DCDT+ User Manual

Version 2.4.1
September 4, 2013

DCDT+ User Manual

Table of Contents

1. Getting Started 1
  1.1. Program overview and organization 1-2
  1.2. How to use this Help file 2-3
  1.3. Standard and advanced modes 3
2. Tutorials 4
  2.1. Tutorial: Calculating the g(s*) distribution 4-8
  2.2. Tutorial: How to select the scans to analyze 8-11
  2.3. Tutorial: Using the g(s*) overlay graph 11-13
  2.4. Tutorial: Single species fit 13-17
  2.5. Tutorial: Multi-species fits 17-22
  2.6. Tutorial: Simulating experiments with the Claverie simulator 22-24
3. Program elements, controls, and features 25
  3.1. User interface 25
    3.1.1. Main menus 25
      3.1.1.1. File menu 25-26
        3.1.1.1.1. Open saved analysis... 26
        3.1.1.1.2. Save analysis 26
        3.1.1.1.3. Save analysis as... 26-27
        3.1.1.1.4. Clone analysis to new window 27
      3.1.1.2. Edit menu 27-28
      3.1.1.3. View menu 28
      3.1.1.4. Tools menu 28-29
        3.1.1.4.1. Mass/diffusion calculator 29-30
        3.1.1.4.2. Options dialog 30-33
      3.1.1.5. Windows menu 33
      3.1.1.6. Help menu 33-34
    3.1.2. Toolbar 34
    3.1.3. Analysis window 34-36
      3.1.3.1. Next and Back buttons 36

3.1.3.2. More >> and << Less buttons 36

3.1.3.3. Log page 36-38

3.1.3.4. User prompt area 38

3.1.4. Navigation tree 38-39

3.1.4.1. Derive g(s*) distribution branch 39-40

3.1.4.2. Fit distribution as N species branch 40

3.1.5. g(s*) overlay graph 40-41

3.2. Analysis pages 41

3.2.1. Load raw scans page 41-42

3.2.1.1. Load raw scans - Preselect data type and cell check box 42

3.2.1.2. Load raw scans - Load new scan files button 42-43

3.2.1.3. Load raw scans - Replace scans for this sample button 43

3.2.1.4. Load raw scans - Add more scans for this sample button 43

3.2.1.5. Load raw scans - Delete selected scans button 43

3.2.1.6. Load raw scans - Scans currently loaded in memory list box 43-44

3.2.1.7. Load raw scans - Scans to be used in g(s*) calculation list box 44

3.2.1.8. Load raw scans - Data loading options 44-45

3.2.1.9. Load raw scans - Sample description 45

3.2.1.10. Load raw scans - Run Claverie simulation to generate data button 45-46

3.2.2. Set meniscus & data region page 46-47

3.2.2.1. Set meniscus & data region - Meniscus position 47

3.2.2.2. Set meniscus & data region - Good data begin X cm from meniscus 47

3.2.2.3. Set meniscus & data region - Good data end at radius = 48

3.2.2.4. Set meniscus & data region - Jitter removal options 48-49

3.2.2.5. Set meniscus & data region - Fringe jump removal options 49-50

3.2.2.6. Set meniscus & data region - Data displayed options 50-51

3.2.2.7. Set meniscus & data region - Zoom meniscus button 51

3.2.2.8. Set meniscus & data region - ID points button 51

3.2.3. Select scans to analyze page 51-53

3.2.3.1. Select scans to analyze - Auto adjust button 53

3.2.3.2. Select scans to analyze - Peak broadening limit 53-54

- 3.2.3.3. Convert to s(20,w) 54-56
- 3.2.3.4. Select scans to analyze - Use true mean time button 56-57
- 3.2.3.5. # of sections 57-58
- 3.2.4. Calculate average dc/dt curve page 58-59
  - 3.2.4.1. Calculate average dc/dt - X-axis point spacing 59
  - 3.2.4.2. Calculate average dc/dt - Set range for g(s*) 59-60
  - 3.2.4.3. Convert to s(20,w) 60-62
  - 3.2.4.4. Calculate average dc/dt - Reset zero & Undo buttons 62-63
- 3.2.5. Calculate g(s*) distribution page 63
  - 3.2.5.1. Calculate g(s*) distribution - Show integrals over distribution 63-64
  - 3.2.5.2. Calculate g(s*) distribution - Set integration region 64
  - 3.2.5.3. Averages and peak areas dialog 64-66
  - 3.2.5.4. Calculate g(s*) distribution - Normalize option 66
  - 3.2.5.5. Convert to s(20,w) 66-68
  - 3.2.5.6. Calculate g(s*) distribution - Display options 68-69
  - 3.2.5.7. Calculate g(s*) distribution - Reset zero & Undo buttons 69
  - 3.2.5.8. Calculate g(s*) distribution - Broad algorithm option 69-70
- 3.2.6. Select fitting model page 70
  - 3.2.6.1. Select fitting model - Number of species 70-71
  - 3.2.6.2. Select fitting model - Improved or classic methods 71
  - 3.2.6.3. Select fitting model - Data choice 71
  - 3.2.6.4. Select fitting model - Fit diffusion or mass 71-72
  - 3.2.6.5. Select fitting model - Weighting of data points 72-73
  - 3.2.6.6. Select fitting model - Constrain M or s ratios among species 73-75
  - 3.2.6.7. Select fitting model - Molecular mass units 75-76
  - 3.2.6.8. Select fitting model - Density and partial specific volumes 76-77
- 3.2.7. Set initial guesses on graph page 77-78
- 3.2.8. Set/alter fitting parameters page 78-79
  - 3.2.8.1. Parameter value text boxes 79
  - 3.2.8.2. Fit control check boxes 79
  - 3.2.8.3. Constrained parameters 79

3.2.8.4. Convergence control 79-80
3.2.8.5. Adjust parameter upper/lower limits button 80-81
3.2.9. Perform least-squares fit page 81-83
3.2.9.1. Perform fit - Do fit button 83
3.2.9.2. Perform fit - 1 iteration button 83
3.2.9.3. Perform fit - Simulate button 83
3.2.9.4. Perform fit - Cancel button 83-84
3.2.9.5. Perform fit - Status 84
3.2.9.6. Fitting parameters monitor tab 84-85
3.2.9.6.1. Increment reduction factor 85-86
3.2.10. Report fit results page 86-87
3.2.10.1. Report fit results - Fitted parameters 87
3.2.10.2. Report fit results - Fixed or constrained parameters 87
3.2.10.3. Report fit results - View parameter cross-correlations button 87-88
3.2.10.4. Report fit results - Compute confidence intervals button 88-90
3.2.10.5. Report fit results - Create reports group box 90
3.2.10.6. Report fit results - Fit statistics 90-91
3.2.10.7. Report fit results - High cross-correlation warning 91
3.2.10.8. Report fit results - Graph bootstrap or Monte Carlo results 91-92
3.2.11. Graph fit results page 92-95
4. How to... 96
4.1. How to avoid common mistakes 96-97
4.2. How to optimize your experiments 98
4.2.1. How to optimize data acquisition for g(s*) analysis 98-101
4.2.2. How (and why) to equilibrate the rotor temperature 101-102
4.3. How to work with graphs 102
4.3.1. How to print graphs 102
4.3.2. How to customize graphs 102-104
4.3.3. How to export graph images or data to other programs 104
4.3.4. How to zoom with the mouse 104
4.4. How to remove bad data points from scan files 104-105

- 4.5. How to identify and remove rogue scans 105-106
- 5. Frequently-asked questions 107
  - 5.1. About g(s*) distributions 107
    - 5.1.1. What is the difference between s* and s? 107
    - 5.1.2. What is the difference between g(s*) and g^(s*)? 107-108
    - 5.1.3. What is the difference between the broad and conventional g(s*) algorithms? 108-109
  - 5.2. About fitting 109
    - 5.2.1. Should I fit the offset? 109-110
    - 5.2.2. Why do I get different answers when fitting g(s*) versus dc/dt? 110-111
    - 5.2.3. What constitutes a good fit? 111-112
    - 5.2.4. What does high parameter cross-correlation mean? 112
  - 5.3. How should I cite this program? 112-113
- 6. Reference information 114
  - 6.1. Symbols used 114
  - 6.2. About using radial intensity scans 114-115
  - 6.3. About this g(s*) implementation 115
  - 6.4. About the fitting functions 115-116
  - 6.5. About the fitting algorithm 116-117
  - 6.6. About the meniscus wizard 117
  - 6.7. About the jitter wizard 117-118
  - 6.8. About the bootstrap method 118
  - 6.9. About the Monte Carlo method 118-119
- 7. Program license and registration 120
- 8. Bibliography 121
  - 8.1. reference 1 - Philo (1997) 121
  - 8.2. reference 2 - Johnson & Faunt (1992) 121
  - 8.3. reference 3 - Stafford (1997) 121
  - 8.4. reference 4 - Durschlag (1986) 121
  - 8.5. reference 5 - Stafford 'rule of thumb' 121
  - 8.6. reference 6 - Stafford (1999) 121

8.7. reference 7 - Stafford (1994) 121-122

8.8. reference 8 - Philo (2000) 122

8.9. reference 9 - Behlke & Ristau (2002) 122

8.10. reference 10 122

8.11. reference 11 - Efron & Tibshirani (1986) 122

8.12. reference 12 - Johnson & Faunt 1994 122

8.13. reference 13 - Philo 2011 122

8.14. reference 14 - Stafford (1992) 122-123

DCDT+ User Manual

1 Getting Started

1.1 Program overview and organization

Below is a picture to illustrate various parts of the program user interface, followed text explaining more about the program design and operation. (See How to use this Help file for information about *expandable sections* and other features of this Help file).

The **See Also section** (at the bottom, or use the button at upper left) will lead you to tutorials and other information to help get you started.

Program screen

Main menu & toolbar

DCDT+ has a main program menu with the standard File, View, Window and other main headings you find in many Windows programs. Analysis documents can be created, opened and saved using the familiar New, Open, and Save commands under the File menu, or using the familiar icons on the Toolbar.

An analysis document within DCDT+ is contained within an Analysis window. In the example above, three analysis windows are open (which happen to be 3 different cells from the same run in this case). Each analysis window concerns the analysis of one sample. In each analysis window you can work through the analysis in a step-by-step manner, and different controls and/or graphs are shown as different pages within the window at different steps.

Navigating within and between analysis pages

1 Movement among the steps (pages) within one analysis window can be accomplished using either the **Next button** and **Back button** at the lower right or by right-clicking on the analysis window with the mouse and using the Go to step popup menu.

DCDT+ also provides an Explorer-style Navigation tree along the left side. This navigation tree provides an easy mechanism to move both between different analysis windows, and also to specific pages (steps) within each analysis window.

Five steps are needed to complete derivation of the g(s*) distribution, and those are grouped to form the derive g(s*) distribution branch within the navigation tree.

Once a *g(s*)* distribution has been computed, you may optionally fit that distribution as a mixture of one to five independently-sedimenting components. Such fitting may involve up to 6 steps, which are grouped to form the fit distribution as N species branch within the navigation tree.

Wizards and Standard Mode make it easy, but more options are there when needed

Although many options and opportunities to fine-tune the analysis are available, whenever possible the program uses intelligent 'wizards' to choose initial settings and parameters for you. The program even pre-selects the buttons you will probably need to hit next.

2 In Standard mode (shown here, and the initial default) many options are hidden to keep things simpler, but you can easily switch to Advanced mode using the **More >> button** when you need more power.

Analysis the easy way:

1. Load the raw scans
2. Push the **Enter** key to confirm the choices the program has made. Pushing **Enter** 10 times will take you all the way from raw data to a *g(s*)* distribution and on through fitting your sample as a single species to give you its sedimentation coefficient and estimated molecular mass. **Yes it is that easy!**

True, for a final result you will probably want to go back and tweak things a bit, like exactly which scans are analyzed. For an accurate molecular mass of course you also have to stop and input the exact solvent density and partial specific volume. However the automatic choices should be more than good enough to tell you whether your sample is a monomer, dimer, trimer, *etc*.

1.2 How to use this Help file

Within certain topics you will see *expandable sections* like the one below that when clicked expand to provide more details about that subject.

Show Getting context-sensitive Help from within the program

As is true for many Windows® programs, pushing the F1 key at any point in the program will bring up a Help screen relevant to that context (page within the analysis window, dialog box, or control).

Remember too that virtually every button or other control will display brief "tool tip" help if you hover over that control with your mouse.

For pages with multiple expandable links you may also see a **Show All/Hide All** link like the one at the right below that expands or contracts all such material within that page.

Show All Show All

Hide All Hide All

Special text sections

Short blocks of text that is specific for the Advanced mode will be indicated by a colored block with the Einstein symbol, as shown below:

This is text concerning a feature that is only available in **Advanced mode**.

Long blocks of text or whole topics specific to Advanced mode may be indicated simply by the Einstein icon.

Tips about easier or better ways to accomplish tasks will be indicated by a colored block with the idea symbol, as shown below:

To best way to do X is to use the Y button and then enter...

Cautions about potential mistakes or pitfalls will be indicated by a colored block with a caution symbol, as shown below:

Be sure the rotor temperature is equilibrated before starting your run---even the best software cannot overcome bad data!

Material that provides **background information or details** will be indicated by a colored block with the teacher symbol, as shown below:

The intellectually curious user will want to know that...

Such information may also be found within an expandable text block in some cases.

Send Comments link

At the upper right corner of every topic is a link that allows you to send an e-mail to the author if you find errors in the Help file, if you have requests for new topics or tutorials, or if you have other comments.

See Also links

A **See Also** section like that below that leads you to related topics and further information will be found at the bottom of many topics. The **See Also** link symbol, See Also symbol, at the upper left corner of the topic page, takes you directly to the **See Also** list.

1.3 Standard and advanced modes

There are two program states or modes for each analysis window:

1. In **Standard mode** certain program options and steps are hidden to keep the user interface simple and keep features that may not be needed out of the way.
2. In **Advanced mode** all program options and steps are shown and available for use.

Each analysis window can be toggled between these modes using the and buttons (the button text and graphic toggle between these two states depending on the current mode).

Within this Help file information that is specific for the **Advanced mode** will be indicated as shown below, or simply by the Einstein symbol:

This is text concerning a feature that is only available in **Advanced mode**.

By default when the program is installed each new window is opened in standard mode. That behavior can be turned off using an entry on the General Options tab of the Options dialog.

2 Tutorials

2.1 Tutorial: Calculating the g(s*) distribution

This tutorial will walk you through computing the *g(s*)* distribution, using a set of interference data as the example. This tutorial will be done in standard mode.

Show All Show All

Hide All Hide All

Show Load from disk the scans to be analyzed

(1) Begin on the Load raw scans page of a new or existing analysis window. You may wish to preselect the data type and cell number for the data you wish to load. In this example, interference data from cell 2 are pre-selected.

Initiating scan load

(2) Push the Load new scan files button to bring up a multiple File Open dialog box.

Selecting all the scans

Unlike other DCDT implementations (and version 1.xxx of this program) you will usually want to load scans covering the entire range of boundary movement, and then later you will choose a subset of those scans that you are going to actually analyze (perhaps 20% of the total). So generally you will want to highlight the file names for all the scans (click on the first scan, hold the Shift key down, and click on the last scan).

- If you know the data acquisition went long after all the sample was pelleted you probably want to omit those scans where the cell is empty.
- If you have hundreds of interference scans and a relatively slow computer you may want to use the Load only every Nth scan option to reduce the total scans loaded, at least initially. Later, after you have found an appropriate subset for analysis, you can go back and load every scan from the beginning to end of that subset to get the maximum signal/noise.
- At a minimum you will want to select at least 4 scans.

(3) Click **Ok** in the File Open dialog to load these scans. Loading may require 10 seconds or more depending on the number of scans (and possibly network speed if you are accessing through a network). The progress bar will advance as the files load.

Analysis page 1 after scans are loaded

At this point all the scan names will show in the Scans currently loaded in memory and Scans to be used in g(s*) calculation list boxes, and your description of that sample will appear near the bottom of the page in the sample description area.

(4) Push the Next button to move to the Set meniscus & data region page and see a graph overlaying all of the scans you just loaded.

Show

Set the meniscus position and define the region containing the 'good' data

Analysis page 2 before meniscus position is set

(1) When you first enter the the Set meniscus & data region page the program's meniscus wizard will attempt to locate the meniscus position and will place the red vertical line at that position, as shown above. That position will be filled into the meniscus text box at the upper left, and that text box is automatically given the focus.

- At this point the *fringe jumps* in these scans make it difficult to see the boundary movement. Those will automatically be removed as soon as a meniscus position has been set.
- Note also that in this experiment the data acquisition region extended past the base of the cell, producing lots of garbage data beyond 7.15 cm. It is recommended that you start the acquisition a little to the left of the reference meniscus and end it a little to the left of the base of the cell.

(2) If you are satisfied with the wizard's choice of meniscus position then all you need to do is push the **Enter** key on the keyboard to accept that choice. Otherwise you can click on the red meniscus marker and drag it to the position you want, or simply type a value into the text box.

- To set the meniscus more accurately (and that is important for obtaining accurate sedimentation coefficients) you may wish to zoom the graph around the meniscus position. You can do that manually by dragging the mouse over the region on the graph (left-click and drag).

Analysis page 2 after the meniscus position is set

(3) For interference scans as soon as a meniscus position is set the default method for fringe jump removal will be applied, the 'jitter wizard' will determine the best position for removing *jitter* and remove it, and the graph will alter to display the corrected data, as shown above. At this point the scans should align vertically so the boundary movement can be seen, but in some cases adjustments may be needed. See jitter removal and fringe jump removal options for more information.

- Note that it is normal that the scans do not overlay in the air-air space (to the left of both meniscii) because of the fringe jumps.

Analysis page 2 after adjusting cutoff at cell base

(4) The region between the vertical green bars defines the 'good' data that will be used in the analysis. By default a small region to the right of the meniscus is excluded because the data there are usually distorted by

optical artifacts.

It is important to exclude the region near the base of the cell (high radii) where solutes are accumulating, and therefore it may be necessary to drag the right-hand green marker further to the left, as is true for these data for a protein of fairly low molecular mass. The image above shows this analysis page after that adjustment.

- The positions of the green markers can also be altered by typing values in the text boxes near the top center and top right of the page. See Good data begin X cm from meniscus and Good data end at radius = for more information.

(5) Push the Next button to move to the Select scans to analyze page.

Show Calculate the *dc/dt* curves for each scan pair and select the subset of scans to be analyzed

Analysis page 3 initial state when entire run was loaded

The image above shows the initial state when all 110 scans covering the entire run were loaded. The graph shows the individual *dc/dt* curves calculated from each of the 55 scan pairs (as dashed lines) and the average of all 55 of those curves (as a heavy black line). Basically at this stage the results are quite poor because we are doing the calculations using scans from the entire run, including those very early in the run when the boundary is still near the meniscus and the separation is quite poor.

Remember, for *g(s*)* analysis we are trying to capture a snapshot of the distribution in the cell at one particular time during the run. If we do the *dc/dt* calculations using all the scans, that is just like trying to take a picture of a runner with the camera shutter locked open: all the details get blurred out.

Thus in this case the 55 *dc/dt* curves are quite divergent because we are including way too many scans. Note that because the first scan pair (black dashed line) includes data at the earliest times, it provides information about species out to nearly 500 S, but none of the other scan pairs give data beyond ~60 S. Note too that the peak broadening limit says that peaks for all species larger than only 0.22 kDa (!!) are being significantly broadened.

Typically we will want to analyze scans covering ~15-20% of the total boundary movement. The selection of the optimal subset of scans to analyze is a complex topic that is covered in depth in Tutorial: How to select the scans to analyze. But here we will simply let the program's 'Auto adjust wizard' do the work for us by pushing the Auto adjust button (this is pre-selected when you enter this page so actually you can just hit the **Enter** key).

Analysis page 3 after Auto adjust procedure

As shown above, after the Auto adjust procedure average *dc/dt* curve shows a nice peak that is roughly centered along the X-axis, and the individual *dc/dt* curves from each scan pair (now only 8 pairs) all lie fairly close to that average. The positions of the two slider controls now reflect the fraction of the total scans that is being analyzed (upper slider) and the position of that group of scans within the total scans that were loaded (lower slider). Note that the peak broadening limit is now 128 kDa, so there should be no significant peak broadening of any species within this range of sedimentation coefficients.

Actually in this case the Auto adjust was probably too cautious about avoiding peak broadening and it would probably be fine to push up the number of scans to ~28, which would give somewhat higher signal/noise. However these automatic settings are giving perfectly usable results, which is exactly what they are supposed to do. As you gain more experience you will probably want to at least 'tweak' the values set by Auto adjust. It is quick and easy to see what happens as you move the sliders, and you will learn a lot from that too.

Now push the Next button to move to the Calculate average dc/dt curve page

Show Display the average *dc/dt* curve; adjust data range for *g(s*)* if needed

Analysis page 4 (average dc/dt curve)

This page displays the average *dc/dt* curve, with experimental error bars on each data point (gray vertical bars, barely visible in this case). Usually no adjustments are needed, but if necessary you can reduce the range of data that will used in calculating the *g(s*)* distribution (trimming off one or both ends), by using the Set range for g(s*) button or manually altering the limits in the text boxes.

- It is not uncommon to see some bad data points near zero S (which come from the data points near the meniscus). This is especially true for interference data if the reference buffer is not a good match to the sample.
- If you see the curve go below zero at the right end, that probably means your data are still being influenced by the accumulation of solutes at the base of the cell. Use the Back button twice and move the cut-off at the cell base (green bar) further to the left.

Now push the Next button to move to the Calculate g(s*) distribution page.

Show Calculate and display *g(s*)*; optionally adjust axis scaling

Analysis page 5 with computed g(s*) curve

This page calculates the *g(s*)* distribution and displays it, along with error bars. The title line of the graph shows the weight-average sedimentation coefficient (integrated over the distribution) and the total loading concentration (in instrument units).

In this case (image above) we now can see that there are some bad data points near zero S. Clearly *g(s*)* should not go negative (a negative concentration)! This problem is probably caused by including data too close to the meniscus.

Adjusting the lower sedimentation coefficient limit for the g(s*) computation

To clean this up we can just go Back one step, push the Set range for g(s*) button and drag the left magenta bar to set the minimum range to ~0.42 S, as shown above. (Alternatively we could navigate back to the Set meniscus & data region page and move the left hand green marker further away from the meniscus).

- Note that when the range has been reduced from the maximum possible that value is displayed in the text back with a white background rather than the yellow background used to indicate computed values.

g(s*) distribution after raising the lower limit to 0.26 S

After that range adjustment the **Next button** brings us this result shown above. This is basically a final result, but to create a final distribution for presentation or publication you would probably want to convert these 'raw' sedimentation coefficients to the $s_{20,w}$ scale.

Averages over the distribution function dialog box

Pushing the Convert to s(20,w) button brings up the dialog box shown above where you enter the correct partial specific volume, solvent density, and solvent viscosity.

When those values are correct push the **Ok** button and we now get the graph above in *s(20,w)* units.

At this point if your data shows more than one peak you may wish to integrate that peak to compute the concentration of that species, its fractional area, and average sedimentation coefficients. If you push the Set integration region button and then click and drag with the mouse over the sedimentation coefficient region of interest, those calculations will be done and displayed in this dialog box shown above (these results are for the middle portion of the single peak)

- If you instead push the Show integrals over distribution button the same calculations are done averaging over the entire range of the distribution.

g(s*) distribution after normalization

The one remaining thing you *might* wish to do is to normalize the distribution (divide by the total concentration so the area under the curve is exactly 1). That is quite useful if you want to compare distributions for samples at different concentrations on the g(s*) overlay graph (see Tutorial: Using the g(s*) overlay graph). To normalize simply push the normalize button and you will get the result shown above.

- The normalized curve of course looks the same; only the Y axis scaling is different. For normalized data the total concentration is not displayed (since it is always 1.000).

At this point you are done with this analysis unless you wish to fit the peak(s) as individual, independent species. See Tutorial: Single species fit and Tutorial: Multi-species fits.

2.2 Tutorial: How to select the scans to analyze

Normally when you the time derivative technique you are trying to take a 'snapshot' of what is in the cell at a particular time during the run. Thus typically only a fairly small portion of the scans from the run (~15-20%) are analyzed to produce a particular *g*(*s**) distribution. Using too many scans is like leaving the shutter on a camera open too long---the result is a blurred, smeared-out distribution. This tutorial concerns how to optimally select which group of scans should be used.

Below three drop-down text blocks discuss differences from earlier implementations and present the principles behind deciding at what point in the run the analysis should be done, and how many scans should be included in that analysis. Then the tutorial describes how to actually interact with the two slider controls on the Select scans to analyze page in order to achieve those goals. This link jumps immediately to that 'How to' tutorial section.

Show All Show All

Hide All Hide All

Show What's new?

Previous users of DCDT+, and users of the Beckman Coulter software or other DCDT implementations, will find that data selection procedures are quite different in this version. In other implementations the data selection takes place primarily when the scans are selected for loading from disk. That can make make data selection awkward, and often requires several re-starts.

In this version, you generally will want to initially load scans encompassing the entire range of boundary movement, and then use the two slider controls on the Select scans to analyze page to select a portion of those scans for analysis. This makes it very easy to "tune" into the region you want, without needing to start over and load different sets of scans.

Show When during the run should the analysis be done?

The general rule is that optimal resolution (separation of species) occurs just as the fastest sedimenting species reaches the base of the cell. (Actually it is not truly the base of the cell that counts, but rather the region near the base where solutes accumulate).

Thus if your sample contains only one species, the optimal situation is when the last scan used in the analysis was taken just before the plateau region disappears.

If a given sample contains species whose sedimentation coefficients differ by more than a factor of ~2, this means that there is no time during the run at which all species are still present in the cell and at which the properties of the slowly sedimenting component(s) can be determined optimally. Under these circumstances there may be no single "best" set of scans for the analysis, and determining which scans should be analyzed depends on the nature of the sample and what you are trying to learn:

- If you are studying an interacting system (self-association, mixed association, or ligand binding) which you are going to characterize by calculating the weight-average sedimentation coefficient of the whole *g^(s*)* distribution, then it is important that the derived distribution function comes fully to zero at both the low and high sedimentation coefficient ends of the distribution. This means the scans should be early enough that no species has reached the cell base, yet late enough that all species have cleared the meniscus.
- If you are studying a system in which you are trying to resolve the properties of multiple individual components (a mixture, or a strongly-interacting system that can be treated as a mixture), then usually the best choice is a time period just before the most rapid component reaches the cell base. However, when there is a large spread in sedimentation coefficients, it may also be worthwhile to run an additional analysis later in the run (after the faster components are pelleted) in order to obtain more accurate values for the sedimentation and diffusion coefficients of the slower-sedimenting component(s).

Show How many scans should be included in the analysis?

Adding more scans to the analysis involves a trade-off between improving the signal/noise ratio and the loss of resolution as a consequence of peak broadening.

A loss of resolution (peak broadening) occurs as more scans are added because the amount of boundary movement between the first and last scan used in the analysis becomes excessive. (Again using our snapshot analogy, there is too much movement while the 'shutter' is open, so the picture is blurred.) This causes the $\Delta c/\Delta t$ values that are calculated to be a poor approximation of the true derivative, dc/dt. This topic is discussed further under peak broadening limit which uses a "rule of thumb" calculation to determine when too many scans are being included.

In the past when fitting peaks using done using the "classic method" the peak broadening directly eroded the accuracy of the *D* or *M* values, so selection of the number of scans was a critical parameter. However when using the default new "improved method" the peak broadening effects do not compromise the accuracy of the fitted hydrodynamic parameters. Nonetheless even with the improved method the peak broadening is still present, and it can still compromise the resolution between species and thus the ability to obtain accurate properties for each species when multiple species are present. Thus it is still generally best to avoid using so many scans that the peak broadening limit is far below the molecular mass of the species in your sample.

In some circumstances it may be quite appropriate to accept significant peak broadening in return for higher signal/noise. For example, if you are unsure whether a small peak in the distribution truly represents an additional species, you may wish to add many more scans in the analysis to improve the signal/noise and see whether this peak is still present. In such a test the peak for a real species would be broadened, but it should not disappear, so the qualitative result (peak or no peak) is definitely enhanced by using more scans even though the shape of the peak may be significantly distorted.

How to adjust the slider controls

This tutorial section will use some real absorbance data for a monoclonal antibody sample run at 45,000 rpm as an example. First we'll do it the hard way (totally manually).

Initial state with entire run analyzed

(1) If you have loaded scans covering the entire range of boundary movement, then you will almost surely want to move the top slider to the left to reduce the total range of boundary movement and improve the resolution (sharpen the peaks).

The image above shows the initial state when all 60 scans covering the entire run were loaded. Note that the individual curves from each scan pair are quite divergent, and the average *dc/dt* curve (heavy black line) is flat on top and very broad. These are signs that the time span (number of scans used) is much too broad. Note too that the peak broadening limit says that all species larger than only 1 kDa (!!) are being significantly broadened.

Result of dropping to 30 scans used

(2) The image above is the result of moving the top slider to the middle (giving 30 scans used). Now the average *dc/dt* curve starts to resemble a peak, but the curves from the 15 individual scan pairs are still fairly different, and the peak broadening limit is still only 17 kDa.

Result after dropping to 10 scans used

(3) Above are the results after moving the slider further to the left, dropping to 10 scans used. Now we are in the right range---the individual *dc/dt* curves are close to the average (some spread near the top of the peak is normal), and the peak broadening limit is >= the mass of an antibody monomer. Reducing the number of scans further down to 4-6 scans would increase the resolution for very high mass species, but also would degrade the signal/noise significantly.

Note that for absorbance scans it is not unusual that only a small number of scans can be used before significant peak broadening occurs. At 45K rpm the boundary for an antibody (150 kDa) is moving quite rapidly compared to the time between scans so to keep our 'snapshot' from being blurry we can only keep our 'shutter' open a short time.

Result after moving lower slider forward 9 scans

(4) After getting the top slider in roughly the right range, you also may also want to move the lower slider left or right to analyze data earlier or later in the run (to cover higher or lower sedimentation coefficients).

A general rule of thumb is that the main peak should be roughly in the middle of the graph. However that rule may not produce a sedimentation coefficient range that includes all of the species that are present in your sample. In the image above the average *dc/dt* curve is not flat over the region from 9 to 12 S, suggesting there may be aggregates at or beyond 12 S present and those will not be analyzed with these settings. The image above shows the results after moving the lower slider such that the group of 10 scans being analyzed starts 9 scans earlier. At this earlier time we can assess species out to nearly 17 S. A close inspection (by zooming in on the graph, which you can't do here) would show that the average *dc/dt* curve is still dropping out to the limit near 17 S, which means this sample does contain traces of aggregates out to at least 17 S. If we want to include those, we would need to start the analysis at least this early in the run. On the other hand, if our primary interest was to measure the properties of the major peak, we should probably go back to the earlier setting, and perhaps even somewhat later in the run.

(5) Once you have decided on the right position within the run (lower slider position) you may need to re-adjust the upper slider. Overall with that adjustment you are trying to find a compromise between loss of signal/noise (as you use fewer scans) and loss of resolution due to peak broadening (as you use more scans).

Now let's do it the easier way, using the Auto Adjust button

If we start from the same initial state shown in the first image (all 60 scans analyzed) and simply push the Auto Adjust button the wizard will produce the result shown below.

Result from using Auto Adjust wizard

As you see, this puts the main peak roughly in the center of the graph, and sets the number of scans included to prevent that peak from being overly broadened. In this case the resulting peak broadening limit is a bit above the actual molecular mass of that main peak.

Thus these are certainly good first-pass settings, but on the other hand by focusing on the main peak the wizard algorithm is picking scans that are too late in the run to cover the aggregates sedimenting faster than 11 S that are present in this sample. So don't rely totally on the Auto Adjust wizard, you need to put your own brain to work too to get the best results!

2.3 Tutorial: Using the g(s*) overlay graph

What data is needed to use the overlay graph?

The first requirement for using the *g(s*)* overlay graph is that you need two or more open analysis windows that contain computed *g(s*)* distributions.

How do I generate the overlay graph?

To display the *g(s*)* overlay graph you can either click on the top node of the Navigation tree, use the entry under the View menu, or use the entry under the Windows menu.

When the overlay window is first opened it automatically graphs all the available data from non-minimized analysis windows; minimized windows are ignored so you can exclude certain distributions without having to close that analysis window. After the overlay window is opened you can force updates of the graph using the **Graph all** button.

The example below shows a comparison graph for the same protein stock at 0.11, 0.32, and 1.1 mg/ml. The intent is to get a quick qualitative picture of any changes arising from reversible self-association.

Overlay plot for a dilution series

The problem with the graph above is that the large difference in sample concentration makes the scales of the three curves very different and therefore it is hard to directly compare them. This is a situation where normalization of the distributions is very helpful.

After going to the Calculate g(s*) distribution page for each of these analysis windows and pushing the Normalize button, when we return to the *g(s*)* overlay plot window we get the following, much improved, result.

overlay plot after normalizing all 3 distributions

Note that the graph updated automatically as we normalized the distributions!

From this graph it is apparent that at 1.1 mg/ml the fraction of the shoulder at ~5.5 S has increased significantly, and the main peak near 3.5 S also shifts slightly to the right. Those changes prove this protein exhibits reversible association. On the other hand, the fact that the shoulder changes little upon diluting from 0.32 to 0.11 mg/ml suggests some of this material may be irreversible rather than reversible.

For such comparisons it is best to be sure the scans chosen for analysis give nearly equal peak broadening limits for all the samples. For samples within the same run that will usually be true if the same scan numbers are used for each sample. Thus once you have analyzed the first sample in that run, an easy approach to generating the appropriate comparisons is to first note which scans were used in for that first analysis (this is displayed in the title bar at the top of that analysis window). Then simply load only that same range of scans for each of the other cells.

The quantitative test for reversible association is comparing the weight-average sedimentation coefficients at different concentrations (see Show integrals over distribution). If the weight-average value increases with concentration, reversible association is present.

How can I use the overlay graph in a publication or presentation?

To use this result in a presentation one other thing you would probably want to do is modify the figure legend to indicate the actual protein concentrations. (The default legend uses the captions of each analysis window.)

You can easily modify the legend text for each distribution by using the up/down spinner control to select the data set and typing a new description into the adjacent text box, as was done to generate the graph below.

overlay graph after modifying the legend

You also can easily modify the plotting style using the radio buttons within the **Graph style** group. For other graph customization options such as adding a title see How to customize graphs.

When you are satisfied with the graph there are buttons provided for printing it and for exporting the image or the actual data table to the Windows Clipboard or a file.

A heterogeneous (mixed) association example

These overlay graphs can also be quite useful in studying heterogeneous associations. The example below shows data for two different proteins when run by themselves and then data when the they are mixed to give the equivalent concentration of each protein. Obviously the new peak near 7 S represents a protein-protein complex, and most of both proteins has been incorporated into complexes.

2.4 Tutorial: Single species fit

This tutorial assumes you have already calculated the *g(s*)* distribution; see the procedure under Tutorial: calculating the *g(s*)* distribution. Here we will fit *g(s*)* data for a monoclonal antibody.

Show All Show All

Hide All Hide All

Show **First, here is the easy way, in Standard mode:**

1. Navigate to the Select fitting model page, shown below.

Select fitting model page

2. If needed set the number of species to 1 (it will probably already be at 1). Be sure to enter the correct solvent density and partial specific volume if you want accurate molecular masses.
3. Push the **Next** button to advance to the Set initial guesses on graph page, shown below.

Set initial guesses on graph page

4. The wizard should automatically find the top of the main peak in the data and enter that as the initial guess

for the sedimentation coefficient. The **Next** button should already be highlighted so just hit **Enter** to advance to the Perform least-squares fit page. This shows you a graph (below) with an overlay of the data being fitted (circles) and the theoretical function corresponding to the initial guesses for sedimentation coefficient, mass, and concentration made by the wizard. In this case those guesses are fairly good ones. The bottom panel of the graph shows the *residual* plot corresponding to those initial guesses.

Perform least-squares fit page

5. The Do fit button is automatically selected so all you have to do hit **Enter** again to complete the fit. As the fitter iterates to improve the fit the graphs will update, but this will probably happen too quickly for you to see.

6. When the fit converges it will automatically advance to the Report fit results page, shown below. This page shows the final best-fit parameters, the values for any parameters that were held fixed, and summarizes the fit statistics.

Report fit results page

7. You may wish to print a report and/or push **Next** again to display various graphs. The default 'species plot' area graph is shown below.

Graph fit results page

8. If a single species does not fit the entire range of data well (for example notice the divergence of the data and the fit above ~8.5 S on the graph above due to the presence of irreversible aggregates in this sample) you may wish to go back to the Set initial guesses on graph page and narrow the range of data to be fitted to just the central portion of that single peak (thus excluding other minor components that might be present).

More details, and Advanced mode

Below is a more detailed and general procedure for a single-species analysis, in advanced mode. It looks more complicated than it is---this takes only a minute or two.

Show 1. Select an appropriate fitting model

A. Navigate to the Select fitting model page. In advanced mode it looks as shown below.

Select fitting model page, advanced mode

B. Set the number of species to 1 (it will probably already be at 1).

C. In the Data choice section choose whether you are going to fit the *g(s*)* or the *dc/dt* data. Fitting to *dc/dt* is sometimes advantageous for very low molecular weight samples (see reference 8).

