Journal of Visualized Experiments

Analysis of SEC-SAXS data via EFA deconvolution and Scatter

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE61578R1
Full Title:	Analysis of SEC-SAXS data via EFA deconvolution and Scatter
Section/Category:	JoVE Biology
Keywords:	Bio-Small Angle X-ray Scattering (BioSAXS); inline size-exclusion chromatography (SEC); background subtraction; E9; vaccinia; Scatter; deconvolution of SEC-SAXS data; BioXTAS RAW; SAXS; EFA
Corresponding Author:	Stephanie Hutin, PhD Universite Grenoble Alpes Grenoble, Isere FRANCE
Corresponding Author's Institution:	Universite Grenoble Alpes
Corresponding Author E-Mail:	stephaniehutin82@gmail.com
Order of Authors:	Mark D Tully
	Nicolas Tarbouriech
	Robert P Rambo
	Stephanie Hutin, PhD
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Open Access (US$4,200)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	European Synchrotron Radiation Facility, Grenoble, France

Stephanie Hutin, PhD
Laboratoire de Physiologie Cellulaire and Végétale
Université Grenoble Alpes/CNRS/CEA/INRA/BIG
Grenoble, France
stephanie.hutin@ univ-grenoble-alpes.fr

Benjamin Werth
Sr. Science Editor - Chemistry | Biochemistry
JoVE

Dear Dr. Benjamin Werth,

Thank you very much for the invitation to publish in JoVE. In 2016/2017/2018/2019 we have worked together on the manuscripts “Online Size-exclusion and Ion-exchange Chromatography on a SAXS Beamline”, “Structure Solution of the Fluorescent Protein Cerulean Using MeshAndCollect” and “Fully Autonomous Characterization and Data Collection from Crystals of Biological Macromolecules” and due to our good experiences with JoVE we would be happy to publishing in your journal, again.

We would like to propose a manuscript describing a project of high interest to the structural biology and general biology community, using top notch innovative new techniques. We describe the data collection and analysis of difficult samples using Bio-Small Angle X-ray Scattering using a combination of deconvolution of the curves and data analysis. We are using the program/interface Scatter, which is commonly used by SAXS specialist and non-specialists to analyse and interpret their SAXS data. Until now, there is only an online tutorial but no publication about the program nor how to use or cite it in the literature. We would like to change this and would like to submit to you our manuscript “Analysis of SEC-SAXS data via deconvolution”.

I am looking forward to hearing from you at your earliest convenience

Stephanie Hutin

TITLE:
Analysis of SEC-SAXS data via EFA deconvolution and Scatter

AUTHORS AND AFFILIATIONS:
Mark D. Tully[1,*], Nicolas Tarbouriech[2], Robert P. Rambo[3] and Stephanie Hutin[4,*]

[1]European Synchrotron Radiation Facility, Structural Biology Group, Grenoble, France
[2]Institut de Biologie Structurale, University Grenoble Alpes, CEA, CNRS, Grenoble, France
[3]Diamond Light Source, Oxford, UK
[4]Laboratoire de Physiologie Cellulaire and Végétale, Université Grenoble Alpes/CNRS/CEA/INRA/BIG, Grenoble, France

*These authors contributed equally to this work

Corresponding Author:
Stephanie Hutin (stephanie.hutin@univ-grenoble-alpes.fr)

Email Addresses of Co-authors:
Mark D. Tully (mark.tully@esrf.fr)
Nicolas Tarbouriech (nicolas.tarbouriech@ibs.fr)
Robert P Rambo (robert.rambo@diamond.ac.uk)
Stephanie Hutin (stephanie.hutin@univ-grenoble-alpes.fr)

KEYWORDS:
Bio-Small Angle X-ray Scattering, BioSAXS, inline size-exclusion chromatography, SEC, background subtraction, E9, vaccinia, Scatter, deconvolution of SEC-SAXS data, BioXTAS RAW.

Inline size-exclusion chromatography coupled with biological small-angle x-ray scattering (SEC-BioSAX)

SUMMARY:
SEC-BioSAXS measurements of biological macromolecules are a standard approach for determining solution structure of macromolecules and their complexes. Here, we analyze SEC-BioSAXS data from two types of commonly encountered SEC traces—chromatograms with fully resolved and partially resolved peaks. We demonstrate the analysis and deconvolution using scatter and BioXTAS RAW.

ABSTRACT:
BioSAXS is a popular technique used in molecular and structural biology to determine the solution structure, particle size and shape, surface-to-volume ratio and conformational changes of macromolecules and macromolecular complexes. A high quality SAXS dataset for structural modeling must be from monodisperse, homogeneous samples and this is often only reached by a combination of inline chromatography and immediate SAXS measurement. Most commonly, size-exclusion chromatography is used to separate samples and exclude contaminants and

aggregations from the particle of interest allowing SAXS measurements to be made from a well-resolved chromatographic peak of a single protein species. Still, in some cases, even inline purification is not a guarantee of monodisperse samples, either because multiple components are too close to each other in size or changes in shape induced through binding alter perceived elution time. In these cases, it may be possible to deconvolute the SAXS data of a mixture to obtain the idealized SAXS curves of individual components. Here, we show how this is achieved and the practical analysis of SEC-SAXS data is performed on ideal and difficult samples. Specifically, we show the SEC-SAXS analysis of the vaccinia E9 polymerase minus exonuclease mutant.

INTRODUCTION:

Biological macromolecules are too small to be seen even with the best light microscopes. Current methods to determine their structures generally involve crystallizing the protein or measurements on vast numbers of identical molecules at the same time. While crystallography provides information on the atomic level, it represents an artificial sample environment, given that most macromolecules are not presented in a crystalline form in the cell. During the last couple of years cryo-electron microscopy delivered similar high-resolution structures of large macromolecules / macromolecular complexes, but although the samples are closer to physiological condition, they are still frozen, hence immobile and static. Bio-small angle X-ray scattering (BioSAXS) provides a structural measurement of the macromolecule, in conditions that are relevant to biology. This state can be visualized as a low resolution 3-D shape determined on nanometer scale and captures the entire conformational space of the macromolecule in solution. BioSAXS experiments efficiently assess oligomeric state, domain and complex arrangements as well as flexibility between domains[1–3]. The method is accurate, mostly non-destructive and usually requires only a minimum of sample preparation and time. However, for the best interpretation of the data, the samples need to be monodisperse. This is challenging; biological molecules are often susceptible to contaminations, poor purification and aggregation, for example from freeze thawing[4]. The development of inline chromatography followed by immediate SAXS measurement helps mitigate these effects. Size-exclusion chromatography separates the samples by size thus excluding most contaminants and aggregations[5–10]. However, in some cases even SEC-SAXS is not sufficient to produce a monodisperse sample, because the mixture may consist of components that are too close in size or their physical properties or their fast dynamics lead to overlapping peaks in the SEC UV trace. In these cases, a software-based deconvolution step of the obtained SAXS data might lead to an idealized SAXS curve of the individual component[5, 11, 12]. As an example, in protocol section 2, we show the standard SEC-SAXS analysis of the vaccinia E9 polymerase minus exonuclease mutant (E9 exo^{minus}) in complex with DNA. Vaccinia represents the model organism of the Poxviridae, a family containing several pathogens, for example the human smallpox virus. The polymerase was shown to bind tightly to DNA in biochemical approaches, with the structure of the complex recently solved by X-ray crystallography[13].

Most synchrotron facilities will provide an automated data processing pipeline that will perform data normalization and integration producing a set of unsubtracted frames. But the approach described in this manuscript could also be use with a lab source provided SEC-SAXS is performed.

Furthermore, additional automation may be available that will reject radiation-damaged frames and perform the buffer subtraction[14]. We will show how to perform primary data analysis on pre-processed data and make the most of the available data in section 2.

In section 3, we show how to deconvolute SEC-SAXS data and analyze the curves efficiently. While there are several deconvolution methods such as the Gaussian peak deconvolution, implemented in US-SOMO [15] and the Guinier optimized maximum likelihood method, implemented in the DELA software[16], these generally require a model for the peak shape[12]. The finite size of individual peaks we are investigating allows the use of evolving factor analysis (EFA), as an enhanced form of singular value decomposition (SVD) to deconvolute overlapping peaks, without relying on the peak shape or scattering profile[5, 11]. A SAXS-specific implementation can be found in BioXTAS RAW[17]. EFA was first used on chromatography data when 2D diode array data allowed matrices to be formed from absorbance against retention time and wavelength data[18]. Where EFA excels is that it focuses on the evolving character of singular values, how they change with the appearance of new components, with the caveat that there is an inherent order in the acquisition[10]. Fortunately, SEC-SAXS data provides all the necessary ordered acquisition data in organized 2D data arrays, lending itself nicely to the EFA technique.

In section 4, we will demonstrate the basics of model-independent SAXS analysis from the buffer-background subtracted SAXS curve. Model-independent analysis determines the particle's radius-of-gyration (Rg), volume-of-correlation (Vc), Porod Volume (Vp), and Porod-Debye Exponent (PE). The analysis provides a semi-quantitative assessment of the particle's thermodynamic state in terms of compactness or flexibility via the dimensionless Kratky plot[2, 4, 19].

Finally, SAXS data are measured in reciprocal space units and we will show how to transform the SAXS data to real-space to recover the pair-distance, P(r), distribution function. The P(r)-distribution is the set of all distances found within the particle and includes the particle's maximum dimension, d_{max}. Since this is a thermodynamic measurement, the P(r)-distribution represents the physical space occupied by the particles' conformational space. Proper analysis of a SAXS dataset can provide solution-state insights that complement high-resolution information from crystallography and cryo-EM.

PROTOCOL:

1. Protein expression, purification and SEC-SAXS measurement is based on the published protocol[13]

1.1. Follow the inline SEC-SAXS data collection protocol (Brennich et al.[6]) in brief.

1.1.1. Equilibrate the SEC-column with at least 2 column volumes of SEC running buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl).

1.1.2. Prepare 50 µL of sample of E9 exo^{minus} at 8–10 mg/mL with 20% molar excess of a partial dsDNA (TCAGGAAGATAACAGCGGTTTAGCC and GGCTAAACCGCTGTTATCTT). E9 exo^{minus} binds with an K_D of 12 ± 6 nM (see **Supplementary Data**).

