SEDANAL

USERS' MANUAL

Walter F. Stafford
and
Peter J. Sherwood

Analytical Ultracentrifugation Research Laboratory
Boston Biomedical Research Institute
64 Grove Street
Watertown, MA 02472

stafford@bbri.org
sherwood@bbri.org

17 May 2010
SEDANAL v5.5x

[A4]

TABLE of CONTENTS

1 Introduction 5

1.1 Background 5

1.2 The SEDANAL software package 5

1.3 Software modules and MAIN MENU 6

1.3.1 ModelEditor (a separate program) 6
1.3.2 Preprocessor Screen 6
1.3.3 Control Screen and control file creation 6
1.3.4 DCDT and WDA Screens 6
1.3.5 BIOSPIN 6
1.3.6 General Purpose Kinetics Simulator 6
1.3.7 Program and Data Files - paths 6
1.3.8 In summary, the directory structure should look like: 7

2 Modules of SEDANAL 9

2.1 The ModelEditor (a separate program) 9

2.1.1 MODELS 10
2.1.2 One Component - Single Ideal Species Model 10
2.1.3 Two Component - Two Species Ideal Model 11
2.1.4 Three Component - Three Species Ideal Model 11
2.1.5 Two Component - Three Species Interacting Model 12
2.1.6 Two Component - Four Species Interacting Model 13
2.1.7 Three Component - Four Species Interacting Model 14
2.1.8 Hydrodynamic Concentration Dependence and Thermodynamic Non-ideality 15
2.1.9 Single species with non-ideality 16
2.1.10 Two Species with Non-ideality 17
2.1.11 Indefinite self-association reactions 17
2.1.12 Non-interacting systems of fixed known stoichiometry 18
2.1.13 Editing models for the SedEq fitter 19
2.1.14 User Auxiliary Parameters for fitting: 19

2.2 MAIN MENU - Other Modules: Access by Main Menu 21

2.3 Preprocess Centrifuge Data 22

2.3.1 Loading Data 23
2.3.2 Adjusting Data for off-sets and jitter 26
2.3.3 Selection of Scans to be displayed. 29
2.3.4 Selection of scans to be fitted. 29
2.3.5 Summary of the preprocessing procedure: 31
2.3.5.1 Zoom. 31
2.3.5.2 Interference Data. 31
2.3.5.2.1 Remove optical jitter: 31
2.3.5.2.2 Remove integral fringe shifts: 31
2.3.5.3 Absorbance and Interference Data: 32
2.3.5.4 Multi-wavelength Data - 32
2.3.5.5 Preprocessing multi-speed data sets for Wide Distribution Analysis (WDA): 34
2.3.5.6 Preprocessing Sedimentation Equilibrium data 35
2.3.5.6.1 Equilbrium run in standard double sector centerpiece. 36
2.3.5.6.2 Multi-channel data with 6 channel centerpieces 38

2.4 Fit Preprocessed Data 39

2.4.1 The Control Screen 39
2.4.1.1 Create a New Control File 39
2.4.1.2 Choose between Analyze Data and Simulate Data 41
2.4.1.3 The Equation Editor 43
2.4.1.4 Start the fitting: 45
2.4.1.5 Advanced Parameters Button 46
2.4.1.5.1 Error Estimation Control 46
2.4.1.5.2 Lamm Equation Solutions: 46
2.4.1.5.3 Kinetic integrator control 48
2.4.1.5.4 Fitting Tab 48

2.4.1.6 Concentration dependence of s and D. 49
2.4.1.7 Selecting data sets to be fitted: 50
2.4.1.7.1 Removing a data set from the fit. 50
2.4.1.7.2 Weighting Factors 51
2.4.1.8 Selecting parameters for the finite element solutions to the Lamm equation. 51
2.4.1.9 Total Number of iterations. 51
2.4.1.10 Extinction coefficients: Global fitting with multiple optical systems - 52
2.4.1.11 Local vs global parameters - changing 54
2.4.1.12 Indefinite Self-association 56
2.4.1.12.1 Isodesmic case: 56
2.4.1.12.2 Isoenthalpic indefinite self-association: 57
2.4.1.13 Constraining the range of parameter values during fitting. 58
2.4.1.14 Exiting the Control Screen to start fitting 59
2.4.1.14.1 Fitting Screen 59
2.4.1.14.2 Screen Dumps 60
2.4.1.14.3 Toggling Plots During Fitting 61
2.4.1.14.4 Displaying the Residuals Bitmap 61
2.4.1.15 Fitting to Sedimentation Equilibrium Data: 62
2.4.1.15.1 Loading the data set - load equilibrium data sets before choosing the model 62
2.4.1.15.2 Choose a model 64
2.4.1.15.3 Enter initial guesses and other parameters 66
2.4.1.15.4 Start the fit 67
2.4.1.15.5 Global fit to a dilution series with linked cells 69
2.4.1.15.6 Global fit with un-linked cells 71
2.4.1.16 Error Analysis Boot-strap with Replacement, Monte Carlo and F-statistics 74
2.4.1.16.1 Error Estimation: Parameter standard deviations. 74
2.4.1.16.1.1 Bootstrap with replacement 74
2.4.1.16.1.2 Monte Carlo Analysis 75
2.4.1.16.2 Error Estimation: Parameter Confidence Limits 77
2.4.1.16.2.1 F-statistics. 77
2.4.1.16.2.1.1 General F-stat preferences 77
2.4.1.16.2.1.2 Turning on F-stats for a particular parameter 77
2.4.1.16.2.1.3 Displaying F-stat progress 78

2.5 DCDT and WDA 78
2.5.1 DCDT 78
2.5.1.1 Switching between g(s*) vs s* and s*.g(s*) vs ln(s*) 80
2.5.1.2 Smoothing of g(s*) vs s* for presntation purposes 80
2.5.2 Wide Distribution Analysis (WDA) 82
2.5.2.1 Bad Scans 84
2.5.2.2 Smoothing after numerical differentiation 84

2.6 Preferences 86
2.6.1 General Preferences 86
2.6.2 Preprocessor preferences 87
2.6.3 DCDT preferences 88
2.6.4 Weighting Factor preferences - Control Tab 90
2.6.5 Error estimation control: Advanced, Control Extended tab 92
2.6.6 Reports: Output, Control Extended tab 92
2.6.7 The other "Output …" tabs 93
2.6.7.1 Minima 93
2.6.7.2 Concentration as a function of radius 94
2.6.7.3 Concentration as a function of time for the initial equilibration step 94
2.6.7.4 Disposition of various log files 95
2.6.7.5 Preferences for indefinite self-association models 95
2.6.8 Fitting preferences 97
2.6.9 Simulation Preferences 97

2.7 BIOSPIN 98

2.8 SIMULATING DATA 98

2.9 Other features: 100
2.9.1 Output of initial reaction time course. 100

2.10 Kinetics Simulator *100*
2.11 Scripting sedimentation fitting and simulation *105*
2.11.1 Scripting commands: 105
2.11.2 Specifying control files in scripts 106
2.11.3 Storing scripting control files 106
2.12 Help *107*
2.13 Exit *107*
3 Background Theory 108
3.1 Sedimentation Velocity Theory: *108*
3.1.1 Concentration time-difference curves. 108
3.2 Sedimentation Equilibrium Theory *111*
3.2.1 Ideal Case: 111
3.2.1.1 Monomer-dimer: 111
3.2.1.2 Hetero-Association: 112
3.2.2 Non-ideal Case: 113
4 APPENDIX A: CHANGE LOG. 114
5 INDEX 115

1 Introduction

SEDANAL is a computer program for analysis sedimentation velocity and sedimentation equilibrium data that are produced by the BeckmanCoulter Optima Series XLA/I ProteomeLab[TM] analytical ultracentrifuge system.

1.1 Background

Starting from the raw data files produced by absorbance, interference or fluorescence experiments, the SEDANAL sedimentation velocity module (**SedVel)** uses numerical solutions to the Lamm equation and chemical kinetics models to fit parameters to the experimental data. SEDANAL can also produce simulated data, in the same format as the raw data files produced by the XLA/I system. SEDANAL fits to time difference curves to remove time invatiant systematic errors from the data completely. It also allows preprocessing of the data to essentially eliminate time dependent, radially independent systematic instrumental errors characteristic of the interference optics as well as integral fringe jumps that can occur at the meniscus.

The SEDANAL sedimentation equilibrium module (**SedEq)** fits sums of exponentials to sedimentation equilibrium data to obtain estimates of molecular masses and equilibrium constants for interacting and relative amounts for non-interacting systems

The output of the SEDANAL **SedVel** curve fitter is a file containing the root mean squared deviation (rmsd) and best-fit values for the parameters, as well as the experimental and calculated difference curves and the residuals of the fit for each of the difference curves. One specifies which parameters are to be fitted and which to be held constant.

The output of the SEDANAL **SedEq** curve fitter is a file containing the root mean squared deviation (rmsd) and best-fit values for the parameters, as well as the experimental and calculated equilibrium concentration curves and the residuals of the fit for each of the curves. One specifies which parameters are to be fitted and which to be held constant.

1.2 The SEDANAL software package

The software package is very CPU-intensive: it runs best on the fastest possible CPU with lots of memory and at least 100 Mb of the hard drive space for installation. A Pentium 4, 1.6GHz, with 512 Kbytes of memory running Windows98 is an OK starting point. A less frustrating setup would be a Pentium 4, 3.4 GHz, with 2 Gigabytes of memory running Windows XP Pro and is even more fun.

1.3 Software modules and MAIN MENU

1.3.1 ModelEditor (a separate program)

Create reaction schemes
Choose default parameters to be fit

1.3.2 Preprocessor Screen

Read in a data set
Designate whether it is a **SedVel** or a **SedEq** run.
Eliminate systematic error
Choose meniscus and base
Choose radial fitting range
Interpolated optical blank subtraction can be carried out on equilibrium data
Save preprocessed data in a "cell data file" (*.abr) for fitting and DC/DT analysis.

1.3.3 Control Screen and control file creation

Read either an existing Control file or Select "New"
Select either "Analyze data" or "Simulate data"
Chose a model
Choose data sets to be fit and scans to be included in the fits.
Enter fitting/simulating parameters (# of points, time increment,...)
Enter initial guesses
Fit the model to the data

1.3.4 DCDT and WDA Screens

Standard time derivative g(s*) analysis (DCDT)
Multi-speed, wide distribution analysis (WDA)

1.3.5 BIOSPIN

Computes number weight and z-average molecular weight as a function of local cell concentration in a sedimentation equilibrium run. Also computes M_{Y1} and M_{Y2}, molar mass moments independent of the second virial coefficient.

1.3.6 General Purpose Kinetics Simulator

A general purpose kinetics simulator can be accessed from the Main Menu and can be used to simulate any reaction scheme that can be represented in the ModelEditor.

1.3.7 Program and Data Files - paths

SEDANAL expects to find a specific path structure for files, (this can be changed to some extent in the Preferences menu), shown in the preferences menu.

The top level directory (also called a "folder") may be named anything you wish; we'll call it "SEDANAL". In the "SEDANAL" directory, create three subdirectories called "**ModelEditor**", "**Program**" and "**User_data**". These names must be written exactly as indicated.

The application "**ModelEditor.exe**" and associated file "**ModelInfo.txt**" must be placed in the "**ModelEditor**" directory. The application "**SEDANAL.exe**" must be placed in the "**Program**" directory along with "**Preferences.txt**" If preferences.txt is omitted, SEDANAL will create a default preferences file.

The "**User_data**" directory will contain sub-directories in **3 levels**. At the highest level within "**User_data**", any number of experiment directories can be created. Each experiment directory, in turn, **must** contain the data directories in the standard BeckmanCoulter XLA/I date/time format. See Figure 1 for an example. Usually one will store control files (which have extension *.**abc**) in an experiment folder. In this case (see example below) it would be the "experiment 1" folder. The run files (with extension ***.abr**) are automatically stored at the level of the XLA/I data files.

1.3.8 In summary**, the directory structure should look like:**

C:\SEDANAL\ModelEditor\ModelEditor.exe
C:\SEDANAL\ModelEditor\ModelInfo.txt
C:\ SEDANAL \Program\ SEDANAL nnn.exe
C:\ SEDANAL \Program\Preferences.txt
C:\ SEDANAL \User_data\my_experiment#1\mmddyy\hhmmss*.IPn
C:\ SEDANAL \User_data\my_experiment#2\mmddyy\hhmmss*.Ipn

See next page for a graphical representation of this path structure:

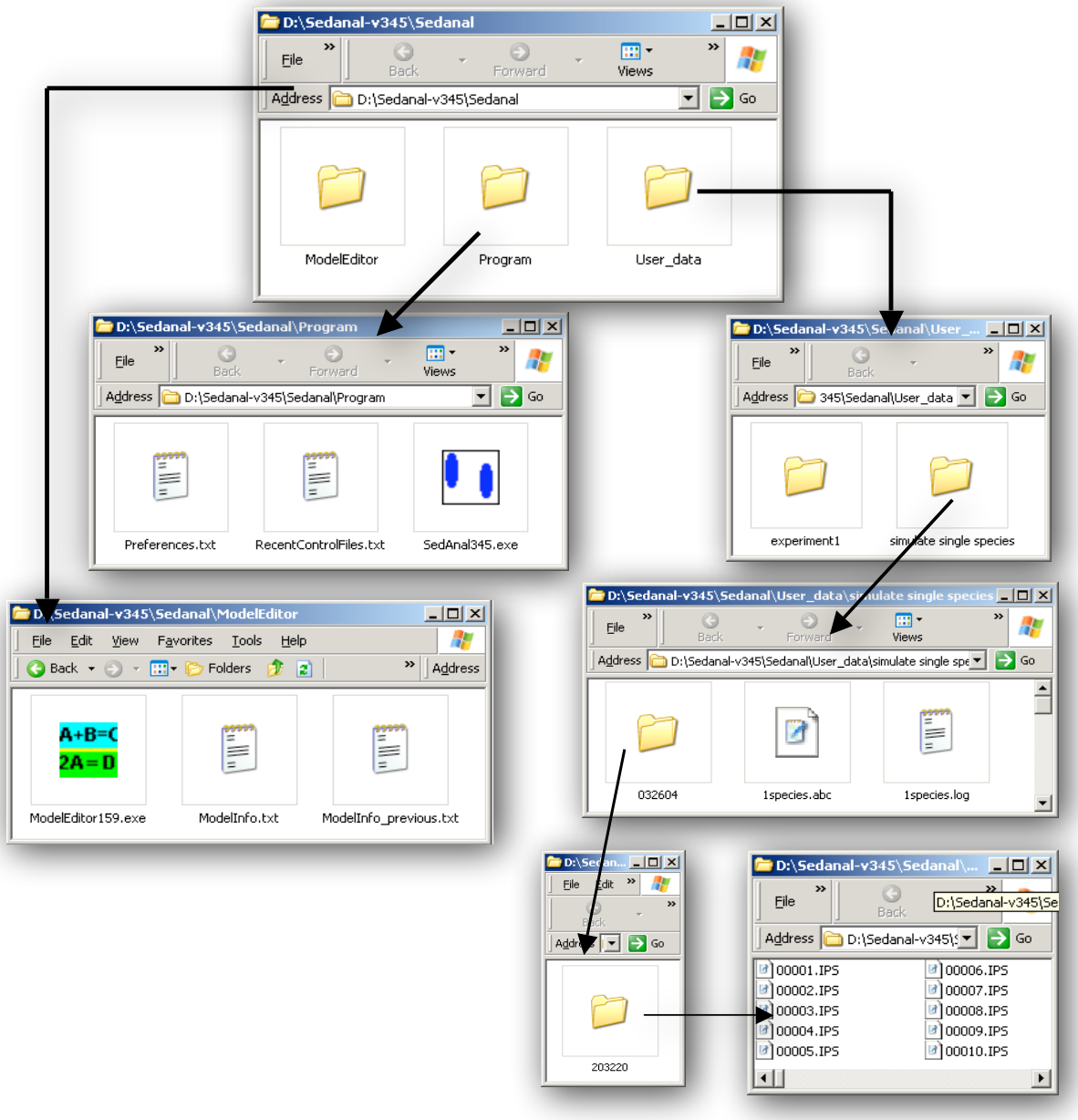


Figure 1-1. Diagram of the default path structure for the SEDANAL suite of programs.

