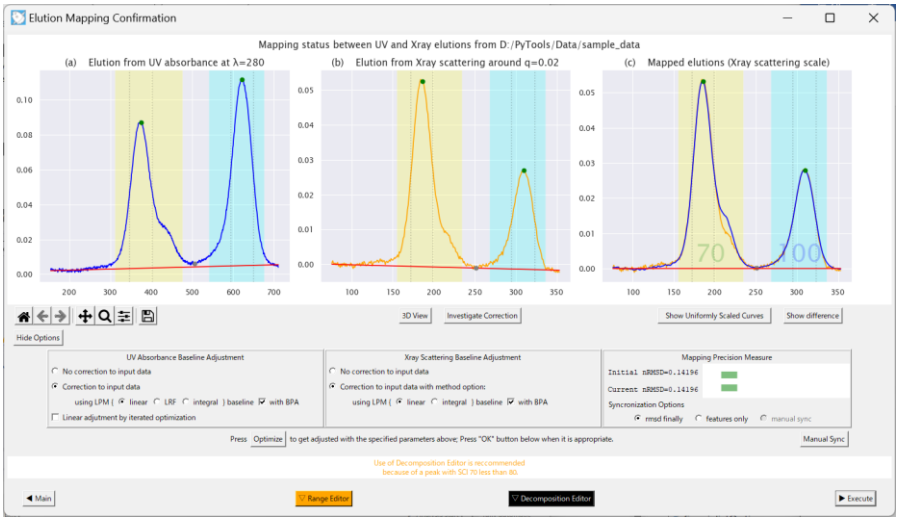


Fig. 7-16 Yet Another Mapping Example where it has overlapping Peaks


Elution Mapping Confirmation
Mapping status between UV and Xray elutions from D:/PyTools/Data/20170307/Backsub
(a) Elution from UV absorbance at λ=280
(b) Elution from Xray scattering around q=0.02
(c) Mapped elutions (Xray scattering scale)
71
97
3D View
Investigate Correction
Show Uniformly Scaled Curves
Show difference
Hide Options
UV Absorbance Baseline Adjustment
No correction to input data
Correction to input data
using LPM (linear LRF integral) baseline with BPA
Linear adjustment by iterated optimization
Xray Scattering Baseline Adjustment
No correction to input data
Correction to input data with method option:
using LPM (linear integral) baseline with BPA
Mapping Precision Measure
Initial nRMSD=0.1710$
Current nRMSD=0.1710$
Syncronization Options
rmsd finally
features only
manual sync
Press Optimize to get adjusted with the specified parameters above; Press "OK" button below when it is appropriate.
Manual Sync
Use of Decomposition Editor is reccommended
because of a peak with SCI 71 less than 80 and an overlapping interval of length 28.3.
◀ Main
▽ Range Editor
▽ Decomposition Editor
▶ Execute

7.2.3 Data Investigation

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

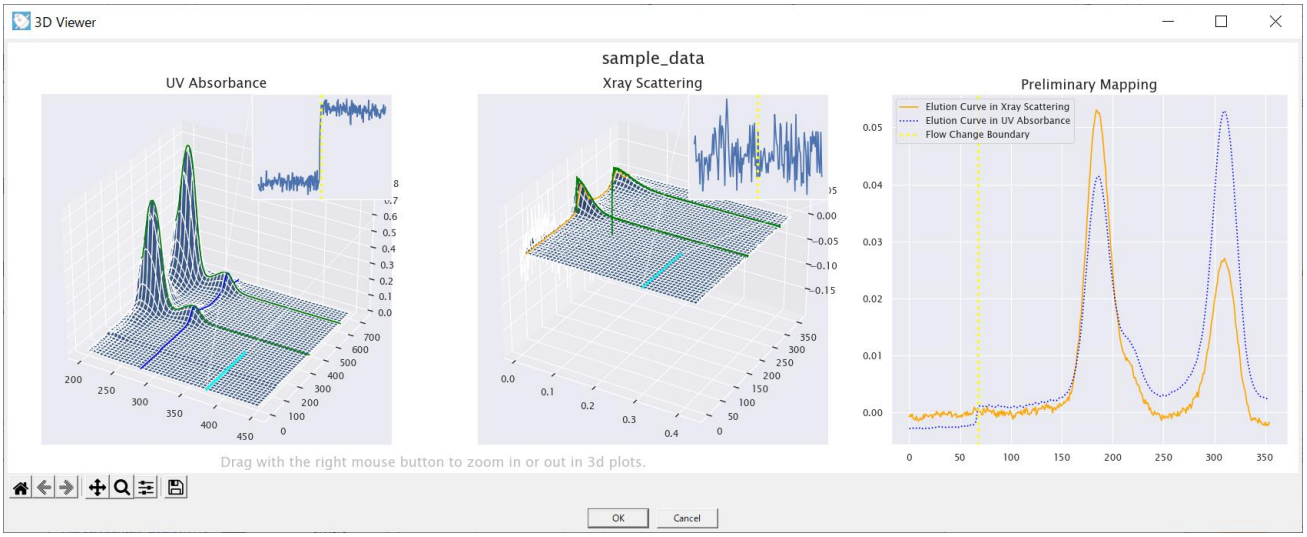
Tab. 7-3 Buttons for data investigation and range selection (from left to right)

Name	Description	Figure
3D View	Shows both sets of data in 3D plot	Fig. **7-17**
Investigate Correction	Shows X-ray scattering data at each Q	Fig. 7-31
Show uniformly scaled curves	Shows a uniformly scaled UV curve overlaid	Fig. 7-18
Show difference	Shows difference of the mapped elution curves	Fig. 7-19

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 "3D View" is to observe the entire sets of data before restriction. (See also Section 7.1.3)

Fig. 7-17 3D View



d As for the "Investigate Correction" button, see Section 7.2.13.

e The uniformly scaled curves, shown in Fig. 7-18, 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. 7-19)

Fig. 7-18　Mapping Dialog with Uniformly Scaled Curves

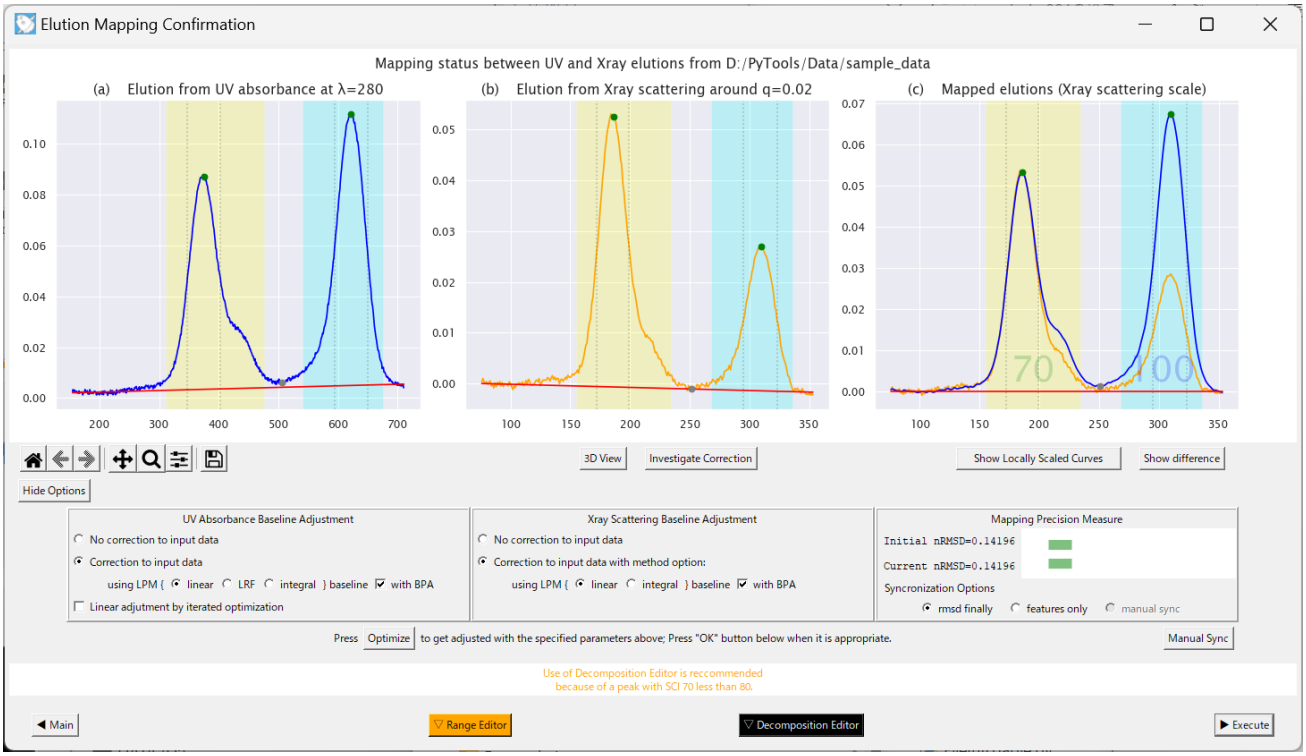
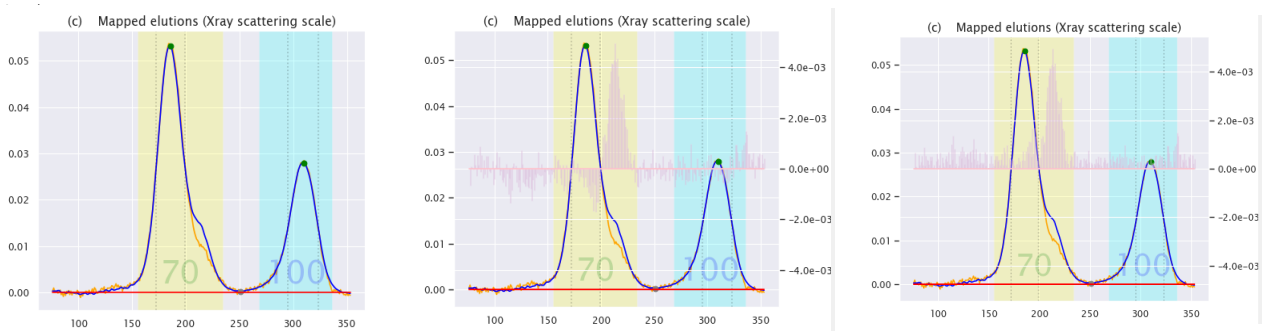


Fig. 7-19　Three states toggled by “Show difference” button



7.2.4 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. 7-4. See Fig. 7-20 for such an example.

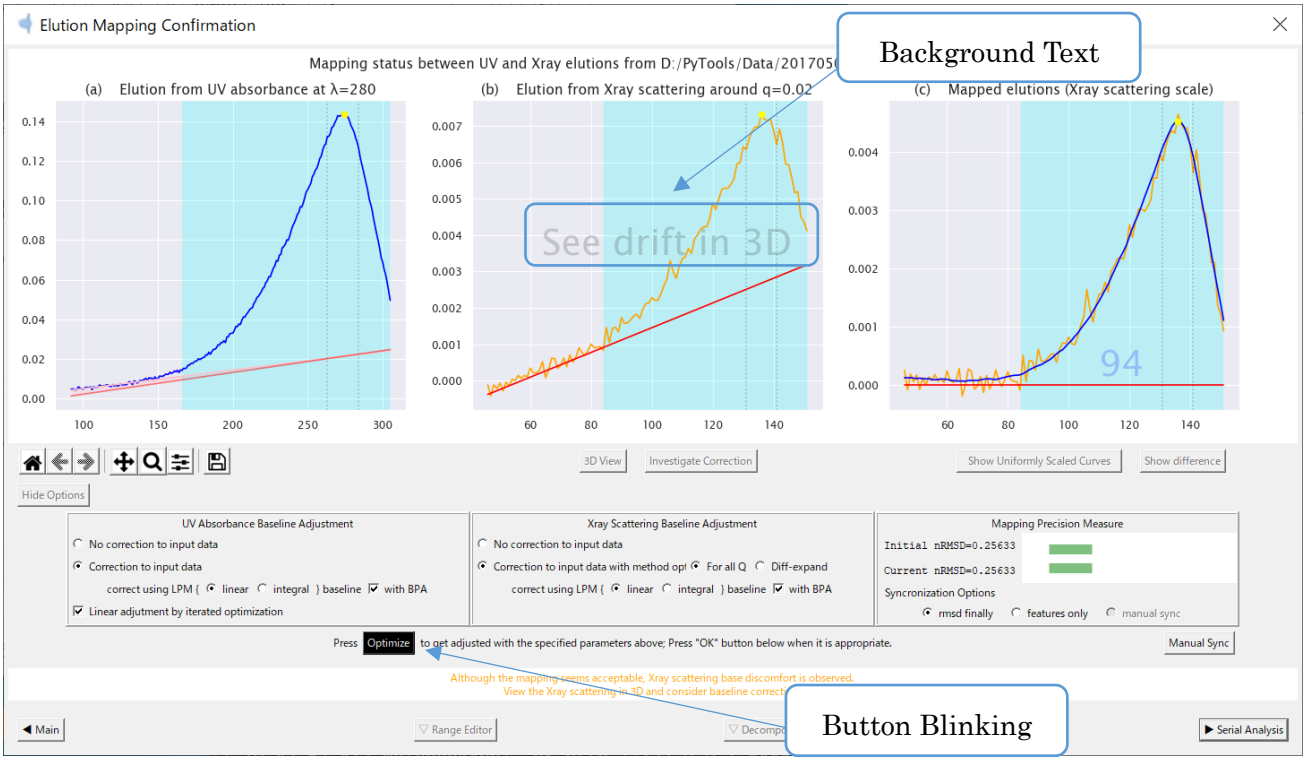
Tab. 7-4 Situations requiring user's attention

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. 7-20 Mapping Dialog with Button-Blinking and Background Text



7.2.5 Baseline Correction Hierarchy

a For both of UV and X-ray data, baseline correction is implemented similarly in the following two hierarchical ways.

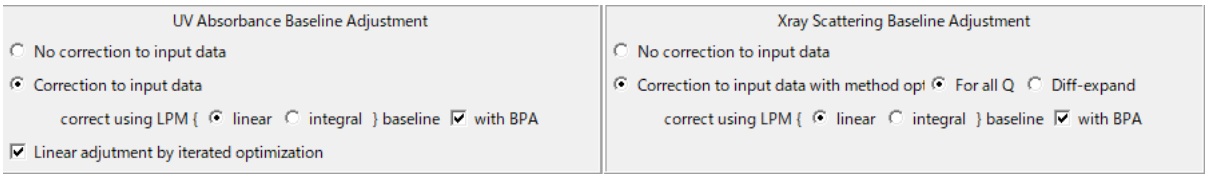
1. Correction without adjustment
2. Correction with adjustment

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. 7-21 Baseline Correction Options (in default selection)

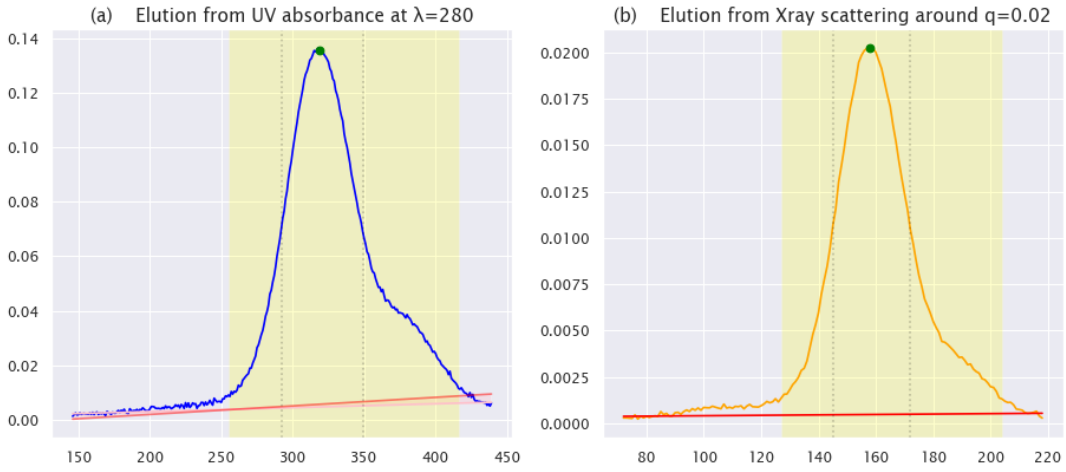


e Therefore, there are, in all, $6 = 3 \times 2$ patterns of combination if "No correction" options for both data are counted.

f Corrected baselines and adjustment amounts are shown in red lines and pink belts in the figures (See Fig. 7-22).

g Detailed descriptions of this section are provided in the following three sections.

Fig. 7-22 Corrected Baselines and Adjustment Amounts shown in the figures



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

a UV and X-ray data have similar 3D structures, i.e., values for each 2D-point, (wavelength, elution No) for UV, (scattering vector length, elution No) for X-ray, 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 where its wavelength=280.

e On the other hand, X-ray data are, of course, fully used. Therefore, the baseline correction is applied to every elution curve at every Q value, not only in the mapping plane where scattering vector length Q=0.02 in the default setting.

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.

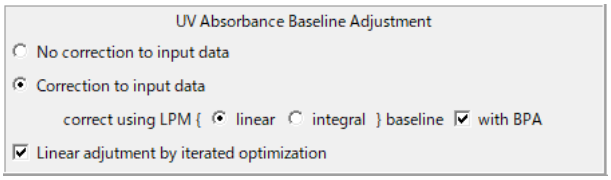
7.2.7 UV Absorbance Baseline Correction Options

a With "No correction" option, input data are used without any correction.

b With "Correction to input data", the method LPM (linear or integral) is automatically selected by default to give a better mapping. Select the other and re-optimize if it seems inappropriate. See sections 7.2.9 and 7.2.10 for explanation.

c If chosen, "Linear adjustment" for UV elution curve is determined to minimize the RMSD between the mapped elution curves. In this version of the program, such adjustments are allowed only to UV side.[14]

Fig. 7-23 UV Absorbance Baseline Adjustment Options



d When "No correction" is chosen manually to change the automatically determined default, "Linear adjustment" is forced to be OFF after the warning message as shown below.

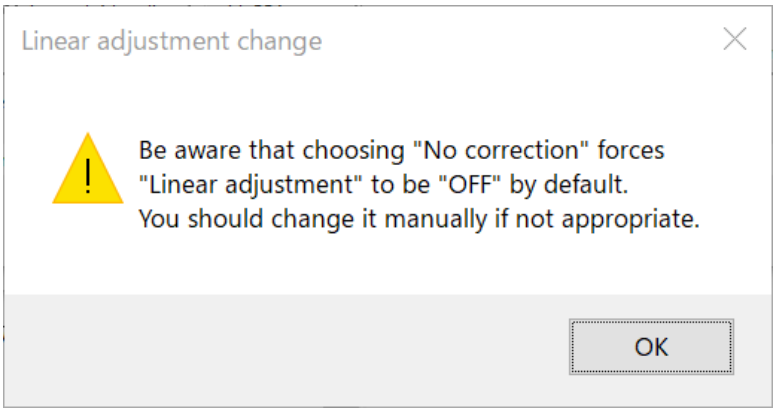
e A similar message will appear when you change from "No correction" to "Correction".

f In either such case, change the forced state manually if it against your intention.

---

[14] This "linear adjustment" was possible both for UV and X-ray in older versions of the program. However, that for X-ray has been omitted since it is usually considered not appropriate.

Fig. 7-24 Forced linear adjustment warning



g See 7.2.12 for “with BPA” option.

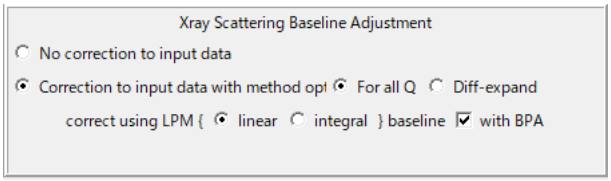
7.2.8 X-ray Scattering Baseline Correction Options

a With “No correction” option, input data are used without any correction.

b With “Correction to input data”, the method is automatically selected by default to give a better mapping. Select the other and re-optimize if it seems inappropriate. See sections 7.2.9 , 7.2.10 and 7.2.11 for explanation.

c See 7.2.12 for “with BPA” option.

Fig. 7-25 X-ray Scattering Baseline Adjustment Options



7.2.9 Low Percentile Method (linear)

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. 7-26)
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.[15]
6. Apply the above to all elution curves.

Fig. 7-26 Explanation of 25 percentile point

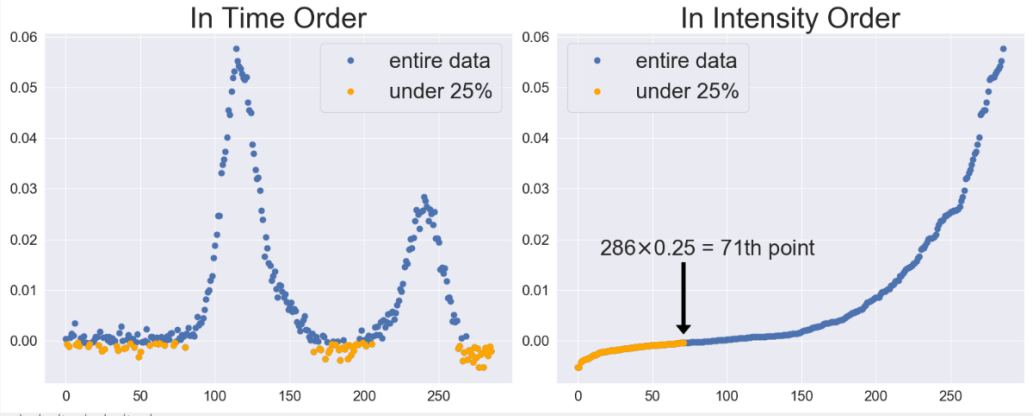
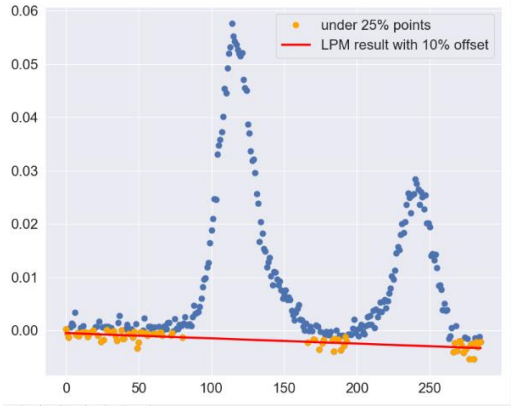


Fig. 7-27 Baseline determined at 10 percentile point



[15] 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.

7.2.10 Low Percentile Method (integral)

a While the assumed baseline is usually linear, it is sometimes observed that drifts have been caused by the accumulation of some maybe broken samples, in which case the baseline forms the curve such as shown in the figure below.

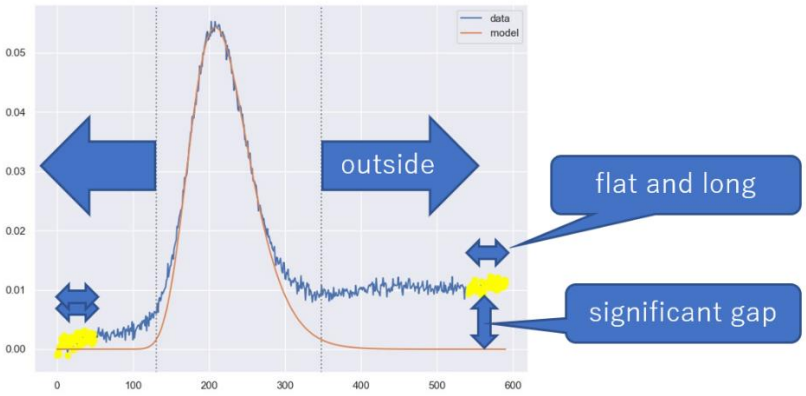
Fig. 7-28 Example likely suitable to integral baseline correction

b With this “integral” option, the program generates the curve on the assumption that those broken substances accumulate proportionally to the intensity of scattering.

c In the automated control, this “integral” option is selected only if all of the following conditions are observed (See Fig. 7-29 on the next page).