D. In the Fit diffusion or mass section choose whether you prefer to use diffusion coefficient as a fitting parameter or *M* (the ratio of *s*/*D*) instead.

E. In the Function computation section you will probably want to use the new **improved method**, but you can choose the less-accurate '**classic**' method if you wish.

F. In the Weighting of data points section choose whether you want to use a **weighted fit** (data points with low standard deviation count more) or a **non-weighted fit** (all points count equally). The new **theoretical weights** option is recommended when you are fitting to the *g(s*)* data.

G. If you have chosen *M* as the fitting parameter, be sure you have entered the correct partial specific volume and solvent density. If those values are not well known it may be preferable to work in units of *buoyant molecular mass*.

H. When you are done push the **Next** button to go to the Set initial guesses on graph page.

Show

2. Set initial guesses for peak positions and the range of s values to be included in the fit

This page (below) has the same appearance in both Advanced and Standard modes.

Set initial guesses on graph page, advanced mode

A. If needed drag the magenta markers indicating the initial guess for the sedimentation coefficient to the proper position. The program wizard automatically sets this at the highest point in the curve.

B. Since the Standard mode analysis of these data (above) indicated the presence of some aggregates, this time we'll try narrowing the range to just the center of the main peak. As shown above the the green markers that define the left and right ends of the data to be fitted have been dragged to positions marking roughly 50% of the maximum peak height.

If you want to reduce the influence of other species you should **not** do this by narrowing the range of sedimentation coefficients used in computing the *g(s*)* distribution (that is, do not use the Set range for g (s*) procedure on the Calculate average dc/dt curve page for this purpose). Change the fitting range, not the *g(s*)* distribution itself.

C. Push the **Next** button to go to the Set/alter fitting parameters page.

Show

3. Optionally alter the initial guesses or which parameters are being fitted

The page for altering the initial guesses for the parameter values or changing which parameters will be fitted is shown below.

Set/alter fitting parameters page

A. You may wish to alter the initial guesses for the sedimentation coefficient and *D* or *M* if you think the estimates made from the graph are poor. The estimates for *D* or *M* may be fairly far off when peaks are poorly resolved.

- The fit should converge as long as the initial guess for *s* is not too far off (more than a factor of ~2), but good guesses will always make it converge faster.

B. You might want to fix the values for certain parameters rather than 'floating' them (allowing the fitter to vary them). This may be helpful at the early stages of a complex, multi-species fit, but will probably not be needed for a single species fit. To fix a parameter un-check its box in the 'Fit this parameter' column, and the value will be held at whatever value is entered in the 'Value' text box.

C. A baseline offset parameter is available, but should not be used routinely. See Should I fit the offset?

D. A program wizard automatically chooses upper and lower limits for each fitting parameter (to keep the values bounded within a range that makes physical sense). Usually those limits do not need to be manually adjusted, but if needed you can do so using the Adjust parameter upper/lower limits button.

E. When you are done push the **Next** button to go to the Perform least-squares fit page.

Show

4. Fit the data

When you enter this page the Graphical fit monitor tab is initially shown, displaying an overlay graph of the data and the theoretical curve resulting from the initial guesses, with a residual plot below.

Graphical fit monitor tab

A. Rather than completing the fit automatically using the Do fit button let's try stepping through one iteration at a time using the 1 iteration button. The image below shows the result of the first iteration. Because the initial guesses were close this first step has already produced quite a good fit in this case.

Graphical fit monitor after first iteration

B. Switching to the Fitting parameters monitor tab (by pushing the tab control located along the bottom) allows you to follow the fit progress by monitoring the actual values of the parameters and the reductions of the *sum of squared residuals* or *chi squared* as the fit improves. The image below shows these data during the pause after the first iteration.

Fitting parameters monitor tab after first iteration

C. Either continue using the 1 iteration button until the fit converges, or use the Do fit button to iterate to convergence without further pauses. When the fit converges the program will automatically advance to the Report fit results page.

Show

5. Review fit results

The Report fit results page (below) shows the final best-fit parameters, the values for any parameters that were held fixed, and summarizes the fit statistics. (This page is the same in either Advanced or Standard modes).

Report fit results page

A. Notice that by limiting the fitting range to the center of the main peak the molecular mass value has now risen to 153 kDa versus the 146 kDa obtained when the entire range (including the aggregates) was fitted as a single species (see the 'easy way' fit above). By not fitting the data that are strongly influenced by aggregates (data that obviously can't be properly fitted by a single-species model) we obtain a much better molecular mass estimate. This value is also close to the 150 kDa expected for an antibody monomer (but the mass calibration is probably uncertain by a few percent in this case because neither the density or partial specific volume are accurately known).

B. You may want to push to the **Next** button to generate a species graph, an overlay of the data and fitted curves, or a *residual* plot to examine the quality of fit more closely. The graph shown below, where the theoretical fit is extended outside the actual range that was fitted, helps to show that there is clearly a significant fraction of the sample sedimenting above ~8 S that is not accounted for by this single species. In this case the presence of those aggregates is also easily seen in the area type of species plot, but low levels of minor components are often more easily seen in this line plot.

Species plot for entire data range (line)

C. If you are fitting *D* rather than *M* and wish to know what molecular mass is implied by the returned *s* and *D* values, use the Mass/diffusion calculator menu command or push the calculator button on the Toolbar. The *s* and *D* values from the last fit will automatically be placed onto the form.

D. If you are satisfied with the fit, you will probably want to determine the true confidence intervals for the parameters and/or print out a report, both of which can be done from the Report fit results page.

E. It is highly recommended that you use the Save analysis as... command under the File Menu (or the equivalent Save button on the Toolbar) so you can quickly reload everything about this fit in the future (and continue fitting with altered data or models if you choose to do so).

You are done!

2.5 Tutorial: Multi-species fits

Below is a general procedure for a multi-species analysis, in advanced mode. In many cases you would first do a single-species analysis. However, for some samples it may be obvious from the shape of the *g(s*)* distribution, or known from other data, that two or more species are present and thus starting with a multi-species fit may make the most scientific sense.

A. Calculate the data to be fitted [g(s*) or dc/dt]:

See the procedure under Tutorial: calculating the g(s*) distribution.

For this tutorial we are going to fit the data shown below from a sample that probably contains a small amount of dimer, producing the shoulder on the right side of the main peak as shown below.

g(s*) distribution to be fitted

B. Set the fitting model to the desired number of species:

Select fit model page

1. Navigate to the Select fitting model page (shown above)
2. Set the number of species using the up/down spinner or by entering the number directly in the text box beside it.
    - Generally you would increase the number of species by one for each round of fitting, unless you have independent information on how many species are present.
3. If you have already fitted these data and set the other aspects of the fitting model appropriately, you may skip steps 4-6.
4. In the **Function computation** group box select a function computation method (classic or improved) and from the **Data choice** group select whether you are going to fit the *g(s*)* or the *dc/dt* data.
5. In the **Fit diffusion or mass** group box choose whether you prefer to use diffusion coefficient as a fitting parameter or *M* (the ratio of *s/D*) instead. If you are using *M* then choose the molecular mass units and enter the solvent density and partial specific volumes if needed.
6. In the **Weighting of data points** group box choose the type of weighting for the experimental points, if any. For the improved method theoretical weights (the default) are always preferred.

C. If you believe these species are small oligomers of a common subunit, you may wish to constrain the ratios of the molecular masses and/or sedimentation coefficients for the different species to values appropriate for oligomers.

The use of constraints and how to set them is discussed in detail under Constrain M or s ratios among species. Using these constraints reduces the number of fitting parameters and therefore makes the fit converge much more easily and rapidly, and is particularly advantageous when the fraction of a particular species is low. The use of constraints will also give lower uncertainties for the remaining fitting parameters. However, the danger of the constraints is that your assumptions may be wrong, and if they are wrong the results will be nonsense (even if it appears to be a good fit).

It is obvious that the *g(s*)* distribution shown above shows a shoulder at roughly 5.0-5.5 S indicative of a second species, but this species is in fairly low abundance and is poorly resolved from the main peak at ~3.5 S. The ratio of the sedimentation coefficient for this shoulder relative to the main peak is around the 1.4-1.5 range that is typical for dimers. Therefore it makes sense to try constraining this second species to have the mass of a dimer (twice the monomer mass, whatever that might be). In other words, we will test the hypothesis that these data can be well fitted as a mixture of monomer and dimer.

To implement this constraint we push the Set constraint values button, which brings up the dialog box shown below.

Set M or sedimentation coefficient constraints dialog

To force species 2 to be a dimer, simply click on the radio button labeled **1** within the **Species 2** group box to indicate that species 2 is constrained as a multiple of species 1, as shown below. You can then enter the appropriate ratio in the text box to the right (the default in this case is already 2.000).

Constraint dialog box after species 2 is constrained as dimer

It would also be possible to constrain the sedimentation coefficient of our putative dimer to be an appropriate multiple of the monomer sedimentation coefficient, but because sedimentation coefficients depend on conformation it is hard to predict precisely what that ratio should be. However if a species is present only in a very low amount constraining its sedimentation coefficient may be necessary to get the fit to converge.

Note that in some situations it is better to constrain a species to be a fragment of another (mass ratio less than 1). For example, suppose you had a protein where the native structure is a tightly-associated dimer, but your sample contains a small amount of a partially-denatured "incompetent" (non-associating) monomer. In that situation you would want to constrain the monomer to be half the molecular mass of the dimer. That is, the "reference" species should be one that comprises a substantial fraction of the total, and whose properties will be well-determined by the fit, rather than a minor component.

The **Ok** button will then close the dialog and set the constraint, returning you to the Select fitting model page. When you return the **Constrain *M* or *s* ratios among species** check box will be checked to remind you that one or more constraints is active.

Select fitting model page after constraint has been set

D. Set the initial guesses for sedimentation coefficients and the range of s values to be included in the fit:

The **Next** button will bring you to the Set initial guesses on graph page.

Set initial guesses on graph page

When you first enter this page with new data, a program wizard will find the top of the main peak and use that position as the initial guess for species 1. It then positions the sedimentation coefficients for any additional species in positions appropriate for a dimer, trimer, tetramer, hexamer series.

1. If needed drag the magenta markers indicating the initial guesses for the sedimentation coefficients to the proper position.
    - In this case the wizard has done all the work for us and these estimates are reasonable starting values for our monomer + dimer mixture.
2. If needed drag the green markers that define the left and right ends of the data to be fitted. Your sample might, for example, contain aggregates or small fragments that you wish to exclude from the analysis by narrowing the range of sedimentation coefficients to be used for fitting.
3. Push the **Next** button to go to the Set/alter fitting parameters page.

E. Optionally alter the initial guesses or which parameters are being fitted:

Set/alter fitting parameters page

1. You may wish to alter the initial guesses for the sedimentation coefficient and *D* or *M* if you think the estimates made from the graph are poor. The estimates for *D* or *M* may be fairly far off when peaks are poorly resolved.
    - The fit should converge as long as the initial guesses for *s* and *M* are not too far off, and provided there really is sufficient information in the raw data to uniquely define that species. However good initial guesses can make the difference between converging or not converging, and will always make it converge faster.
2. You might want to fix certain parameters rather than 'floating' them (allowing the fitter to vary them). This

can be very helpful at the early stages of a complex, multi-species fit, but should not be needed here with only 2 species plus a constraint.

3. When you are done push the **Next** button to go to the Perform least-squares fit page. In this case we will simply accept the guesses made by the program wizard.

F. Fit the data:

When you enter this page the Graphical fit monitor tab is initially shown, displaying an overlay graph of the data and the theoretical curve resulting from the initial guesses, with a residual plot below.

Graphical fit monitor

In this case the initial guesses seem to give us a reasonable starting point. If you see a theoretical curve that is grossly different than the experimental data you should probably go **Back** and modify the initial guesses and/or the fitting model.

Fit the data (use the **Do fit** button or step through 1 iteration at a time using the **1 iteration** button). Hopefully you will see the theoretical curve move to overlap the experimental data.

If it appears the fit is having trouble converging you may wish to switch to the Fitting parameters monitor tab to follow how the parameter values are changing. If you are trying to fit more species than are justified by the data it is likely that the concentration of one species may be dropping to zero, or its sedimentation coefficient may be trying to go out of the range covered by the data. If so halt the fit using the **Cancel** button and go **Back** 3 steps to modify your fitting model.

When the fit converges it will automatically advance to the Report fit results page.

G. Examine whether this seems to be a reasonable fit:

Report fit results page

The fit results in our example indicate this sample consists of about 90% of a 58 kDa monomer at 3.56 S plus about 10% dimer at 5.45 S.

The first thing we want to know if whether this monomer-dimer mixture model actually gives a good fit of the data. One way to examine the quality of fit more closely is to push to the **Next** button to generate various graphs of the fitting results.

Species plot

The image above shows a species graph summarizing this fit. There are no obvious regions where these two species do not explain the experimental data fairly well, except possibly in the region from ~1.4 to 2.2 S.

Residual plot

Similarly the residual plot above suggests this is statistically a fairly satisfactory fit of these data except for the region below 2.2 S. The deviations below 2.2 S might indicate there is a trace of a third slowly-sedimenting component, but note that these are interference data and thus the low sedimentation coefficient region is subject to systematic errors if the *jitter* is not perfectly removed. See also What constitutes a good fit?

Often though a more significant guide to whether this is a reasonable fit is based on whether the results make scientific sense. For example, is the returned monomer mass reasonable based on a known sequence mass or mass spectrometry data? Is the sedimentation coefficient reasonable for that mass?

In this example we find that the ratio of the dimer to monomer sedimentation coefficients is 1.53. That is somewhat higher than the 1.45 ratio for a dimer of spheres, but certainly seems to fall within a range that is hydrodynamically reasonable.

H. If you were using constraints on *M* or *s* values, you may wish to relax the constraints one-by-one and refit:

Relaxing the constraint is one way to test whether the hypothesis that is implicit in the constraint is really true. If the species are well-resolved and none of their concentrations are too small, it should be possible for the fit to converge even with one or all the constraints removed (particularly when you start with values giving a reasonably good fit). If the constrained parameter(s) do not change much when the constraints are removed (and especially if the confidence interval for the unconstrained parameter includes the value implied by the constraint), you can be reasonably confident that your hypothesis that these species are oligomers is true.

With regard to constraints on sedimentation coefficients, it is important to remember that the ratios for oligomers are sensitive to the shape and configuration of the oligomers, and thus it is hard to predict the "correct" ratio. For this reason, for "final" results it may be desirable to relax such constraints. However, if you are trying to detect small amounts of one species, and are more interested in knowing the amount of this species than its hydrodynamic properties, you will obtain lower uncertainties for the concentrations (and for the hydrodynamic properties of other species) if the *s* ratio constraint is used.

In this example, if we remove the constraint on the mass of the second species and re-fit the data, the fit does converge, but in some cases the constraint(s) may be required for convergence. Here we get the following results:

Fit results after removing constraints

This un-constrained fit is not too different from the earlier one, although the best-fit mass for species 2 has now dropped to be only ~60% larger than species 2. Note however the very wide confidence interval on that parameter.

In this sample at this concentration there is probably some reversible dimer present as well as an irreversible one, and the dynamic interconversion of the reversible dimers might be affecting the calculated masses. This too is a good lesson---be cautious about interpreting the results unless you are certain your sample is truly behaving as a mixture!

I. If the fit is still unsatisfactory, you may wish to try increasing the number of species and fitting again.

It is certainly not always easy to judge whether systematic residuals really represent another species or are just due to experimental error. Based on simulations, in principle under favorable circumstances one should be able to resolve three species without knowing anything in advance about the hydrodynamic properties of any of them, if all are present at levels over 10-20%. Even under favorable circumstances, however, determining four or five species is probably impossible unless you know (and fix during the analysis) the properties of one or more of them.

Also, remember that failure to obtain a good fit (or implied molecular masses that make no sense) may indicate that the species are not truly independent and that they interact and/or inter-convert.

J. When you are satisfied with the fit, you will probably want to determine the true confidence intervals for the parameters, print out a report, and save the fit to disk.

Computing confidence intervals and printing reports is done from the Report fit results page.

It is highly recommended that you use the Save analysis as... command under the File Menu (or the equivalent Save button on the Toolbar) so you can quickly reload everything about this fit in the future (and continue fitting with altered data or models if you choose to do so).

TIP 1: If you are uncertain whether a particular peak in your data represents a real species (as opposed to noise or an artifact), you may wish to increase the number of scans used in calculating *dc/dt* by a factor of two or more to improve the signal/noise ratio, even if doing so results in significant broadening of the peaks. The point of adding the additional scans is to see whether this additional peak persists through a larger portion of the experiment, or instead just fades away as additional data are added. Once you are sure of the qualitative conclusion that the species is real, before actually fitting you will probably want to reduce the number of scans again to avoid excessive broadening.

TIP 2: Sometimes samples contain a broad distribution of either very large or very small species (aggregates or small contaminants). Ignoring these species may distort the results for the species of interest, yet trying to fit many, many species is also unreasonable. Under such circumstances, try limiting the range of *s* values used in the fit to exclude contributions from such species. If this does not work, consider modeling the group of large or small species as a single species with some average *s* value and a very high diffusion coefficient. The drawback to the latter approach is that the broad peak used to model the group of species is likely to overlap the peak(s) for the species of interest, and therefore the accuracy of the returned properties for those major peaks may be poor.

2.6 Tutorial: Simulating experiments with the Claverie simulator

This tutorial first covers how to simulate a sample containing a single species, and then how to do multi-species simulations.

Show All Show All

Hide All Hide All

Show Step-by-step: single species simulation

1. Decide if you are simulating absorbance, interference, or fluorescence data and check the appropriate radio button (at the lower right on the form).
    - This choice affects the units of concentration, the default number of radial points, the allowed range for the loading concentration and added noise, and the file names for the generated data sets. For absorbance mode the default 601 data points corresponds to approximately the number of data points actually collected when using the default 0.003 cm point spacing for velocity scans. For interference and fluorescence modes the default 1641 and 601 data points, respectively, correspond approximately to the radial spacing of of those optical systems.

2. Decide if you are simulating conventional cells or a synthetic boundary cell. For a conventional cell the inner radius and the meniscus position are set to be equal; for a synthetic boundary cell the meniscus is at a radius greater than that of the cell inner radius.
    - Making the meniscus and inner radius equal for conventional cells essentially prevents diffusion of solutes to the region above the meniscus. In a synthetic boundary cell solutes can diffuse to radii inside the initial boundary position.

3. Enter the desired concentration, sedimentation coefficient, and diffusion coefficient for the species being simulated.

- If you are trying to simulate a particular molecular mass, use the Mass/diffusion calculator form to calculate the corresponding diffusion coefficient.

4. Enter the run time (elapsed time) for the first data set, the time interval between data sets, the total number of data sets (1000 maximum), and the rotor speed. (See **Tip** below.) For *dc/dt* analysis you will want the interval between sets to be fairly small (perhaps 60 s to simulate interference data or 240 s to simulate absorbance data in a 4-hole rotor).
5. Enter the calculation interval in seconds. This is the time increment between successive calculations of the concentration distribution.
    - Smaller intervals will give a more accurate, but slower, simulation. Faster moving boundaries will require smaller intervals; for low speeds or small sedimentation coefficients larger intervals are fine. A 1 sec interval should give good accuracy in most cases, but for sharp boundaries to get accurate boundary widths (diffusion coefficients) you may need to go to ~0.1 s. The minimum interval is 0.01 sec, maximum 50 sec.
    - **For highly accurate simulations of conditions giving sharp boundaries you will need a small interval and the simulation will likely require up to a minute to complete.** Absorbance simulations require smaller intervals because the radial spacing is larger (but they calculate faster at a given interval for the same reason).
6. Enter the number of data points to cover the radial range between the inner radius and cell bottom. Again, more points gives more accuracy but a longer calculation.
    - Sharper boundaries will require more radial points for accuracy. The data files that are generated will have data at these radial positions.
7. If desired, specify a nonzero value for the concentration dependence coefficient of the sedimentation coefficient.
    - This algorithm assumes a linear dependence of *s* with concentration, given by the formula , where *k* is the coefficient entered on the form. For globular proteins an empirical value for *k* is often taken to be 0.009 ml/mg (corresponding to *c* in mg/ml). In this program the units for *k* are per AU or per fringe. The *D* value has no concentration dependence.
8. If desired, specify a nonzero value for the random Gaussian noise to be added to the simulated curves.
9. Press the 'OK' button to start the simulation. The progress bar will show approximately how far the calculations have progressed, and they can be aborted using the 'Cancel' button.
10. When the simulation is done, the data are automatically loaded into the internal data arrays, just as though they had been read in as scan files from a disk. You will be asked whether you wish to also save these results to disk as scan files so they could be reloaded at a later date.
    - If you save the scans you may specify which folder to put them into using the File Save dialog box that appears. This dialog box also allows you to create a new folder in which to store the simulation. The scans are normally written in sequence starting with 00001. You will also be asked to provide a comment line for the scans. Also please note that in order to provide an obvious meniscus position for simulations of conventional cells, when the data are written to disk two extra data points to the left of the meniscus are added, and the data point at the meniscus position is set to a value equal to the loading concentration. This produces a 'spike' at the meniscus, similar to that seen in real absorbance data.
11. Finally, you will automatically be taken to the Set meniscus & data region page to view a graph of the simulation you just created.
    - Note that in simulated data the concentration at the cell base often builds up to very high values. Thus the graph will be automatically zoomed to give a range from 0 to the loading concentration (which means the region where solutes accumulate will go off scale).

Finding the right scan interval and number of scans to get the proper range of boundary movement can take some trial and error. Do some quick, low accuracy, simulations with a long calculation interval to get the time range correct, and then do a final accurate simulation with a smaller time interval. If you are going to be summing data for more than one species, remember you will want to pick a run time that is appropriate for all the species.

Show

Step-by-step: multi-species simulation

Multi-species simulations are done by summing single-species simulations. This feature is governed by the 'Add to previous Claverie data' check box near the bottom of the form. This check box becomes active when Claverie data are currently in memory. It is only possible to sum simulations that generate compatible data sets (same radii, start time, *etc.*). That is, in summing simulations the only parameters changed between species are their *s* and *D* values and concentration.

The steps for a multi-species simulation are:

1. Perform a simulation for the first species.
    - You may save the results to disk if you want to keep that data, but it is not necessary to save to disk.
2. Immediately after the first simulation, invoke the Claverie simulation command again.
3. Enter the desired concentration, sedimentation coefficient, and diffusion coefficient for the next species.
4. Be sure to check the 'Add to previous Claverie data' check box (it is automatically cleared when the form is loaded).
5. Run the simulation for the next species.
6. Note that the graph being displayed shows only the current simulation being generated, not the sum.
7. If there are additional species, return to step **2**.
8. After the last species has been calculated, you may wish to save the summed data to disk.

If you are going to add noise, you will probably not want to add it to each species, but rather add it only to the last species. If you added the same amount of noise to each species, the total r.m.s. noise will grow as the square root of the number of species.

3 Program elements, controls, and features

3.1 User interface

3.1.1 Main menus

3.1.1.1 File menu

New Analysis Window

start a fresh analysis in a new Analysis window. This menu command is duplicated by a button on the Toolbar. Shortcut = CTRL+N

Clone Analysis to New Window

copy the current analysis into a new analysis window so it can be modified without affecting the current window. This menu command is duplicated by a button on the Toolbar.

Close Current Analysis Window

close the active analysis window

Close All Windows

close all open analysis windows

Open Saved Analysis...

restores results and/or fits previously saved to disk. This menu command is duplicated by a button on the Toolbar. Shortcut = CTRL+O

Save Analysis

update analysis previously saved to disk. This menu command is duplicated by a button on the Toolbar. Shortcut = CTRL+S

Save Analysis As...

assign file name and save dc/dt and g(s*) curves and fitting results (if any) in a permanent record on disk

Exit

exits program after reminder to save unsaved or altered analyses to disk (if there are any)

(Recent Files)

list of 6 most recently saved or restored analysis files; click on these entries to restore that analysis (in a manner equivalent to the Open Saved Analysis... command)

3.1.1.1.1 Open saved analysis...

This command will restore the data, results, and states of various program variables from dc/dt or g(s*) calculations and/or fitting results that were previously saved using the Save Analysis As... command. This menu command is duplicated by a button on the Toolbar.

The command brings up a standard Windows® file dialog that allows the user to specify the file name and folder of a previously saved ".dcdt" file (saved by version 2) or ".fit" file (saved by version 1).

After the file is loaded, if the saved results include a completed fit, the program will bring up the Report fit results page to show the results from that fit. Otherwise the program will display a graph of the _g(s*)_ or *dc/dt* results, depending on how many steps toward calculating _g(s*)_ had been completed.

The path to this restored fit will be added to the top of the recent files list at the bottom of the File menu.

Tip 1: Once you have opened a saved analysis, you may want to use the Clone analysis to new window command to create a second copy that can be altered while preserving the original analysis in an open window for reference.

Tip 2: Since the .dcdt file extension will be associated with this program, you can also restore a fit by double-clicking on the file name from Windows Explorer, or in the Recent Documents list on the Start Menu.

Note that when a saved analysis is loaded, the program attempts to set the default folder for scan file loading to the folder from which the data were originally loaded. If that folder is not currently available (for example if the data were on a removable drive or a network drive to which you are not currently connected) there is often a delay of several seconds while Windows tests for what drives and folders are currently available.

3.1.1.1.2 Save analysis

This command saves an updated version of an analysis that was previous saved using the Save analysis as... command. This command is enabled when the previously saved version is altered.

3.1.1.1.3 Save analysis as...

This command will write out a file that contains all the data (raw scans, dc/dt and g(s*) curves) and any fit results that have been obtained, up to and including the last calculation or fit that was performed. This provides a permanent record of your work that can be quickly reloaded if further processing is needed or to print graphs and reports, and to serve as an archival record of your work.

The command brings up a standard Windows® file dialog that allows the user to specify the file name and folder. The user can specify a filename of any length; the program will automatically add the extension of ".dcdt" to that name.

This command does not become active until after at least the individual dc/dt curves have been calculated (Select scans to analyze page).

The data file is written in a proprietary binary format. All scans which have been loaded into memory are saved, independent of whether that scan is presently included in the analysis. The scan data require only 8 bytes of storage per data point (much more compact than the original ASCII scan files).

The path to this saved results file will be added to the top of the recent files list at the bottom of the File menu.

3.1.1.1.4 Clone analysis to new window

This command will open a new analysis window containing an exact copy of the current analysis window. That copy includes all the raw data (in memory, not the original scan files on disk), calculated *dc/dt* and *g(s*)* data, settings, *etc*. This menu command is duplicated by a button on the Toolbar.

This allows you to alter the analysis while preserving an open copy of your original analysis. For example, you might want to analyze scans early or later in the run, and then overlay that result with the earlier one in the g(s*) overlay graph.

The name of the cloned window is derived by adding an underscore character and a sequence number to the name of the current analysis. Thus the first cloned version of a window called "cell2" would be "cell2_1".

An alternative way of altering an analysis while preserving a saved original version is to make the alterations and then save the new version under a new name using the Save analysis as... command. However with that approach you might accidentally overwrite the old version if you forget to alter the name.

3.1.1.2 Edit menu

These commands export data from the current analysis as text tables. The data can either be copied to the Clipboard so these values can be pasted into some other program, or written to disk as text (ASCII) files.

These Copy or Write menu items under the Edit menu

Main edit menu

lead to a submenu, as shown below, that determines which data to export:

Edit menu

When writing to disk, a standard File Save dialog allows you to specify a file name and the folder for the file. By default the file extension will be .dat, but you can also select .txt or any other extension.

When the data is exported the columns are separated by tab characters, and rows are ended by a carriage return and line feed.

Scan data from graph

This menu selection will export the scan data (signal versus radius) from the graph shown on the Set meniscus & data region page. For interference scans this will give the data after removal of jitter and fringe jumps if that is what is currently shown on the graph. This menu item is only enabled when this graph is currently being displayed.

If the graph shows *N* scans then the data table will have 2*N* columns, with the first two columns holding the x and y values for the first scan, the 3rd and 4th columns holding the x and y values for the second scan, *etc*.

Note that copying the data into a large text string can take a minute or more when the number of scans is large. An hourglass is displayed while the text string is being created.

Because absorbance scans have different numbers of data points some of the columns in the data table will have fewer entries than others. Table positions with non-existent data are copied as empty strings.

***dc/dt* data (all scan pairs)**

This menu selection will copy the *dc/dt* data from each scan pair.

If there are *N* scan pairs then the data table will have 2*N* columns, with the first two columns holding the x and y values for the first scan pair, the 3rd and 4th columns holding the x and y values for the second scan pair, *etc*.

Note that copying the data may take 10 seconds or more when the number of scans is large.

Because individual *dc*/dt curves have different numbers of data points some of the columns in the data table will have fewer entries than others. Table positions with non-existent data are copied as empty strings.

Average dc/dt curve

This menu selection will copy the average *dc/dt* curve for the scans currently in use (3 columns with x, y, and standard deviation values).

g(s*) distribution

This menu selection will copy the *g(s**) distribution table (3 columns with x, y, and standard deviation values).

g^(s*) distribution

This menu selection will copy the *g^(s**) distribution table (3 columns with x, y, and standard deviation values).

The menu items below are only enabled after a fit or fit simulation has been completed.

Data being fitted

This menu selection will copy the actual data values being fitted (x,y pairs).

Calculated fit

This menu selection will copy the theoretical values from the fitted functions (x,y pairs).

Residuals

This menu selection will copy the residuals from the fit (x,y pairs).

Species table

This menu selection will copy the total of all species, followed by the contribution from each individual species (x, total, species1, species2,...).

This menu item is only enabled if 2 or more species were fitted.

3.1.1.3 View menu

View menu

The **View menu** provides one of several means to display the *g(s*) overlay graph*. This menu item will only be active when two or more analysis windows that contain *g(s*)* distributions are open (enough data to compare and overlay).

This menu also allows you to toggle on or off the display of the Toolbar and the Navigation Tree. Turning off the Toolbar and/or the Navigation Tree will provide additional screen space for display of multiple analysis windows.

3.1.1.4 Tools menu

The **Tools menu** contains submenus to bring up either of two tools:

1. the Mass/diffusion calculator for calculating masses when the sedimentation and diffusion coefficients are known, or diffusion coefficients when the mass and sedimentation coefficient are known
2. the Options dialog for setting user preferences (defaults) for various program options

Each of these menu items is mirrored by an icon on the Toolbar.

3.1.1.4.1 Mass/diffusion calculator

This View Menu command or its equivalent button on the Toolbar will bring up a small calculator-like form (shown below) which

1. **calculates solute molecular masses** given values for sedimentation and diffusion coefficients, solute partial specific volume, solvent density, and temperature.
2. **calculates diffusion coefficients** if the mass and sedimentation coefficient are known.

mass/diffusion calculator

This form is a handy way to determine the molecular mass corresponding to the *s* and *D* parameters from a fit where *D* is used as the fitting parameter, or to find the *D* value corresponding to a molecular mass when *M* has been used as a fitting parameter, without changing the fitting model and re-fitting.

Show All Show All

Hide All Hide All

Show How it works

This dialog box implements the Svedberg equation:

where *R* is the gas constant, vbar is the partial specific volume, ρ is the solvent density, and *T* is temperature (in Kelvin).

The user may simply enter any desired values and the corresponding molecular mass will be calculated. When numbers are manually entered in the molecular mass box, then the corresponding *D* is calculated (and the label of this *D* value changes to 'calculated *D*'.

Show Species buttons

After a fit has been completed, a number of buttons at the top of the form will become active to correspond to the number of species in that fit (2 in the case shown above). Pressing these buttons will fill in the parameters from that fit. The partial specific volume value associated with that species will also be used if that has been filled in on the select fitting model page.

Show Copy button

The 'Copy all values to Clipboard' button will copy the labels and values for these entries as a small table to the clipboard, with the labels and values separated by tab characters and a separate line for each type of entry.

Show About temperature

The program automatically fills in the sample temperature based on the temperature recorded in the header of the last scan used in calculating the *dc/dt* data, but since the temperature may drift during the run you may wish to adjust that value manually. The program internally converts the temperatures from Celsius to Kelvin scales.

Please note that changing the temperature on this form overrides the temperature recorded with the scans, and also alters the value used elsewhere in the program (for example, for conversions from raw *s* to *s(20,w)*).

Show About s(20,w) units

If conversion to *s(20,w)* units is currently active, the sedimentation and diffusion coefficients will be *s(20,w)* and *D(20,w)* values, the vbar value will be value at 20 °C, the solvent density will automatically be set to that of water at 20 °C (and cannot be altered), and the temperature will be set to 20 °C (and cannot be altered). While conversion to *s(20,w)* units is active, changing the vbar on this form will not alter the vbar value used elsewhere in the program.

3.1.1.4.2 Options dialog

The **Options** item on the Tools menu will bring up the **Program Options dialog box** which uses multiple tabs to list various user preferences for program defaults. The options on each tab will be described below, with the default setting at program installation noted in square brackets [].

The **Cancel button** will close the dialog without altering any settings. The **Ok button** will save any new settings (which in general apply only to newly-opened analysis windows).

Show All Show All

Hide All Hide All

Show Default sample & solvent properties tab

- Sample partial specific volume [0.73 ml/g]
- Solvent density [0.998234 g/ml] - the default is the density of water at 20 °C
- Solvent viscosity [1.002 cp] - the default is the viscosity of water at 20 °C. Note that these values are in centipoise, not poise as used in SEDNTERP and SEDFIT.

These default values are used for the first analysis window opened each time the program is started. When subsequent analysis windows are created, the last values that were entered in any previous window are used, which saves time when you are analyzing multiple samples in the same buffer or of the same protein.

Show *dc/dt* options tab

- Default meniscus clearance [0.015 cm] - This is the distance between the meniscus and the start of the data to be analyzed.
- Default data cutoff at cell base [7.15 cm] - Data beyond this radius will not be analyzed because that region is influenced by the accumulation of solutes at the base of the cell.
- Default cell region to test for and correct integer fringe shifts [middle half] - see Set meniscus & data region - fringe jump removal options
- Use Philo (2000) algorithm to compute g(s*) by default [off] - see What is the difference between the broad and conventional g(s*) algorithms?

Show Fitting options tab

- Confidence probability for parameter confidence limits [68.3%] - This sets the default confidence level for parameter confidence limits. **Usually this is the only setting on this tab the user will want or need to alter.**
- Fit converged when parameters change less than [0.05%] - The fit iterations will halt when the last iteration changes all parameters less than this amount. See convergence control.
- Fit converged when variance change < % of statistical limit [10%] - This provides a second criterion for fit convergence, halting the iterations when the improvement of the fit is no longer statistically significant. See convergence control.
- Parameter cross-correlation defined as "high" [97%] - cross-correlations between this level and the "very high" setting will produce a mild warning on the Report fit results page and in printed reports. See What does high parameter cross-correlation mean?
- Parameter cross-correlation defined as "very high" [99%] - cross-correlations above this setting will produce a stronger warning on the Report fit results page and in printed reports. See What does high parameter cross-correlation mean?
- Maximum number of fit iterations before quitting [60] - This defines the maximum number of iterations before the fit halts with a warning message. Generally if a fit has not converged within 60 iterations it probably never will converge, and this limit is to prevent infinite or near-infinite loops.

Show Fitting model defaults tab

- Fit diffusion or mass [fit M] - this default governs whether to fit the diffusion coefficient for each species or the species molecular mass (the *s*/*D* ratio). See Select fitting model - fit diffusion or mass
- Molecular weight units [true molecular mass] - this default sets the units to be used for molecular mass. See Select fitting model - molecular mass units

Show File locations tab

- Default location for scan files [remember last location] - These options govern which folder is initially displayed in the File Open dialog when new scan files are loaded.
  - The "Always start in this folder" option specifies that scan browsing always begins in a particular specified folder. For example, this might be a network drive, or a folder where all experiments for an individual XL-I user are stored. When this option is selected the Browse button is active and may be used to select the default folder.
  - The "Remember last location and start there" option has the program remember the last folder from which scans were loaded and start there.
- Default location for saved analyses [store with scans] These options govern which folder is initially displayed in the File Save dialog during the Save Analysis As... procedure.
  - The "Always store in this folder" option specifies that the File Save dialog always begins in a particular specified folder, such as the user's "My Documents" folder. When this option is selected the Browse button is active and may be used to select the default folder.
  - The "Store in same folder as scan files" option has the File Save dialog start in the folder from which the last scans were loaded. That approach has the advantage of keeping the saved analysis together with the raw data.

Show Other options tab

- Initial size of analysis windows [default 100%] - this sets the size of newly-created analysis windows and other program dialog boxes as a percentage of the program default size. Smaller sizes will allow more analysis windows to be open without crowding, larger sizes will enhance readability (font size increases with window size). Allowed range 50%-150% of normal.
- Always start in standard mode [default checked] - if checked new analysis windows will be in Standard mode; otherwise they will start in Advanced mode.
- Always start in Scan Preselect mode [default checked] - if checked when loading scans for analysis the File Open dialog box will show only scans from the cell and data type that are pre-selected
- Font size for reports [12] - the font size, in points, to be used in printed reports of fitting results. Report text is always printed in the Arial font.