2 Modules of SEDANAL

2.1 The *ModelEditor* (a separate program)

The Model Editor allows one to create and edit various models and reaction schemes. It is also where the default fitting parameters and their default limits are chosen. (The defaults can be adjusted later if necessary on the Control Screen.)

Models may consist of any combination of reaction schemes ranging from a single species to combinations of hetero- and self- associations. It will allow up to 10 thermodynamic components and up to 28 species related by up-to 27 chemical reactions.

When the ModelEditor is first invoked, it starts up with an empty box (upper left hand corner) in which the model name is entered as a text string. The drop-down window will show the other models that have been entered previously. The new model will be added to this list when the "STORE" button is clicked or the "EXIT" button is selected. The user will select one of these models from the Control Screen.

The three small boxes to the right of the Model Name Window allow the selection of the number of chemical reactions and the number of species.

The parameters include:

- molar mass
- sedimentation coefficient
- density increment, $(d\rho/dc)_{T,\mu 3}$ which for most practical purposes is just $(1-v\rho)$,
- or partial specific volume (actually the product of $v\rho$, as in $(1-v\rho)$, since the density is set to 1.000 internally)
- mass extinction coefficient, either A.U.-$(mg/mL)^{-1}$ or fringes-$(mg/mL)^{-1}$ for a 12 mm optical path length.
- concentration dependence of the frictional coefficient., $k_{s,}$ $[s(c)=s_o/(1 + k_s*c)]$ or $[D(c)=D_o/(1 + k_s*c)]$; where k_s, in this case, applies to both s and D.
- thermodynamic non-ideality expressed through the second virial coefficient, BM_1, as in the expression $D(c)=D_o(1 + 2BM_1*c)$

The simplest model is a single ideal species. In this case the parameters of interest are the sedimentation coefficient and molar mass (or diffusion coefficient). A screen dump of the ModelEditor for a single species is shown below. One species has been selected in the species box.

In the lower left portion of the window is a list of molecular parameters and their scale factors. In this example, the first parameter is the molar mass and its scale factor has been set to unity. Therefore, future references to this parameter must be in units of daltons or grams/mole. If one had entered 1000 in the scale factor field, future references to the molar mass would be in kilodaltons or kg/mole. The second parameter is the sedimentation coefficient and has been scaled to $1x10^{-13}$ so that future references to it must be in units of **svedbergs**. At the same time it would also be a good idea to rename the parameter to reflect the new units; for example, you might change it to "Sedimentation coefficient, (S)". The third parameter is either the density increment or partial specific volume times the density, and the fourth, is the extinction coefficient on the mass

concentration scale, (g/L). For absorbance optics this is the specific extinction coefficient adjusted for the 1.2 cm path length of the centrifuge cell (**i.e. one has to multiply the usual value (normally corresponding to a 1 cm optical path) by 1.2 before entering it into the box**). For interference optics, it is the number of fringes produced by a 1 mg/mL solution and is on the order of 3.29 for typical proteins, but varies depending on whether the macromolecule is a protein, carbohydrate or nucleic acid. Each species may have a different extinction coefficient. SEDANAL will allow data from different optical systems and wavelengths to be combined for global fitting with a different extinction coefficient to be entered for each species in each data set (or cell) obtained with a different optical system or wavelength or with a different path length..

2.1.1 MODELS

2.1.2 One Component - Single Ideal Species Model

The simplest model is a single species with no non-ideality. The necessary parameters are entered into the ModelEditor control screen as shown in Figure 2-1.

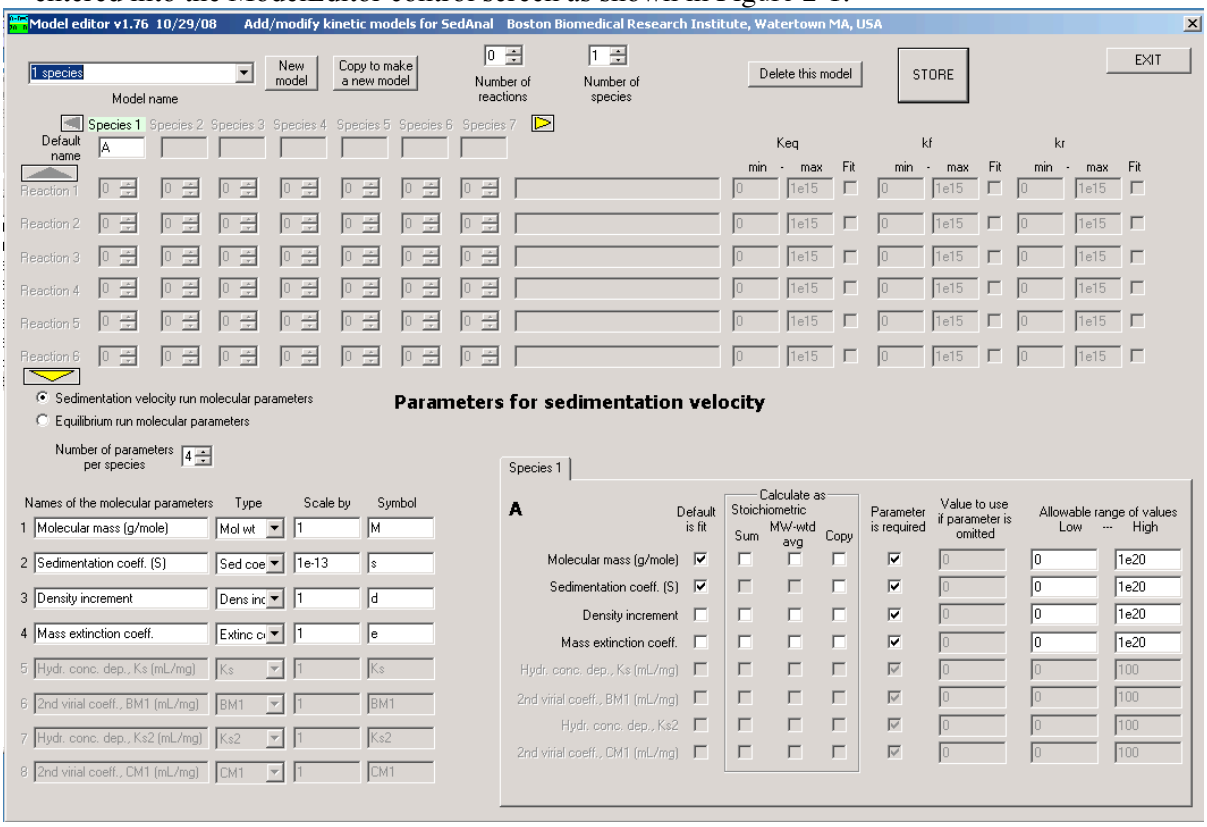


Figure 2-1. **Screen showing parameters for a One Component - Single Ideal Species Model.**

In this example, the model is named "1 comp 1 species". The number of species is "1", the number of reactions is "0", and the number of parameters is “4”. One might want to fit for the molar mass and sedimentation coefficient by default. So, these buttons would be checked. They will appear with a light gray background on the control screen. If they aren't checked here, they will have a blue background on the control screen indicating that the default is to hold them constant. However, their status can be changed later from "hold" to "fit" by right clicking on the corresponding box on the control screen.

The light green high-lighted species are the **components**.

One may also select the allowable range for each parameter to be checked during entry of values on the control screen. Allowed ranges to be used during fitting are set by right clicking on the parameter value on the control screen.

2.1.3 Two Component - Two Species Ideal Model

The next example, "2 comp 2 species", comprises two independent ideal species (i.e. two components) with no interaction between them. Necessary parameters in the ModelEditor control box are shown in Figure 3.

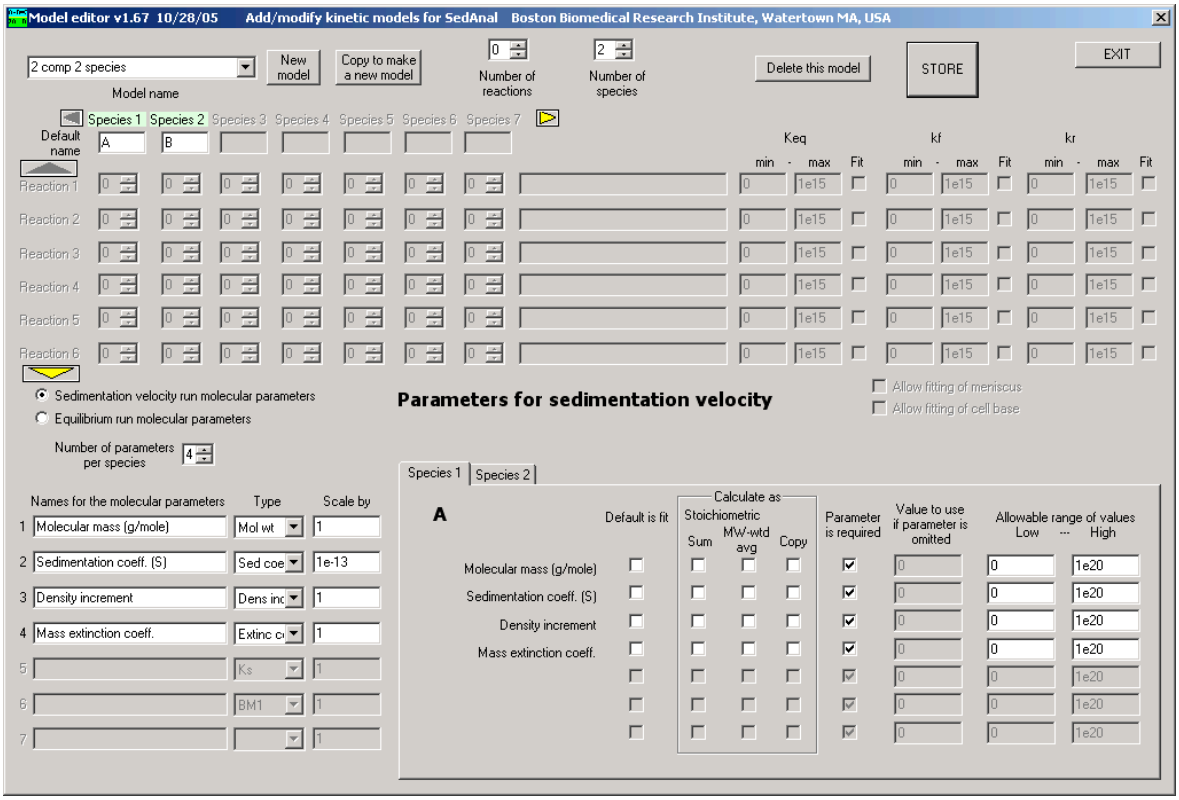


Figure 2-2. Screen showing parameters for the Two Independent Ideal Species Model.

Since we are fitting for A and B as independent species in this example, the model tab for species 2 would be the same as for species 1.

2.1.4 Three Component - Three Species Ideal Model

Similarly, 3 component, non-interacting 3 species system, let's call it "3 comp 3 species", would be designated by choosing 3 species and no reactions. The ModelEditor screen would look like Figure 4 with each of species tabs the same having molecular mass and sedimentation coefficient checked.

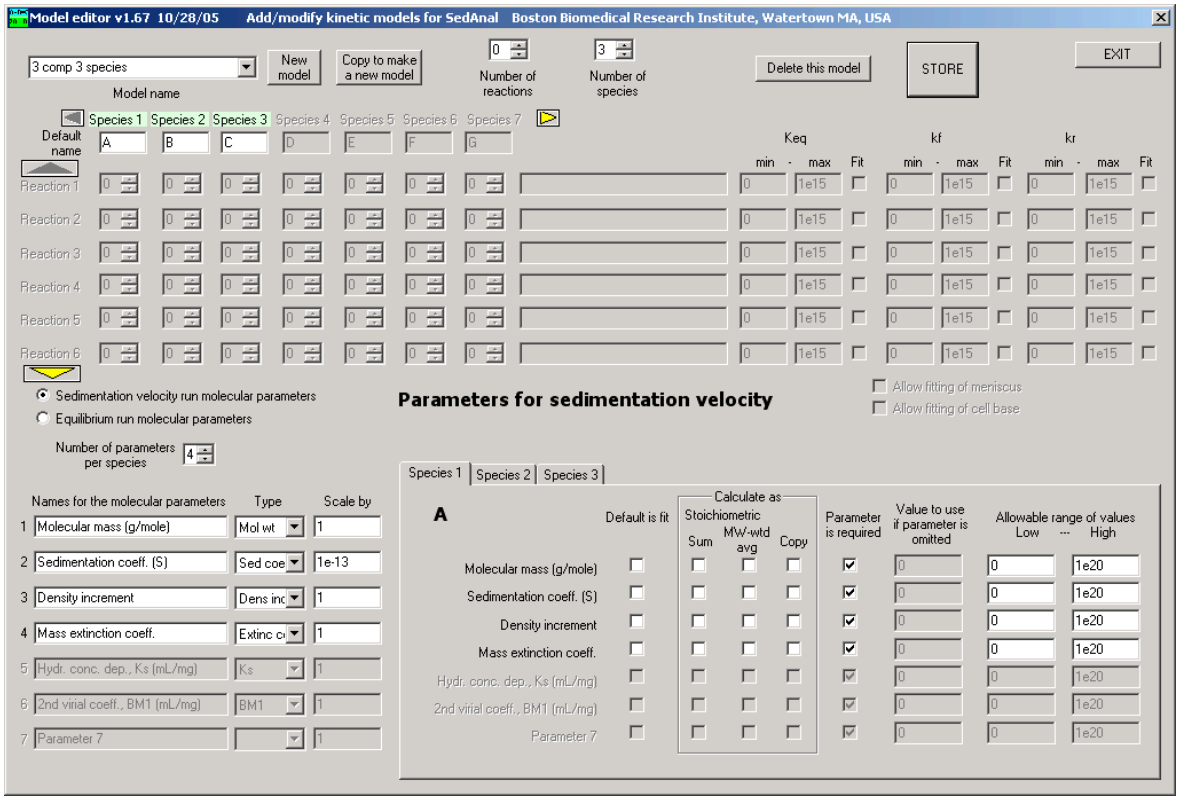


Figure 2-3. Control box showing parameters for the Three Independent Ideal Species Model.

2.1.5 Two Component - Three Species Interacting Model

In this case, there is simple complex formation between species A and species B:

$A + B = C \qquad K_{eq} = k_f/k_r$

Where K_{eq} is the association constant, k_f is the forward rate constant and k_r is the reverse rate constant.

In this case, we would have studied A and B separately, and therefore, would not have to fit for their properties. So none of the “Default is fit” boxes are checked for either A or B (Figure 5). In general, in this case, we would be fitting for the sedimentation coefficient of C as well as the association equilibrium constant for complex formation. Note that “Species 3” is not highlighted because it is not an independent component; the two components are A and B on the "species" tabs. (Note: the two components could have been chosen as A and C but that is not convenient in this case.) The species tab for A and B would be filled in the same way (Figure 5).