1. The baseline increases from left to right with a significant gap.
2. The data includes enough base region outside of peaks.
3. Such base regions are sufficiently flat.

Fig. 7-29　Conditions suitable for integral correction



7.2.11　Correction Application Order

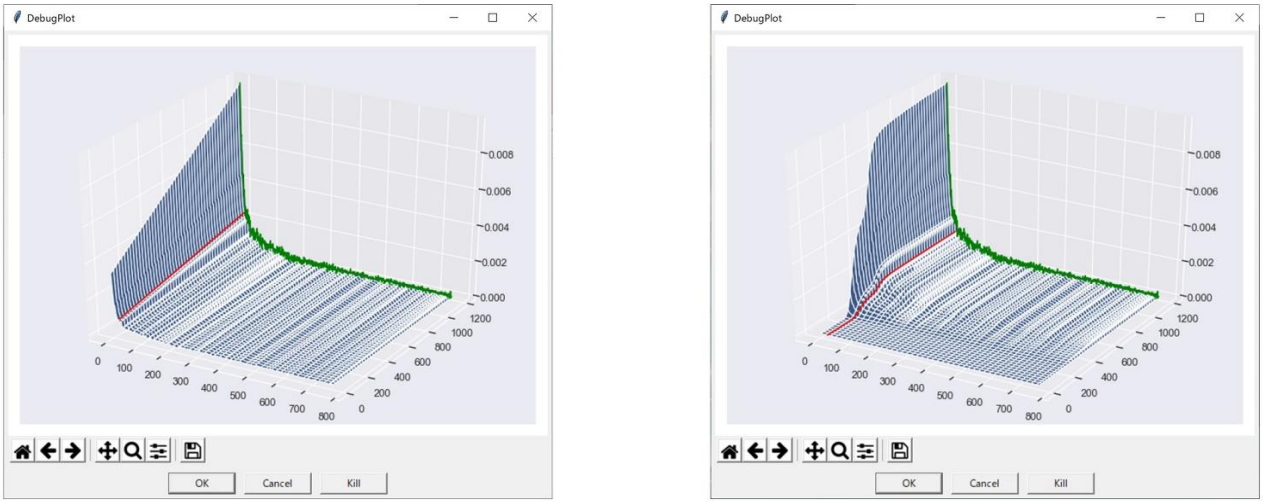
a　As described in the above, LPM can be applied at each fixed Q (scattering vector length) and applying “for each Q” separately is the default way of correcting the whole data.

b　On the other hand, it can be performed by first determining the baseline only in the standard mapping plane, which is indicated with a red line or curve in the figure below, and expanding the line or curve along the Q-axis using the total difference scattering curve, which is indicated with green curves in the same figures.

c　Which option is suitable for the case can be inspected using the Baseline Correction Inspection (Section 7.1.6), and be changed in the Mapping Dialog options (Section 7.2.8).

Fig. 7-30　Diff-expanded base surface

Linear　　Integral



7.2.12 Base Plane 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.

$$\underset{a,b,c}{\operatorname{argmin}} \left[\left\| \mathrm{P}_{a.b,c} \cdot \mathrm{C}_{a.b.c} - \mathrm{M}_{a,b,c} \right\|_{\mathrm{F}}^{2} \right] \text{ or} \tag{7-1}$$

$$\underset{a,b,c}{\operatorname{argmin}} \left[\left\| \mathrm{M}_{a.b,c}[*, \mathrm{j}] \cdot \mathrm{M}_{a.b.c}[\mathrm{i},*] - \mathrm{M}_{a,b,c} \right\|_{\mathrm{F}}^{2} \right] \tag{7-2}$$

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

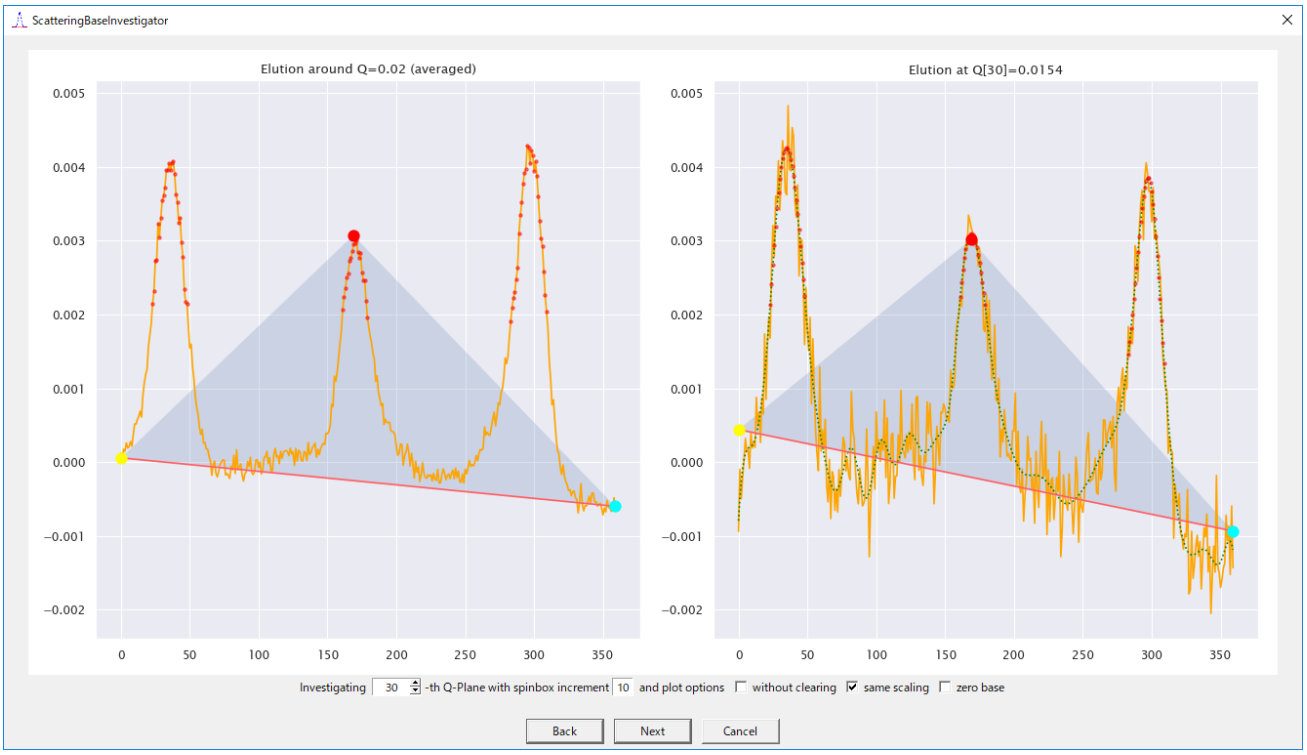
7.2.13 X-ray Scattering Baseline Investigation

a As mentioned in Section 7.2.6, 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. 7-31 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.

7.2.14 UV-Xray Mapping 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 mapping process is roughly made of the following two steps.

1. Feature point matching
2. Minimization of the RMSD

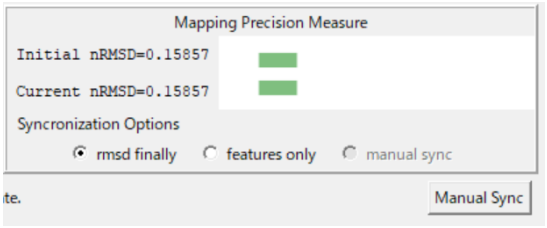
d Keeping the above in mind, users can choose from the three options in the table below.

Tab. 7-5　Mapping Options

Option Name	Step1	Step 2	Description
rmsd finally	○	○	Adopt the result of step 2 after conducting step 1
features only	○		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. 7-32)

Fig. 7-32　Radio Buttons for Synchronization Control

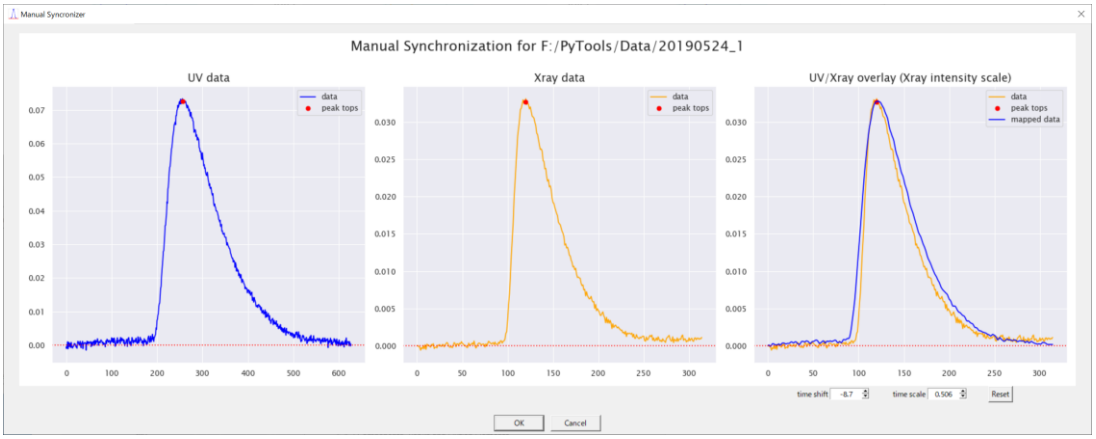


f　Users can choose another option with those radio buttons.

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

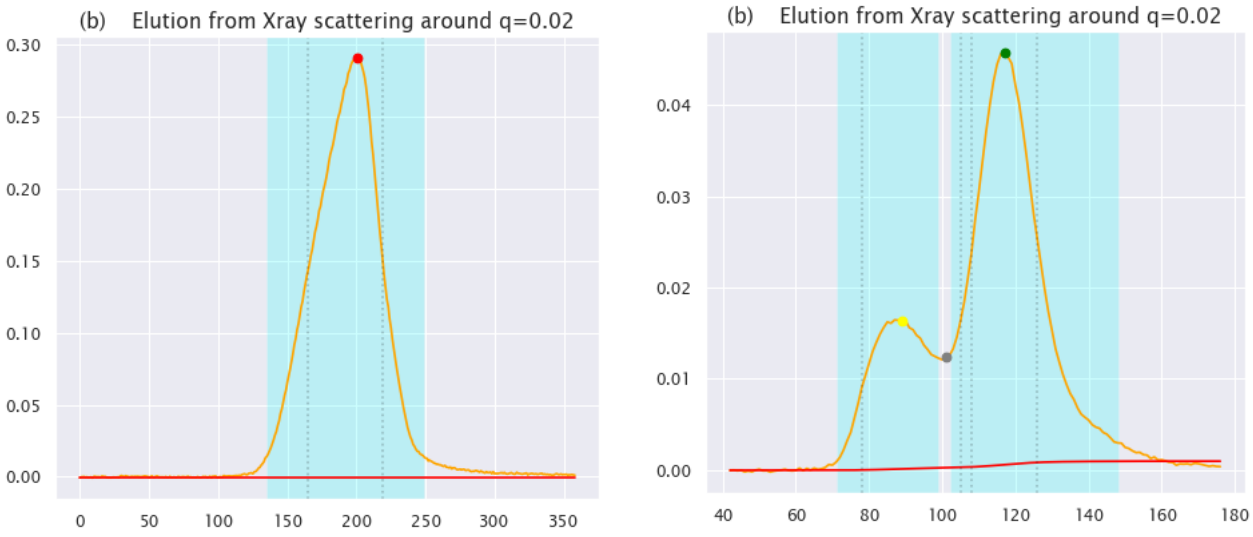
Fig. 7-33　Manual Mapping Dialog



7.2.15 Concentration Dependency Inspection

a In the mapping dialog figures, recognized peak tops are marked with dots in different colors according to their levels of concentration dependency. (See Fig. 7-34 and Tab. 7-6 below)

Fig. 7-34 Peak tops marked with different colors



Tab. 7-6 Colors distinguishing concentration dependency levels

Color of peak top mark	Concentration dependency	SCD range
● Red	Significant	0 ～ 2
● Yellow	Weak	2 ～ 5
● Green	Ignorable	5 ～

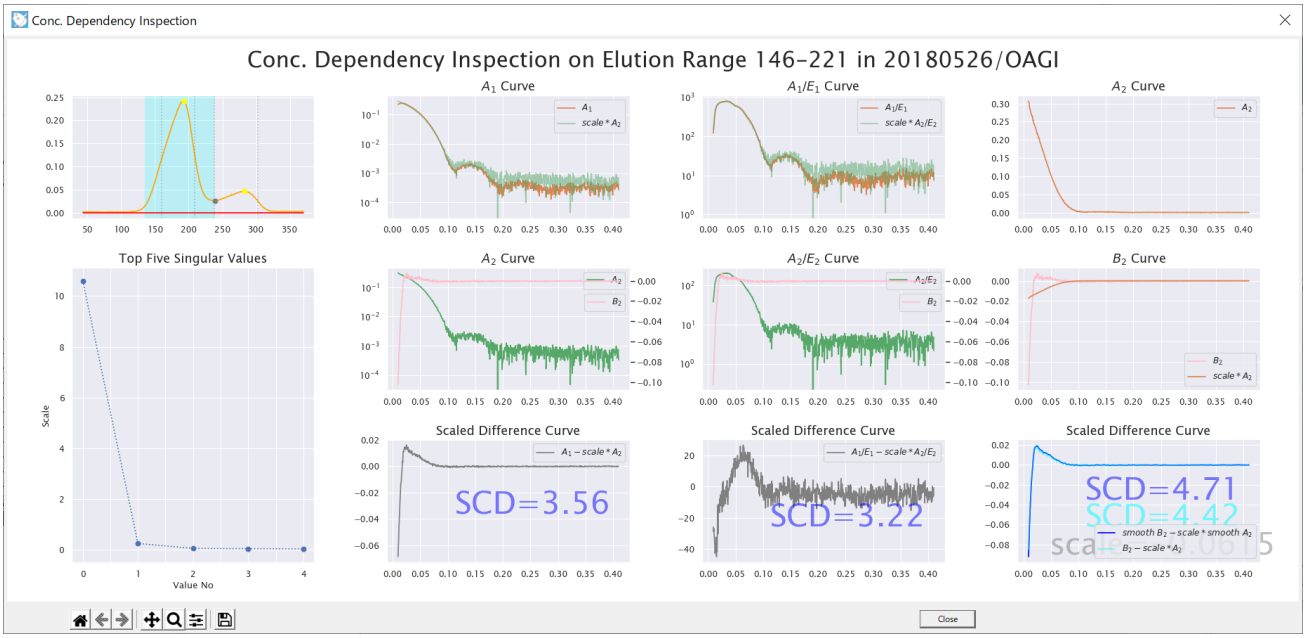
b More detailed information, such as SCD[16], is available in the "Conc. Dependency Inspection" (Fig. 7-35 below), which can be shown from the popup menu appearing with right-click.

c Although SCD definition and this inspection dialog design given below are in a state of "groping" due to the difficulty of defining the concentration dependency level, they are presented "as is" as of this writing.

d At the left-most are shown the top five singular values from SVD of the elution range indicated with cyan rectangle, which can be used to observe the rank of the data matrix in that range. For example, a significant ratio of the second singular value to the first could be interpreted to suggest that the rank is two, which indicate the existence of significant interference effects. However, the problem is that we cannot easily distinguish between rank one and two only by using such singular values, which has led us to the SCD described below.

[16] SCD is an acronym for Score of Concentration Dependency, which is described in the next page.

Fig. 7-35　Concentration Dependency Inspection Dialog



e　In this dialog, three types of SCD's are given, which are summarized in the table below.

Tab. 7-7　Colors distinguishing concentration dependency levels

SCD type name	Definition	Adoption State
A(q) difference	Normalized RMSD of A_1 and A_2	For human reference only
A(q)/E(q) difference	Normalized RMSD of A_1/E_1 and A_2/E_2	For human reference only
Modified B(q) difference[17]	Normalized RMSD of $B_1 - s_1 A_1$ and $B_2 - s_2 A_2$ where s_i is determined to minimize $\|B_i - s_i A_i\|$ respectively	Adopted in the program

A_1 : First column of rank 1 SVD-reconstructed solution $P = M_1 \cdot {C_1}^+$

A_2 : First column of rank 2 SVD-reconstructed solution $P = M_2 \cdot {C_2}^+$

E_1 : Propagated error corresponding to A_1

E_2 : Propagated error corresponding to A_2

B_1 : Second column of rank 1 SVD-reconstructed solution $P = M_1 \cdot {C_1}^+$

B_2 : Second column of rank 2 SVD-reconstructed solution $P = M_2 \cdot {C_2}^+$

f　In the program, only the last type of SCD in the above table is used to distinguish the concentration dependence levels, which accordingly determine the ranks used in SVD-reconstruction and concentration matrix C.

[17] The modification term $-s_i A_i$ which appears in the definition is a kind of work-around to avoid a possible inconvenience, which may arise when the rank of the data matrix of this range is nearly one, where A_i and B_i could be computed as if almost linearly dependent as a result of confusing noises and concentration effects. (i.e., linear algebra does not care such distinction)

7.3 Range Editor

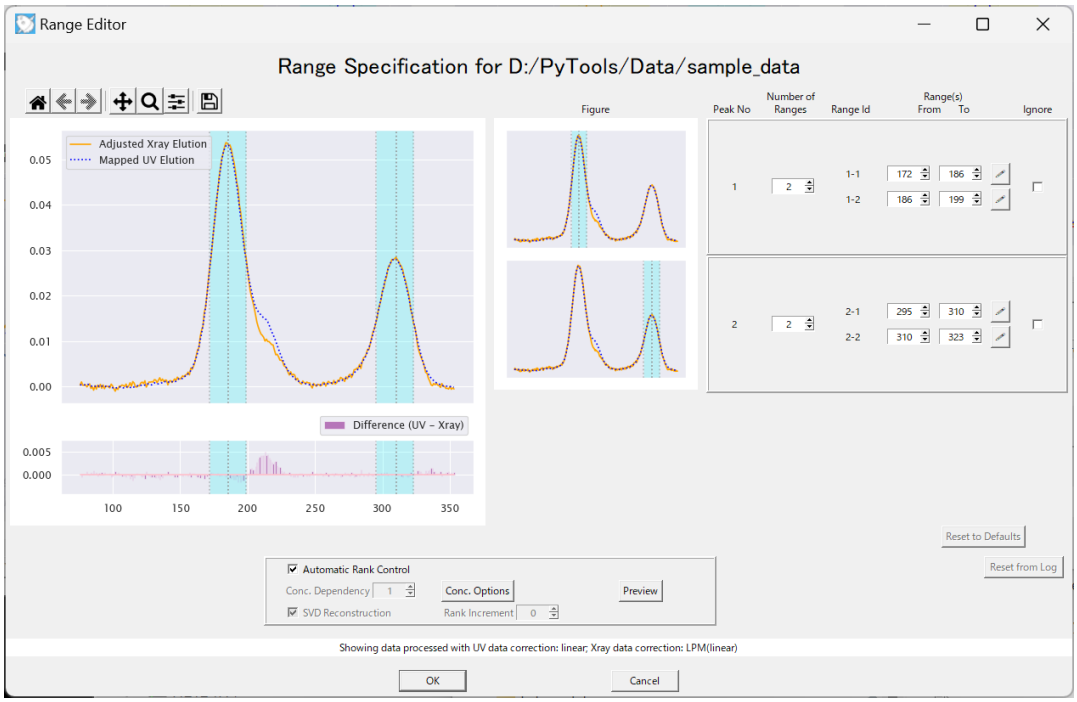
7.3.1 Purpose and Manipulation Flow

a The default analysis ranges, depicted with cyan bars in Fig. 7-36 below, are automatically determined by the program, and this dialog is for changing them.

b It is invoked by the "▽ Range Editor" button in the Mapping Dialog.

c As for the "Preview" button with the preceding options, except "Conc. Opts" button described soon below, see Section 7.6.

Fig. 7-36 Range Editor



7.3.2 How to Change Analysis Ranges

a "Number of Ranges" per peak can either be one or two depending on your needs.

b In other words, you should select "double ranges" if you want separate extrapolation results for each side of the peak, otherwise select "single range".

c Using this spin box, you can select either of these numbers, as shown in the following figures.

Fig. 7-37 Double Ranges per Peak

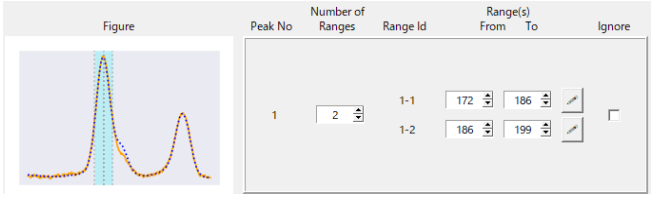
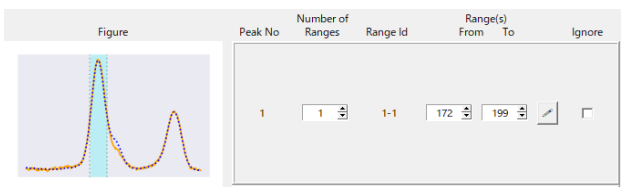


Fig. 7-38 Single Range per Peak

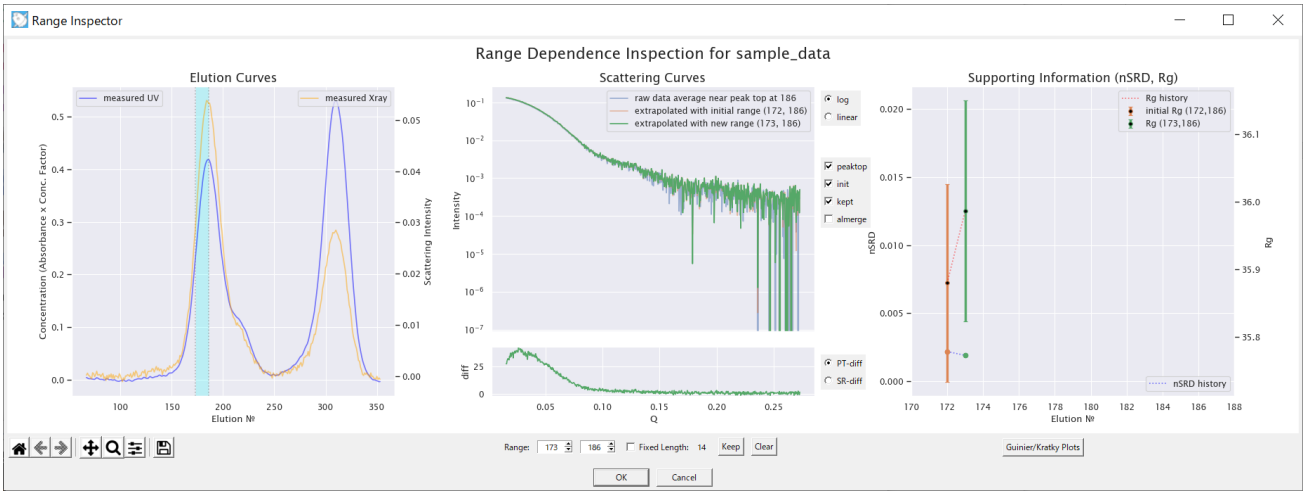


d To change the width of each range, use the spin boxes corresponding to the end points of the

range.

e Another way of changing the ranges is to invoke Range Inspector by clicking "✎" button, which shows a dialog shown below.