- Printer for reports [Windows default printer] - The text box shows the name of the current printer to be used for reports of fitting results. To alter this default use the **Browse...** button to open a Print dialog box and select from the drop-down list of available local or network printers.

3.1.1.5 Windows menu

The **Windows menu** provides control over the display of the analysis windows and the g(s*) overlay graph.

Window menu

New Analysis Window

This will create a new empty analysis window, and is equivalent to the File...New in the File menu or the New button on the Toolbar.

Show g(s*) Overlay Window

This will open the *g(s*)* overlay graph if it is not yet open, or bring it to the foreground if it is minimized or hidden by analysis windows. This command is only available when two or more analysis windows contain *g(s*)* data.

Close All Windows

This provides a quick way to close all analysis windows (prompting you to save any containing unsaved results) and the *g(s*)* overlay graph, freeing up screen space for new analyses.

Cascade

This is the usual Windows command to make all the open windows have a uniform size and arrange them is an overlapping cascade.

Arrange Icons

This is the usual Windows command that arranges the icons corresponding to any minimized windows along the bottom of the document area.

(Window List)

The bottom of this menu contains a list of all open windows, with a check mark next to the active window. You can activate a different window and bring it to the front by clicking its name, or by pressing the corresponding number on the keyboard.

3.1.1.6 Help menu

Help menu

Contents

This brings up the Table of Contents tab for this Help file, and opens the Program Overview and Organization page.

Index

This brings up the Index tab for locating specific topics in the Help file.

Search

This brings up the Search tab for locating specific words or phrases that are not in the Index.

About DCDT+...

This brings up a dialog box showing the current version, registration information, and how to contact the author.

Registration

This brings up a tabbed dialog box with information about why and how to register the program, a page for entering and validating your program serial number, and buttons to display the software license agreement or print an order form.

Registration dialog

Check for updates

This opens the Download page in your web browser.

Program web site

This opens the main DCDT+ web page in your web browser.

E-mail author

This opens an e-mail for your questions, comments, or suggestions.

3.1.2 Toolbar

The **Toolbar** provides buttons as convenient alternatives for some of the menu commands.

The usual New, Open, and Save buttons are equivalents of the **New Analysis Window**, Open saved analysis..., and Save analysis as... items in the File menu. (Once an analysis has been saved and assigned a file name, the **Save button** becomes equivalent to the Save analysis command rather than Save analysis as....)

The **Clone button** will copy the current analysis into a new window so it can be modified without affecting the current window.

The **Calculator button** is equivalent to the Mass/diffusion calculator entry in the Tools menu and pops up the Calculate Mass or D dialog box.

The **Options button** is equivalent to the Options dialog entry in the Tools menu and pops up the Program Options dialog box for setting user preferences.

The **Help button** is equivalent to the **Index** entry in the Help menu and will bring up the index for the program Help file.

The display of the **Toolbar** is optional. It can be turned off to provide more screen area for Analysis windows using an entry under the View menu.

3.1.3 Analysis window

This program allows multiple analyses to be open at one time. For example, those might be:

1. different time spans of data from a single cell and run
2. multiple cells from one velocity run
3. data from multiple velocity runs

Each of those analyses has an independent analysis window within the main program screen area. Each analysis window shows up to 11 different pages of controls and graphs, depending on the steps within the analysis that have been completed and/or which results you wish to see.

Those different pages occupy the area that is outlined and washed-out within the image below. The washed-out area shows the first page, the Load raw scans page, a page used for loading the scans to be analyzed.

Analysis window

Moving from step to step

Movement among the steps (pages) within one analysis window can be accomplished using either

A) the Next button and Back button at the lower right,

B) the Navigation tree, or

C) by right-clicking with the mouse on the analysis window (in between the buttons and other controls) and using the **Go to page** pop-up menu shown below. This context menu also has commands for saving this analysis to disk, closing this analysis window, or closing all analysis windows except this one.

Go to step popup menu

Additional tools along window bottom

Along the bottom of each window is a panel containing the or buttons to toggle between Standard and Advanced modes,

a **progress bar** that indicates the fractional completion of data loading or lengthy calculations, and

a User Prompt area that displays instructions or other information. This panel is part of every analysis page.

The panel containing the **Next/Back** and **More/Less buttons**, progress bar, and user prompt can optionally be moved to the top of the analysis window. Right-clicking the mouse within this panel will bring up a context menu that will toggle this panel between top and bottom locations.

Switching between analysis windows

When multiple analysis windows are open you can switch to a different analysis window by either clicking on that window with the mouse, using the Navigation tree, or using the Windows menu.

About window names

New analysis windows are initially named 'Analysis1', 'Analysis2', etc. The current name for each window is shown at its top in its Title Bar area, and also in the corresponding root node within the Navigation tree. Once an analysis has been saved to disk (see Save Analysis As...) then that path and filename are displayed in the Title Bar. As in the example shown above, when the full path and file name are too long to display, the path is abbreviated.

After scans have been loaded and selected for analysis, the names of the first and last scan that are being used are also shown in that window's Title Bar (.IP3 scans 120 through 164 in the example pictured above).

3.1.3.1 Next and Back buttons

The Next button and Back button **buttons** move forward or back one analysis step (page) at a time.

If the **Next button** is currently disabled this indicates the program is waiting for you to complete some operation on this page that is required before moving to the next step (loading scans, setting the meniscus position, completing a fit, *etc*.). The **Back button** will always be disabled on the first page (Load raw scans page).

3.1.3.2 More >> and << Less buttons

The and buttons toggle between Standard and Advanced modes.

The button text changes to indicate whether **More** options will be shown when the button is clicked (switch to **Advanced mode**) or **Less** options will be shown (switch to **Standard mode**).

3.1.3.3 Log page

As you perform an analysis DCDT+ creates a Log to record exactly what you have done, and when it was done. This Log can be displayed at any time using the Log button button along the bottom of the Analysis window.

Log page

The example above shows the first portions of the Log generated during the analysis of some absorbance scans. As you see, each of the steps taken, analyst inputs, and options selected is documented. If the *g(s*)* distribution is fitted, the full report from that fit (including a summary graph for the fit) is placed into the Log.

Show The full Log from the above analysis is contained in this drop-down section as an example.

```
[4/26/2012; 20:48:39]

Loaded 54 new scans from folder C:\XLAWIN\XLIDATA\antibody:

    00001.RA1, 00002.RA1, 00003.RA1, 00004.RA1, 00005.RA1, 00006.RA1, 00007.RA1, 00008.RA1, 00009.RA1, 00010.RA1, 00011.RA1,
    00012.RA1, 00013.RA1, 00014.RA1, 00015.RA1, 00016.RA1, 00017.RA1, 00018.RA1, 00019.RA1, 00020.RA1, 00021.RA1, 00022.RA1,
    00023.RA1, 00024.RA1, 00025.RA1, 00026.RA1, 00027.RA1, 00028.RA1, 00029.RA1, 00030.RA1, 00031.RA1, 00032.RA1, 00033.RA1,
    00034.RA1, 00035.RA1, 00036.RA1, 00037.RA1, 00038.RA1, 00039.RA1, 00040.RA1, 00041.RA1, 00042.RA1, 00043.RA1, 00044.RA1,
    00045.RA1, 00046.RA1, 00047.RA1, 00048.RA1, 00049.RA1, 00050.RA1, 00051.RA1, 00052.RA1, 00053.RA1, 00054.RA1

[4/26/2012; 20:48:40]

Scans despiked, spike threshold = 0.2.

[4/26/2012; 20:48:42]

Meniscus position set as 5.9521 cm, as suggested by the Meniscus Wizard.

[4/26/2012; 20:48:49]
```

Maximum radius being fitted set to 7.070 cm.

[4/26/2012; 20:48:54]

Scans used for calculating dc/dt were set by the Auto Adjust wizard to 12 total scans, from 00022.RA1 to 00034.RA1

[4/26/2012; 20:49:08]

Lower sedimentation coefficient limit for g(s*) computation was manually set to 0.3 S.

[4/26/2012; 20:49:18]

Current fitting model is 1 species, fitting the g(s*) data, using the improved method. The species molecular parameters are sedimentation coefficient and molecular mass (in true mass units).

[4/26/2012; 20:49:36]

The vbar for all species was changed to 0.7326mL/g.

[4/26/2012; 20:49:40]

Solvent density changed to 1.003000 g/mL.

[4/26/2012; 20:49:56]

Lower limit of fitting range was set to 4.95 S.

[4/26/2012; 20:50:02]

Upper limit of fitting range was set to 7.15 S.

[4/26/2012; 20:50:13]

Initial guesses for the loading concentration and molar mass were automatically set to 6.000 S and 0.7653 OD for species 1.

[4/26/2012; 20:50:16]

Full fit started...

[4/26/2012; 20:50:16]

Fit Completed (4 iterations).

Results of fit:

Parameter	**Best Value**	**Standard Error**	**Starting Value**
Co(1) (OD)	0.8045	+/- 0.0011	0.7653
s(1) (S)	6.064	+/- 0.001	6.000
M(1) (kDa)	159.41	+/- 0.71	150.03

Parameters held constant:

Parameter	**Fixed Value**
offset (OD/S)	0.00000

Fit statistics and convergence settings:

Final sum of squared residuals = 0.0002945381 (starting value = 0.04889639); rms = 0.001776434, for 89 data points and 86 degrees of freedom; runs test Z = 2.52.

Convergence limit = 0.1; parameter precision limit = 0.05%.

The r.m.s. residual corresponds approximately to 4.1E-3 OD.

Sedimentation coefficients are uncorrected (raw) values.

The diffusion coefficient corresponding to the best fit molecular mass is 3.50 F.

[4/26/2012; 20:50:47]

Pasted 'Species plot for entire data range (area)' graph into log.

The Log is a permanent record that is stored along with all the other fit information during the Save analysis procedure.

If your IT system provides user authentication (the user's identity is validated during a log-in procedure) then the analyst's name (as given by the authentication system) is also added into the log. The user name is added whenever a new analysis is started, or when a previously saved analysis is opened. Thus if user A does the initial analysis, and later user B re-opens the analysis file and re-analyzes the data in a different manner, the log will show who did what (and when).

Buttons and other controls on the Log page

The **Print log** button along the bottom of the log page allows the entire Log, or a selected portion of it, to be printed (to the printer you select from a dialog box).

The **Copy to clipboard** button will place the entire Log, or a selected portion of it, onto the Clipboard in rich text (RTF) format so that it can be pasted into a word processor or electronic notebook software.

- If you want to print or copy only a portion of the Log, first highlight (select) the desired portion by dragging over it with the mouse or using the keyboard movement keys while the SHIFT key is pressed, and then press either the Print or Copy buttons. Copying of the selected portion can also be done using the standard Windows CTRL-C shortcut key combination.

The **Magnification up/down spinner** allows you to increase or decrease the size of the text and images being displayed within the rich text box that displays the Log. This is sometimes useful because:

1. The rich text box sometimes fails to detect that the images are larger than its width and therefore it should display a scroll bar along the bottom.
2. The size of images imbedded in rich text is absolute, which can cause problems when the analysis is saved on one computer and then opened on a computer with a different display resolution.

The **Return** button simply hides the Log page and brings the current analysis page back to the foreground.

ding graphs to the Log

Typically the various graphs generated during intermediate steps of generating the *g(s*)* distribution, as well as the graphs that can be generated and displayed on the Graph fit results page or by the Graph bootstrap or Monte Carlo results dialog, are considered temporary results and are not placed into the Log.

However if for example one of those graphs illustrates why further refinement of a fit is needed, it may be appropriate to document that finding in the Log before proceeding with such refinements. Therefore buttons are provided on both the Graph fit results page and the Graph bootstrap or Monte Carlo results to add those graphs to the Log when it is desirable to permanently document them.

Loading saved fits lacking a Log

When a fit saved using an earlier program version which lacked the **Log page** feature is loaded a new Log will be created describing the old fit, listing the various parameters such as the fitting region, the names of the scan files, etc. An example of such a case is shown below.

Log page with graph

The file name and full path of the restored fit, and the program version that created the new Log, are also entered into the Log (these are not visible in the image above).

3.1.3.4 User prompt area

The **User prompt area** provides textual prompts for certain actions and displays a position read-out when the user is specifying positions on graphs.

Those prompts will automatically be cleared after being displayed for 5 seconds.

3.1.4 Navigation tree

The **Navigation Tree** is a heirarchical tree similar to that used in Windows Explorer and other programs. It provides a root node (branch) for each analysis window, two major sub-nodes for the major tasks, Derive *g(s*)* distribution branch and Fit distribution as N species branch, and minor sub-nodes for the individual steps within those tasks.

Any open analysis window can be selected as the active analysis, and any available step within that analysis can be selected, by clicking on the corresponding node within the tree. Steps which are not yet available are shown as gray (inactive). The topmost node in the tree provides access to the g(s*) overlay graph whenever more than one non-minimized analysis window has calculated a $g(s^*)$ distribution.

The Navigation Tree can optionally be closed to provide more room for analysis windows. Its visibility can be controlled using an entry in the View menu. When the width of the Navigation Tree panel is altered by clicking the mouse at the edge and dragging, that width will be remembered the next time the program is started.

Navigation tree context menu

Navigation tree context menu

Right-clicking on any branch of the tree with the mouse will bring up the context menu shown above. This allows you to:

1. save the analysis to disk
2. close this analysis window
3. close all analysis windows except this one
4. collapse all branches of the tree except the one that was clicked
5. expand any collapsed branches in all analyses
6. alter the font size used to display the tree to either make the tree easier to read or to make it more compact

3.1.4.1 Derive g(s*) distribution branch

This branch of the Navigation Tree contains five sub-branches corresponding to the 5 steps in calculating a $g(s^*)$ sedimentation coefficient distribution. Each of those steps corresponds to one page of the analysis window:

1. Load raw scans page
    - load scan files to analyze, or numerically simulate a velocity run using the Claverie simulator
2. Set menisus & data region page
    - set the meniscus position and the radial region for data to be used in calculating the $g(s^*)$ distribution
    - for interference scans when needed make adjustments to the removal of "jitter" and fringe jumps
3. Select scans to analyze page

   - use scroll bars to select the appropriate time window in the run to use for analysis, or just push the Auto Adjust button
   - optionally change sedimentation coefficient scale to *s(20,w)* units

4. Calculate average dc/dt curve page
   - display final average *dc/dt* curve
   - optionally adjust sedimentation coefficient range used for calculating *g(s*)* distribution

5. Calculate g(s*) distribution page
   - display graph of final *g(s*)* distribution
   - optionally integrate distribution to calculate peak areas and sedimentation coefficient averages
   - optionally normalize the distribution to facilitate comparisons of samples at different concentrations

3.1.4.2 Fit distribution as N species branch

This major branch of the Navigation Tree for each analysis window holds the sequence of steps needed to set up and complete the fitting of either the *g(s*)* or *dc/dt* data:

1. **Select fitting model page**
   - Select the fitting function, number of species, fitting of D or molecular mass, and weighted or unweighted fits. Also select the molecular mass units, and enter the partial specific volume and solvent density to give the correct molecular mass values. For multi-species fits you can optionally constrain M or s ratios among species

2. **Set initial guesses on graph page**
   - enter initial guesses for the sedimentation coefficients of each species from a graph of the data being fitted, and establish the range of sedimentation coefficients to be included in the fit

3. **Set/Alter fitting parameters page**
   - review and optionally alter which parameters are being fitted and the initial guesses for the fitting parameters; set criteria for fit convergence
   - this page is shown only in Advanced mode

4. **Perform least-squares fit page**
   - perform the fit, optionally 1 iteration at a time, and monitor the fit progress either through graphs or a display of the numerical values of the fitting parameters

5. **Report fit results page**
   - view a display of the best fit parameters, their uncertainties, the fit statistics, and optionally print a report

6. **Graph fit results page**
   - view various graphs of the fit results and data, and optionally print or export them

3.1.5 g(s*) overlay graph

The ***g(s*)* overlay graph** is used to compare distributions for different samples, or for the same sample analyzed at different times in the run. See Tutorial: Using the g(s*) overlay graph.

g(s*) overlay graph form

The **Overlay graph** is enabled whenever there are two or more analysis windows open where *g(s*)* distributions have been computed.

Any minimized analysis windows (shrunk to an icon) are ignored, however, which makes it possible to add or subtract curves from the graph without closing or opening analysis windows.

The **Graph All** button will update the graph with whatever distributions are currently available.

The **Print**button will bring up a Print dialog box to allow selection of the printer, orientation, and use of color.

The **Export** button will bring up a dialog that allows export of the graph image in several formats as a file or to the Clipboard. This export function can also export all of the data values.

The radio buttons within the **Graph Style** box allow easy choice of different plotting styles: points, lines, or both points and lines.

The **Edit Legend** text box and up/down spinner allow quick alteration of the description of each distribution to help create graphs for presentations or publication.

3.2 Analysis pages

3.2.1 Load raw scans page

This page allows you to load data from scan files, and is the starting point for any new analysis. Each control on the image below is a clickable *hotspot* that will lead to a separate page with further information, or just use the numbered list below the image.

Note that the **Next** button will not be active unless at least 4 scans have been loaded.

Load raw scans page

Controls on this page:

Preselect data type and cell check box

Load new scan files button

Replace scans for this sample button

Add more scans for this sample button

Delete selected scans button

Scans currently loaded in memory list box

Scans to be used in g(s*) calculation list box

Data loading options

Run Claverie simulation to generate data button

Sample description

3.2.1.1 Load raw scans - Preselect data type and cell check box

When this box is checked then the **Data type** radio buttons and **Cell number** up-down spinner controls within this group will control the data type and cell number of the scans that are initially displayed in the File Open dialogs.

Using this option makes it easier to find the data you are seeking among the hundreds of scan files that may be present. If you change your mind after you are already working within the File Open dialog you can still choose different data types or cells by making a different choice in the 'Files of type' drop-down list near the bottom of that dialog box.

Note that the 'Absorbance' choice also includes **radial intensity** scans, which can be converted to true absorbance or pseudo-absorbance (for when you are using two samples per cell). See About using radial intensity scans.

The cell selection spinner allows cell number 0 because SEDFIT creates cell 0 files when it saves simulated data. It also allows cell 8 for those who replace the counterbalance with an additional cell.

The preselect mode is on by default when the program is installed. You can turn off this default on the General Options tab of the Options dialog .

3.2.1.2 Load raw scans - Load new scan files button

This button is used to initially load scan files into a new analysis window, or to replace all the data that was previously loaded into that window. It will bring up a standard Windows file open dialog which allows you to select and load multiple scan files (you will want to load a minimum of 4).

Tip 1: Generally you will want to load scans covering nearly the entire range of boundary movement, *i.e.* all or nearly all the scans. If you know your latest scans go long after the cell is empty, there is no point in loading those, and you also may want to skip the first ~10% of the run where the meniscus region is barely cleared.

Tip 2: If you have hundreds of interference or fluorescence scans, at least initially to save time you may want to load only every second or third scan by using the Load every Nth scan option.

If the Preselect data type and cell option is on (checked) then the File Open dialog will initially show only those scan files that match that data type (absorbance, interference, or fluorescence) and cell number.

You can always change the data type and cell number that are being shown, or display all scan files, by picking appropriate choices from the 'Files of type' drop-down list near the bottom of the File Open dialog box.

The initial scan folder that will be displayed depends on settings within the File Locations tab of the Options dialog. You may choose either to always start in a particular folder (for example the folder where your instrument deposits new data), or to have the program remember the location of the last data that were loaded and to always start there.

3.2.1.3 Load raw scans - Replace scans for this sample button

This button is used to load a different set of scans for a particular sample, retaining sample-dependent settings such as the meniscus location. It is the retention of those settings that distinguishes this operation from using the Load New Scan Files button to replace the scans.

This button is only active after scans have been loaded using the Load New Scan Files button.

3.2.1.4 Load raw scans - Add more scans for this sample button

This button allows you to load additional scans while retaining sample-dependent settings such as the meniscus position.

The added scans can be from times either earlier or later in the run than those already loaded. The scan files are sorted based on the run times at the end of this operation, and any duplicate scans are removed.

This button is only active after scans have been loaded using the Load New Scan Files button.

3.2.1.5 Load raw scans - Delete selected scans button

This button will delete the scans whose names are currently highlighted in the Scans currently loaded in memory list box. (If nothing is highlighted a beep is emitted but nothing else happens.)

This command is useful if you need to remove an entire scan where the data is bad (for example an interference "*rogue scan*" that bears no resemblance to the others).

3.2.1.6 Load raw scans - Scans currently loaded in memory list box

This list box shows the scan file names for all scans currently in memory for this analysis window, including ones that might currently be excluded from the analysis.

Scans in memory list box

If you right-click on any scan name a pop-up context menu will appear with options to:

1. display the scan header information (comment, wavelength, run time, total number of data points, etc.)
2. edit the scan file using NotePad

If you choose to edit the scan file in NotePad the program will minimize itself and launch NotePad with that scan loaded. The program then waits for NotePad to close. When NotePad closes, the program detects whether the file was altered, and if so it automatically re-loads the altered data. For other data editing options see How to remove bad data points from scan files.

3.2.1.7 Load raw scans - Scans to be used in g(s*) calculation list box

This list shows the scan file names for the subset of scans currently designated to be used in calculating the *g (s*)* distribution. (These are the ones actually used if the distribution has already been calculated.)

Scans to be used list box

Scans which are loaded into memory but which are being excluded are shown as a "-" in the same position in the list as that scan appears to the left in the Scans currently loaded in memory list box. In the example shown scans 100-119 have been loaded but are being excluded.

If you right-click on any included scan name a pop-up context menu will appear with options to:

1. display the scan header information (comment, wavelength, run time, etc.)
2. edit the scan file using NotePad

3.2.1.8 Load raw scans - Data loading options

This group of controls specifies optional parameters that alter how scan files are loaded.

If the **Load every Nth scan** check box is activated (checked) then only every Nth scan among the scans selected within the File Open dialog will be loaded when new scans are read. This applies to the Load New Scan Files button, Replace Scans for this Sample button, and the Add More Scans for this Sample button. The value of **N** (default = 2) can be altered using either the up-down spinner or by directly typing a number within its text box.

Normally hundreds of scans can be loaded and processed if desired (limited only by available memory). Thus the use of this option is almost never really required, and for absorbance scans it is never recommended. This option is primarily intended to speed up data processing for situations where (1) a very large number of interference scans have been acquired and there is very little boundary movement between successive scans, and (2) the signal/noise of those scans is sufficiently high that averaging the *dc/dt* data from a large number of scan pairs is not required.

The **Despike options** are only visible in Advanced mode, and only active for absorbance scans. By default despiking is always on. The purpose of despiking is to remove the single-point spikes that sometimes occur in absorbance scans (often due to lamp miss-firing). These spikes are visible as single-point spikes within the dc/dt curve from one pair of scans (and a corresponding smaller spike in the average dc/dt curve), as displayed on the 'Select scans to analyze' page.

The despiking routine defines spikes as data points which differ from the preceding and following points by more than a threshold value, and removes those data points (from the data being processed, not in the original scan file). This **spike threshold** defaults to 0.2 OD, but can be altered using the corresponding text box. The default value works well in most situations, but may fail to detect some small spikes. The danger of setting the threshold too low is that false spikes may be detected (for example normal random noise in the data, or the top of the meniscus region). All of the original raw data is always retained in memory (and will be stored by the Save Analysis As... command), so you can alter this setting as desired and reprocess the data until you are satisfied with the results displayed on the Select scans to analyze page.

Note that the despiking routine ignores a small region around the meniscus so the meniscus spike is not affected.

If you want to see all the bad data in the graph on the Set menisus & data region page (for example if you want to edit the raw scans to eliminate bad points) then you will want to turn despiking off.

3.2.1.9 Load raw scans - Sample description

This text display area shows the sample description from the comment line of the last scan that was loaded (the information that was entered for that cell during data acquisition), or the description of the last Claverie simulation if simulated data are currently loaded.

3.2.1.10 Load raw scans - Run Claverie simulation to generate data button

A Claverie simulation uses finite-element methods to solve the Lamm equation. This approach can be give highly accurate simulations of velocity experiments, but is computationally intensive.

The purpose of providing this simulator is to provide a tool for "what if" explorations and tests of the accuracy of the fits. Some examples of questions that can be approached this way are:

1. Is the difference between the mass value returned by fitting g(s*) and an expected result really significant, or just due to inaccuracy of the method?
2. To what extent will adding more scans to your analysis cause broadening of the peaks and therefore inaccurate D and M values?
3. What would the data and residuals look like if my sample contained a small amount of a second species, but was analyzed as a single species?

This Claverie simulator (code kindly provided by David Cox and Walter Stafford) allows you to generate accurate data for these scenarios, which you can then use for fitting, comparing the fitting results to the known correct answers. It also allows for concentration-dependent sedimentation coefficients (something not dealt with by the fitting functions) and thus permits you to see how concentration dependence will affect the results. It does not, however, include concentration dependence of diffusion coefficients.

The dialog box used to control and run the simulations is shown below as it appears while a simulation run is in progress. This run corresponds to the default parameters, a quick (moderate accuracy) simulation of ovalbumin at 60K rpm.

Claverie simulation dialog

For instructions on how to use this simulator see Tutorial: Simulating experiments with the Claverie simulator.

3.2.2 Set meniscus & data region page

This page allows you to view a graph of the scan data, set the meniscus position, and set the radial limits defining the data to be used in generating *g(s*)* distributions. The radial positions or the meniscus and for the data range can be specified either directly in the text boxes at the top, or by clicking and dragging the vertical red or green markers on the graph itself.

When this page is first displayed with new data, the meniscus wizard automatically attempts to set the meniscus position, but manual adjustment is always needed for fluorescence scans, and may also be needed for interference or absorbance scans.

Each control on the image below, as well as the vertical position markers on the graph, are clickable *hotspot*s that will lead to further information, or just use the numbered list below the image.

The bottom row of controls is visible only in Advanced mode. These controls allow you to zoom in on the meniscus region to make a more accurate placement (although you can always zoom with the mouse), to display the analyzed data region as a 3-D waterfall graph, or to display whether the rotor temperature changed over the course of the run. For interference scans two groups of controls are active to allow fine-tuning the removal of *jitter* and *fringe jumps*. Lastly the ID points button displays the scan data as individual clickable data points for read-out and/or removal of bad points.

Set meniscus & data region page

Controls on this page:

Meniscus position (red marker on graph)

Good data begin X cm from meniscus (left green marker on graph)

Good data end at radius = (right green marker on graph)

The controls below are shown and available only in Advanced mode

Zoom meniscus button

Jitter removal options (gray markers on graph, interference data only)

Fringe jump removal options (interference data only)

Data displayed options

ID points button

SNotes about some features of the graph display

To keep the graph legend from occupying too large a fraction of the graph the legend is suppressed when more than 40 scans are displayed (as shown above). Also as shown above, when more than 40 scans are displayed the sequence of colors and line styles that is normally used by the graphing routines is suppressed (since it becomes impossible to distinguish individual scans when there are so many) and instead the colors are varied to produce a rainbow effect. Note that when more than 40 scans are shown you can force display of a legend by right-clicking on the graph and selecting a legend position from within the Legend Style submenu. The legend text size is reduced when there are 20 or more scans shown.

Note also that if you wish to export the scan data being displayed on this graph for use in another program, whenever this page is being displayed that can be done using the **Scan data from graph** items on the Edit menu.

3.2.2.1 Set meniscus & data region - Meniscus position

This setting defines the radial position of the meniscus. The value can be altered either by directly entering a value in the text box, or by dragging the red vertical line displayed on the 2D graph.

- For absorbance scans, if you manually drag the meniscus marker on the graph you will then be asked whether you want the wizard to re-calculate the mean position of the positive spike (the mean of the actual spike positions for every scan that has been loaded). This is helpful when the scans contain two positive spikes, and the wizard has identified the wrong one as the true meniscus.

Upon first entry to this analysis page after new scans have been loaded the meniscus wizard will try to locate the meniscus for you, and the *focus* will be moved to the meniscus entry text box. Thus to confirm the value set by the wizard all you need to do is push the **Enter** key on the keyboard.

- If these are interference scans using the **Enter** key will then also trigger the jitter wizard, an automatic fringe jump correction, and an update of the graph after these systematic errors are removed. If the wizard has not done a good job and rather than accepting its choice you immediately use the mouse to drag the meniscus marker to a new position, those automatic systematic noise corrections will be postponed until the end of that meniscus drag operation.

To obtain accurate and reproducible sedimentation coefficients you must set the meniscus position carefully and reproducibly. To do this you may need to zoom in on the region around the meniscus (see How to zoom with the mouse). In Advanced mode the Zoom meniscus button can do this for you.

3.2.2.2 Set meniscus & data region - Good data begin X cm from meniscus

This setting is used to exclude the region near the meniscus where the data are distorted by optical artifacts. As you adjust the meniscus position the left hand green marker moves along with it to maintain a constant separation.

This value can be altered either by directly entering a value in the text box, or by dragging the left-hand green vertical line displayed on the 2D graph.

Failure to exclude this region will cause significant errors in the low sedimentation coefficient region of the *dc/dt* data and *g(s*)* distribution. Due to the mathematics of the data transformations, the data points near zero sedimentation coefficient in the *g(s*)* distribution are often quite noisy anyway, so excluding those points may be desirable simply for that reason.

Obviously if you find that you often need to adjust this value then you may need to revise the default value via the Preferences... dialog.

3.2.2.3 Set meniscus & data region - Good data end at radius =

This setting determines the upper radius limit on raw scan data to be used in calculating *dc/dt* curves and *g(s*)* distributions. This setting is used to exclude the region near the cell base where solutes are accumulating. That region must be excluded because the theory treats the cell as semi-infinite. On the other hand, choosing too low a radius will cause loss of potentially valuable data.

The upper radial limit can be altered either by directly entering a number in the text box, or by dragging the right-hand green vertical marker displayed on the 2D graph to a new position.

For low molecular weight solutes back diffusion from the cell base may extend out to 7 cm or even further. In such cases you must cut out the data influenced by the cell base because the calculations do not account for the build-up of solutes there. Failure to remove this region will cause the *dc/dt* and *g(s*)* curves to go negative near the upper limit of sedimentation coefficients.

The default radial limit is controlled by a setting within the '*dc/dt* Options' tab of the Options dialog.

Example showing problems to avoid

Failure to exclude this region will cause significant errors in the high sedimentation coefficient region of the *dc/dt* data and *g(s*)*. As an example, the graph below shows the *dc/dt* curves for individual scan pairs from some data for a sample that unexpectedly contained some low mass fragments from proteolysis.

dc/dt graph with build-up at cell base

Since the low mass material has a much higher diffusion coefficient, it accumulates over a region at the cell base which extends to radii significantly below the default cutoff of 7.15 cm. The result is a negative spike or 'tail' at the high sedimentation coefficient end of the *dc/dt* curve from each scan pair.

An important (but often missed) point is that although these regions of bad *dc/dt* data represent only a small fraction of each individual curve, the different sedimentation times for each scan pair mean that these bad data occur at different sedimentation coefficients. Thus if these 4 data sets were averaged together, the entire region above ~12.3 S would be contaminated with bad data (about 1/6 of the total curve!).

Therefore, although it is true that one can manually remove the effects of the region at the cell base by excluding the right end of the sedimentation coefficient range from the analysis, that approach (which is the only option is most DCDT implementations) throws away a lot of good data. Further, that approach is dangerous because it may not be obvious from examination of the averaged data where the bad region begins. This example also illustrates why it is a good idea to display and examine the *dc/dt* data from each scan pair before averaging!

If you see *dc/dt* curves data that look like this example then move back to the Set meniscus and data region page, drag the right-hand green marker further to the left, and then use the **Next button** to move forward and display all the individual *dc/dt* curves to confirm that all the negative downward "hooks" at their right end are absent.

Why not simply set the default cutoff to a radius small enough to always avoid this problem? The problem with that approach is that in many experiments you would be throwing away good data, and this data may represent a significant fraction of the total potential range of sedimentation coefficients (particularly if the cell is not completely filled).

3.2.2.4 Set meniscus & data region - Jitter removal options

Jitter removal options are displayed only in Advanced mode and active only for interference data and pseudo-absorbance data.

Jitter options

The purpose of *jitter* removal is to correct for small fringe shifts (vertical displacements of small fractions of a fringe) that often occur from one interference scan to another. The correction is done by selecting a region (a range of radii) where the scans are expected to align vertically, and then forcing them to do so. This type of matching is also called "normalization" in the Beckman Coulter software, and "fine adjusting" in some versions of Walter Stafford's DCDT program.

A similar phenomenon also seems to arise when absorbance data are collected in pseudo-absorbance mode (two samples per cell). In that case the vertical shifts typically occur in a few discrete jumps per run.

The jitter matching region is generally placed either in the air-air region or a region just to the right of the meniscus. Using a region to the right of the meniscus generally works better. However, this will only work correctly if the meniscus is completely cleared and there is no sedimentation of low-molecular-weight components that will produce signals in this region.

By default when new interference data are loaded the jitter wizard tries to automatically determine the best matching region, marks that region on the graph using two vertical gray bars to mark the region limits, and then does the correction using that region. The **Auto button** will show as depressed (on) to indicate this automatic correction is active.

- The jitter wizard will be invoked as soon as you set a meniscus position, either by using the Enter key to confirm the program's automatic determination of the meniscus location, or by manually dragging the red meniscus marker.

If the *g(s*)* distribution is significantly negative (not just the expected negative points due to random noise) this is a sign that the jitter correction may be poor. Try moving the correction region to another location.

To manually select the match region:

Two gray vertical bars are displayed on the graph to mark the inner and outer radii for the match region. When the **Manual button** is pressed a prompt is given to drag those gray markers to a new position. Whenever the markers are dragged away from the position set by the jitter wizard, the Manual button will change to the depressed (on) state.

When you drag either gray bar both bars will move together, maintaining the same separation between them.

If you drag the gray markers **while holding the Shift key down** the two gray bars can be moved independently.

Once the Manual button is pushed **this operation must be completed** by dragging a gray marker; until it is completed other controls will be inactive.

Note that moving the right marker far to the right of the meniscus will affect the lowest sedimentation coefficient that is included in the *dc/dt* curves and *g(s*)* distribution. Only data points which are to the right of both the leftmost green marker and the rightmost gray marker will be used for data analysis. (Since the data in the match region are being forced to be zero, it makes no sense to include them in the analysis.)

To turn off jitter correction:

Jitter correction is off whenever both the **Auto button** and the **Manual button** are up (off). Simply press whichever button is down to turn jitter correction off.

3.2.2.5 Set meniscus & data region - Fringe jump removal options

Fringe jump removal options are displayed only in Advanced mode and active only for interference data.

Fringe jump removal

This group of controls selects the method used to compensate for any *fringe jumps* that occur within a group of interference scans that are being analyzed. The methods differ only in the region of the scan that is used in comparing successive scans for a possible fringe shift.

Four 'automatic' methods are selectable via a drop-down list box. Those are:

1. **Entire scan** compares the scans across the entire sample area to the right of the meniscus. This will work well as long as the amount of sample movement between scans is not too large (moderate concentrations and sedimentation rates), which is generally true for scans appropriate for the dc/dt method.
2. **Near meniscus** confines the comparison to the part of the sample near the meniscus (60 data points). This will work well if all the scans are taken after the meniscus region has cleared.
3. **Middle half** compares the scans in the region from 1/4 to 3/4 of the sample column. This is the default method when the program is initially installed.
4. **Outer eighth** makes the comparison only for the high radius portion. This will work well if all the scans have a well-defined plateau region, provided the scans do not include too much data from the region where sample is accumulating at the bottom of the cell. The outer end of the range used for this comparison is limited by the 'good data end at radius' setting.

A default automatic method will automatically be used whenever this analysis window is first entered after new scans have been loaded.

Whenever an automatic method has been used the **Auto button** within this group will be depressed. If a different method is selected via the drop-down list box while the Auto button is depressed, the new method will be calculated and applied to the data displayed on the graph.

Manual fringe jump removal

The **Manual button** initiates a procedure where the position used for fringe jump compensation is specified by double-clicking on the graph at the desired radial position. The actual region used in the calculation is then a group of 11 points centered around that position. This manual setting will change the 'Auto' button to the off (raised) position.

Once the Manual button is pushed **this operation must be completed** by double-clicking on the graph; until it is complete other controls will be inactive.

All these methods are implemented in the same manner as the corresponding methods in Walter Stafford's original DCDT program. The default method can be altered via the '*dc/dt* Options' tab within the Options dialog.

3.2.2.6 Set meniscus & data region - Data displayed options

This group of three buttons controls options for the type of graph and what data is displayed.

- The **Corrected button** is only active for interference data, and only after the meniscus has been set. This options displays the scan data after the *jitter* and *fringe jumps* have been removed. This button acts as a toggle switch, with the uncorrected 'raw' scans shown when it is off (up).
- The **3D button** will bring up a 3-dimensional "waterfall" graph of the scans. In this graph only the portions of the scans within the analysis region (between the green vertical bars on the 2D graph) is shown. The viewing angle for the 3D graph can be rotated using the scroll bars on the right and

bottom of the graph; the axes are suppressed during the rotation operation.

