

RESEARCH ARTICLE

An automated interface for sedimentation velocity analysis in SEDFIT

Peter Schuck^{1*}, Samuel C. To, Huaying Zhao

Laboratory of Dynamics of Macromolecular Assembly, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, Maryland, United States of America

* schuckp@mail.nih.gov

Abstract

Sedimentation velocity analytical ultracentrifugation (SV-AUC) is an indispensable tool for the study of particle size distributions in biopharmaceutical industry, for example, to characterize protein therapeutics and vaccine products. In particular, the diffusion-deconvoluted sedimentation coefficient distribution analysis, in the software SEDFIT, has found widespread applications due to its relatively high resolution and sensitivity. However, a lack of suitable software compatible with Good Manufacturing Practices (GMP) has hampered the use of SV-AUC in this regulatory environment. To address this, we have created an interface for SEDFIT so that it can serve as an automatically spawned module with controlled data input through command line parameters and output of key results in files. The interface can be integrated in custom GMP compatible software, and in scripts that provide documentation and meta-analyses for replicate or related samples, for example, to streamline analysis of large families of experimental data, such as binding isotherm analyses in the study of protein interactions. To test and demonstrate this approach we provide a MATLAB script mISEDFIT.

OPEN ACCESS

Citation: Schuck P, To SC, Zhao H (2023) An automated interface for sedimentation velocity analysis in SEDFIT. *PLoS Comput Biol* 19(9): e1011454. <https://doi.org/10.1371/journal.pcbi.1011454>

Editor: Susumu Uchiyama, Graduate School of Engineering, JAPAN

Received: May 15, 2023

Accepted: August 22, 2023

Published: September 5, 2023

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: SEDFIT version 16.50 equipped with the command line interface, as well as the source code of an accompanying family of MATLAB scripts mISEDFIT can be freely downloaded from sedfitsedphat.nibib.nih.gov/ software. Example input and output files can be downloaded from sedfitsedphat.nibib.nih.gov/ tools. All data can be downloaded from the Harvard Dataverse <https://doi.org/10.7910/DVN/CZDWRZ>.

Funding: This work was funded through the National Institute of Health (ZIA EB000051-12 to PS). The funders had no role in study design, data

Author summary

Sedimentation velocity analytical ultracentrifugation (SV-AUC), a classical first-principles based method to study size-distributions of macromolecules and particles in solution, has become a popular technique in a variety of disciplines, such as physical biochemistry, structural biology, supramolecular chemistry, and nanoparticles due to its high resolution, wide size-range, and label-free detection. It has also assumed an important role in pharmaceutical development of therapeutics and vaccines. One factor complicating the implementation of SV-AUC in biotechnology is the lack of compatibility of the most widely used software, SEDFIT, with regulatory requirements in the good manufacturing practices (GMP) environment. In this paper we present a solution to this problem, by introducing a command line interface that permits automated configuration of SEDFIT, data loading, and retrieval of results. In this way, SEDFIT may be used as a computational module integrated in GMP-compliant software. We demonstrate the validity of the results from this approach. In addition, this automation facilitates meta-analyses of families of SV-AUC experiments, for example, in binding isotherm analysis of interacting macromolecules.

collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Sedimentation velocity analytical ultracentrifugation (SV-AUC) is a classical first-principles based method to study particles that sediment or float as a result of a gravitational field generated in a centrifuge [1–3]. In the last two decades, this technique has been used in a growing number and types of applications, coinciding with a significant advance in the computational data analysis [4]. Specifically, the combination of efficient Lamm equation modeling [5], novel size-distribution techniques for direct data fitting with and without diffusional deconvolution [6–8], and systematic noise analysis [9,10] as combined in the $ls-g^*(s)$ and the high-resolution $c(s)$ approach implemented in the software SEDFIT has garnered widespread applications. These include analyses of size-distributions, hydrodynamic properties, and interactions of particles across the entire size-range accessible to SV-AUC from below 1 kDa to above 10 GDa, ranging from small carbohydrates and peptides [11–13] to proteins and protein complexes [14–17], carbohydrates [18], synthetic polymers [19], small and large nanoparticles [20,21], multi-protein complexes and interacting systems [15,22–25], lipid vesicles and emulsions [26,27], viral particles [28–34], and entire cellular organisms [35], at concentrations from picomolar to millimolar [3,36,37]. Due to the high resolution and sensitivity, and the measurement of particle sedimentation free in solution in the absence of surfaces or labels, this approach has also proven to be advantageous in biotechnology for the characterization of therapeutic and vaccine products, for example including antibody and other protein therapeutics [38–47], protein/polymer conjugates [48,49], and AAV products [28–32].

One of the challenges of applying SV-AUC in the context of biopharmaceutical production is the lack of software compliant with current good laboratory practices (cGLP) and current good manufacturing practices (cGMP) that satisfies part 11 of Title 21 of the Code of Federal Regulations; Electronic Records; Electronic Signatures (21CFR11) by FDA [50]. Other existing limitations are the required expertise and the laboriousness of serial analyses, for example, of replicate sets or of titration series in the study of interacting systems [43,44,51]. Both would greatly benefit from some degree of automation and streamlining of the data analysis. For this purpose, the present work presents an extension of the SEDFIT software for partially automated operation through the use of a command line interface. The interface allows starting SEDFIT in a specific state, as well as communicating data, analysis parameters, and results. Any secondary software utilizing this interface can spawn SEDFIT, automatically load specific data, execute a specific model for analysis, and retrieve the best-fit results of SEDFIT for documentation, further aggregation, and to carry out meta-analyses of multiple sedimentation coefficient distributions. While this allows streamlined SV-AUC analysis of replicates and titration series, importantly, this approach will be particularly suitable for preserving custody of data and associated analysis results and generating audit trails consistent with 21CFR11.

As illustrated in the present work, the results from such command-line controlled analyses are equivalent to the standard entirely manual use of SEDFIT. Due to the universality of SV-AUC, this approach can be applied to the analysis of therapeutic proteins, carbohydrates, nucleic acids, protein/polymer conjugates, any viral vectors (recombinant or wildtype) including adeno-associated virus (AAV), adenovirus, and lentivirus, lipid vesicles and lipid-based nanoparticles, metal- and other nanoparticles with and without protein conjugates or associated macromolecules such as loaded nucleic acids or proteins.

Method

The general strategy for command line operated SEDFIT is for it to be spawned by a secondary software, initiating a controlled input of data and providing a controlled return of the analysis

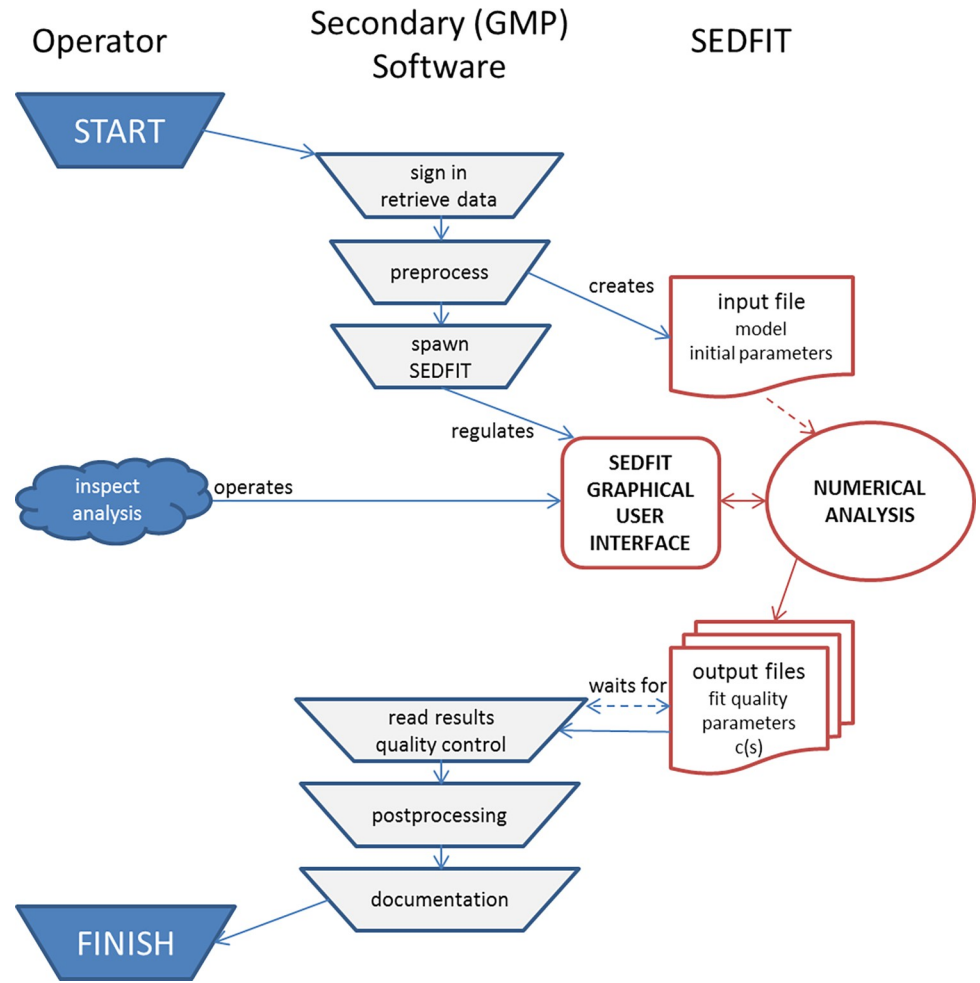


Fig 1. Flowchart for the use of SEDFIT in command line operation with a secondary software. The secondary software organizes access control and preprocesses data. After SEDFIT is spawned by the secondary program, it reads a specifically formatted input file, and provides a graphical user interface with options controlled by the secondary program. Upon termination of SEDFIT analysis, its output files are read, and quality control, postprocessing and documentation by the secondary program can take place. This flow allows a single or multiple copies of SEDFIT to be utilized solely as a computational module within a framework of the secondary program, which may enforce GMP compatibility, incorporate results into meta-analyses, and/or provide an expert system or AI for automated analysis and quality control.