1.1.3. Inject 50 µL of this mix onto a SEC-column (S200 Increase) inline with the flow cell for SAXS measurements at 0.3 mL/min.

1.1.4. Collect 1000 frames at 1 s exposure each.

NOTE: On BioSAXS beamline BM29, at the European Synchrotron Facility (ESRF) the individual frames are processed automatically and independently within the EDNA framework[14]. After the data collection, open the ISPyB database[20] and under the **Data acquisition** tab press the **Go** button to access the data set and the results of the automatic analysis[21].

1.2. Download the data.

2. Primary data analysis

2.1. Open the Java-based program **Scatter IV** (see the **Table of Materials**) and perform a background subtraction for size exclusion chromatography (SEC) data.

2.1.1. Open the **SEC** tab. Drag and drop the reduced data files (**.dat*) into the "Drop Data below" window. Set output directory "Out Dir ::" by clicking the blue-labeled **Output Dir** button.

NOTE: If your data was collected in nm^{-1}, a conversion box will need to be checked (bottom left of panel) when dropping files into the window or the subtraction tab.

2.1.2. Edit the experimental details, use the **Edit Details** button and fill out as many fields as possible, these include sections on which source/beamline was used to collect data, the collection parameters and sample details. These will be saved with the data and allow to more easily populate the "Data collection parameters" section in future publications.

2.1.3. Enter the sample name in the **Save as** box. Click on **TRACE**.

NOTE: This has two effects. Firstly, it will create a **.sec* file for the data. This is a single text file that will collate all the experimental observations from the separate **.dat* files. In addition, the **.sec* file contains the averaged set of frames that is the buffer-background, all the frames used in the averaging as well as the buffer-background subtracted frames across the entire SEC-SAXS experiment. Secondly, a signal plot is created which plots the frame number versus Integral ratio to the background. This shows the selected frames (gray) that were averaged for the buffer subtraction. The points for the average buffer are determined from the whole range of the data. However, it is advisable to manually choose the buffer frames for averaging as a poorly defined background may occur due to poorly equilibrated or dirty columns or capillary fouling[22].

2.1.4. Select buffer frames manually. Click **Clear Buffers** then reselect a buffer region with a left click-drag, on the trace curve. Ideally, this should be a flat region before the void volume of the SEC column of approximately 100 frames. Click **SET BUFFER** and then **Update** to recalculate the **.sec* file which may take a few minutes.

2.1.5. Identify a region of interest (ROI). On the signal plot, select the region of the peak of interest, with a left click-drag.

NOTE: This populates three plots in the right-hand panel. The top two plots are linked with crosshairs moving between them, a second signal plot (top right) shows only the ROI selected, with the intensity of each frame in blue and the corresponding Rg of each frame in red and a corresponding heat map below, showing the residuals for each frame colored according to the Durbin-Watson auto-correlation analysis. Regions of high similarity are colored cyan (Durbin-Watson, d = 2) while dissimilar frames will follow darker blues to pinks and finally to reds depending on the severity of the dissimilarity (d > 2). The bottom plot is a subtracted *I* versus *q* curve for the central selected frame (also denoted by a vertical line). The arrow keys can be used to navigate through the subtracted frames. The *I* versus *q* plot will demonstrate the quality of the subtracted frames from the SEC experiment.

2.1.6. Select frames to merge. Click on the crosshairs in the heat map plot to select the subset of frames that will be used for merging. The crosshairs will identify a triangular area of predominately cyan that falls to the bottom right-hand side of the crosshairs. Use a mouse-click to set these frames as selected and highlight the frames in the corresponding area of the Signal plot above. These frames should ideally highlight a region with a stable Rg.

NOTE: As needed, zoom in on the heat map with a left-click drag and zoom out with a left-click swipe to the right.

2.1.7. When satisfied with the selected frames click **MERGE**. This will merge the subtracted frames and present them in the **ANALYSIS** tab.

3. Data deconvolution

3.1. Open the deconvolution program (e.g., BioXTAS Raw 2.0.0).

3.2. In the deconvolution program, load the dataset, under **Files** tab, in the **Control Panel**, use foldether symbol to locate the data or copy and paste the location into the address bar.
NOTE: Make sure the folder contains only the raw **.dat* files and no processed or average data files.

3.3. Highlight all the **.dat* files, hit the **Plot Series** button, a plot of integrated intensity versus frame number will be drawn in the “Series Plot”.

3.4. In the **Control Panel** select the **Series** tab and then click to highlight the curve. Open up the **LC Analysis** pop-up window using the button at the base of the control panel. This window gives access to several options, such as selecting varying molecule types (protein or RNA). It also allows the user to select the buffer region for the plot. In the first instance click **Auto;** this should select a suitable buffer region.

NOTE: If this fails, possibly due to an unstable baseline, then "Add region" to optimize the buffer region. This populates the Buffer box with a smaller box in which one can manually add the frame numbers to use for the buffer. Alternatively, click "Pick" to give the option of selecting an area on the plot. Locate the area, left-click once for the start position, move the cursor to next position and left-click again. It may be necessary to add more than one buffer location. Click **Set buffer** and the curves will be subtracted and the Rg calculated across the SEC peak. If a pop-up box appears, click **OK**.

3.5. To start the Evolving Factor Analysis (EFA), right-click on the highlighted file at the bottom of the **Control Panel** and then select **EFA** from the menu.

3.5.1. Check that a pop-up window opens which shows the single value decomposition (SVD) of the data set. In the controls box, check the **Use Frames** box so that the whole peak area to deconvolute is covered in the intensity plot. The "Singular Values" plot, top right, shows the intensity of the singular values (separate peaks/species) above the baseline.

NOTE: The number of points present above the baseline represents the number of scattering species present. With the caveat that it is the relative magnitude of the singular value to the flat area/baseline that matters.

3.5.2. To help validate the number of single values, use the bottom **AutoCorrelation** plot. This shows the right and left single correlation vectors. Click **Next**.

NOTE: These essentially represent scattering or concentration profiles for the vector in the solution. Where the absolute size represents the significance of the vector. A significant component will have an autocorrelation near 1 (a rule of thumb cutoff is >0.6–0.7). RAW helpfully calculates this and is shown in the #Significant SVs box, bottom left, though you can change this if necessary. If there are several single values (e.g. 4+), it may be necessary to look at just 2 or 3 of the components only, altering the range of the data used. The lower the number of components the easier the EFA analysis will be but at the cost of using less data.

3.5.2. Check that the EFA is calculated by generating plots in the forward and backward directions for each vector. These plots show when components start (forward plot) and exit (backwards plot) the solution profile for the selected SEC-SAXS data. RAW tries to identify these ranges; change these using the arrows next to the counters so that each circle is at the start of an inflection point rising from or falling to baseline. Click **Next**.

NOTE: The last stage of the EFA turns the SVD vectors back into scattering curves. On the left of the window, the previously defined ranges are plotted at the top. These ranges are the constraints to define where to rotate the singular vectors back into scattering curves. The right hand panel shows these corresponding scattering curve profiles, for each separated peak. A plot for the concentration of each peak, that should be representative of elution profiles and a plot for the mean error weighted chi^2. The chi^2 plot is measuring the deconvolution data set to the original data set. Ideally, this will be flat, however spikes can often be seen.

3.5.3. Try to reduce or eliminate spikes by altering the **Component Range Controls**, first identify approximately which frame corresponds to the spike (from chi^2 plot) and then, in the **Range Controls,** which component contains this frame (it could be more than one), using the arrows, move up or down the corresponding range.

NOTE: This should produce a response, increasing or decreasing the spike. If the spike frame was present in more than one component then a little trial and error between each component may be necessary.

3.5.4. When a minimum chi^2 has been achieved, perform a validation check by clicking **back**, the previous window appears to allow verifying if the changes made have drastically changed the original EFA plots. If they still look valid, click **Next**. Click **Save EFA Data** to save the plots and then click **Done** to close the EFF window.

NOTE: A second validation is to click off the check-box next to each component range, in turn. These provide a positive concentration constraint to each component and turning off will check if these significantly affect the data set. If no change is seen in the concentration plot then the data are valid.

3.6. Back in the RAW window, click the **Profiles** tab in the **Control Panel** to view the curves and in the **Manipulation** tab of the **Control Panel**, manipulate the curves further or save the curves as **.dat* files by right-clicking on the file and selecting **save selected file(s)** from the menu pop up. Save the file. Use Scatter IV for further analysis.

NOTE: Further information and instructions on deconvolution and EFA BioXTAS RAW is found at https://bioxtas-raw.readthedocs.io/en/latest/

4. Determine SAXS properties

NOTE: An in-depth tutorial for SAXS determination is found at Bioisis.net. Here we show a basic step by step approach, highlighting the most useful buttons in Scatter.

4.1. In the Scatter **ANALYSIS** tab, press the **G** button for the manual Guinier analysis tool, to the right of each sample file. The plot that opens shows the $\ln[I(q)]$ versus q^2 in the top box and the corresponding residuals in the bottom box. Add or remove points such that the residuals do not

have a “smile” or “frown” feature. The selected data in the Guinier fit should not exceed the maximum *q* x Rg limit of 1.3.

4.2. Press the **Normalized Kratky** button; the plot that pops up provides a semi-quantitative assessment of macromolecule’s thermodynamic state, normalized for mass and concentration.

NOTE: The crosshairs designate the Guinier-Kratky point at (√3, 1.1)[19]. A compact, spherical protein will show a single peak with the maximum value at the Guinier-Kratky point. An intrinsically disordered or cylindrical biopolymer would have a maximum greater than the crosshairs and would not decrease. A protein that had both folded domains and long elongated unstructured regions might present with an increased maximum through the crosshairs but would also show an obvious decreasing trend at higher *q* x Rg.

4.3. Click on the **Vc** button (Volume-of-correlation), which brings up two plots, the total scattered intensity and an integrated area of the total scattered intensity as a function of q. The plots are used as a quick reference to validate the quality of the scattering curve.