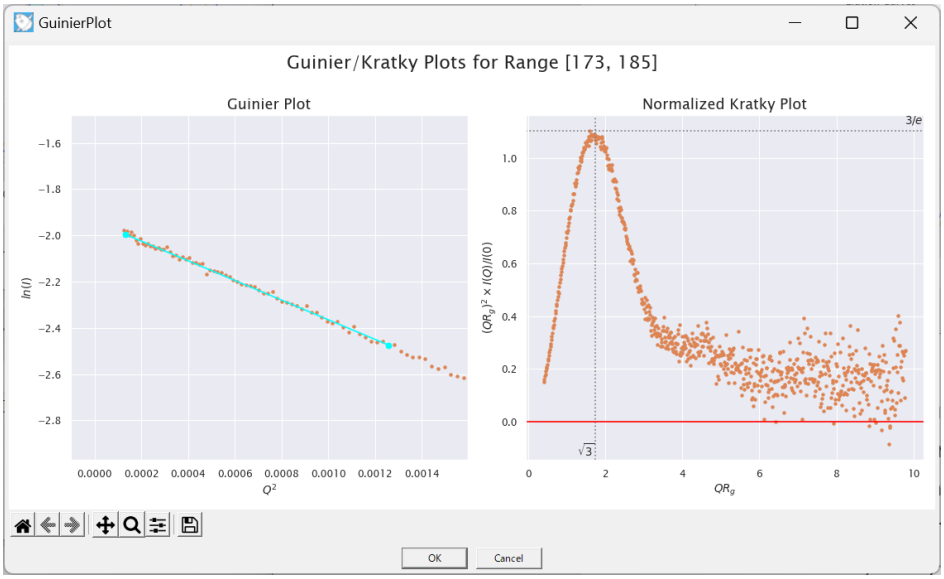
Fig. 7-39 Range Inspector Dialog



f In this dialog, users can instantaneously observe the LRF result of the modified range, either by comparing the resulted scattering curve with its raw peak top ridge curve in the middle figure, or by observing the change of its estimated Rg value.

g For more information, Guinier plot or Kratky plot as shown below are available from the "Guinier/Kratky Plot" button.

Fig. 7-40 Guinier / Kratky Plot



7.3.3 Concentration Options

a In the current version of the program, the elution curve used as concentration is selectable.

b By default, the mapped 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. 7-41 below.

f To observe the difference of these two options, three types of figures, shown in the figures below, are provided

Fig. 7-41 Concentration Options Dialog

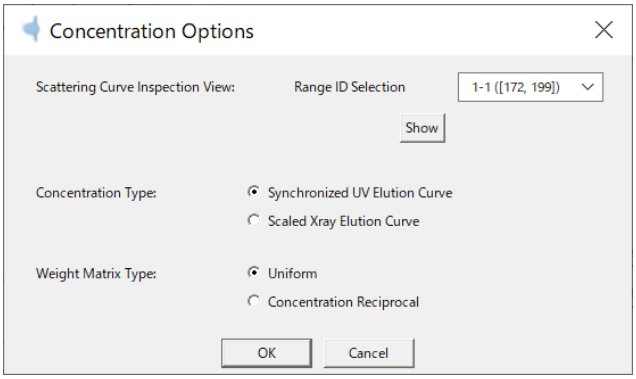


Fig. 7-42 Linear Plot of the Curves

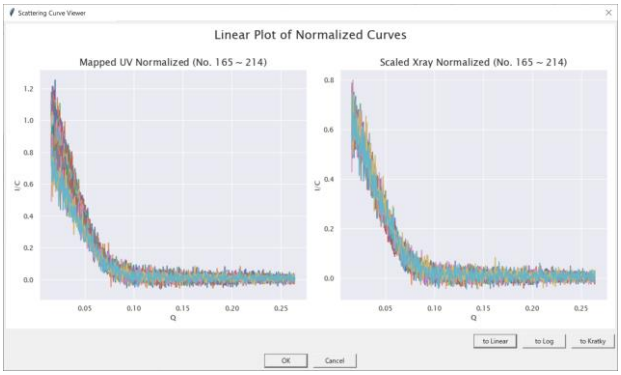


Fig. 7-43 Log Plot of the Curves

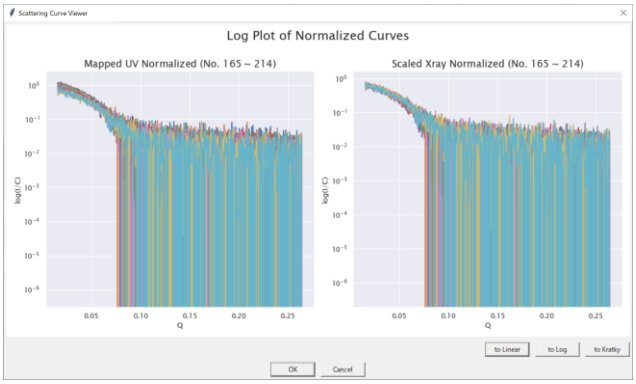
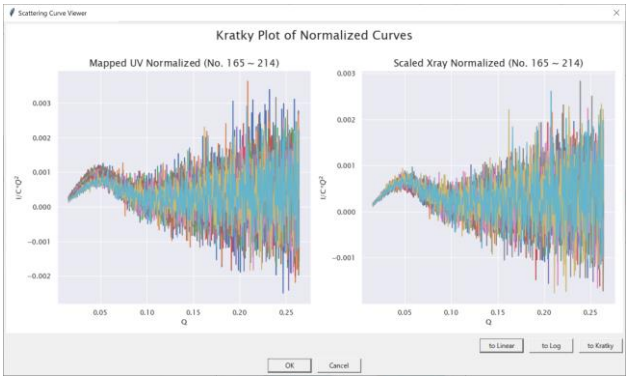


Fig. 7-44 Kratky Plot of the Curves



g "Weight Matrix Type" switches the formulation used for the extrapolation (or low rank factorization) between the following two.

1. Uniform (default): minimize $\|P \cdot C - M\|$
2. Concentration Reciprocal: minimize $\|P \cdot C/c_0 - M/c_0\|$ where X/c_0 means dividing each row of matrix X by the first row c_0 of concentration matrix C.

7.4 Decomposition Editor

7.4.1 Purpose and Manipulation Flow

a This dialog, as shown in Fig. 7-45 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.

b For this purpose, elution curve models such as EMG[18] or EGH[19] are employed.

c It is invoked by the "▽ Decomposition Editor" button in the Mapping Dialog.

d Alternatively, you can set the "Dialog Navigation Preferences" in the Settings Dialog (see Section 7.7) to let it automatically navigate you along to this editor.

e As for the "Preview" button with the preceding buttons, see Section 7.6.

Fig. 7-45 Decomposition Editor (EMG with X-ray data)

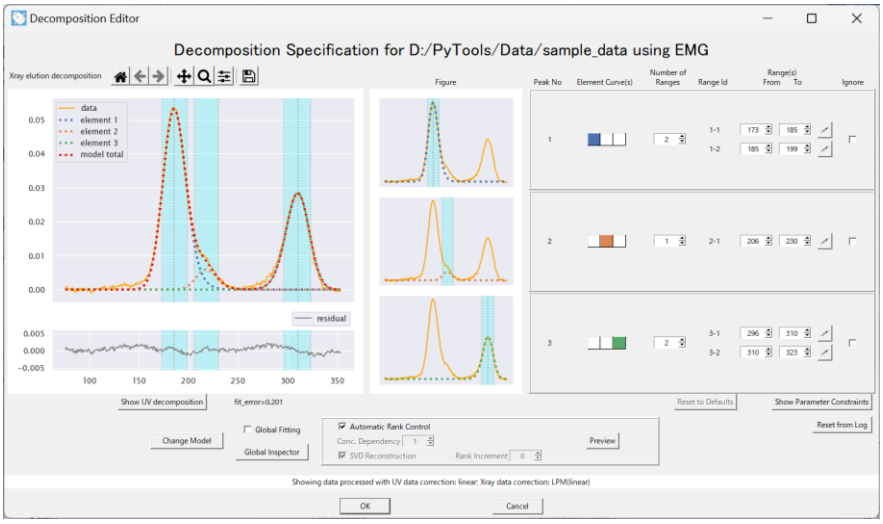
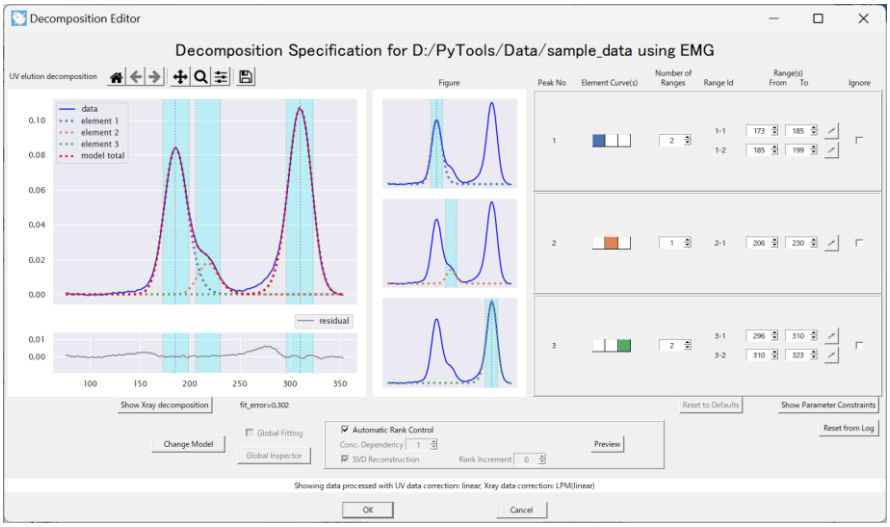


Fig. 7-46 Decomposition Editor (EMG with UV data)



[18] Exponentially Modified Gaussian

[19] Exponential Gaussian Hybrid

7.4.2 Two Types of Data by Two Types of Models

a 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.

b Therefore, in all, you have four types of views to choose which model is better.

c Confirm the meaning of the above statement and observe the slight differences among them by the following table and figures (Fig. 7-45, Fig. 7-46, Fig. 7-47, Fig. 7-48).

Tab. 7-8 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. 7-47 Decomposition Editor (EGH with X-ray data)

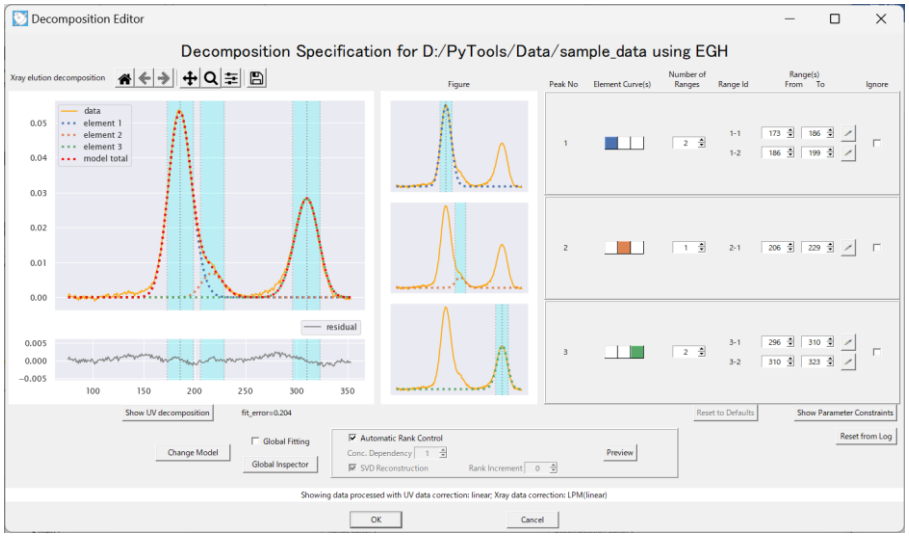
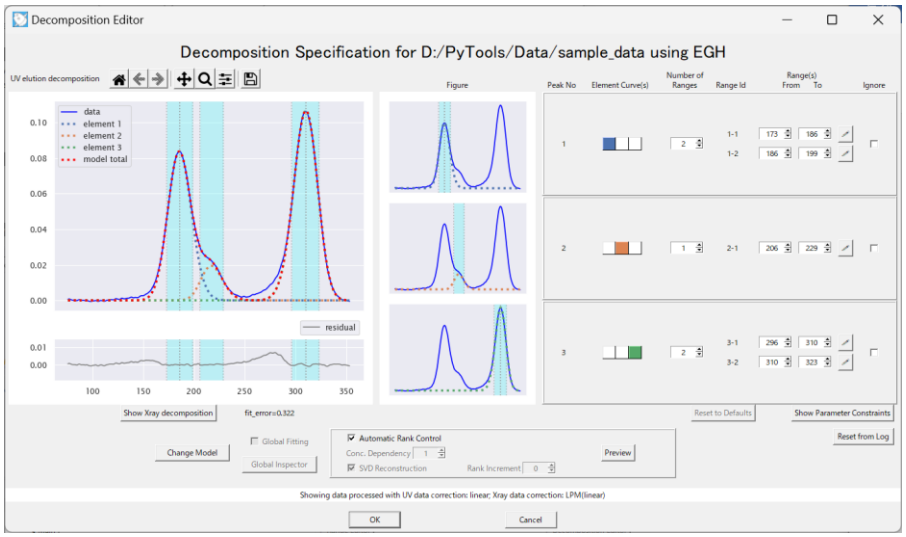


Fig. 7-48 Decomposition Editor (EGH with UV data)



7.4.3 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.[20]

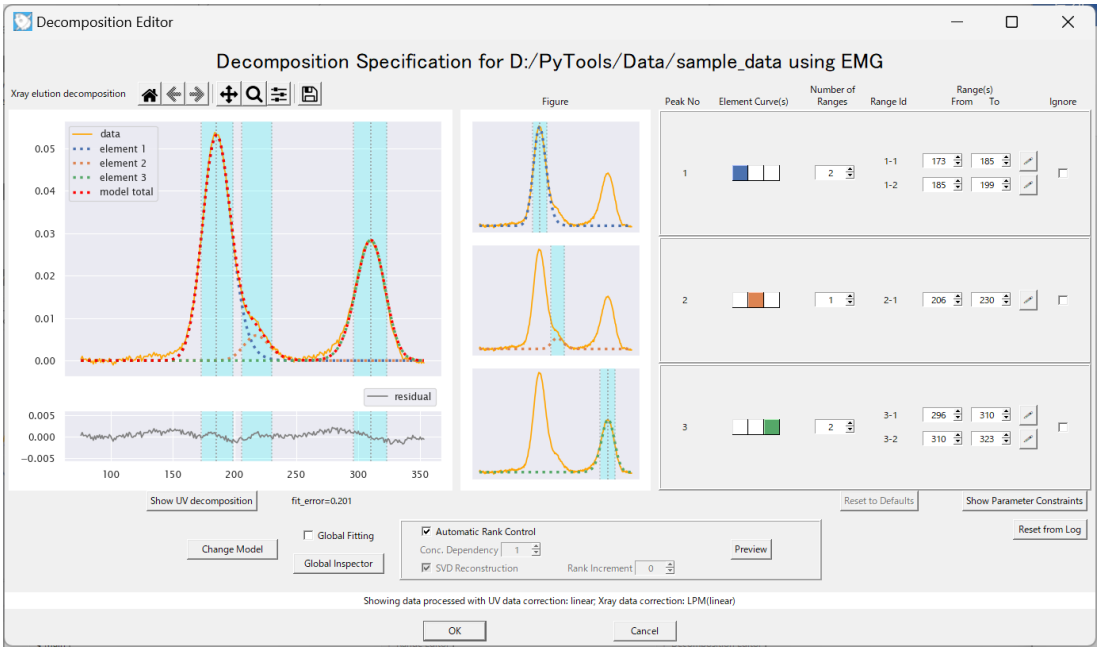
c In the example shown in the above figures, there observed are four elution elements identified by colors: ■, ■, ■, ■.

7.4.4 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. 7-49 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.

[20] A composite of multiple elution elements is used when a single element is not enough to model a peak.

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.

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.

7.4.5 Editing Combination of Elution Elements

a Now, using the EGH model in another example, 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. 7-50 One major peak consisting of two elements

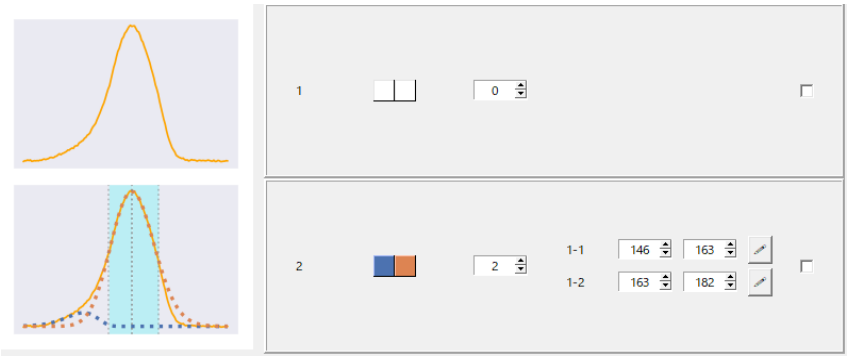
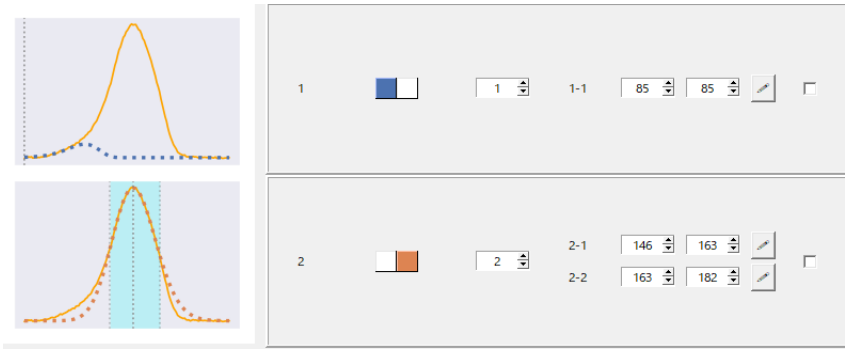


Fig. 7-51 Two peaks (minor and major) with each single element



c With the first interpretation (Fig. 7-50), 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. 7-51), 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.

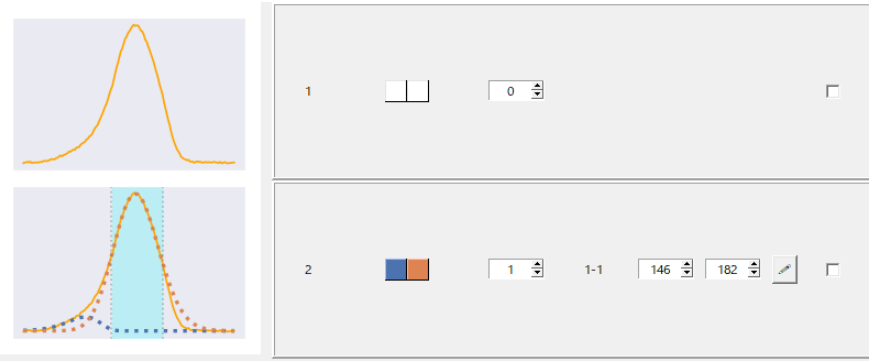
e To change the interpretation between the above two, we use “number of ranges” of minor peaks (which correspond to the upper frame — peak No. 1 — in this case).[21]

f If you reduce the number to zero, you will get the combination of Fig. 7-50.[22].

g If you increase the number to one, you will get the combination of Fig. 7-51.

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. 7-52 One major peak with one analysis range



7.4.6 How to Change Analysis Ranges

a Changing the width of ranges here is possible in the same way as with Range Editor described in Section 7.3.2, except that the concentration curves used in the LRF are modeled (i.e., EMG or EGH) curves described above in this chapter.

[21] For this toggle purpose, we use the number of ranges for minor peaks because it is either one or zero and easy to indicate whether the peak exists alone or the element is a part of another.

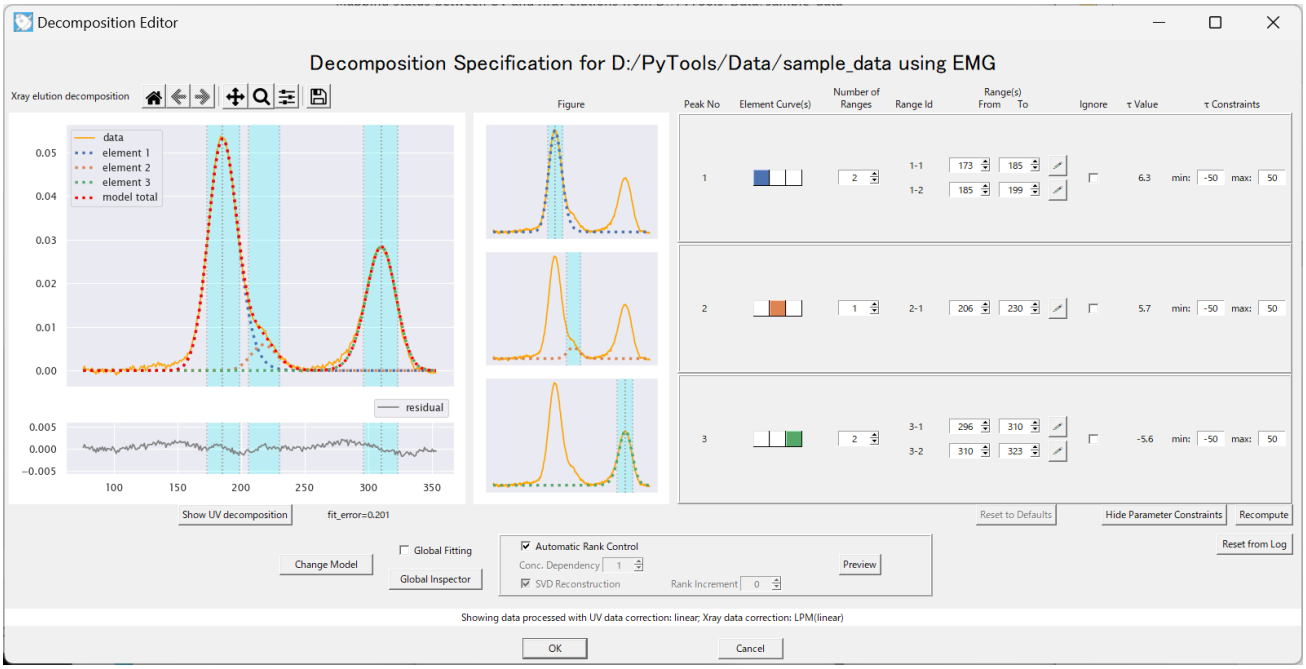
[22] When the number of ranges is set to zero and the element is indicated to be a part of another, there are, in general, two possibilities whether it belongs to a left or right adjacent major peak. It is determined by the distance between the centers of peaks which side it should belong.

7.4.7 Elution Model Parameter Control

a By pressing the “Show Parameter Constraints” button, you can observe and set constraints for the τ parameter of the selected elution model.

b After specifying the minimum or maximum value, press “Recompute” button to get an updated view of the figures and parameter values.