<https://doi.org/10.1371/journal.pcbi.1011454.g001>

results (Fig 1). This will allow logging of all activities and safely record results for further processing. SEDFIT may be also manually executed from the command line, for example, from within the DOS command prompt in Windows, although this would not take advantage of much of the graphical user interface for loading and saving data that is provided in the conventional stand-alone operation of SEDFIT. As a roadmap for the envisioned incorporation of command line operated SEDFIT into secondary software, we have created a set of MATLAB scripts mlSEDFIT, which can be freely downloaded from sedfitsedphat.nibib.nih.gov/software.

The command line-controlled run of SEDFIT requires three command line parameters. The first is a number that identifies the SEDFIT mode of operation, at present with possible values of “111” and “112”. The mode “111” will invoke SEDFIT with a standard menu allowing the full range of SEDFIT analyses models, and “112” will present SEDFIT with a reduced set of menu functions more narrowly tailored to elementary size-distribution analyses. In this

restricted mode, non-standard settings are inaccessible to the operator. Different values of this command line parameter are possible in future implementations with alternate behavior. The second command line parameter is an arbitrary string that serves to identify the SEDFIT call. It will be repeated by SEDFIT after completion of the analysis, and thereby provides a layer of security that the reported results do belong to the requested analysis. The third parameter is the path of the input file. For example, the command line could be "**sedfit.exe 111 handshakestring c:\datafolder\TestInput.xml**" (without the quotes) where 'handshakestring' can be replaced by any string, and 'c:\datafolder\TestInput.xml' should be replaced by the appropriate directory and filename to point to a valid xml input file in the correct format.

An example for an input file is provided in [S1 File](#). For convenience in generating and parsing the input file, it is in xml format version 1.0, as indicated in the first line by the prolog string `<?xml version="1.0" encoding="UTF-8"?>`. This is followed by the root 'cGMPSedfitCall' as written in an initial `<cGMPSedfitCall>` call, which is paired with a final `</cGMPSedfitCall>` statement at the end of the file. In between these opening and closing root identifiers are the analysis parameters. They can be in any order, but must adhere to the xml syntax of `<ParameterName>value string</ParameterName>`, where 'ParameterName' is any of the identifiers described below, and 'value string' can be a number, a file path, the string TRUE, FALSE, or other string, as appropriate for the parameter. A complete list of all parameters, their purpose and possible values can be found in [Table 1](#). Unless the menu of SEDFIT is restricted (in mode "112", see above), all parameters can also be set later as usual in the SEDFIT user interface by the operator.

While most of the input parameters are strictly related to the data analysis, some are related to the SEDFIT interface and analysis flow. Most importantly, the parameter **AllDoneFlagFile** will establish the path and name of the file that will be created upon completion of the SEDFIT analysis. For example, it might be called 'c:\datafolder\finished.txt'. Importantly, this file must not exist prior to the SEDFIT call, such that its creation can be used as a convenient flag to the secondary software that SEDFIT has created the results files and was exited. The only content of this file will be the handshake string (second command line parameter, see above), which may be checked by the secondary software for consistency with the calling parameter. The parameter **OutputResultsDirectory** is a path that points to the directory where SEDFIT will create several output files describing the results after conclusion of the analysis, as detailed below. For example, the string 'c:\datafolder\analysisresults\results1' will cause SEDFIT to create the subfolder 'results1' within the existing 'c:\datafolder\analysisresults' directory, and all results files will be located in this 'results1' subfolder. Finally, it is possible to use a parameter **PassThrough** to declare any string that, for convenience and clarity, should be repeated in the output xml file. For example, this could be a sample name, user notes, or some additional control strings to be recorded alongside the results. It may also be itself an xml-formatted string, but in this case it should not contain any regular parameter identifiers used in the SEDFIT interface to avoid conflicts.

To define the data to be analyzed, the input file must specify the folder containing the SV-AUC scan files in the parameter **DataDirectory**, the parameter **Channel** that identifies the scan file extension, and the parameters **FirstScan**, **LastScan**, and **ScanInterval** that specify the set of scans to be loaded. It is assumed that the scan files are in the customary SV-AUC numeric format with leading zeros. This information substitutes for manually invoking the Load New Files function in the Data menu of SEDFIT. For example, if **DataDirectory** is 'c:\datafolder\Run158', **Channel** is 'IP2', **FirstScan** is '1', **LastScan** is '101', and **ScanInterval** is '10', SEDFIT will load 11 files named 'c:\datafolder\Run158\00001.ip2', 'c:\datafolder\Run158\00011.ip2', 'c:\datafolder\Run158\00021.ip2', . . . , 'c:\datafolder\Run158

Table 1. List of Input Parameters. Input parameters will be read from a file named as third command line parameter. It is in xml format and parameters are case sensitive. An example can be found in [S1 File](#).

Parameter Name	Values	Purpose
AllDoneFlagFile	path to non-existent file in existing directory	creation of this file will serve as a flag that SEDFIT analysis has concluded
OutputResultsDirectory	path to non-existent directory	folder location where results files will be saved
PassThrough	arbitrary string	for additional information, such as sample name, to be repeated in output file; can be xml formatted
DataDirectory	path to data folder	for accessing scan files
Channel	file extension of scan file	defines cell number and scan type
FirstScan	integer	first scan to be loaded
LastScan	integer	last scan to be loaded
ScanInterval	integer	interval of scans to be loaded
FilterDataSpikes	TRUE/FALSE	switch to filter adventitious spikes in scan files
DataSpikeThreshold	floating point number	threshold for recognition of isolated data spikes
Meniscus	floating point number	air/solution interface distance from center of rotation, in cm
MeniscusLowerLimit	floating point number	lower limit of meniscus during fit
MeniscusUpperLimit	floating point number	upper limit of meniscus during fit
Bottom	floating point number	distal end of the solution column, in cm from center of rotation, in cm
BottomLowerLimit	floating point number	lower limit of bottom position during fit
BottomUpperLimit	floating point number	upper limit of bottom position during fit
LeftFitLimit	floating point number	minimum radial position of the data range included in the analysis, in cm
RightFitLimit	floating point number	maximum radial position of the data included in the analysis, in cm
AutoRun	TRUE/FALSE	switch to automatically call the "Run" command after loading data
AutoFit	TRUE/FALSE	switch to automatically call the "Fit" command after loading data and optional "Run"
AutoSubtractSystematicNoise	TRUE/FALSE	switch to subtract all best-fit systematic noise components from the data in the scan file display of the SEDFIT window
ShowResidualsHistogram	TRUE/FALSE	switch to display the residuals histogram next to the residuals overlay in the SEDFIT window
Model	'cofs' or 'lsgofs' (no quotation marks)	model selection
Resolution	integer	number of grid points in the distribution
Smin	floating point number	minimum <i>s</i> -value of the distribution
Smax	floating point number	maximum <i>s</i> -value of the distribution
GridFromFile	TRUE/FALSE	switch to read the distribution grid from "sdist" file
UseLogSpaceSgrid	TRUE/FALSE	switch to abandon equidistant spacing of grid points in favor of logarithmic spacing
RegularizationType	'maxent' or 'Tikhonov' (no quotation marks)	defines the regularization used in the distribution analysis
RegularizationPvalue	floating point number	p-value for scaling the extent of regularization, between 0.5 and 1.0
SuppressBaselineCorrelation	TRUE/FALSE	switch to use Bayesian prior to reduce correlation of smallest <i>s</i> -value in the distribution with baselines
NumberComputationThreads	integer	number of computing cores available for SEDFIT
FittingAlgorithm	'Simplex' or 'Levenberg-Marquardt' (no quotation marks)	choice of algorithm for non-linear regression
StartingFrictionalRatio	floating point number	frictional ratio value to be used for <i>c(s)</i> analysis
FrictionalRatioFitted	TRUE/FALSE	switch to optimize the frictional ratio in non-linear regression
MeniscusFitted	TRUE/FALSE	switch to optimize meniscus position
BottomFitted	TRUE/FALSE	switch to optimize bottom position
BaselineFitted	TRUE/FALSE	switch to optimize radial- and time-invariant baseline
RINoiseFitted	TRUE/FALSE	switch to optimize radial-invariant baselines
TINoiseFitted	TRUE/FALSE	switch to optimize time-invariant baselines