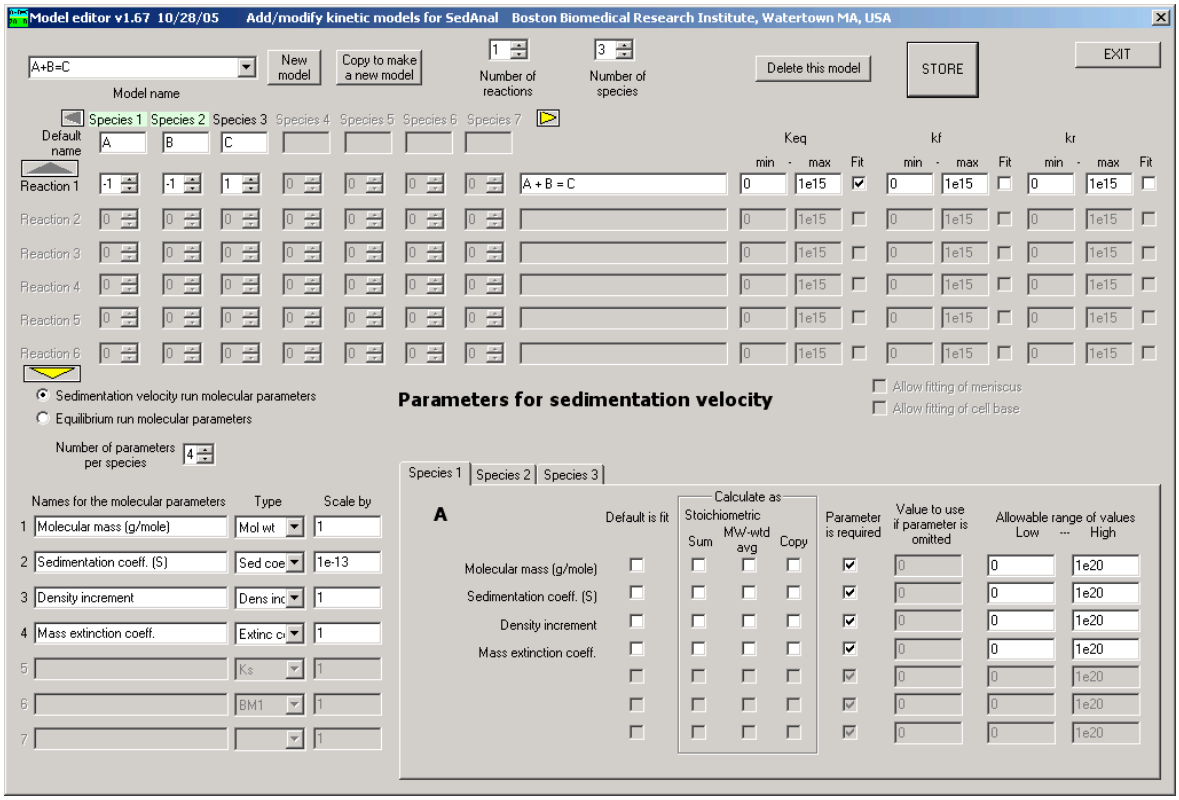


Figure 2-4. Screen showing parameters for the Interacting Species Model

For **Species C**, we have (by selecting the "Species 3" tab) parameters as shown in Figure 2-5.

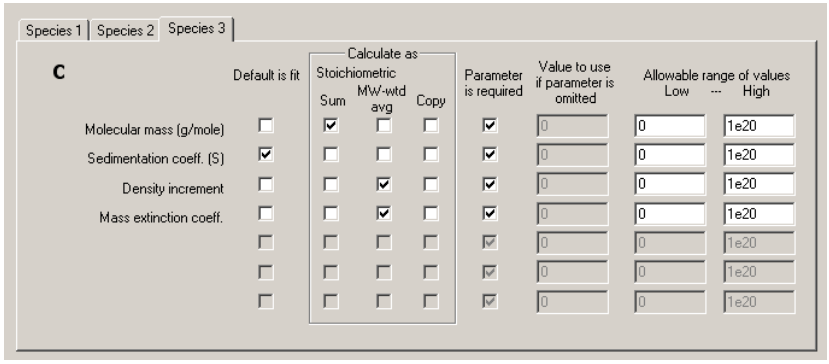


Figure 2-5. Species tab showing parameters for Species 3 ("C").

We have checked the "MW-wtg avg" boxes for Density Increment and Mass extinction coefficient to indicate that these parameters are calculated as weight averages of those for species A and B.

In this example, on the row labeled "Reaction 1", the "fit" box for Keq is checked to indicate that we are going to fit for the equilibrium constant by default. For this particular model, SEDANAL uses an analytical solution to the mass action equations. SEDANAL uses analytical solutions for monomer-dimer and A+B=C.

2.1.6 Two Component - Four Species Interacting Model

A more complicated example is a 2 component system, comprised of 4 species, and 2 reactions. This might represent an antigen-antibody reaction.

$A + B = C \qquad K_1=k_{1f}/k_{1r}$
$C + D = D \qquad K_2=k_{2f}/k_{2r}$

Selection of parameters is shown in Figure 2-6.

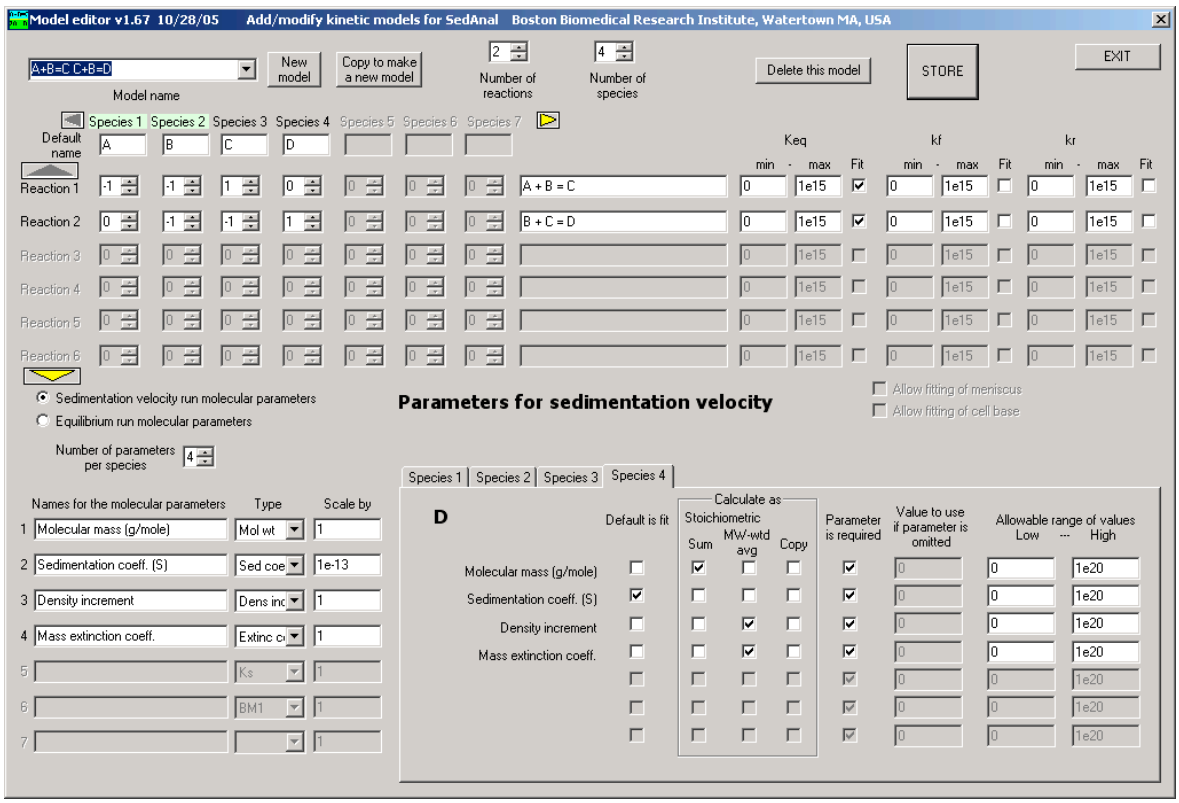


Figure 2-6. Selection of parameters in the control box for the two component - four species interacting model: A + B = C, C + B = D.

NOTE: Setting allowable ranges for the parameters shown in the Model Editor affects only the range allowed be inputted on the Control Screen. One may also select the allowable range for each parameter to be used during the fitting procedure by left-shift clicking in the parameter's box on the control screen. If one of the limits is hit, the function evaluator will return a large value for the r.m.s. deviation for that iteration.

2.1.7 Three Component - Four Species Interacting Model

An example of a 3 component (paragraph 3.1.5), 4 species system would be a system like A+B=C with an impurity or aggregate, D. A and B are the first 2 components and D will be the third. The ModelEditor will correctly understand that D is not participating in a reaction and is, therefore, an additional component.

When this model is chosen (see below), the Control Screen, will show the three components with their appropriate labels. (The species' names can be edited to reflect the three components. For example species C could be renamed, "AB". So the model editor screen would show the third species as AB, as in the figure below.) The figure also shows the tab for species 4 now labeled as D.

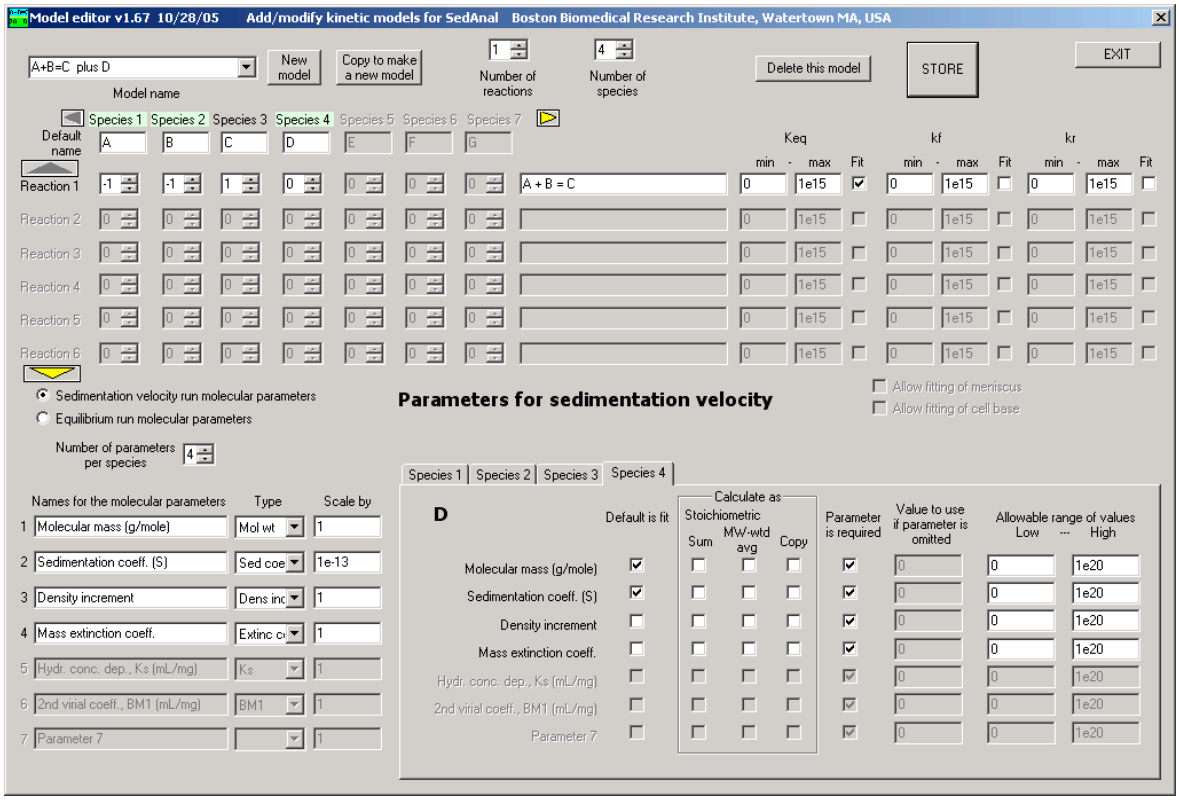


Figure 2-7. Selection of parameters in the control box for three component - four interacting species model.

2.1.8 Hydrodynamic Concentration Dependence and Thermodynamic Non-ideality

Non-ideality is treated through two parameters, K_s, the concentration dependence of the frictional coefficient, and BM_1, the second virial coefficient term and included in the modeling in the following way for sedimentation and diffusion coefficients:

$$s(c) = \frac{s^o}{(1 + K_s c)},$$

$$D(c) = \frac{D^o(1 + 2BM_1 c)}{(1 + K_s c)}.$$

The factor $(1+K_sc)$ represents the concentration dependence of the frictional coefficient. The factor $(1+2BM_1c)$ represents the contribution from thermodynamic non-ideality through the second virial coefficient.

$$s(c) = \frac{s^o}{(1 + k_s c + k'_s c^2)}$$

$$D(c) = \frac{D^o\left(1 + 2BM_1c + 3CM_1c^2\right)}{\left(1 + k_sc + k'_sc^2\right)}$$

We have now added second order non-ideality terms to allow fitting to data at much higher concentrations than before.

2.1.9 Single species with non-ideality

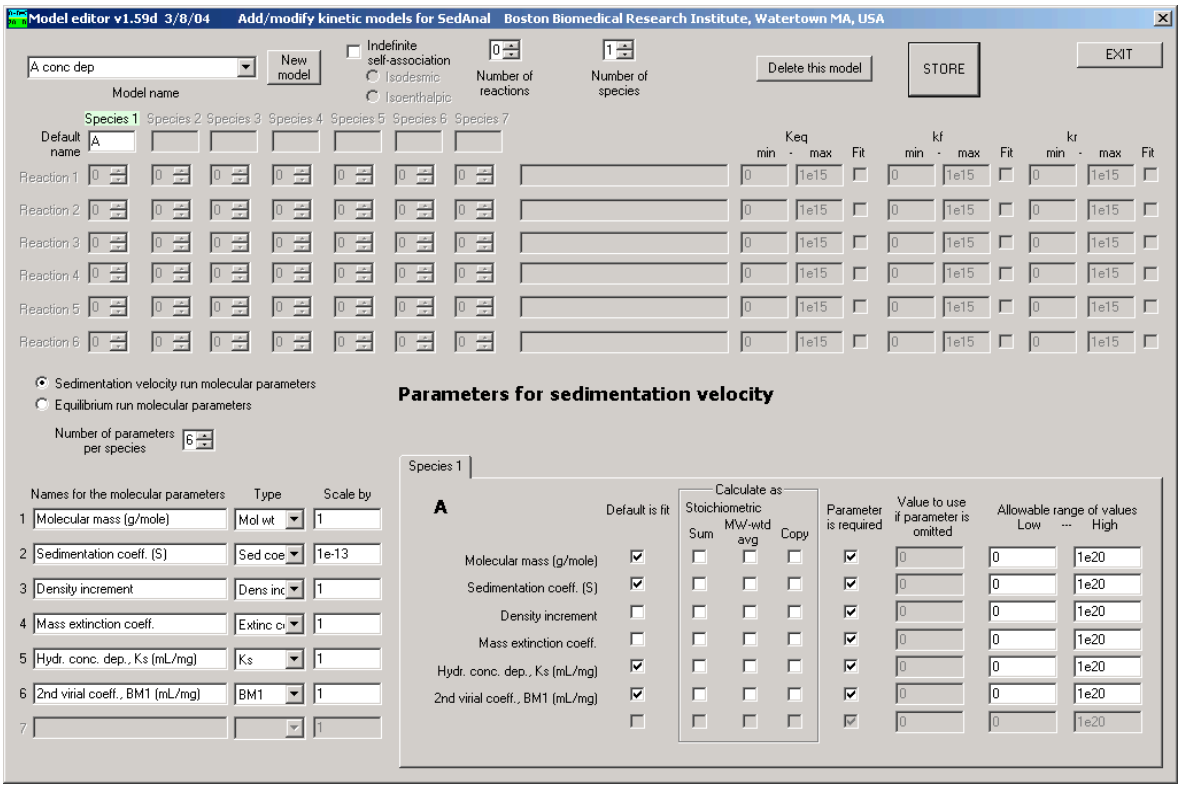


Figure 2-8. Selection of parameters in the control box for single species with non-ideality.