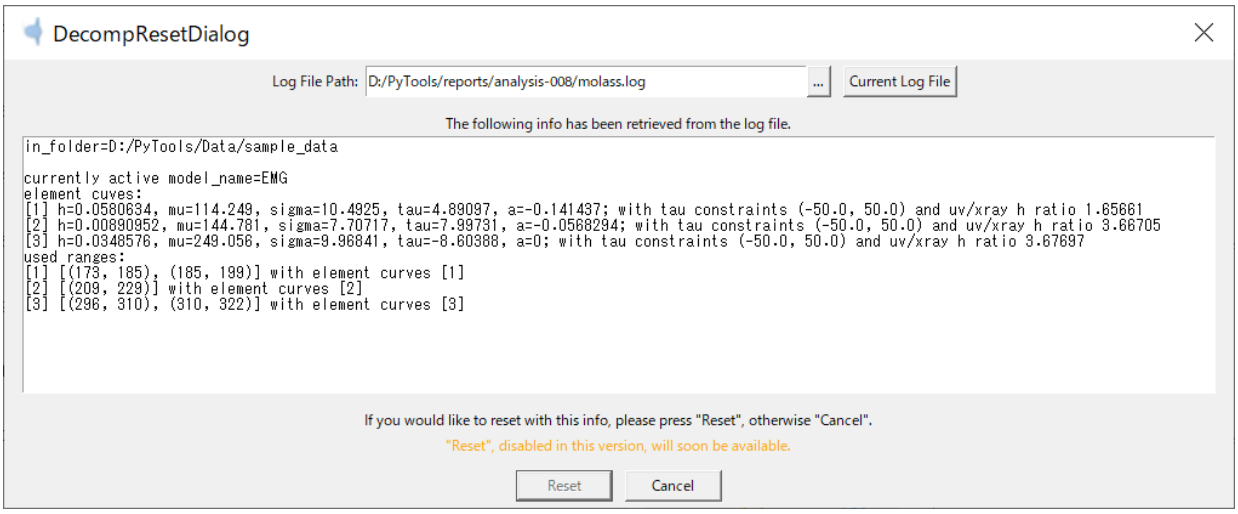
Fig. 7-53 Decomposition Editor with the Parameter Constraints shown



c 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.

d This dialog can be used to restore the set of parameters previously used if you choose the corresponding old log file.

Fig. 7-54 “Reset from Log” Dialog



7.5 Progress Dialog

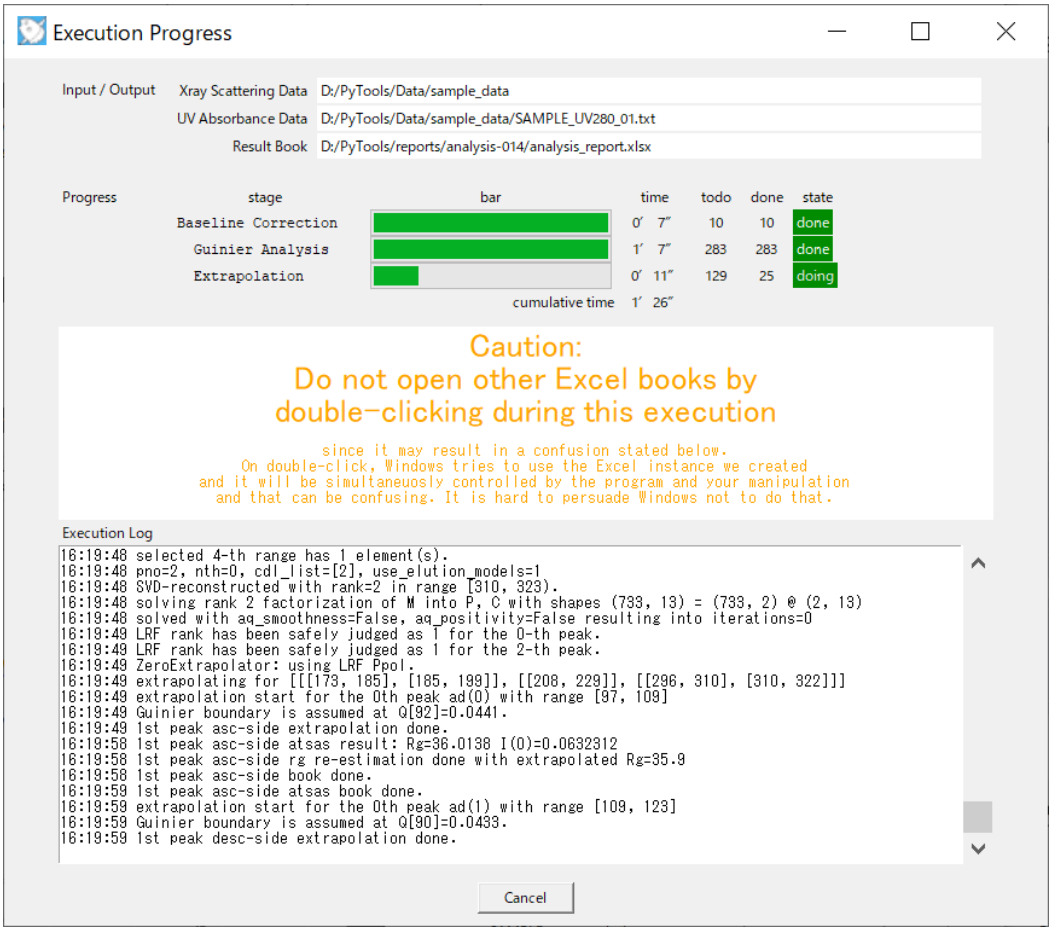
7.5.1 Purpose and Manipulation Flow

a This dialog appears just after pressing “Run” button in Analyzer Dialog.

b It shows the progress with progress bars and log messages.

c You can use the “Cancel” button to stop the execution on the way.

Fig. 7-55 Progress Dialog (in progress)



7.5.2 Description

a The two figures (Fig. 7-55 above, Fig. 7-56 below) show the dialog in the states of “in progress” and “when complete”.

b Input/output information in the top three rows is read-only just for confirmation.

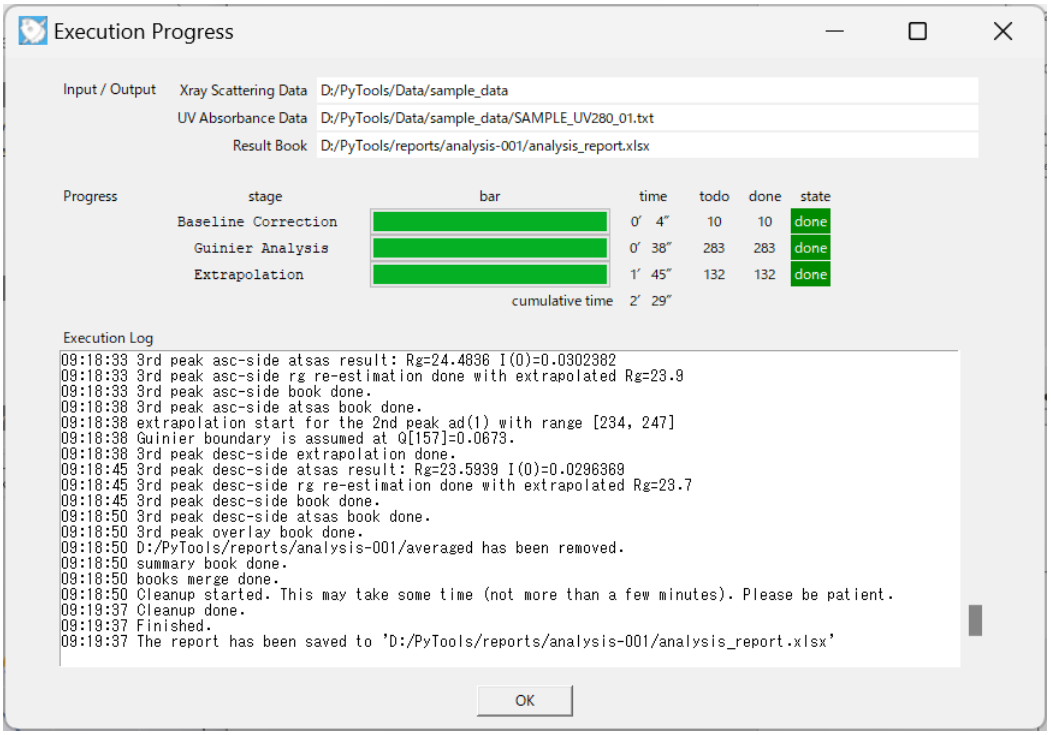
c The numbers, which appear under the column title “todo” and “done”, show rough counts in internal process units, and should not be interpreted strictly.

d Other items are self-explanatory.

e Users are advised not to open other Excel books during this execution to avoid troubles from

simultaneous use of the same Excel instance by human and program.[23]

Fig. 7-56　Progress Dialog (when complete)



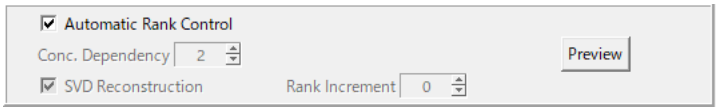
[23] Although the program creates an invisible Excel instance for its exclusive use, Windows usually forces the use of the same instance when requested by humans by double clicking books. It is not easy to persuade the OS so as not to do that.

7.6 Extrapolation (LRF) Preview

7.6.1 Preview Button Frame

a The button which invokes the preview explained in the next section is placed in a frame shown below with its preceding control buttons.

Fig. 7-57 Preview Button Frame in the default state

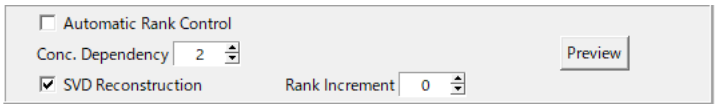


b This "Preview" button frame is placed in each of the two dialogs, Range Editor (7.3) and Decomposition Editor (7.4).

c While the "Automatic Rank Control" is selected by default, what users can do is only to press the "Preview" button since all other option entries are disabled.

d Otherwise, there are available two more options, one for the concentration dependence level, the other for the rank used in SVD-reconstruction.

Fig. 7-58 Preview Button Frame for optional inputs



e For the understanding of the following description, recall the formula below which was introduced in Section 5.1.2.

$$P = M \cdot C^{+} \qquad (5\text{-}7)$$

f "Conc. Dependency" forces the number of rows of the above matrix C as specified. It should be either "1" or "2" corresponding to the following two types of the matrices.

$$\mathrm{C}_1 = [c_1 \quad c_2 \quad \cdots \quad c_n]$$

$$\mathrm{C}_2 = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ {c_1}^2 & {c_2}^2 & \cdots & {c_n}^2 \end{bmatrix},$$

g "SVD reconstruction" indicates the use of reconstructed matrix $\tilde{\mathrm{M}}$ in place of the original data matrix M. It is checked by default. If you uncheck this, reconstruction process will be skipped so that the original M be directly used instead of reconstructed $\tilde{\mathrm{M}}$.

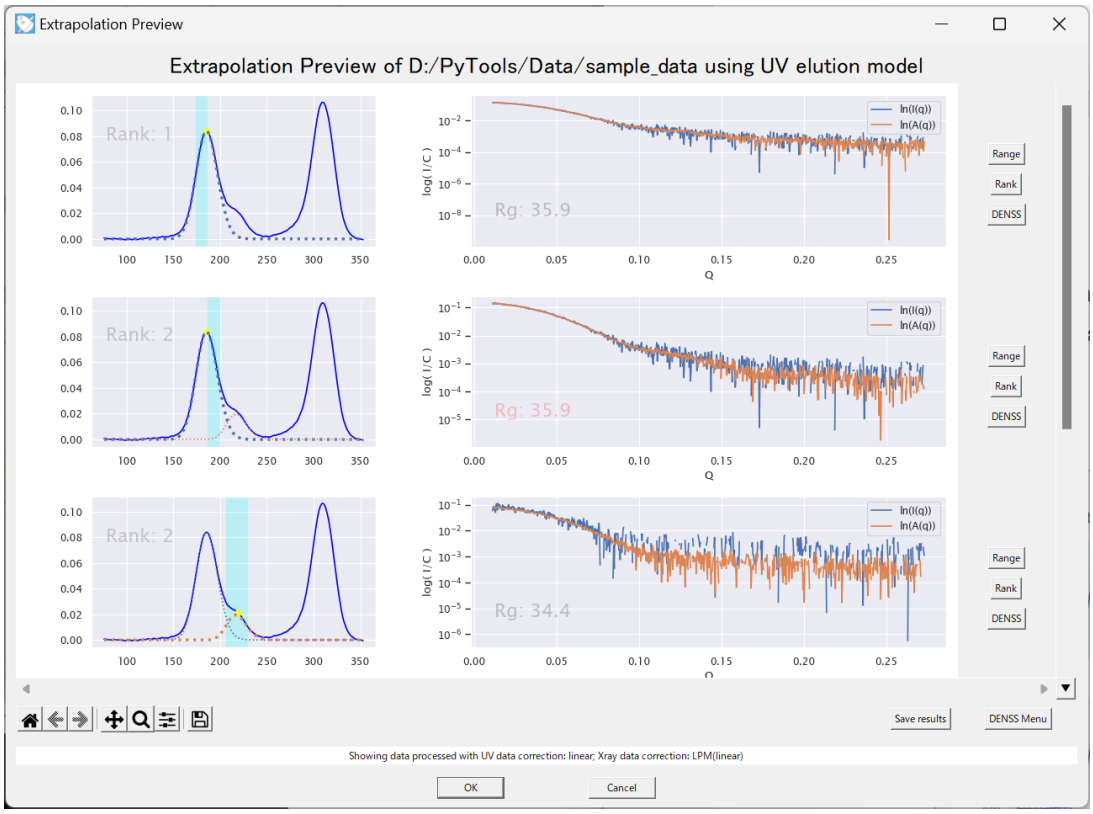
h The rank used in the reconstruction is the same as that of C by default. You can increase it with "Rank Increment".

7.6.2 Purpose and Manipulation Flow

a This preview dialog is for quickly observing the results of extrapolation (LRF) process.

b It is invoked by the “Preview” button mentioned in the previous section.

Fig. 7-59 Extrapolation Preview



7.6.3 Description

a The preview consists of rows, each of which contains three figures and some buttons.

b Each row corresponds to an analysis range as an optimization unit of extrapolation process.

c The 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 Rg value,
3. All the three curves including the B(q) curve in the entire angle region.

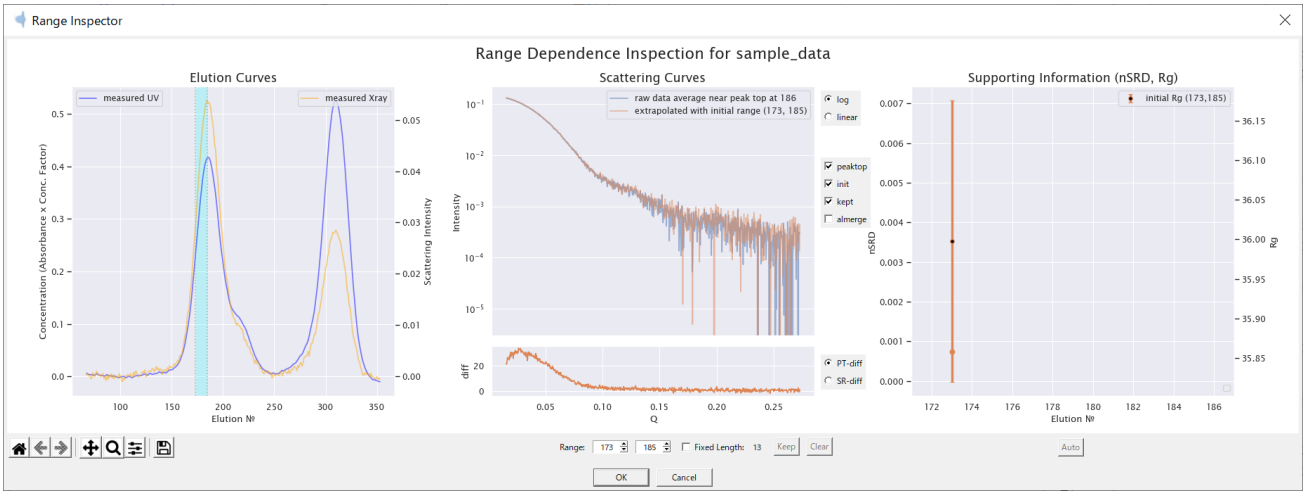
d See the description in 7.6.4 when you encounter a “Low Quality Warning”.

e The usage of buttons on the right-side panel is described in the following table.

Tab. 7-9 Accessory Tools

Button Name	Action
Range	Shows “Range Inspector” (Fig. 7-60) which is used to inspect range dependency of the extrapolation result. After inspecting, the last range will be applied as your analysis range if you press “OK”. (See also Section 7.3.2)
Rank	Shows “Concentration Dependency Inspection Dialog” (Fig. 7-61), which can be used to inspect the rank[24] of the data matrix within the range used in the optimization. See also 7.2.15, paying attention to the difference of the range of inspection, i.e., the range here is usually limited to one side of the peak for optimization purpose, while the range there includes usually the both sides.
DENSS	Invokes a GUI dialog (Fig. 7-62) to run DENSS program, which is yet an experimental feature in this version.

Fig. 7-60 Range Inspector



[24] In this context, inspecting the rank of the data matrix is equivalent to inspecting the score of concentration dependency (SCD) because the rank is regarded as one if the concentration dependency is ignorable, or two otherwise, for example, in case of a single component peak.

Fig. 7-61　Concentration Dependency Inspection Dialog

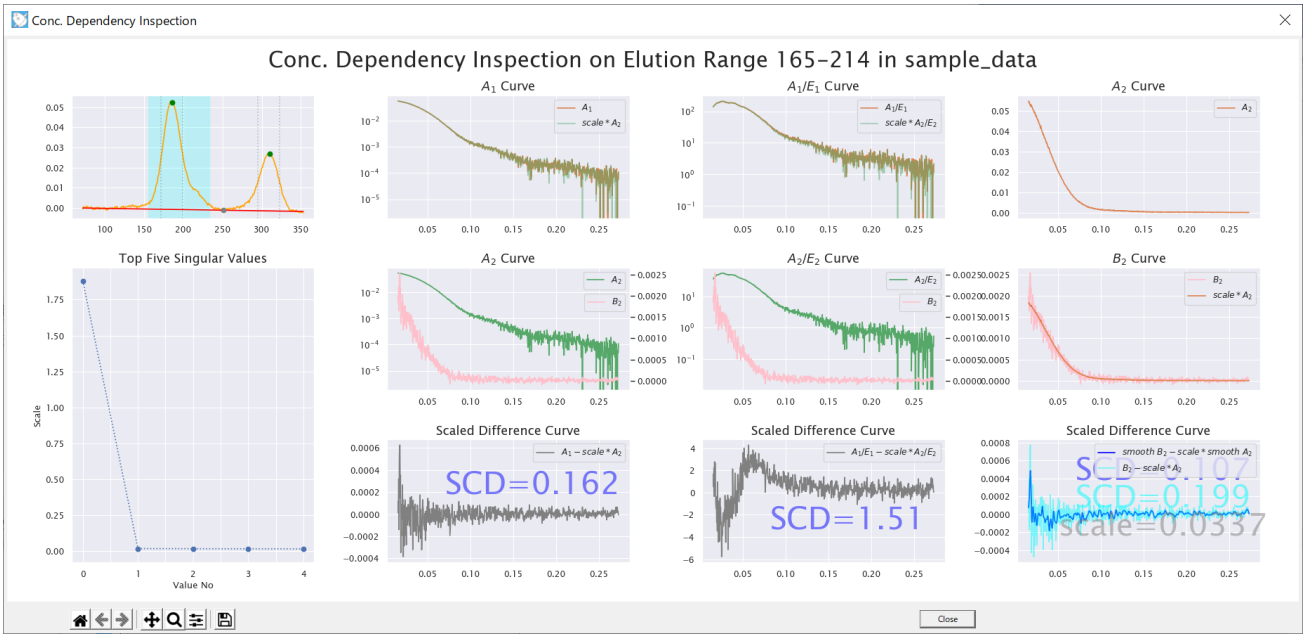
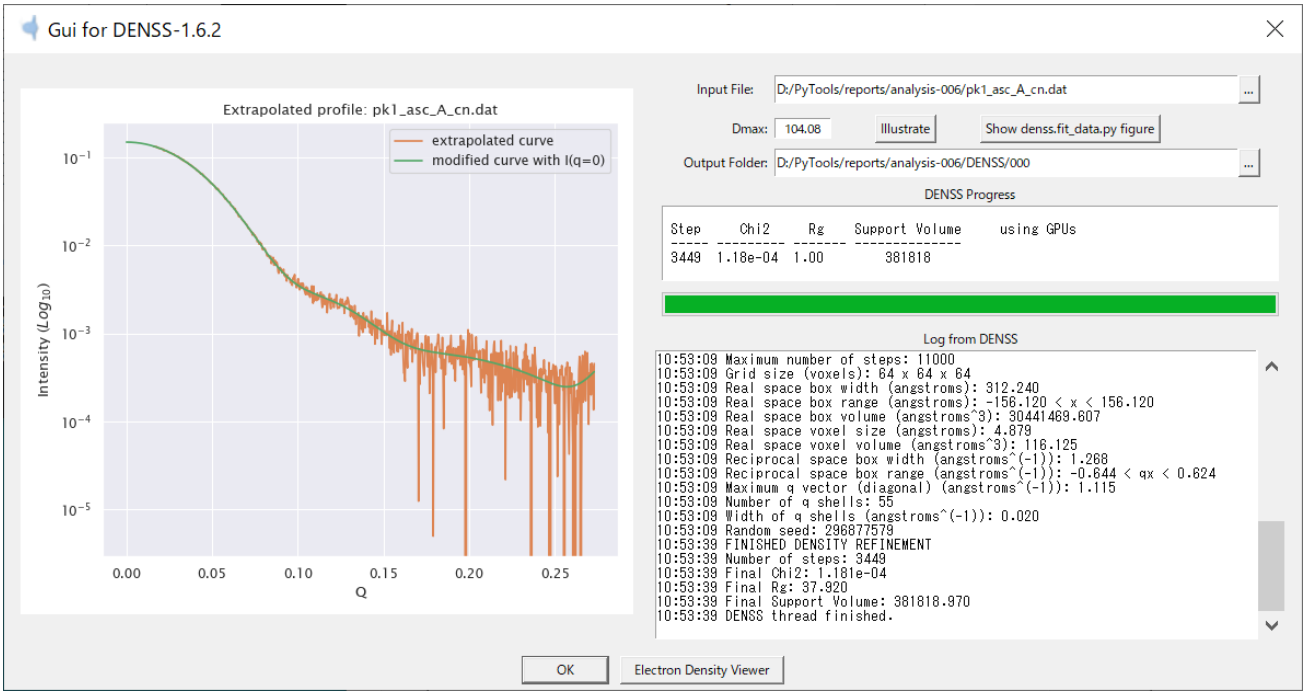


Fig. 7-62　DENSS invocation GUI

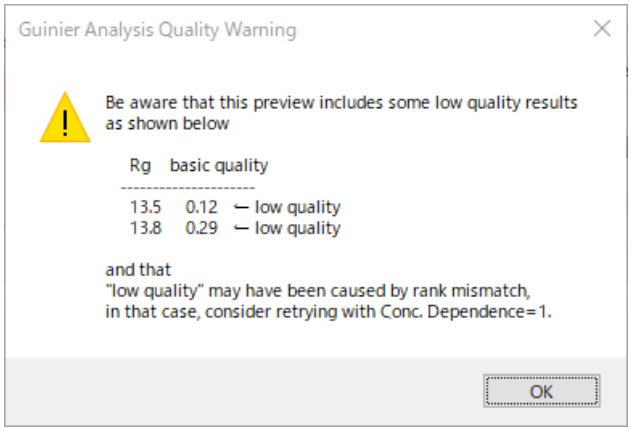


a　See also Section 7.9.1 for DENSS GUI description.

7.6.4 Low Quality Warning

a The program warns you with a message like below when low quality LRF results have been detected.

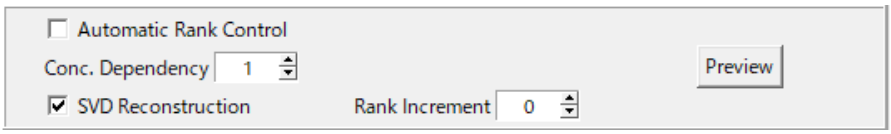
Fig. 7-63 Low Quality Warning



b Such situations can result not only from really "low quality" input data, but also from inadequate rank control of the program where "Conc. Dependency" degree has been estimated as 2, indicating inter-particle effects[25].

c When the latter case is suspected, please consider to retry by manually unchecking "Automatic Rank Control" and changing the degree to 1 in the preview button frame (see below), either in the "Range Editor" or in the "Decomposition Editor".