(Continued)

Table 1. (Continued)

Parameter Name	Values	Purpose
Vbar	floating point number	partial-specific volume to be used in hydrodynamic scaling law to scale diffusion coefficient from frictional ratio in $c(s)$ analysis, in mL/g
BufferDensity	floating point number	density of solvent, in g/mL
BufferViscosity	floating point number	viscosity of solvent, in g/mL

<https://doi.org/10.1371/journal.pcbi.1011454.t001>

\00101.ip2'. Scan data not conforming to the assumed filename convention should be renamed accordingly by the secondary software prior to spawning SEDFIT. The parameters **FilterDataSpikes** and **DataSpikeThreshold** will control whether to ignore isolated spikes in the scan data and the threshold for defining a spike, analogous to the corresponding Loading Options and Tools function of the Options menu of SEDFIT. If the parameters relating to spike filtering are not set then the default filtering with a threshold of 0.4 will be applied.

The input parameters related to the SEDFIT workflow include **AutoRun** and **AutoFit**, both of which take Boolean values TRUE or FALSE. If TRUE they will automatically initiate the SEDFIT Run or Fit command, respectively, directly after loading the data. This will be equivalent to manually invoking these commands from the SEDFIT menu. (It should be noted that when using **AutoFit** the analysis is carried out before the SEDFIT window appears on the screen.) To enhance the performance of SEDFIT, a parameter **NumberComputationThreads** can be set according to the computational cores that SEDFIT should employ for multithreaded computations. In $c(s)$ distribution analyses in SEDFIT, key computational steps will be approximately n -fold faster if $n > 1$ threads are specified and available [52]. While the default is 2, the optimal value will depend on the computer hardware and on concurrently running software, which may be other instances of SEDFIT. The best choice for n is usually the total number of processing cores of the computer used. Finally, parameters related to the graphical data presentation are **ShowResidualsHistogram** and **AutoSubtractSystematicNoise**, which will take Boolean values TRUE or FALSE and control the visual appearance of the SEDFIT analysis window as in their corresponding SEDFIT menu functions [9,10,53]. In addition to changing the graphical display of the SEDFIT window during data analysis, an image of this window will be saved for documentation at the end of the analysis. Scan data, best-fit values, and residuals will also be saved in an output file, so that the graph showing the quality of fit can be recreated (see below).

Ancillary parameters needed for the analysis are the estimated locations of the meniscus and bottom of the solution column, and their fitting limits. **Meniscus** and **Bottom**, as well as **LeftFitLimit** and **RightFitLimit**, are specified as numerical values in cm from the center of rotation, and will function equivalently to their customary graphic input in SEDFIT. During the SEDFIT analysis, they can still be readjusted by the operator or by the fitting routine. If the parameter **MeniscusFitted** is set TRUE, the meniscus value will be optimized during the fit, remaining within the bounds specified in the **MeniscusLowerLimit** and **MeniscusUpperLimit** parameters (which should bracket the **Meniscus** value and be smaller than **LeftFitLimit**). Fitting for the meniscus is usually recommended, and therefore coarse initial estimates may suffice, for example, from a table of expected solution column heights dependent on filling volume in standard double-sector cells provided in [3], from a simple identification of the meniscus artifact in scan files, or from a preliminary analysis. Analogous parameters are available for fitting the bottom of the solution column (Table 1), however, unless significant back-diffusion is affecting the sedimentation process, typically the bottom position does not need to be fitted in standard analysis and the corresponding value for **BottomFitted** would be FALSE.

Most importantly, the parameter **Model** specifies the data analysis model. It currently can take string values of 'lsgofs' for the $ls-g^*(s)$ model [7,8] and 'cofs' for the $c(s)$ model [6,7]. Specifying these models is equivalent to their selection in the Model menu of SEDFIT. Both are sedimentation coefficient distribution models that require definition of the discrete grid of s -values through **Smin**, **Smax**, and **Resolution**. Analogous to the manual entry of these parameters in the model parameter boxes of SEDFIT, they take numerical values describing the range of the distribution and the number of grid points. Alternatively, if the parameter **GridFromFile** is set **TRUE**, the grid of s -values can be read from a file 'sdist' with the same extension as the scan files (such as 'sdist.ip2') that must be located in the same folder as the scan data. This text file will automatically be created after each analysis and contains a single column of s -values of the distribution grid, but it can also be edited and augmented and serve as a template for new distribution analyses with custom-spaced grid, for example, to efficiently describe very small or very large sedimenting species outside the range of the majority particles of interest. Another option to modify the grid is **UseLogSpaceSgrid**, which when **TRUE** abandons equidistant grids in favor of logarithmically increasing s -value intervals. When using the command line interface these grid functions operate identically with their common stand-alone use of SEDFIT.

The $c(s)$ analysis applies diffusional deconvolution to achieve sedimentation coefficient distributions with high hydrodynamic resolutions. When using the SEDFIT command line interface, the most commonly used standard $c(s)$ variant is applied where the diffusion coefficients associated with each sedimentation coefficient are based on a hydrodynamic scaling law *via* a constant hydrodynamic frictional ratio. Thus, the parameter **StartingFrictionalRatio** must be set. The numerical values of the predicted diffusion coefficients also depend on the partial specific volume **Vbar** (to be specified in mL/g) and the buffer density **BufferDensity** (in g/mL). During the analysis, non-linear regression will be used to optimize the frictional ratio during the fit when **FrictionalRatioFitted** is **TRUE**. It should be noted that, unless the final frictional ratio is to be interpreted quantitatively, the partial specific volume values can be rough estimates. The distribution analysis also requires regularization to avoid spurious peaks and error amplification [6,52], which is set through the parameter **RegularizationType**. It currently can take values 'maxent' for maximum entropy regularization or 'Tikhonov' for Tikhonov-Philips regularization [6]. As is standard in SEDFIT, the tolerated increase in the root-mean-square deviation (rmsd) of the fit allowed for regularization is scaled by F-statistics and a p-value specified in the parameter **RegularizationPvalue** [7]. Since sedimentation patterns of very small particles are very similar to baseline offsets, a correlation between baselines and distribution values at small s -values can exist. As shown previously, this correlation can be suppressed through a Bayesian prior at the smallest s -value [54], and this is specified by setting the parameter **SupressBaselineCorrelation** to **TRUE**.

Lastly, details of the fit can be specified through the command line interface. The fitting algorithm is chosen by setting the parameter **FittingAlgorithm** to either 'Simplex' or 'Levenberg-Marquardt'. In addition to the above mentioned fitting of meniscus and bottom of the solution column and the frictional ratio, the treatment of baselines can be specified in the parameters **BaselineFitted**, **RINoiseFitted**, and **TINoiseFitted**. If set to **TRUE** this will cause a spatio-temporally uniform baseline, a time-dependent baseline, and/or a radial-dependent baseline to be fitted during the nonlinear regression, respectively [9,10]. It should be noted that even when the SEDFIT operation is set to automatically execute a Run and Fit command, these procedures can be interrupted and/or manually executed by the SEDFIT operator as usual. Even in the reduced SEDFIT menu it is possible to readjust solution column parameters and other fitting parameters to achieve the best fit prior to concluding the analysis. The

usual side effect of the SEDFIT analysis is the creation of the above mentioned 'sdist' file containing the s -value grid of the distribution analysis, as well as a file named `~tmppars` that, when manually reloading the scan files will allow to restore the last best-fit analysis. Both files are located in the data directory and have the same file extension as the scan files loaded.