The "Species 1" tab has both Ks and BM1 checked.

2.1.10 Two Species with Non-ideality

The "Species 1" and "Species 2" would both look the same with the Ks and BM1 both checked.

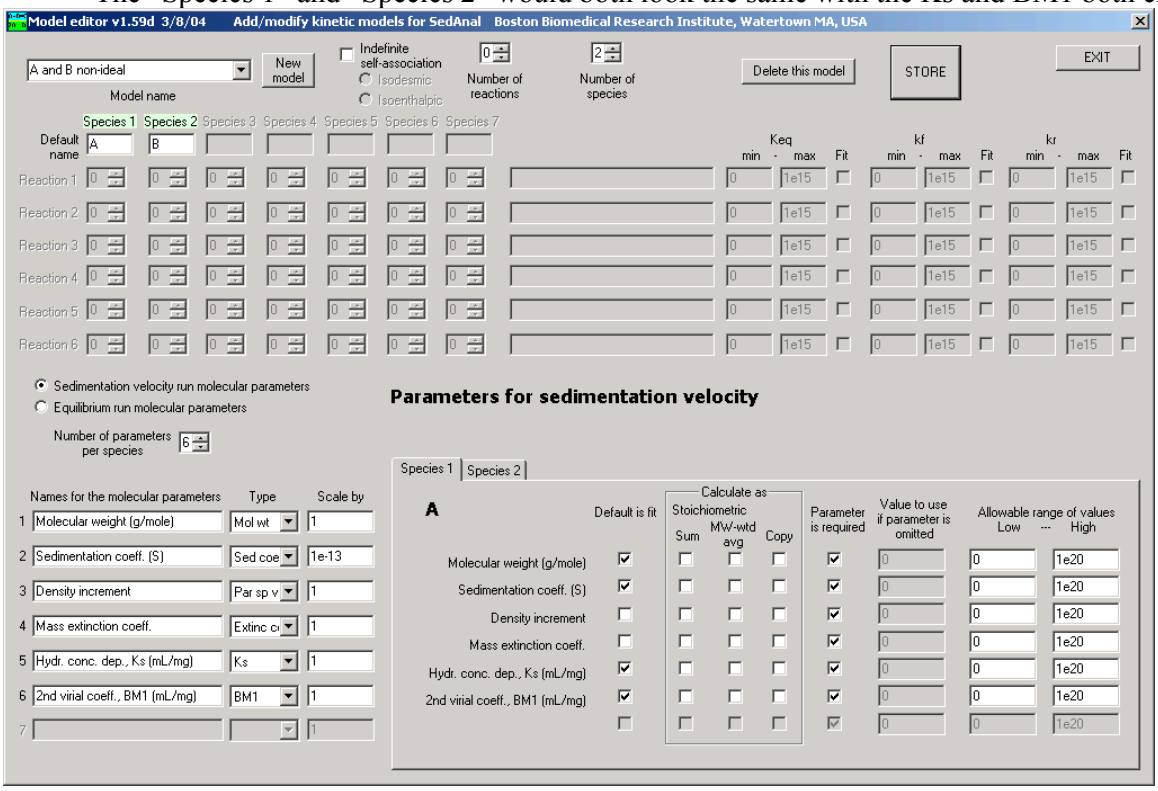


Figure 2-9. Selection of parameters in the control box for two species with non-ideality.

If you suspected that the non-ideality is the same for both species A and B, you would check the "copy" box for "Species 2" and uncheck the "Default is fit" boxes. This would allow for fitting for a single value of Ks and BM1 for both species.

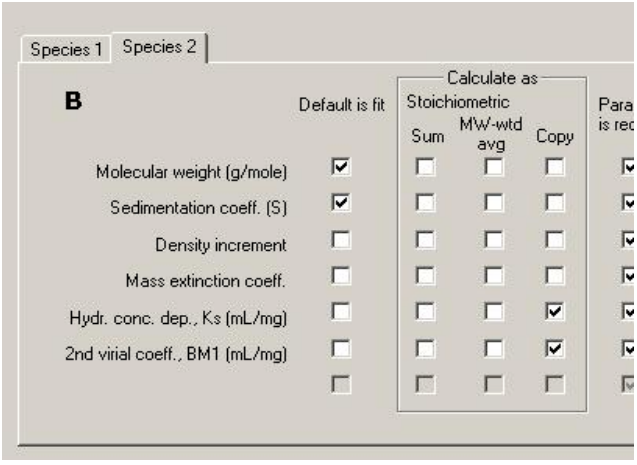


Figure 11. "Copying" values for "Species 2" from "Species 1"

2.1.11 Indefinite self-association reactions

The model editor can create isodesmic and isoenthalpic models for use with SEDANAL. Click on the "Number of reactions" to indicate 1 reaction and the type the letters "sa" in the window below the "Default name" for "species 1", name the reaction either "isodesmic" or "isoenthalpic", click the corresponding button to the right of the row and click the "Store" button.

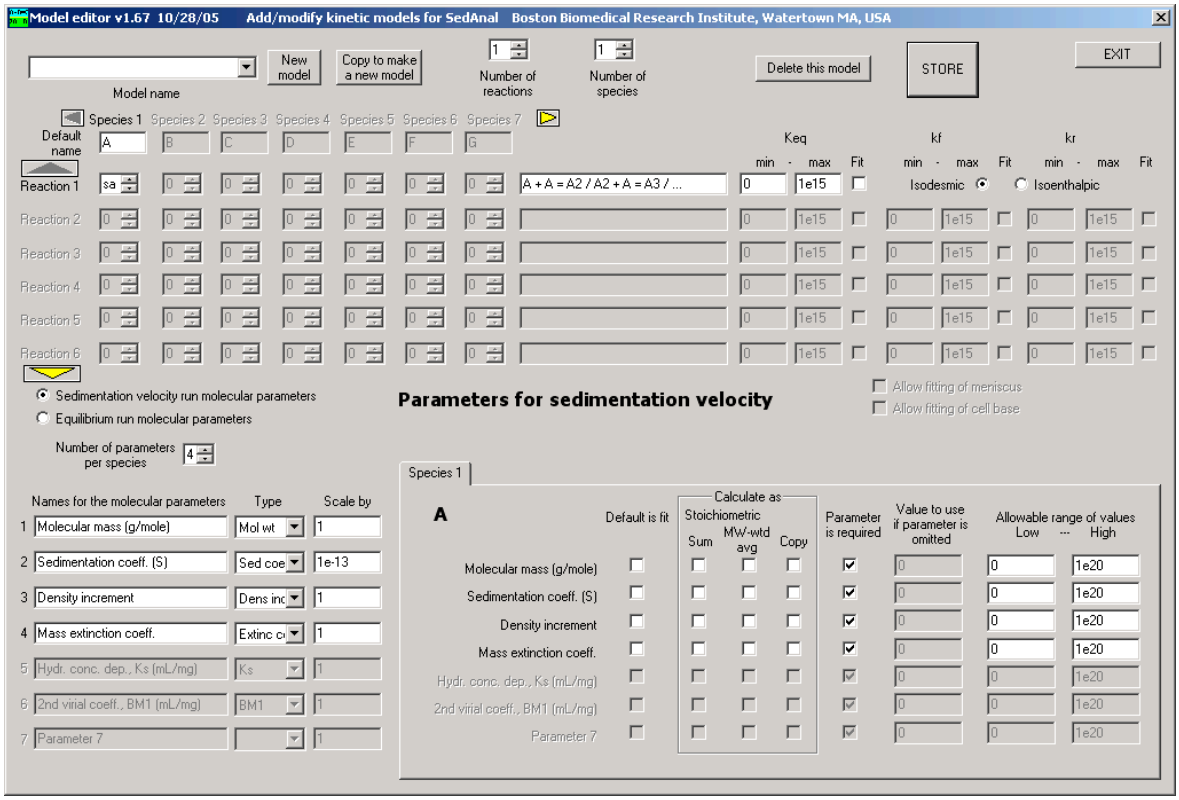


Figure 2-10. Screen for indefinite self-association.

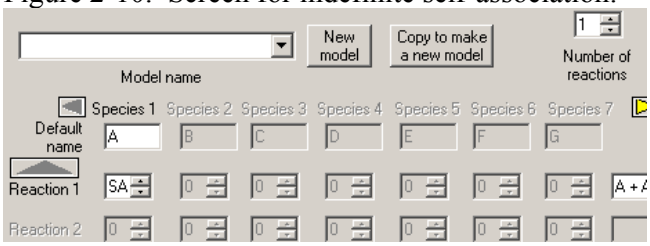


Figure 2-11. Blow-up of upper left of screen

Figure 2-12. Isodesmic case

For the isoenthalpic case (Figure 2-12), the ModelEditor window looks the same except that the "Isoenthalpic" button should be clicked to indicate to the fitter that it should use the equations for that case.

2.1.12 Non-interacting systems of fixed known stoichiometry

Models for non-interacting (i.e. not in rapidly reversible equilibrium) systems composed of components with known molar mass ratios can be treated using the equation editor on the control screen. Fixed relationships between parameters can be established by typing in FORTRAN-like equations in the **Equation Editor Page** ("Eqn" button"). - More on this later when we get to the control screen.

2.1.13 Editing models for the SedEq fitter

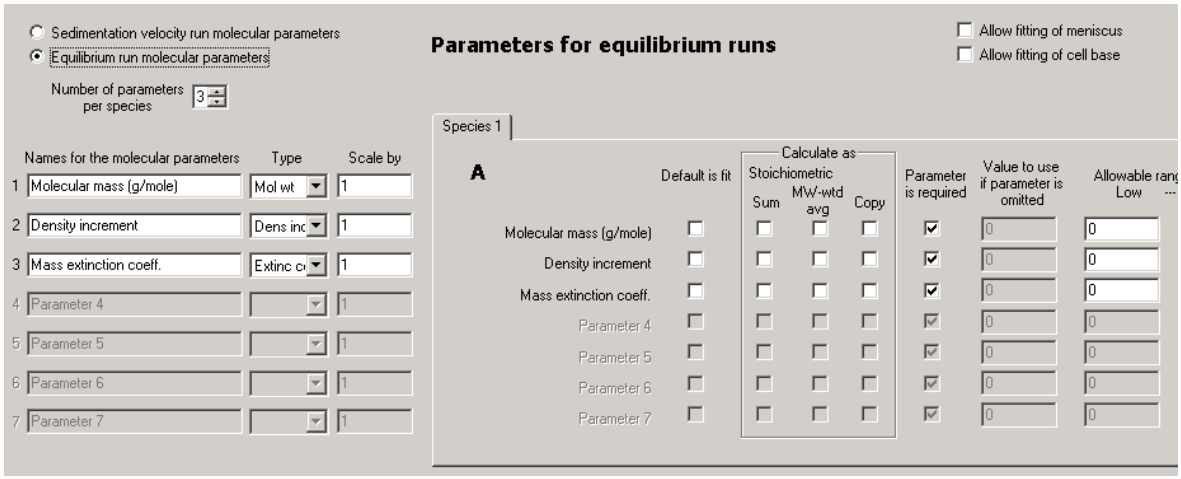


Figure 2-13. SEDANAL will determine from information in the cell data file (*.abr) whether the data are from a SedVel or SedEq run and display the appropriate boxes on the control screen for a particular model after reading in the cell data file. The type of run is **indicated when preprocessing the data. (This shows the bottom half of the model editor screen).**

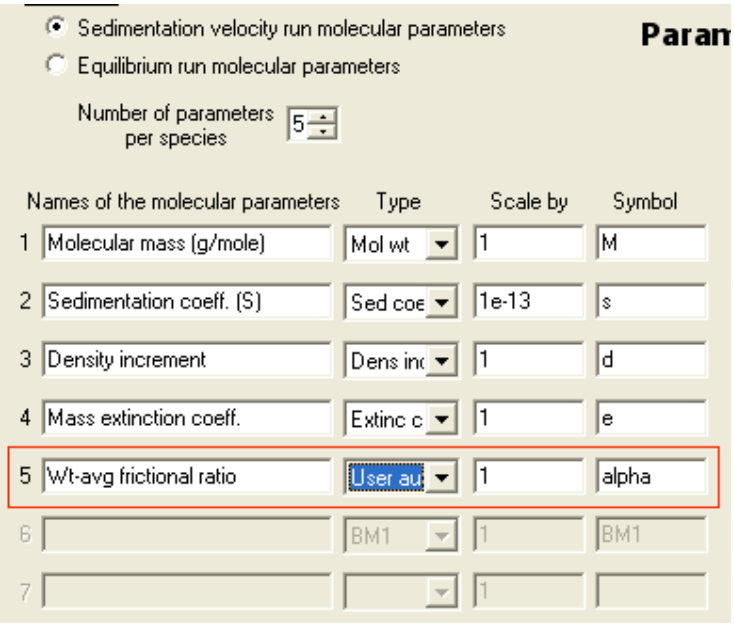
Figure 16. SEDANAL will determine from information in the cell data file (*.abr) whether the data are from a SedVel or SedEq run and display the appropriate boxes on the control screen for a particular model after reading in the cell data file. The type of run is indicated when preprocessing the data. (This shows the bottom half of the model editor screen).

2.1.14 User Auxiliary Parameters for fitting:

The fitting process for sedimentation velocity experiments always uses the molecular parameters molar mass, sedimentation coefficient, density increment (or partial specific volume), and extinction coefficient. For equilibrium experiments, sedimentation coefficients are not used. Other molecular parameters, such as virial coefficients, are optional.

SEDANAL v4.86 introduced a new type of molecular parameter, the *user auxiliary* parameter, which is not used directly for fitting. Instead, it is used indirectly via the equation editor. For example, suppose we want to fit for the weight-average frictional ratio, α, for each species, and then compute the molar mass for each species i as $M_i = 1000\ (3.2\alpha_i s_i)^{3/2}$, with s_i in Svedbergs.

First, we create a model in the MODEL EDITOR (MODEL EDITOR must be v1.70 or later) with a single user auxiliary parameter, which we will name "Wt-avg frictional ratio", entering the symbol `alpha` for it. Now `alpha`(*i*) can be used in the equation editor.



Next, create a control file, specifying the new model. In the equation editor, enter

```
M(1)=1000.*pow(3.2e13*s(1)*alpha(1),1.5)
M(2)=1000.*pow(3.2e13*s(2)*alpha(2),1.5)
M(3)=1000.*pow(3.2e13*s(3)*alpha(3),1.5)
```

Note that the internal values `s(i)` have been converted from sec^{-1} (the units used by SEDANAL) to Svedbergs by multiplying by 10^{13}.

For each iteration during the fitting process, the values for `M` will be computed according to the above equations. The values for `alpha` can be either held constant, or fitted.

The symbol for the auxiliary and other molecular parameters used in the equation editor, `alpha` in this example, is given in the model (in the MODEL EDITOR). Symbol names are made up of letters, digits, and underscores ("_"), with the initial character not a digit. Symbol names are case-sensitive, so, e g, `alpha` is different from `Alpha`.

The symbol names must be unique, and cannot conflict with the symbol names used for other fitting parameters; currently these are `K`, `kf`, `kr`, `L`, `r`, `rw`, `m`, `b`, `y` (for equilibrium runs only), `p`, `v` and `T`.

2.2 MAIN MENU - Other Modules: Access by Main Menu

When one launches SEDANAL, one gets the Main Menu with 10 choices, see Fig. 17.

