

An automated interface for sedimentation velocity analysis in SEDFIT

Laboratory of Dynamics of Macromolecular Assembly, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD 20892, USA

For general background see PLOS Computational Biology (2023)

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011454>

Update in Version 17.0 (February 2024), with features described in ACS Nano 2024
(<https://pubs.acs.org/doi/full/10.1021/acsnano.4c05322>)

- **New flotation c(s) model**
- **New regularization type Tikhonov-Phillips modified 3rd order regularization**
- **New option to define scan files**
- **More detailed control over log-spaced s-distributions, and new default behavior to use log-spaced grid for s-values > 10**
- **New option to specify a mode for simulating sedimentation boundaries**

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.

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
nIndividualScanFilePaths	integer	alternative to FirstScan/LastScan/ScanInterval specification, this switches SEDFIT to expect the following parameters to be full pathnames of each individual file; minimum 2. If this option is used, the parameters “Channel”, “FirstScan”, “LastScan” and “ScanInterval” will not be needed anymore
ScanFilePath1	path to first data file to be loaded	must exist
ScanFilePath2, ... ScanFilePath3, ...	path to next data file to be loaded	must exist; must be ordered by scan time; must be sequential and complete until ScanFilePathN if ‘N’ is the number specified in the parameter nIndividualScanFilePaths
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’ or ‘flotationcofs’ or ‘discrete’ (no quotation marks)	model selection
Resolution	integer	number of grid points in the distribution
Smin	floating point number	minimum s -value of the distribution
Smax	floating point number	maximum s -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.
LogSpacedSmin	floating point number	s-value where the scale switches from linear ($s < \text{LogSpacedSmin}$) to log spaced ($s > \text{LogSpacedSmin}$). For flotation distributions the same considerations apply for the absolute value of the s-values. If LogSpacedSmin is specified ≤ 0 then log-spaced grid is turned off. If this parameter is not set, the default behavior is to switch log-spaced grids ON with a threshold of $s = 10$ S.
FlotationDeconvolutionR	floating point number	hydrodynamic radius in nm; alternative to FlotationDeconvolutionD (set either one)
FlotationDeconvolutionD	floating point number	diffusion coefficient in F ($10^{-7}\text{cm}^2/\text{sec}$); alternative to FlotationDeconvolutionR (set either one)
FlotationMinS	floating point number	s-value for floating branch of $c(s)$: include ‘-’ sign;
FlotationMaxS	floating point number	s-value for floating branch of $c(s)$: include ‘-’ sign; note that FlotationMinS < FlotationMaxS
FlotationGridN	integer	similar to ‘Resolution’, but for the floating branch of $c(s)$
FlotationOnly	TRUE/FALSE	set TRUE if there is no sedimentation branch of $c(s)$
RegularizationType	‘maxent’ or ‘Tikhonov’	defines the regularization used in the distribution

	or 'TP3' (no quotation marks)	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 s -value in the distribution with baselines
NumberComputationThreads	integer	number of computing cores available for SEDFIT
SimulateOnly	TRUE/FALSE	Switch SEDFIT from fitting to simulation mode
SimulateS	floating point number	sedimentation coefficient to be simulated, in S; use with Model 'discrete'.
SimulateD	floating point number	diffusion coefficient to be simulated, in F (corresponding to 10^{-7}cm²/sec); use with Model 'discrete'.
SimulateC	floating point number	signal amplitude to be simulated; use with Model 'discrete'.
SaveRadialScanFit	TRUE/FALSE	if true, will save fitted scan files in XLA format
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 $c(s)$ 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
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

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	ω^2t 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