When the operator invokes the Exit function of SEDFIT, several files are created prior to the termination of the SEDFIT process and placed in the designated output folder defined in the **OutputResultsDirectory** parameter. The first is a bitmap image of the SEDFIT window, saved as 'screenshot.bmp'. Besides the graphical display of the scan file overlay and fit, residuals overlay, distribution, and optional residuals bitmap, it shows the customary informational text that provides information about the data files and fitting parameters including the overall rmsd of the fit. Further, it creates text files 'RInoise.dat' and 'TInoise.dat' which contain two columns with the best-fit time- and radial-dependent baseline values. A file 'ScanRMSD.dat' provides information about the rmsd of the fit to each scan file separately. This may help to recognize trends or outliers. The file 'distribution.dat' contains two columns with the distribution in form of grid s -values vs. $c(s)$ values. This distribution file can later be integrated used for further analysis in the secondary software. A file 'dfr.dat' saves the fitted boundary data for recreation of the data and fit overlay plot, and a residuals plot or bitmap. It is an ASCII text file in the form of a matrix with columns: radius (TI noise), TI noise, RI noise, radius (scan 1), raw data (scan 1), fit value (scan 1), radius (scan 2), raw data (scan 2), fit value (scan 2), etc., and rows corresponding to consecutive radii or scan times, respectively. Finally, SEDFIT creates an output file 'ResultParameters.xml' in the same xml format as the input file and containing the same parameters, some of which may have changed due to adjustments during the fit or by operator actions. Such changes may be registered in the spawning program. In addition, it reports the SEDFIT version, file paths of all input SV-AUC scan data files and all output files, as well as the **PassThrough** parameter.

As statistical measures of the quality of fit the output file reports the overall rmsd (**RMSD**), the number of data points fitted (**RMSD-points**), the sum of squared residuals (**RMSD-SSR**), the runs test Z-value (**RunsTestZ**), and the histogram H (**HistogramH**) [7,53]. Additional information about the data include whether scan file time stamps could be accessed for correction (**CheckTimeStamps**) [55], the rotor speed (**RotorSpeed**), the time and accumulated $\omega^2 t$ value of the last scan (**tLastScan** and **w2tLastScan**), the rotor temperature at the time of the first and last scan (**TemperatureStart** and **TemperatureEnd**), as well as the temperature average and largest temperature difference (**TemperatureAverage** and **TemperatureDiffMax-Min**). These parameters may be used for experimental quality control to flag the possible presence of convection artifacts [56,57]. A full list of the output parameters can be found in [Table 2](#).

To indicate termination of the SEDFIT analysis and to allow hand-over to the secondary software, the **AllDoneFlagFile** is created. As mentioned above, this ASCII text file contains as sole entry the handshake string from the command line starting SEDFIT. Creation of this file also indicates that the other results files have been created, which will not be the case if SEDFIT is prematurely exited.

In summary, to access the SEDFIT command line interface any secondary software must carry out the following tasks ([Fig 1](#)): 1) organize data such as scan files and starting analysis parameters; 2) generate an input xml file that contains the desired SEDFIT completion flag file (which must not exist yet) and designate the directory for results files; 3) execute SEDFIT with the command line parameters, including a handshake string; 4) wait for completion of the analysis by periodically checking for the creation of the specified completion flag file containing the handshake string; 5) read the results from the xml and other output files created by SEDFIT; 6) perform optional quality checks, optional integration of the distribution results,

Table 2. List of Additional Output Parameters. Output will be written in an xml formatted file in the designated output folder. Output parameters include the same parameters regarding data, model, and solution conditions as the input parameters, but also include the additional parameters in this table.

Parameter Name	Values	Purpose
SEDFITVersion	string	document the SEDFIT version
ScreenShotPath	path to screenshot.bmp file	documented screenshot at end of analysis
CofsDataPath	path to distribution.dat file	two-column ASCII text file with the best-fit sedimentation coefficient distribution for further analysis
TINoisePath	path to TInoise.dat file	two-column text file with TI noise profile as function of radius to aid in graphical display of scan files
RINoisePath	path to RInoise.dat file	two-column text file with RI noise as function of time to aid in graphical display of scan files
ScanRMSDpath	path to ScanRMSD.dat file	two-column text file with local rmsd for each scan file, to detect outliers and trends
ScanDataFilesLoaded	list of paths	document all the scan files that were fitted
PassThrough	arbitrary string	for additional information, as specified in the input parameter file
CheckTimeStamps	TRUE/FALSE	for documentation whether time stamps of the scan files were accessible for scan time correction
RotorSpeed	integer	rpm of the experiment, as read from scan files
Wavelength	floating point number	wavelength entry read from the scan files
w2tLastScan	floating point number	$\omega^2 t$ entry read from last scan file
tLastScan	floating point number	time entry read from last scan file
TemperatureStart	floating point number	temperature entry read from first scan, to document temperature discrepancies as a possible source of convective artifacts
TemperatureEnd	floating point number	temperature entry read from last scan
TemperatureAverage	floating point number	average temperature entries from all scans
TemperatureDiffMax-Min	floating point number	range of temperature variation during the SV run
FittingStepsText	string	top line of the SEDFIT fitting information display, containing detailed information about the algorithm and steps during non-linear regression
RMSD	floating point number	final rmsd of the fit
RMSD-points	integer	total number of data points fitted
RMSD-SSR	floating point number	sum of squared residuals of the fit
RunsTestZ	floating point number	result of the runs test for statistics of runs of positive and negative residuals, as a measure for randomness of residuals
HistogramH	floating point number	statistical analysis of the similarity of the residuals histogram to a Gaussian
FrictionalRatio	floating point number	best-fit frictional ratio after $c(s)$ analysis

<https://doi.org/10.1371/journal.pcbi.1011454.t002>

carry out secondary analyses, and/or write automated reports. These tasks can be wrapped in access controlled environment, as necessary in the GMP setting. To allow efficient analyses of a large number of samples, the SEDFIT interface can be run in multiple instances side-by-side; if care is taken to create unique **AllDoneFlagFile** files and **OutputResultsDirectory** locations, different SEDFIT instances will operate completely independently of each other.

Results

To test the command line interface we wrote a family of MATLAB scripts for data input and retrieval of results, termed ‘mlSEDFIT’. The script may be taken as a template for further modification. We applied it to the analysis of stressed NISTmAb monoclonal antibody [58] that is partially denatured and presents a series of oligomeric populations and a multimodal sedimentation boundary (Fig 2). Using the **AutoRun** and **AutoFit** option, the non-linear regression converges at an rmsd of 0.006743 OD (Fig 2, solid lines) with a best-fit meniscus at 6.1657 cm, a best-fit frictional ratio of 1.37, and the $c(s)$ distribution shown in Fig 3.

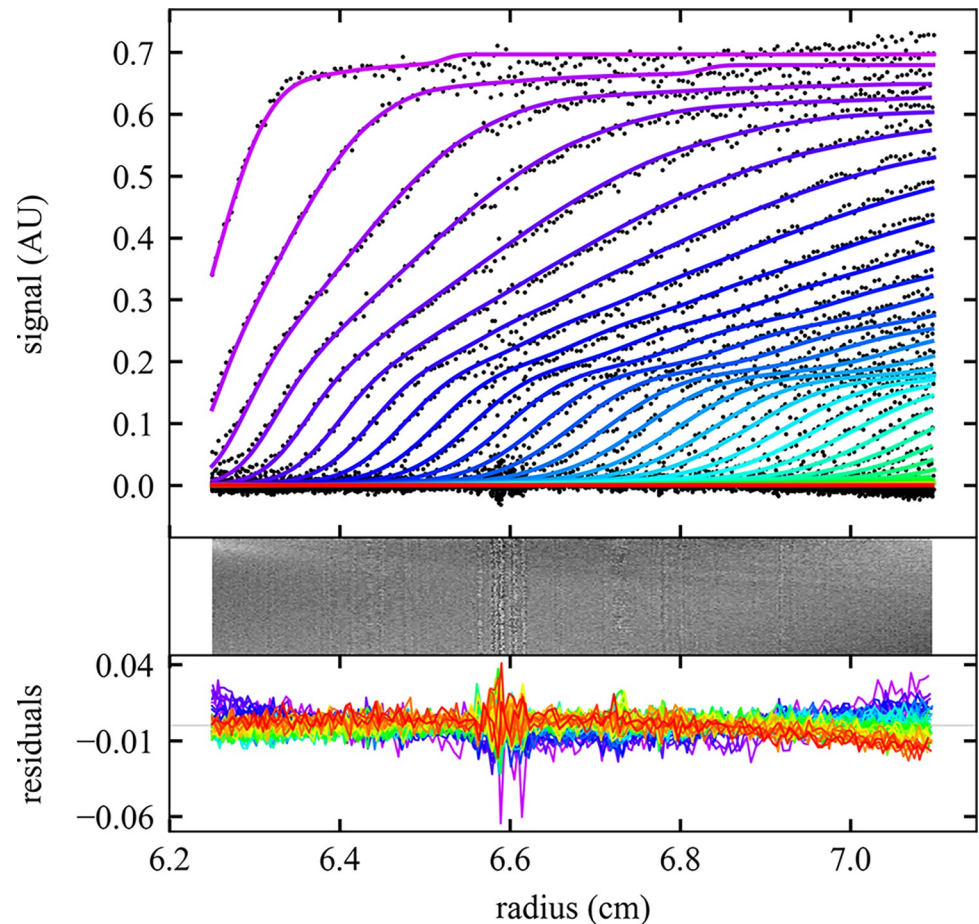


Fig 2. Sedimentation analysis of a stressed NISTmAb sample at 50,000 rpm and 20°C using the command line operation of SEDFIT. Top: Scan files and best fit (for clarity, showing black dots only for every 2nd data point of every 2nd scan) with a $c(s)$ model automatically converged to a final rmsd of 0.006743 OD (colored lines). Progression of scan time is indicated by color from purple to red. Middle and Bottom: Residuals bitmap and residuals overlay. Plot was made using the software GUSSE [59], which is spawned from the script mlSEDFIT.