NOTE: The total scattered intensity is sensitive to the *I*(0) and if this has not been measured correctly then the plot will not show a continuous line. The integrated area plot, ideally, should show a sigmoidal line with an extended plateau for each SAXS curve. If there are buffer mismatch/subtraction, aggregation or interparticle interference in the sample a sharp slope will be observed at higher q-values.

4.4. Press the **Flexibility** button to start the flexibility analysis. This will open a window with four panels and a slider at the bottom. Each opened panel shows a plot exploiting a power-law relationship that exists between compact and elongated/flexible biopolymers[23]. To use, move the slider at the bottom of the box from right to left with the left mouse button pressed. Keep moving slowly to the left until a plateau in one of the plots is reached.

NOTE: If the plateau is seen in the Porod-Debye plot, then the sample is compact in nature, which should be consistent with a single peak at the Guinier-Kratky point in a normalized Kratky plot. If the plateau is reached first in the Kratky-Debye plot then the sample is most likely elongated or flexible. If the SIBYLS plot is first to plateau, then the sample most likely contains areas of both compactness and flexibility, a particle with mixed states. The theory for this flexibility relationship with the Porod-Debye Law is exquisitely addressed in Rambo, et al.[23]

4.5. Click on **Volume**. Volume determination should be performed immediately after the flexibility analysis from above. When opened after the flexibility analysis, a pop up with three more graphs is generated. In the bottom, left corner the Porod-Debye plot remembers where one left the slider from the flexibility plot, showing the plateaued area.

4.5.1. To calculate the volume of the particle, move the start and endpoints using the arrow buttons or type in the boxes, so that the blue line on the plot fits the plateaued region. For an

unbiased result, the residuals in the top right Porod-Debye exponent power-law fit, should show no pattern.

4.6. Press the **P(r)** tab. The real-space distribution is in the left panel and the scattering curve for the sample in the right-hand panel. The objective is to create a real-space representation of the sample from the reciprocal space SAXS curve. Ideally, the distribution curve will be smooth with no waves present and should just gently kiss the x-axis.

NOTE: The instrument's measured q-range may not be entirely useable due to poor buffer matching, aggregation, radiation damage, sub-optimal exposure times and low particle concentrations. The P(r)-determination step will fundamentally determine the useable q_{min} and q_{max} range of the SAXS dataset and it should be this range of data that is used for any subsequent modeling or fitting.

4.6.1. Right-click on the sample name then click **Find DMAX** to open a new window. Limits for the d_{max} are pre-set with the suggested q_{max}, lower and upper d_{max} limits and a lower and upper alpha score. Three models can be chosen (L1-norm, Legendre and Moore) and the use of Background included. Leave these unchanged in the first instance.

4.6.2. Press the **Start** button. A composite distribution is created in the left panel with the suggested d_{max} and alpha level written underneath. If this looks acceptable then close the window and return to the P(r) tab. The reciprocal space plot will have been cropped to match the suggested q_{max}.

4.6.3. Choose the model **Moore**, click on **Background** and then set the alpha level and d_{max} to the suggested values from the pop-up box. Press the **refine** button. A cross-validation plot will pop up showing if any points had to be rejected, marked in red. If there are only a few points rejected and the distribution looks good then the model is good.

NOTE: The cross-validation plot will highlight regions of the data that are inconsistent with the determined P(r)-distribution. If the rejected region is mainly in the low-q region, that is the region near the y-axis, this likely suggests a d_{max} that is too short, presence of aggregation or higher order oligomers. It highlights an inconsistency between the higher and lower resolution information. Here, d_{max} and q_{min} (increasing start value) should be adjusted using a manual, trial-and-error approach. Similarly, if the rejected region is mainly in the high-q region, this may indicate an issue with the background subtraction or that the signal is too weak to be meaningfully explained by the determined P(r)-distribution. In this case, q_{max} should be truncated (decreasing end) until no additional data are rejected. Ideally, rejected points should be randomly distributed and make up less than 5% of the useable data. A properly defined q_{min}, q_{max} and "d_{max}" will produce a smooth distribution where the d_{max} kisses the x-axis. However, do not increase this value so much that it completely removes the Guinier region. This point is easily found by checking the **q x I(q)** box (on the left of the panel above the table). The scattering curve is replaced by the "Total Scattered Intensity plot", on this curve all points before the max inflection are part of the Guinier region. After removing points try again to increase/decrease the "d_{max}" and then

refine once more. If problems persist, especially when many points are being rejected from the start of the validation curve, this strongly suggests the data are not ideal for structural modeling.

4.7. To print a report, navigate back to the **Analysis** tab, left-click to highlight the sample then right-click on the sample name and move to **Create report from single data set** in the menu. A text box opens to allow for comments to be added. A PDF document is produced showing all the figures and values generated.

REPRESENTATIVE RESULTS:

The advantage of using deconvolution over classical frame selection[13] is to remove the influence of species on one another, producing a monodisperse scattering signal. This is also often followed with a better signal to noise ratio. When E9 exo^{minus} is bound to DNA and run using SEC-SAXS, two peaks are observed (**Figure 1**). The first, large peak (approximately frames 420–475) is the E9 exo^{minus}-DNA complex the second (approximately frames 475–540), the unbound state (see **Supplementary Data: Figure 2**). While the classical approach of selecting frames provides a stable Rg of the complex in the first peak (see **Supplementary Data: Figure 3**), the second peak is clearly merged and the Rg across the plot shows that the second peak of interest does not have a stable Rg, due to cross-peak contamination. Only 5 frames could be used that showed a semi-stable Rg, when subtracted they gave an Rg = 35.8 Å (**Figure 2**, green). When the peaks were deconvoluted using EFA the corresponding curve for the second peak (**Figure 2**, blue) was overlaid with the original and showed a clear decrease in signal to noise, and a lower Rg, 34.1 Å was recorded. The Kratky plot (**Figure 3**) shows the complex with the deconvoluted peak (blue) is more globular. This is confirmed by the P(r) curve (**Figure 4**) which gives a d_{max} 105 Å for the deconvoluted curve (blue) while the non-deconvoluted is more elongated with a d_{max} 123 Å (green), this is most likely due to heterogeneity arising from the unbound E9 exo^{minus}.

FIGURE AND TABLE LEGENDS:

Figure 1: Signal plot of E9 exo^{minus} alone and with DNA in complex. The top panel shows a plot of the integral ratio to the background for each frame of a SEC-SAXS run (light blue). The red points show the Rg at each frame over the peak. The bottom panel shows the corresponding heat map showing the residuals for each frame colored according to the Durbin-Watson auto-correlation analysis, regions of high similarity are colored cyan while dissimilar frames follow darker blues to pinks and finally to red depending on the severity of the dissimilarity.

Figure 2: Plot of intensity versus scattering vector. An overlay of the subtracted SAXS data form the E9 exo^{minus} . In green 5 frames (frame 517–522) averaged and subtracted from an area of semi-stable Rg and in blue the representative scattering curve derived from the EFA deconvolution of the SEC-SAXS peak.

Figure 3: Dimensionless Kratky curve. Overlay of the deconvoluted (blue) and non-deconvoluted (green) Kratky curve showing E9 exo^{minus} is globular.

Figure 4: P(r) curve. The overlay of the deconvoluted (blue) and non-deconvoluted (green) curves for the E9 exo^{minus}.

DISCUSSION:

It is desired to have a monodisperse sample before starting a SAXS experiment, but in reality, many data collections do not satisfy this and must be improved by combining the measurement with inline chromatography—SEC in most cases. However, even the shortage of time between purification and data acquisition monodispersity of the sample is not guaranteed. Most commonly, this applies to experiments where components are too close in size or in their physical properties to be separated or are prone to fast dynamics. Here, we have provided a protocol combining single value decomposition with evolving factor analysis to remove the influence of DNAbound E9 exo^{minus} from its unboundform creating a monodisperse scattering profile that we were then able to analyze with the SAXS package Scatter IV.

SVD with EFA of SEC-SAXS data are very powerful methods developed to deconvolute SAXS data and improve analysis, but they do have limitations. They require that noise or drift in the buffer baseline of the SEC-SAXS is kept to a minimum. This may involve extra column equilibration (better to use more than 3 column volumes, depending on the buffer) before sample loading. However, the most critical step is the choice of the number of the singular values and the range of data used, as this will greatly affect the accuracy of the deconvolution. It is for this reason that the results should not be taken on their own but further analyzed using techniques such as analytical ultracentrifugation (AUC) or multi-angle-laser-light-scattering (MALLS) for biological interpretation.

Scatter IV is a new, software package, free for research and industrial use with an intuitive user interface that allows even non-experts to analyze their data. Scatter IV has several new features that help to improve the analysis of SEC-SAXS data, such as the heat map linked to the signal plot, enabling greater accuracy with choice of frame selection. In primary data analysis, the Guinier Peak analysis and the cross-validation plot associated with the P(r) analysis offer an integrated troubleshooting ability in the software.

It should be mentioned that many other programs can be used for primary data analysis; these contain the same basic features and are also updated regularly such as BioXTAS RAW[17] ATSAS package[24] and US-SOMO[15] to name a few.

But regardless of which SAXS package is used for analysis, the major limitations are common: the sample preparation, before collection and analysis. In the E9 exo^{minus} example shown, it is clear to see the improvement in the signal to noise ratio and with a reduction in the Rg the d_{max} associated with a monodisperse sample. This will greatly aid further processing of the data such as fitting or modeling with known high-resolution structures.