  - This 3D display really has no function for data analysis, it is simply an alternate method of displaying the data and the motion of the boundaries. The positions of the meniscus and other region marker are not displayed on this 3D graph.

- The **2D button** displays the normal 2-dimensional graph.
- The **T button** displays a graph of the recorded rotor temperature versus elapsed run time (useful to check whether the temperature was truly stable during this run). The range of temperature data displayed corresponds to all the scans in memory, not just those currently being used in the analysis. See also How (and why) to equilibrate the rotor temperature.

3.2.2.7 Set meniscus & data region - Zoom meniscus button

This button will zoom the 2D graph to show only a radial region 2-3 mm wide centered around the current setting of the meniscus position. This is intended to help position the meniscus more accurately.

This command does not alter the Y axis scales and thus it may still be necessary to manually zoom with the mouse to increase the vertical magnification.

This is a toggle button and will remain down to alert you that the graph is currently in zoomed meniscus mode. You can unzoom either by toggling the button off or by hitting the Z key while the graph has the *focus*.

3.2.2.8 Set meniscus & data region - ID points button

This button will cause the 2D graph of the raw data to be displayed as individual data points rather than continuous lines, and each data point becomes a clickable object. When a data point is clicked the corresponding scan file name and *x,y* data values are displayed within a box above the graph area.

  - Note that for graphs containing large numbers of scans, especially interference scans, there can be a significant delay (10 sec or more) between clicking the button and the display of the graph, and also between clicking a data point and the display of its identity. This is simply because there are tens to hundreds of thousands of clickable regions to be created and then identified.

The main purpose of this display mode is to help identify *rogue scan*s (see How to identify and remove rogue scans) or individual bad data points.

When the same data point is clicked two times in succession (separate single clicks, not a double-click) the program assumes this may be a bad data point that you want to remove. You will be asked a series of questions about whether you want to remove that point only temporarily (from memory), or to permanently alter the scan file (either reversibly or irreversibly). See How to remove bad data points from scan files for more details.

3.2.3 Select scans to analyze page

Overview

This page displays a graph of the *dc/dt* curves calculated from each scan (as dashed lines), as well as the average of all those individual *dc/dt* curves (as a thicker black line).

Analysis page 3

Two slider controls at the top of the page allow you to vary which scans are being used, among all of those currently loaded in memory, to optimize your analysis. As you make adjustments the *dc/dt* curves are immediately re-calculated and displayed.

The top slider varies the total number of scans being used (always an even number).

- Generally you will only want to use a small portion of the total run (perhaps 10-20%).
- As you make adjustments the text boxes to the left of this slider indicate the number being used and the total available

The lower slider will move that group of scans earlier or later in the run.

- Generally you will want to analyze a group of scans where the boundary is about in the middle of the cell.
- As you make adjustments the text boxes below this slider show the file names of the first and last scans

Optimally selecting which scans to analyze takes some judgement and experience; see Tutorial: How to select the scans to analyze for all the details and trade-offs.

However now there is any easy way:

Simply use the **Auto adjust button** and a program wizard will make these adjustments for you. This wizard will first try to find the time in the run when the major species are about in the middle of the cell, and then adjust the number of scans used upward until the peaks start to broaden significantly.

- This automatic choice often produces acceptable results, but cannot be guaranteed to make the best choices. Further, you may be particularly interested in species within a certain range of sedimentation coefficients, but the wizard doesn't know that and might not choose the portion of the run that is optimal for those species. So ultimately your intelligence is likely to be needed to truly obtain the optimum settings.

Note that for interference data when the **Auto jitter removal** option has been selected, each time the group of scans to be analyzed is altered using the sliders or the **Auto adjust button** the program wizards re-evaluate the best position for jitter removal and recompute the fringe shift corrections prior to re-computing the *dc/dt* curves. Thus you may see a change in the jitter removal region if you return to the graph on the Set meniscus & data region page.

The **peak broadening limit**is an important guide to whether the number of scans being used (slider 1) is too high. Species with molecular masses higher than this limit will have their peaks significantly broadened. Generally you will want to have this limit roughly equal to the highest mass present in your sample. See Peak broadening limit for more details.

The **Convert to s(20,w) button** allows you to toggle the sedimentation coefficient scale between "raw" *versus* standardized *s(20,w)* values.

The **Use true mean time button** controls the assignment of the run time to each scan pair when calculating the *dc/dt*curves. When this button is on (down) the algorithms use the true mean of the times for those two scans. When this button is off (up, the default) the algorithms use the harmonic mean, as in the original Stafford implementation. The true mean can give better results for multi-species fits using the new "improved method" when the peaks are severely broadened, but should **not** be used in other circumstances. See Use true mean time button for further details.

The **# of sections up/down spinner** can be used to divide the scan pairing scheme into multiple sections for the purpose of reducing peak broadening (at a penalty of lower signal/noise ratio for the average *dc/dt* curve and the *g(s*)* distribution. This option should only be used if the intent is to fit the data, not if the *g(s*)* distribution itself, or averages over the distribution, is the intended output of the analysis.

How are scans paired when the *dc/dt* curves are calculated?

By default the scans are split into two groups (first half and second half) and pairs are formed from the corresponding members of each group. Thus if there are 16 scans, scan 9 is subtracted from scan 1, 10 from 2, 11 from 3, ..., and 16 from 8. If the number of sections was instead set to 2, then the pairing is (1,5), (2,6), (3,7), (4,8), (9,13), (10,14), (11,15), (12,16). When the number of scan pairs cannot be evenly divided by the number of sections then the last section will have more scan pairs than the other section (s).

3.2.3.1 Select scans to analyze - Auto adjust button

This button invokes a program wizard to automatically select the group of scans to be analyzed. The wizard first calculated a time in the run when the main boundary is approximately in the center of the cell, and selects 4 scans (the minimum number) centered around that time.

The wizard then starts increasing the number of scans, monitoring the degree to which the *dc/dt* curves from different scan pairs deviate from one another. During this adjustment you will see repeated re-calculations of the *dc/dt* curves and re-displays of the graph. The wizard keeps increasing the number of scans until it is able to determine that the different scan pairs are deviating significantly (an indication that the larger time span from the first to last scan is causing peak broadening), and then selects a number of scans that should be a reasonable compromise between the conflicting goals of good signal/noise and low peak broadening.

Remember, this automatic adjustment will not necessarily produce the true optimum settings, but they should be at least a good starting point for your further adjustments. Remember too that the wizard is not as smart as you are and can sometimes be fooled by noisy data.

3.2.3.2 Select scans to analyze - Peak broadening limit

This displayed molecular mass value provides a means to test whether the time span between the first and last scans currently included in the analysis is too large to give an accurate approximation of *dc/dt* for the species you are studying.

If too many scans are being used the peaks in *dc/dt* and *g(s*)* will be broadened, resulting in poor resolution. This will also result in incorrect estimates for *D* and *M* if such peaks are fitted using the "classic" approach of fitting to Gaussian functions (but does not cause such errors when using the new "improved method").

The "rule of thumb"

Whether or not the time span is "too large" depends on the molecular mass (*s/D* ratio) of the species that are sedimenting, with higher mass species requiring shorter time spans and therefore fewer scans. This limit is calculated using a "rule of thumb" test to determine the approximate highest molecular mass that can be present without experiencing significant broadening, *M*(max).

The "rule of thumb" that is used was developed by Walter Stafford [reference 5]. It states that the time difference, Dt, between the first and the last scan used should not exceed the following value:

where *t* is the average of the times for first and last scan, *M* is the molecular mass in kilodaltons and RPM is the rotor speed. This formula assumes a solvent density of 1 and a partial specific volume of 0.725 ml/g.

The elapsed times used in this calculation are actually derived from the $\omega^2 t$ values stored with each scan and the rotor speed, rather than the elapsed time, since the $\omega^2 t$ values reflect the true effective sedimentation time.

This rule can also be re-arranged to calculate for any group of scans the highest mass that will not be significantly broadened, M_{max}, given by

Please note that just because a certain species has a mass <= M_{max} this does not mean that its peak will have no broadening. This 'rule of thumb' was originally developed by Walter Stafford for use in qualitative analysis to assess when the peaks would be visually broadened.

Just because a certain species exceeds M_{max} does not mean that the *dc/dt* or *g(s*)* data are invalid. Indeed, in some circumstances it may be appropriate to strongly violate this rule of thumb. For example, if you are unsure whether a small peak in the data truly represents an additional species, you may wish to add many more scans in the analysis to improve the signal/noise and see whether this peak is still present. In such a test the peak for a real species would be broadened, but it should not disappear, so the qualitative result (peak or no peak) is enhanced even though this rule of thumb would be grossly violated.

For a more detailed analysis of how the accuracy of fitted parameters derived by the "classic" fitting methods is affected by the time span see Fig. 6 in reference 8 and the discussion therein. Briefly, simulations indicate that for M ~= M_{max} the errors in *D*, relative to the value obtained with no peak broadening, are 14-17%, and those in *M* are 12-15%. For smaller time spans these errors are reduced approximately linearly with the ratio of M/M_{max} (or equivalently, approximately with the square of the number of scans).

3.2.3.3 Convert to s(20,w)

This option will convert sedimentation coefficients for *dc/dt* and *g(s*)* curves, as well as fitted values, to the $s_{20,w}$ scale.

Note that $s_{20,w}$ will also sometimes be written in this document as *s(20,w)* to avoid problems in generating subscripts

The **Convert to $s_{20,w}$ buttons** are present on the Set meniscus & data region page, the Calculate average dc/dt curve page, and the Calculate g(s*) distribution page. When the data are in "raw" sedimentation coefficient unit pressing this button will bring up a dialog box (shown below) for entry of the data needed to perform the conversion (discussed in more detail below).

If this conversion has already been done (as indicated by the button remaining in the down state) then pressing the button will turn off the conversion and revert to raw *s* values.

Conversion to s(20,w) dialog

It is generally desirable to present and report $s_{20,w}$ values rather than raw (uncorrected) sedimentation coefficients. Conversion from raw *s* to $s_{20,w}$ compensates for the factors such as buffer viscosity which are relevant only to a particular experiment, giving the sedimentation coefficient that molecule would have if measured in water at 20 °C. Because of the solvent-independent nature of $s_{20,w}$, it describes quantitatively the fundamental hydrodynamic properties of the molecule and it is this value which is most useful in comparing the sedimentation behavior of different molecules.

The partial specific volume that is entered on this form and used in converting to $s_{20,w}$ scale is applied uniformly to the entire *g(s*)* distribution. Therefore if you believe your sample contains species with different vbar values (for example different polypeptides, or both PEGylated and non-PEGylated forms of the same protein) you should not apply this conversion---you must work in raw sedimentation coefficient units.

- For similar reasons, the partial specific volume entered on this conversion form is applied to all species that are fitted.

The actual scaling factors that will be applied to correct the raw sedimentation and diffusion coefficients to standard conditions are displayed in the panel nearest the bottom of the dialog box.

Show All Show All

Hide All Hide All

Show What about extrapolation to zero concentration?

This program does not attempt to correct for concentration dependence of sedimentation coefficients (extrapolation to infinite dilution) to give $s^0_{20,w}$. Often *dc/dt* analysis is applied to samples at concentrations sufficiently low that concentration dependence can be neglected, and if so the corrected results can be considered to be $s^0_{20,w}$ values, but the program will not explicitly designate them as such.

Show How this conversion affects the *g(s*)* distribution and average *dc/dt* curve

Please be aware that conversion to $s_{20,w}$ also alters the magnitudes of the *g(s*)* and *g^(s*)* distributions. This arises because these distributions are defined such that the area under the curve gives the concentration. Thus if the X-axis is expanded or contracted as a consequence of conversion to the *s20,w* scale, the Y-axis must be contracted or expanded to keep the area unchanged.

Internally in the program the *dc/dt*, *g(s*)*, and *g^(s*)* data are always kept in raw units---the conversion is only applied as graphs are generated and values are displayed and printed. Generally this is transparent to the user, but with regard to setting the δ*s* intervals for the average *dc/dt* and *g(s*)* curves it is not possible to have conveniently rounded intervals between data points in both raw and $s_{20,w}$ scales. Consequently switching between raw and $s_{20,w}$ units will usually force a re-calculation of the average *dc/dt* curve and *g(s*)* distribution because the X-axis point spacing will change if that is being determined by the default **Automatic option** (see Calculate average dc/dt - X-axis point spacing). Thus the average *dc/dt* curve and *g(s*)* distribution will be slightly different in the different units (the difference is not simply a multiplicative factor for the X-axis). However those differences will be negligible compared to the noise in the data.

Show

How this conversion affects results from fitting

When this conversion is used, the sedimentation coefficients obtained by fitting will also be $s_{20,w}$ values, and the diffusion coefficients from fitting will be $D_{20,w}$ values. They will be marked as such on all program outputs, and the values used to make the conversion will be printed along with the printed reports of fit results.

Switching between raw and $s_{20,w}$ units will usually force a re-calculation of the average *dc/dt* curve and *g(s*)* distribution because the X-axis point spacing will change if that is being determined by the default Automatic option (see Calculate average dc/dt - X-axis point spacing). Thus making this switch will invalidate any previous fitting results.

Show

How this conversion affects reported diffusion coefficients

Activating this option will also cause raw diffusion coefficients to be converted to $D_{20,w}$ values, via the formula

where *T* is the sample temperature (Kelvin).

Show

Data required for conversion

Conversion to $s_{20,w}$ requires knowledge of the solvent density at the experiment temperature, the solvent viscosity, h, at the experiment temperature, the solute partial specific volume (vbar) at the experiment temperature, and the solute partial specific volume at 20 °C, through the relation:

where the subscripts refer to the temperature and solvent conditions (b=buffer, w=water). The density and viscosity of water at 20 °C are 0.998234 g/ml and 1.002 centipoise, respectively.

If all those data are not available, or cannot be estimated (*e.g.* by SEDNTERP), then the conversion cannot be made.

The program will optionally automatically compute a temperature correction to the partial specific volume, using the same formula used in SEDNTERP [reference 4]:

This option is controlled by the check box on the lower frame, and when it is checked the input box for the vbar at 20 °C is disabled (but the computed value is displayed).

The viscosity values are to be entered in centipoise. Since SEDNTERP displays these values in poise, to facilitate copying and pasting numbers from SEDNTERP, whenever a number is entered that is less than 0.1 the program will assume it is a value in poise and multiply it by 100.

The program automatically fills in the experiment temperature based on the temperature recorded in the header of the last scan used in calculating the *dc/dt* data, but since the temperature may drift during the run you may wish to adjust that value manually.

Please note that changing the temperature on this form overrides the temperature recorded with the scans, and also alters the value used elsewhere in the program (*e.g.* for conversions to buoyant or true mass units).

3.2.3.4 Select scans to analyze - Use true mean time button

This button is only visible in Advanced mode and this option is **not recommended for general use**.

Briefly, this option is sometimes helpful when you will be doing multi-species fits to *g(s*)* distributions and for data where the peaks have been severely broadened and distorted because the range of boundary movement is very large (for example a large range may be needed because the signal/noise of the data is very low).

What does it do?

This button toggles between two alternative algorithms for calculating the *dc/dt* curves resulting from each pair of scans that is included in the analysis.

When a pair of scans is subtracted the result is a table of Δc *versus* radius. Those radius values can be converted to apparent sedimentation coefficients, *s**, using the relation

where r_m is the radius of the meniscus. However since we are using a pair of scans, which value for $\omega^2 t$ should be used?

In the original algorithm developed by Walt Stafford (reference 7), and in versions 1.x of this program, the $\omega^2 t$ value used in the formula above is the harmonic mean of the $\omega^2 t$ values for the two scans:

geometric mean

When the **Use true mean time** option is on, this program instead uses the true mean (the arithmetic mean) of the $\omega^2 t$ values for the two scans in the pair.

Why is the true mean option sometimes useful?

The use of the harmonic mean can significantly distort the shapes of the average *dc/dt* and *g(s*)* curves when the range of boundary movement is quite large (when the peaks are severely broadened). In particular, the use of the harmonic mean results in *g(s*)* distributions that do not fully return to zero on the right-hand side of a peak (falsely making it appear there is a small amount of faster-sedimenting components present).

With the "improved" fitting algorithm the fitted hydrodynamic parameters will be accurate no matter whether the harmonic mean or true mean is used, but under conditions of severe peak broadening the peaks will usually be less broadened and distorted when the true mean is used. That reduction of broadening can be helpful for multi-species fits.

When should this option not be used?

This option is only helpful for multi-species fitting using the "improved" fitting algorithm. You should definitely not use it if your primary interest is in the *g(s*)* distribution itself or calculating properties such as the weight-average sedimentation coefficient from it. It's use for those purposes would result in sedimentation coefficients that are less accurate.

3.2.3.5 # of sections

# of sections

This control is only visible in Advanced mode and this option is **definitely not recommended for general use.**

Briefly, this option is sometimes helpful when you will be fits to *g(s*)* distributions for samples containing more than one species to help avoid severe broadening of the peaks because the range of boundary movement is very large (for example a large range might be needed because the signal/noise of the data is very low). The reasons why this option was developed are discussed, and some examples of its use can be found, in reference 13 (a PDF of this article can be obtained from John Philo).

What does this do?

By default the scans are split into two groups (first half and second half) and pairs are formed from the corresponding members of each group. Thus if there are 16 scans, scan 9 is subtracted from scan 1, 10 from 2, 11 from 3, ..., and 16 from 8.

If the number of sections is instead set to 2, then the pairing is (1,5), (2,6), (3,7), (4,8), (9,13), (10,14), (11,15), (12,16). This reduces the time span between the two scans in a pair by a factor of 2, and thus reduces the peak broadening by approximately a factor of 4.

When the number of scan pairs cannot be evenly divided by the number of sections then the last section will have more scan pairs than the other section(s). This is done because peak broadening is less of an issue for the later scans.

Why is using multiple sections sometimes useful?

Peak broadening can significantly degrade the resolution between species, as well as distorting the shapes of the average dc/dt and *g(s*)* curves. This occurs when the range of boundary movement is quite large. With the "improved" fitting algorithm the fitted hydrodynamic parameters will be accurate no despite the broadening, but the broadening can still make it difficult to accurately measure the properties of poorly-resolved peaks. The peak broadening is also problematic when you are trying to 'focus' your fit on a single peak in order to remove the influence of other minor components (aggregates, fragments, impurities, *etc*.).

When should this option not be used?

You should definitely not use multiple sections if your primary interest is in the *g(s*)* distribution itself, or calculating properties such as the weight-average sedimentation coefficient from it. It's use for those purposes would result in incorrect proportions for multiple components and sedimentation coefficients that are less accurate.

3.2.4 Calculate average dc/dt curve page

This page shows the average *dc/dt* curve corresponding to the scans selected via the Select scans to analyze page, and provides options such as converting the sedimentation coefficients to $s_{20,w}$ values or limiting the range of data to be used in calculating the *g(s*)* distribution.

Each control on the image below is a clickable *hotspot* that will lead to further information, or just use the numbered list below the image.

Analysis page 4

To calculate the average the *dc/dt* curve from each scan pair is first interpolated onto the uniform grid of X-axis values. Those interpolated curves are then averaged and the standard deviation for each point of the average is calculated. The range of sedimentation coefficients used for the average corresponds to the range of values that are present in at least 2 of the individual *dc/dt* curves.

Controls on this page:

Set range for g(s*) button - Push this button to manually set the sedimentation coefficient range for the *g(s*)* calculation by dragging position markers on the graph.

Range limit text boxes - these text boxes display, and allow you to manually alter, the sedimentation coefficient range for the *g(s*)* calculation.

Convert to s(20,w) button - This button toggles on or off the conversion of the X axis to $s_{20,w}$ units. When toggled on, it brings up a dialog box for entry of the property values needed to make this conversion.

The controls below are shown and available only in Advanced mode

Show error bars button - This button toggles on or off the display of error bars for each data point on the graph. The error bars are shown as vertical gray line segments extending from each point; they are not visible in the graphic above because for these data they are smaller than the data points.

X-axis point spacing options - These control and display the sedimentation coefficient interval between adjacent data points

Reset zero and Undo buttons - The Reset zero button allows manual adjustments of the zero level. Use with caution! Undo will remove the manual adjustment.

3.2.4.1 Calculate average dc/dt - X-axis point spacing

X-axis point spacing options

This group of controls is used to set δs, the increment of sedimentation coefficient between data points for both the average *dc/dt* curve and the *g(s*)* curve.

The default **Automatic** option (which is on when the button appears depressed) invokes a program wizard to set this value automatically. The wizard will choose a value for δs that gives a total number of data points along the sedimentation coefficient axis between 1 and 2 times the number of data points in the corresponding region of the raw scans. These automatic values are rounded to give intervals in a 1, 2, 2.5, 3, 4, 5, 10... sequence.

Alternatively any point spacing value can be entered manually into the text box to the left of the Automatic button, which will toggle the Automatic button off.

If the Convert to s(20,w) option is on, δs will be in *s(20,w)* units. If you manually set δs to a rounded value (.02, .05, ...) and then turn the conversion on or off, the interval will no longer be a rounded value. When the Automatic option is on, as the the conversion to *s(20,w)* units is turned on or off the program will recalculate the interval to keep it as a rounded number.

If the δs value is altered after the average *dc/dt* curve has already been calculated, that calculation will be re-done, and any previously calculated *g(s*)* curves or fitting results will be cleared.

Setting δs too large will result in poor resolution; setting it too small will yield too many data points. The consequence of "too many" points is that there will be far more data points along the X-axis than were present in the original scans, which requires the use of interpolation to create 'new' data points in between the experimental ones. Thus the data points in the average *dc/dt* and *g(s*)* curves will not be truly statistically independent (successive data points will be highly correlated). If these curves were subsequently used in fitting, then the statistical estimates of the uncertainties in the parameters would be too small because those statistical calculations assume all data points are independent.

3.2.4.2 Calculate average dc/dt - Set range for g(s*)

Fairly often you may want to exclude some data at low or high sedimentation coefficients from the *g(s*)* calculation.

- The data at the smallest *s* values may be noisy or unreliable because of the presence of low molecular weight contaminants or sedimentation of salts or other buffer components.
- The data at the highest *s* values may need to be excluded if that region is being influenced by the buildup of solutes at the base of the cell. However it is preferable to eliminate that problem by altering the Good data end at radius value from the Set meniscus & data region page.

By default the initial range for calculating *g(s*)* is the full range of the average *dc/dt* curve. If you want to reduce that range you can either:

1. manually enter new values into the text boxes, or
2. push the **Set range for g(s*) button** to display two magenta vertical lines on the graph marking the range. Drag those lines with the mouse to alter the limits.

Whenever the current upper or lower limit is at the full limit of the average *dc/dt* data the background color of the corresponding text box becomes a light yellow as a visual indication of that state.

3.2.4.3 Convert to s(20,w)

This option will convert sedimentation coefficients for *dc/dt* and *g(s*)* curves, as well as fitted values, to the $s_{20,w}$ scale.

Note that $s_{20,w}$ will also sometimes be written in this document as *s(20,w)* to avoid problems in generating subscripts

The **Convert to** $\boldsymbol{s_{20,w}}$ **buttons** are present on the Set meniscus & data region page, the Calculate average dc/dt curve page, and the Calculate g(s*) distribution page. When the data are in "raw" sedimentation coefficient unit pressing this button will bring up a dialog box (shown below) for entry of the data needed to perform the conversion (discussed in more detail below).

If this conversion has already been done (as indicated by the button remaining in the down state) then pressing the button will turn off the conversion and revert to raw *s* values.

Conversion to s(20,w) dialog

It is generally desirable to present and report $s_{20,w}$ values rather than raw (uncorrected) sedimentation coefficients. Conversion from raw *s* to $s_{20,w}$ compensates for the factors such as buffer viscosity which are relevant only to a particular experiment, giving the sedimentation coefficient that molecule would have if measured in water at 20 °C. Because of the solvent-independent nature of $s_{20,w}$, it describes quantitatively the fundamental hydrodynamic properties of the molecule and it is this value which is most useful in comparing the sedimentation behavior of different molecules.

The partial specific volume that is entered on this form and used in converting to $s_{20,w}$ scale is applied uniformly to the entire *g(s*)* distribution. Therefore if you believe your sample contains species with different vbar values (for example different polypeptides, or both PEGylated and non-PEGylated forms of the same protein) you should not apply this conversion---you must work in raw sedimentation coefficient units.

- For similar reasons, the partial specific volume entered on this conversion form is applied to all species that are fitted.

The actual scaling factors that will be applied to correct the raw sedimentation and diffusion coefficients to standard conditions are displayed in the panel nearest the bottom of the dialog box.

Show All Show All

Hide All Hide All

Show

What about extrapolation to zero concentration?

This program does not attempt to correct for concentration dependence of sedimentation coefficients (extrapolation to infinite dilution) to give $s^0_{20,w}$. Often *dc/dt* analysis is applied to samples at concentrations sufficiently low that concentration dependence can be neglected, and if so the corrected results can be considered to be $s^0_{20,w}$ values, but the program will not explicitly designate them as such.

Show

How this conversion affects the *g(s*)* distribution and average *dc/dt* curve

Please be aware that conversion to $s_{20,w}$ also alters the magnitudes of the *g(s*)* and *g^(s*)* distributions. This arises because these distributions are defined such that the area under the curve gives the concentration. Thus if the X-axis is expanded or contracted as a consequence of conversion to the *s20,w* scale, the Y-axis must be contracted or expanded to keep the area unchanged.

Internally in the program the *dc/dt*, *g(s*)*, and *g^(s*)* data are always kept in raw units---the conversion is only applied as graphs are generated and values are displayed and printed. Generally this is transparent to the user, but with regard to setting the δs intervals for the average *dc/dt* and *g(s*)* curves it is not possible to have conveniently rounded intervals between data points in both raw and $s_{20,w}$ scales. Consequently switching between raw and $s_{20,w}$ units will usually force a re-calculation of the average *dc/dt* curve and *g(s*)* distribution because the X-axis point spacing will change if that is being determined by the default **Automatic option** (see Calculate average dc/dt - X-axis point spacing). Thus the average *dc/dt* curve and *g(s*)* distribution will be slightly different in the different units (the difference is not simply a multiplicative factor for the X-axis). However those differences will be negligible compared to the noise in the data.

Show

How this conversion affects results from fitting

When this conversion is used, the sedimentation coefficients obtained by fitting will also be $s_{20,w}$ values, and the diffusion coefficients from fitting will be $D_{20,w}$ values. They will be marked as such on all program outputs, and the values used to make the conversion will be printed along with the printed reports of fit results.

Switching between raw and $s_{20,w}$ units will usually force a re-calculation of the average *dc/dt* curve and *g(s*)* distribution because the X-axis point spacing will change if that is being determined by the default Automatic option (see Calculate average dc/dt - X-axis point spacing). Thus making this switch will invalidate any previous fitting results.

Show

How this conversion affects reported diffusion coefficients

Activating this option will also cause raw diffusion coefficients to be converted to $D_{20,w}$ values, via the formula where T is the sample temperature (Kelvin).

Show Data required for conversion

Conversion to $s_{20,w}$ requires knowledge of the solvent density at the experiment temperature, the solvent viscosity, h, at the experiment temperature, the solute partial specific volume (vbar) at the experiment temperature, and the solute partial specific volume at 20 °C, through the relation:

where the subscripts refer to the temperature and solvent conditions (b=buffer, w=water). The density and viscosity of water at 20 °C are 0.998234 g/ml and 1.002 centipoise, respectively.

If all those data are not available, or cannot be estimated (*e.g.* by SEDNTERP), then the conversion cannot be made.

The program will optionally automatically compute a temperature correction to the partial specific volume, using the same formula used in SEDNTERP [reference 4]:

This option is controlled by the check box on the lower frame, and when it is checked the input box for the vbar at 20 °C is disabled (but the computed value is displayed).

The viscosity values are to be entered in centipoise. Since SEDNTERP displays these values in poise, to facilitate copying and pasting numbers from SEDNTERP, whenever a number is entered that is less than 0.1 the program will assume it is a value in poise and multiply it by 100.

The program automatically fills in the experiment temperature based on the temperature recorded in the header of the last scan used in calculating the *dc/dt* data, but since the temperature may drift during the run you may wish to adjust that value manually.

Please note that changing the temperature on this form overrides the temperature recorded with the scans, and also alters the value used elsewhere in the program (*e.g.* for conversions to buoyant or true mass units).

3.2.4.4 Calculate average dc/dt - Reset zero & Undo buttons

Reset zero & Undo buttons

The **Reset zero button** is used to manually redefine the zero level for the average *dc*/dt curve. This may be necessary to compensate for uncorrected "jitter" in interference data and/or other effects that are causing shifts of the zero level. If uncorrected, such shifts may cause the *g(s*)* curve to go below zero or remain above zero in regions away from the peaks.

Since this manual zero adjustment could bias the results, its use on a routine basis is **not** recommended.

When the button is pressed a red horizontal bar used to designate the new zero level is displayed on the graph of the average *dc/dt* curve. The bar can be dragged up and down using the mouse. At the end of the drag operation the new zero position is used to recalculate and re-display the average *dc/dt* curve.

The **Undo button** will restore the zero level to the original, experimental value.

When one or more peaks are fitted to *dc/dt* or *g(s*)* data which have been manually re-zeroed, the amount of the zero shift will be printed in the report of the fitting results.

3.2.5 Calculate g(s*) distribution page

This page shows the *g(s*)* distribution, and provides options to instead display the *g^(s*)* distribution or the integral distribution, *G(s*)*. It also provides commands to integrate specific regions or the entire distribution and display the resulting average sedimentation coefficients.

Each control on the image below is a clickable *hotspot* that will lead to further information, or just use the numbered list below the image.

Calculate g(s*) distribution page

Below are brief descriptions of the controls on this page; more details can be found by following the links.

Show integrals over distribution button - Push this button to calculate and display the sedimentation coefficient averages (s_w, s_z, and s_{z+1}) and total concentration for the entire range of the *g(s*)* distribution.

Set integration region button - This button calculate and displays the sedimentation coefficient averages (s_w, s_z, and s_{z+1}) and concentration for a region of the *g(s*)* distribution (for example one peak). That region is set by dragging with the mouse.

Normalize button - This button toggles on or off normalization (scaling) of the distribution so the entire area under the curve is 1 (100%). When toggled on, the area under a peak represents the fraction of that species.

Convert to s(20,w) button - This button toggles on or off the conversion of the X axis to $s_{20,w}$ units. When toggled on, it brings up a dialog box for entry of the property values needed to make this conversion.

The controls below are shown and available only in Advanced mode

Display options - These options control which type of distribution is displayed [*g(s*)*, *g^(s*)*, or *G(s*)*] and whether error bars are included on the graph.

Reset zero and Undo buttons - The Reset zero button allows manual adjustments of the zero level. Use with caution! Undo will remove the manual adjustment.

Broad algorithm option - This option toggles on or off the use of an alternative algorithm to calculate *g(s*)*. That alternative algorithm gives better estimates of the error bars for the points in the distribution when the peaks are significantly broadened due to excessive boundary movement during the selected scans.

3.2.5.1 Calculate g(s*) distribution - Show integrals over distribution

This command calculates and displays the total concentration and the number-, weight-, z-, and z+1-average sedimentation coefficients (s_n, s_w, s_z, and s_{z+1}) for the entire range of both the *g(s*)* and *g^(s*)* distributions.

Result of Show integrals over distribution command

The results are displayed in a dialog box (example shown above). For more details about the options within this dialog box and how the calculations are done, see Averages and peak areas dialog.

If you want these calculations done only over a portion of the distribution (for example one peak) use the Set integration region button instead.

3.2.5.2 Calculate g(s*) distribution - Set integration region

This command calculates and displays the concentration and the number-, weight-, *z*-, and *z*+1-average sedimentation coefficients (s_n, s_w, s_z, and s_{z+1}) for a selected portion of both the _g(s*)_ and _g^(s*)_ distributions.

The integration region is selected by dragging with the mouse over a region on the graph; only the sedimentation coefficient range is significant during the drag operation, the Y values are irrelevant.

Although in the end the Y range is irrelevant, you must move the mouse sufficiently in the Y direction to change the cursor from the 'no' symbol to a magnifying glass symbol before raising the mouse button to complete the drag operation.

Results after Set integration region operation

The results are displayed in a dialog box (example shown above). For more details about the options within this dialog box and how the calculations are done, see Averages and peak areas dialog.

If you want these calculations done over the entire range of the distribution use the Show integrals over distribution button instead.

3.2.5.3 Averages and peak areas dialog

Averages over the distribution dialog

This dialog box is used to display the results from either the Show integrals over distribution or Set integration region commands within the Calculate g(s*) distribution page.

Note that the values calculated for the _g^(s*)_ distribution are model-independent and are valid for interacting as well as non-interacting systems. The values from the _g(s*)_ distribution involve extrapolation back to the initial loading concentration, and this extrapolation assumes each species is independent (a mixture).

Integration range

The **Integration range** controls at the top govern the range of sedimentation coefficients used in the calculations.

- If the **Full range** option is checked then the entire range for the distribution is used. This is selected automatically when you use the Show integrals over distribution command to bring up this dialog box.
- When you use the Set integration region command the range you selected by dragging over the graph is

filled into the integration range text boxes.

- You may manually alter the integration limits within the text boxes. When you are done editing the value the program will calculate the nearest actual data point value and display that. Then use the **Redo button** to force a new calculation with the modified range.

Calculated properties

The following properties are calculated and displayed, along with their estimated standard deviation:

- The total concentration (area under the curve) in instrument units
  - this is not calculated when the Normalize option is on (the case shown above) since it is the same as the fraction of the total
- The fraction of the total sedimenting material (fraction of the total area under the curve for the full range)
  - this will always be 100% if the Full range option is checked
- The number-, weight-, z-, and z+1-average sedimentation coefficients (s_n, s_w, s_z, and s_{z+1})

Each property is calculated for both the *g^(s*)* and *g(s*)* distributions.

The averages computed and displayed are the so-called "formal" averages, where the *g(s*)* distribution is weighted by the sedimentation coefficient raised to the appropriate power (see equations below). For the so-called "true" averages the *g(s*)* distribution is weighted by powers of the molecular mass. Computing the "true" averages would require an exact relationship or model to assign a molecular mass to each sedimentation coefficient.

The Broad algorithm option may give more accurate estimates of the standard deviations of these computed quantities.

Other features

A **Copy all values to Clipboard button** is provided to copy all the results to the Windows Clipboard so they can be pasted into another document. Individual values may be copied by highlighting them with the mouse and using the CTRL-C copy keyboard shortcut.

The note at the bottom of the form will indicate whether the sedimentation coefficient averages represent uncorrected (raw) values or $s_{20,w}$ values.

Show How are these values calculated?

The average sedimentation coefficients and the concentrations are given by

where the integrations run over the entire range of sedimentation coefficients covered by the distribution. Thus the results will only be valid thermodynamic quantities if the distributions are calculated at a time such that all species are included.

The error bars on these quantities are calculated from the estimated uncertainty for each data point in the sedimentation coefficient distribution functions, σ_g, (which derives from the differences in *dc/dt* data for different scan pairs) through the following formulae derived from the standard propagation-of-error rules:

where n = -1,0,1,2 correspond to s_n, s_w, s_z, and s_{z+1}, respectively.

The uncertainty in loading concentration c(0) is given by

3.2.5.4 Calculate g(s*) distribution - Normalize option

Normalization is used to re-scale the *g(s*)* distribution so the total area under the curve is exactly 1 (100%). For a normalized distribution the area of each peak gives the fraction of that species.

Normalization is particularly useful when you want to visually compare *g(s*)* distributions from samples at different loading concentrations, such as a dilution series, on the g(s*) overlay graph.

Affects on fitting results

When you fit individual species to a normalized distribution, the fit returns the fraction of that species rather than its concentration in instrument units.

Note that the *G(s*)* distribution (the integral distribution) is always normalized, and thus technically the *g(s*)* distribution should also be normalized since it is supposed to be the derivative of *G(s*)*. That is, we should have $g(s^*) = \partial(G(s^*))/\partial s$. However the *g(s*)* distributions created by Walter Stafford's original DCDT algorithm were not normalized, so un-normalized distributions become the usual convention for this method.

3.2.5.5 Convert to s(20,w)

This option will convert sedimentation coefficients for *dc/dt* and *g(s*)* curves, as well as fitted values, to the $s_{20,w}$ scale.

Note that $s_{20,w}$ will also sometimes be written in this document as *s(20,w)* to avoid problems in generating subscripts

The **Convert to $s_{20,w}$ buttons** are present on the Set meniscus & data region page, the Calculate average dc/dt curve page, and the Calculate g(s*) distribution page. When the data are in "raw" sedimentation coefficient unit pressing this button will bring up a dialog box (shown below) for entry of the data needed to perform the conversion (discussed in more detail below).

If this conversion has already been done (as indicated by the button remaining in the down state) then pressing the button will turn off the conversion and revert to raw *s* values.