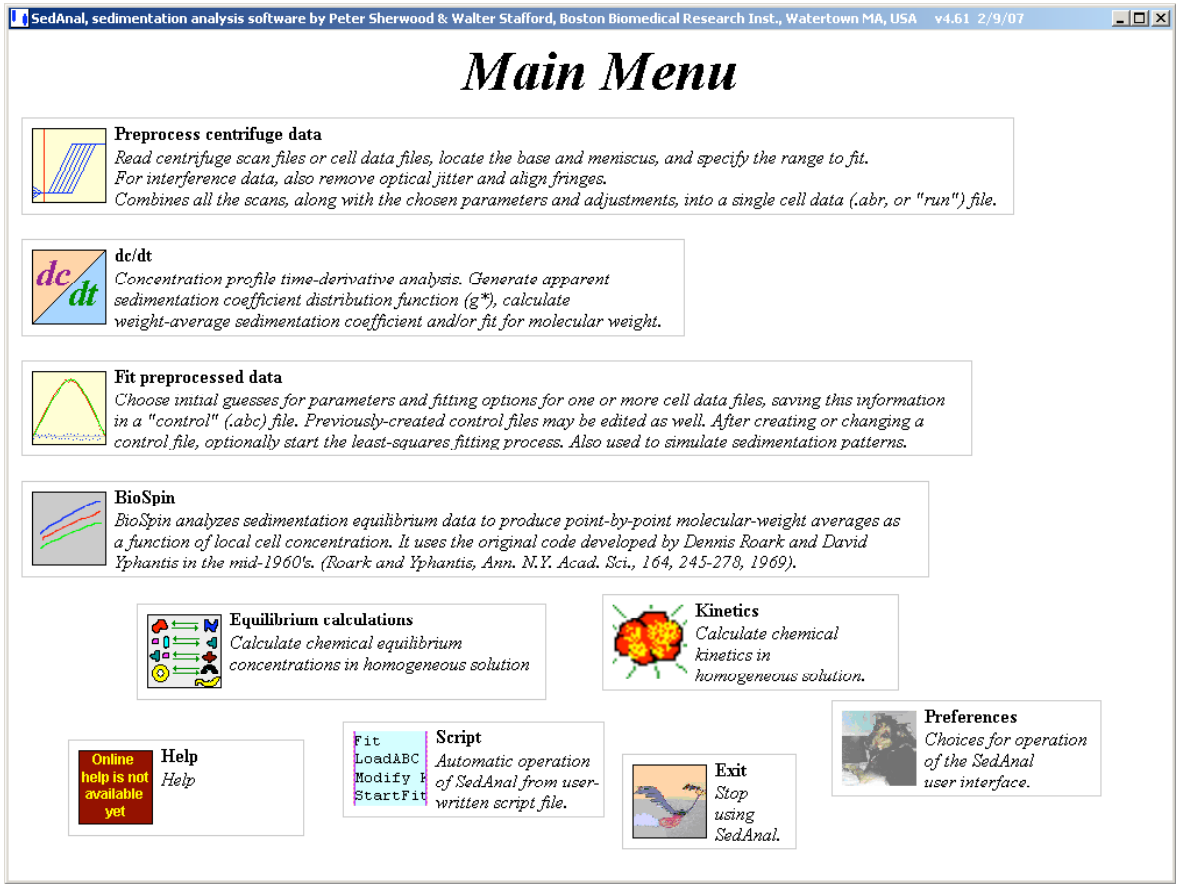


Figure 17. The Main Menu

2.3 Preprocess Centrifuge Data

The preprocessor screen presents the following window (Figure 18).

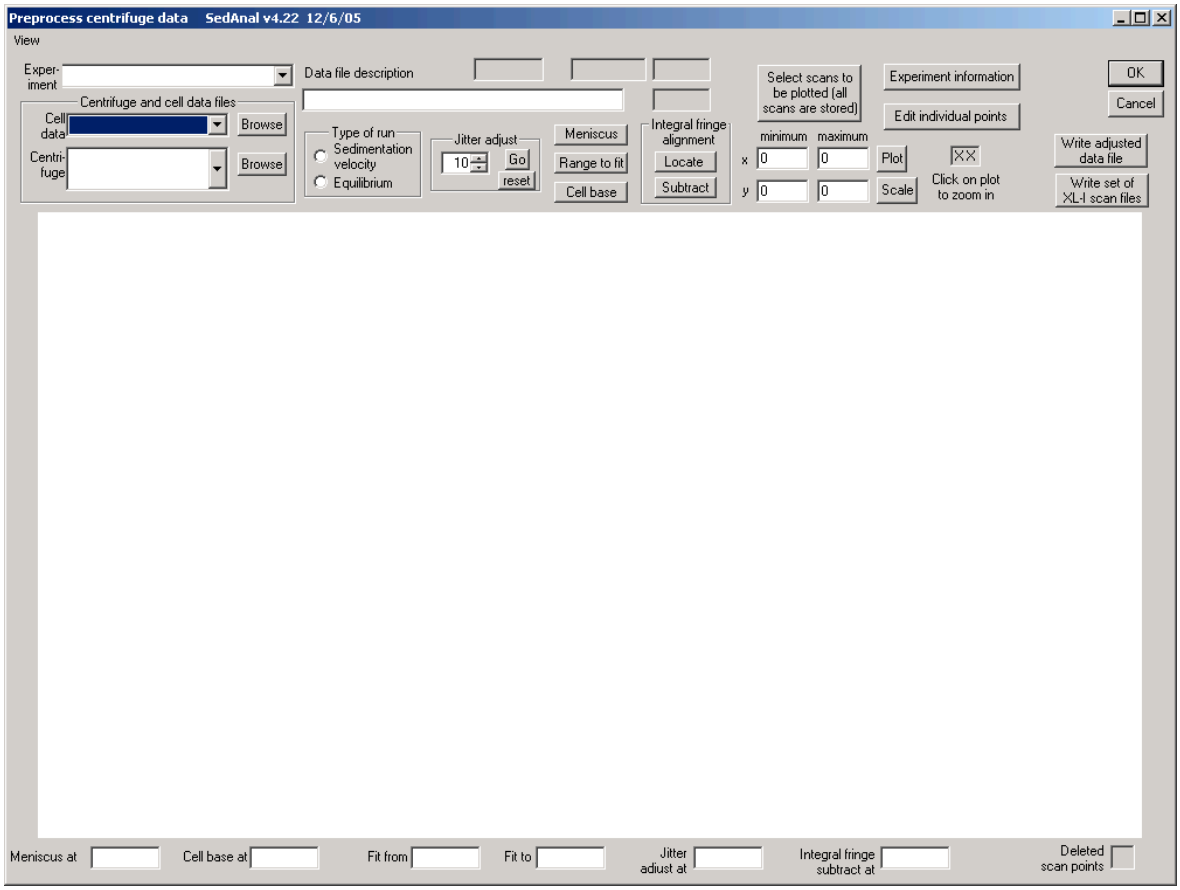


Figure 18. Preprocessor screen.

The first step is to select the experiment folder by clicking on the top upper left box to reveal a list of experiments: in the top drop-down window, one must select the experiment folder. Then, once the experiment folder has been selected, in the bottom drop-down window, one will see the available data sets and select the data set to be preocessed.

Processing is different for absorance data and interference data. For both absorbance data and interference data, one must select the meniscus, the base and range of data to be fitted. Additionally, for interference data, one must remove optical jitter (time dependent, radially independent noise) by aligning the fringes in the air-air space to the left of (centripetal to) the meniscus and then remove any integral fringe jumps by selecting a spot either in the plateau, near the meniscus, or near the hinge point depending on the data set.

In addition to absorbance and interference data, the preprocessor recognizes both **intensity** data and **fluorescence** data. Each set of intensity data is presented in the preprocessor as two separate data sets, one from the sample side and one from the reference side. The intensity data are converted to "pseudo-absorbance" data by taking their logarithm before being processed. Pseudo-absorbance data are linearly related to absorbance data and differ from the corresponding absorbance data by an arbitrary additive constant offset.

2.3.1 Loading Data

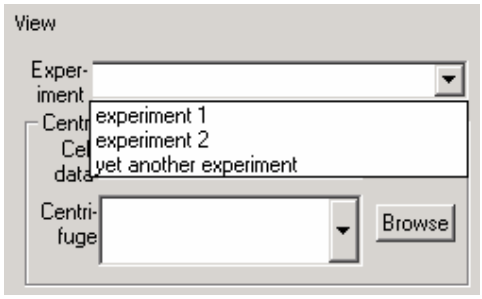


Figure 19. In this example, we'll select "experiment1" which contains XL data stored in the standard format. Now click on the "Centrifuge" drop-down window. To reveal the data sets available for fitting.

(N.B: Standard format is …\User_data\experiment 1\"date"\"time"*.IPn).

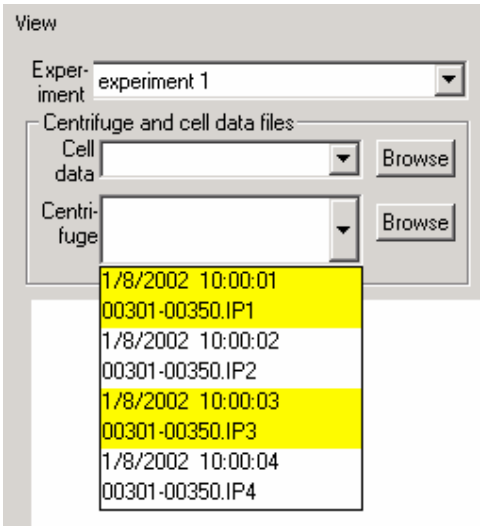


Figure 20. The data sets (in this case from a 4 cell run) are shown with alternating yellow and white backgrounds. The time and date shown in the "Centrifuge" drop down window are derived from the folder names.

When one of the data sets is selected, the entire data set will be read in and will be used for fitting. The actual files (i.e. scans) to be used in fitting will be chosen from the control screen by right-clicking on the run file name window (see below)

Note: If a file or files are missing from a data set, SEDANAL will break the set up into two sets. For example, if the run generated scans 00001.IP1 thru 00099.IP1 but scan 00050.IP1 were missing for some reason, SEDANAL would show two batches of scans from that cell, one from 00001.IP1 thru 00049.IP1 and another from 00051.IP1 thru 00099.IP1.

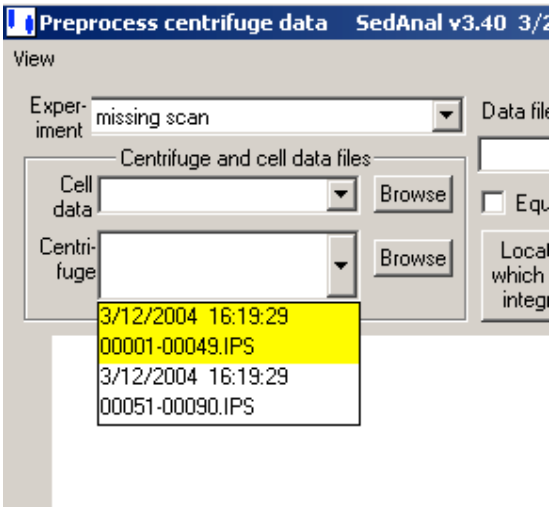


Figure 21.

In order to rejoin the two parts of the run, a dummy file would be created with the name 00050.IP1. Now the data set will appear as one contiguous set of files and would be treated as a single data set. Later, on the control screen one would have to exclude that dummy file from the analysis. The dummy file must at least contain the two header lines at the top. The resulting series of scans (showing only those from 00040.IP1 to 00060.IP1) would look like but now the data can be treated as a single set.

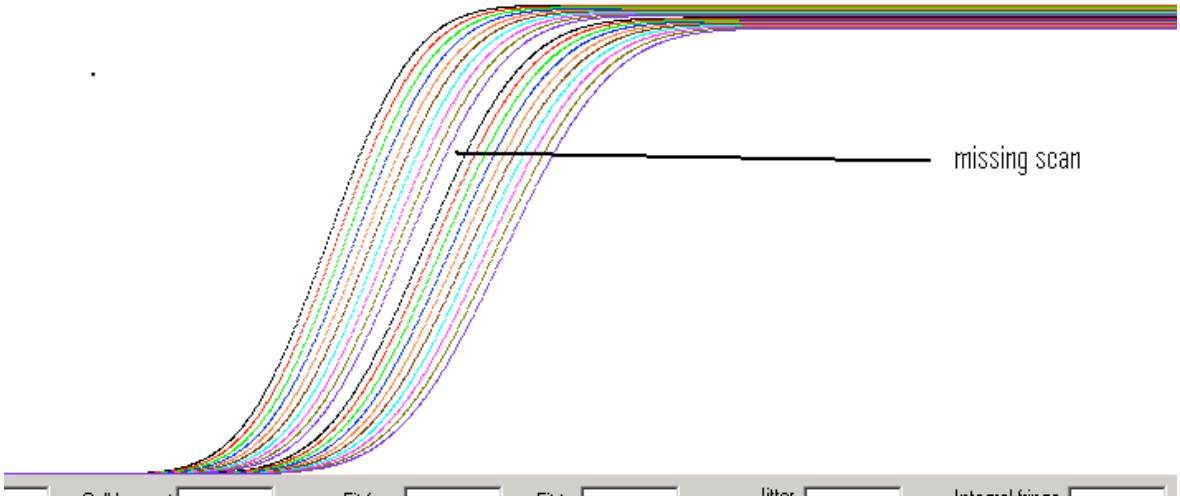


Figure 21b.

Occasionally, the XL-A/I will write an empty file and the data will appear as shown above. The name (i.e. number) of the missing file must be noted and that file must be excluded from the fitting or from the DCDT analysis.

After a data set has been selected and read in, the screen will look something like this:

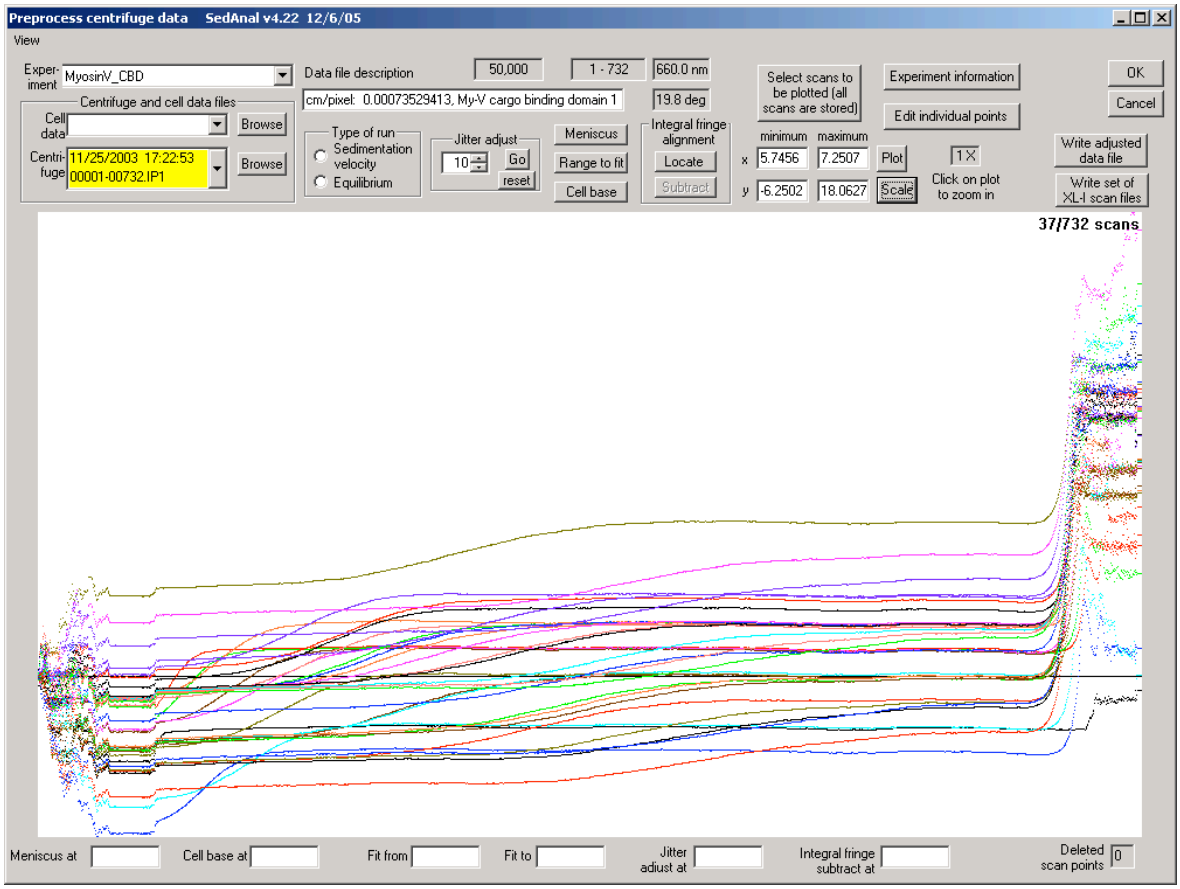


Figure 22a. Initially when the data (in this case, interference data) are read in, they will look like above and will require alignment in the air-air space and removal of integer fringe jumps in the solution column that may occur at the meniscus.