ACKNOWLEDGMENTS:

We acknowledge the financial support for the project from the French grant REPLIPOX ANR-13-BSV8-0014 and by research grants from the Service de Santé des Armées and the Délégation

Générale pour l'Armement. We are thankful to the ESRF for the SAXS beam time. This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA). We thank Wim Burmeister for financial and scientific support and we also thank Dr. Jesse Hopkins from BioCAT at the APS for his help and for developing BioXTAS RAW.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1. Pelikan, M., Hura, G., Hammel, M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *General Physiology and Biophysics*. **28** (2), 174–189, doi: 10.4149/gpb_2009_02_174 (2009).
2. Brosey, C.A., Tainer, J.A. Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Current Opinion in Structural Biology*. **58**, 197–213, doi: 10.1016/j.sbi.2019.04.004 (2019).
3. Gräwert, M., Svergun, D. A beginner's guide to solution small-angle X-ray scattering (SAXS). *The Biochemist*. **42** (1), 36–42, doi: 10.1042/BIO04201036 (2020).
4. Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J.A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*. **40** (03), 191–285 (2007).
5. Meisburger, S.P. et al. Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. *Journal of the American Chemical Society*. **138** (20), 6506–6516, doi: 10.1021/jacs.6b01563 (2016).
6. Brennich, M.E., Round, A.R., Hutin, S. Online Size-exclusion and Ion-exchange Chromatography on a SAXS Beamline. *Journal of Visualized Experiments*. (119), doi: 10.3791/54861 (2017).
7. Watanabe, Y., Inoko, Y. Size-exclusion chromatography combined with small-angle X-ray scattering optics. *Journal of Chromatography A*. **1216** (44), 7461–7465 (2009).
8. Graewert, M.A. et al. Automated Pipeline for Purification, Biophysical and X-Ray Analysis of Biomacromolecular Solutions. *Scientific reports*. **5** (2015).
9. David, G., Pérez, J. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *Journal of applied crystallography*. **42** (5), 892–900 (2009).
10. Ryan, T.M. et al. An optimized SEC-SAXS system enabling high X-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis. *Journal of Applied Crystallography*. **51** (1), 97–111, doi: 10.1107/S1600576717017101 (2018).

11. Gampp, H., Maeder, M., Meyer, C.J., Zuberbühler, A.D. Calculation of equilibrium constants from multiwavelength spectroscopic data—III: Model-free analysis of spectrophotometric and ESR titrations. *Talanta*. **32** (12), 1133–1139, doi: 10.1016/0039-9140(85)80238-1 (1985).
12. Maeder, M., Neuhold, Y.-M. *Practical Data Analysis in Chemistry.* at <http://www.123library.org/book_details/?id=35069>. Elsevier. Burlington. (2007).
13. Tarbouriech, N. et al. The vaccinia virus DNA polymerase structure provides insights into the mode of processivity factor binding. *Nature Communications*. **8** (1), doi: 10.1038/s41467-017-01542-z (2017).
14. Brennich, M.E. et al. Online data analysis at the ESRF bioSAXS beamline, BM29. *Journal of Applied Crystallography*. **49** (1) (2016).
15. Brookes, E., Rocco, M. Recent advances in the UltraScan SOlution MOdeller (US-SOMO) hydrodynamic and small-angle scattering data analysis and simulation suite. *European Biophysics Journal*. **47** (7), 855–864, doi: 10.1007/s00249-018-1296-0 (2018).
16. Malaby, A.W. et al. Methods for analysis of size-exclusion chromatography–small-angle X-ray scattering and reconstruction of protein scattering. *Journal of Applied Crystallography*. **48** (4), 1102–1113, doi: 10.1107/S1600576715010420 (2015).
17. Hopkins, J.B., Gillilan, R.E., Skou, S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*. **50** (5), 1545–1553, doi: 10.1107/S1600576717011438 (2017).
18. Maeder, Marcel. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Analytical Chemistry*. **59** (3), 527–530, doi: 10.1021/ac00130a035 (1987).
19. Durand, D. et al. NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of Structural Biology*. **169** (1), 45–53, doi: 10.1016/j.jsb.2009.08.009 (2010).
20. De Maria Antolinos, A. et al. ISPyB for BioSAXS, the gateway to user autonomy in solution scattering experiments. *Acta Crystallographica Section D*. **71** (1), 76–85 (2015).
21. Brennich, M.E. et al. Online data analysis at the ESRF bioSAXS beamline, BM29. *Journal of Applied Crystallography*. **49** (1), 203–212, doi: 10.1107/S1600576715024462 (2016).
22. Kirby, N. et al. Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta Crystallographica Section D Structural Biology*. **72** (12), 1254–1266, doi: 10.1107/S2059798316017174 (2016).
23. Rambo, R.P., Tainer, J.A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*. **95** (8), 559–571, doi: 10.1002/bip.21638 (2011).
24. Franke, D. et al. *ATSAS 2.8* : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*. **50** (4), 1212–1225, doi: 10.1107/S1600576717007786 (2017).