<https://doi.org/10.1371/journal.pcbi.1011454.g002>

It is possible to manually reload the command line generated analysis and inspect the fit further. However, as an independent control we loaded the same data separately in standard operation of SEDFIT and performed an analysis with the same model. It converged to an rmsd of 0.006738 OD, with a best-fit meniscus at 6.1655 cm, a best-fit frictional ratio of 1.35, and a $c(s)$ distribution that is virtually identical in all aspects to the distribution from command line operated SEDFIT analysis (Fig 3). Differences of < 0.01 S in sedimentation coefficients and $< 0.1\%$ in population for all peaks are observed, which is better than the typical accuracy and statistical precision of SV-AUC analysis.

Generally, one practical limitation in comparing analysis results can be the required graphical input of the fitting limits, which is replaced by numerical control in the command line parameters. Similarly, detailed loading option preferences may differ in the two operation modes. Furthermore, small and insignificant numerical differences in repeat analyses should be expected with the Simplex algorithm for fitting, since this involves initial randomization of fitting parameters. Small differences may also be found when adopting different paths in the error surface during non-linear regression. In the present case, remaining insignificant differences in the fit are a result of a locally very flat error surface for the precise numerical value of

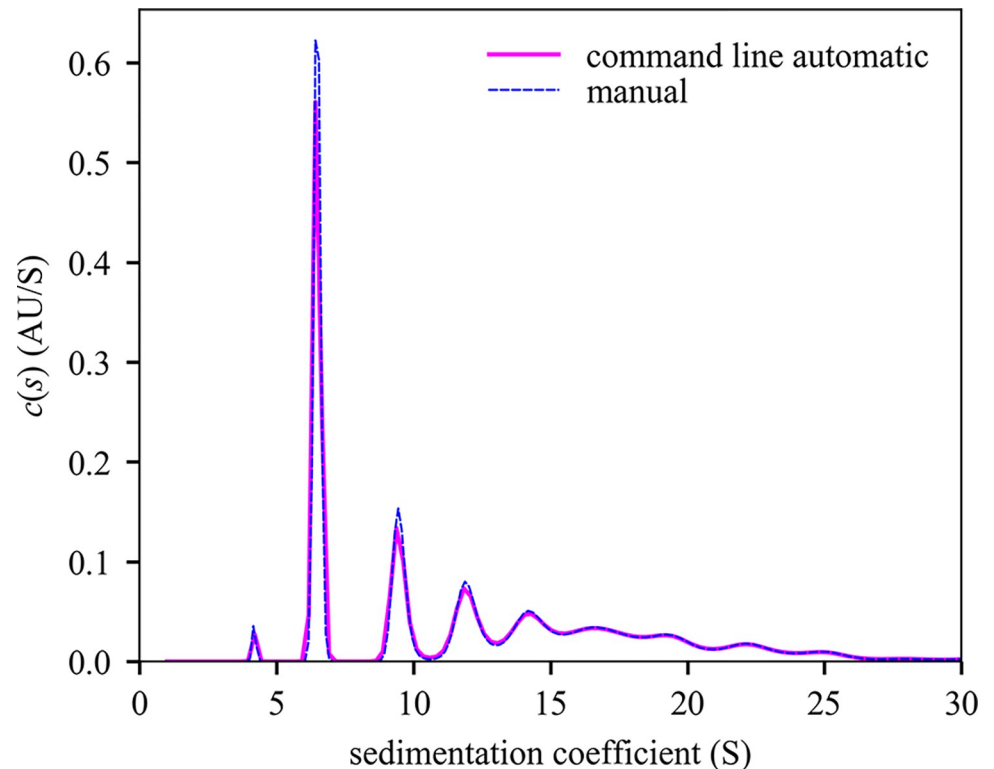


Fig 3. Comparison of $c(s)$ distributions computed with the command line initialization of SEDFIT and with manual operation. The distribution from command line operation (Fig 2), and exhibits a monomer peak at 6.477 S with 29.20% of signal, a trace degradation product at 4.199 S with 0.95% of signal, a dimer peak at 9.473 S with 12.51% of signal, and higher aggregates with collective s_w 16.799 S and 51.77% of signal. The analogous manually operated analysis producing a monomer peak at 6.481 S with 29.27% of signal, a degradation product of 4.178 S with 0.92% of the signal, a dimer peak at 9.488 S with 12.51% of signal, and higher aggregates with collective s_w of 16.81 S with 51.73% of signal. Integration and plot were made using the software GUSSE [59], which can be spawned from the script mlSEDFIT.

<https://doi.org/10.1371/journal.pcbi.1011454.g003>

the frictional ratio parameter, which ultimately reflects the limit of diffusion information content of the broad sedimentation boundaries of the sample dataset used.

The script mlSEDFIT for testing and demonstrating the new interface is equipped with functions for automatically pre-determining an estimate for the meniscus position for absorbance data prior to spawning SEDFIT, and for creating high-quality illustrations by spawning the software GUSSE [59] utilizing the data, fit, and residuals values retrieved from SEDFIT after the data analysis. In addition, it can integrate the distributions alternatively through a graphical process or through pre-determined integration limits (Fig 4). Finally, it will save the results alongside the SEDFIT output parameters. mlSEDFIT can be easily customized and extended, and may be compiled to prevent further modification.

Discussion

SV-AUC has become an indispensable tool to study particle size distributions in science and biopharmaceutical industry [38–40,56,60,61]. Therefore, implementation of SV-AUC analyses in the GMP environment would be desirable. The lack of SV-AUC analysis compatible with the GMP environment was previously discussed by Savelyev and colleagues [62]. Their software ULTRASCAN GMP provides data access and analysis workflow control, but unfortunately the SV-AUC data analysis in ULTRASCAN includes *ad hoc* algorithms that are

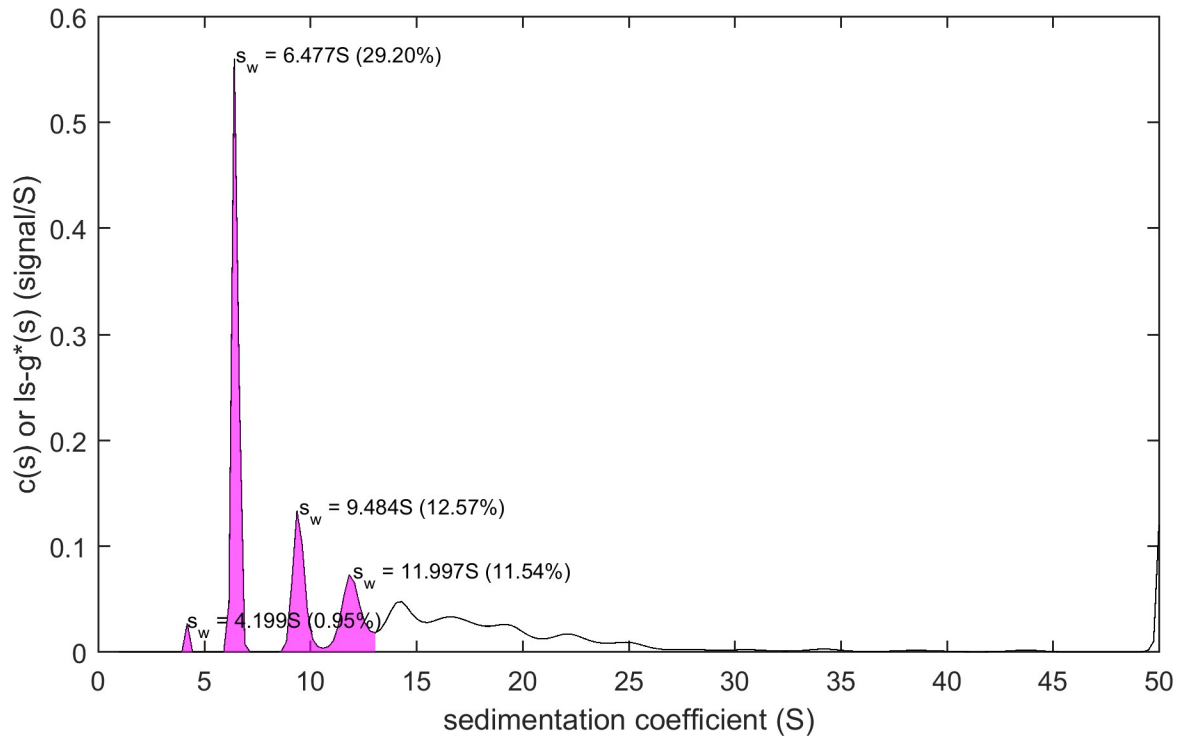


Fig 4. Example for postprocessing of results from SEDFIT analysis in mlSEDFIT. The output generated through the command line interface can be read in the mlSEDFIT script. For example, integration of distribution peaks can be carried out in this script after mouse clicks on the peaks in the distribution plot, as shown.