Now indicate what type of run it is:

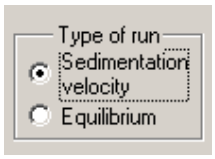


Figure 22b. Click the appropriate button.

Next, we take care of vertical "jitter" by aligning the fringes in the air-air space. Under the "Jitter adjust" label click on "**Go**" to activate the cursor and click in the air-air space. The number "10" indicates that a range of plus and minus 10 pixels will be used for the alignment. If you need to undo what you have done, click on "reset".

2.3.2 Adjusting Data for off-sets and jitter

Figure 23. The "Go" button.

When the "Go" button is clicked the cursor is activated to indicate the spot at which the jitter adjustment will be carried out. the number "10" in the window next to the "Go" button indicates the number of pixels around the clicked position (i.e. +/- 10 points) that will be used in the jitter correction process.

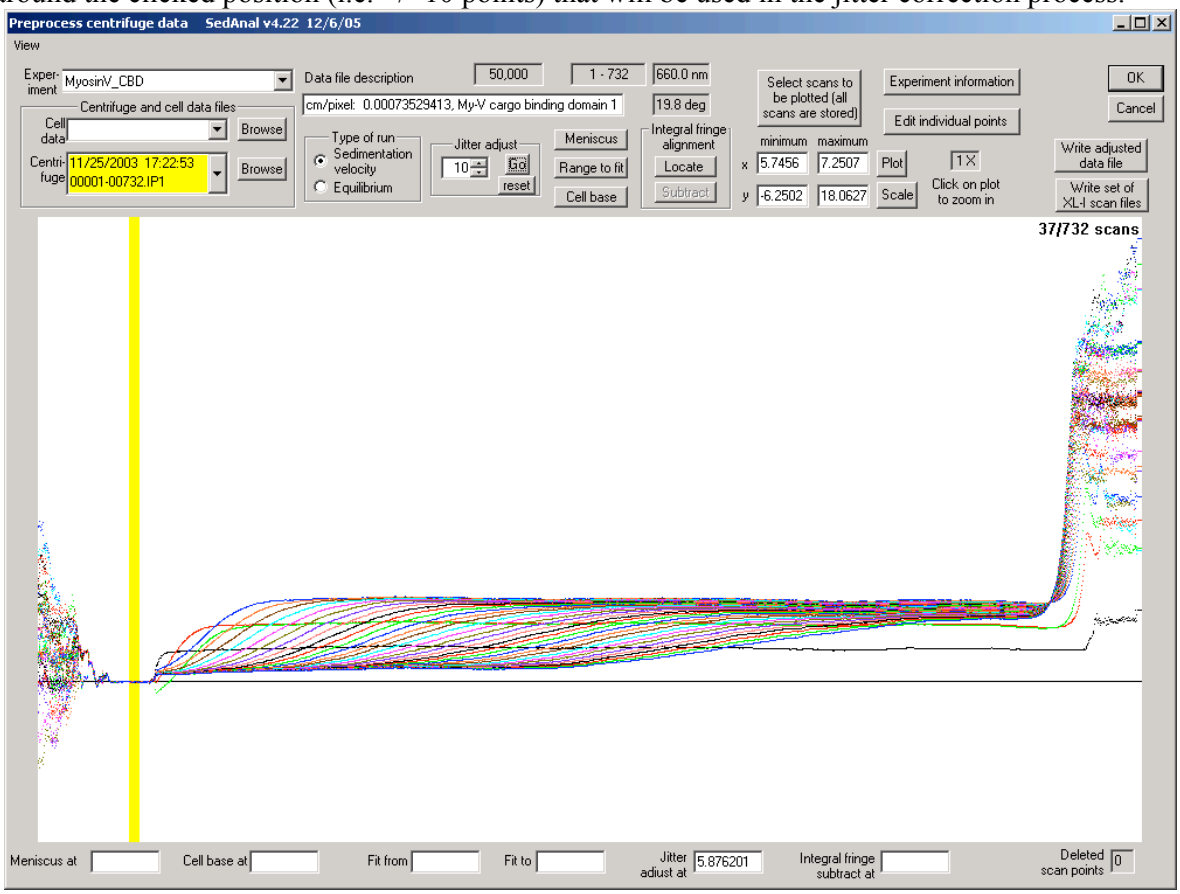


Figure 24. After the jitter adjustments. The scans are now all aligned in the air-air space.

Once the data have been pre-processed, they will be written into a "run" file (with the extention ".abr") and will be available for re-editing by selecting them in the middle window labeled "Cell Data". The example below was actually reloaded from a "run" file whose name is shown in the middle window. All the parameters, meniscus, base etc are stored in the run file and are used by the fitter to perform the fits. Run files can be re-editted at any time, for example, for re-fitting with a different range or better estimate of the base or meniscus positions.

The preprocessing procedure is explained in detail below.

The fringes are aligned first on the air-air space to remove vertical variations due to time dependent instrumental noise. Under lick on the "Go" button and then click in the air-air space. The fringe patterns will all be aligned as shown in Figure 24.

Now we take care of the integral fringe adjustment:

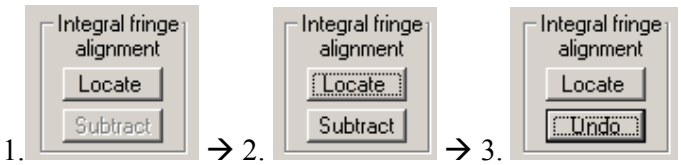


Figure 25. The next step is to re-adjust the data for integral fringe shifts introduced by the data acquistion software at the meniscus. Next (1) click on the "Locate" button and select a spot – usually someplace near the bottom in the plateau region – by clicking there. Then (2) click the "Subtract" button. The result will be as shown in Figure 26, and the buttons will like (3) above. Clicking the "undo" button will undo the last adjustment.

After manually re-scaling the plot, click on "Meniscus" and 'click and drag and unclick' at the meniscus position. Repeat the process to choose the base position by clicking on the "cell base" button and then 'click and drag and unclick' at the base position. Note that the base of the normal double sector centerpieces ia very close to 7.20 cm and so the actual base postion will probably be at a higher radius than you are at first inclined to choose. (You can verify this by running some simulations. You'll be surprised.)

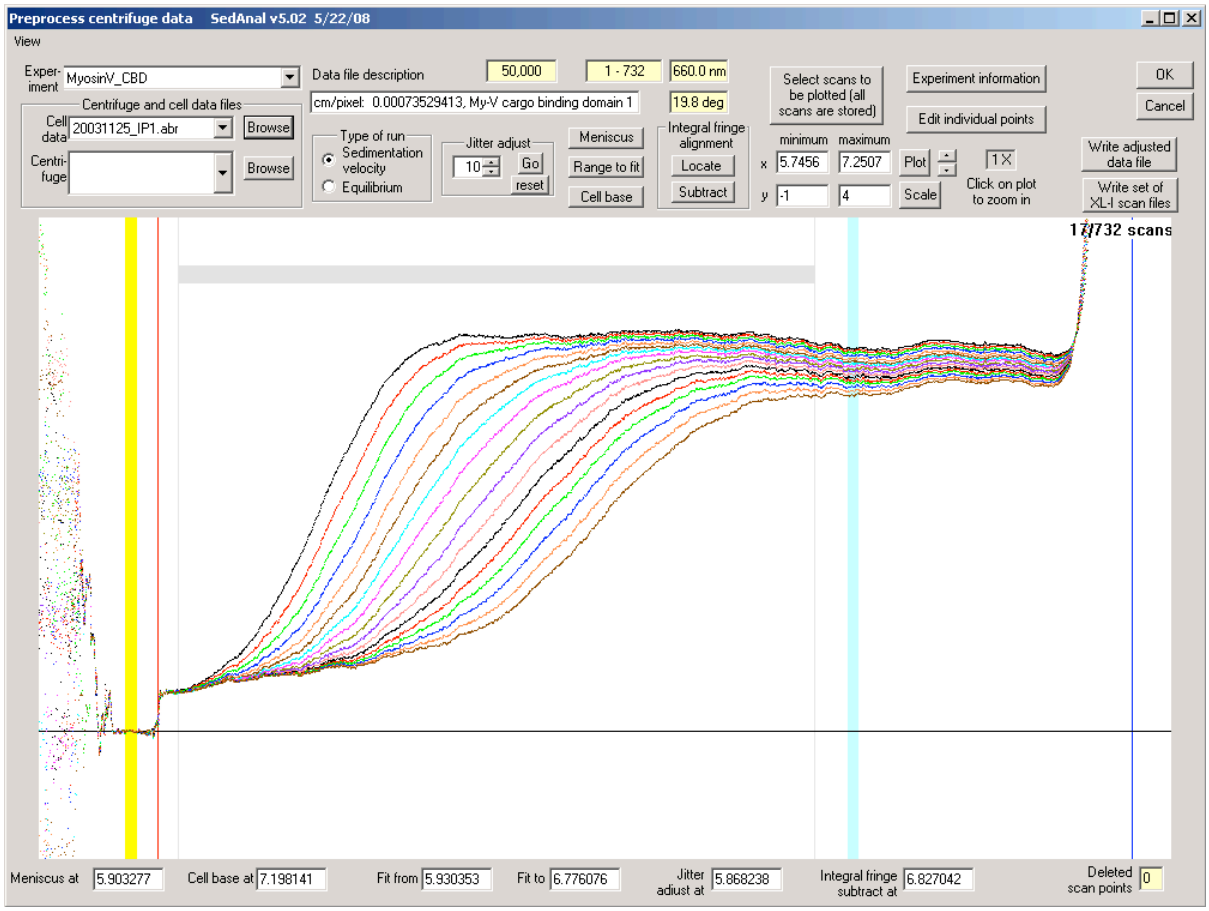


Figure 26. After the fringe adjustments have been completed and after the meniscus, base and range to be fitted have been selected, the screen will look like one presented in this figure with all five windows at the bottom filled in their corresponding values.

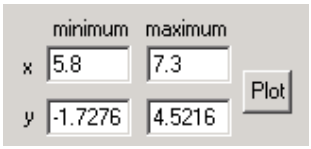


Figure 27. The plot can be rescaled by manually by selecting the y-axis and x-axis ranges in the four windows at the top of the plot. Click on the "Plot" button to refresh the screen.

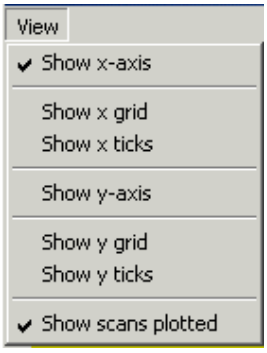


Figure 28. The "View" button in the upper left corner of the preprocessor screen allows one to choose various charcteristics of the plot, as well to show how many of the loaded scans are currently being plotted. The number of scans being plotted is shown in the upper right corner of the plot.

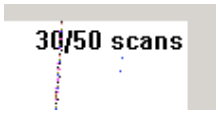


Figure 29. The number of scans being plotted is shown in the upper-right corner of the plot.

2.3.3 Selection of Scans to be displayed.

Click on the "Select scans to be plotted" button to select which scans will be displayed on the screen. Note that all scans will be stored in the "run" file and can be used in the fitting process. Check or uncheck each scan as shown below: Figure 31.

A range of scans and the increment can be selected by typing in the small boxes in the middle right of the window.

2.3.4 Selection of scans to be fitted.

After the meniscus, base, etc have been selected, it's a good idea to choose a range of scans to be fitted at this point while you have them on the screen. Although these can be selected later in the fitting control window by right clicking on the file name for the cell whose scans are to be chosen (see below), it's easier to do it at this stage to see what is being chosen. For example, you might decide to fit the followling scans after seeing them in the preprocessor (i.e. scans 400 to 499 by 1):

Later on the fitting screen this range of scans can be recalled by right clikcking on the data file name and reloading them as they were displayed in the preprocessor. This also gives you opportunity to select the appropriate radial “range-to-fit” for the cosen set.. Bad scans can be excluded from fitting and plotting by right clicking on the scan box as indicated by the green box in Figure 31.

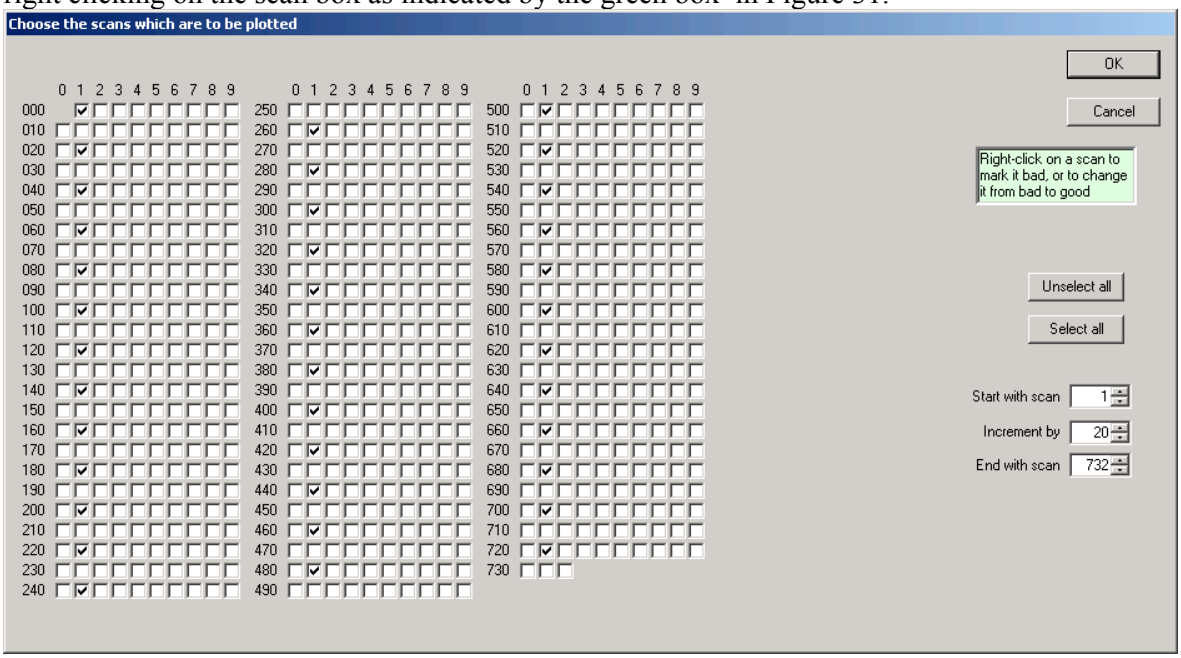


Figure 31. Scan selection for plotting.