Fig. 7-64 Changing manually "Conc. Dependency"



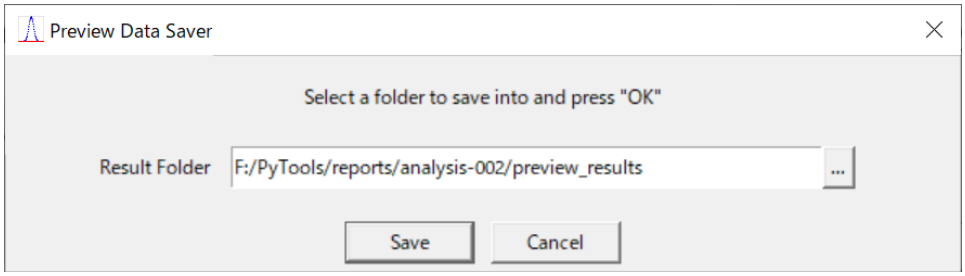
7.6.5 Saving the Preview Results

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.

[25] It can be problematic when the effects are not so strong enough to make the data suitable for rank 2 LRF.

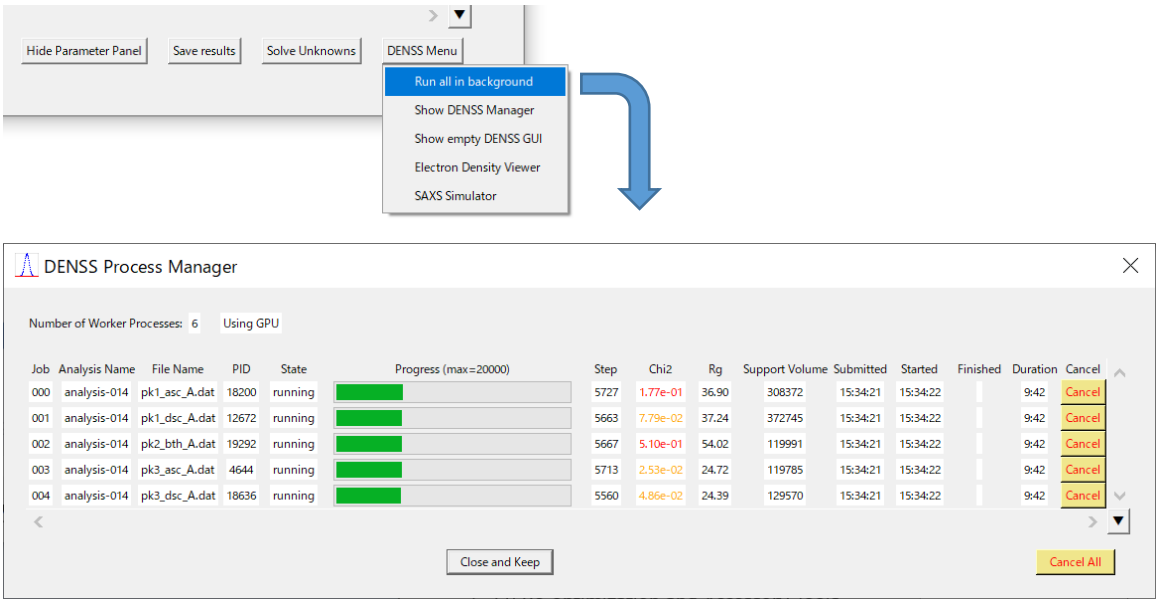
Fig. 7-65 Preview Results Saver



7.6.6 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. 7-66 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 7.9 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.

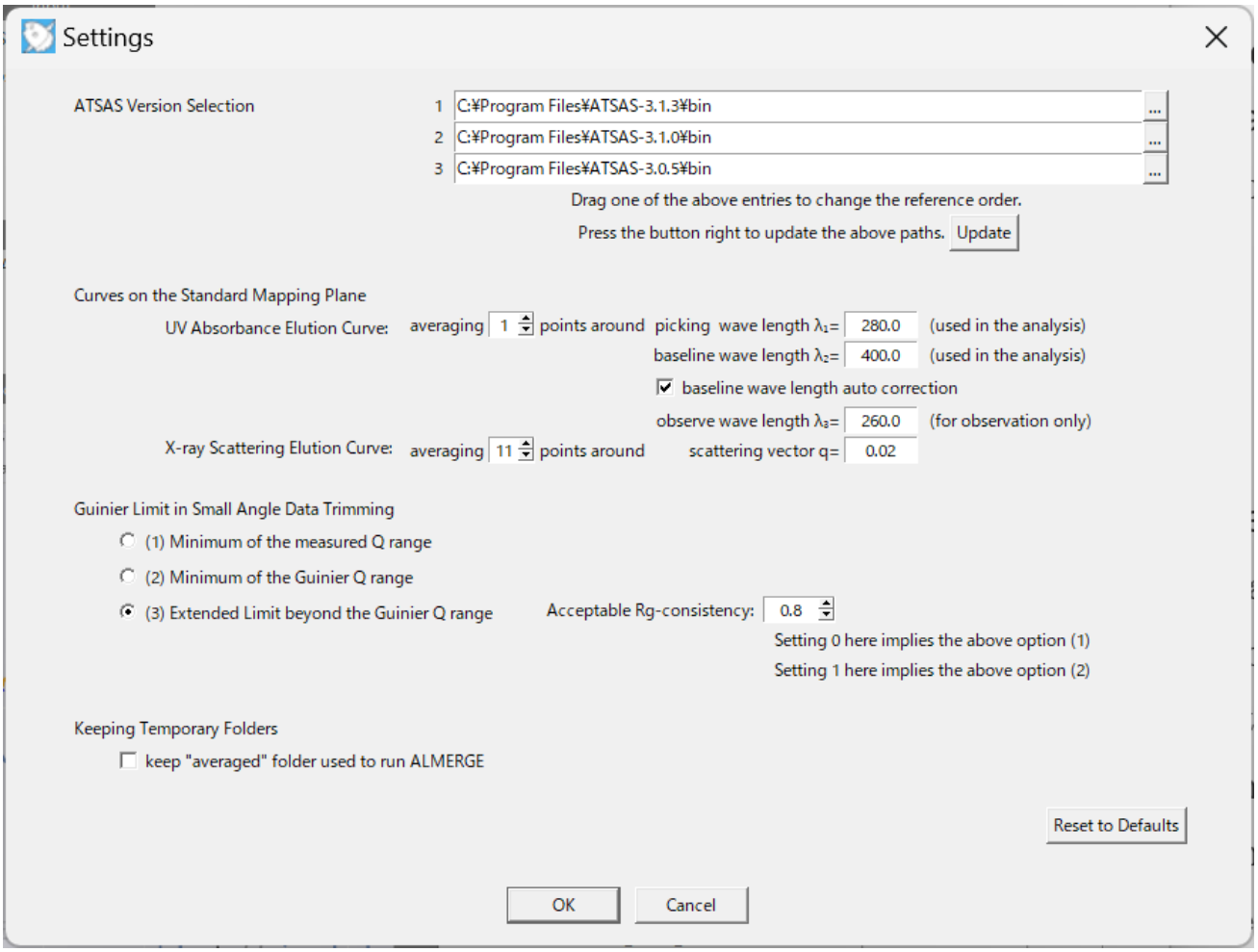
f Execution using GPUs will be enabled if the machine is equipped with NVIDIA compatible GPUs and you have CUDA Toolkit 10.2 installed.

7.7 Settings Dialog and Setting Menu

7.7.1 Purpose and Manipulation Flow

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

Fig. 7-67 Settings Dialog



7.7.2 ATSAS Version Selection

a At top is a list showing the location of ATSAS exe files, which are recognized and to be used by the program. By default, the latest version will be listed at top and be used.

b In order to change such default, drag a desired entry to the top of the list.

c In case there do not appear any expected entries, manually select your correct paths.

d This path list will not be updated automatically until the user press the “Update” button below the list manually.

7.7.3 Standard Mapping Plane (Elution Curve Picking Positions)

a You can change the positioning of the Standard Mapping Plane (elution curve picking

positions) by changing values for wave length (λ_1=280) or scattering vector length (q=0.02) or numbers of points for averaging.

b The second UV wave length value (λ_2=400) will be used to determine the baseline. This wavelength value is automatically changed properly unless the UV absorbance value there has vanished as in the case of common protein. You can disable this automatic change by unchecking the check button just below the value entry if you want it to be used as is.

c The third UV wave length value (λ_3=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. 7-4)

7.7.4 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 Rg value remains within the acceptable consistency specified. For example, suppose the Rg in the Guinier region be 50 and you set 0.8 (which is default) to the acceptable Rg-consistency, then, the limit angle (Q) is determined as small as possible while the estimated Rg[26] in a small interval beyond the Guinier region remains within a range around from 40 to 60.

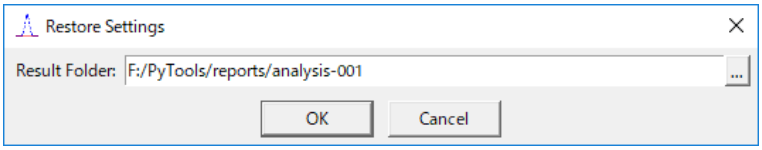
c Rg-consistency between two Rg's, say Rg_1 and Rg_2 based on Rg_2, is calculated as follows.

$$R_g \text{ consistency} = \exp\left(-\text{abs}\left(1 - \frac{Rg_1}{Rg_2}\right)\right).$$

7.7.5 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. 7-68 Restoring Settings from a Result Folder



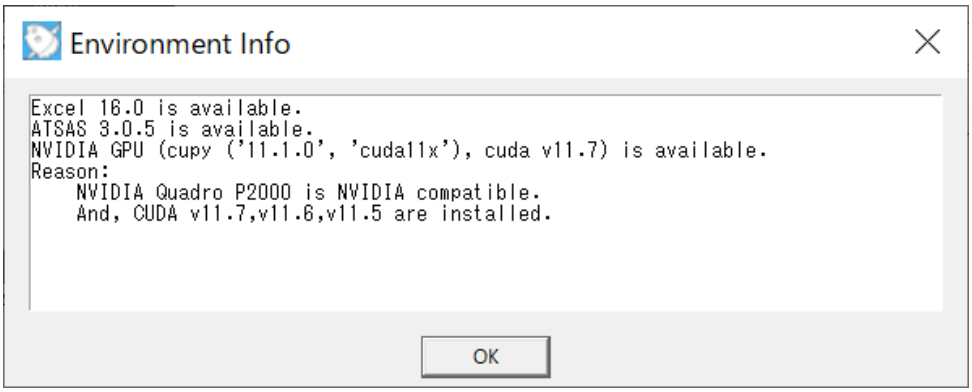
b Note that this is a risky operation since the settings parameters may change from version to version and this restoring process does not support any conversion.

[26] This Rg 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.

7.7.6 Showing Environment Information

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

Fig. 7-69 Showing Environment Information



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

7.8 SEC Tools

7.8.1 Elution Curve Picker

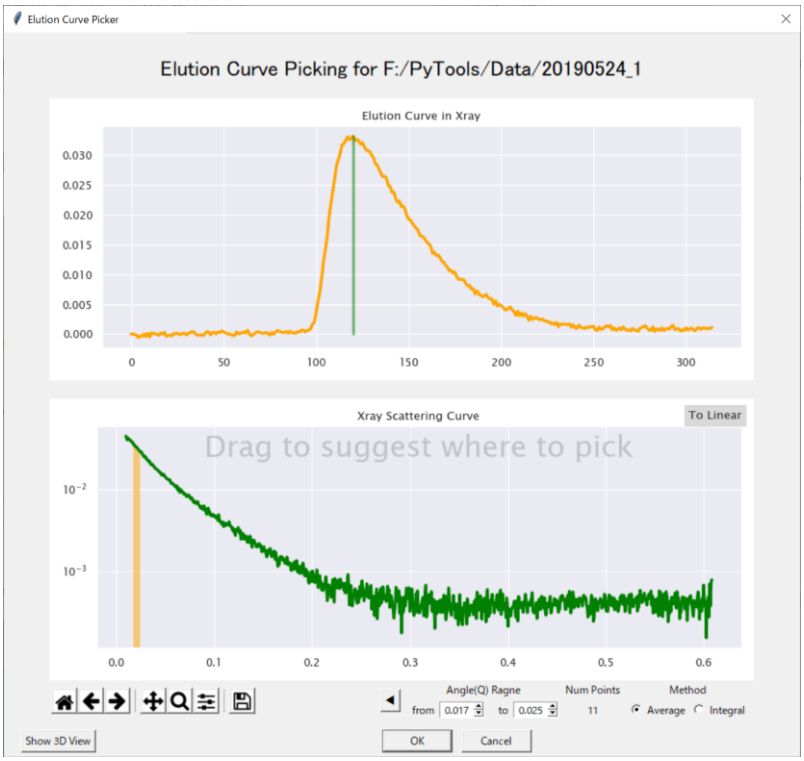
a By default, X-ray Data by averaging 11 curves near the point (Q=0.02) on the angular axis.

b While you can change these default parameters in “Settings Dialog”, you can do it visually by using “Elution Curve Picker” (Fig. 7-70) 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. 7-70 Elution Curve Picker



7.8.2 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. 7-71 Scattering Curve Plotter

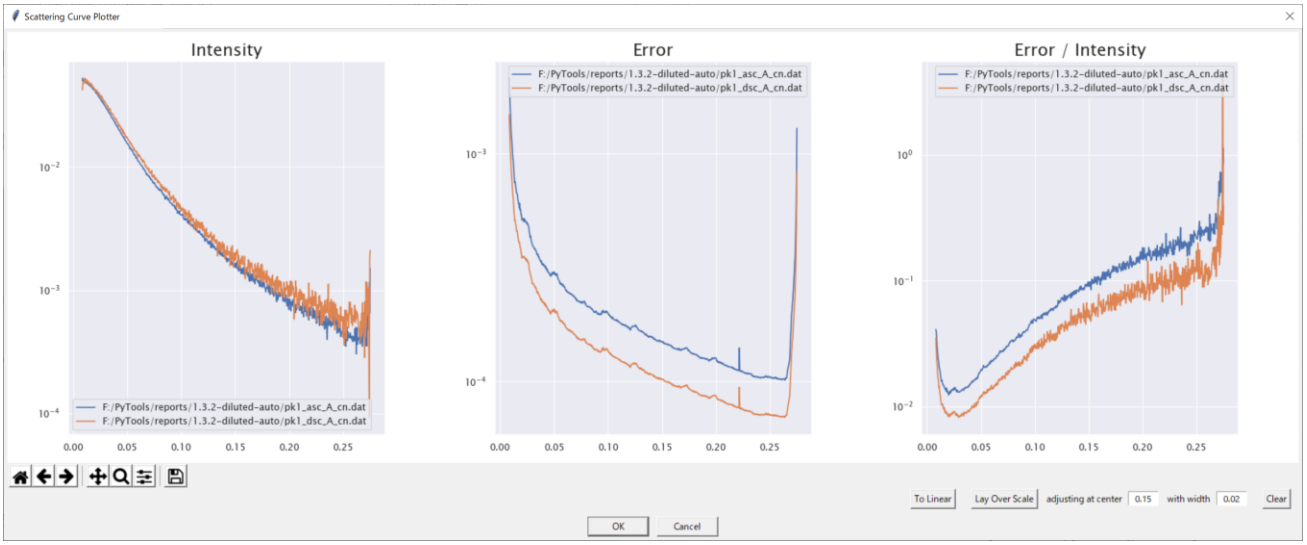
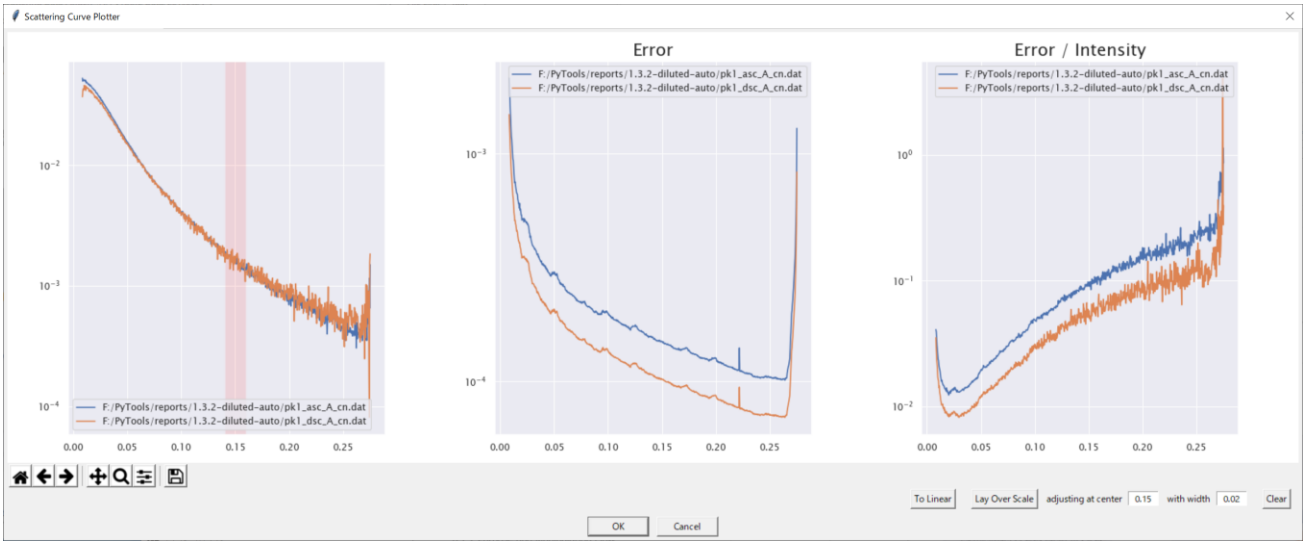


Fig. 7-72 Scattering Curve Plotter (after “Lay Over Scale” press)



7.8.3 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. 7-73 SVD results for Xray data

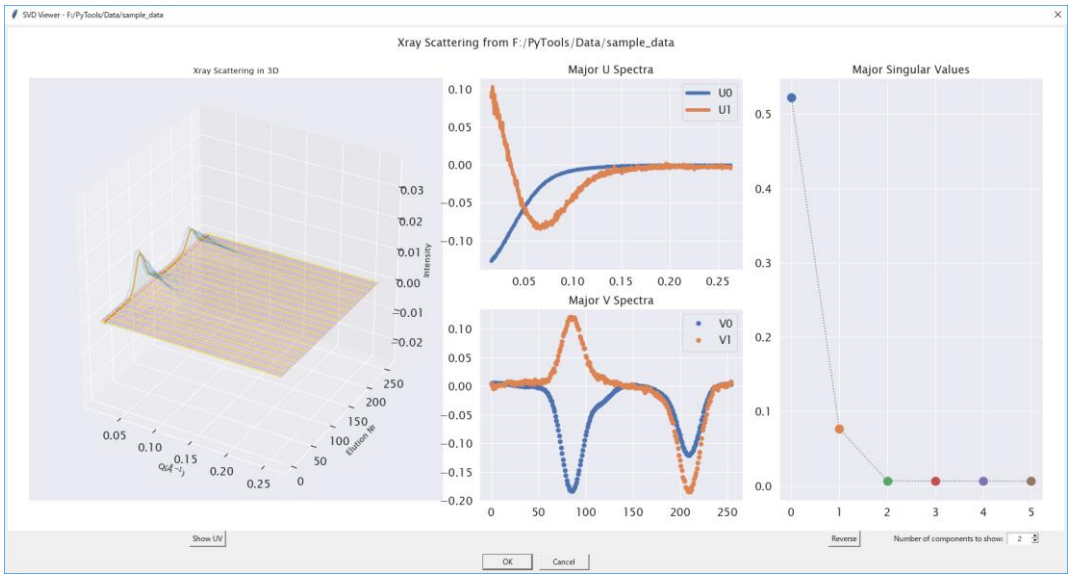
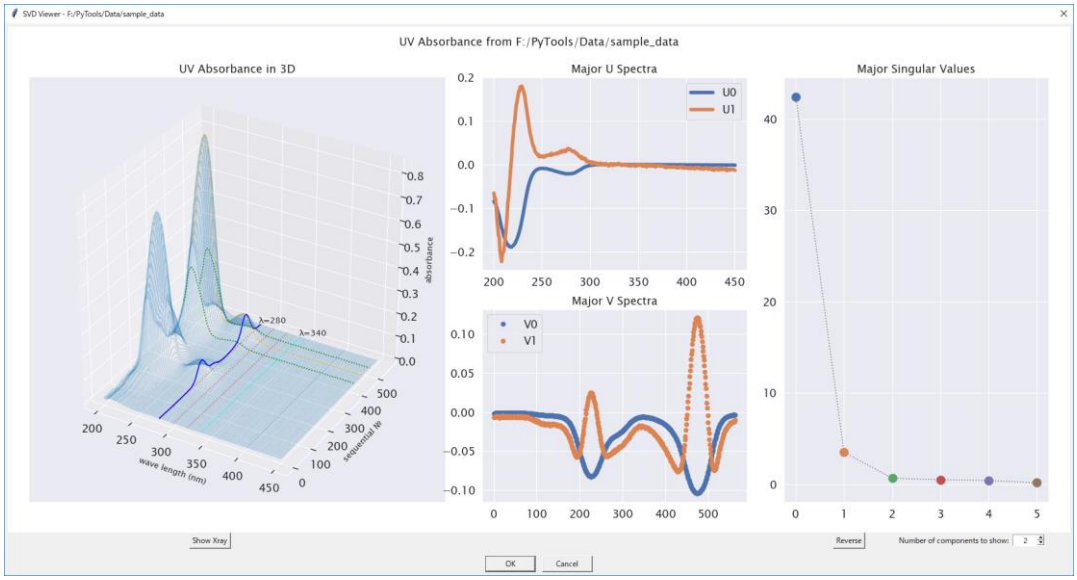


Fig. 7-74 SVD results for UV data



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.

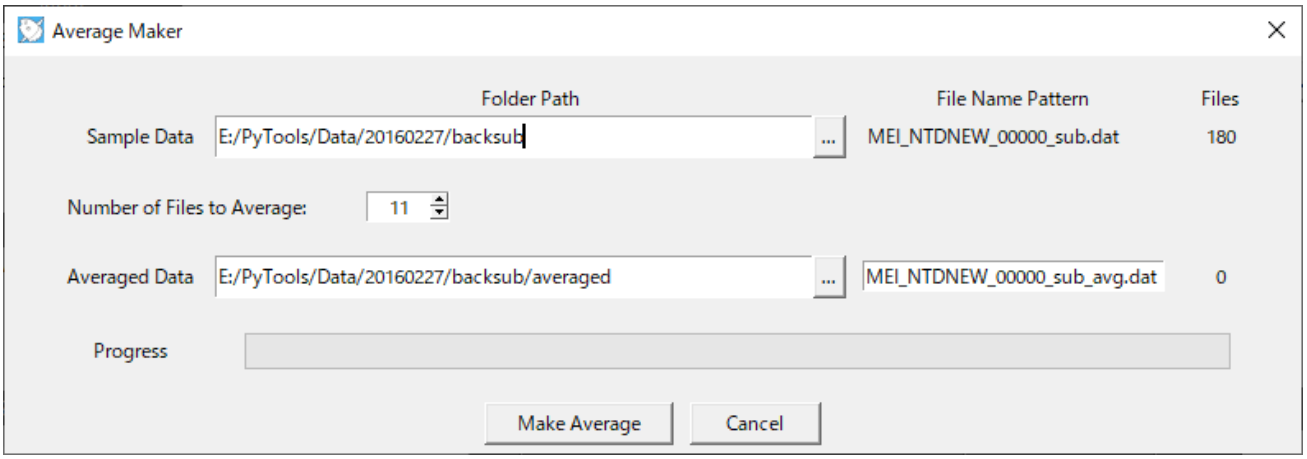
7.8.4 Average Maker

a In earliest versions of MOLASS (or Serial Analyzer), the program averaged input scattering files to get data with better S/N.

b For later versions of MOLASS, such facility has been abandoned in the main stream process, due to incompatibility with LRF.

c This tool provides the same old-days averaging facility by just conducting that single process.