Figure 1

Click here to access/download;Figure;Fig_1_signalplot.png

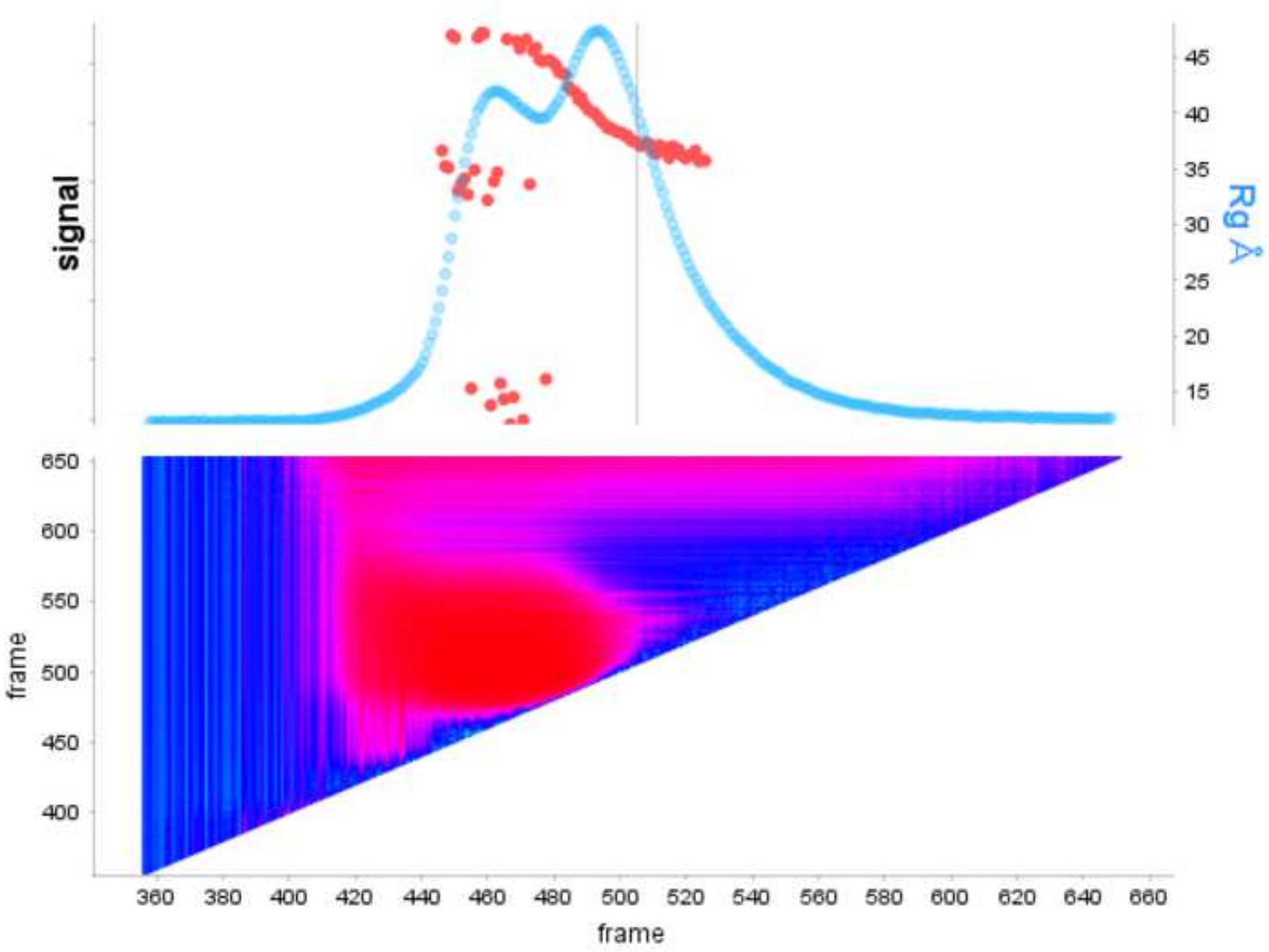


Figure 2

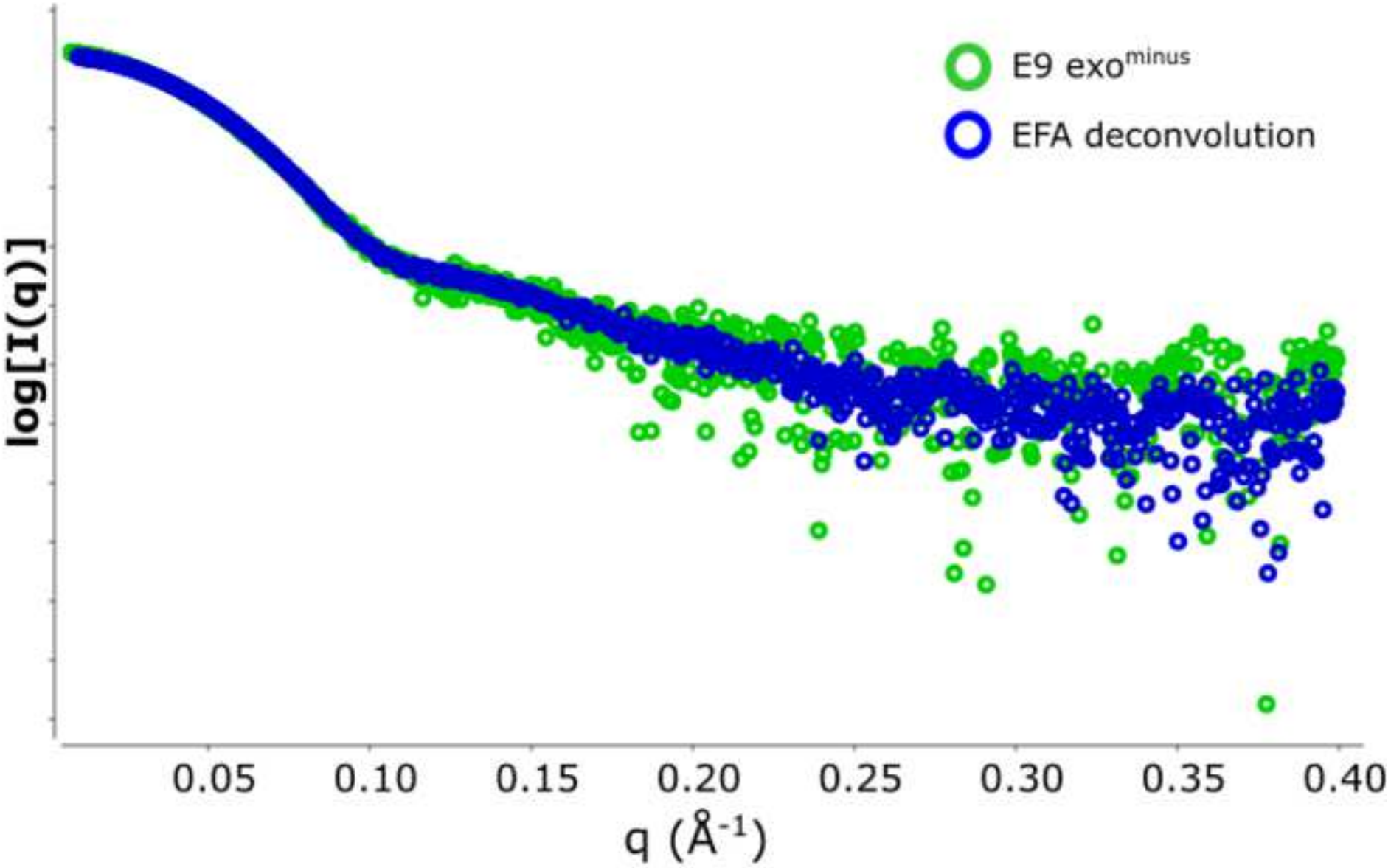


Figure 3

Click here to access/download;Figure;Fig_3_kratkyedit.png

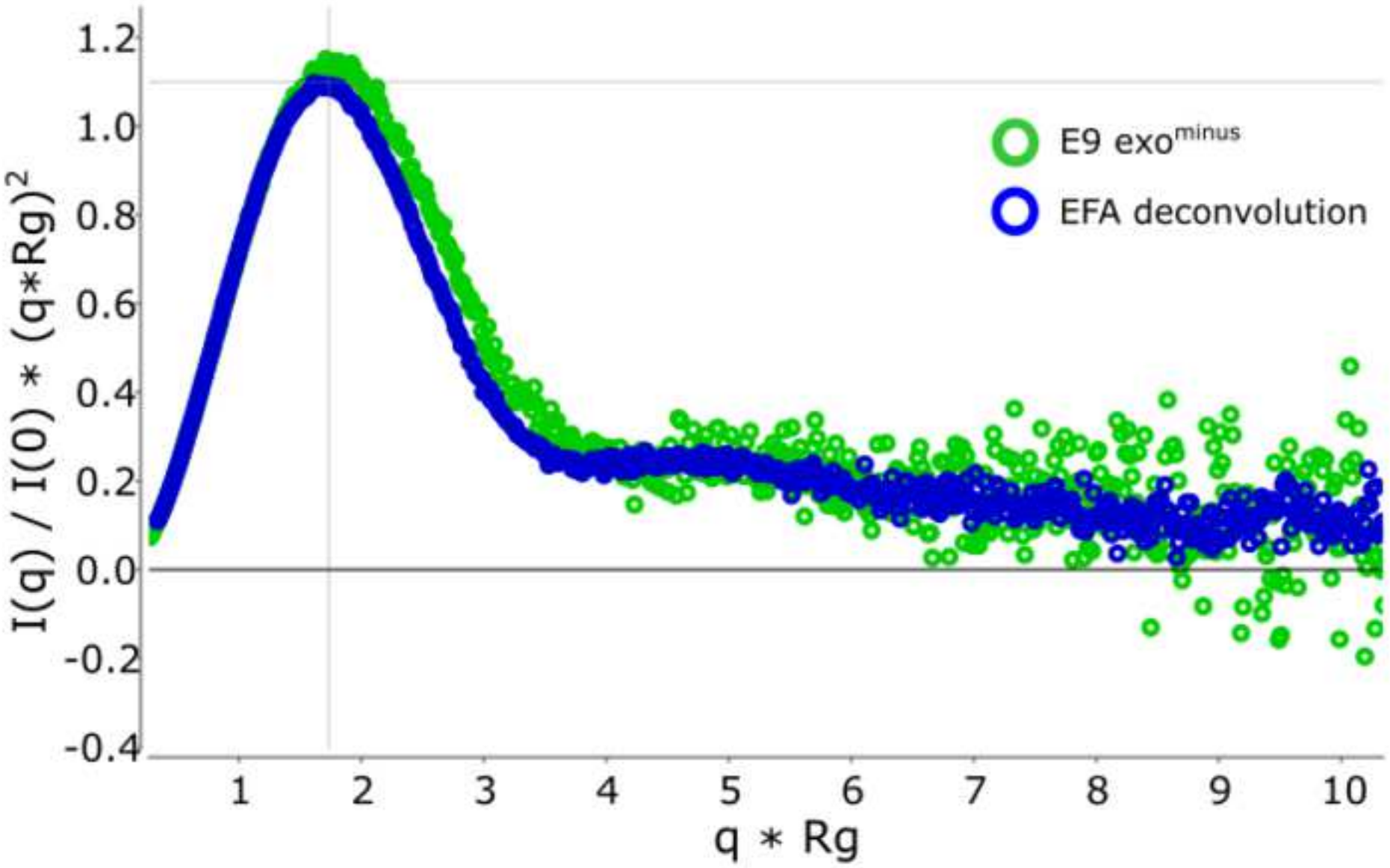
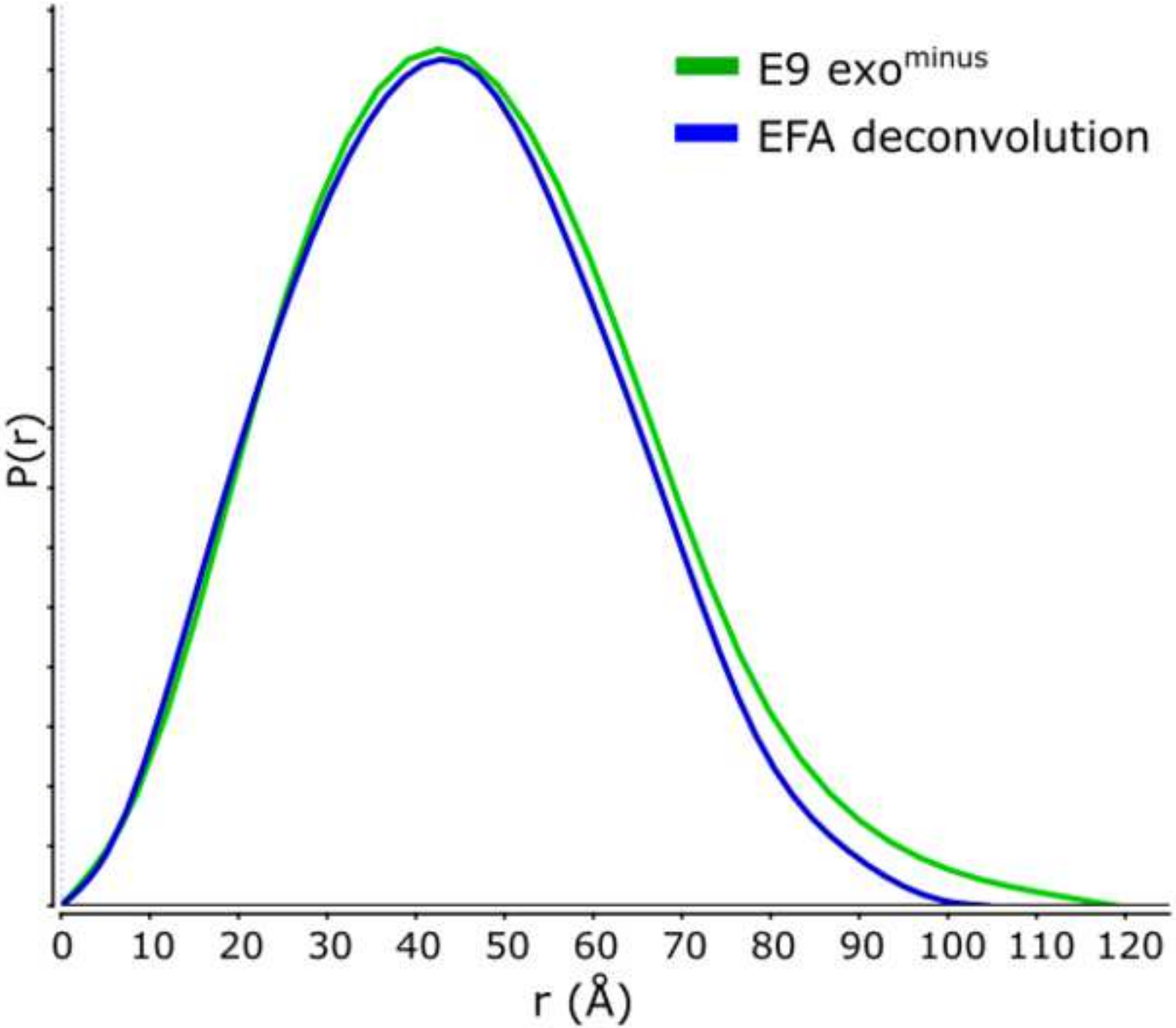


Figure 4

Click here to access/download;Figure;Fig_4_pr_edit.png



Click here to access/download

Video or Animated Figure

Fig_3_kratky.svg

Click here to access/download

Video or Animated Figure

Fig_4_pr.svg

Name of Material/ Equipment	Company	Catalog Number
Beamline control software BsXCuBE	ESRF	Pernot et al. (2013), J. Synchrotron Rad. 20, 660-664
BioXTAS Raw 1.2.3.	MacCHESS	http://bioxtas-raw.readthedocs.io/en/latest/index.html
HPLC program LabSolutions	Shimadzu	n.a.
ISPyB	ESRF	De Maria Antolinos et al. (2015). Acta Cryst. D71, 76-85.
NaCl	VWR Chemicals (BDH Prolabo)	27808.297
Scatter	Diamond Light Source Ltd	http://www.bioisis.net/tutorial/9
Superdex 200 Increase 5/150 GL column	GE Healthcare	28990945
Tris base	Euromedex	26-128-3094-B

	Comments/Description
local development	
First developed in 2008 by Soren Skou as part of the biological x-ray total analysis system (BioXTAS) project. Since then it has been extensively developed, with recent work being done by Jesse B. Hopkins	
local development	
Supported by SIBYLS beamline (ALS berkeley, Ca) and Bruker Cororation (Karlsruhe, Germany)	
SEC-SAXS column used	

Stephanie Hutin, PhD
Laboratoire de Physiologie Cellulaire and Végétale,
Université Grenoble Alpes/
CNRS/CEA/INRA/BIG,
Grenoble, France
Tel: +33 (0) 6 47 25 23 16

Alisha DSouza
Senior Review Editor
JoVE

Dear Dr. DSouza,

Thank you very much for the possibility to improve the manuscript to publish in JoVE. We tried to address each of the questions and comments of the editorial board as well as the reviewers.