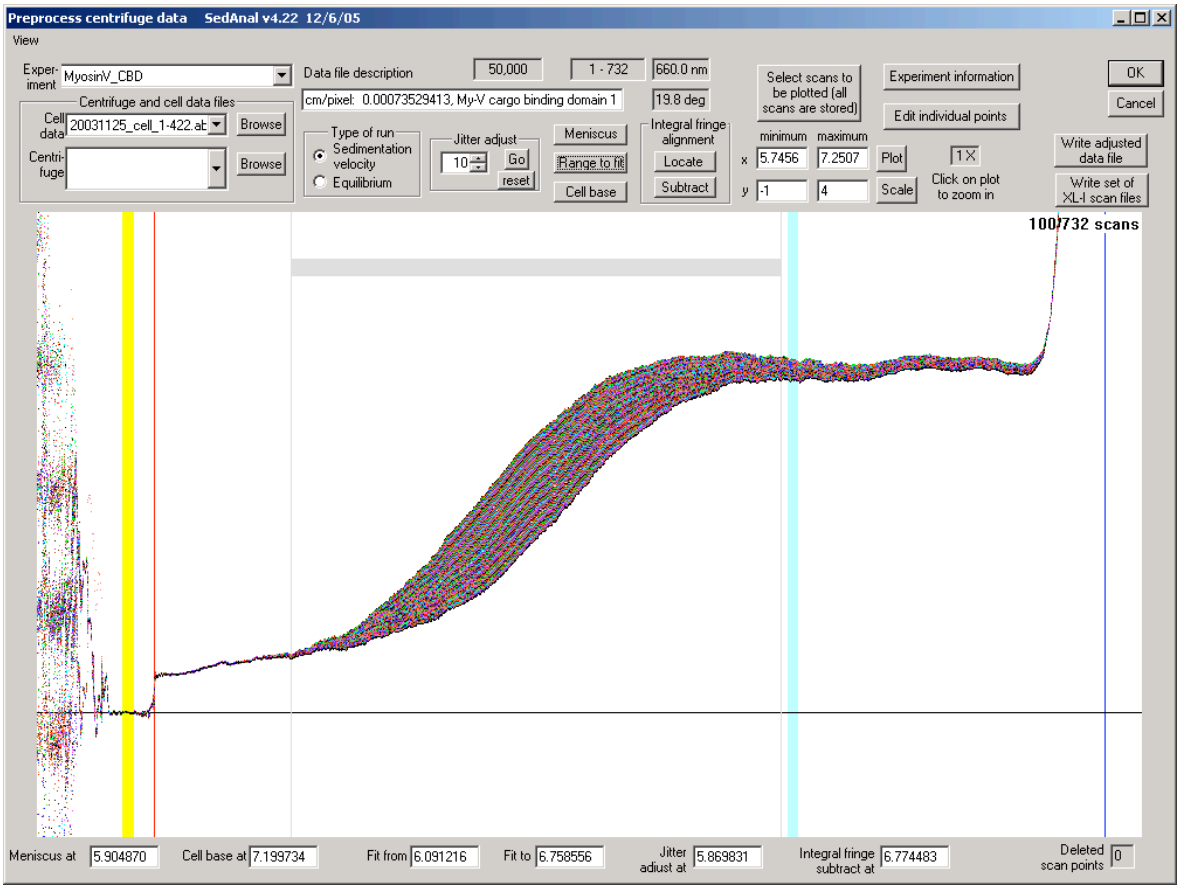


Figure 32. Scan plotted are from 400 to 499. You might want to limit the fitting to just this subset because of an aggregate you want to ignore, for example.

2.3.5 Summary of the preprocessing procedure:

2.3.5.1 Zoom.

The image on the plot can be expanded by clicking at the point around which you'd like the picture to grow. There are three levels of scale expansion, the fourth click will return the image to its original size.

2.3.5.2 Interference Data.

2.3.5.2.1 Remove optical jitter:

To remove optical jitter click on the button labeled "Jitter adjust" and then click in a relatively flat region to the left of the meniscus. The fringes will be aligned to each other by least squares fitting the curves to each other over a small region (default is +/- 10 points).

2.3.5.2.2 Remove integral fringe shifts:

Integral fringe shifts can sometime occur at the meniscus. Because there may be a horizontal region of several pixels where the are no fringes, the fringe tracing algorithm can get confused and loose count by one or more integer jumps. These shifts are eliminted by comparing fringe pattterns in a region where little transport is taking place. To remove integral shifts, click on the "Locate" button under Integral fringe alignment. Click on a region of the pattern to the right of the mencscus where the pattern changes least between scans; the plateau region is usually a

good spot. A green band will appear. Then click on “Subtract” to eliminate the shifts. This is done simply by subtracting or adding integers to the data as required.

2.3.5.3 Absorbance and Interference Data:

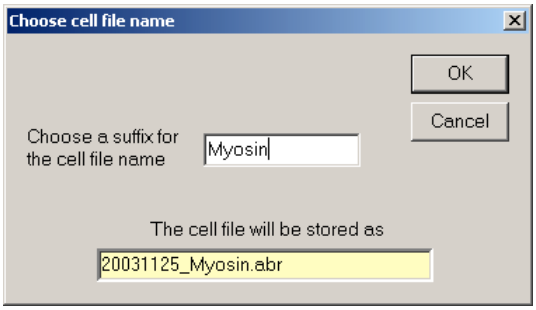
To set the meniscus position, first click on the mensicus region to expand the image. Then click on the button labeled “Meniscus” and then click and while holding the mouse down move the black vertical line until it coincides with the meniscus, and release the mouse button. Now a red vertical line will appear at the meniscus position and the radius of the meniscus will appear in a box at the bottom of the screen. The value in the box at the bottom of the screen will track the mouse while the button is held down. You can also type a vlue for th emeniscus postion into this box.

To set the base position, repeat the same process for the location of the base of the cell by clicking on “Cell base” first.

To select the range to be fittted: The range to be fitted involves a “click-and-drag-from-left-to-right-and-release” operation. Click on the “Range to fit” button, then click on the centripetal point and drag the pointer to the centrifugal point. This selects the range of the data to which the model will be fitted.

Save Preprocessed data.

Once the data have been preprocessed, the data are ready to be fitted. The preprocessed data are stored in a “run” file with the extension “.abr”.



As shown above, the file name will be of the form “yyyymmdd_<user_suffix>.abr”, where <user_suffix> is supplied by the user when either the “OK” or the “Write adjusted data file” button is clicked.

2.3.5.4 Multi-wavelength Data -

The advantages of using absorbance optics on the ultracentrifuge is the high sensitivity and specificity; by choosing an appropriate wavelength, you can often observe one species independently of others. This is not the case with interference optics, where all species contribute to the signal.

To differentiate among species, it is convenient to use the multi-wavelength feature of the XL-A or XL-I. This allows you to have every *n*th scan be done at a different wavelength (*n* can be up to 3). This is accomplished by moving the monochromator grating between scans. The nominal wavelength is recorded in the file header. Beckman’s control software allows you to choose among a single wavelength for all

scans, or alternate scans at two or three different wavelengths. On the XL-I, you can also do an interference scan, as well as absorbance at up to 3 different wavelengths, in the same run.
When doing scans at multiple wavelengths, the monochromator reproducibility is about 2-4 nm; the reported wavelength accuracy is unknown. Therefore, it is desirable to avoid wavelengths where the extinction coefficient is changing rapidly with wavelength, although some compromise may be necessary to take advantage of the peak intensities of the Xenon flash lamp.

When preprocessing the scan files produced by the centrifuge, SedAnal will allow you to separate the scans for each wavelength. This is necessary because the extinction coefficients vary. Normally, the choice of wavelengths is designed to maximize the difference between extinction coefficients at each wavelength for the species of interest.

There are two ways to do the separation using SedAnal. Remember that the first step in analyzing scans is to store the data in a run file (.abr). As an example, suppose you have collected 100 scans of absorbance data at 280 and 350 nm in cell 2. The XL-A will use 280 nm for the odd-numbered scans (00001.RA2, 00003.RA2, ...), and 350 nm for even-numbered (00002.RA2, 00004.RA2, ...).

Create a single run file containing all the scans for the cell. During preprocessing you choose a single meniscus, cell base, and range to fit. On the fitting control screen, you will choose the run file *n* times, selecting sets of scans at each wavelength. For the example, we would analyze two instances of the run file; for the first, we select "scans to be fitted" of 1, 3, ..., 99, and for the second, 2, 4, ..., 100. On the control screen, the two cells will be *linked* (see below) indicating they are physically identical. For the extinction coefficients of the various species, you will put in different values for the two cells, corresponding to the different wavelengths (see below).

Create multiple run files containing only the scans at a particular wavelength. Each wavelength is treated independently. During preprocessing you choose a meniscus, cell base, and range to fit for each wavelength. When fitting the data, you use the *n* run files, and no scan selection is needed, although you may want to do so to speed processing or restrict the part of the run being analyzed. For the example, we have SEDANAL create two separate run files; the first containing the 280 nm data, the second, 350 nm data. These cells will also be *linked* (see below), because they have identical loading concentrations. For the extinction coefficients of the various species, you will put in different values for the two cells (see below), corresponding to the different wavelengths.

The difference between these choices is only how many run files are created from the original scans, whether the scans need to be separated by wavelength, and whether different menisci, etc, can be used for different wavelengths.

The XL-A monochromator does not reposition itself to the exact specified wavelength for successive scans. In addition, the measurement of the actual wavelength, stored in the header of the scan file, is subject to error. SedAnal has a tolerance for wavelength matching: two scans are considered to be at the same wavelength if they are within this tolerance. The value of this tolerance is set in Preferences/Preprocessor, by checking the box "Preprocessor should offer the option to separate scans..." and entering a tolerance below. The default tolerance is 4.0 nm, with the option enabled.

When the preprocessor is reading in a set of scans, it looks at the wavelengths, and if it detects more than one, it will offer the choice to create separate run files, assuming this option has been enabled in Preferences, as described above.

If the wavelength tolerance is too small, there may be more different wavelengths than intended. In this case, SedAnal will list the first six wavelengths seen, and suggest increasing the wavelength tolerance.

2.3.5.5 Preprocessing multi-speed data sets for Wide Distribution Analysis (WDA):

When loading a multi-speed data set, the preprocessor will stop after loading the scans from each speed and allow you to choose a radius at which to align the fringes in the air-air space and to choose a meniscus position at that speed. The integral finge subtraction will be done after the scans from the last speed have been loaded.

After the scans from the first speed have been loaded the boxes at the top of the screen will look something like Figure 33.

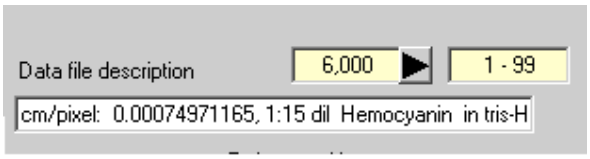


Figure 33.

One clicks on the little black arrow to advance the preprocessor to the next speed's data set where the next meniscus and vertical alignment in the air-air space are carried out. The ""Range to fit" does not need to be set for WDA data.

Figure 34.

After all the speeds have been processed, carry out the integral fringe alignment and click "OK" to save the abr file. Next proceed to the DCDT button on the Main menu.

Click the "Wide distribution" button (left side above the dcdt window) to enter the WDA mode. Then load the abr file. Enter 1.0 for the smoothing window. The data from each speed will appear in a different color (Figure 35). See below for more on WDA.

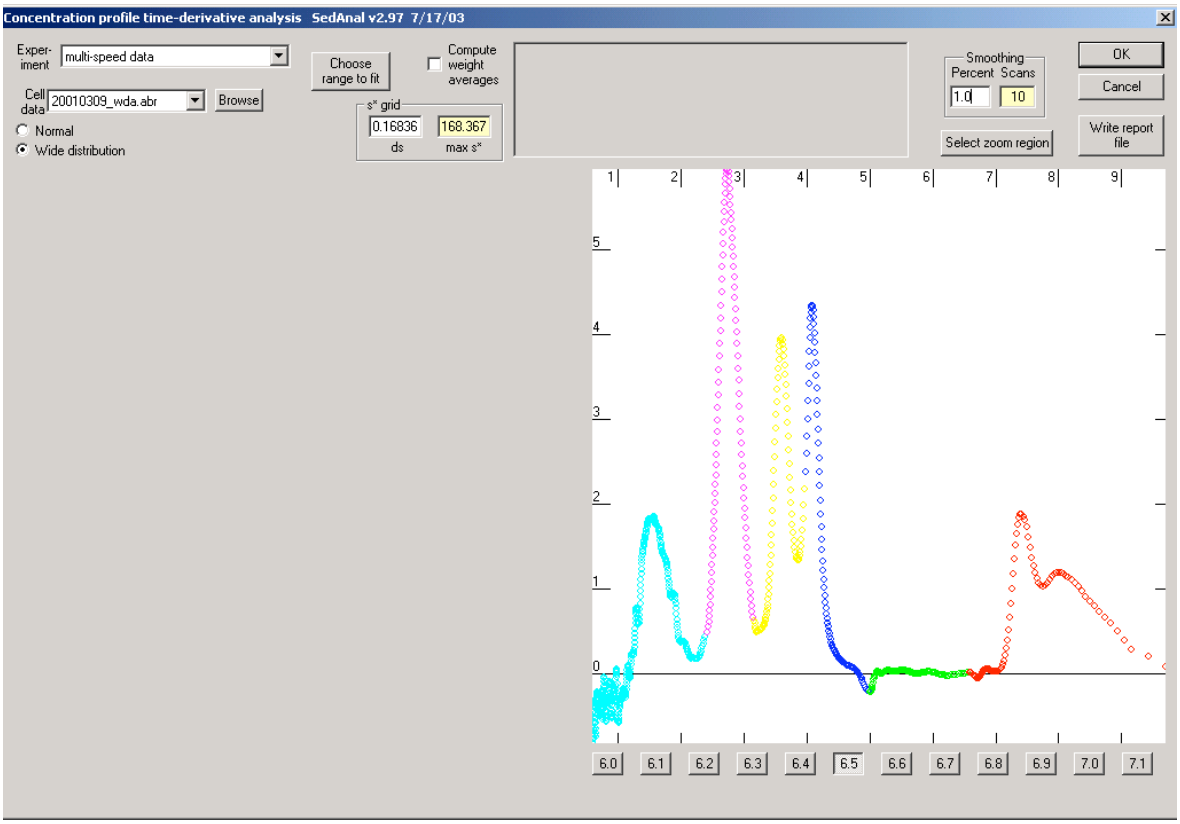


Figure 35. Multi-speed wide distribution analysis.

2.3.5.6 Preprocessing Sedimentation Equilibrium data

The scan or scans to be precessed for SedEq analysis should be put in a folder on the same “date/time” path are for sedimentation velocity. Only the scan(s) to be precessed should be made visible in that folder. SEDANAL will allow the user to load an equilbiurm scan and an optical baseline scan (a background scan taken usually with water vs water to be used to correct for systematic errors that arise from irregularities in the optical system). SEDANAL will subtract the optical background scan from the equilibiurm data scan. This is mandatory for interfernce data and often helpful for absorbance data if there is a scratch on a window or dirt on the optics.

2.3.5.6.1 Equilbrium run in standard double sector centerpiece.

Here is an example from a run performed in a standard double sector cell with a loading volume of 0.25 mL (Figure 37).