Fig. 7-75 Average Maker



d For example, size 5 average for input file numbers 0, 1, 2, …, 99 is calculated according to the number correspondence suggested as follows.

Average file 0 ← [0, 1, 2]
Average file 1 ← [0, 1, 2, 3]
Average file 2 ← [0, 1, 2, 3, 4]
Average file 3 ← [1, 2, 3, 4, 5]
Average file 4 ← [2, 3, 4, 5, 6]
…
Average file 96 ← [94, 95, 96, 97, 98]
Average file 97 ← [95, 96, 97, 98, 99]
Average file 98 ← [96, 97, 98, 99]
Average file 99 ← [97, 98, 99]

e Each input file usually should contain the same number of scattering values with the same q-values, although lacks of q-values if any can be handled by adjusting to the common q-values (i.e., discarding lacking q-values).

7.9 DENSS Tools

7.9.1 DENSS GUI

a This GUI is almost identical to the one introduced in Section 7.6.3, where previewed curves could be passed to a local DENSS program.

b 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.

c 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. 7-76 DENSS GUI (initial state)

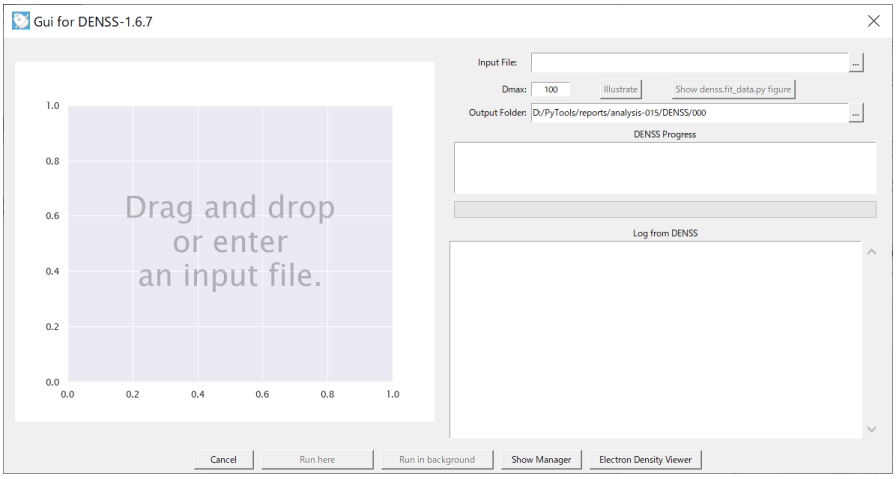
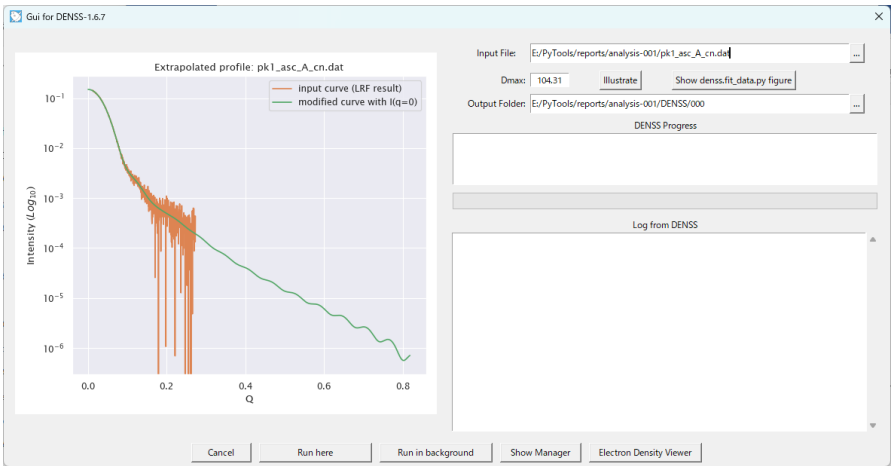


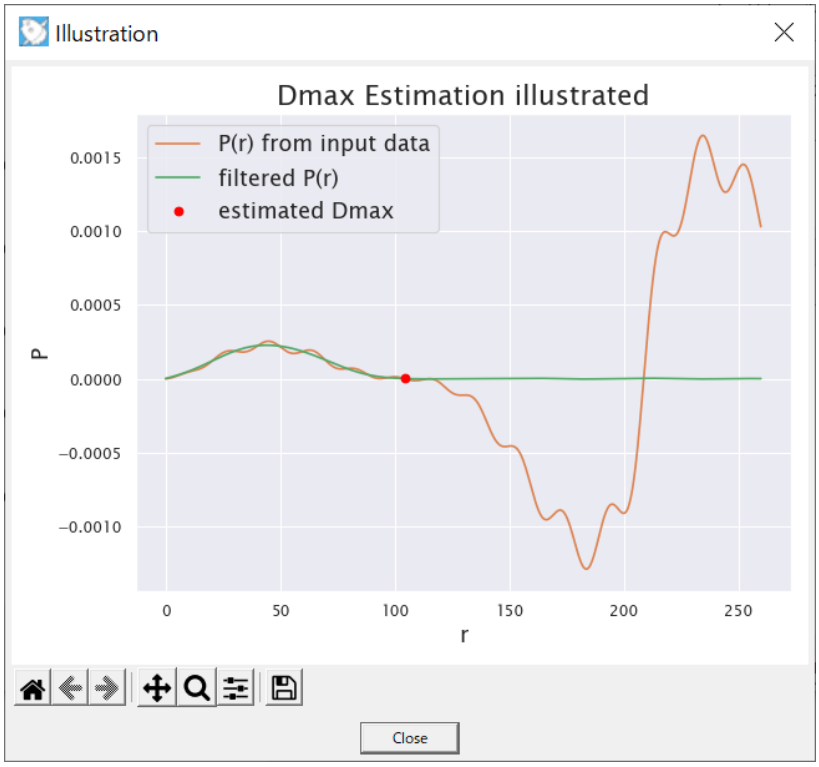
Fig. 7-77 DENSS GUI (when given a curve)



a After the input, the run buttons should be active, and users can run the job either here in the GUI or in background.

b The initial value of “Dmax” is automatically estimated by a DENSS program and set to the entry by this GUI.

Fig. 7-78　Dmax Estimation illustrated



c　You can change the estimated “dmax” after manually determining its proper value using either the dens.fitr_data.py from the button here (Fig. 7-79), or GNOM in ATSAS program suite. (See also DENSS tutorial)

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

Fig. 7-79　denss.fit_data.py figure

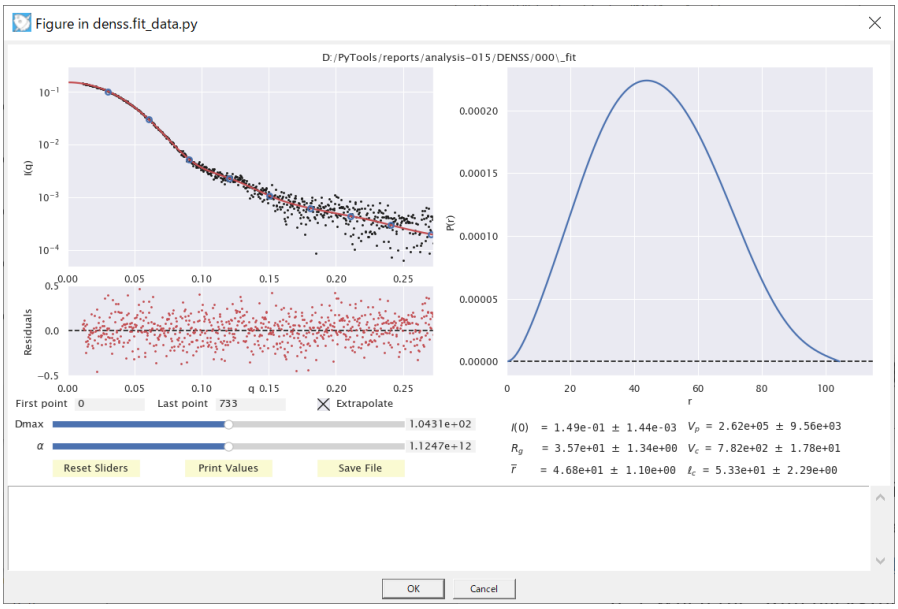
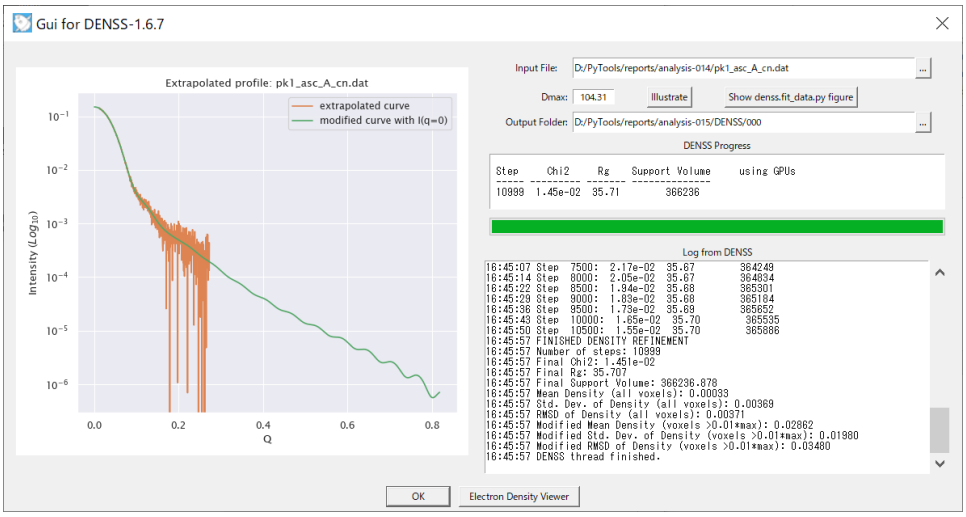


Fig. 7-80　DENSS GUI (when run here)



e　Note that the Rg value in the DENSS progress text is not correct when run using GPUs, while the final value in the log text is correct. It is a known inconvenience due to the current limitation of Cupy (7.8.0 as of this version) support for Scipy ndimage functions.

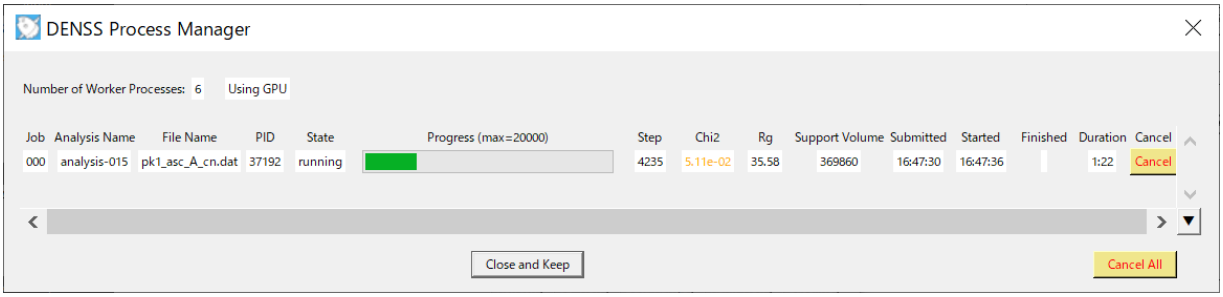
7.9.2　DENSS Manager

a　“DESS 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 7.6.6 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. 7-81.

Fig. 7-81　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.

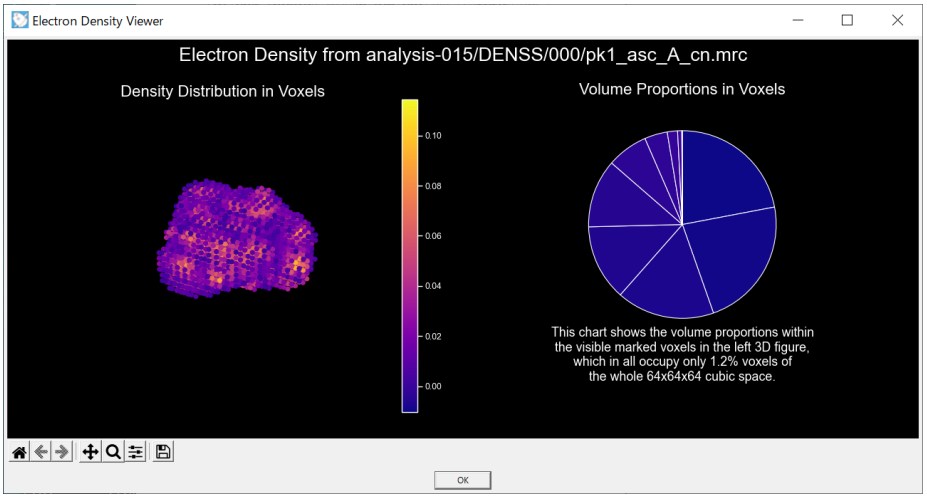
e Re-opening can be performed from DENSS Tools menu in the Main Dialog or the Extrapolation Preview.

f Execution using GPUs will be enabled if the machine is equipped with NVIDIA compatible GPUs and you have CUDA Toolkit 11.x installed. See Section 7.7.6 to verify GPU availability in your environment.

7.9.3 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. 7-82 Electron Density Viewer



7.9.4 SAXS Simulator

a Users can also use "SAXS Simulator" to verify the DENSS results as shown in Fig. 7-83.

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. 7-84 and Fig. 7-85.

Fig. 7-83 SAXS Simulator (comparing the simulated and experiment)

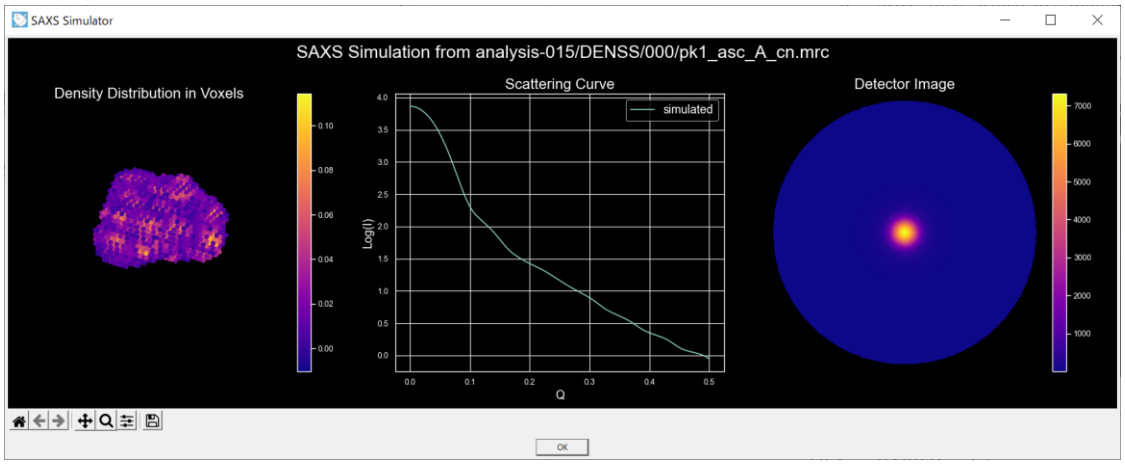


Fig. 7-84　SAXS Simulator (PDB query)

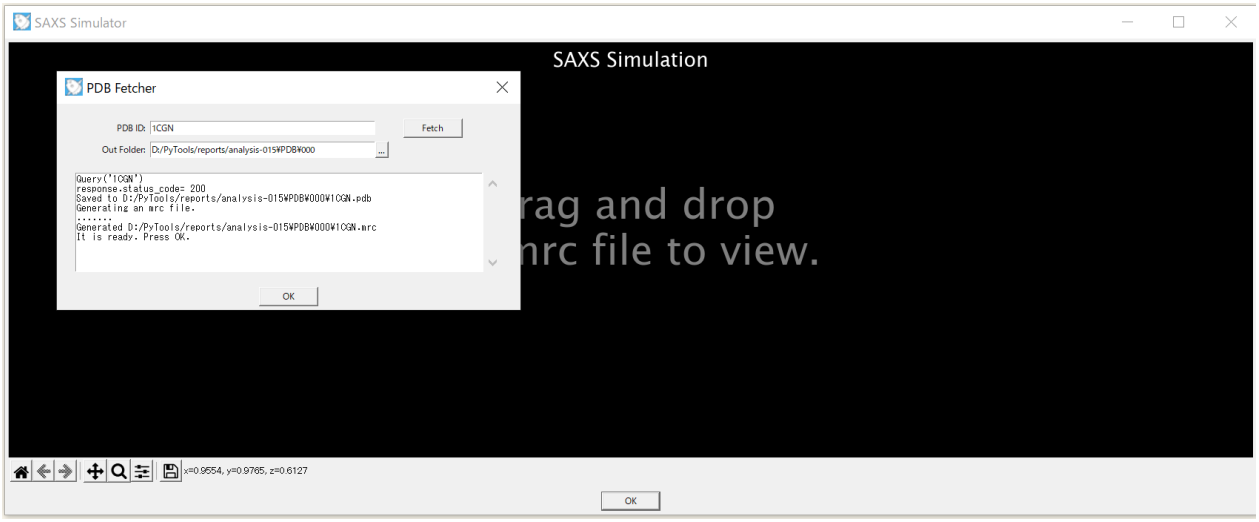
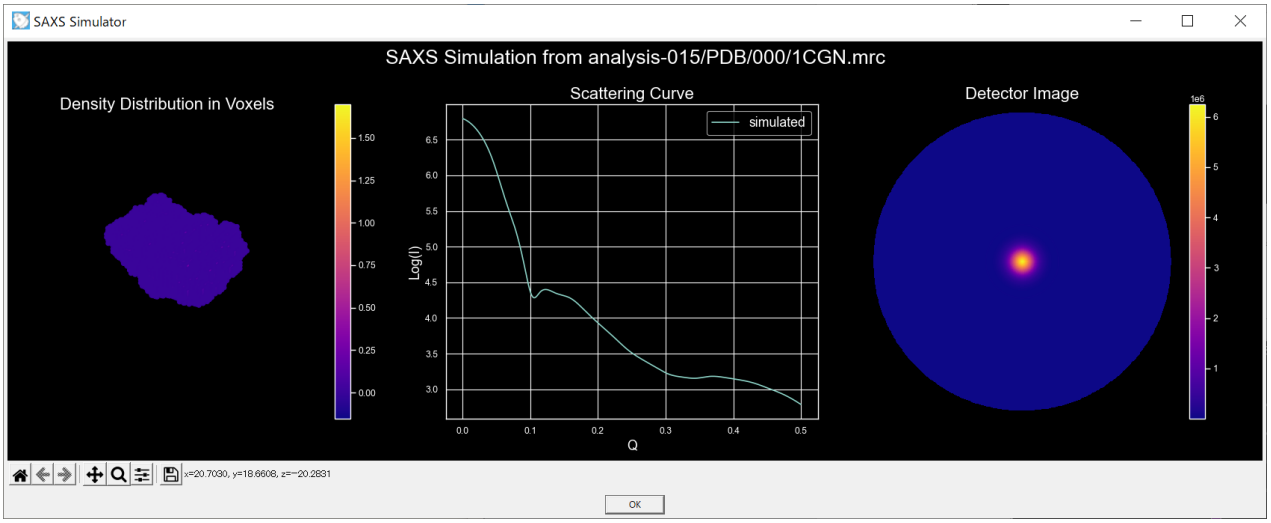


Fig. 7-85　SAXS Simulator (PDB query result)



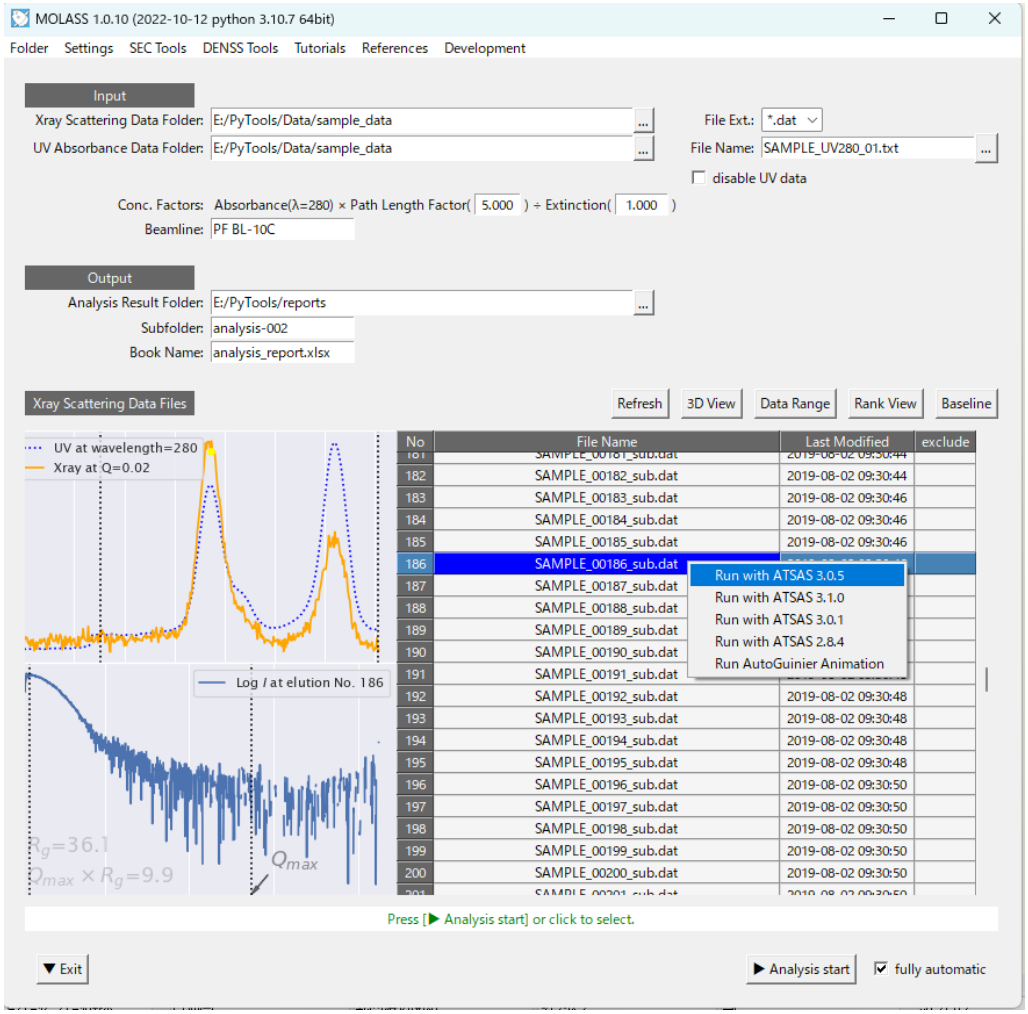
7.10 AutoGuinier Interface

a The main dialog (Fig. 7-86) can be used as a GUI to AutoGuinier program.

b For such a usage, double click a desired row to get the result in a dialog as shown in Fig. 7-87.

c 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 7.7. (Double clicking invokes the first option)

Fig. 7-86 Single Invocation of AutoGuinier Program



d The result dialog shows the AutoGuinier result and AUTORG result in a table and three figures as shown in Fig. 7-87.

e The figures at center and right are zoomed ones of the left figure.

f The table shows the estimated Rg, I(0) and related quantities and quality evaluation.