Please find our comments attached to this letter and the new version of the manuscript uploaded.

I am looking forward to hearing from you at your earliest convenience

Mark Tully and Stephanie Hutin

Please note that the reviewers raised some significant concerns regarding your method and your manuscript. Please revise the manuscript to thoroughly address these concerns. Additionally, please describe the changes that have been made or provide explanations if the comment is not addressed in a rebuttal letter. We may send the revised manuscript and the rebuttal letter back to peer review.

You will find Editorial comments and Peer-Review comments listed below. Please read this entire email before making edits to your manuscript.
NOTE: Please include a line-by-line response to each of the editorial and reviewer comments in the form of a letter along with the resubmission.

We thank the editor and reviewers for their useful and thoughtful revision. Please find below our comments and corrections concerning each point.

Editorial Comments:

• Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammatical errors.

Thank you very much for the advice, we ask native speakers to control the text.

• **Protocol Numbering:** All steps should be lined up at the left margin with no indentations. There must also be a one-line space between each protocol step.
We corrected this.

• **Protocol Highlight:** Please highlight ~2.5 pages or less of text (which includes headings and spaces) in yellow, to identify which steps should be visualized to tell the most cohesive story of your protocol steps. Please see JoVE's instructions for authors for more clarification. Remember that the non-highlighted protocol steps will remain in the manuscript and therefore will still be available to the reader.

1) The highlighting must include all relevant details that are required to perform the step. For example, if step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be included in the highlighting.
2) Some of your shorter protocol steps can be combined so that individual steps contain 2-3 actions and maximum of 4 sentences per step.
3) The highlighted steps should form a cohesive narrative, that is, there must be a logical flow from one highlighted step to the next.
4) Please highlight complete sentences (not parts of sentences). Include sub-headings and spaces when calculating the final highlighted length.
5) Notes cannot be filmed and should be excluded from highlighting.

2.5 pages are highlighted

• **Discussion:** JoVE articles are focused on the methods and the protocol, thus the discussion should be similarly focused. Please ensure that the discussion covers the following in detail and in paragraph form

(3-6 paragraphs): 1) modifications and troubleshooting, 2) limitations of the technique, 3) significance with respect to existing methods, 4) future applications and 5) critical steps within the protocol.

We rewrote the discussion accordingly.

• **Figures:** Please provide each figure (if multiple panels are present per figure, keep them within 1 file) as an individual SVG, EPS, AI, TIFF, or PNG file.

We uploaded all images as single PNG files.

• **Table of Materials:** Sort this alphabetically.

Please find the table in alphabetical order.

• If your figures and tables are original and not published previously or you have already obtained figure permissions, please ignore this comment. If you are re-using figures from a previous publication, you must obtain explicit permission to re-use the figure from the previous publisher (this can be in the form of a letter from an editor or a link to the editorial policies that allows you to re-publish the figure). Please upload the text of the re-print permission (may be copied and pasted from an email/website) as a Word document to the Editorial Manager site in the "Supplemental files (as requested by JoVE)" section. Please also cite the figure appropriately in the figure legend, i.e. "This figure has been modified from [citation]."

Our Figures are originals and produced by us and not published before.

Reviewer: 1

Manuscript Summary:
The article describes a procedure for deconvolving overlapping peaks in SEC-SAXS data using widely available software. It will be a useful resource to the community when published. However, I have several major concerns detailed below, as several steps of the procedure failed when I attempted to do them. Some of these had little effect on my ability to do the rest of the analysis, but some of them, such as the incorrect description of how to set the buffer region in BioXTAS RAW (using the currently available version), and the failure of the Dmax window to open in Scatter, represent major issues. Scatter is also not available from the website provided by the authors, which is another major issue. The manuscript needs major revision, and then should be reviewed again to verify they have fixed these issues in the protocol.

Other notes:
Because I found several bugs in the Scatter software, of which one of the article authors is the author, I have included the following information for their use in debugging:
Downloaded from: sibyls webpage (bioisis.net doesn't work)
Version: Scatter IV (downloaded April 27, 2020), no other version information is obvious.
Java JDK: 14.0.1, 64 bit
Operating system: MacOS 10.14.6

Thank you for your thoughtful reviewing, we are trying to incorporate your suggestions.

Major Concerns:
-bioisis.net is not available, you can't download Scatter from there. Nor is the in-depth protocol for analysis mentioned in section 4 available. This should be made available elsewhere and appropriately mentioned in the paper or added to the paper itself.

We are moving server, but the latest version of the program can always be made available by contacting Rob Rambo directly. By the time the manuscript will be published the new server should be operational.

-The protocol fails at step 2.2.3, where I was unable to select an output directory for the SEC-SAXS data. After loading files as directed, I cannot select 'Choose' in the save directory dialog.

We are sorry that this step failed, but we retested it and in the latest version it works.

-The protocol fails at 2.2.4, clicking 'Edit Details' does nothing.

Again, we are sorry that this step failed, it works for us.

The authors are using an outdated version of BioXTAS RAW (version 1.2.3 is 3 years old, according to the BioXTAS RAW release notes). Authors should download and use the latest version, currently 1.6.4, as that is what will be available for readers of the article. This will change parts of the protocol, as noted below, and the protocol needs to be updated appropriately.

We updated BioXTAS RAW to the version 2.0.0 and changed the protocol accordingly. Please see the main manuscript.

-In section 3, the authors need to provide some guidance on how to pick the correct range for deconvolution and how to pick the right number of significant singular values for deconvolution. That seems to be at the heart of this method, but is glossed over.

Thank you for pointing this out. Using the new version of BioXTAS RAW, this problem is solved within the program and the protocol is adjusted to it.

-The protocol fails at step 4.6. In 4.6.1, right clicking and selecting 'Find DMAX' does not open a new window. It appears the options described are available in the bar below the plot, but there doesn't seem to be a 'Start' button. The rest of 4.6 cannot be carried out.

This problem is now fixed in the new version of Scatter 03Mar2020B1.

-The protocol fails at step 4.7. No text box opens for comments. No PDF is saved.

This potential problem is now fixed in the new version of Scatter 03Mar2020B1 as well.

Minor Concerns:

-There are several available methods for deconvolving SEC-SAXS data, primarily EFA as implemented in BioXTAS RAW and the Gaussian peak deconvolution as implemented in US-SOMO but also other methods such as the Guinier optimized maximum likelihood method implemented in the DELA software (Malaby et al. J. Appl. Cryst 2015). Authors only use one of these methods. They should adjust their title, abstract, and manuscript comments to clarify this fact, comment on the existence of other methods, and if appropriate should mention why they use the EFA method over the Gaussian deconvolution or other methods.

We added a paragraph in the introduction discussing other methods and our choice and changed the title accordingly to "Analysis of SEC-SAXS data via EFA deconvolution and Scatter" and adjusted the abstract.

-While reference 10 is one of the original references for EFA, a more modern and useful reference that could be included is Practical Data Analysis in Chemistry by Maeder & Neuhold (2007), published by Elsevier Science. I suggest including both.

Thank you, we added the citation.

-Line 65, 107: It is not clear to me that the phrase thermodynamic state or thermodynamic measurement is appropriate here. While technically true, usually the use is accompanied by definition of state variables. The use also often, but not always implies, a system in equilibrium, which is not required for SAXS. It would be useful if the authors can clarify what they mean by these phrases.

You are correct, a more rounded explanation of thermodynamic should be incorporated, but as this is not really required for this this paper, we have removed the term.

-I suggest that the authors provide more citations in their introduction. Citing several recent review papers for SAXS would be useful for readers who are not as familiar with SAXS, and there are several appropriate places those could be added.

People who are not familiar with working with SAXS are invited to take a look in our first JoVE publication Brennich et al. as cited right at the start of the method. Reviews and other citations are now added in the introduction.

-Line 76: The authors cite a SEC-SANS paper without ever mention SANS. I think that this either ought to be mentioned or the citation removed.

We took the reference out.

-Recently SEC-SAXS has been demonstrated on a lab source. The authors should comment on whether these methods are restricted to SEC-SAXS at a synchrotron or can be applied to laboratory SEC-SAXS as well.

Technically, the method will work with any lab source, given that we do not have any experiences we did not go into detail here. Still, it is mentioned in the introduction now

-Lines 95-96: When discussion EFA, the authors state "A SAXS specific implementation can be found in several programs for example in BioXTAS RAW." If there are other SAXS specific implementations, those should be cited as well. If appropriate, the authors should discuss why they chose to use BioXTAS RAW for this paper.

We added an entire paragraph to discuss our choice.

-Lines 99-110: There are a lot of different techniques mentioned in this paragraph. Citations to relevant literature that explain the how/why of these techniques should be included.

We added citations accordingly.

-Line 118: Type of column (10/300, 5/150, etc) should be included here.

The 5/150 was included in the text

-Line 139: Output directory is labeled "Save as ::" in Scatter IV, not "Out Dir ::".