Conversion to s(20,w) dialog

It is generally desirable to present and report $s_{20,w}$ values rather than raw (uncorrected) sedimentation coefficients. Conversion from raw s to $s_{20,w}$ compensates for the factors such as buffer viscosity which are relevant only to a particular experiment, giving the sedimentation coefficient that molecule would have if measured in water at 20 °C. Because of the solvent-independent nature of $s_{20,w}$, it describes quantitatively the fundamental hydrodynamic properties of the molecule and it is this value which is most useful in comparing the sedimentation behavior of different molecules.

The partial specific volume that is entered on this form and used in converting to $s_{20,w}$ scale is applied uniformly to the entire $g(s^*)$ distribution. Therefore if you believe your sample contains species with different vbar values (for example different polypeptides, or both PEGylated and non-PEGylated forms of the same protein) you should not apply this conversion---you must work in raw sedimentation coefficient units.

- For similar reasons, the partial specific volume entered on this conversion form is applied to all species that are fitted.

The actual scaling factors that will be applied to correct the raw sedimentation and diffusion coefficients to standard conditions are displayed in the panel nearest the bottom of the dialog box.

Show All Show All

Hide All Hide All

Show What about extrapolation to zero concentration?

This program does not attempt to correct for concentration dependence of sedimentation coefficients (extrapolation to infinite dilution) to give $s^0_{20,w}$. Often dc/dt analysis is applied to samples at concentrations sufficiently low that concentration dependence can be neglected, and if so the corrected results can be considered to be $s^0_{20,w}$ values, but the program will not explicitly designate them as such.

Show How this conversion affects the $g(s^*)$ distribution and average dc/dt curve

Please be aware that conversion to $s_{20,w}$ also alters the magnitudes of the $g(s^*)$ and $g^\wedge(s^*)$ distributions. This arises because these distributions are defined such that the area under the curve gives the concentration. Thus if the X-axis is expanded or contracted as a consequence of conversion to the *s20,w* scale, the Y-axis must be contracted or expanded to keep the area unchanged.

Internally in the program the dc/dt, $g(s^*)$, and $g^\wedge(s^*)$ data are always kept in raw units---the conversion is only applied as graphs are generated and values are displayed and printed. Generally this is transparent to the user, but with regard to setting the δs intervals for the average dc/dt and $g(s^*)$ curves it is not possible to have conveniently rounded intervals between data points in both raw and $s_{20,w}$ scales. Consequently switching between raw and $s_{20,w}$ units will usually force a re-calculation of the average dc/dt curve and $g(s^*)$ distribution because the X-axis point spacing will change if that is being determined by the default **Automatic option** (see Calculate average dc/dt - X-axis point spacing). Thus the average dc/dt curve and $g(s^*)$ distribution will be slightly different in the different units (the difference is not simply a multiplicative factor for the X-axis). However those differences will be negligible compared to the noise in the data.

Show How this conversion affects results from fitting

When this conversion is used, the sedimentation coefficients obtained by fitting will also be $s_{20,w}$ values, and the diffusion coefficients from fitting will be $D_{20,w}$ values. They will be marked as such on all program outputs, and the values used to make the conversion will be printed along with the printed reports of fit results.

Switching between raw and $s_{20,w}$ units will usually force a re-calculation of the average *dc/dt* curve and *g(s*)* distribution because the X-axis point spacing will change if that is being determined by the default Automatic option (see [Calculate average dc/dt - X-axis point spacing]). Thus making this switch will invalidate any previous fitting results.

Show [How this conversion affects reported diffusion coefficients]

Activating this option will also cause raw diffusion coefficients to be converted to $D_{20,w}$ values, via the formula

where *T* is the sample temperature (Kelvin).

Show [Data required for conversion]

Conversion to $s_{20,w}$ requires knowledge of the solvent density at the experiment temperature, the solvent viscosity, h, at the experiment temperature, the solute partial specific volume (vbar) at the experiment temperature, and the solute partial specific volume at 20 °C, through the relation:

where the subscripts refer to the temperature and solvent conditions (b=buffer, w=water). The density and viscosity of water at 20 °C are 0.998234 g/ml and 1.002 centipoise, respectively.

If all those data are not available, or cannot be estimated (*e.g.* by SEDNTERP), then the conversion cannot be made.

The program will optionally automatically compute a temperature correction to the partial specific volume, using the same formula used in SEDNTERP [reference 4]:

This option is controlled by the check box on the lower frame, and when it is checked the input box for the vbar at 20 °C is disabled (but the computed value is displayed).

The viscosity values are to be entered in centipoise. Since SEDNTERP displays these values in poise, to facilitate copying and pasting numbers from SEDNTERP, whenever a number is entered that is less than 0.1 the program will assume it is a value in poise and multiply it by 100.

The program automatically fills in the experiment temperature based on the temperature recorded in the header of the last scan used in calculating the *dc/dt* data, but since the temperature may drift during the run you may wish to adjust that value manually.

Please note that changing the temperature on this form overrides the temperature recorded with the scans, and also alters the value used elsewhere in the program (*e.g.* for conversions to buoyant or true mass units).

3.2.5.6 Calculate g(s*) distribution - Display options

g(s*) distribution display options

This group of controls governs which type of sedimentation coefficient distribution is displayed, and whether errors bars (+/- 1 standard deviation) are shown on the graph.

- When the **G(s*) button** is depressed the integral distribution ("big G of s") is displayed rather than the differential distribution ("little g of s"). This *G(s*)* distribution measures the fraction of the sample that sediments with sedimentation coefficients less than or equal to each value of *s**.
    - Note that this *G(s*)* distribution is calculated directly from the *g(s*)* distribution by integration, and is not corrected for diffusion; i.e., it is not equivalent to the *G(s)* distribution obtained by the van Holde-Weischet method. *G(s*)* is always normalized (the values are fractions and range from 0 to 1).
- When the **g^(s*) button** is depressed the *g^(s*)* data is displayed rather than *g(s*)*.
- When both the **G(s*) button** and **g^(s*) button** are depressed the integral distribution *G^(s*)* is displayed.
- The **With error bars button** is a toggle switch that turns on display of error bars when it is depressed.
    - Error bars are not displayed for the integral distributions.

3.2.5.7 Calculate g(s*) distribution - Reset zero & Undo buttons

Reset zero & Undo buttons

The **Reset zero button** is used to manually redefine the zero level for the *g(s**) distribution. This may be necessary to compensate for uncorrected "jitter" in interference data and/or other effects that are causing shifts of the zero level. If uncorrected, such shifts may cause the *g(s*)* distribution to go below zero or remain above zero in regions away from the peaks.

Since this manual zero adjustment could bias the results, its use on a routine basis is **not** recommended.

When the button is pressed a red horizontal bar that will define the new zero level is displayed on the graph of *g(s*)*. The bar can be dragged up and down using the mouse. At the end of the drag operation the new zero position is used to recalculate and re-display the *g(s**) distribution.

The **Undo button** will restore the zero level to the original, experimental value.

When one or more peaks are fitted to *g(s*)* data which have been manually re-zeroed, the amount of the zero shift will be printed in the report of the fitting results.

Bringing the *g(s*)* curve to zero at any X-axis position can only be done approximately, due to the fact that the *dc/dt* value at any value of *s** contains contributions from the plateau regions of all species with lower values of *s*. Thus two or more iterations of pushing the **Reset zero** button may be necessary to bring that point completely to zero, particularly if the chosen *s* value is to the right of a significant peak.

Because the transformation from *dc/dt* to *g(s*)* involves dividing each point in *dc/dt* by its corresponding *s* value, the effects of re-zeroing *g(s*)* at a certain *s* value will be to shift the *g(s*)* data more strongly at lower *s* values and less strongly at higher *s* values. As a consequence, picking a re-zeroing point far to the right in the data may have undesirable effects at small *s* values.

3.2.5.8 Calculate g(s*) distribution - Broad algorithm option

This button will toggle on or off the use of the alternate "broad" algorithm for calculating the *g(s*)* distribution and the average *dc/dt* curve. This choice primarily affects only the estimates of the error bar for each data point. See What is the difference between the broad and conventional g(s*) algorithms? for a detailed explanation.

3.2.6 Select fitting model page

This page is used for selecting the theoretical model which will be used to fit the data (number of species, choice of fitting to *g(s*)* or *dc/dt* curves), as well as certain options which affect the way the data and fitting parameters are treated.

Each control or group of controls on the image below is a clickable *hotspot* that will lead to further information, or just use the numbered list below the image.

Select fitting model page

Below are brief descriptions of the controls on this page; more details can be found by following the links.

Number of species - Set using the up/down spinner or type a number into the text box (allowable range 1-5).

Molecular mass units - Choose among true mass, buoyant mass, or *s/D* ratio. Only relevant when the Fit M (s/D ratio) instead of D option is on (always true in Standard mode).

Density and partial specific volumes - Set the solvent density and sample partial specific volume(s). Only relevant when the Fit M (s/D ratio) instead of D option is on (always true in Standard mode).

The controls below are shown and available only in Advanced mode

Data choice - You may fit either to the *g(s*)* distribution or the average *dc/dt* data.

Fit diffusion or mass - Choose either *D* or molecular mass (*s/D* ratio) as the second hydrodynamic parameter used to characterize each species.

Improved or classic methods - Choose between alternative methods for computing the fitting functions. The improved method is recommended for all new analyses; the classic method is provided for backward compatibility with versions 1.xx.

Weighting of data points - Use theoretical weighting (recommended), experimental weighting, or no weighting.

Constrain *M* or *s* ratios among species - Use constraints to force species to behave as expected for a series of oligomers.

3.2.6.1 Select fitting model - Number of species

Number of species selection

You may select 1 to 5 species by either typing the number into the text box or using the up and down arrows on the spinner control to increment the number up or down.

It is important to keep in mind that for multi-species fits

1. the different species are assumed to sediment independently (behave as a simple mixture)
2. the different species are assumed not to inter-convert over the time course of the velocity run

Thus this program is generally not appropriate for analysis of interacting systems unless the rates of the association-dissociation process are very slow and the sample is effectively a mixture.

Further, the fact that the program implements five species is not meant to imply that you will be successful at fitting that many to your data! Resolving five species would only be possible for very high molecular mass species, and even then it would likely be necessary to impose constraints on their masses or sedimentation coefficients

3.2.6.2 Select fitting model - Improved or classic methods

Function computation options

These radio buttons choose between alternative methods for computing the fitting functions. They are shown only in Advanced mode.

The **Classic method** is provided primarily to provide backward compatibility with DCDT+ versions 1.xx. The **Improved method** is recommended for all new analyses. The differences between these approaches are discussed fully in About the fitting functions.

- In the **Classic method** the fitted values for *D* or *M* are quite sensitive to the peak broadening that occurs as the scans used for analysis cover a larger and larger time span (see reference 8). The sedimentation coefficients are also subject to systematic errors, but that is a smaller effect. The **improved methods** eliminate those systematic errors.

The **Improved method** is the default for all new analyses.

3.2.6.3 Select fitting model - Data choice

Data choice options

These radio buttons determine whether the data points to be fitted are from the *g(s*)* distribution or from the average *dc/dt* curve. In Standard mode the *g(s*)* data are automatically chosen. For a discussion of the advantages and drawbacks of fitting to *dc/dt* versus *g(s*)* see Why do I get different answers when fitting g(s*) versus dc/dt?

3.2.6.4 Select fitting model - Fit diffusion or mass

Fit diffusion or mass options

Internally DCDT+ always uses sedimentation coefficient, *s*, and the diffusion coefficient, *D*, as the fitting parameters which characterize the hydrodynamic properties of each species.

However, instead of using *D* as a fitting parameter, it is often desirable to instead use the ratio *s/D*, a quantity which is proportional to molecular mass, *M*. In fact, fitting to mass rather than *D* is the initial program default when installed, but that can be altered via the Options dialog.

One significant advantage of fitting *M* rather than *D* is that this allows using constraints on the mass values for different species (for example fitting to a mixture of monomer and dimer, with the monomer mass as a fitting parameter). See Constrain M or s ratios among species.

Why this works

The **Svedberg equation** shows that the *s*/*D* ratio is directly related to molecular mass, *M*:

where *R* is the gas constant, is the partial specific volume, ρ is the solvent density, and T is temperature (Kelvin).

How this works

As stated above, internally the program always uses the value of *D* to determine the shape of the corresponding peak. When you choose to use *M* as the fitting parameter, during each iteration of the fitter the correct *D* value is calculated from the current values of *s* and *M*. Since variations in the *s* value will also alter the apparent mass (the *s*/*D* ratio), it is also necessary to adjust the derivatives of the fitting function to account for that effect.

3.2.6.5 Select fitting model - Weighting of data points

Weighting options

This group of three radio buttons determines whether you wish to assign a statistical weight to each data point, or instead to weight all the data points equally (the **None** option).

For weighted fits there are two choices for determining the weight to give each point:

1. The **Experimental weights** option uses the standard deviation, σ, assigned to each data point in the *g(s*)* or *dc*/*dt* data. The weights are then given by $1/\sigma^2$, so the data points with the highest precision (lowest σ) are weighted most heavily.
2. The **Theoretical weights** option (not available prior to version 2.0) assigns a weight to each point in the *g(s*)* distribution on a theoretical basis, assuming the uncertainties for all points in the original raw scan files are equivalent. This is now the program default when fitting to *g(s**) distributions.

Show All Show All

Hide All Hide All

Show Why experimental weights are not recommended

The drawback to experimental weights is that the σ value for each data point is determined from the differences among the *dc*/*dt* curves for individual scan pairs that went into computing the average *dc*/*dt* curve. The problem with that approach to estimating σ is that the amplitude and width of each *dc*/*dt* curve varies with time during the run, and unless the scans cover a very narrow time span much of the variation among the *dc*/*dt* curves for individual scan pairs is due to those systematic changes rather than noise in the raw data. (This is discussed in detail in reference 8.)

Consequently the σ values are often grossly over-estimated, particularly when the number of scans grows larger, and particularly at the tops of the peaks. Thus when experimental weights are used, it is often the case that the most important data points near the top of a peak are assigned very low weights. It is also true that as you add more scans into the analysis to improve the signal/noise, ironically the estimated σ values often increase. Because of these phenomena **the use of experimental weights is not recommended**; this option is offered primarily for backward-compatibility with DCDT+ versions 1.xx.

Show How are the theoretical weights calculated?

Because each data point in the average *dc/dt* curve is divided by its corresponding *s** value during the transformation from *dc/dt* to *g(s*)*, the theoretical uncertainty for the *g(s*)* values is much higher when the value of *s** is near zero (the left end of the distribution) than it is when *s** is large (the right end of the distribution). Thus the theoretical weight of each point is simply given by *A* x (*s**)^2, where the constant *A* is chosen so the sum of all weights equals N, the total number of data points (as is true when all points are weighted equally). These theoretical weights generally give a much more realistic estimate of the relative precision of data from different regions of the *g(s*)* distribution than the experimental ones, and therefore this is now the default for fitting *g(s*)*.

When fitting to the *dc/dt* data the theoretical weight for all points is the same (the same as for an unweighted fit) so the theoretical weight option is disabled. An unweighted fit is now the default when fitting to the *dc/dt* data.

Show Consequences for reported fit statistics

If you elect an experimentally weighted fit, a *chi squared* value will be calculated and displayed with the fit results; for both unweighted and theoretically-weighted fits the *sum of squared residuals* will be calculated and displayed.

3.2.6.6 Select fitting model - Constrain M or s ratios among species

Constraint options

Parameter constraints are used with multi-species fits when there is reason to believe two or more species represent oligomers of a common monomer.

The **Set constraint values button** brings up a dialog box that allows the sedimentation coefficients and/or masses of the different species to be linked together and constrained to the ratios expected for oligomers. By constraining these ratios, the number of fitted hydrodynamic parameters is greatly reduced, often resulting in more accurate values for the other parameters that are fitted.

Whenever constraints are currently active, the **Constrain *M* or *s* ratios among species** check box will be checked. Clearing this check box will inactivate all constraints. Setting this check box when no constraints are active will bring up the dialog to set constraints.

The actual dialog box and procedure for establishing the constraints is described in detail below, but first it may be useful to explain the intended use for these constraints.

Show All Show All

Hide All Hide All

Show Why are constraints useful?

As an example, a common degradation pathway for proteins upon storage is the formation of disulfide-linked dimers or higher oligomers, and thus many protein samples could contain a small amount of dimer. If the fraction of dimer is quite small, it will be very difficult for the fit to accurately evaluate its sedimentation and diffusion coefficients (especially *D*).

Therefore if you want more accuracy for the amount and sedimentation coefficient of the dimer fraction (and there is a good reason for believing this is a dimer, such as evidence from non-reducing gels or SEC-MALLS) it may be a good idea to constrain the *D* value of the dimer to be consistent with a mass twice that of monomer. That is, if s_1 and D_1 are the current values of *s* and *D* for the monomer, and s_2 and D_2 are the current *s* and *D* values for the dimer, then at every iteration of the fit the program can force the s_2/D_2 ratio to be exactly twice the s_1/D_1 ratio; i.e., $D_2 = s_2 \times D_1 / (2 \times s_1)$.

Similarly, if the fraction of dimer is very small (a few percent) or the signal/noise is poor, it may also be appropriate to constrain the dimer sedimentation coefficients to be an appropriate multiple of the monomer s_1 value. If each monomer is a hard sphere, then theory says that $s_2 = 1.45 \times s_1$ (reference 10). At the other extreme, if two spherical monomers merged to form a spherical dimer, then $s_2 = 2^{2/3} * s_1 = 1.59 * s_1$. Thus the difficulty with constraining sedimentation coefficients is that the ratios depend on conformation.

These constraints can be very useful for hypothesis testing. If we see that a sample fits poorly as a single species, and/or the fit implies a molecular mass significantly less than a known value (broader boundaries), then you may wish to test the hypothesis that the poor result is due to the presence of a small amount of oligomer. This can be done by first constraining both the *M* and *s* value of the second species and fitting. If this gives a good fit and reasonable values for the monomer parameters, then you could conclude that the hypothesis is satisfied. At that point you would probably also want to try releasing the constraint on the oligomer *s* value, re-fitting, and noting whether the fitted dimer *s* value is still consistent with the assumed oligomer size. If this is also true you would have still more confidence that your hypothesis is true. See Tutorial: Multi-species fits for an example.

Show Why are constraints sometimes needed?

The need for constraints arises when:

1. The separation between two species is too poor to uniquely define all the properties of both of them.
2. Only a small amount of a species is present, and either the signal/noise or the separation is too poor to permit uniquely determining all three of its properties (*c, s, and D*).
3. You are more concerned about obtaining accurate values for certain properties, but others are less important. The use of constraints (if they are appropriate ones) can often greatly increase the precision of the remaining fitted parameters.

The constraint dialog box:

The dialog box used for establishing or removing the constraints is shown below. In this example, the fitting model includes three species, so the controls related to species 4 and 5 are inactive (grayed).

Constraint dialog box

Mass constraints

The image shows an example of a 3-species fit where species 2 is believed to be a dimer of species 1. Hence on the left side the molecular mass of species 2 is set up to be constrained to be exactly twice that of species 1 (whatever that may be).

The desired ratio of masses for the constrained species is entered into the text box

Once the constraint link is established to species 1, species 1 cannot itself be linked to other species (if there were any), so its linking box becomes inactive. As shown, the molecular mass of species 3 is not constrained, but it could still be linked to that of species 1 (perhaps as a trimer or tetramer) and thus the group of radio buttons for linking the mass of species 3 to another species is still active.

An alternative type of mass constraint is also available. For **constant f/f_0 ratio mass scaling** the mass of the constrained species is calculated assuming this species has the same hydrodynamic shape (same frictional coefficient ratio relative to an anhydrous sphere, f/f_0) as the species to which it is linked. This constraint is therefore equivalent to the assumption used when calculating the *c(s)* distribution in Peter Schuck's program SEDFIT. When the f/f_0 ratio is constant the mass scales with the sedimentation coefficient to the 3/2 power. Thus if species 2 is linked to species 1 with this type of constraint we have

To use this constant f/f_0 ratio mass scaling simply enter -1 as the mass ratio in the text box.

Sedimentation coefficient constraints

On the right side, the sedimentation coefficient of species 2 is also constrained to be a multiple of that for species 1.

The ratio of sedimentation coefficients for the constrained species is entered into the text box .

The table at lower right shows theoretical ratios for oligomers of hard spheres. The table has room only for oligomers up to tetramer, but ratios for pentamers, hexamers, and cubic octamer replace that table when the **Show bigger oligomers** button is pressed.

In this example species 2 is being constrained at the 1.45 ratio appropriate for a dimer of hard spheres. These ratios for oligomers come from Garcia de la Torre, J. and V. A. Bloomfield (1981), reference 10.

In some cases it may be better to constrain the smaller species to be a sub-multiple of a larger one. For example, if you had a sample where the native protein is a dimer, but it contains a small fraction of a denatured or "incompetent" monomer, it would be best to constrain the monomer mass to be half that of the dimer, and perhaps set the monomer *s* value to be 0.69 that of the dimer (i.e., the inverse of the 1.45 ratio shown in the table).

Other buttons

When you are done setting constraints the **Ok button** will accept the current settings and return you to the Select fitting model page.

If you want to undo any modifications and return to the settings present when this dialog box was opened, use the **Cancel button**, which will close the dialog and return you to the Select fitting model page.

This group of 3 buttons provides quick mechanisms to clear out unwanted constraints. The usage is obvious from the names of the buttons.

Note that if you reduce the number of species from the Select fitting model page after constraints are active, any constraints to or from species that no longer exist in the model will be automatically cleared. You will not need to bring up this dialog box to clear them.

3.2.6.7 Select fitting model - Molecular mass units

Molecular mass units

This group of controls allows you to choose among three systems of units for molecular mass. This choice is only relevant if you have chosen to use molecular mass (the *s/D* ratio) as the second hydrodynamic parameter for the fit within the Fit diffusion or mass group box.

Generally you will want to use units of true molecular mass, and this is the initial program default.

However, in some cases you may not know the values for both ρ and ☒ and thus cannot calculate the true mass. In other cases you may wish to work directly with the buoyant molecular mass or in units of the *s/D* ratio (Svedbergs per *Fick*).

To set the molecular mass units:

1. Use the radio buttons to select among the three choices of units.
2. To use units of buoyant molecular mass or true mass, the sample temperature must be known.
    - The program automatically fills in the sample temperature based on the temperature recorded in the header of the last scan used in calculating the *dc/dt* data. However, because the temperature may drift during the run you may wish to adjust that value manually. The program internally converts the temperatures from Celsius to Kelvin scales.
    - Please note that changing the temperature on this page overrides the temperature recorded with the scans, and also alters the value used elsewhere in the program (*e.g.* for conversions from raw *s* to *s (20,w)*.
3. If you select true molecular mass, go to the Density and partial specific volumes group on the lower right side of the page and fill in the values for solvent density and the partial specific volume(s).

☒ To obtain accurate true molecular mass you must know the solvent density and solute partial specific volume accurately.

Note that changing the molecular mass units on this page will automatically alter the results displayed on the Report fit results page if a fit has been completed.

3.2.6.8 Select fitting model - Density and partial specific volumes

☒ Density and partial specific volume group

This group of controls is used to set values for the solvent density, ρ, and sample partial specific volume, ☒. Those values are only relevant and needed when you have chosen to use molecular mass (the *s/D* ratio) as the second hydrodynamic parameter for the fit within the Fit diffusion or mass group box, and when you have chosen units of true molecular mass within the Molecular mass units group box.

- When checked the **Use the same partial specific volume for all species** check box causes the program to assign the same ☒ to all species (as would be true, for example, if they are all oligomers of a single type of monomer). This is the default state of the program.
    - If your sample contains more than one chemical species, uncheck the box and then fill in the correct values for the other species.
    - If you later turn on the **Use the same partial specific volume for all species check box**, the value entered for species #1 will automatically assigned to all the other species.
- When the program starts, ρ and ☒ are set to default values. These default values may be altered using the Options dialog.
- When you alter ρ or ☒ on this page, those values will also be used in other calculations applied to this

analysis window, such as converting raw sedimentation coefficients to *s(20,w)* values.

3.2.7 Set initial guesses on graph page

This page allows you to:

1. set the sedimentation coefficient range of the data to be fitted (indicated by vertical green lines)
2. set an initial guess for the sedimentation coefficient of each species (vertical magenta lines)

Set initial guesses on graph page

How to set the fit range:

You may wish to exclude some data at low or high sedimentation coefficients, either because those regions are noisy and unreliable, or because the data in those regions are influenced by species that are too poorly defined to be included in the fit.

The actual setting of the sedimentation coefficient range is usually done by dragging the green vertical bars that mark the limits of data to include in the fit.

- During the range setting procedure the actual position of the limit bar is constantly displayed in the user prompt area at the bottom of the analysis window. At the end of the drag operation the limit will be set to the nearest value for which there is an actual data point.
- Upon entry into this procedure, if a range has not previously been set for these data the program defaults to the full data range. When the upper or lower limit are at the limit of the data, the background color of the corresponding text box is set to light yellow.
- If you have previously set a range for fitting *g(s*)* data, and then change to fitting the *dc/dt* data (or vice versa), the program will assume you wish to use the same range and sets this range for you.

Alternatively you can directly type the desired sedimentation coefficient range into the text boxes at the upper left.

How to set the initial guesses for peak positions:

A magenta vertical bar will be displayed for each species (2 in the image above), which you may drag with the mouse to mark an initial guess for the sedimentation coefficient of that species (usually the center of the peak).

- If this is new data a program wizard will automatically find the highest peak and mark this as the position of the first species.
- You are allowed to drag the markers over top of one another---the positions will be sorted in order of increasing sedimentation coefficient and assigned to species numbers in that order.

When are these new guesses applied?

The two radio buttons within the **Species properties updating** group box govern whether new guesses for concentration and *D* or *M* will be made based on the currently marked peak positions when the **Next** button is used, or alternatively whether the current values (from a previous fit) should be retained.

- By default whenever the magenta peak position markers are dragged the program assumes this means you will want new estimates for c and (D or M), so it automatically selects the **Make new concentration and peak width guesses based on new marked positions** option.

- If you instead want to force the program to retain the current values then manually select the **Proceed using previous concentration and peak width values** option.

How are the initial guesses for c and (D or M) computed?

An initial guess for the concentration is made based on the magnitude of the data at the marked peak position (the peak height), and a guess for D or M is made based on the peak width at half-height.

Since the peak widths are used in calculating the D or M value, these guesses will probably not be very accurate for peaks that are poorly resolved. The peak width is also used in determining the guess for the concentrations, so those values may also be significantly in error for poorly-resolved peaks. Nonetheless these initial guesses are generally more than adequate starting points and will usually allow the fit to converge.

3.2.8 Set/alter fitting parameters page

This page is used to manually enter or alter the initial guesses for the fitting parameters, for specifying whether those parameters will be fitted or held constant, and for specifying the criterion for convergence of the fit. In Standard mode this page is skipped. If you move to this page from the Set initial guesses on graph page you will see the initial guesses that were computed based on your inputs there.

Each control or group of controls on the image below is a clickable *hotspot* that will lead to further information, or just use the numbered list below the image.

This particular example is that for a fit involving 2 species, and where the molecular mass of species 2 is being constrained to be a multiple of that of species 1.

Set/alter fitting parameters page

Below are brief descriptions of the controls on this page; more details can be found by following the links.

Parameter value text boxes - Input or alter the initial guess (for parameters being fitted) or the desired value for that parameter (for parameters being held constant).

Fit control check boxes - When checked that parameter will be fitted ("floated"); when unchecked that parameter is held constant ("fixed").

Constrained parameters - The values for parameters that are constrained are neither fitted nor fixed. Their values cannot be altered from this dialog; they are computed based on the values of other parameters. In the case shown species 2 is constrained to have twice the molecular mass of species 1 (species 2 is a dimer). See Constrain M or s ratios among species.

This is an example of a parameter (the zero offset) that is being held fixed (at zero in this case). That is the program default for this parameter. See Should I fit the offset?

Advanced topic Convergence control - These two values are used in determining when a fit has converged. The default values work well in most situations. Lower values will produce more iterations, causing the fitter to come ever closer to the best possible fit, but those extra iterations will probably not produce changes that are significant compared to the statistical precision of the parameter values. Higher settings on these values may help the fitter reach convergence on particularly long and difficult fits.

Advanced topic Adjust parameter upper/lower limits - This brings up a dialog box for altering the default upper and lower bounds on each parameter. Generally changes in these values are not needed.

Cancel changes - This is an 'Undo' button that restores all values and settings as they were when this analysis page was first entered.

3.2.8.1 Parameter value text boxes

Parameter value text boxes

Use these text boxes to input starting guesses for the parameters (or fixed values if that parameter is not being fitted).

- The units for concentrations are absorbance units or fringes, sedimentation coefficients are in Svedbergs, diffusion coefficients are in *Ficks*, and masses are either in kDa or Svedberg/Fick (depending on your choice of molecular mass units).
- You may enter values in scientific notation if you choose, using formats such as '1.23E1' or '1.23e1', '1.23D1' or '1.23d1'. The ESCape key will return you to the previous value. Warnings will be issued if you enter values outside of the current upper and lower bounds on that parameter.

3.2.8.2 Fit control check boxes

Fit control check boxes

These check boxes control whether a parameter will be varied iteratively during the fit ("floated") or held constant at the value specified in the text box to the left ("fixed"). The text next to the check box will correspondingly read "Yes" or "No".

TIP: When a fit is having difficulty converging it is often helpful to temporarily fix one or more parameters so the fitter can get the other parameters into a reasonable range (giving a rough fit of the experimental data). Then return to the Set/alter fitting parameters page and turn on fitting of one more parameter, and re-fit, continuing this process until all parameters are actively fitted.

The boxes are inactive (grayed) and read "constrained" for constrained parameters.

3.2.8.3 Constrained parameters

Constrained parameters

Users cannot input values for those parameters whose values are being computed via constraints (see Constrain M or s ratios among species). The current values of those parameters are displayed with a light yellow background to indicate that these are computed quantities. The corresponding Fit control check box is also inactive (since the parameter is automatically being fitted) and reads "constrained" as a reminder of that condition.

Note that the displayed value will not change if you manually alter the parameter to which this parameter is linked---the constrained values are recomputed just prior to each fit iteration.

3.2.8.4 Convergence control

Convergence control

The **Convergence factor** and **Parameter precision factor** are numerical factors used in determining when a fit has converged. The criteria used to detect convergence are either

1. the fit iterations are producing insignificantly small changes in all the fitting parameters
2. the fit iterations are lowering the *sum of squared residuals* by an amount that is too small to be statistically significant.

Specifying a lower value for these factors will produce more iterations, but those iterations will be less and less likely to produce significant changes in parameter values. Specifying a larger value for them can help avoid wasted iterations, especially in situations where the parameters are highly correlated.

In any iterative fitting scheme the iterations can proceed indefinitely. However, if the fit is successful, at some point the sum of squared residuals will approach a minimum value and the changes in parameters between successive iteration will be very small, so further iterations are pointless. Therefore it is necessary to have a rule to define when the fit has converged.

The first criterion for convergence is based on a percentage precision on the parameters. Fitting continues until the changes in all parameters with each iteration are below this precision. By default this program will define the fit to have converged if the change in all parameters is less than 0.05%, but this can be altered from the Set/alter fitting parameters page.

The problem with using parameter precision as the only convergence criterion is that usually some of the parameters can be determined with much higher precision than others, and many many iterations can be spent trying to get high precision on a parameter whose standard error will be very large. In this program, it is almost always possible to determine the sedimentation coefficients with much higher precision than the diffusion coefficient (or molecular mass).

This program avoids this problem by also basing convergence on a second test, the size of the changes in the sum of squared residuals (and in general it is this criterion which halts the iterations first). Statistics tells us that a change in the value of a parameter will not be statistically significant unless that change increases the sum of squared residuals by a factor of approximately 1/N, where N is the number of data points being fitted. Thus once successive iterations of the fit produce a fractional drop in the sum of squared residuals that is smaller than 1/N, further iterations will produce refinements in the parameters that are not statistically significant.

The actual implementation is that the fit will continue until the fractional drop in the sum of squared residuals is less than 1/N times the "convergence factor". Thus if the convergence factor is 0.01 (the default value), the fit continues until the changes are 100 times smaller than the smallest change that is likely to be statistically significant.

For multi-species fits, or in situations where the parameters are very highly correlated, it may be worthwhile to increase the values of the convergence factor or parameter precision factor to reduce computation time.

3.2.8.5 Adjust parameter upper/lower limits button

Adjust parameter upper/lower limits button

This button will bring up the dialog box below for altering the default upper and lower limits bounds) on parameter values.

Adjust parameter limits dialog

Show All Show All

Hide All Hide All

Show What is the purpose of these limits?

The intent of bounding the parameters is to prevent the fitter from straying into areas where the parameters would make no physical sense. Generally the fitter never needs to impose bounding, and the user will not need to alter the default values for the bounds, but in certain situations the program defaults may be inappropriate and could potentially prevent convergence.

Show How are the default limits calculated?

1. Default limits for concentrations are calculated based on the data type (absorbance, interference or fluorescence).
2. Default limits for sedimentation coefficients are set based on the range of values within the data (0.8 x the lowest value, 1.5 x the highest value).
3. Default diffusion coefficient or molecular weight limits are then calculated based on those sedimentation coefficient limits and estimates of values for anhydrous spheres.
    - For example the maximum diffusion coefficient is assumed to be 3 times the anhydrous sphere value corresponding to the lowest sedimentation coefficient; the maximum *M* value is set as 10X that for an anhydrous sphere with the maximum sedimentation coefficient.)

Show How to alter the limits

The user is presented with the dialog box shown above, containing a list of the current upper and lower bounds. These can be manually altered as desired.

Note that the alterations will apply only to the current analysis window---they do not change how the program computes default values.

The **OK button** will accept any alterations that have been made and exit the form; the **Cancel button** will leave the values unchanged.

Show What happens if a parameter goes outside the limits?

During each round of fitting the new parameters are tested against the limits. If any parameters will go out of bounds, the increments for all parameters are reduced proportionally using the increment reduction factor until none will go out of bounds. That procedure moves all parameters in the direction, and in the relative proportions, that should improve the quality of the fit. However this bounding procedure may prevent the 'good' parameters (the ones not exceeding the bounds) from ever reaching the values needed for an optimum fit, i.e. often when any parameter is going out of bounds the fit will never converge.

Therefore it is often useful to warn the user that parameters are trying to go out of bounds. This is done by automatically switching the Fit Monitor to the Fitting parameters monitor tab and displaying the value of that parameter with a bright yellow background.

3.2.9 Perform least-squares fit page

This analysis page controls and monitors the non-linear least-squares fitting process.

- In most cases you will simply push the Do fit button and watch the theoretical curve move toward the experimental data as the fit converges.
- However you are also provided options of proceeding one iteration at a time, or aborting the process, and of monitoring the actual parameter values and fit statistics rather than the graph.

This page is divided into two tabs:

1. the **Graphical fit monitor tab**, which will be shown and described below, and
2. the Fitting parameters monitor tab, which is described in detail in its own Help topic.

The Do fit button, 1 iteration button, Simulate button, and Cancel button are common to both tabs, as are displays of the fit Status and the current iteration number.

Both tabs are available in both Standard and Advanced modes, although the Fitting parameters monitor tab will be of more interest to advanced users.

Graphical fit monitor

The image above shows this analysis page immediately after entry from the Set initial guesses on graph page for a two-species fit.

Fit overlay and species plot - The upper graph panel displays an overlay of the data to be fitted (displayed as circular points) and the initial theoretical fit (cyan line) based on the initial peak positions that were set on the Set initial guesses on graph page. The graph also shows the individual theoretical contribution from each species (blue and green lines, which when summed give the cyan line).

Residual plot - The lower graph panel displays a *residual* plot (red line). The vertical scale of this plot will change as the fit proceeds and the residuals grow smaller.

Status box - The Status box will display status messages as the fit proceeds. In this case the program is waiting for the Do fit button, 1 iteration button, Simulate button, or Cancel button to be pressed.

Iteration counter - The iteration counter simply displays the total number of iterations in this fitting session (0 in this case since fitting has not started).

Simulate button - The Simulate button simply computes the theoretical curves based on the current parameter settings. The primary purpose is that this will enable movement to the Graph fit results page. Thus by manually modifying parameters on the Set/alter fitting parameters page and then using the Simulate button you can generate theoretical plots, species plots, *etc.* for any desired set of parameters (not necessarily ones that give good fits).

Do fit button - The Do fit button tells the fitter to proceed, without halting after each iteration. This is the normal mode, and thus this button is given the *focus* by default when you enter this page so you can initiate the fit by pressing the **Enter** key.

1 iteration button - The 1 iteration button is used to manually step through the fitting procedure one iteration at a time. This can be a useful learning exercise, but generally this would only be used when you are having trouble obtaining a fit and want to carefully monitor what is happening, step-by-step.

Cancel button - The Cancel button will abort the current fit, stopping any calculations that are in progress. The Back button can then be used as needed to alter parameter guesses or the fitting model before trying to fit again.

Fitting parameters monitor tab - Select this tab to monitor the actual values of the parameters and the decrease of the *sum of squared residuals* as the fit progresses.

3.2.9.1 Perform fit - Do fit button

Do fit button

This button will allow the fit to proceed without pausing after each iteration.