g From the popup menu, you can invoke an animation that explains the algorithm uses in AutoGuinier program (See Fig. 7-88).

h Scoring factors used to determine the appropriate Guinier region are listed in

Fig. 7-87　AutoGuinier Result Dialog

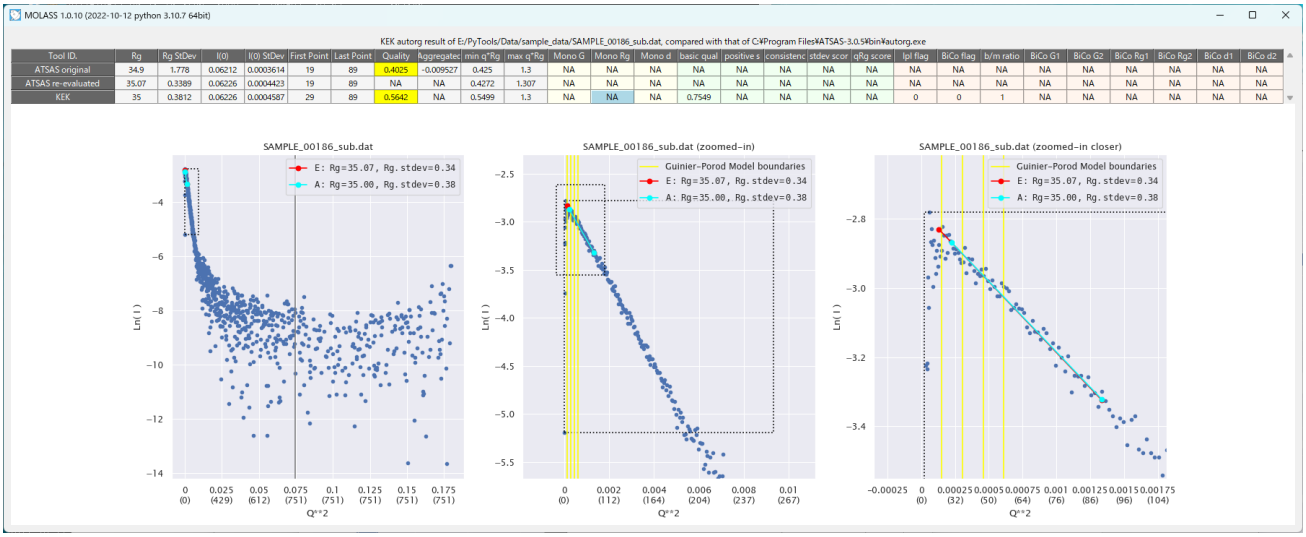
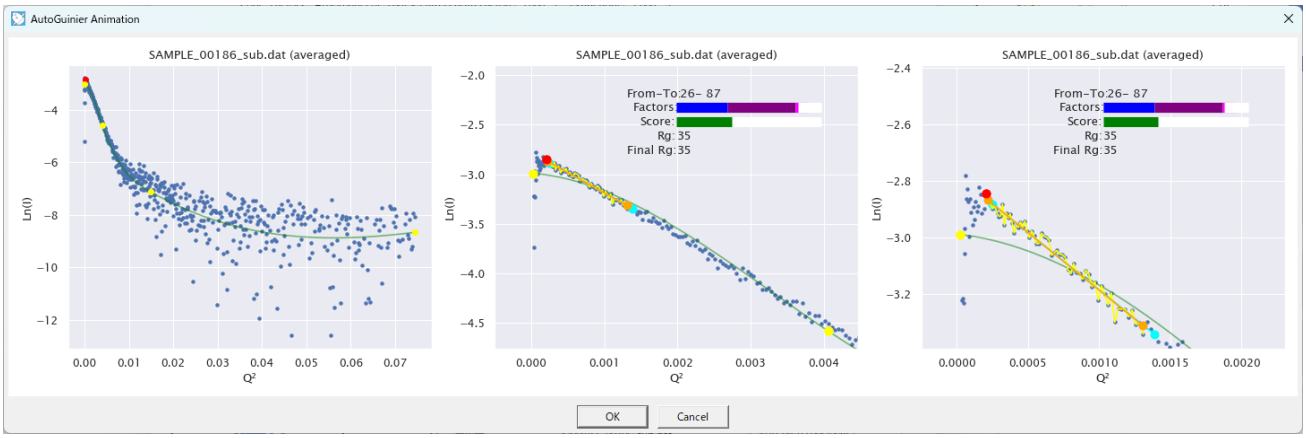


Fig. 7-88　AutoGuinier Animation



Tab. 7-10 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	$\mathrm{Rg\_Consistency(Spline\_Rg, Measured\_Rg)}$	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	$\exp(-(r\_value + 1) \times \mathrm{LINEARITY\_SCALE})$	Computed from the r_value of linear regression with. $\mathrm{LINEARITY\_SCALE} = 5$
3	end consistency	0 – 1	0.06	0.06	0.0	0.0	$\mathrm{Rg\_Consistency(SA\_End\_Rg, LA\_End\_Rg)}$	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	$\mathrm{Rg\_Consistency(Extended\_SA\_End, SA\_End)}$	Computed from the Rg values at extended small angle side end and the small angle side end.

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

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

8 X-ray Data only Mode

8.1 Design Philosophy

a MOLASS 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).

8.2 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.

8.3 Different Manipulation Points from the Normal Usage

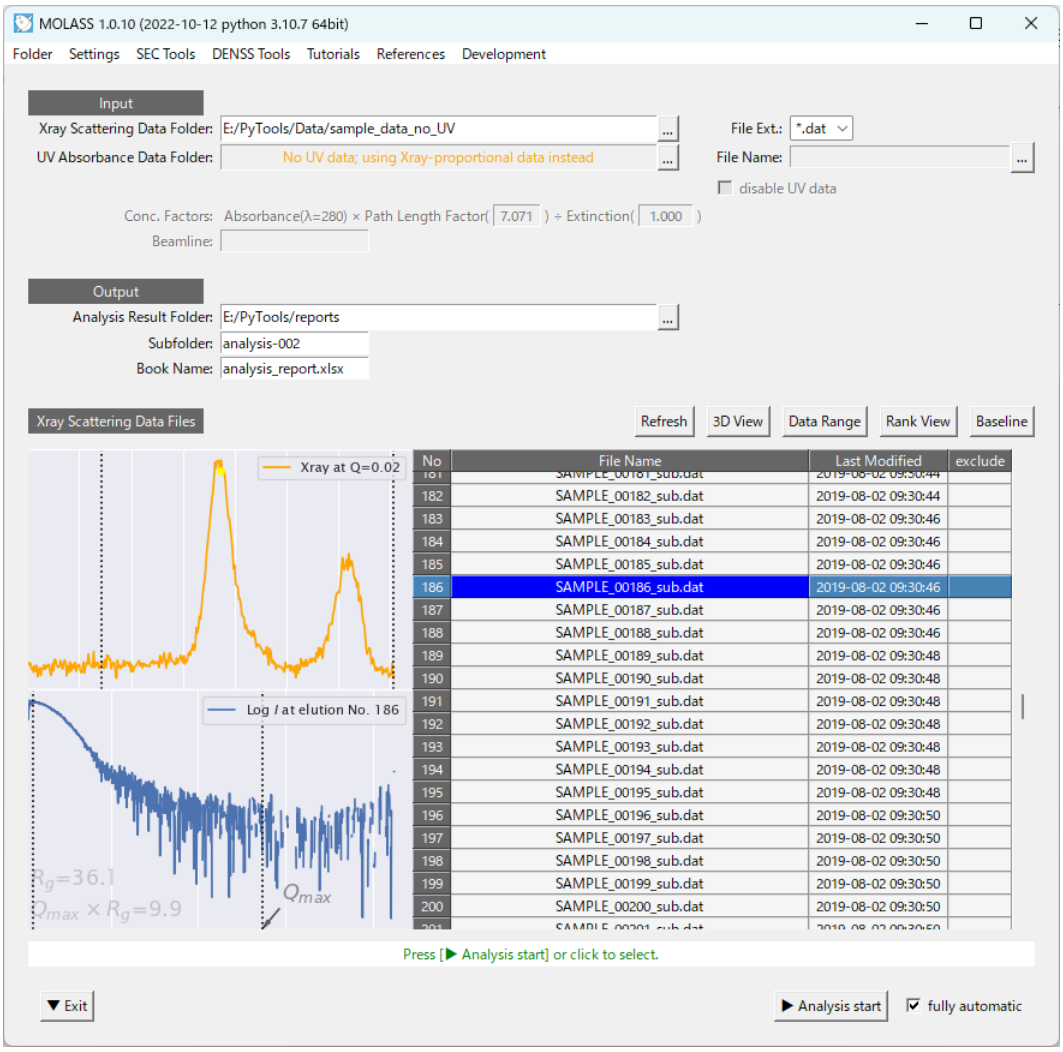
8.3.1 Main Dialog

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. 8-1.

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 7.1.7)

c Usage of “Restriction” tool is limited to the X-ray data.

Fig. 8-1　Main Dialog in X-ray Data only Mode



8.3.2　Mapping Dialog

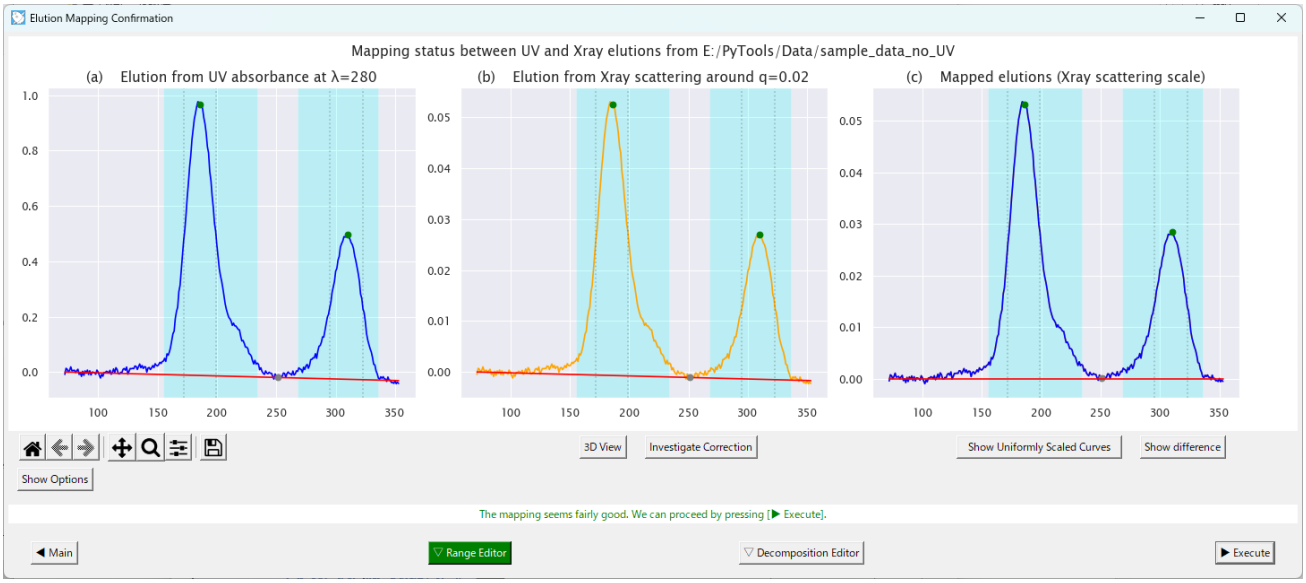
a　Mapping is not necessary when you are using X-ray data only.

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. 8-2.

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. 8-2　Mapping Dialog in the Case of X-ray only Data



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.

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

$$\text{Stdev(Rg)} \fallingdotseq \sqrt{(50-55)^2 \cdot \tfrac{1}{2} + (80-55)^2 \cdot \tfrac{1}{4} + (40-55)^2 \cdot 1/4} = 15 \qquad \text{(A-2)}$$

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.

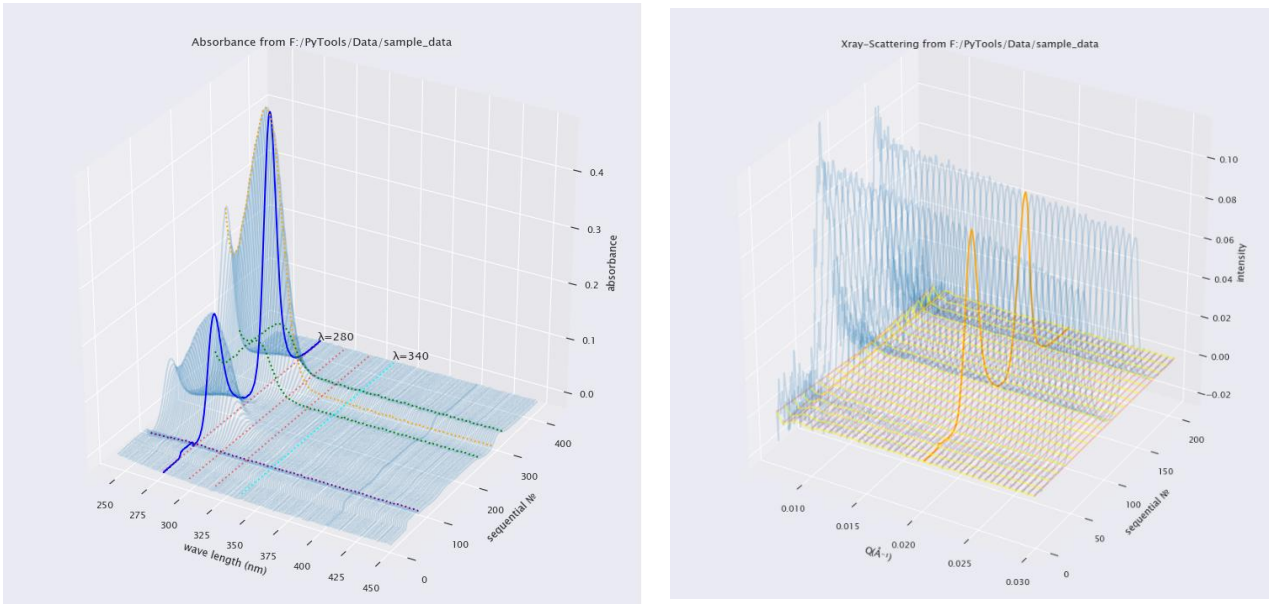
1. Errors computed like (A-2) indicate a different type of variations than Gaussian noises.
2. The distributions for adoption probability in real world are not so simple as this example, and difficult to model, implement and understand.
3. It is better to separate different kind of errors, namely Gaussian 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 Terminology Notes

B.1 Standard Mapping Plane

a Both UV absorbance and X-ray scattering data from a serial experiment have similar 3D structures as seen in Fig. B.1-1, 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. B.1-1 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 λ =280nm for UV absorbance and the scattering vector length Q=0.02 for X-ray scattering.

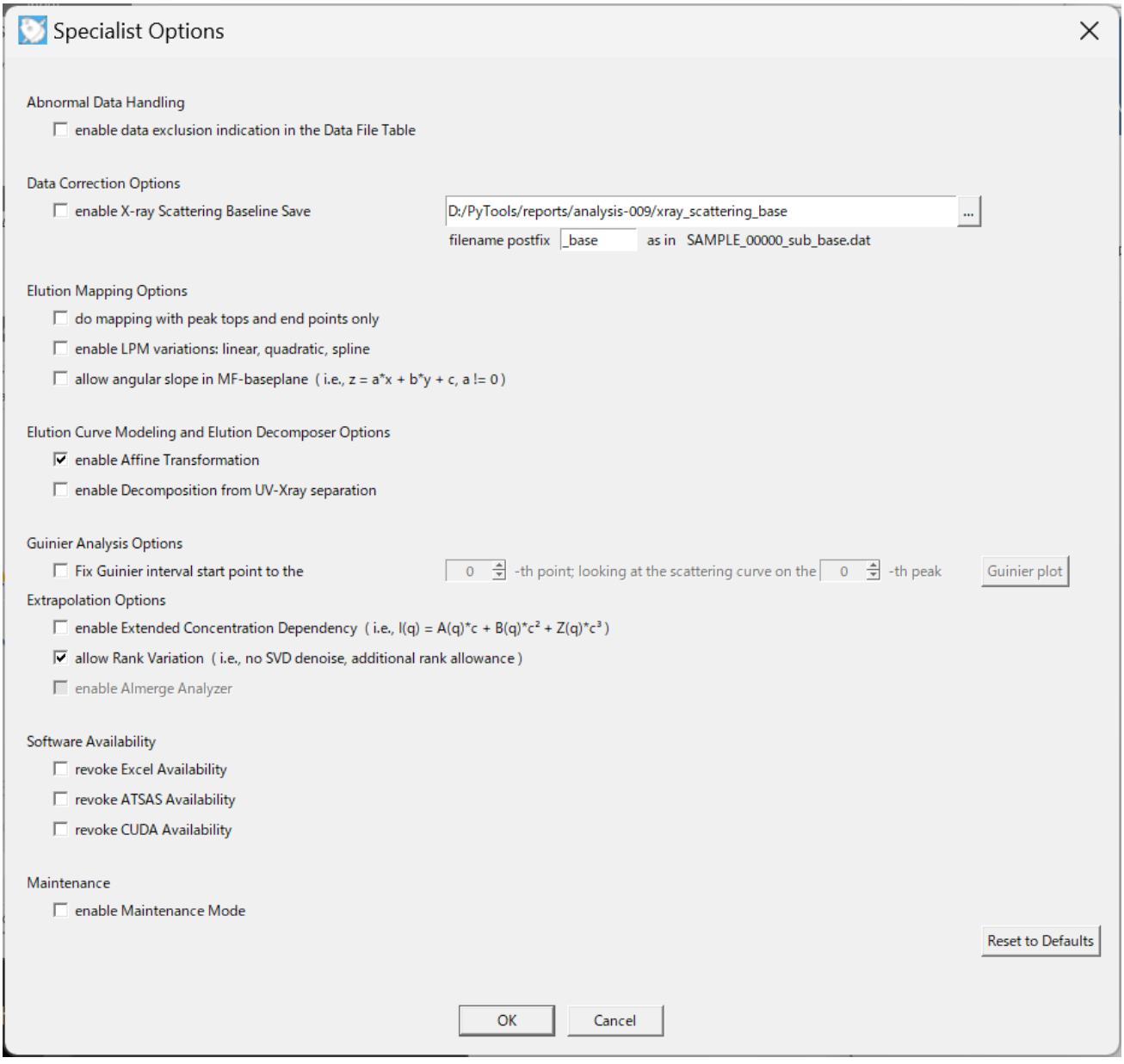
g In later descriptions, unless otherwise stated, elution curves in this standard plane are considered.

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



C.1.1 Absorbance Data Treatment

C.1.2 Data Correction Options

C.1.3 Elution Mapping Options

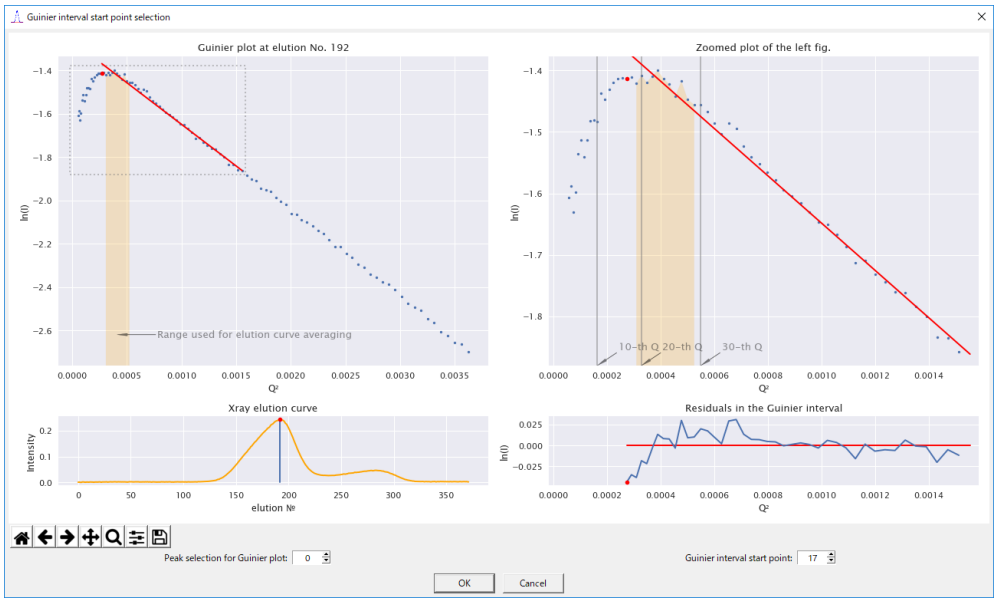
C.1.4 Elution Curve Modeling

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

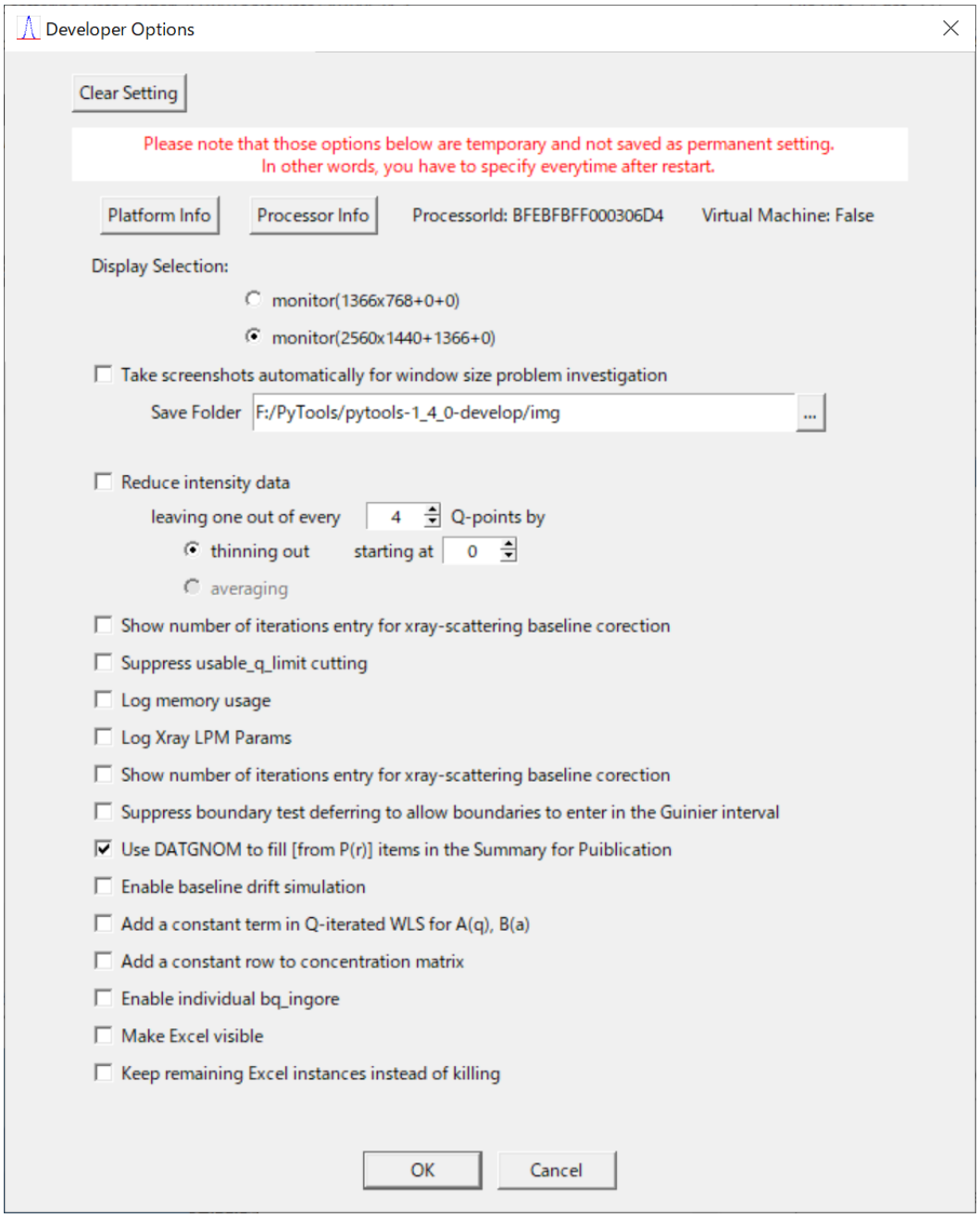
C.1.7 Deprecated Extrapolation Options

C.1.8 Maintenance Mode (for Guinier Analysis)