This problem was taken care of and the manuscript and the new version are now aligned.

-Line 139: Unable to select an output directory in Scatter IV.

This problem was fixed in the new version.

-Protocol step 2.2.8 is unclear. Which plot is the main plot? If it is the signal plot, then why refer to it with two different names? How do you select an ROI? Is it the same as selecting a buffer region? Is the buffer region just an ROI? Please clarify.

We changed it to: 2.2.8 Identify a region of interest (ROI). On the main signal plot, select the general region of the peak that you are interested in. This populates three plots in the right hand panel. The top two plots are linked with crosshairs moving between them, a signal plot (top right) showing only the ROI selected with the intensity of each frame in blue and the corresponding Rg of each frame in red and a corresponding heat map showing the residuals for each frame colored according to the Durbin-Watson auto-correlation analysis. Regions of high similarity will be colored cyan (Durbin-Watson, d=2) while dissimilar frames will follow darker blues to pinks and finally to reds depending on the severity of the dissimilarity (d>2). The bottom plot is a subtracted I vs. q curve for the central selected frame. The arrow keys can be used to navigate through the subtracted frames. The I vs. q plot will demonstrate the quality of the subtracted frames from the SEC experiment.

-In step 2.2.6 the authors use signal plot to refer to the integral ratio to background plot, whereas in 2.2.8 they use it to refer to the plot labeled 'signal'. This is confusing and should be clarified.

The text is changed now to main plot to signal plot.

-In Step 2.2.8 I would find it helpful if the authors mentioned that the vertical line in the 'signal' plot corresponds to the subtracted profile displayed in the bottom plot (at least, that's my assumption).

This is correct, we have added extra wording to highlight this.

-In step 2.2.9, zooming out with a left click swipe to the right compressed the correlation plot into uselessness.

We do not see that problem. We assume that you will not have this problem in the new version any longer.

-Line 185: It is BioXTAS not Bio-XTAS

We corrected this typo.

-Line 189: In BioXTAS RAW 1.6.4 it is the 'Plot Series' button, not 'Plot SEC'.

We corrected this in the text.

-Line 190: It displays on the 'Series' plot in the Plot Panel.

We corrected this in the text.

-Line 191: In BioXTAS RAW 1.6.4 it is the 'Series' tab not the 'SEC' tab

We corrected this in the text.

-In step 3.4, this does not work in BioXTAS RAW 1.6.4. You now use the LC Series Analysis panel to do buffer subtraction for a series, as described here: https://bioxtas-raw.readthedocs.io/en/latest/tutorial/s1_sec.html

Step 3.4 adapted to BioXTAS RAW 2.0.0 and was rewritten.

-Line 195, 200, 203, 208, 314, 331: 'EFA' not 'EFF'

We corrected this in the text.

-Line 212: Should read 'selecting "save selected file(s)" from the menu pop up'?

We corrected this in the text.

-In the section 'representative results' the authors state "The advantage of using deconvolution over

classical frame selection is the better signal to noise ratio". While this is true, it is a somewhat limited perspective. Often the peak overlap may be sufficiently strong that you cannot obtain monodisperse data for one or more of the eluted components through 'classical frame selection'. The authors should correct/discuss this. They actually observe it in their results, but do not sufficiently comment on it.

We changed it to: The advantage of using deconvolution over classical frame selection is to remove the influence of other species on to one another, producing a monodisperse scattering signal.

-Figures are poor quality. Can the authors generate higher resolution figures? Also, some of this may be from the generation of the review PDF, but the authors should look into this.

We will upload them in svg and png to ensure high quality.

-Figures 2, 3 and 4 legend should be EFA not EFF.

We corrected this in the text.

-Ideally all links in the document would be provided as https, not http, when available.

Bioisis was not a https but this will be solved by moving the server

-The link for Scatter IV in the table of materials doesn't work.

This will be solved with the new server

-In the table of materials, it should be 'Sibyls' beamline not 'Sibylis' beamline. Also, I believe it is capitalized (SIBYLS).

We corrected this in the table.

Scatter IV specific suggestions or bugs, not required to be fixed for publication:
-Program suggestion: I would find it extremely helpful to be able to fine tune the selected ROI range without having to click and drag on the main plot. For example, when testing the protocol for the paper I found myself wanting to remove ~4 frames from the start of the ROI, but found it hard to click that precisely (e.g. changing start frame from 687 to 691 while keeping the same end frame). It is possible to zoom in on the plot, but I found that not terribly useful, as you couldn't control the amount of zoom.
-Program bug: There is a bug in Scatter IV where the selected ROI displayed on the main plot disappears when zooming in on the main plot.
-Program suggestion: It would be very useful if the ROI was displayed dynamically while clicking and dragging on the plot.

We thank you very much for this useful and thoughtful revision and your comments to improve Scatter IV.

Reviewer #2:

Manuscript Summary:

The authors reported the usability of EFA for SEC-SAXS data in this manuscript. The reviewer admits the importance of such a trial for obtaining the reliable mono-disperse data. On the other hand, the detailed processes applied for real sample is totally missing in the present version of manuscript. The reviewer believes that complete set of these information will be quite helpful for future readers, hence they should describe them in more detail. Other comments are described below.

Thank you for your input.

Major Concerns:

1. The authors should explain the principle of EFA. Without it, it is very difficult to catch up with the advantage of this method beyond conventional SVD methods.

We added a paragraph in the introduction discussing why we choose EFA and mention different alternatives.

2. They should describe the dissociation contant of E9 exominus and DNA.
Without it, feature readers might consider that two peaks are originated from pure E9 exominus and pure DNA. Other experimental technique such as AUC is needed for validating the existence of complex of E9 exominus and DNA. Only the SEC chart is not sufficient for proving the detection of complex.

We agree that a SEC chart is not sufficient for proving the detection of complex, but the complex was described in: The vaccinia virus DNA polymerase structure provides insights into the mode of processivity factor binding. Tarbouriech N, Ducournau C, Hutin S, Mas PJ, Man P, Forest E, Hart DJ, Peyrefitte CN, Burmeister WP, Iseni F. Nat Commun. 2017 Nov 13;8(1):1455. doi: 10.1038/s41467-017-01542-z. Further we added suggestions in the discussion and provided supplementary informations.

3. 3.5.4 A bit of trial and error is required here.
>> The reviewer believes that this procedure
is the crucial step for successful EFA,so the authors should describe more information about obtaining reliable minimum chi2 value.

We agree, we have completely changed this section to remedy this.

4. Detailed process for EFA is totally missing. Hence the reviewer could not judge the validity of this analysis for E9 exominus and DNA system. The authors should describe the procedures step by step in more detail.

We agree, we have completely changed this section to remedy this

Minor Concerns:

1. It is better to include the information of HPLC.

The HPLC procedure was described in: The vaccinia virus DNA polymerase structure provides insights into the mode of processivity factor binding. Tarbouriech N, Ducournau C, Hutin S, Mas PJ, Man P, Forest E, Hart DJ, Peyrefitte CN, Burmeister WP, Iseni F. Nat Commun. 2017 Nov 13;8(1):1455. doi:

10.1038/s41467-017-01542-z. A detailed description of how to perform a HPLC run on the beamline can be found in one of our earlier JoVE publications: Brennich, M.E., Round, A.R., Hutin, S. Online Size-exclusion and Ion-exchange Chromatography on a SAXS Beamline. Journal of Visualized Experiments. (119), doi: 10.3791/54861 (2017). Hence we only cite the two papers.

2. At all the sections, it is better to include the brief flow charts.
It will be helpful for understanding the procedures at each section.

Given that the protocol will be filmed, we assume that a flow chart will not be necessary.

3. The reviewer considers that the globular protein have a maximum at QRg=root(3) in the Kratky plot. But the reviewer could not understand the meaning of Guinier-Kratky point at (root(3), 1.1) especially for 1.1. Please explain it in more detail.

The scattering pattern of a globular protein follows Guinier's law $I(q) = I(0) \exp(-(qR_g)^2/3)$. The corresponding dimensionless Kratky plot exhibits a maximum with a value $3 \exp(-1) = 1.104$ at $qR_g = \sqrt{3}$. A link to the original paper where this was first defined has been added.

Reviewer #3:
In this manuscript Tully et al describe a protocol to analyze SEC-BioSAXS data. SAXS is an incredibly powerful structural biology technique but achieving high quality data can be challenging. This detailed protocol describes the steps to deconvolute scattering curves from SEC profiles with partially resolved peaks. Aside from the minor comments below this protocol is acceptable for publication in JoVE.

Thank you for your comments.

Minor Comments:
1.1.2 - Please provide the suggested concentration of E9 exo as this is an important parameter for SAXS

The concentration is now included in point 1.1.2.

2.2.4 - Please elaborate on what the experimental details entail

We elaborated point 2.2.4 to Edit your experimental details, use the "Edit Details" button and fill out as many fields as possible, these include sections on which source/beamline was used to collect data, the collection parameters and sample details. These will be saved with your data and allow you to more easily populate the "Data collection parameters" section in your future publications.

3.5 - Please define the EEF acronym

This was a typo we corrected. It should have been EFA.

Representative Results - It would be very useful to show an example SEC-BioSAXS trace where this method fails.

Although we are thankful for the comment we disagree with the idea of showing negative results because we do not see the point why that would help in a protocol describing a method. We are giving here a complex example.

Discussion - This discussion seems too brief. I suggest that the authors elaborate on more of the benefits of this particular method. Aside from Scatter couldn't a similar analysis also be carried using the ATSAS data suite?

The discussion was rewritten including this point.

Reviewer #4:
Manuscript Summary:
The authors report a detailed method for the deconvolution of SEC-SAXS data. Briefly, structural characterization of biological molecules by SAXS requires that the scattering curve corresponds to a monodisperse entity. However, this is not always possible to obtain even by using size-exclusion chromatography. Species of different sizes can overlap and undergo chemical exchange. To overcome this, Meisburger et al (2016) reported an algorithm for the deconvolution of SEC-SAXS data, which is accomplished using the software BioXTAS. In this manuscript, the authors described a detailed protocol for 1. acquisition, 2. Primary data processing, 3. Data deconvolution, 4. SAXS properties analysis of SEC-SAXS.