This is the normal way to complete a fit. The fitting algorithm will iterate, seeking parameters that give the lowest *sum of squared residuals* (lowest χ^2 for weighted fits). The iterations will halt when the iterations no longer give improvement (the fit has converged). At that point the program will automatically move to the Report fit results page to display the results.

Once fitting is started, it can be aborted using the Cancel button.

3.2.9.2 Perform fit - 1 iteration button

1 iteration button

This button will allow the fit to proceed through only one iteration, after which it will pause.

This procedure allows time for examination of trends in the parameter values, and whether the initial guesses for the parameters are reasonable. This may be a good choice for the initial stages of a difficult fit. If the parameters are moving in a direction that makes no physical sense, you can go back to the Set/alter fitting parameters page and perhaps try holding one or more parameters constant until the other parameters get into a better range, or perhaps try using parameter constraints.

At any point the fit may be continued without pausing by using the **Do Fit button**.

If the requested single iteration produces convergence the program will automatically move to the Report fit results page to display the results.

3.2.9.3 Perform fit - Simulate button

Simulate button

The **Simulate button** forces the program to calculate the theoretical fit and the *residual*s using the current parameters and model, as though a fit has been completed. Those calculations will enable the display of the graphs on the Graph fit results page.

By manually altering the parameters via the Set/alter fitting parameters page and then using the **Simulate button** you can generate theoretical curves for specific values, graph those results, and if desired even export the theoretical data.

3.2.9.4 Perform fit - Cancel button

Cancel button

This will abort the current fit or single iteration, stopping any calculations that are in progress.

The **Back button** can then be used as needed to go to the Set/alter fitting parameters page to alter parameter guesses or change which parameters are being fitted, or to go to the Select fitting model page to change the fitting model before trying to fit again.

3.2.9.5 Perform fit - Status

This **status box** shows status messages as the fit progresses:

- **Waiting for input** indicates the program is halted waiting for the Do Fit, 1 Iteration, or Simulate button
- **Calculating sums** indicates the program is looping through the data points collecting the *sum of squared residuals* and partial derivatives that are needed to calculate improved parameters
- **Inverting matrix** indicates the matrix inversion needed for calculating improved parameters is being done
- **Adjusting increment** indicates that the new parameters are being tested to be sure the *sum of squared residuals* is improved. During this time the increment reduction factor may change.
- **Converged, last sum** indicates that convergence has been obtained and a final set of sums needed for calculating parameter errors is being accumulated

3.2.9.6 Fitting parameters monitor tab

This tab is used to monitor the actual values of the fitting parameters as they are updated during the fit.

Fitting parameters monitor tab

The image above shows the analysis page after the first iteration of a fit, which was initiated via the 1 iteration button.

Fit control buttons and displays - These controls are shared with the **Graphical fit monitor tab**. Follow this link for descriptions and links to further details.

Current value column - This column shows the parameter values for the current iteration of the fit. Comparisons to the **Previous value** column to the right will show whether this parameter is changing rapidly or converging to a definite value.

- Before the first iteration is completed these values will correspond to those entered or computed as the initial guesses
- If a parameter tries to go outside the current upper or lower limits the background color of its text box will flash momentarily bright yellow. It will stay yellow once that parameter reaches the limit. (Note that when parameters go out of bounds the program will switch from the **Graphical fit monitor tab** to show this **Fitting parameters monitor tab** if needed.)

Previous value column - This column shows the parameter values from the previous iteration of the fit.

- Before the first iteration is completed this column will be empty.

Chi squared or sum of squared residuals - This row displays the total *chi squared* or the *sum of squared residuals* calculated for the current and previous iteration, depending on whether an experimentally weighted, theoretically weighted, or unweighted fit was selected.

Graphical fit monitor tab - Select this tab to monitor fit overlay and residual plots as the fit progresses.

Increment reduction factor - The increment reduction factor is used within a "modified Gauss-Newton fitter" to try to ensure that each iteration actually lowers the variance. When the value falls below 1 that indicates the fitter is having difficulty finding improved parameters. It is normal, however, for this value to fall just before the fit converges because further improvement is not possible.

3.2.9.6.1 Increment reduction factor

The **increment reduction factor** is a numerical multiplier used within a so-called "modified Gauss-Newton" fitter to improve the rate of convergence [reference 2]. When this value falls below the nominal value of 1, this is an indication that the fitter either

1. is having some difficulty finding a good fit (there are too many species or very poor initial guesses), or
2. it is so near to convergence that it simply cannot find better parameters values

Show

How it works

Let {p1, p2, ...} be the set of current fitting parameters. Then at each iteration the Gauss-Newton method calculates a set of increments to those parameters, $\{\delta p1, \delta p2, ...\}$, that will define a new set of improved parameters, $\{p1+ \delta p1, p2+ \delta p2, ...\}$, which should give a lower *variance*.

In the modified method these increments are each multiplied by a scalar factor, α, the "increment reduction factor", and hence the new parameters will be $\{p1 + \alpha \times \delta p1, p2 + \alpha \times \delta p2, ...\}$.

This modification is used because when the fit is far from the optimal parameters the Gauss-Newton method tends to pick a reasonable direction for the increment vector, but not necessarily the right magnitude. Thus the idea is to test how the variance changes along this direction, and pick a value of α that will, at least, reduce the variance.

When each fit begins α is initially set to 1. At each iteration, the variance is first calculated for a new set of parameters, using an initial value of α based on that from the preceding iteration (see rule 1 below). If these new parameters give a lower variance than that for the previous parameter set, these parameters are accepted and the next iteration is begun.

If instead the new parameters give a higher variance, then the value of α is reduced by factors of two until finally a lower variance is obtained (unless the value of α falls below 0.001, which triggers exception 1 below). Once this lower variance is obtained, this variance together with those for the preceding values of α are used to fit a parabolic function of variance *versus* α and thereby to predict what value for α will give the minimum variance. This predicted value for α is then used to calculate the parameter set for the next iteration of fitting.

Although these procedures are intended to guarantee that the variance will drop at each and every iteration, occasionally that will not be true.

Rule 1 adjusts the starting value for α when α from the preceding iteration is small. In this circumstance it is hoped that this previous iteration succeeded in moving into a region of parameter space where convergence will be better, and therefore α can be increased back toward 1. Further, since it is much faster to calculate variances in order to test values for α than it is to carry out a full iteration, it is usually worthwhile to try keeping α large in order to move more rapidly to convergence. Therefore the rule applied is that when α from the preceding round is < 0.005 it is set to 0.02; for $0.005 < \alpha < 0.25$ it is multiplied by 4, and for $0.25 < \alpha < 1$ it is set to 1.

Exception 1 occurs when the sequential reductions of α during the attempt to find a lower variance reduce α below 0.001. This means the fitter is having major trouble! That may be because too many species are being included in the fitting model, or because one or more of those species are poorly resolved or in very low abundance. In that circumstance the program tries making α negative (taking the parameters in the opposite direction, since even small positive values of α just make things worse). If a negative α lowers the variance, that value is accepted; otherwise the sign of α is inverted again and that positive value accepted. If a negative value of α is used 5 iterations in a row, the program assumes convergence has been reached.

3.2.10 Report fit results page

The **Report fit results** page is displayed automatically upon completion of a fit, or you can navigate to it whenever fit results are available.

This page displays the best fit parameter values, the values for parameters that were held constant or constrained, the fit statistics, and buttons for printing a report.

Each area or control on the image below is a clickable *hotspot* that will lead to further information, or just use the numbered list below the image.

Report fit results page

The image above shows the results from a 4-species fit where the masses of species 3 and 4 have been constrained as trimer and tetramer of species 1.

Fitted parameters - This table lists the best fit parameters, the estimated uncertainty for that parameter (the standard error or true confidence limits after those have been computed), and the initial value for that parameter at the start of the fit procedure.

Fixed or constrained parameters - This table lists the values for parameters that were held constant during the fit and also any parameter constraints that were imposed.

View parameter cross correlations - This button will bring up a dialog box displaying the parameter cross-correlation matrix. See What does high parameter cross-correlation mean?.

Compute confidence intervals - This button initiates a procedure to calculate more rigorous estimates of the uncertainties in the parameter values. Generally you will only do this when you are satisfied this is your final result. (The button is disabled in the example above because this computation has already been completed).

Create report group box - Either of the two buttons will generate a report which documents every aspect of your analysis. This report is then either printed or copied to the Clipboard (in rich text format) for pasting into a word processor or electronic laboratory notebook. The **Include Graph checkbox** controls whether a summary graph showing an overlay of the data and the fit as well as a residual plot is included at the end of the report.

Graph bootstrap or Monte Carlo results - This button is active whenever confidence intervals have been computed using either the bootstrap or Monte Carlo methods (the name on the button changes to whichever method was last used). It will open a dialog box that generates graphs of the parameter distributions (either histograms of a single parameter, or two-parameter correlation plots).

Fit statistics - This area displays some of the statistical parameters characterizing the fit.

High cross-correlation warning - This warning panel is only displayed when there are one or more pairs of parameters that are highly correlated (there is 1 in this example). This is a warning that some of the parameters may be poorly determined, and you may want to consider fixing or constraining some of the parameters.

3.2.10.1 Report fit results - Fitted parameters

Fitted parameters

This area of the Report fit results page displays:

1. the returned best-fit values for the fitted parameters.
    - Note that the units applicable to the fitting parameters are shown; these will depend on other program choices and options, such as whether the *g(s*)* distribution has been normalized.
2. an estimate of the error (uncertainty) for each parameter
    - When the fit is first completed that error estimate will be a simple estimate of the standard error based on the covariance matrix
    - After the Compute confidence intervals button has been used, this area will instead show the true confidence intervals, with the confidence level shown in the column heading.
3. The starting value for each parameter (the initial guess at the start of that fitting session).
    - This is provided primarily to document the initial conditions of the fit.
    - Note that fits saved from version 1.xx of this program did not store that information, so this column will not be displayed.

3.2.10.2 Report fit results - Fixed or constrained parameters

Fixed or constrained parameters

This table lists:

1. The values for parameters that were held constant during the fit (parameters whose values were not optimized during the fit). In this example the baseline offset has been held fixed at zero.
2. Equations summarizing any parameter constraints that were imposed. In this example the mass of species 3 is constrained to be 3 times the mass of species 1 (species 3 is a trimer of species 1), and species 4 is constrained to be a tetramer of species 1.

3.2.10.3 Report fit results - View parameter cross-correlations button

This button will bring up a window showing the cross-correlation matrix. Parameters with 'high' or 'very high' correlations will be flagged with yellow or red backgrounds. The graphic below shows this window.

Correlation matrix display

In this particular example there is a high (in absolute value) negative correlation between the concentration and mass parameters for species 2. The negative correlation between the concentration and sedimentation coefficients of species 1 with the concentration of species 2 are also quite strong (-0.9363 and -0.9362), although not quite strong enough to flag in this case.

For this sample species 2 (a dimer of an antibody) is present at a level of only 2%, and there is simply insufficient information in the data to determine the concentration, sedimentation coefficient, and mass accurately for species 2. This is a case where it would be better to constrain the mass of species 2, which will greatly improve the precision of the estimates for the concentration and sedimentation coefficient of not only species 2 but also species 1.

3.2.10.4 Report fit results - Compute confidence intervals button

Compute confidence intervals button

This button will initiate computation of confidence intervals for the parameters, a much more statistically rigorous way to express the uncertainties in the best fit parameters (but time consuming to calculate).

TIP: Don't bother to compute confidence intervals until you are sure this is your final fit.

You are first presented with a dialog box for setting the desired confidence level, and selecting a computation method, as shown below. The default level of confidence that will be shown in this dialog (68.3%, 95%, *etc.*) can be altered via the Options dialog. The Ok button initiates the confidence level calculations.

Set confidence level dialog

The confidence limits are often not symmetric around the best fit value, and thus true confidence limits usually cannot be reported as a ± value.

Selecting a computation method:

Three radio buttons allow the user to select a preferred mathematical approach for computing the confidence intervals. These are:

Search error ellipse: This method moves each parameter away from its best-fit value until the *chi squared* increases to a limit that is bases on *F* statistics (it evaluates the shape of the error surface near the minimum). The implementation is based on reference 2. The search of parameter space is made along axes of a hyperellipse that represent eigenvectors of the covariance matrix. The calculation of the critical *chi squared* level is based on equation 35 in Johnson & Straume (1994), reference 12.

This approach is relatively fast to calculate. It's major drawback is that it is based on certain statistical models and assumptions, including an assumption that all the data points are statistically independent. Further, there is some disagreement among different sources about exactly how the critical *chi squared* level should be calculated.

Bootstrap with replacement: In this method data points are selected randomly from the original fitted data set to make a new data set with the same number of points (but due to the random selection some data points will be selected more than once). This new data set is then fitted, those parameters are stored, and the process is repeated many times to build up a distribution of parameter values. The confidence limits are then calculated from that parameter distribution (based on options as described below).

The up/down spinner next to this radio button sets the number of bootstrap rounds (default 500). See About the bootstrap method for more details.

Monte Carlo: In the Monte Carlo method random noise is added to a set of theoretical scans that are calculated based on the current best-fit parameters. Those simulated scans (with noise) are then used to calculate a new average *dc/dt* and *g(s*)* curves and then a new data set for fitting. This new data set is then fitted, those parameters are stored, and the process is repeated many times to build up a distribution of parameter values. The confidence limits are then calculated from that parameter distribution (based on options as described below).

The up/down spinner next to this radio button sets the number of Monte Carlo rounds (default 500).

The rms noise level to add to the raw scans can be entered manually in the text box below this radio button. If you are fitting simulated data you should enter the actual value used in the simulation. When you first bring up this dialog after a new fit an estimate of the noise in the raw data is made and automatically entered into this text box. If you later want to re-calculate and re-enter this estimate use the **Recalc button**. See About the Monte Carlo method for more details about how this estimate is made.

Finding the confidence limits from the bootstrap or Monte Carlo results

Once a table of parameter values is built from the *N* rounds of bootstrap or Monte Carlo, there are two choices for calculating the parameter confidence intervals:

Computed standard deviations of the parameter distribution: With this choice first for each parameter the mean and standard deviation of the set of *N* values for that parameter is calculated, and then the confidence limit is found assuming that distribution is random (a Gaussian distribution). This option is generally the best choice when the number of rounds is 200 or less.

Because in reality for non-linear fits the parameter distribution may be asymmetric around the mean, two standard deviations are calculated, one for those parameters that exceed the mean value, and another for those that are less than the mean value. Those two standard deviations are then used to calculate an upper and lower confidence limit, respectively, and those values are not in general symmetric about the mean.

- It is also sometimes true that the mean parameter value from the bootstrap or Monte Carlo rounds is significantly shifted away from the corresponding best-fit parameter. When this is true (specifically when the shift is greater than 20% of the standard deviation) the confidence limits are also shifted (to make the mean match the best-fit value).

Observed parameter frequencies: With this choice no statistical assumptions are made about the shape of the parameter distribution. Instead, the *N* parameter values are first sorted and then to find the confidence limit the list is searched inward from the beginning or end (the lowest or highest observed values) until the appropriate confidence frequency has been exceeded. For example, for 95% confidence probability and *N* = 1000, the confidence limits correspond to frequencies of 2.5% (giving 5% total for the upper and lower wings of the distribution), which is 25 data points out of 1000. Therefore the confidence limits are the 25th and 975th values in the sorted list.

The drawback of this approach is to accurately find the desired confidence limit a large number of rounds, *N*, is required. If, for example, *N* is only 100, then it is not possible to distinguish a parameter frequency of 2.5% (the desired value for the lower wing for 95% confidence) from a frequency of 2% or 3%. That is, if you wanted the confidence probability to be accurate to 0.1%, it is necessary to have *N* >= 2000.

What happens after I have selected these options?

The **Ok button** will proceed to actually computing the confidence, and **Cancel** will abort that process. During the subsequent computations a small dialog box is displayed.

If you have selected the **Search error ellipse** option you will see a dialog box like that below, showing which confidence limit is currently being evaluated, an estimate of remaining time, and a progress bar.

Confidence limit calculation progress dialog

For either the **Bootstrap** or **Monte Carlo** options the dialog box shown below which indicates the current simulation round, an estimate of remaining time, and a progress bar.

Display during bootstrap or Monte Carlo

The **Cancel** button on these dialog boxes will abort the calculations (leaving the previous confidence limits intact if they exist).

What if I want to re-compute the confidence limits with different choices?

You can always change your selection and re-compute with a different confidence level or different computation method simply by pushing the **Compute confidence button** again.

3.2.10.5 Report fit results - Create reports group box

Print reports buttons

Either of the two buttons within this group box will generate a complete, nicely-formatted report defining what data were fitted, how they were fitted, the best fit parameters, fit statistics, etc. That report is then either printed or copied to the Clipboard (in rich text format) so that it can be pasted into a word processor or electronic lab notebook.

This report is intended to document every aspect of your analysis. All user choices and program defaults that affect the calculations of the data or the fitting procedure are listed in the report.

The **Include Graph checkbox** governs whether a summary graph showing an overlay of the data and the fit, with a residual plot underneath (as shown on the Graphical fit monitor tab of the Perform least-squares fit page) will be included at the end of the report.

As an aid to judging the quality of the fit, the program will translate the *rms residual* from the fit into an equivalent noise level for the original raw scans and include that value in the report.

- This calculation is only accurate to ~20%. For *dc*/*d*t ftting this calculation is done by multiplying the observed r.m.s. residual by the mean time interval between scan pairs and the square root of half the number of pairs. For *g(s*)* fitting it is done by computing the root mean square value of the product of the residual for each point and the corresponding sedimentation coefficient, and then multiplying this by the mean time interval between scan pairs times the square root of half the number of pairs and then dividing this result by the mean effective sedimentation time for the scans.

If the **Print report** button is selected then a standard Windows printer dialog box will open allowing selection of the printer to be used for the report (this will by default be the printer set as the default report printer via the Options dialog. The graph will be in color if that printer is a color printer.

The size of the text font used for reports (whether printed or sent to the Clipboard) may be altered using the Options dialog.

Caution: Although the **Send report to Clipboard** button does not actually generated printed output, when a graph is to be included the program must have access to a valid printer in order to size the graph correctly relative to the text. If the default report printer is not valid either because it is not currently connected or no printer with that name exists, you will get a message asking for a valid printer (which could be a PDF printer or other virtual printer if you aren't currently connected to a real printer). Also note that the graph produced by this rich-text copy operation will be in color whether or not this printer is actually capable of color.

3.2.10.6 Report fit results - Fit statistics

Fit statistics

This area of the Report fit results page will report a statistical summary of the fit.

For **experimentally weighted** fits the values for *chi squared*, the *degrees of freedom*, the *rms residual*, and the *reduced chi squared* will be displayed.

For **unweighted** or **theoretically weighted** fits (as shown here) the *sum of squared residuals*, the *degrees of* *freedom*, and the *rms residual* will be displayed.

3.2.10.7 Report fit results - High cross-correlation warning

High cross-correlation warning

This area of the Report fit results page is used to display a warning whenever some of the parameters are highly cross-correlated. See What does high parameter cross-correlation mean?.

There are two levels of warning, one for "high" correlation and one for "very high" correlation. The numbers of correlations exceeding each level is listed, unless there are none at that level.

- The default levels that trigger those warnings are 0.97 and 0.99, respectively, but these levels can be set within the Fitting Options tab of the Options dialog.
- These warnings will also appear in printed reports.

The actual cross-correlation matrix can be viewed by using the view parameter cross-correlations button on this page.

3.2.10.8 Report fit results - Graph bootstrap or Monte Carlo results

Graph bootstrap or Monte Carlo results | Graph bootstrap results

These buttons become active whenever confidence limits have been calculated using either the Monte Carlo or bootstrap methods, respectively (the button name changes to match the method that was last used). Pressing the button will bring up a dialog box that displays graphs of the parameter distributions (using the table of parameter values from each round of bootstrap or Monte Carlo).

Initially this dialog box will display a histogram of the mass values for species 1 (assuming that parameter was fitted), like that shown below. The vertical red line marks the best-fit value. The vertical green lines on the histogram show the ±1σ and ±2σ limits (as calculated separately for the values above and below the mean). The magenta curve shows the cumulative frequency (the area under the distribution curve).

parameter distribution histogram

Corresponding graphs for the other fitting parameters can be obtained using the drop-down list labeled 'Parameter 1'.

Note that this example shows the results from a rather large number of bootstrap rounds (2000) and consequently the width of the bins in the histogram can be small, and the overall shape of the distribution is fairly well-defined. With lower numbers of bootstrap or Monte Carlo rounds the width of the bins must be larger in order to have enough events in each bin to give a statistically meaningful result. The program automatically adjusts the number of bins displayed to try to account for this statistical issue.

TIP: The number of bins used for generating the histogram can be altered by clicking on any of the blue bars with the mouse. Using fewer bars will tend to average out the statistical fluctuations, but at a cost of fewer positions to define the true shape of the distribution.

Parameter cross-correlation graphs

Clicking the **Two parameter cross-correlation plot radio button** will generate a scatter plot like that shown below, where each point represents a pair of parameter values for each round of bootstrap or Monte Carlo simulation. The vertical and horizontal red lines mark the best-fit value for that parameter.

Parameter cross-correlation plot

For this example the data points are clearly clustered in an elliptically-shaped region tilted roughly 30 degrees to the left of vertical. This pattern shows that the mass values are moderately highly correlated with the concentration values (a negative correlation in this case). As the mass value is increased or reduced from the best-fit value it is possible to partially compensate for that change, and minimize the increase in the *variance* of the fit, by simultaneously decreasing the concentration parameter. That is, the two parameters cannot be determined completely independently.

If the points are very tightly clustered along a line (corresponding to a correlation coefficient near +1 or -1) this is an indication that too many parameters are being fitted (the information content of the data is not sufficient to determine all the parameters independently). In some cases the non-linearity of the fitting may produce clustering along a curve rather than a line. In any situation where the distribution is very tightly clustered along a diagonal line or a curve, the fitting results are probably not trustworthy and it may be appropriate to reduce the number of parameters by introducing parameter constraints, or by reducing the number of species.

If these parameters were not significantly correlated then the points would cluster along an ellipse that is oriented parallel to either the vertical or horizontal axis. (If the distribution was truly random the shape of the pattern would in principle be circular around the best-fit values if the axes were scaled in units of standard deviations, but in general the graph scaling will produce an ellipse rather than a circle).

This graph provides information that is quite similar to the cross-correlation coefficients that can be viewed using the View parameter cross-correlations button. However those values are based on a linear statistical model and assume the noise in the fitted data is random. The cross-correlation information obtained from these plots of bootstrap or Monte Carlo results is independent of such statistical assumptions.

Different pairs of parameters can be selected using the two drop-down list boxes. (Obviously when the same parameter is chosen for both axes the result will be a diagonal line.) The size of the data points on this graph can be altered by clicking on any of them with the mouse; the default size is automatically reduced as the number of data points increases to try to avoid too much overlap on the screen, but this choice may not be optimum when the graph is exported or printed.

Other features of this dialog box

The **Copy button** will place a copy of the current graph onto the Clipboard (in Windows enhanced *metafile* format).

The **Print button** will send this graph to the printer.

The **Add to log button** will add the current graph at the end of the Log page.

The **Readout data values checkbox** will turn on a readout of the graph coordinates corresponding to the current mouse position. This could for example be used to find the lower and upper 90% confidence limits for the actual distribution (without assuming a Gaussian shape) by locating the mass at the points where the cumulative frequency (magenta line) reaches 5% and 95%, respectively (giving the desired 10% probability that the true value lies outside the confidence interval).

3.2.11 Graph fit results page

This analysis page displays nine different types of graph of the fit results, fitted data, or theoretical fitted curves. The type of graph is selected from a drop-down list at the top of the page.

Also located at the top of the page are convenient buttons for printing the graph (**Print** button), copying the graph image to the Windows Clipboard as a high-resolution *metafile* so it can be pasted into another document (**Copy** button), and adding the graph image at the end of the Log page (**Add to Log** button).

The drop-down sections below show examples of each type of graph (all from the same analysis of an antibody sample), with brief comments about the intended use of that plot type. Note that although these descriptions refer to the fitted data, if you have reached this page after doing a simulation rather than a fit then the plots show the results of the simulation.

Show All Show All

Hide All Hide All

Show 1. Species plot for entire data region (area)

This type of plot shows each fitted species as an area graph, and extends the plot over the entire sedimentation coefficient range (not just the range that is fitted).

In this example below the area plot of the fit to the antibody monomer region highlights the fact that this one species does not fit the high sedimentation coefficient region (due to the presence of aggregates). The area to the right of the single species peak, between the theoretical fit and the experimental data, is directly a measure of the aggregate content.

Species plot for entire data range (area plot)

It is also apparent that there is material sedimenting in the region from 0.5 to 3.5 S that is not accounted for by this single-species fit (this sample contains both antibody fragments and detergent micelles that sediment in that region).

Because of the presence of those other species, for this fit the fitting region was limited to sedimentation coefficient between 3 S and 6.88 S (the latter is about midway down the right side of the peak). That was done deliberately in order to get a good fit to the monomer properties by excluding regions strongly influenced by these other species. By extending the plot outside the fitted region we can now visualize the fact that this single species does not provide a good fit outside that fitted region, and also get a good estimate of the aggregate content. If instead the aggregate region had been included in the fit, the fitter would broaden the single species peak (lower estimated mass) in order to reduce the residuals in the aggregate region, and we would actually get a much poorer estimate of the aggregate content as well as poor values for the properties of the main peak.

This graph displays a small title line that lists the first and last scan files included in the analysis, and the folder from which they were loaded.

Show 2. Species plot for entire data range (line)

This plot is also a species plot for the entire data range, like #1, but the individual species are shown as simple line plots rather than filled areas. While the area plots are pretty, when you are trying to look closely at whether the fit really follows the data well, the area plots can sometimes obscure the experimental data, so this type of plot is often better for that purpose.

Species plot for entire data range (line)

See the discussion under plot #1 above about why it is often useful to limit the range of data that is fitted but to extend the theoretical plots over the entire data range.

Show 3. Species plot for fitted range (area)

This plot is like #1 above but is limited to the sedimentation coefficient range that was actually fitted.

Species plot for fitted range (area)

As noted under #1 above, in this example the fitting range was deliberately limited to the range from 3 to 6.88 S to avoid fitting data that is strongly influenced by the presence of other species. This allows obtaining accurate properties for the major species (antibody monomer) without having to accurately model the other species.

Show

4. Species plot for fitted range (line)

This plot is like #2 above but is limited to the sedimentation coefficient range that was actually fitted.

Species plot for fitted range (line)

Show

5. Overlay of data and fitted curve

This plot overlays the data points that were fitted (shown as circles) with the best-fit curve (as a solid line). This is a good basic summary of the fit and whether it described the data well, without actually plotting the contributions from each individual species.

Overlay of data and fitted curve

The number of species and the type of data being fitted are indicated in the graph title line.

This plot is essentially the upper panel of the plot shown on the Graphical fit monitor tab of the Perform least-squares fit page.

Show

6. Summary plot - overlay + residuals

This plot creates a summary of the fit by combining an overlay plot (#5 above) with a residual plot (#7 below) into a single graph object with two sets of Y axes. This plot is the same as the plot shown on the Graphical fit monitor tab of the Perform least-squares fit page and the one that is used in printed reports and automatically added to the Log page.

Summary plot - overlay + residuals

The relative vertical size of the overlay and residual plot portions can be altered using the mouse. When the mouse hovers over the region in between these plots the cursor changes to a vertical arrow and the division point between the graphs can be moved by clicking and dragging up or down with the mouse.

Show

7. Residual plot

This residual plot is functionally equivalent to, but formatted differently than, the one that is shown in the lower panel of the **Summary** plot (#6 above) and also on the Graphical fit monitor tab of the Perform least-squares fit page.

Residual plot

The value of the *rms residual* for this fit is displayed as the plot title.

Note that if you prefer alternate plotting methods, such as plots with points connected by lines, or stick plots, the plotting method can be quickly changed by right-clicking on the graph and using the Plotting Method sub-menu on the pop-up graph context menu.

Show

8. Fitted function

This plot simply shows the best-fit curve for the data region that was actually fitted.

Fitted function plot

Show

9. Data being fitted

This plot shows only the actual experimental data that were fitted.

Data fitted plot

4 How to...

4.1 How to avoid common mistakes

Problems with strange results or bad fits usually arise either from a violation of the assumptions that are inherent in this type of analysis, or instrumental problems. Below is a list of mistakes and pitfalls to avoid.

Show All Show All

Hide All Hide All

Show 1. Wrong assumptions

It is essential to keep in mind that fitting the *g(s*)* or *dc/dt* data to a sum of species requires certain assumptions about the nature of the sample and the way in which the experiment is conducted:

1. the sample has no significant concentration dependence of *s* or *D* values
2. if there are multiple species, they do not interact or inter-convert (or if they do inter-convert that process occurs slowly compared to the time of the velocity run)
3. solution non-ideality is minimal
4. the rotor is directly and rapidly accelerated to run speed
    - With regard to the last assumption regarding rotor acceleration, it is not possible to obtain correct diffusion coefficients or molecular masses for experiments where the rotor speed is altered after significant sedimentation has occurred (although the sedimentation coefficients would still be correct). Thus it is not a good idea to start the rotor spinning and then wait for the temperature to equilibrate. See How (and why) to equilibrate the temperature.
    - However, in most cases it is quite acceptable to pause at low speed (~3K) to take a quick scan to check that the samples are loaded properly. The exception to this would be for samples with very high *s* values that will sediment significantly even at low speed.

Show 2. Influence of cell base

One common problem in *dc/dt* analysis is failure to properly exclude the region at the cell base where solutes are accumulating.

- Optimally this region should be excluded at the time the data acquisition is set up and therefore it never appears in the data. Proper adjustment of the Good data end at radius = setting should take care of this problem. However, for low mass solutes or different cell types this default may not exclude all the bad data, resulting in negative *dc/dt* and *g(s*)* values at the high sedimentation coefficient region. See Good data end at radius = for further discussion and an example graph.

Show 3. Insufficient sample volume

A common cause of poor resolution of multiple species in an analysis is loading too little sample into the cell (too little column height). A longer sedimentation distance improves the physical separation between species having different sedimentation coefficients, and also reduces the relative influence of diffusion on each species. It is recommended that you load enough sample to give a meniscus position <= 6.0 cm (about 430 microliters of sample in a 12 mm charcoal-epon cell).

Show

4. Cell leakage

Another common cause of bad fits and strange results is slow sample leakage, which is easily detected by shifts of the meniscus position to higher radii over the course of the run. Slow sample leakage usually results from evaporation via a leak at the red cell plug gasket rather than a liquid leak through the window-centerpiece interface.

Show

5. Wavelength shifts

With absorbance data you really cannot take data at different wavelengths during the same run.

- Because the monochromator never returns to exactly the same wavelength, if you use different wavelengths then the apparent concentration will vary from scan to scan (because the extinction coefficient will be slightly different as the wavelength varies +/- 1-2 nm from scan to scan). This effect can most easily be seen as a jumping up and down of the plateau region from scan to scan. If you need data at different wavelengths, it is highly recommended that you do each wavelength as a separate run.

Show

6. Drifting sample temperature

Another instrumental problem that can cause bad fits and poor or inconsistent results is changes in sample temperature during the run.

- It is important to thoroughly equilibrate the rotor temperature before starting the run (and remember that you are not really reading the rotor temperature until the vacuum drops below 100 microns!). See How (and why) to equilibrate the temperature.
- You can check the temperature stability for your run by clicking the **T button** within the Data displayed options box on the Set meniscus & data region page to display a graph of the recorded temperature versus elapsed time (this requires being in Advanced mode).

Show

7. Interference from buffer components

Especially for interference scans, you may see systematic deviations in the *g(s*)* curves at low sedimentation coefficient values resulting from redistribution of salts or other low molecular weight components in the cell region near the meniscus.

- To minimize these problems, the sample and reference column heights can be matched exactly if you use a synthetic-boundary centerpiece (or better still the Walter Stafford meniscus-matching centerpiece design available through Spin Analytical) and follow this procedure:
    1. Fill the cell as you would a conventional cell.
    2. Bring the rotor up to about 10K rpm to force sufficient solvent to cross through the capillary to equalize the volumes in both channels (you can see this happen on the Raptor fringe display).
    3. Stop the run, remove the rotor, and turn it over several times to mix the sample channel. This procedure gives a slight dilution of the sample.
    4. Reload the rotor, and re-equilibrate the temperature. When you come up to running speed the reference and sample meniscus will be exactly matched.

4.2 How to optimize your experiments

4.2.1 How to optimize data acquisition for g(s*) analysis

There are a number of things that can be done to optimize the experimental setup to improve the results of time derivative analysis. This section will first discuss what factors limit the signal/noise, and then describe specific things for optimizing either absorbance or interference data.

Show All Show All

Hide All Hide All

Show What factors govern the signal/noise of a *dc*/*dt* analysis?

A common misconception is that using scans taken very closely together in time will limit the signal/noise because the differences between successive scans are smaller. In fact, the faster scans are acquired, the better the signal/noise ratio.

It is important to remember that the algorithm does not subtract successive scans. Rather, the group of scans is broken into two halves and the corresponding members of each half are subtracted. Thus if there are 10 scans, scan 6 is subtracted from scan 1, scan 7 from scan 2, *etc*.

The resulting concentration differences, Δc, are divided by the time differences between the scan pairs, Δt, to give $\Delta c/\Delta t$ which will closely approximate the true derivative dc/dt provided there is not excessive boundary movement between the two scans being subtracted.

For a given level of instrument noise, then, the signal/noise for any particular $\Delta c/\Delta t$ computation will grow approximately linearly with time. What ultimately limits this increase in signal/noise is the need to keep the boundary movement from becoming excessive, causing broadening of the peaks in the distribution and loss of resolution between species. The maximum time span that can be covered without broadening is related to the mass(es) of the solute(s), as discussed under peak broadening limit This leads us to Rule 1.

Rule 1: the first thing to do to optimize signal/noise is to use the largest time span possible between the first and last scan in the analysis, while staying within the restrictions necessary to avoid peak broadening.

Once this criterion is met, the overall signal/noise is improved by including and averaging more scans within this maximum time span (and will improve roughly as the square root of the number of scans assuming the noise is random). This leads to Rule 2.

Rule 2: scan as fast as possible so more scans can be included in the analysis.

For both absorbance and interference data the time to take a scan is roughly proportional to the radial range being scanned. A common mistake is to leave the radial range set at the default values of 5.8 to 7.3 cm. This results in a significant fraction of the instrument time being spent taking data above the meniscus and beyond the base of the cell. Remember, too, that you don't want to include the region near the base of the cell where solutes are accumulating.

Corollary to Rule 2: adjust the regions being scanned so you don't waste time taking data for regions of the cell that you can't include in the analysis.

Show How to speed up data acquisition for interference data

While interference scanning is inherently much faster than absorbance, there are still things that can be done to maximize the scan rate, and these may be particularly important if your instrument control computer has a slow processor and you are scanning many cells or trying to measure samples at very low concentrations.

The display of the scans within the data acquisition program (the "GUI") takes a significant amount of time, particularly if you set N > 1 on the 'Overlay last N scans' option (under the Options button).

The absolute maximum scan rate is obtained by setting N = 0 on the 'Overlay last N scans' option, which results in no display of the new data. The drawback is, of course, that you cannot as easily monitor the course of the run. (You can however monitor the experiment by displaying scans using some other program, such as the plotting commands in the Beckman software or John Philo's public domain program XLGraph.)

The extent to which scan rate is affected by the graphing of the incoming data depends on the computational and graphics speed of your computer. These settings are more important for slower computers, and are now much less important because computers have gotten so much faster since the XL-I was first released in 1995.

Show

How to speed up data acquisition for absorbance data

The scan rate for absorbance data is more complex than for interference data. It depends on a number of instrument settings, including the scan mode (step or continuous), radial step size, and number of replicates.

Below is a detailed discussion of the issues involved, but the bottom line is that the default settings of continuous mode, .003 step size, and 1 replicate work well in most situations, so generally it is probably best to just stick with the default.

One common misconception is that the scan rate is inversely proportional to the radial step size. It is also important to realize that the radial step size settings are only approximate. The actual steps vary from position to position and scan to scan, and even the overall average step size does not necessarily match this setting.

Another confusing issue is whether or not to ask for replicates. From strictly a signal/noise perspective, it can be advantageous to use replicates, as discussed in some detail below. However, **the use of replicates with continuous mode is not recommended**. In continuous scan mode the slit drive motor moves continuously. If replicates are used in this mode, a single average radius position is assigned to each averaged data point, and that radius assignment assumes the motion is uniform, which is probably not true. Thus the use of replicates with continuous scans can compromise the precision of the radius values, which in turn compromises the precision of the sedimentation coefficients, which is one of the key strengths of sedimentation velocity.

Show

Signal/noise considerations for different absorbance scan settings

Is it better to have more data points with a higher noise level, or fewer with a lower noise level? In theory, the "value" of each data point is inversely proportional to the square of its uncertainty. For the moment, let us assume that the number of data points per scan is independent of whether we use replicates. If we ask for four replicates, and actually obtain the two-fold lower noise level for those points expected by theory, then each of those points is "worth" four times as much as a data point with 1 replicate. Thus if our scan with four replicates takes more than four times as long to produce, we are worse off using replicates; if our scan with four replicates takes less than four times as long, we are better off using replicates, even though we will have fewer scans to use in the analysis.