D Developer Tools

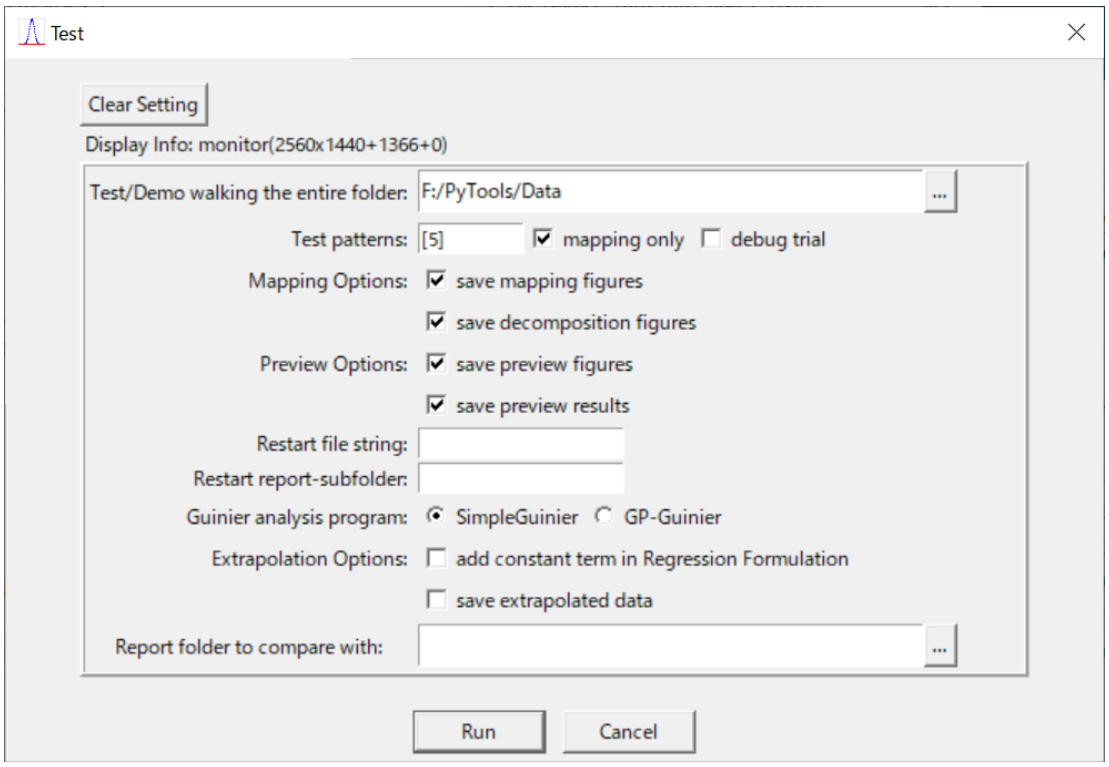
D.1 Developer Options Dialog

Fig. D.1-1 Developer Options Dialog



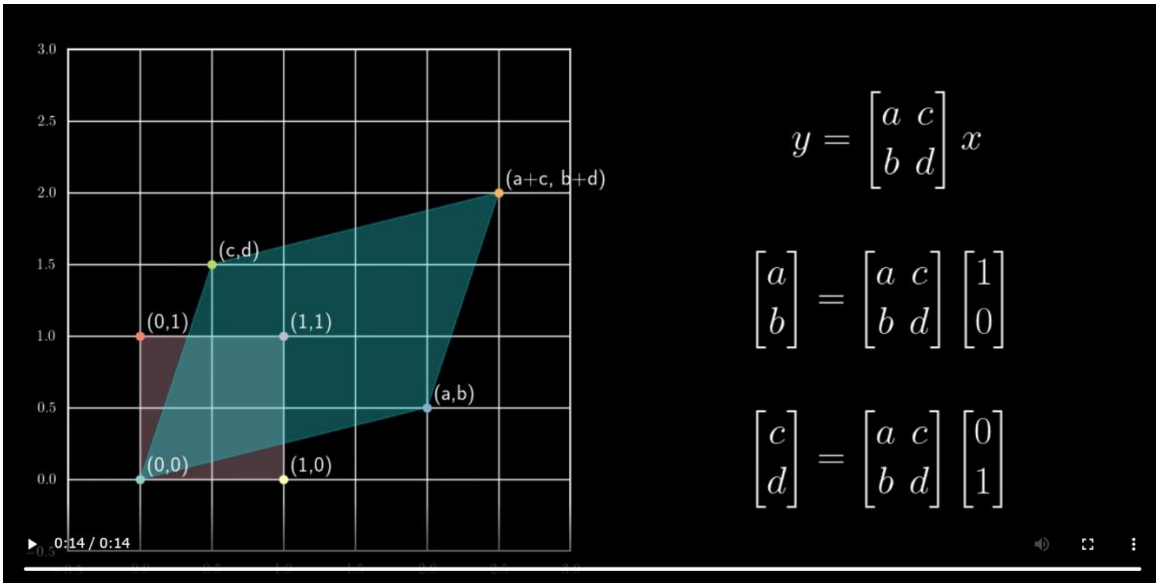
D.2 Tester Dialog

Fig. D.2-1 Developer Options Dialog

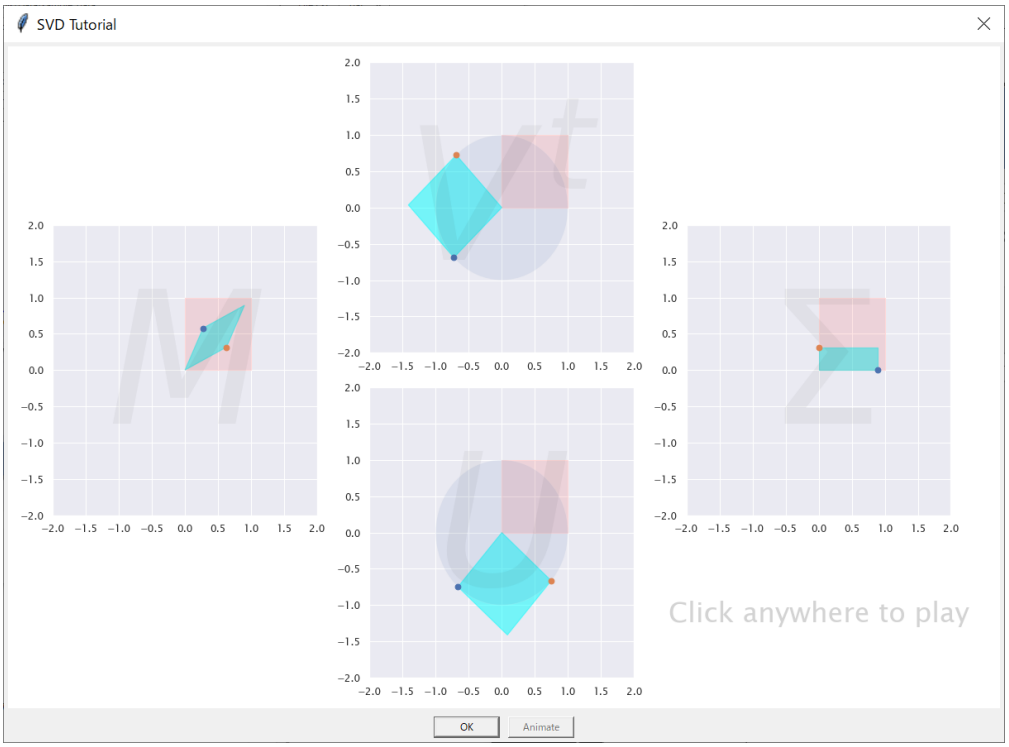


E Tutorials

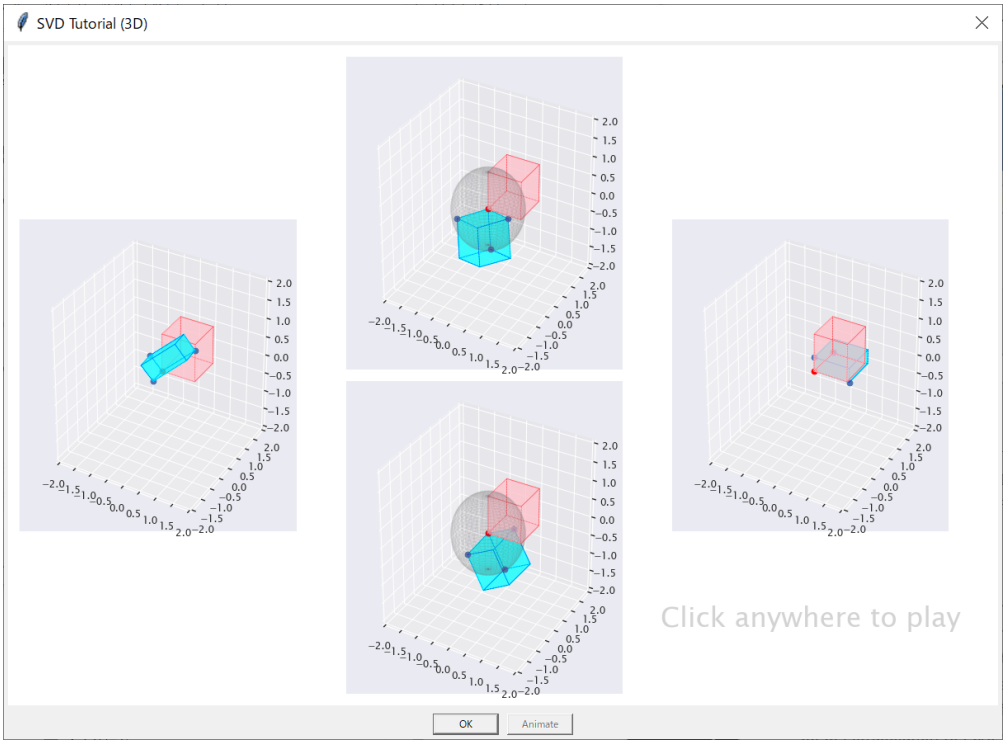
E.1 Linear Transformation



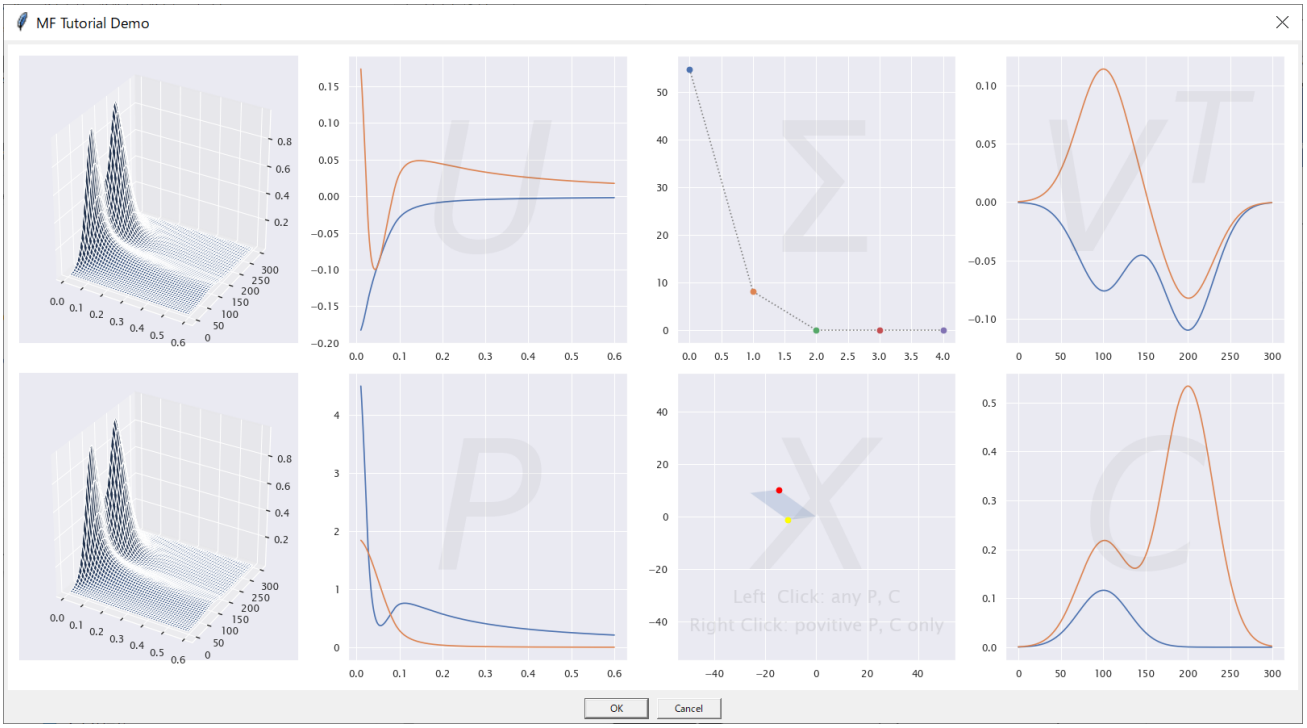
E.2 SVD (2D)



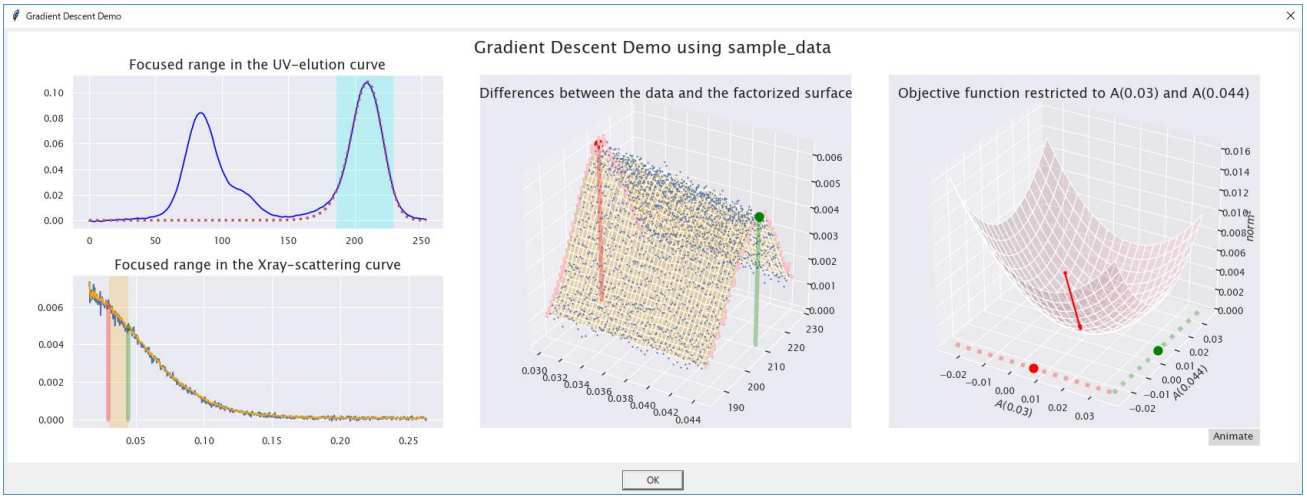
E.3 SVD (3D)



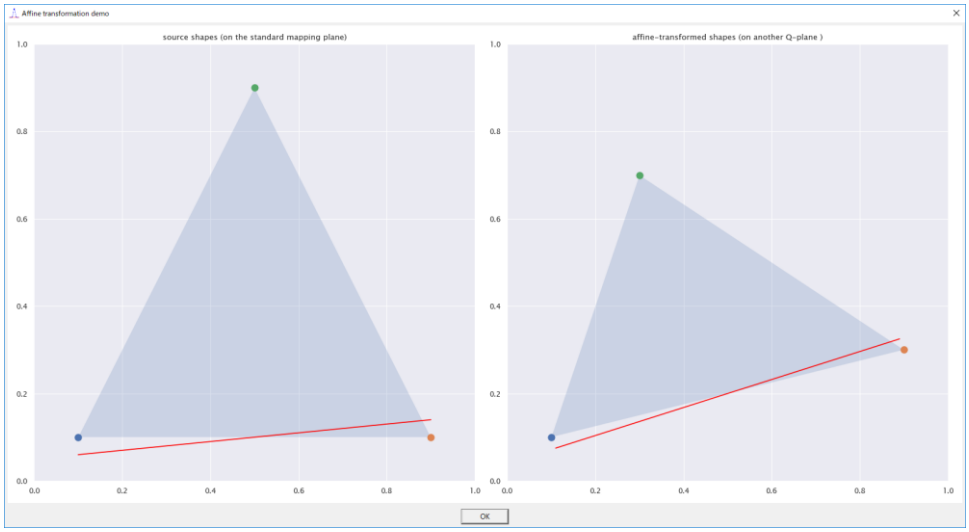
E.4 Matrix Factorization



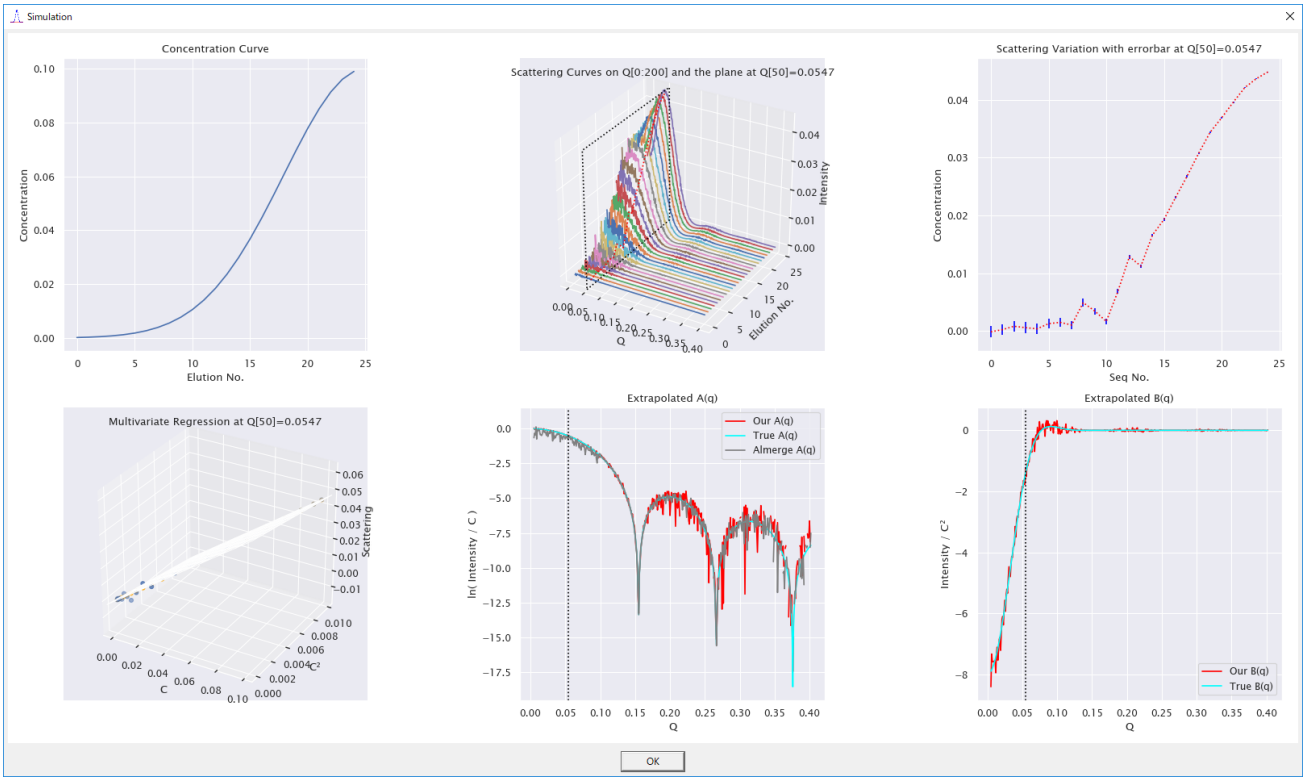
E.5 Conjugate Gradient Algorithm



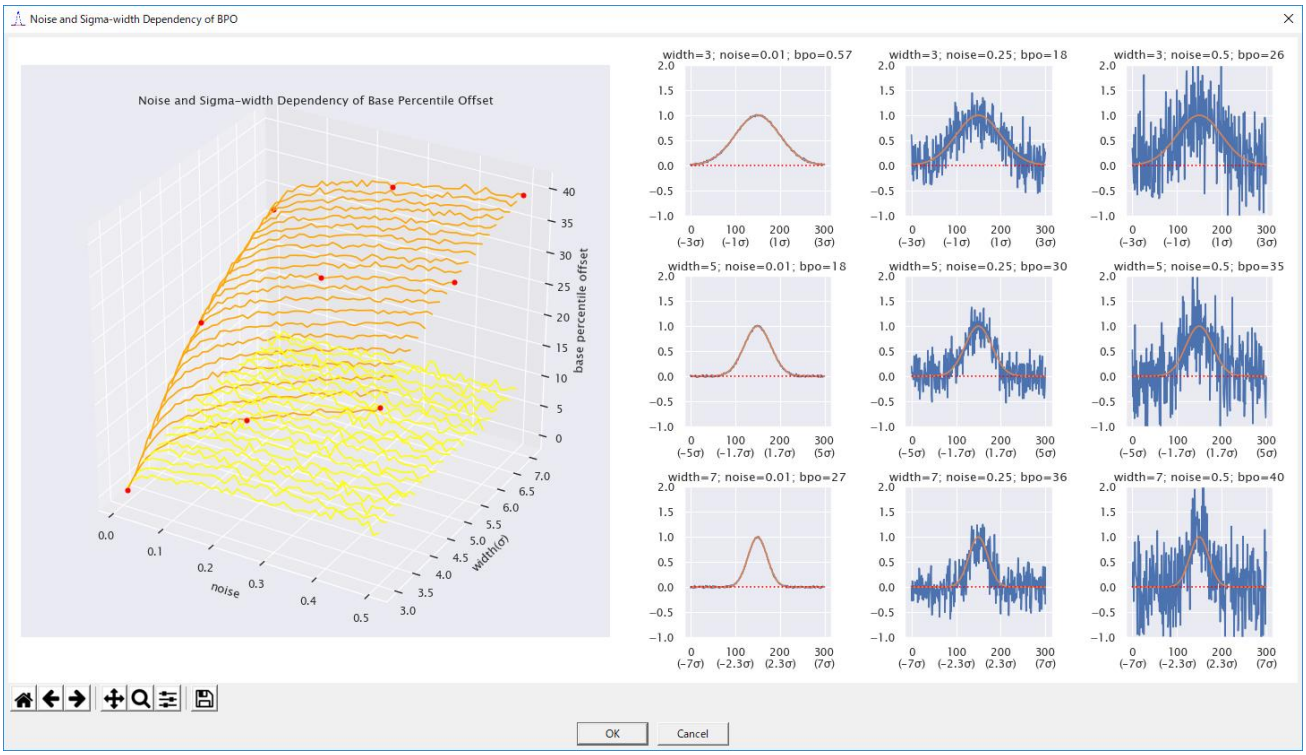
E.6 Affine Transformation of Baseline Adjustment



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 Trewhella (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