Major Concerns:
No major concerns.

Minor Concerns:
Overall, this is a very technical and detailed protocol destined at users that are already familiar with BioSAXS, but have never performed deconvolution. I found that the following minor points to be corrected.

Thank you for pointing them out.

Line 64: "but again the samples are not in a close to physiological condition using this technique." Please rephrase. Also, because of the flash-freezing process, cryoEM is arguably as close to physiological as SAXS. I don't think it is necessary at all to state that.

During the last couple of years cryo-electron microscopy delivered similar high-resolution structures of large macromolecules / macromolecular complexes, but although the samples are closer to physiological condition they are still frozen, hence immobile and static. Bio-Small Angle X-ray Scattering (BioSAXS) provides a structural measurement of the macromolecule’s thermodynamic state in conditions that are relevant to biology.

Line 85: "...the complex recently solved". Please add "... by X-ray crystallography".

We added it.

Lines 86-90: For this paragraph, please indicate in which section "data rechecking" is shown.

We added it.

Line 132: The Bioisis website is currently down, so I could not verify the section 2
We are moving server, but the latest version of the program can always be made available by contacting Rob Rambo directly. By the time the manuscript will be published the new server should be operational.

Line 249: Please define the Sibyls plot, this is not common knowledge even in the SAXS field.

This is a relatively new plot, the paper on the Porod-Debye law by Rambo and Tainer gives a great explanation. We have directed users to this.

Lines 307-319: Please define better the "first" and "second" peaks by providing their frame numbers. Also please indicate exactly which 10 frames were used for generating the green curve in Fig. 2. I thought this should correspond to the peak of interest (the complex), which I assume would elute earlier in SEC. Please clarify.

We rewrote the entire explanation and discussion on this point.

Again, we would like to thank all reviewers and our editor for their thoughtful and helpful suggestions.

Sincerely,

Mark Tully and Stephanie Hutin

Supplementary data

Material :

Primers :

5'biot-temp-30 : biotin-CCGAATCAGGAAGATAACAGCGGTTTAGCC

temp-25 : TCAGGAAGATAACAGCGGTTTAGCC

prim-20 : GGCTAAACCGCTGTTATCTT

Primer combinations:

for SAXS : temp-25 + prim-20

for SPR : 5'biot-temp-30 + prim-20

Methods:

Annealing:

The primers 5'biot-temp-30 and prim-20 or temp-25 and prim-20 were mixed at a final concentration of 100 µM each in 20 mM Tris pH 7.0, 75 mM NaCl, 2 mM Mg Ac and annealed using a linear temperature gradient from 98°C to 23°C at 0.5°C temperature decrease per minute using an Eppendorf Mastercycler pro.

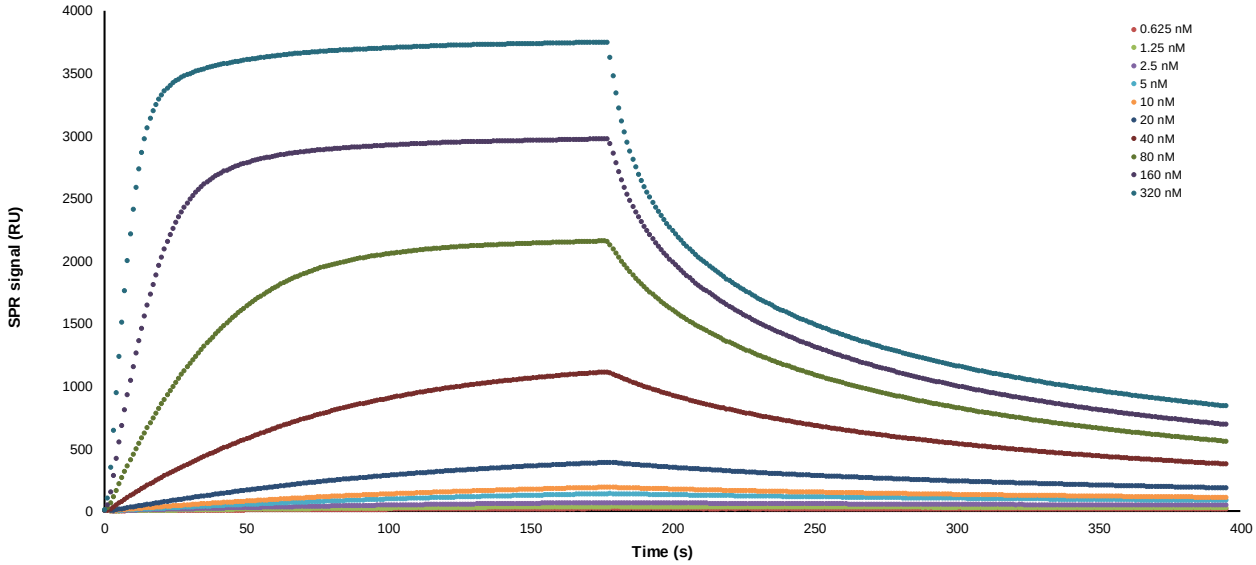
SPR experiment:

Data were collected on a BIAcore 3000 on a streptavidin coated chip. All experiments were performed in 20 mM Tris pH 7.0, 100 mM NaCl with a flow rate of 15 µl/min.

Sensor preparation : 75 µL of a 1 µM solution of 5'biot-temp-30 + prim-20 annealed primers were loaded at 5µL/min on one flow cell, a second flow cell was used for background subtraction. Unbound DNA was eliminated by a 2 min wash with 0.05% SDS.

SPR run: A dilution series from 0.625 to 320 nM of E9 exo^{minus} (2-fold step) was injected each during 3 min followed by a 5 min dissociation. The protein was washed between injections with a 2 min 0.05% SDS wash before being re-equilibrated in running buffer for 5 minutes.

Data analysis: Background subtracted signals were imported to LibreOffice Calc for curve fitting using the Solver non-linear function. Single exponential fit gives a K_D of 12 ± 6 nM. Figure was generated using MS Excel.



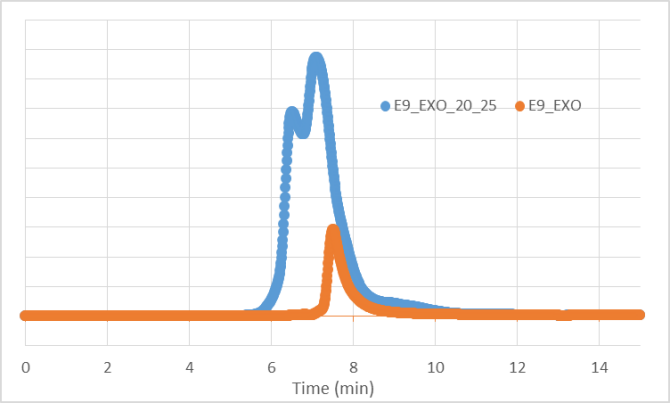
Supplementary Figure 1: SPR signal curves for binding of E9 exo^{minus} to DNA.

Sample preparation for SAXS:

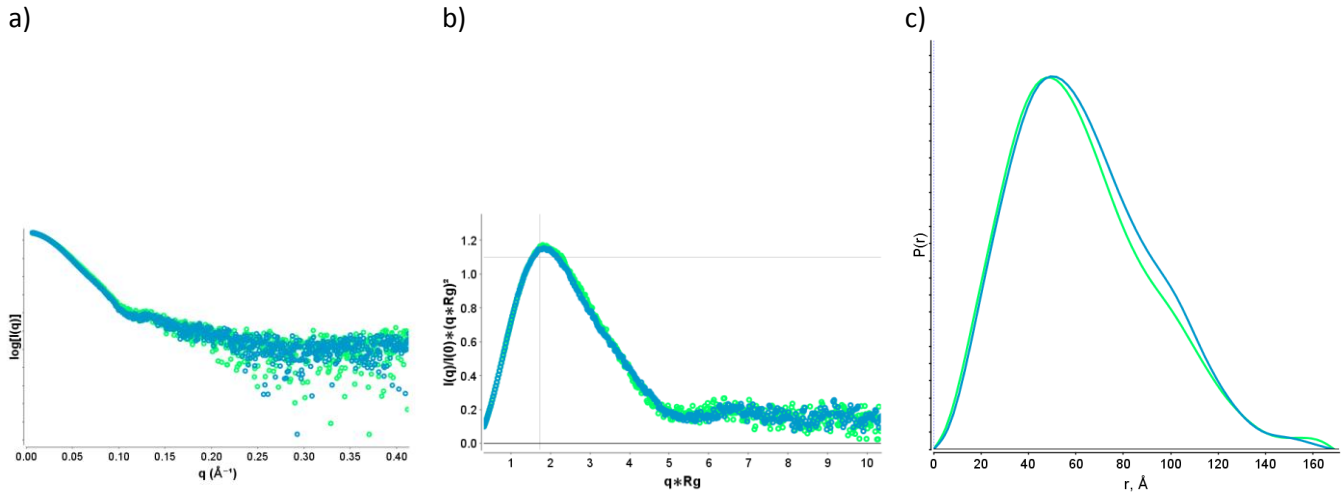
800 µl of E9exo^{minus} (7.8 µM) were mixed with 80 µl annealed primers at 100 µM. The salt was diluted to 75 mM and the sample then concentrated to 60 µl (~ 8-10 mg/ml protein, DNA 1.2 x).

<u>Concurrent SAXS analysis of the first peak:</u>

The analysis of the first peak representing the E9 exo^{minus} +DNA complex using classical frame selection and EFA analysis shows that in this case there is no significant improvement of the scattering signal using deconvolution.



Supplementary Figure 2: UV spectra from SEC-SAXS of the E9 exo^{minus} with (blue) versus without (orange) DNA (data not shown) showing that the second peak corresponds to E9 exo^{minus} unbound form.



Supplementary Figure 3: a) I versus q curve, b) Kratky plot, c) P(r) of E9 exo^{minus} +DNA. The blue curve represents the analysis using EFA and the green curve represents the analysis without EFA. The R_gs differ slightly (49.1Å versus 48.4 Å in real space).