In practice the signal/noise issue with replicates is even more complex because asking for replicates also affects the number of data points produced per scan. In theory our ability to define the properties of individual species by fitting is proportional to the total number of data points. If we make the number of data points in the *dc/dt* and *g(s*)* curves approximately match the number of data points in the scans (which corresponds to the default for the ds setting in this program), then it really doesn't matter if a scan takes twice as long provided it produces twice as many data points with an equivalent noise level.

Overall, then, what really counts as a "figure of merit" for signal/noise is the product of the data rate (points recorded in the scan file per sec) times the value factor for those points (inverse square of their noise level). Again, assuming the replicates actual reduce the noise level by the square root of the number of replicates, then our "figure of merit" becomes (data rate) * (replicates), and the relative signal/noise for different conditions should scale as the square root of this figure of merit.

The table below shows some results obtained by John Philo. All scans covered the region from 5.95 to 6.15 cm. A series of 8 scans were done to get the average time per scan. The scan files were examined to determine the typical actual number of data points recorded (this varies from scan to scan), and from this the average step size was computed. The "figure of merit" (data points * replicates per sec) was then calculated, and the r.m.s. noise level over the region from 6.0 to 7.0 cm was determined (averaged over 4 of the 8 scans).

Mode	Radial Step Size (cm)	Replicates	Seconds per Scan	Actual # of Data Points	Average Step Size (cm)	Figure of Merit (data rate * replicates)	Measured Noise (OD r.m.s.)
continuous	.003	1	84	566	.0021	6.7	.00420
continuous	.005	1	64	363	.0033	5.7	.00424
continuous	.010	1	55	184	.0065	2.1	.00399
continuous	.003	4	**157**	758	.0016	19.4	.00213
continuous	.005	4	**157**	512	.0023	13.1	.00212
continuous	.010	4	**157**	238	.0050	6.1	.00202
step	.003	1	96	380	.0032	4.0	.00386
step	.005	1	73	239	.0050	3.3	.00396
step	.010	1	59	121	.010	2.1	.00383
step	.003	4	183	337	.0036	7.4	.00219
step	.005	4	109	227	.0053	8.3	.00207
step	.010	4	77	120	.010	6.2	.00202

Some observations from the above results are:

- Increasing the radial step size above .003 does not reduce the time per scan proportionally to the number of data points produced, and always results in a lower figure of merit
- For the same nominal radial step size, step mode always produces fewer data points than continuous mode (*i.e.* continuous mode produces smaller step sizes than the value requested)
- With 1 replicate, step mode always takes longer (and also produces fewer data points) than continuous mode
- Using 4 replicates actually does lower the noise level by very close to the expected factor of 2 (slightly less than a factor of 2 for the continuous mode)
- Using 4 replicates never increases the time per scan by as much as a factor of 4 (and often by less than a factor of 2).
- With 4 replicates in continuous mode, increasing the step size above .003 did not decrease the time per scan!

Show So is there really one "optimum" absorbance scan setup for *g(s*)* analysis?

Not necessarily! Certainly the conventional wisdom that continuous mode is faster than step mode generally holds true.

If we strictly pay attention to the theoretical "figure of merit" then the conclusion would be that continuous mode, .003 cm step size, and 4 replicates is best (at least among the conditions tried here). However this ignores the fact that using replicates in continuous mode may cause errors in radial position and compromise the accuracy of the sedimentation coefficients.

Further, the "figure of merit" criterion also assumes the overall noise in the *g(s*)* distribution is dominated by random, statistical noise in the absorbance measurement. In practice, the variations in *dc/dt* from different scan pairs seem to be larger than expected due to random deviations, implying the existence of some source of systematic noise.

A direct comparison was made of using 4 replicates versus 1, with continuous mode and a .003 step size, on samples with good signal/noise (OD of ~0.8). The number of scans used in each analysis was chosen to cover a nearly equivalent time span, so the degree of band broadening was kept equivalent. Although in theory the experiment using 4 replicates should have given a lower noise level for the *g(s*)* distribution (as well as more data points in that distribution), the actual result was that the error band on the distribution was as good or slightly better with 1 replicate. Similarly, the uncertainty in *s* and *D* values obtained by fitting the peak were as good, and perhaps slightly better, for the experiment using 1 replicate.

Furthermore, the setup using 4 replicates requires > 2.5 minutes per scan (~8 minutes for 3 cells), and such long scan times have drawbacks that may override signal/noise considerations. The consequence of having a low scan rate is that the total number of scan pairs in the analysis will usually be quite small. Therefore moving the start of the analysis forward or back by a single scan will shift the range of sedimentation coefficients by a substantial amount, and adding or subtracting even one scan pair may have a large effect on peak broadening and signal/noise.

Overall the default velocity-mode setting of continuous mode, 0.003 cm spacing, and 1 replicate is probably best in most situations.

In those rare situations where the random optical signal noise is dominant over the systematic noise (very low or very high absorbance, or when working at a wavelength with low light intensity), and where the low signal/noise means the sedimentation coefficients will not be highly accurate anyway, the "best" setup should more closely follow the "figure of merit" criterion and it therefore will probably be best to use replicates.

4.2.2 How (and why) to equilibrate the rotor temperature

Temperature is the key to obtaining reproducible and accurate sedimentation coefficients!

The "raw" sedimentation coefficient will typically increase by 2.4% per degree (due primarily to the large temperature dependence of the viscosity of water). Since sedimentation coefficients can be measured with a precision of ±0.2% or better, that means temperature control to ±0.1 degree is necessary to avoid compromising the accuracy of the *s(20,w)* values.

You can check the temperature stability for your experiment by displaying a graph of the recorded temperature versus elapsed time (covering all the scans currently in memory) by clicking the **T button** within the Data displayed options box on the Set meniscus & data region page (this requires being in Advanced mode).

How can I tell if the temperature was stable during my run?

A plot of the temperatures recorded during the run can be obtained from the Set meniscus & data region page (the "T" button, available in Advanced mode).

What is required for good temperature control?

1. Sufficient time to achieve temperature equilibration
2. Vacuum below 5 microns to avoid frictional heating of the rotor

Reduce the temperature equilibration time by putting the rotor and monochromator into the chamber to equilibrate, under vacuum, while you are cleaning, assembling and filling cells.

However, the wrong method of equilibrating temperature will also cause problems!

Many people have been taught to bring the rotor to 3000 rpm and wait for temperature equilibrium before going to the final rotor speed. This approach will cause problems for nearly all analysis methods that attempt to derive diffusion coefficients or masses from velocity data (including this one)!

- To handle diffusion these methods must assume instantaneous acceleration to the final rotor speed. If any significant amount of sedimentation occurs at low speed, the derived diffusion coefficients and masses will be compromised (but a quick scan at 3000 rpm to check for proper cell loading should not be a problem for species <~ 10 MDa)

The trick to proper temperature equilibration: "start" a run at zero rotor speed

This will turn on the diffusion pump without spinning the rotor.

- Remember, the instrument only turns on the infrared thermometer to read the actual rotor temperature when the vacuum drops below 100 microns, so it is essential to get the diffusion pump working to get proper temperature equilibration.
- Setting the speed to zero can only be done from the instrument control panel, not from the computer through the control software

1. first push the 'Speed' button then push '0' and 'Enter'
2. then hit 'Enter', then 'Start'

Caution! This trick creates a small problem:

If you need to release the vacuum when "running" at a speed of zero the instrument fails to force you to wait for the diffusion pump to cool (a bug in the instrument firmware). If you need to release the vacuum, hit 'Stop' and then wait ~5 minutes for the diffusion pump to cool before you push the 'Vacuum' button. Failure to cool the diffusion pump may get pump oil on your optics, including the diffraction grating in the monochromator (~$20,000 to replace).

4.3 How to work with graphs

4.3.1 How to print graphs

All graphs can be quickly printed by right-clicking on the graph to bring up this context menu and selecting the bottom 'Print Graph to Default Printer' entry.

Graph context menu

That choice will print a full-page graph in landscape orientation to the current default printer (the Windows default printer, not the printer specified for analysis reports from the Options dialog).

If you want to print to a specific printer, print a graph of a specific size, or control whether the graph is printed with color or in monochrome, you can do this by selecting the **Export Dialog...** menu entry.

4.3.2 How to customize graphs

'Click and edit' customizations

The axes, axis titles, and graph titles are hotspots. When passing over these hot spots the cursor changes into a hand with pointing finger; single clicking will bring up a small dialog box for editing and customizing those features.

Axis scaling dialog

Clicking on an axis brings up the dialog shown above, which allows alterations of the scale range (minimum and maximum), the intervals between major and minor tick marks, and whether minor ticks are displayed (note that the graphs generated by DCDT+ normally do not show minor tick marks). If the graph is currently zoomed then this dialog will change the axis of the zoomed graph (and won't alter what is displayed when the graph is un-zoomed with the 'z' key).

The major tick marks (the ones which can have their values labeled) are always positioned starting from zero, not necessarily from the start of the axis. Thus in the example above, even though the X axis of this raw data plot starts at 5.9 cm, because that value is not divisible by 0.2 the labeled major tick marks will be at 6.0, 6.2, 6.4... rather than at 5.9, 6.1, 6.3...

TIP: Sometimes a tick label which should be present at the beginning or end of an axis does not display due to round-off error. When this happens you can get the label to display by increasing the axis maximum (or decreasing the minimum) a tiny amount (perhaps 1 part in 100,000), which makes that tick mark clearly within the bounds of the graph. Doing this has no visible effect on the graph scaling but solves the tick label display issue.

Clicking on an axis title or the graph title or subtitle (both located above the graph) will bring up small text entry dialog box that allows editing of the corresponding axis label, similar to that shown below.

Title edit dialog

Note that a completely empty entry is interpreted as the 'Cancel' command, so if you want a blank label simply enter a space character. These graph title and sub-title can also be changed through the Customization Dialog discussed below.

Right-click context menu customizations

Right clicking with the mouse on any graph will bring up a context menu that provides access to numerous graph options and customization features.

The **Display Mouse Coordinates** item toggles on and off a display of the graph coordinates (X, Y values) that correspond to the current mouse position (when the mouse is positioned over the graph).

The **Round Axis Scales After Zoom** item toggles on and off whether after zooming in on a region of the graph (see How to zoom with the mouse) the graph scales are first rounded (rounding on) or instead the graph displays the exact region that was zoomed (rounding off). Rounding avoids very odd values for the scale minimum and maximum, but after zooming into a very small region it sometimes leads to a range that is much larger than you intended to display.

The **Customization Dialog** item brings up a tabbed dialog (shown below) that allows you to alter features such as line/point styles and colors, the plotting method for individual data sets (points, lines, *etc.*), legend display, legend position, graph borders, and graph background colors. This customization dialog also provides a second means of customizing axis scaling (but not tick marks, as can be done by clicking directly on the axis) and graph titles (but not axis titles).

Graph customization dialog

The **Export Dialog** context menu item brings up a dialog that allows printing the graph, exporting the image to a file or the Clipboard, and even exporting the data that creates the graph. Note however that if you want to export data, this is usually done more easily using the Edit menu.

For graphs that contain markers such as the vertical lines to indicate the meniscus position or regions of interest, if you don't want those markers to appear in the exported graph first uncheck the **Show Annotations** menu item before you export the graph.

The **Help** menu item brings up a small graphing-specific Help file that discusses some of these options in more detail.

The **Print Graph to Default Printer** item sends a full-page graph in landscape orientation to the current default printer (the Windows default printer, not the printer specified for analysis reports from the Options dialog). If you want to print to a specific printer, print a graph of a specific size, or control whether the graph is printed with color or in monochrome, you can do this by selecting the **Export Dialog...** menu entry.

Most of these context-menu customization features are implemented by the Gigasoft ProEssentials graphing library used by this program.

4.3.3 How to export graph images or data to other programs

Right click with the mouse on any graph to bring up a context menu that provides access to various graph options and customization features, including the **Export dialog**.

The **Export dialog...** menu item brings up a dialog box that allows

1. exporting the image to a file or the Clipboard
2. exporting the data that creates the graph
3. printing the graph to any available printer

With regard to exporting data values, note that the data contained in many of the graphs can be more easily copied to the Clipboard using the Edit menu.

Exported image types supported include Windows metafile, Windows bitmap (BMP), JPEG, and portable network graphics (PNG).

4.3.4 How to zoom with the mouse

To "zoom in" (enlarge) a section of the graph you can click the left mouse button at the corner of a region within the axes of an existing graph and then drag (move while holding the left mouse button) until you reach the opposite corner of the region you wish to be expanded. As you drag, an inverted area will indicate the region being marked. When you stop dragging (by releasing the mouse button) the axes will be rescaled to show you this region.

If the region you have defined is very small, the zoom is not enabled, as indicated by the cursor shape.

You can undo this zoom and return to the previous scales for the axes by entering a **Z** or **z** from the keyboard, or right-clicking on the graph and using the 'Undo Zoom' entry on the pop-up menu.

Note that by default the axis scales will be rounded to even values after you zoom. Sometimes this behavior is undesirable, and it can also prevent you from zooming to very high magnification. This scale rounding behavior can be turned off by right-clicking on the graph and selecting the 'Round Axis Scales After Zoom' entry to toggle it off.

4.4 How to remove bad data points from scan files

Generally it is not necessary to manually remove individual "bad" data points from a scan file, because those can be temporarily removed by the despiking routine (see Load raw scans - Data loading options). If you do want to edit a scan file to remove bad data (perhaps for use by other software), methods for doing so are discussed below.

If you do need to manually remove bad data point(s), this is very easy to do:

1. If needed switch to Advanced mode.
2. Go to the Load raw scans page and be sure the Despiking option is turned off so that all data points are visible.

3. Advance to the Set meniscus & data region page to graph the raw data
4. Push the ID points button to display every data point as a clickable object.
5. Click on the bad data point and its *x*,*y* coordinates and scan file name will be displayed.
6. Click on this same point a second time and you will be asked to confirm that you want to permanently remove that point, and then asked whether you want to remove it in an irreversible or reversible manner (explained below). If you choose to make an irreversible deletion, a back-up copy of the un-altered scan file will be created (by adding an extra .bak extension on the scan file name).

Although the easiest way to remove bad points is the one described above, from the Load raw scans page you can also right-click on a scan file name and then automatically bring that scan into Notepad for editing. If you do edit scan files using the right-click pop-up menu from the Load raw scans page, the program will automatically reload that edited scan file.

Alternatives for removing bad points from scan files:

Although it possible to simply delete the data row corresponding to a bad data point from scan files, that procedure leaves no record that a bad data point was removed, and does not permit easy restoration of the original data. To address these concerns, this program implements the "University of Connecticut convention" for marking bad data points in a manner where the original information is retained so it can later be restored if needed.

- Note however that not all AUC analysis software follows this convention, and in particular Peter Schuck's SEDFIT and SEDPHAT do not, so if you want to use DCDT+ to edit scans for those programs you must permanently delete the bad data points.

By this UConn convention, any data point for which the value of the third column in the data is negative will be ignored when the file is read. For absorbance data, this third column normally contains either zero (when no replicates are used) or the standard deviation of the absorbance value (when multiple flashes are being averaged). For interference data, there normally is no third column entry. However, this software checks each line of data for a third column, and will ignore any data point where a third column exists and is negative.

Thus to mark a data point as "bad" the entry in the third column is set to a negative value (by adding a third column for that line with -1 as the value in the case of interference data). For absorbance data, simply changing the sign of the standard deviation allows you to later restore the original raw data if needed, without loss of information, by changing the sign back to positive again.

4.5 How to identify and remove rogue scans

The interference optical systems may sometimes produce "*rogue scan*s". Needless to say, you don't want to use these aberrant scans in your analysis!

You may see such a scan when you view the scans on the Set meniscus & data region page. The image below shows an example where there is a fringe jump within the sample region.

Example of rogue scan

Because there may be dozens or even hundreds of scans in this display, it can be difficult and frustrating to try to identify which is the bad one.

When there aren't too many scans you may be able to identify the rogue scan from the line color and pattern shown in the legend. However by default this graph does not display a legend when there are more than 40 scans shown (as is true above) because the legend occupies too much space. To force display of the legend, right-click on the graph and select a legend position from within the Legend Style sub-menu, as shown below.

Bring up the graph pop-up context menu to force display of a legend

That results in the following graph, from which you may be able to see that scan number 92 (file 00092.ip3) is the bad one.

Scan display with the legend forced on

However an alternate procedure that works even when there are hundreds of scans is to push the ID points button. This turns every point in every scan into a clickable object, and places a window above the graph that will display the scan file name and X,Y data values for any point that is clicked, as shown below. (Generating this display can take a fair amount of time, so be patient.)

Scan display in ID points mode

Now if you click anywhere along the right portion of the rogue scan (the easily-identified region to the right of the fringe jump), then as shown below the window above the graph identifies that the rogue scan is number 92.

Scan display after clicking on data point

Once the rogue scan is identified, use the Back button to return to the Load raw scans page. Click on the file name of the rogue scan within the Scans currently loaded in memory list box, as shown below.

Preparing to use the Delete selected scans button

Then push the Delete selected scans button, the rogue scan will be deleted from memory, and you can continue with your analysis.

5 Frequently-asked questions

5.1 About g(s*) distributions

5.1.1 What is the difference between s* and s?

The usage of the symbol ***s**** is intended to indicate that this is an "apparent" sedimentation coefficient. That is, it is one derived from the transformation from radial coordinates to sedimentation coefficient coordinates through the relation

where r_m is the radius of the meniscus.

The designation as "apparent" will be clearer if we use the following *g(s*)* distribution as an example:

This distribution shows one very broad peak with a maximum at ~1.8 S. Let us consider some sedimentation coefficient away from this maximum, say at 1.2 S. The distribution indicates a value of ~1.2 AU/S at this position. Does this mean there is really this amount of material with a sedimentation coefficient of 1.2 S in this sample? Possibly, but not necessarily.

This distribution could arise from a truly polydisperse sample containing many components covering a broad range of sedimentation coefficients between 0 and 5 S, in which case there really could be a significant amount of a species with a true sedimentation coefficient of 1.2 S.

On the other hand, this distribution might also be broad primarily because of diffusion. In that case the non-zero value at 1.2 S might not be an indication that there is any real species with $s = 1.2$, and instead the value of *g(s*)* at 1.2 S may arise totally from the influence of diffusion on species having other sedimentation coefficients.

That is, diffusion may make it only appear that there is material sedimenting at 1.2 S because it makes the boundaries broad and therefore spreads the contribution of each species over a range of sedimentation coefficients. Indeed this is true in this case, since this is actually the result from a Claverie simulation for a ~18 kDa protein at $s = 1.88$ plus 5% dimer at $s = 2.82$.

Hence the terminology "apparent sedimentation coefficient" or *s** arises from the fact that the influence of diffusion has not been removed from these distributions, and the resultant uncertainty in interpretation.

However, when the *g(s*)* or *dc/dt* distributions are fitted as a mixture of species, the values of sedimentation coefficient returned by the fit represent the values for true individual species, and thus should no longer be designated as *s** or "apparent" values.

You should be aware that some scientists in this field also use the "*s**" symbol and "apparent" sedimentation coefficient terminology to indicate raw sedimentation coefficients, *i.e.*, values which are not *s(20,w)* values that have been corrected to standard conditions and/or which have not been extrapolated to zero concentration.

5.1.2 What is the difference between g(s*) and g^(s*)?

The *g^(s*)* ["g hat"] distribution reflects the composition of the sample at the time the scans used in the analysis were taken. Thus the total area under the curve represents the plateau concentration at that time in the run, and the concentration of any one species in a mixture (area under any particular peak) is its concentration in the plateau region at that time.

The *g(s*)* distribution is determined by extrapolating the *g^(s*)* distribution back to time zero, accounting for the radial dilution effects. Thus the total area under the *g(s*)* curve represents the loading concentration, and the concentration of any one species in a mixture (area under any particular peak) is its loading concentration.

The *g^(s*)* distribution is therefore the more fundamental distribution and is more closely tied to the raw experimental data, but it is the *g(s*)* distribution that is more commonly presented.

For **interacting systems** (reversible self-association or mixed association) the composition will in general change during the velocity run, rendering an extrapolation to time zero invalid. Therefore only the *g^(s*)* distribution should be presented or used in calculating averages over the distribution for such samples.

For **non-interacting systems** (mixtures) the *g(s*)* distribution is preferred since one usually wants to know the composition that was initially loaded into the cell.

5.1.3 What is the difference between the broad and conventional g(s*) algorithms?

The program offers two alternative ways of calculating *g(s*)*:

1. the conventional algorithm developed by Walter Stafford (reference 7)
2. the "broad" algorithm developed by John Philo (reference 8)

This alternative broad algorithm is primarily provided

- for backward compatibility with DCDT+ versions 1.xx; and
- to give better estimates of the precision of the weight- and z-average sedimentation coefficients from the *g(s*)* distribution in some circumstances

Selection of the broad algorithm is only available in Advanced mode. Einstein

Fitting to *g(s*)* or *dc/dt* calculated using the broad algorithm is implemented only for the classic method, not for the new improved method.

Show All Show All

Hide All Hide All

Show How do the algorithms differ?

The conventional algorithm first averages together the *dc/dt* data from different scan pairs, then calculates *g^(s*)* from that average, and then calculates *g(s*)* from *g^(s*)*.

The broad algorithm calculates a *g^(s*)* from each scan pair, then averages these to get an average *g^(s*)*, and then calculates *g(s*)* from the average *g^(s*)*. The broad algorithm also allows an alternate calculation of the average *dc/dt* curve (see below).

This broad algorithm has some advantages when the time span of the data is long and the peaks are being signficantly broadened (hence the name "broad algorithm"). When the time span is very short and the peaks are not being broadened the two are equivalent.

The actual *g^(s*)* and *g(s*)* curves from the two algorithms are essentially identical; what differs are the error bars on those values (and therefore the uncertainties in derived fitted parameters, weight-average sedimentation coefficients, etc.). The conventional algorithm tends to over-estimate those error bars when the time span grows long, and this is at least partially corrected by the broad algorithm.

Because the broad algorithm gives smaller error bars on *g(s*)*, it also tends produce lower uncertainties for the parameters derived by fitting to *g(s*)* when the fit uses the classic method and is being weighted using the estimated error bars for each data point. This improvement in precision is greatest for sedimentation coefficients and loading concentrations; there is little improvement in the precision of *D* or *M*.

The graph below (figure 3 from reference 8) shows how the broad algorithm (panel B) reduces the error bars as more scans are used, whereas with the conventional algorithm the error bars actually increase in going from 12 to 16 scans (panel A).

The broad algorithm may be switched on a case-by-case basis (even after *g(s*)* has already been calculated) using a button on the calculate g(s*) distribution page.

Show

So should I use the broad algorithm?

Generally no. Because the use of experimental weights during fitting is no longer recommended, and because the improved method is now the default, the issue of improved estimates for the uncertainties of the data points in the *g(s*)* or *dc/dt* data is no longer very important.

Probably the one circumstance where the broad algorithm will continue to be useful is when you want to calculate and report the weight-, z-, or z+1-average sedimentation coefficient for the whole sample or a portion of the distribution. The use of the broad algorithm will improve the accuracy of the estimates of the uncertainties for those quantities.

Show

How the broad algorithm affects the average *dc/dt* data

The broad algorithm also calculates a revised average *dc/dt* curve, again providing an improved estimate of the error bars on *dc/dt* when the time span is grows longer. This revised average *dc/dt* curve is back-calculated from the average *g^(s*)* distribution, as described in reference 8.

Because the broad algorithm gives smaller error bars on *dc/dt*, it also tends to improve the precision of parameters derived by fitting to *dc/dt* using the classic method (the parameters will have lower uncertainties). This improvement in precision is greatest for sedimentation coefficients and loading concentrations; there is little improvement in the precision of *D* or *M*. The broad algorithm also somewhat reduces the systematic errors in *D* and *M* arising from peak broadening (i.e., it improves their accuracy). For a more complete discussion of this issue see reference 8.

One complication arising in back-calculating *dc/dt* from *g^(s*)* is that the range of sedimentation coefficients used in calculating *g^(s*)* and *g(s*)* is often less than the full range of the *dc/dt* data. This complication is handled by using the conventional algorithm to calculate the average *dc/dt* for the regions that were excluded from the *g(s*)* calculations.

5.2 About fitting

5.2.1 Should I fit the offset?

The offset parameter allows for some error in defining the zero level of the *g(s*)* or *dc/dt* data. Whether or not to allow a non-zero offset is largely a matter of scientific judgement.

If you have no reason to believe there is any uncertainty in the zero level, then you probably should keep the offset at zero. This should nearly always be true for absorbance data (unless perhaps there are low mass absorbing species that are not well matched between sample and reference).

On the other hand, if the *g(s*)* data do not return to zero or go negative in regions away from the peak(s), then it will be impossible to obtain a good fit unless you include an offset. With the *dc/dt* data it is more difficult to judge the zero level since the data will not return to zero on the right side of peaks. Be careful though, because the presence of species sedimenting outside the range of sedimentation coefficients covered by your data can give the appearance of a zero offset problem.

With interference data, and particularly for low concentration samples, even the jitter removal procedure may not get the zero level set correctly.

This issue is of course also related to whether or not you choose to manually adjust the zero levels of the data using either the Reset zero for average dc/dt or Reset zero for g(s*) procedures. In principle if that manual adjustment is done correctly then the offset fitting parameter should be left at zero.

5.2.2 Why do I get different answers when fitting g(s*) versus dc/dt?

The short answer is that these two fitting approaches differ in the approximations involved, and their sensitivity to certain errors. Therefore the results should be somewhat different, and neither is necessarily 'right' or 'best' under all circumstances.

The differences between fitting to *g(s*)* versus *dc*/dt data are far less when using the new improved method than they were using the classic method. We will first discuss issues that are common to both the improved method and the classic method, and then give a detailed and lengthy discussion that applies strictly to the classic method.

You can (and should!) confirm the phenomenon discussed below for yourself by fitting to Claverie simulations where the correct answer is known.

Show Differences between fitting g(s*) versus dc/dt with either improved or classic methods

In multi-species fits there may be some differences in sensitivity to minor components and resolving power. Minor species that sediment faster than the main component will produce a more prominent peak in the *dc/dt* data than in *g(s*)*, and conversely minor species that sediment slowly will be more prominent in *g(s*)* than in *dc/dt*. In principle this effect is compensated by the fact that the signal/noise for the minor component is equivalent in both distributions, so fitting either should be equivalent. However, when the separation is poor, the higher amplitude for the minor component may help the fitting, so one option may be significantly better than the other.

Another fundamental difference between the fitting methods arises because the *dc/dt* curves for each species do not return to zero at high sedimentation coefficients, whereas the individual peaks in *g(s*)* are non-zero only over a finite range. When fitting to *dc/dt*, a poor fit of a slowly sedimenting component can lead to a error in the high *s* 'tail' from that component that will affect the results for the faster sedimenting components.

Show Differences between fitting g(s*) versus dc/dt using the classic method

For a full and detailed discussion of the accuracy and advantages/disadvantages of both approaches when using the the classic method see reference 8. That discussion is summarized below.

In general, when the time span of the scans is short and peak broadening is minimal, fitting to *dc/dt* will give more accurate values for *s* and *D* (or *M*). This is true primarily because certain approximations are involved in transforming from *dc/dt* to *g(s*)*, and in part because the function used for fitting the *dc/dt* data is a better approximation to the true shape of the curves than is the Gaussian function used in fitting the *g(s*)* data.

Fitting to *g(s*)* tends to underestimate *D* and overestimate *M* by a few percent (and in some circumstances these errors can be as large as ~10%). In theory the accuracy for *s, D,* and *M* when fitting to *dc/dt* is ~1%, but in real experiments the boundaries are always a bit broader than expected (for reasons not fully understood) so even under optimal circumstances *D* usually comes back a few percent high and *M* a few percent low.

For the classic method in many cases the accuracy of *D* and *M* is strongly affected by peak broadening, for both *g(s*)* and *dc/dt* fits. Peak broadening as the time span of the scans grows too large causes an overestimate of *D* and an underestimate of *M*. See Peak broadening limit about defining when the time span is "too large".

When fitting to *g(s*)* the peak broadening error tends to compensate for the opposing tendency to underestimate *D,* so the apparent *M* may actually be closer to the "expected" or true value when fitting to *g(s*)* than when fitting to *dc/dt*, but this will be true only over a certain range of broadening. When the broadening is fairly strong the absolute errors are similar for both fitting approaches.

For slowly sedimenting species ($s < 4$) fitting to *g(s*)* also underestimate the *s* value, and therefore such results should probably not be used for shape analysis unless you manually correct for this effect (for correction factors see reference 6). This error is absent when you fit the *dc/dt* data. These errors in sedimentation coefficients when fitting *g(s*)* can be significant even for larger species (6 S or greater), particularly if they are run at low rotor speeds, and also cause problems in trying to compare sedimentation coefficients from runs at different rotor speeds.

On the other hand, when species are poorly resolved, the fact that curve shapes are more accurate for *dc/dt* fitting can lead to significantly enhanced ability to resolve the species and get accurate values for their properties.

It cannot be over-emphasized that the best way to determine what these types of analyses can and cannot do, and to assess their potential accuracy, is to run Claverie simulations of your experiments (including noise) and fit the simulated data.

5.2.3 What constitutes a good fit?

Judging "good" and "bad" fits is somewhat of an art, and ultimately relies on the accumulated experience of the experimenter. However, in general a good fit is one where:

- the residuals are small and randomly distributed (they are not systematically above or below zero over substantial portions of the radial range)
- the *rms residual* is consistent with the intrinsic optical noise of the instrument
- the returned values make scientific sense based on other information you have about this sample

With regard to the last point, one important criterion is whether the apparent molecular weight implied by the *s/D* ratio is consistent with the sequence weight, values from mass spectrometry, or sedimentation equilibrium results.

For truly homogeneous, non-associating samples the apparent *M* from fitting using the improved method should probably be within 2-5% of the "correct" value.

It is important to remember that the presence of extra species or self-association will make the boundaries and the peaks in *dc/dt* and *g(s*)* broader, which makes the apparent *D* larger and therefore the apparent *M* smaller. Thus if the apparent *M* is more than ~5% below the "correct" value, it is likely that your sample is heterogeneous or self-associates. If this is due to self-association, you should see shifts in *s* values by changing sample concentration.

A second important criterion is to ask whether the results make hydrodynamic sense. Is a species sedimenting faster than is theoretically possible for its molecular mass ($f/f_0 < 1$)? Is a species sedimenting much more slowly than is reasonable for its molecular mass (f/f_0 very high)?

With regard to whether the *rms residual* is consistent with the instrumental noise, this may be difficult to judge because of the raw data have been substantially transformed in deriving the *dc/dt* and *g(s*)*distributions. As an aid in making this judgement, when a fit report is printed a calculation is made to back-transform the observed rms residual and calculate an approximate equivalent noise level in the original scans.

Causes of "bad" fits

A number of causes of bad fits are discussed under How to avoid common mistakes .

Another factor that can cause boundary broadening and thus overestimates of *D* is scratches or roughness on the inside of the centerpiece.

5.2.4 What does high parameter cross-correlation mean?

Parameter cross-correlation values near 1 or -1 indicate that the best-fit value for that parameter is extremely sensitive to the value for another parameter. Thus these two parameters cannot really be independently determined, and neither is truly well-defined.

Technically, the high correlation means that the increase in the sum of squared residuals that arises when a parameter is moved away from its optimum value can be almost completely compensated by simultaneously adjusting the value of the other parameter.

It is not at all uncommon for the concentration parameters in multi-species fits to be highly negatively correlated, and this is generally not a cause for concern. High cross-correlations for the other parameters is, however, usually a significant concern, and probably means that the data contain insufficient information to determine all the parameters uniquely. This condition might be relieved, for example, by reducing the number of species or by using the procedure.

The likelihood of high negative correlations between species concentrations arises because it is generally true that it will be possible to largely compensate for the effect of decreasing the concentration of one species by adding the same amount to another species.

One important way to reduce parameter cross-correlation in multi-species fits is to reduce the number of fitting parameters by constraining the ratios of molecular masses or sedimentation coefficients between species.

5.3 How should I cite this program?

1. The primary citation for this program is
    - Philo, J. S. (2006). Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. Anal. Biochem. 354, 238-246.
2. Because the original development of the *dc/dt* method was by Walter Stafford, **please always also cite Walter Stafford's original paper**:
    - Stafford, W.F., III. 1992. Boundary analysis in sedimentation transport experiments: A procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Analytical Biochemistry* 203:295-301.
3. The program itself should be mentioned in your *Methods* section as DCDT+ by John Philo, with the version number.

Specialized method citations

The reference for the special multi-segment dc/dt calculations is

- Philo, J. S. (2011). Limiting the sedimentation coefficient range for sedimentation velocity data analysis: Partial boundary modeling and g(s*) approaches revisited. *Anal. Biochem.* 412, 189-202.

The reference for fitting to dc/dt data rather than g(s*), and for the Use true mean time button (the 'broad algorithm') is:

- Philo, J.S. (2000) A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. *Analytical Biochemistry*, 279, 151-163.

6 Reference information

6.1 Symbols used

D	diffusion coefficient
$D(20,w)$	diffusion coefficient corrected to standard conditions (20 °C, water), but not extrapolated to zero concentration
M	molecular mass
R	universal gas constant
s	sedimentation coefficient
s^*	apparent sedimentation coefficient (not corrected for the effects of diffusion)
$s(20,w)$	sedimentation coefficient corrected to standard conditions (20 °C, water), but not extrapolated to zero concentration
ρ	solvent density
	partial specific volume
ω	angular velocity [$\omega = 2\pi \times (\text{RPM}/60)$]
	product of angular velocity squared times elapsed time in the run (proportional to total sedimentation force)

6.2 About using radial intensity scans

In addition to normal absorbance scans, this program can also directly load radial intensity scans, including so-called **pseudo-absorbance scans** where two different samples were loaded in one cell (one in the sample channel, one in the reference channel).

When the program detects that you are trying to load intensity scans, it brings up the following dialog box for you to select how you want the data to be handled.

Intensity conversion dialog

Processing and handling of pseudo-absorbance data

When you select either of the pseudo-absorbance options, the program assumes a maximum intensity value of 1000 and converts all the data points relative to that value using the formula $OD = \log(1000/I)$.

When the "Write converted data as scan file in same folder" check box is checked (which is true by default), as the intensity scan files are read the converted absorbance or pseudo-absorbance values will be written out as a normal radial absorbance scan file, into the same folder from which the intensity scans are being read. For pseudo-absorbance data, the file name will be preceded by an "S" or "R" to indicate whether this data is from the sample or reference channel, respectively. That is, pseudo-absorbance scan file 00001.RI1 will generate new files S00001.RA1 or R00001.RA1.

If these converted files are later read by the program, the only way it can distinguish that these were pseudo-absorbance scans is the "S" or "R" as the first letter of the file name. Therefore whenever scans with names starting with "S" or "R" are read a dialog box will ask you to confirm whether these are in fact pseudo-absorbance data.

- The important functional difference between pseudo-absorbance data and true absorbance data is that pseudo-absorbance data exhibit systematic shifts somewhat analogous to "jitter" in interference scans. Thus when the program recognizes pseudo-absorbance data it turns on jitter correction by default.

6.3 About this g(s*) implementation

The calculation algorithms used here differ slightly from those used in Walter Stafford's versions of DCDT and the WinDCDT version from the Univ. of Connecticut by Jeff Lary and Walt Stafford. The differences relate only to the treatment of the data near the meniscus and near the base of the cell.

1. This implementation excludes data points at radii just to the right of the meniscus because the data near the meniscus are always distorted by optical aberrations and because the data in this region are also usually noisy.
   - The width of this excluded region near the meniscus is determined by a "meniscus clearance" parameter, with a default value of .015 cm. This default value can be altered using the Options dialog. For interference data when the region used for jitter removal is to the right of this excluded region near the meniscus, then the data within the jitter alignment region will also be excluded from the *dc/dt* calculations, since these data have already been forced to be zero (on average).
2. This implementation also excludes data near and beyond the base of the cell based on a maximum radius for inclusion (initial default value 7.15 cm). This default value can be altered using the Options dialog; the value can be manually altered for each analysis from the set meniscus & data region page. A similar exclusion is used Walt Stafford and Jeff Lary's versions of DCDT, but the value cannot be altered by the user.
3. Another way in which this implementation is different concerns how the *dc/dt* curves from each scan pair are combined into an average *dc/dt* curve. Scan pairs from later in the run will yield *dc/dt* curves that begin and end at lower sedimentation coefficients than those from earlier in the run. Thus at the extremes of the sedimentation coefficient range there are data points from only one of the scan pairs, which makes defining the average value and standard error for those points problematic.
   - In this implementation, the program computes the average *dc/dt* curve only for the range of sedimentation coefficients that is covered by at least two of the scan pairs. Because there are fewer values to average near the extremes of the sedimentation coefficient range, you will find that the error bars for those points are generally larger.