<https://doi.org/10.1371/journal.pcbi.1011454.g004>

mathematically uncertain in important aspects, and computationally excessively wasteful requiring supercomputers [52]; therefore it has only a minor share in SV-AUC applications in biopharmaceutical applications, the vast majority of which are carried out with the $c(s)$ and $|s-g^*(s)$ methods implemented in SEDFIT [7,41]. Furthermore, ULTRASCAN GMP presents a closed system with pre-conceived workflow strategies and analyses that may not be suitable or adaptable to different objectives. Moreover, it is linked to a particular version of the analytical ultracentrifuge instrument, of which deficiencies for certain applications have been reported [56], and which is currently incapable of providing independent operating system time-stamps of the scan files to the analyst to verify time accuracy [55].

The availability of the computational interface for SEDFIT provides a computational core for flexible state-of-the-art SV-AUC analysis as a module that can be easily embedded into scripts and software satisfying GMP requirements, including auditable analysis trails and custody of data and results. For demonstration and customization, a generic MATLAB script for spawning SEDFIT was developed, and similar access could conceivably be incorporated into user-friendly software such as GUSI [59], or into ULTRASCAN GMP [62] or other custom-written GMP software. The extent to which 21CFR11 requirements are met will depend on the spawning software, and the SEDFIT spawning mode chosen.

At present, the analysis still needs to be supervised, since manual adjustments to the fitting parameters and model may be required to arrive at the best-fit analysis. This allows adventitious scan files or other possible artifacts from experimental imperfections to be recognized and their effect to be alleviated. Detailed protocols and instructions can be found in the literature [3,7,41,47,63–65], and for reliable results this guidance is equally valid when using the command line interface. Nonetheless, the interface described here can provide a platform for

future improvements that conceivably may allow fully unsupervised analyses by expert systems or AIs with automated meniscus recognition, adjustment of fitting limits, and judgment of fit quality, which may be applied, for example, to replicate experiments. For series of equivalent experiments, the present version already allows the result (for example, meniscus position or frictional ratio) of an initial analysis to be automatically entered as starting parameters for a following data set, thereby improving the efficiency of the analysis. Besides the GMP environment, this may be useful for analyzing large families of experimental data sets designated for meta-analyses, such as collective integration of distributions for binding isotherms and their analysis [23,51,65]. This ties in with developments of higher throughput experimental techniques, such as pseudo-absorbance data acquisition without need of a reference sector [66] and 3D-printed multi-sector centerpieces [67,68].

Another area for future expansion of the SEDFIT interface is the extension to allow pre-selection of more sophisticated analysis models. For example, biopharmaceutical samples often contain small co-solutes that sediment and create dynamic density and signal gradients, both of which can be taken into account when analyzing macromolecular sedimentation [69–71]. Similarly, different sedimentation configurations, such as analytical zone or band sedimentation can be highly desirable for certain applications [31,34,72,73], and advanced regularization methods may be advantageous [7,74,75]. While the corresponding analysis methods are currently available if SEDFIT is spawned in the unrestricted mode, future releases may allow passing their relevant parameters directly through the command line interface.

Importantly, since none of the computational functions from SEDFIT have been altered, the results will remain the same as in the equivalent standard operation of SEDFIT. The command line interface solely modifies the data input and output, replacing manual startup and loading of analysis files with automated pre-loaded SEDFIT. For this reason, the command line mode of SEDFIT will be applicable to the same range of current and future applications. With regard to the biopharmaceutical industry this includes studies of therapeutic peptides and proteins, polymer conjugates, nucleic acids, carbohydrates, vectors for therapeutics or vaccines based on metal nanoparticles, lipid nanoparticles, viral vectors such as adenovirus, AAV or lentivirus, and others. More generally, due to the universal nature of buoyant mass-based separation in SV-AUC and the high sensitivity and hydrodynamic resolution of $c(s)$ analysis in SEDFIT, it will be applicable to study mass- and size-distributions of macromolecules and particles that differ in density from that of the formulation buffer across a mass range from 1 kDa to >10 GDa, or a sedimentation coefficient range between 0.1 and 100,000 S [76,77].

Supporting information

S1 File. Example input xml file.
(XML)

Author Contributions

Conceptualization: Peter Schuck.

Funding acquisition: Peter Schuck.

Investigation: Peter Schuck, Samuel C. To, Huaying Zhao.

Methodology: Peter Schuck.

Software: Peter Schuck.

Writing – original draft: Peter Schuck.

Writing – review & editing: Peter Schuck, Huaying Zhao.

References

1. Svedberg T, Pedersen KO (1940) *The Ultracentrifuge*. London: Oxford University Press.
2. Schachman HK (1959) *Ultracentrifugation in Biochemistry*. New York: Academic Press.
3. Schuck P, Zhao H, Brautigam CA, Ghirlardo R (2015) *Basic Principles of Analytical Ultracentrifugation*. Boca Raton, FL: CRC Press. 302 p.
4. Schuck P (2013) Analytical ultracentrifugation as a tool for studying protein interactions. *Biophys Rev* 5: 159–171. <https://doi.org/10.1007/s12551-013-0106-2> PMID: 23682298
5. Brown PH, Schuck P (2008) A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. *Comput Phys Commun* 178: 105–120. <https://doi.org/10.1016/j.cpc.2007.08.012> PMID: 18196178
6. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* 78: 1606–1619. [https://doi.org/10.1016/S0006-3495\(00\)76713-0](https://doi.org/10.1016/S0006-3495(00)76713-0) PMID: 10692345
7. Schuck P (2016) *Sedimentation Velocity Analytical Ultracentrifugation: Discrete Species and Size-Distributions of Macromolecules and Particles*. Boca Raton, FL: CRC Press. 244 p.
8. Schuck P, Rossmanith P (2000) Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* 54: 328–341. [https://doi.org/10.1002/1097-0282\(20001015\)54:5<328::AID-BIP40>3.0.CO;2-P](https://doi.org/10.1002/1097-0282(20001015)54:5<328::AID-BIP40>3.0.CO;2-P) PMID: 10935973
9. Schuck P, Demeler B (1999) Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. *Biophys J* 76: 2288–2296. [https://doi.org/10.1016/S0006-3495\(99\)77384-4](https://doi.org/10.1016/S0006-3495(99)77384-4) PMID: 10096923
10. Schuck P (2010) Some statistical properties of differencing schemes for baseline correction of sedimentation velocity data. *Anal Biochem* 401: 280–287. <https://doi.org/10.1016/j.ab.2010.02.037> PMID: 20206114
11. Pavlov GM, Korneeva EV, Smolina NA, Schubert US (2010) Hydrodynamic properties of cyclodextrin molecules in dilute solutions. *Eur Biophys J* 39: 371–379. <https://doi.org/10.1007/s00249-008-0394-9> PMID: 19159925
12. Pechar M, Pola R, Laga R, Braunová A, Filippov SK, et al. (2014) Coiled coil peptides and polymer-peptide conjugates: Synthesis, self-assembly, characterization and potential in drug delivery systems. *Biomacromolecules* 15: 2590–2599. <https://doi.org/10.1021/bm500436p> PMID: 24857680
13. Zhao H, Wu D, Hassan SA, Nguyen A, Chen J, et al. (2023) A conserved oligomerization domain in the disordered linker of coronavirus nucleocapsid proteins. *Sci Adv* 9. <https://doi.org/10.1126/sciadv.adg6473> PMID: 37018390
14. Naue N, Curth U (2012) Investigation of protein-protein interactions of single-stranded DNA-binding proteins by analytical ultracentrifugation. *Methods Mol Biol* 922: 133–149. https://doi.org/10.1007/978-1-62703-032-8_8 PMID: 22976181
15. Manna A, Zhao H, Wada J, Balagopalan L, Tagad HD, et al. (2018) Cooperative assembly of a four-molecule signaling complex formed upon T cell antigen receptor activation. *Proc Natl Acad Sci* 115: 201817142. <https://doi.org/10.1073/pnas.1817142115> PMID: 30510001
16. Ebel C (2011) Sedimentation velocity to characterize surfactants and solubilized membrane proteins. *Methods* 54: 56–66. <https://doi.org/10.1016/j.ymeth.2010.11.003> PMID: 21112401
17. Padrick SB, Brautigam CA (2011) Evaluating the stoichiometry of macromolecular complexes using multisignal sedimentation velocity. *Methods* 54: 39–55. <https://doi.org/10.1016/j.ymeth.2011.01.002> PMID: 21256217
18. Grube M, Dinu V, Lindemann H, Pielenz F, Festag G, et al. (2020) Polysaccharide valproates: Structure—property relationships in solution. *Carbohydr Polym* 246: 116652. <https://doi.org/10.1016/j.carbpol.2020.116652> PMID: 32747284
19. Pavlov GM, Knop K, Okatova OV, Schubert US (2013) Star-brush-shaped macromolecules: Peculiar properties in dilute solution. *Macromolecules* 46: 8671–8679. <https://doi.org/10.1021/ma400160f>
20. Bekdemir A, Stellacci F (2016) A centrifugation-based physicochemical characterization method for the interaction between proteins and nanoparticles. *Nat Commun* 7: 13121. <https://doi.org/10.1038/ncomms13121> PMID: 27762263
21. Sousa AA, Schuck P, Hassan SA (2021) Biomolecular interactions of ultrasmall metallic nanoparticles and nanoclusters. *Nanoscale Adv* 3: 2995–3027. <https://doi.org/10.1039/d1na00086a> PMID: 34124577