6.4 About the fitting functions

The fitting function that is used depends on the choice between the improved or classic methods, and the choice of fitting to either the *g(s*)* distribution or the average *dc/dt* curve.

The classic method is provided primarily to provide backward compatibility---the improved method is recommended for all new analyses.

Show All Show All

Hide All Hide All

Show Function calculation for the 'improved' method

The 'improved method' is based around calculating the actual raw sedimentation boundaries that would occur for the current sum of individual species, at the exact time points corresponding to the experimental scans being analyzed. Those theoretical raw scans are then processed to generate either average dc/dt data or the g (s*) distribution using the identical algorithms that were applied to the experimental data.

Because this approach exactly mimics the processing of the experimental data, the peak broadening that inevitably occurs as the boundary movement between the first and last scans analyzed grows larger does not cause errors in the fitted properties for each species (as it does for the 'classic' method). In essence this approach gives accuracy comparable to whole boundary fitting approaches while retaining the advantages of time-derivative approaches.

The theoretical sedimentation boundaries are calculated using an approximate solution to the Lamm equation described by Joachim Behlke in reference 9 that is also used in his whole boundary analysis program LAMM. The actual function used is the first three terms in Eq. 28 of that paper (the terms which describe the moving boundary). This function was shown to give errors for sedimentation coefficient of less than 0.2%, and errors for D or M or less than 0.5%, for proteins of 2 kDa or larger, and errors of 0.1% or below for all parameters for proteins larger than 10 kDa.

Show Function calculation for the 'classic method'

When fitting to the g(s*) data the fitting function is a Gaussian. A Gaussian function describing the *g(s*)* curve for a single species with sedimentation coefficient s may be defined by

where *s* is the center of the peak, *A* is the total area under the curve (the loading concentration), and σ is the width of the Gaussian.

Note that many definitions of the Gaussian function employ a definition of the width of a Gaussian which is twice that used here; this definition is used for consistency with the one used by Walter Stafford.

The width of the Gaussian, σ, is directly related to *D* [reference 3] through the relation

where r_m is the meniscus position, ω is the rotor angular velocity, and $\omega^2 t$ is the integrated product of ω^2 and time, as recorded in the header of each scan.

When fitting to the *dc/dt* data the fitting function is the analytical time derivative of the modified Fujita-MacCosham function [reference 1]. The modified Fujita-MacCosham function is an approximate solution of the Lamm equation, and is used for direct boundary fitting in SVEDBERG. This function gives good results for proteins of ~5 kDa and above.

6.5 About the fitting algorithm

This program uses a modified Gauss-Newton iterative least-squares fitting routine, essentially corresponding to the "preferred method" as described by Johnson & Faunt (1992) [reference 2].

The user may select whether all data points are equally weighted or to apply either theoretical or experimental weights. That choice is made within the Select fitting model page.

There are three modifications to the standard Gauss-Newton method, all of which occur at the end of each iteration:

1. Each new parameter is tested to be sure it falls within defined upper and lower bounds. If that parameter

would exceed the bounded limit, it is set at the limiting value. When this occurs a yellow warning box appears next to that parameter on the Fitting parameters monitor tab of the Perform least-squares fit page.

2. An increment reduction factor is applied to insure that the new set of parameters actually gives a lower variance and to improve the rate of convergence. The current value of the increment reduction factor is also displayed on the Fitting parameters monitor.
3. If any constraints have been set on the fitting parameters those constrained parameters are updated to reflect the new values of the parameters to which they are linked.

6.6 About the meniscus wizard

1. For **absorbance scans** the program logic first locates a starting meniscus position by finding the highest OD value within the leftmost 20% of the last scan being used in the analysis (trying to find the top of the positive spike at the meniscus). It then examines a region within 0.02 cm of that initial position to find the top of the positive spike for each of the scans currently in use, and sets the meniscus at the arithmetic mean of those positions (rounded to the nearest 0.0001 cm after averaging).
    - This usually works well, since conventionally for absorbance scans the meniscus position is considered to at the top of the positive spike. When the sample OD is quite high the plateau is sometimes above the meniscus spike, and this algorithm may fail to find the correct meniscus position in such cases.
    - If you manually drag the meniscus marker you will then be asked whether you want the wizard to re-calculate the mean position of the spike. This is helpful when the scans contain two positive spikes, and the wizard has identified the wrong one as the true meniscus.
2. For **interference scans** there is usually a slight negative dip at the sample meniscus. Thus the logic positions the meniscus at the minimum fringe reading within the first 1/4 of the scan.
3. For **fluorescence scans** there is not yet sufficient data to discern a pattern to allow automatic location of the meniscus. Therefore at this time the program simply puts the meniscus at 6.0 cm (or whatever is the first data point if the scan starts to the right of 6.0 cm).

6.7 About the jitter wizard

To make the *jitter* correction automatically, the program tests two general regions, using spans of 20 data points:

1. a region in the air-air space
    - if the scans start to the right of 5.83 cm, the first 20 data points in each scan are used
    - otherwise a region started at 5.83 cm is tested; if the scan-to-scan variance is high at 5.83 cm (probably because the reference meniscus is to the left of 5.83 cm) the wizard compares regions starting at 5.82, 5.81, and 5.80 cm and picks the best one (lowest variance).
2. the region to the right of the meniscus starting at the left green "good" data marker (a region that will be constant if the meniscus is fully cleared at the time of analysis)
    - the initial trial region begins 0.007 cm to the right of the green data marker
    - four additional regions each 20 points further to the right are also examined to see which of these five possible regions is best (if the boundary is well clear of the meniscus generally a position further to the right, closer to the boundary, works best)

The wizard examines the deviations within each scan over these regions and picks the one where the variance is lower (after removing the average Y value for that scan over that region).

6.8 About the bootstrap method

In the 'bootstrap with replacement' (reference 11) method (usually shortened to just 'bootstrap method') for evaluating confidence limits data points are selected randomly from the original fitted data set to make a new data set with the same number of points. The random selection means that some of the original points will be selected more than once (and obviously others not at all). That is, each bootstrap data set is really a subset of the original data (but some points are repeated).

This new data set is then fitted, and the new fitted parameters are stored in an array. This whole process is repeated *N* times, and then for each fitted parameter the array of *N* values is used to evaluate the confidence limits.

Note that in the implementation used here some random noise is added to the best-fit parameters before each round so that each bootstrap round begins from slightly different starting values (to avoid biasing the results).

What are the advantages and drawbacks of this approach?

The principal advantage of the bootstrap method is that it makes no assumptions about the magnitude or randomness of the noise for the data being fitted. Thus for example if the *residual*s for the original best fit show a systematic pattern, that pattern is retained by the data subsets created during the bootstrap rounds.

A further advantage over the Monte Carlo method is that in real experiments the noise level of the raw data is often not constant across a scan or for scans at different times in the run. That is particularly true for fluorescence scans, where the noise magnitude is usually substantially higher in the plateau regions than in the low concentration regions to the left of the boundary. That same phenomenon also exists for absorbance scans, but the change in the noise magnitude is smaller than for fluorescence scans.

One drawback of this method is that it operates on the *g(s*)* or *dc/dt* data, not the original raw scans. Thus it cannot directly evaluate whether the data transformations themselves introduce any statistical bias. Reference 13 shows that statistical bias resulting from the data transformation is in practice not a significant issue.

6.9 About the Monte Carlo method

In the Monte Carlo method for evaluating confidence limits first a set of theoretical raw scans is generated based on the number of species and the parameters from the best fit. Then random noise at a specified rms level is added to those theoretical raw scans, and new average *dc/dt* data and a new *g(s*)* distribution are calculated, and then a new data set is created (covering the same range of sedimentation coefficients as the original data).

This new data set is then fitted, and the new fitted parameters are stored in an array. This whole process is repeated *N* times, and then for each fitted parameter the array of *N* values is used to evaluate the confidence limits.

Note that in the implementation used here some random noise is added to the best-fit parameters before each round so that each Monte Carlo round begins from slightly different starting values (to avoid biasing the results).

HHow is the magnitude of the random noise to be added determined?

One challenge for applying the Monte Carlo method to time-derivative analysis is that the *rms residual* of a fit to *g(s*)* or *dc/dt* data does not directly tell us the rms noise level of the original raw scans. Therefore this implementation uses an iterative process. First an *estimated* level of noise is used for *N* rounds of Monte Carlo simulation. The mean *rms residual* from these Monte Carlo rounds is then compared to that from the original best fit. If these two values differ by more than 2%, the user is asked whether to re-adjust the noise level for the Monte Carlo to try to better match the best fit. If so, then the noise is scaled up or down, another *N* rounds of Monte Carlo are performed, and the noise levels are again compared.

This iterative process then continues until noise levels agree within 2% (or the user is satisfied that some poorer match is sufficient). Usually a match within 2% is achieved with one or two iterations.

For the *dc/dt* data it is fairly trivial to estimate the noise in the raw scans since it is reasonable to assume that the noise in the average *dc/dt* curve is that of the raw data, times the square root of 2 (from the subtraction of a pair of scans), divided by the average time difference between the scan pairs.

For the *g(s*)* data the noise in the raw scans is estimated by first assuming that the noise in *g(s*)* varies inversely with sedimentation coefficient (the pattern expected if the noise in the raw scans is random), using that assumption to calculate the apparent noise in the average *dc/dt* curve, and then calculating as indicated in the preceding paragraph.

What are the advantages and drawbacks of this approach?

The principal advantage of the Monte Carlo method is that it evaluates the confidence limits based on noise in the raw scans (and noise that is truly random).

Its main drawbacks are:

1. we don't usually know the true noise level of the raw scans
2. the noise level may not be constant across each scan, or from scan to scan
    - in particular note that for fluorescence scans the noise magnitude usually is substantially lower in the region to the left of the boundary than in the plateau region---for that reason the bootstrap method may be more appropriate for fluorescence data
    - the noise in absorbance scans is also usually significantly lower in the low OD regions to the left of the boundary than in the plateau region
3. we generally don't know whether the experimental noise is truly random (except when we are fitting simulations)

7 Program license and registration

8 Bibliography

8.1 reference 1 - Philo (1997)

Philo, J.S. (1997). An improved function for fitting sedimentation velocity data for low-molecular-weight solutes. *Biophysical Journal* **72**, 435-444.

(This article can also be downloaded as an Adobe Acrobat® document called sveddoc2.pdf from the program download page.)

8.2 reference 2 - Johnson & Faunt (1992)

Johnson, M.L. And L.M. Faunt (1992). Parameter estimation by least-squares methods. *Methods Enzymol.* **210**, 1-37.

8.3 reference 3 - Stafford (1997)

Stafford, W.F., III. (1997). Sedimentation velocity spins a new weave for an old fabric. *Curr. Opin. Biotechnol.* **8**, 14-24.

8.4 reference 4 - Durschlag (1986)

H. Durchschlag (1986). In *Thermodynamic Data for Biochemistry and Biotechnology*, H.-J. Hinz, ed., Springer-Verlag, New York, Chapter 3, p. 45.

8.5 reference 5 - Stafford 'rule of thumb'

This rule of thumb is described in a document by Walter Stafford posted on the RASMB web site at http://www.bbri.org/dcdt/Rule.pdf

8.6 reference 6 - Stafford (1999)

W.F. Stafford, III (1999). Analysis of reversibly interacting macromolecular systems by time derivative sedimentation velocity. *Methods in Enzymology* **323**, 302-325.

8.7 reference 7 - Stafford (1994)

Stafford, W. F., III (1994). Boundary analysis in sedimentation velocity experiments. *Methods Enzymol.* **240**, 478-501.

8.8 reference 8 - Philo (2000)

Philo, J.S. (2000) A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. *Analytical Biochemistry* **279**, 151-163.

(A pre-print of this article is available for downloading from the program download site as the file http://jphilo.mailway.com/dcdtplus.pdf)

8.9 reference 9 - Behlke & Ristau (2002)

Behlke, J. and Ristau, O. (2002). A new approximate whole boundary solution of the Lamm differential equation for the analysis of sedimentation velocity experiments. *Biophys. Chem.* **95**, 59-68.

8.10 reference 10

Garcia de la Torre, J. and V. A. Bloomfield (1981). Hydrodynamic properties of complex, rigid, biological macromolecules: Theory and applications. *Q. Rev. Biophys.* **14**: 81-139.

8.11 reference 11 - Efron & Tibshirani (1986)

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1, 54-77.

8.12 reference 12 - Johnson & Faunt 1994

Johnson, M. L. and Straume, M. (1994). Comments on the analysis of sedimentation equilibrium experiments. In: *Modern analytical ultracentrifugation: acquisition and interpretation of data for biological and synthetic polymer systems*. T.M.Schuster and T.M.Laue, eds. Birkhäuser, Boston, pp. 37-65.

8.13 reference 13 - Philo 2011

Philo, J. S. (2011). Limiting the sedimentation coefficient range for sedimentation velocity data analysis: Partial boundary modeling and g(s*) approaches revisited. *Anal. Biochem.* 412, 189-202.

8.14 reference 14 - Stafford (1992)

Stafford, W.F., III. 1992. Boundary analysis in sedimentation transport experiments: A procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Analytical Biochemistry* 203:295-301.

Index

- **# of sections up/down spinner, _of_sections_updown_spinner.htmlUnknown**
- **1 iteration button, hlp_fit_monitor2.html**
- **About the bootstrap method, About_bootstrap.html**
- **About the fitting algorithm, cii9q4.html**
- **About the fitting functions, 2baj0ix.html**
- **About the jitter wizard, jitter_wizard.html**
- **About the meniscus wizard, men_wizard.html**
- **About the Monte Carlo method, About_Monte_Carlo.html**
- **About this g(s*) implementation, cii9uv.html**
- **About using radial intensity scans, intensity.html**
- **Absorbance scans, hs790.html, Clav_tutorial.html**
- **Add more scans for this sample button, hs820.html**
- **Adjust parameter upper/lower limits button, rdfil..html**
- **Advanced mode, hs200.html**
- **Analysis page 1**
    - Load raw scans - Add more scans for this sample button, hs820.html
    - Load raw scans - Data loading options, hs860.html
    - Load raw scans - Delete selected scans button, hs830.html
    - Load raw scans - Load new scan files button, hs800.html
    - Load raw scans - Preselect data type and cell check box, hs790.html
    - Load raw scans - Replace scans for this sample button, hs810.html
    - Load raw scans - Run Claverie simulation to generate data button, hs870.html
    - Load raw scans - Sample description, hs880.html
    - Load raw scans - Scans currently loaded in memory list box, hs840.html
    - Load raw scans - Scans to be used in g(s*) calculation list box, hs850.html
    - Load raw scans page, hs780.html
- **Analysis page 10**
    - Report fit results - Compute confidence intervals button, hlp_results_x6.html
    - Report fit results - Create reports group box, hlp_results_x7.html
    - Report fit results - Fit statistics, hlp_results_x3.html
    - Report fit results - Fitted parameters, hlp_results_x.html
    - Report fit results - Fixed or constrained parameters, hlp_results_x2.html
    - Report fit results - Graph bootstrap or Monte Carlo results , Graph_bootstrap_or_Monte_Carlo_results.html
    - Report fit results - High cross-correlation warning, hlp_results_x4.html
    - Report fit results - View parameter cross-correlations button, hlp_results_x8.html
    - Report fit results page, hs1360.html

Analysis page 11

- Graph fit results page, hs1440.html

Analysis page 2

- Set meniscus & data region - Data displayed options, hs940.html
- Set meniscus & data region - Fringe jump removal options, 6r6460.html
- Set meniscus & data region - Good data begin X cm from meniscus, 417ij0.html
- Set meniscus & data region - Good data end at radius =, cellbase.html
- Set meniscus & data region - ID points button, ID_points.html
- Set meniscus & data region - Jitter removal options, 1_9pj8m.html
- Set meniscus & data region - Meniscus position, meniscus.html
- Set meniscus & data region - Zoom meniscus button, zoom_men.html
- Set meniscus & data region page, hs890.html

Analysis page 3

- Convert to s(20w), 12uc04r.html
- Select scans to analyze - # of sections up/down spinner, _of_sections_updown_spinner.html
- Select scans to analyze - Auto adjust button, autoadjust.html
- Select scans to analyze - Peak broadening limit, 9dszbc.html
- Select scans to analyze - Use true mean time button, use_mean.html
- Select scans to analyze page, hs950.html

Analysis page 4

- Calculate average dc/dt - Reset zero & Undo buttons, hs995.html
- Calculate average dc/dt - Set range for g(s*), hs990.html
- Calculate average dc/dt - X-axis point spacing, 1lelyh.html
- Calculate average dc/dt curve page, hs970.html

Analysis page 5

- Averages and peak areas dialog, hs1045.html
- Calculate g(s*) distribution - Broad algorithm option, usebroad.html
- Calculate g(s*) distribution - Display options, hs1050.html
- Calculate g(s*) distribution - Normalize option, hs1030.html
- Calculate g(s*) distribution - Reset zero & Undo buttons, e.ppbj.html
- Calculate g(s*) distribution - Set integration region, hs1040.html
- Calculate g(s*) distribution - Show integrals over distribution, 22avez4.html
- Calculate g(s*) distribution page, hs1020.html

Analysis page 6

- Select fitting model - Constrain M or s ratios among species, 11a2ula.html
- Select fitting model - Data choice, hs1100.html
- Select fitting model - Density and partial specific volumes, hs1090.html
- Select fitting model - Fit diffusion or mass, hlp_select_model_M_or_D.html
- Select fitting model - Improved or classic methods, hlp_select_model1.html

  - Select fitting model - Molecular mass units, f.a3iy.html
  - Select fitting model - Number of species, hlp_select_model.html
  - Select fitting model - Weighting of data points, hlp_fit_xsxd.html
  - Select fitting model page, hs1070.html

Analysis page 7

  - Set initial guesses on graph page, hs1150.html

Analysis page 8

  - Adjust parameter upper/lower limits button, rdfil..html
  - Constrained parameters, hlp_constrained_parameters.html
  - Convergence control, hlp_convergence_factor.html
  - Fit control check boxes, hlp_fit_control.html
  - Parameter value text boxes, hlp_input_parameter.html
  - Set/alter fitting parameters - Cancel changes button, hlp_cancel_x.html
  - Set/alter fitting parameters page, hs1160.html

Analysis page 9

  - Fitting parameters monitor tab, hs1260.html
  - Increment reduction factor, hlp_increment_reduction.html
  - Perform fit - 1 iteration button, hlp_fit_monitor2.html
  - Perform fit - Cancel button, hlp_fit_monitor3.html
  - Perform fit - Do fit button, hlp_fit_monitor1.html
  - Perform fit - Simulate button, hs1265.html
  - Perform fit - Status, hlp_fit_monitor4.html
  - Perform least-squares fit page, hs1240.html

Analysis pages

  - Calculate average dc/dt curve, hs970.html
  - Calculate g(s*) distribution, hs1020.html
  - Graph fit results, hs1440.html
  - Load raw scans page, hs780.html
  - Perform least-squares fit, hs1240.html
  - Report fit results, hs1360.html
  - Select fitting model, hs1070.html
  - Select scans to analyze, hs950.html
  - Set initial guesses on graph, hs1150.html
  - Set meniscus and data region, hs890.html
  - Set/alter fitting parameters, hs1160.html

Analysis window, hs150.html

Auto adjust, ni.su1.html

Auto adjust button, autoadjust.html

Average dc/dt curve, hs970.html

Averages and peak areas dialog, hs1045.html

Back button, hs170.html

Bad data points, c2asgg.html, ID_points.html

Bad scans, rogue.html

Baseline offset, faq6.html

Bibliography

- reference 1, krn4g8.html
- reference 10, 3_sixl.html
- reference 11, Reference_11.html
- reference 12, reference_12.html
- reference 13, ref13.html
- reference 14, reference_14.html
- reference 2, hg3u9i.html
- reference 3, 24ylhkp.html
- reference 4, hg3u9k.html
- reference 5, 3_sj2k.html
- reference 6, hg3u9m.html
- reference 7, hg3u9n.html
- reference 8, 3_sixt.html
- reference 9, hs590.html

Bootstrap, Graph_bootstrap_or_Monte_Carlo_results.html

Bootstrap method, hlp_results_x6.html, About_bootstrap.html

Broad algorithm, usebroad.html

Broad g(s*) algorithm, faq5.html

Buoyant mass, f.a3iy.html

Calculate average dc/dt - Reset zero & Undo buttons, hs995.html

Calculate average dc/dt - Set range for g(s*), hs990.html

Calculate average dc/dt - X-axis point spacing, 1lelyh.html

Calculate average dc/dt curve page, hs970.html

Calculate g(s*) distribution - Broad algorithm option, usebroad.html

Calculate g(s*) distribution - Display options, hs1050.html

Calculate g(s*) distribution - Normalize option, hs1030.html

Calculate g(s*) distribution - Reset zero & Undo buttons, e.ppbj.html

Calculate g(s*) distribution - Set integration region, hs1040.html

Calculate g(s*) distribution - Show integrals over distribution, 22avez4.html

Calculate g(s*) distribution page, hs1020.html

Calculator, 7l7or5.html

Cancel button, hlp_fit_monitor3.html

Cancel changes button, hlp_cancel_x.html

Cell base, cellbase.html
Cell number, hs790.html
Citation, faq8.html
Classic method, hlp_select_model1.html
Claverie simulation, hs870.html
Clone, Clone analysis.html
Clone analysis to new window, Clone analysis.html
Confidence limits, hlp_results_x6.html
Constrained parameters, hlp_constrained_parameters.html
Constraints, 11a2ula.html, hlp_constrained_parameters.html
Conventional g(s*) algorithm, faq5.html
Convergence, hlp_convergence_factor.html
Convergence control, hlp_convergence_factor.html
Convert to s(20,w), 12uc04r.html
Copy data, hs60.html
Copy report, hlp_results_x7.html
Correlation matrix, hlp_results_x8.html
Create reports, hlp_results_x7.html
Cross-correlation, faq3.html
Data choice, hs1100.html
Data displayed options, hs940.html
Data loading options, hs860.html
Data range, hs1150.html
Data simulation, hs870.html, Clav_tutorial.html
Data type, hs790.html
DCDT+ User Manual, Manual cover page.html
Defaults, m151qg.html
Delete selected scans button, hs830.html
Density, hs1090.html
Derive g(s*) distribution branch, hs210.html
Despiking, hs860.html
Diffusion, hlp_select_model_M_or_D.html, 7l7or5.html
Display options, hs1050.html
Do fit button, hlp_fit_monitor1.html
Edit data, c2asgg.html, hs840.html
Edit menu, hs60.html
Error ellipse, hlp_results_x6.html
Export, exporting_graphs.html
F statistics, hlp_results_x6.html

FAQs, faq5.html
  Broad g(s*) algorithm, faq5.html
  Citing this program, faq8.html
  Fitting g(s*) versus dc/dt data, faq7.html
  g(s*) versus g^(s*), faq2.html
  Good fit, faq4.html
  High parameter cross-correlation, faq3.html
  s* versus s, faq1.html
  Should I fit the offset?, faq6.html
File menu, 338q0xc.html
File open, 3ze3b.html
File save, 16r_0yl.html
Fit control check boxes, hlp_fit_control.html
Fit distribution as N species branch, 3b9ugl.html
Fit statistics, hlp_results_x3.html
Fit status, hlp_fit_monitor4.html
Fitted parameters, hlp_results_x.html
Fitting parameters monitor, hs1240.html, hs1260.html
Fitting parameters monitor tab, hs1260.html
Fixed or constrained parameters, hlp_results_x2.html
Fluorescence scans, hs790.html, Clav_tutorial.html
Frequently-asked questions
  How should I cite this program?, faq8.html
  Should I fit the offset?, faq6.html
  What constitutes a good fit?, faq4.html
  What does high parameter cross-correlation mean?, faq3.html
  What is the difference between g(s*) and g^(s*)?, faq2.html
  What is the difference between s* and s?, faq1.html
  What is the difference between the broad and conventional g(s*) algorithms?, faq5.html
  Why do I get different answers when fitting g(s*) versus dc/dt?, faq7.html
Fringe jump removal options, 6r6460.html
Fringe jumps, 6r6460.html
g(s*) overlay graph, hs770.html, hs760.html
g^(s*), faq2.html
Geometric mean time, use_mean.html
Getting Started
  How to use this Help file, 2p05re5.html
  Program overview and organization, 5mvbyz0.html
  Standard and advanced modes, hs200.html

Good data end at radius =, cellbase.html

Good fit, faq4.html

Graph bootstrap results, Graph_bootstrap_or_Monte_Carlo_results.html

Graph customization, hs330.html

Graph fit results page, hs1440.html

Graph Monte Carlo results, Graph_bootstrap_or_Monte_Carlo_results.html

Graphical fit monitor, hs1240.html

Graphs

- g(s*) overlay graph, hs760.html

Help menu, 1ns5.4n.html

Help on Help, 2p05re5.html

High cross-correlation warning, hlp_results_x4.html

How (and why) to equilibrate the rotor temperature, ni.su2.html

How should I cite this program?, faq8.html

How to

- Avoid common mistakes, x1liv4.html
- Customize graphs, hs330.html
- Equilibrate the rotor temperature, ni.su2.html
- Export graph images or data, exporting_graphs.html
- Identify and remove rogue scans, rogue.html
- Optimize data acquisition, ni.stz.html
- Print graphs, print_graphs.html
- Remove bad data points, c2asgg.html
- Select scans to analyze, ni.su1.html
- Zoom with the mouse, 1jdo8b2.html

How to avoid common mistakes , x1liv4.html

How to customize graphs, hs330.html

How to export graph images or data to other programs, exporting_graphs.html

How to identify and remove rogue scans, rogue.html

How to optimize data acquisition for g(s*) analysis, ni.stz.html

How to remove bad data points from scan files, c2asgg.html

How to use this Help file, 2p05re5.html

How to zoom with the mouse, 1jdo8b2.html

How to...

- How (and why) to equilibrate the rotor temperature, ni.su2.html
- How to avoid common mistakes , x1liv4.html
- How to customize graphs, hs330.html
- How to export graph images or data to other programs, exporting_graphs.html
- How to identify and remove rogue scans, rogue.html

    How to optimize data acquisition for g(s*) analysis, ni.stz.html
    How to print graphs, print_graphs.html
    How to remove bad data points from scan files, c2asgg.html
    How to zoom with the mouse, 1jdo8b2.html
ID points button, ID_points.html
Improved method, hlp_select_model1.html
Increment reduction factor, hlp_increment_reduction.html
Initial guesses, hs1150.html
Intensity scans, intensity.html
Interference scans, 1_9pj8m.html, 6r6460.html, hs790.html, Clav_tutorial.html
Jitter, 1_9pj8m.html
Jitter removal options, 1_9pj8m.html
Jitter wizard, jitter_wizard.html
Load every Nth scan, hs860.html
Load new scan files button, hs800.html
Load raw scans - Add more scans for this sample button, hs820.html
Load raw scans - Data loading options, hs860.html
Load raw scans - Delete selected scans button, hs830.html
Load raw scans - Load new scan files button, hs800.html
Load raw scans - Preselect data type and cell check box, hs790.html
Load raw scans - Replace scans for this sample button, hs810.html
Load raw scans - Run Claverie simulation to generate data button, hs870.html
Load raw scans - Sample description, hs880.html
Load raw scans - Scans currently loaded in memory list box, hs840.html
Load raw scans - Scans to be used in g(s*) calculation list box, hs850.html
Load raw scans page, hs780.html
Log page, Log_page.html
Main menus, Main menus.html
Mass/diffusion calculator, 7l7or5.html
Meniscus, meniscus.html, 417ij0.html
Meniscus clearance, 417ij0.html
Meniscus position, meniscus.html
Meniscus wizard, men_wizard.html
Menus
    Edit menu, hs60.html
    File menu, 338q0xc.html
    Help menu, 1ns5.4n.html
    Main menus, Main menus.html
    View menu, hs100.html

Window menu, hs130.html
Mistakes, x1liv4.html
Molecular mass, hlp_select_model_M_or_D.html, 7l7or5.html
Molecular mass units, f.a3iy.html
Monte Carlo, Graph_bootstrap_or_Monte_Carlo_results.html
Monte Carlo method, hlp_results_x6.html, About_Monte_Carlo.html
More >> and << Less buttons, hs180.html
More >> and Less << buttons, hs180.html
Navigation tree, hs40.html
Next and Back buttons, hs170.html
Next button, hs170.html
Normalize button, hs1030.html
Number of species, hlp_select_model.html
Number-average sedimentation coefficient, 22avez4.html
Offset, faq6.html
Open, 3ze3b.html
Open saved analysis..., 3ze3b.html
Options dialog, m151qg.html
Out-of-bounds, rdfil..html
Overlay plot, hs1440.html
Overview, 5mvbyz0.html
Parameter bounds, rdfil..html
Parameter constraints, 11a2ula.html
Parameter cross-correlation, hlp_results_x8.html, hlp_results_x4.html
Parameter errors, hlp_results_x6.html
Parameter limits, rdfil..html
Parameter precision, hlp_convergence_factor.html, hlp_results_x6.html
Parameter value text boxes, hlp_input_parameter.html
Partial specific volume, hs1090.html
Peak broadening limit, 9dszbc.html
Perform fit - 1 iteration button, hlp_fit_monitor2.html
Perform fit - Cancel button, hlp_fit_monitor3.html
Perform fit - Do fit button, hlp_fit_monitor1.html
Perform fit - Simulate button, hs1265.html
Perform fit - Status, hlp_fit_monitor4.html
Perform least-squares fit page, hs1240.html
Pitfalls, x1liv4.html
Preferences, m151qg.html
Preselect, hs790.html

Print graph, exporting_graphs.html

Print graphs, print_graphs.html

Print report, hlp_results_x7.html

Program citation, faq8.html

Program organization, 5mvbyz0.html

Program overview and organization, 5mvbyz0.html

Pseudo-absorbance, intensity.html, 1_9pj8m.html

Reference 1, krn4g8.html

Reference 10, 3_sixl.html

Reference 11, Reference_11.html

Reference 12, reference_12.html

Reference 13, ref13.html

Reference 14, reference_14.html

Reference 2, hg3u9i.html

Reference 3, 24ylhkp.html

Reference 4, hg3u9k.html

Reference 5, 3_sj2k.html

Reference 6, hg3u9m.html

Reference 7, hg3u9n.html

Reference 8, 3_sixt.html

Reference 9, hs590.html

Reference information

- About the bootstrap method, About_bootstrap.html
- About the fitting algorithm, cii9q4.html
- About the fitting functions, 2baj0ix.html
- About the jitter wizard, jitter_wizard.html
- About the meniscus wizard, men_wizard.html
- About the Monte Carlo method, About_Monte_Carlo.html
- About this g(s*) implementation, cii9uv.html
- About using radial intensity scans, intensity.html
- Symbols used, 12ndrxp.html

Replace scans for this sample button, hs810.html

Report fit results - Compute confidence intervals button, hlp_results_x6.html

Report fit results - Create reports group box, hlp_results_x7.html

Report fit results - Fit statistics, hlp_results_x3.html

Report fit results - Fitted parameters, hlp_results_x.html

Report fit results - Fixed or constrained parameters, hlp_results_x2.html

Report fit results - Graph bootstrap or Monte Carlo results , Graph_bootstrap_or_Monte_Carlo_results.html

Report fit results - High cross-correlation warning, hlp_results_x4.html
Report fit results - View parameter cross-correlations button, hlp_results_x8.html
Report fit results page, hs1360.html
Reports, hlp_results_x7.html
Reset zero, e.ppbj.html, hs995.html
Residual plot, hs1440.html
RIN, 1_9pj8m.html
Rogue scans, rogue.html
Rule of thumb, 9dszbc.html
s(20,w), 12uc04r.html
s*, faq1.html
Sample description, hs880.html
Save, 16r_0yl.html, Save analysis as....html
Save analysis, 16r_0yl.html
Save analysis as, Save analysis as....html
Save analysis as..., Save analysis as....html
Scans currently loaded list box, hs840.html
Scans to be used in g(s*) calculation list box, hs850.html
sections, _of_sections_updown_spinner.html
Sedimentation coefficient, faq1.html
Select fitting model - Constrain M or s ratios among species, 11a2ula.html
Select fitting model - Data choice, hs1100.html
Select fitting model - Density and partial specific volumes, hs1090.html
Select fitting model - Fit diffusion or mass, hlp_select_model_M_or_D.html
Select fitting model - Improved or classic methods, hlp_select_model1.html
Select fitting model - Molecular mass units, f.a3iy.html
Select fitting model - Number of species, hlp_select_model.html
Select fitting model - Weighting of data points, hlp_fit_xsxd.html
Select fitting model page, hs1070.html
Select scans to analyze - # of sections up/down spinner, _of_sections_updown_spinner.html
Select scans to analyze - Auto adjust button, autoadjust.html
Select scans to analyze - Peak broadening limit, 9dszbc.html
Select scans to analyze - Use true mean time button, use_mean.html
Select scans to analyze page, hs950.html
Set defaults, m151qg.html
Set initial guesses on graph page, hs1150.html
Set integration region, hs1040.html
Set meniscus & data region - Data displayed options, hs940.html
Set meniscus & data region - Fringe jump removal options, 6r6460.html

Set meniscus & data region - Good data begin X cm from meniscus, 417ij0.html

Set meniscus & data region - Good data end at radius =, cellbase.html

Set meniscus & data region - ID points button, ID_points.html

Set meniscus & data region - Jitter removal options, 1_9pj8m.html

Set meniscus & data region - Meniscus position, meniscus.html

Set meniscus & data region - Zoom meniscus button, zoom_men.html

Set meniscus & data region page, hs890.html

Set range for g(s*), hs990.html

Set/alter fitting parameters - Cancel changes button, hlp_cancel_x.html

Set/alter fitting parameters page, hs1160.html

Should I fit the offset?, faq6.html

Show integrals over distribution, 22avez4.html

Simulate button, hs1265.html

Simulation, hs870.html, Clav_tutorial.html

Solvent density, hs1090.html

Species plot, hs1440.html

Standard and advanced modes, hs200.html

Standard mode, hs200.html

Statistics, hlp_results_x3.html

Summary plot, hs1440.html

Symbols used, 12ndrxp.html

Temperature equilibration, ni.su2.html

Temperature plot, hs940.html

Toolbar, hs30.html

Tools menu, Tools_menu.html

True mass, f.a3iy.html

True mean time, use_mean.html

Tutorial: Calculating the g(s*) distribution, jogtbx.html

Tutorial: How to select the scans to analyze, ni.su1.html

Tutorial: Multi-species fits, cz2ewo.html

Tutorial: Simulating experiments with the Claverie simulator, Clav_tutorial.html

Tutorial: Single species fit, fxhaf0.html

Tutorial: Using the g(s*) overlay graph, hs770.html

Tutorials

- Calculating the g(s*) distribution, jogtbx.html
- Claverie simulation, Clav_tutorial.html
- Multi-species fits, cz2ewo.html
- Single species fit, fxhaf0.html
- Tutorial: Calculating the g(s*) distribution, jogtbx.html

  Tutorial: How to select the scans to analyze, ni.su1.html
  Tutorial: Multi-species fits, cz2ewo.html
  Tutorial: Simulating experiments with the Claverie simulator, Clav_tutorial.html
  Tutorial: Single species fit, fxhaf0.html
  Tutorial: Using the g(s*) overlay graph, hs770.html
  Using the g(s*) overlay graph, hs770.html
Use true mean time button, use_mean.html
User interface elements
  Analysis window, hs150.html
  Clone analysis to new window, Clone analysis.html
  Derive g(s*) distribution branch, hs210.html
  Edit menu, hs60.html
  File menu, 338q0xc.html
  Fit distribution as N species branch, 3b9ugl.html
  g(s*) overlay graph, hs760.html
  Help menu, 1ns5.4n.html
  Log page, Log_page.html
  Main menus, Main menus.html
  Mass/diffusion calculator, 7l7or5.html
  More >> and << Less buttons, hs180.html
  Navigation tree, hs40.html
  Next and Back buttons, hs170.html
  Open saved analysis..., 3ze3b.html
  Options dialog, m151qg.html
  Save analysis, 16r_0yl.html
  Save analysis as..., Save analysis as....html
  Toolbar, hs30.html
  Tools menu, Tools_menu.html
  User prompt area, hs190.html
  View menu, hs100.html
  Windows menu, hs130.html
User prompt area, hs190.html
Vbar, hs1090.html
View menu, hs100.html
View parameter cross-correlations, hlp_results_x8.html
Weight-average sedimentation coefficient, 22avez4.html, hs1040.html, hs1045.html
Weighted fit, hlp_fit_xsxd.html
What constitutes a good fit?, faq4.html
What does high parameter cross-correlation mean?, faq3.html

What is the difference between g(s*) and g^(s*)?, faq2.html
What is the difference between s* and s?, faq1.html
What is the difference between the broad and conventional g(s*) algorithms?, faq5.html
Why do I get different answers when fitting g(s*) versus dc/dt?, faq7.html
Windows menu, hs130.html
X axis point spacing, 1lelyh.html
Z-average sedimentation coefficient, 22avez4.html, hs1040.html, hs1045.html
Zoom, 1jdo8b2.html
Zoom meniscus button, zoom_men.html