22. Schuck P, Zhao H (2013) Biophysical methods for the study of protein interactions. *Methods* 59. <https://doi.org/10.1016/j.ymeth.2013.03.003> PMID: 23522094
23. Schuck P (2010) Sedimentation patterns of rapidly reversible protein interactions. *Biophys J* 98: 2005–2013. <https://doi.org/10.1016/j.bpj.2009.12.4336> PMID: 20441765
24. Chaton CT, Herr AB (2015) Elucidating complicated assembling systems in biology using size-and-shape analysis of sedimentation velocity data. *Methods Enzymol* 562: 187–204. <https://doi.org/10.1016/bs.mie.2015.04.004> PMID: 26412652
25. Ebel C, Birck C (2021) Sedimentation Velocity Methods for the Characterization of Protein Heterogeneity and Protein Affinity Interactions. *Methods Mol Biol* 2247: 155–171. https://doi.org/10.1007/978-1-0716-1126-5_9 PMID: 33301117
26. Perugini MA, Schuck P, Howlett GJ (2002) Differences in the binding capacity of human apolipoprotein E3 and E4 to size-fractionated lipid emulsions. *Eur J Biochem* 269: 5939–5949. <https://doi.org/10.1046/j.1432-1033.2002.03319.x> PMID: 12444983
27. Mehn D, Iavicoli P, Cabaleiro N, Borgos SE, Caputo F, et al. (2017) Analytical ultracentrifugation for analysis of doxorubicin loaded liposomes. *Int J Pharm* 523: 320–326. <https://doi.org/10.1016/j.ijpharm.2017.03.046> PMID: 28342788
28. Maruno T, Usami K, Ishii K, Torisu T, Uchiyama S (2021) Comprehensive Size Distribution and Composition Analysis of Adeno-Associated Virus Vector by Multiwavelength Sedimentation Velocity Analytical Ultracentrifugation. *J Pharm Sci* 110: 3375–3384. <https://doi.org/10.1016/j.xphs.2021.06.031> PMID: 34186069
29. Saleun S, Mas C, Le Roy A, Penaud-Budloo M, Adjali O, et al. (2023) Analytical ultracentrifugation sedimentation velocity for the characterization of recombinant adeno-associated virus vectors sub-populations. *Eur Biophys J*. <https://doi.org/10.1007/s00249-023-01650-3> PMID: 37106255
30. Burnham B, Nass S, Kong E, Mattingly M, Woodcock D, et al. (2015) Analytical Ultracentrifugation as an Approach to Characterize Recombinant AAV Vectors. *Hum Gene Ther Methods*: 1–48. <https://doi.org/10.1089/hgtb.2015.048> PMID: 26414997
31. Khasa H, Kilby G, Chen X, Wang C (2021) Analytical band centrifugation for the separation and quantification of empty and full AAV particles. *Mol Ther—Methods Clin Dev* 21: 585–591. <https://doi.org/10.1016/j.omtm.2021.04.008> PMID: 34095342
32. Yarawsky AE, Zai-Rose V, Cunningham HM, Burgner JW, DeLion MT, et al. (2023) AAV analysis by sedimentation velocity analytical ultracentrifugation: beyond empty and full capsids. *Eur Biophys J*. <https://doi.org/10.1007/s00249-023-01646-z> PMID: 37037926
33. Wawra S, Kessler S, Egel A, Solzin J, Burkert O, et al. (2023) Hydrodynamic characterization of a vesicular stomatitis virus—based oncolytic virus using analytical ultracentrifugation. *Eur Biophys J*. <https://doi.org/10.1007/s00249-023-01649-w> PMID: 37133524
34. Maruno T, Ishii K, Torisu T, Uchiyama S (2023) Size Distribution Analysis of the Adeno-Associated Virus Vector by the c(s) Analysis of Band Sedimentation Analytical Ultracentrifugation with Multiwavelength Detection. *J Pharm Sci* 112: 937–946. <https://doi.org/10.1016/j.xphs.2022.10.023> PMID: 36374763
35. Trachtenberg S, Schuck P, Phillips TM, Andrews SB, Leapman RD (2014) A structural framework for a near-minimal form of life: Mass and compositional analysis of the helical mollicute *Spiroplasma melleiferum* BC3. *PLoS One* 9: e87921. <https://doi.org/10.1371/journal.pone.0087921> PMID: 24586297
36. Zhao H, Mayer ML, Schuck P (2014) Analysis of protein interactions with picomolar binding affinity by fluorescence-detected sedimentation velocity. *Anal Chem* 86: 3181–3187. <https://doi.org/10.1021/ac500093m> PMID: 24552356
37. Chaturvedi SK, Ma J, Brown PH, Zhao H, Schuck P (2018) Measuring macromolecular size distributions and interactions at high concentrations by sedimentation velocity. *Nat Commun* 9: 4415. <https://doi.org/10.1038/s41467-018-06902-x> PMID: 30356043
38. Liu J, Yadav S, Andya J, Demeule B, Shire SJ (2015) Analytical ultracentrifugation and its role in development and research of therapeutic proteins. *Methods Enzymol* 562: 441–476. <https://doi.org/10.1016/bs.mie.2015.04.008> PMID: 26412663
39. Berkowitz SA, Engen JR, Mazzeo JR, Jones GB (2012) Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nat Rev Drug Discov* 11: 527–540. <https://doi.org/10.1038/nrd3746> PMID: 22743980
40. Berkowitz SA, Philo JS (2015) Characterizing biopharmaceuticals using analytical ultracentrifugation. In: Houde DJ, Berkowitz SA, editors. *Biophysical Characterization of Proteins in Developing Biopharmaceuticals*. Amsterdam: Elsevier. pp. 211–260. <https://doi.org/10.1016/B978-0-444-59573-7.00009-9>
41. Bou-Assaf GM, Budyak IL, Brenowitz M, Day ES, Hayes D, et al. (2022) Best Practices for Aggregate Quantitation of Antibody Therapeutics by Sedimentation Velocity Analytical Ultracentrifugation. *J Pharm Sci*. <https://doi.org/10.1016/j.xphs.2021.12.023> PMID: 34986360

42. Gabrielson JP, Brader ML, Pekar AH, Mathis KB, Winter G, et al. (2007) Quantitation of aggregate levels in a recombinant humanized monoclonal antibody formulation by size exclusion chromatography, asymmetrical flow field flow fractionation, and sedimentation velocity. *J Pharm Sci* 96: 268–279. <https://doi.org/10.1002/jps.20760> PMID: 17080424
43. Chaturvedi SK, Parupudi A, Juul-Madsen K, Nguyen A, Vorup-Jensen T, et al. (2020) Measuring aggregates, self-association, and weak interactions in concentrated therapeutic antibody solutions. *MABS* 12: 1810488. <https://doi.org/10.1080/19420862.2020.1810488> PMID: 32887536
44. Parupudi A, Chaturvedi SK, Adão R, Harkness RW, Dragulin-Otto S, et al. (2021) Global multi-method analysis of interaction parameters for reversibly self-associating macromolecules at high concentrations. *Sci Rep* 11: 5741. <https://doi.org/10.1038/s41598-021-84946-8> PMID: 33707571
45. Hopkins MM, Parupudi A, Bee JS, Bain DL (2021) Energetic Dissection of Mab-Specific Reversible Self-Association Reveals Unique Thermodynamic Signatures. *Pharm Res*. <https://doi.org/10.1007/s11095-021-02987-0> PMID: 33604786
46. Philo JS (2009) A critical review of methods for size characterization of non-particulate protein aggregates. *Curr Pharm Biotechnol* 10: 359–372. <https://doi.org/10.2174/138920109788488815> PMID: 19519411
47. Arthur KK, Kendrick BS, Gabrielson JP (2015) Guidance to Achieve Accurate Aggregate Quantitation in Biopharmaceuticals by SV-AUC. 1st ed. Elsevier Inc. 1–24 p. <https://doi.org/10.1016/bs.mie.2015.06.011> PMID: 26412664
48. Lu Y, Harding SE, Turner A, Smith B, Athwal DS, et al. (2008) Effect of PEGylation on the solution conformation of antibody fragments. *J Pharm Sci* 97: 2062–2079. <https://doi.org/10.1002/jps.21170> PMID: 17828753
49. Clardy SM, Lee DH, Schuck P (2021) Determining the Stoichiometry of a Protein–Polymer Conjugate Using Multisignal Sedimentation Velocity Analytical Ultracentrifugation. *Bioconjug Chem* 32: 942–949. <https://doi.org/10.1021/acs.bioconjchem.1c00095> PMID: 33848127
50. FDA (n.d.) Part 11, Electronic Records; Electronic Signatures—Scope and Application. Available: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application>. Accessed 5 May 2023.
51. Schuck P, Zhao H (2017) Sedimentation Velocity Analytical Ultracentrifugation: Interacting Systems. Boca Raton, FL: CRC Press. 271 p.
52. Schuck P (2010) On computational approaches for size-and-shape distributions from sedimentation velocity analytical ultracentrifugation. *Eur Biophys J* 39: 1261–1275. <https://doi.org/10.1007/s00249-009-0545-7> PMID: 19806353
53. Ma J, Zhao H, Schuck P (2015) A histogram approach to the quality of fit in sedimentation velocity analyses. *Anal Biochem* 483: 1–3. <https://doi.org/10.1016/j.ab.2015.04.029> PMID: 25959995
54. Brown PH, Balbo A, Schuck P (2007) Using prior knowledge in the determination of macromolecular size-distributions by analytical ultracentrifugation. *Biomacromolecules* 8: 2011–2024. <https://doi.org/10.1021/bm070193j> PMID: 17521163
55. Zhao H, Ghirlando R, Piszczek G, Curth U, Brautigam CA, et al. (2013) Recorded scan times can limit the accuracy of sedimentation coefficients in analytical ultracentrifugation. *Anal Biochem* 437: 104–108. <https://doi.org/10.1016/j.ab.2013.02.011> PMID: 23458356
56. Berkowitz SA, Laue TM (2021) Boundary convection during velocity sedimentation in the Optima analytical ultracentrifuge. *Anal Biochem*: 114306. <https://doi.org/10.1016/j.ab.2021.114306> PMID: 34274312
57. Zhao H, Balbo A, Metger H, Clary R, Ghirlando R, et al. (2014) Improved measurement of the rotor temperature in analytical ultracentrifugation. *Anal Biochem* 451: 69–75. <https://doi.org/10.1016/j.ab.2014.02.006> PMID: 24530285
58. Schiel JE, Turner A, Mouchahor T, Yandrowski K, Teliikepalli S, et al. (2018) The NISTmAb Reference Material 8671 value assignment, homogeneity, and stability. *Anal Bioanal Chem* 410: 2127–2139. <https://doi.org/10.1007/s00216-017-0800-1> PMID: 29411089
59. Brautigam CA (2015) Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol* 562: 109–133. <https://doi.org/10.1016/bs.mie.2015.05.001> PMID: 26412649
60. Caputo F, Clogston J, Calzolari L, Rösslein M, Prina-Mello A (2019) Measuring particle size distribution of nanoparticle enabled medicinal products, the joint view of EUNCL and NCI-NCL. A step by step approach combining orthogonal measurements with increasing complexity. *J Control Release* 299: 31–43. <https://doi.org/10.1016/j.jconrel.2019.02.030> PMID: 30797868
61. Kirchoff CF, Wang XZM, Conlon HD, Anderson S, Ryan AM, et al. (2017) Biosimilars: Key regulatory considerations and similarity assessment tools. *Biotechnol Bioeng* 114: 2696–2705. <https://doi.org/10.1002/bit.26438> PMID: 28842986

62. Savelyev A, Gorbet GE, Henrickson A, Demeler B (2020) Moving analytical ultracentrifugation software to a good manufacturing practices (GMP) environment. *PLOS Comput Biol* 16: e1007942. <https://doi.org/10.1371/journal.pcbi.1007942> PMID: 32559250
63. Zhao H, Brautigam CA, Ghirlando R, Schuck P (2013) Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation. *Curr Protoc Protein Sci* 7: 20.12.1. <https://doi.org/10.1002/0471140864.ps2012s71> PMID: 23377850
64. Salvay AG, Communie G, Ebel C (2012) Sedimentation velocity analytical ultracentrifugation for intrinsically disordered proteins. *Methods Mol Biol* 896: 91–105. https://doi.org/10.1007/978-1-4614-3704-8_6 PMID: 22821519
65. Zhao H, Li W, Chu W, Bollard M, Adão R, et al. (2020) Quantitative Analysis of Protein Self-Association by Sedimentation Velocity. *Curr Protoc Protein Sci* 101: 1–15. <https://doi.org/10.1002/cpps.109> PMID: 32614509
66. Kar SR, Kingsbury JS, Lewis MS, Laue TM, Schuck P (2000) Analysis of transport experiments using pseudo-absorbance data. *Anal Biochem* 285: 135–142. <https://doi.org/10.1006/abio.2000.4748> PMID: 10998273
67. Juul-Madsen K, Zhao H, Vorup-Jensen T, Schuck P (2019) Efficient data acquisition with three-channel centerpieces in sedimentation velocity. *Anal Biochem* 586. <https://doi.org/10.1016/j.ab.2019.113414> PMID: 31493371
68. To SC, Brautigam CA, Chaturvedi SK, Bollard MT, Krynitsky J, et al. (2019) Enhanced Sample Handling for Analytical Ultracentrifugation With 3D-Printed Centerpieces. *Anal Chem* 91: 5866–5873. <https://doi.org/10.1021/acs.analchem.9b00202> PMID: 30933465
69. Schuck P (2004) A model for sedimentation in inhomogeneous media. I. Dynamic density gradients from sedimenting co-solutes. *Biophys Chem* 108: 187–200. <https://doi.org/10.1016/j.bpc.2003.10.016> PMID: 15043929
70. Zhao H, Brown PH, Balbo A, Fernandez Alonso MC, Polishchuck N, et al. (2010) Accounting for solvent signal offsets in the analysis of interferometric sedimentation velocity data. *Macromol Biosci* 10: 736–745. <https://doi.org/10.1002/mabi.200900456> PMID: 20480511
71. Gabrielson JP, Arthur KK, Kendrick BS, Randolph TW, Stoner MR (2009) Common excipients impair detection of protein aggregates during sedimentation velocity analytical ultracentrifugation. *J Pharm Sci* 98: 50–62. <https://doi.org/10.1002/jps.21403> PMID: 18425806
72. Vinograd J, Bruner R, Kent R, Weigle J (1963) Band-centrifugation of macromolecules and viruses in self-generating density gradients. *Proc Natl Acad Sci USA* 49: 902–910. <https://doi.org/10.1073/pnas.49.6.902> PMID: 13997382
73. Lebowitz J, Teale M, Schuck P (1998) Analytical band centrifugation of proteins and protein complexes. *Biochem Soc Trans* 26: 745–749. <https://doi.org/10.1042/bst0260745> PMID: 10047819
74. Wafer L, Kloczewiak M, Luo Y (2016) Quantifying trace amounts of aggregates in biopharmaceuticals using analytical ultracentrifugation sedimentation velocity: Bayesian analyses and F statistics. *AAPS J*. <https://doi.org/10.1208/s12248-016-9925-y> PMID: 27184576
75. Brown PH, Balbo A, Schuck P (2008) A Bayesian approach for quantifying trace amounts of antibody aggregates by sedimentation velocity analytical ultracentrifugation. *AAPS J* 10: 481–493. <https://doi.org/10.1208/s12248-008-9058-z> PMID: 18814037
76. Ma J, Zhao H, Sandmaier J, Liddle JA, Schuck P (2016) Variable-field analytical ultracentrifugation: II. Gravitational sweep sedimentation. *Biophys J* 110: 103–112. <https://doi.org/10.1016/j.bpj.2015.11.027> PMID: 26745414
77. Mehn D, Rio-Echevarria IM, Gilliland D, Kaiser M, Vilsmeier K, et al. (2018) Identification of nanomaterials: A validation report of two laboratories using analytical ultracentrifugation with fixed and ramped speed options. *NanoImpact* 10: 87–96. <https://doi.org/10.1016/j.impact.2017.12.005>