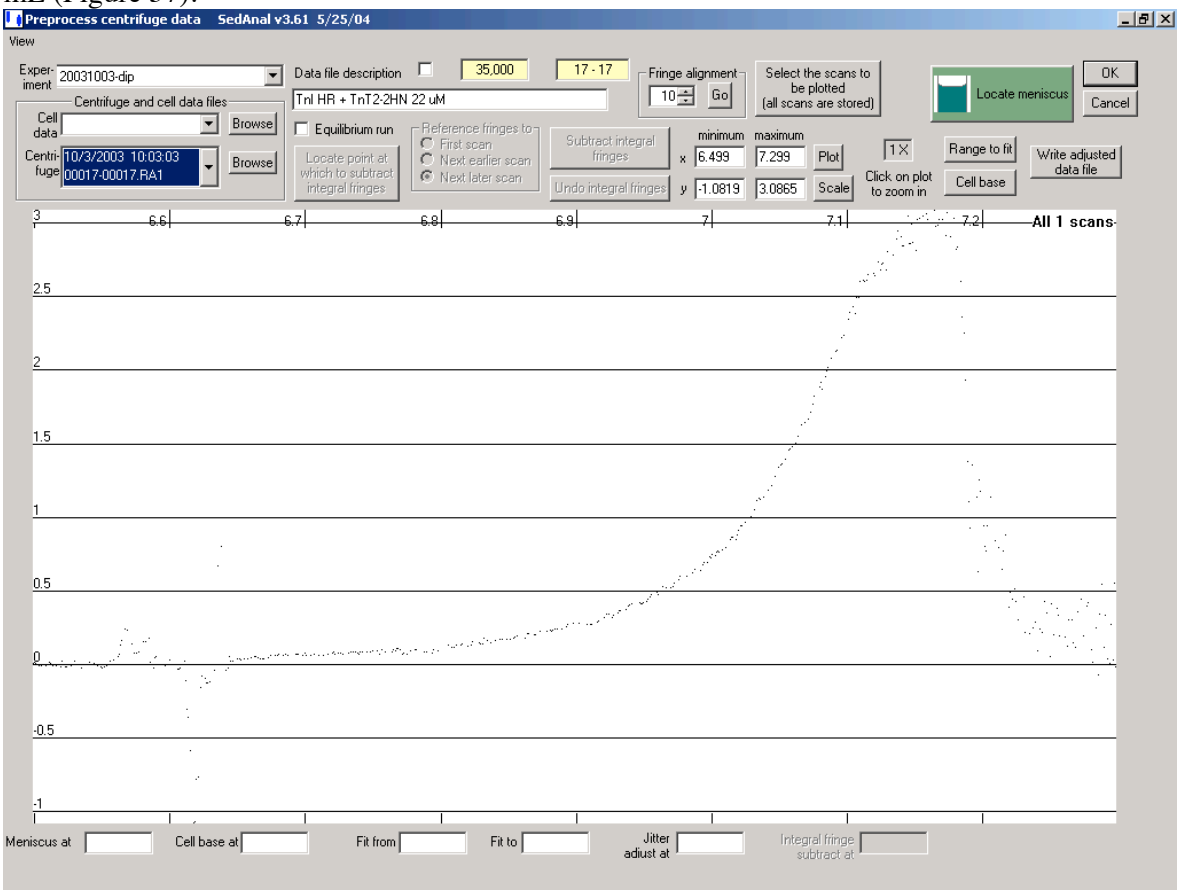


Figure 37.

After selecting the meniscus, base and range-to-be-fitted, the screen will look like in Figure 38, (**Do not forget to click the "Equilibrium" check box**). (This screen is slightly different in v4.6x; you will be remined to select either velocity or equilibrium if you forget)

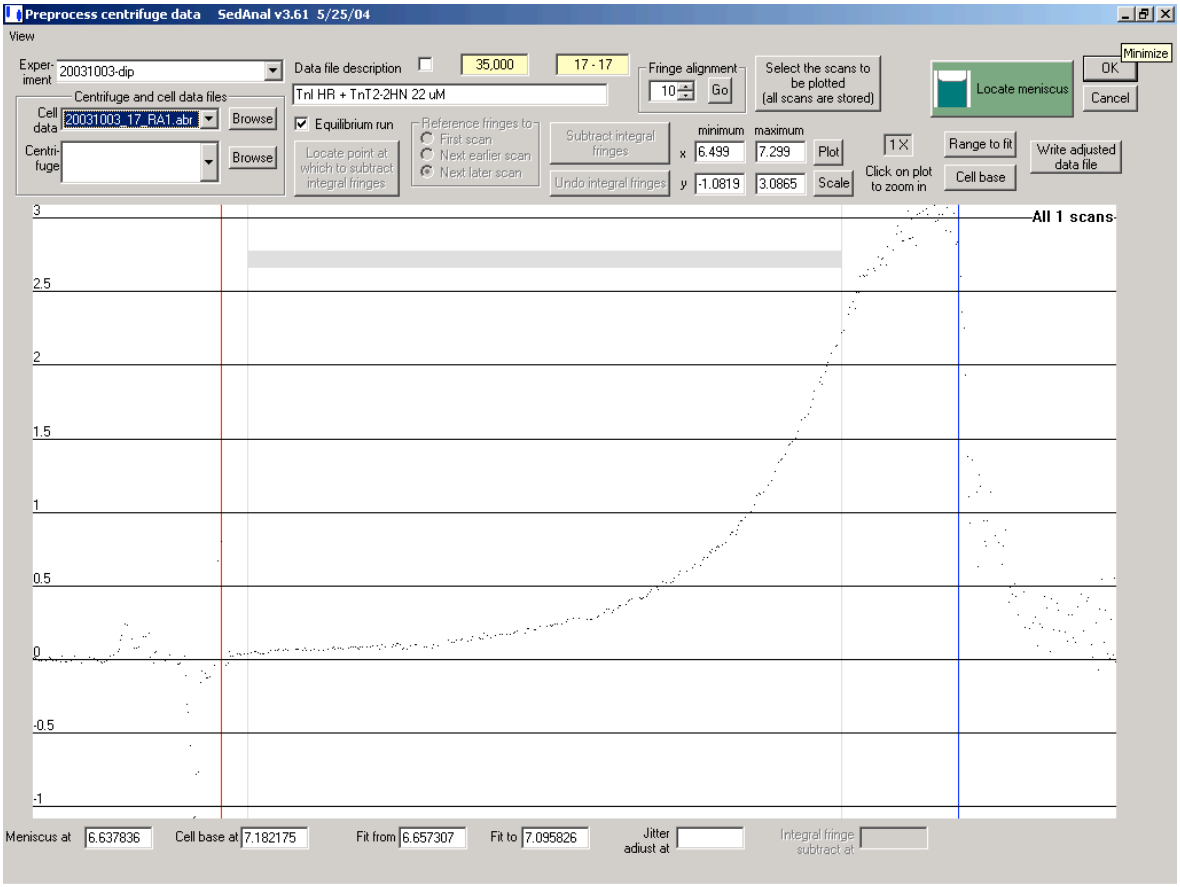


Figure 38.

If an optical blank is to be subtracted from the data run, select it now in the Centrifuge data file drop-down window. Three curves will apear, the equilibrium data, the blank, and the corrected data.

2.3.5.6.2 Multi-channel data with 6 channel centerpieces

After loading a data set from a 6 channel centerpiece, we select the meniscus, base and range for each channel and repeat until three cell data files are produced.

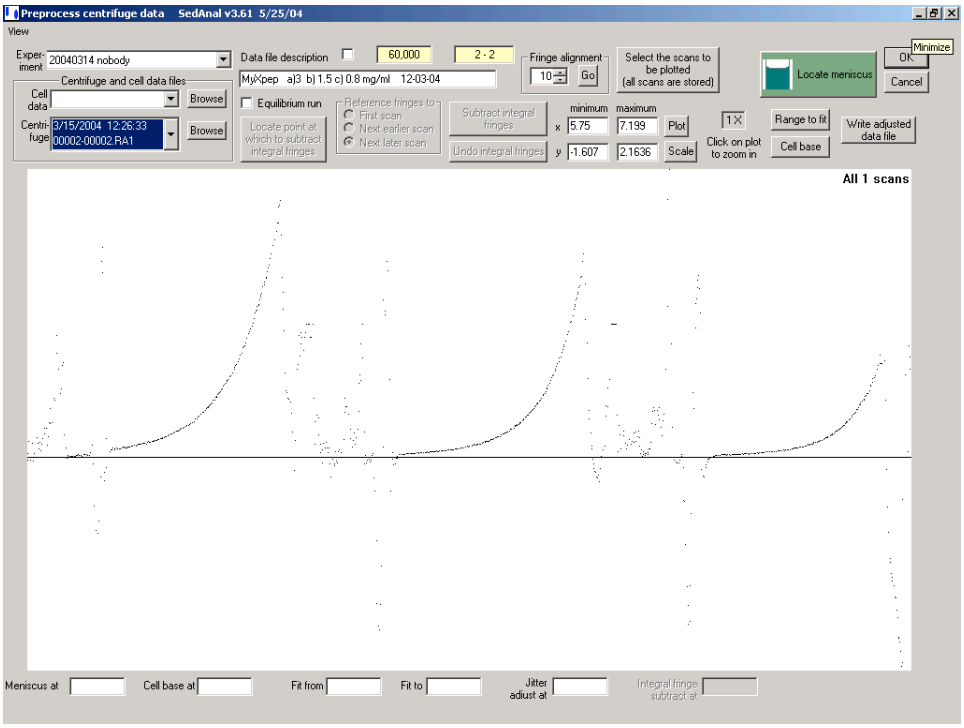


Figure 39. Data from 6 channel Yphantis type equilibrium centerpeice.

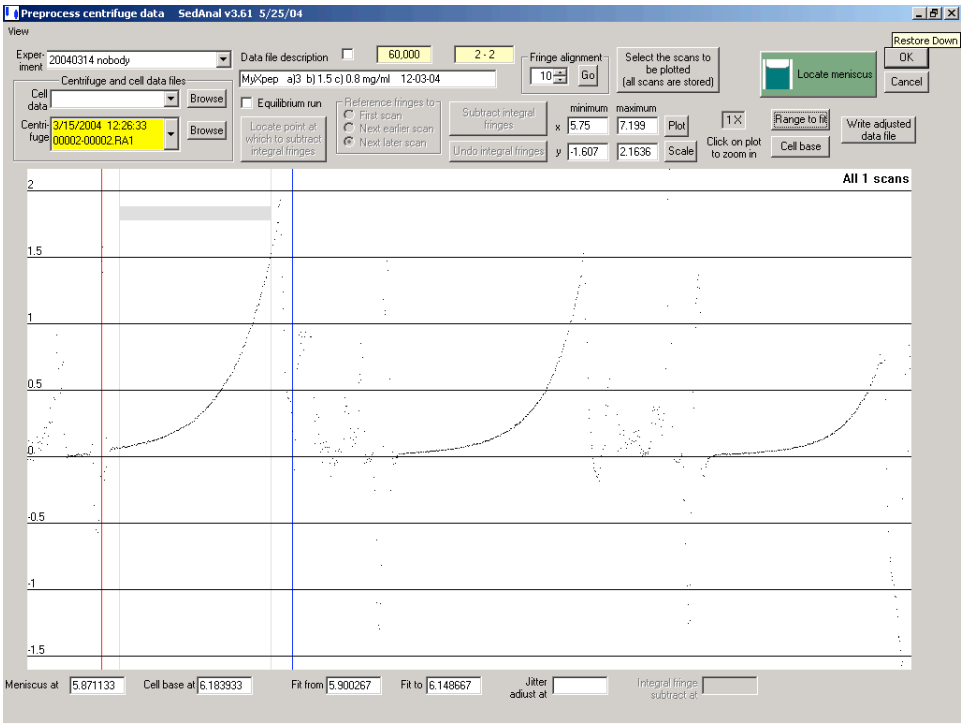


Figure 40.

After the meniscus, base and range-to-fit have been chosen for the first channel, the processed data can be saved with a name like 20040314_60K_CHA, for example. And then these data must be reloaded and the process is repeated for the other channels and each is saved as a separate cell data (*.abr) file using similar names (eg. 20040314_60K_CHB, and 20040314_60K_CHC, etc).

2.4 Fit Preprocessed Data

2.4.1 The Control Screen

When "Fit Preprocessed Data" is selected from the Main Menu, the default control screen (Figure 41) will appear.

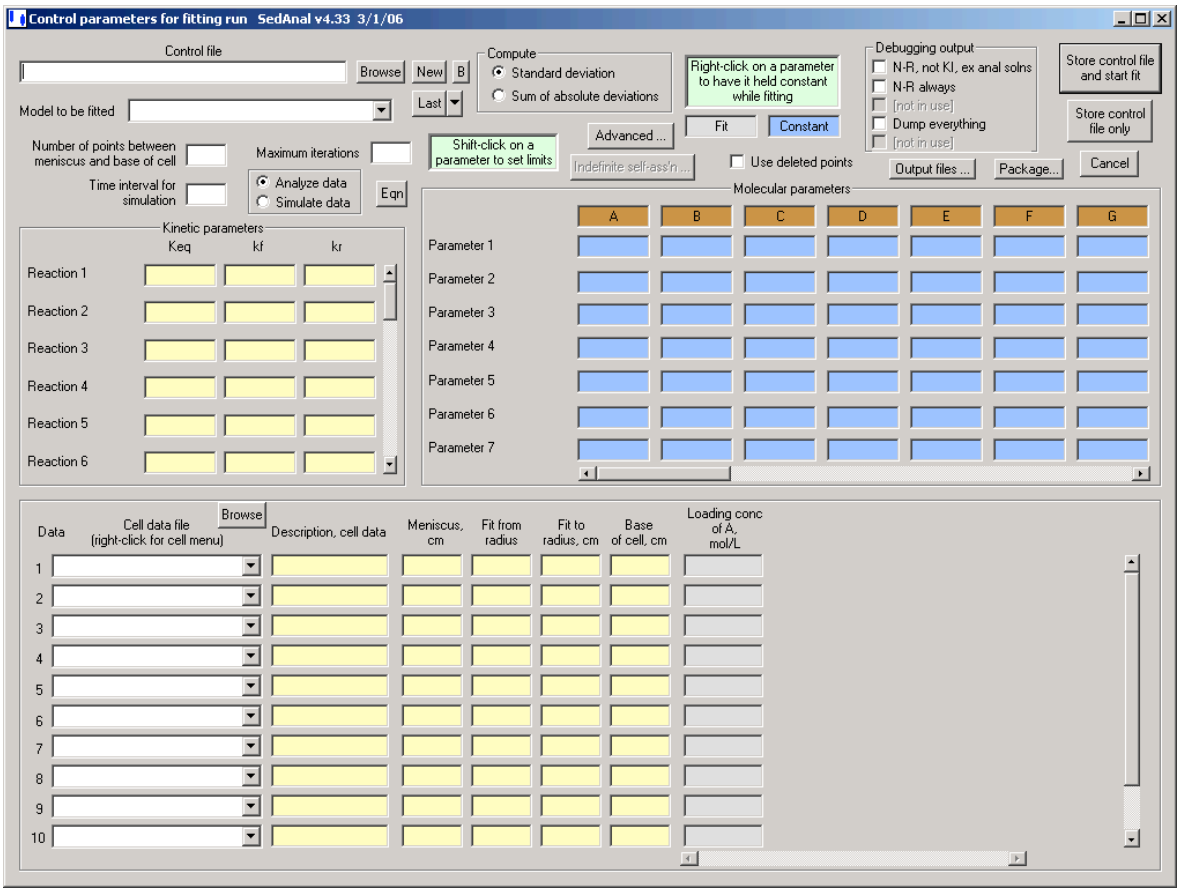


Figure 41. The main control screen for the fitting preprocessed data

2.4.1.1 Create a New Control File

To start a new fit, click on “New” and a dialog box asking you to specify an “experiment” directory will appear. The directory you specify next will be the location in which the control file will be stored with extension “.abc” (Figure 42).

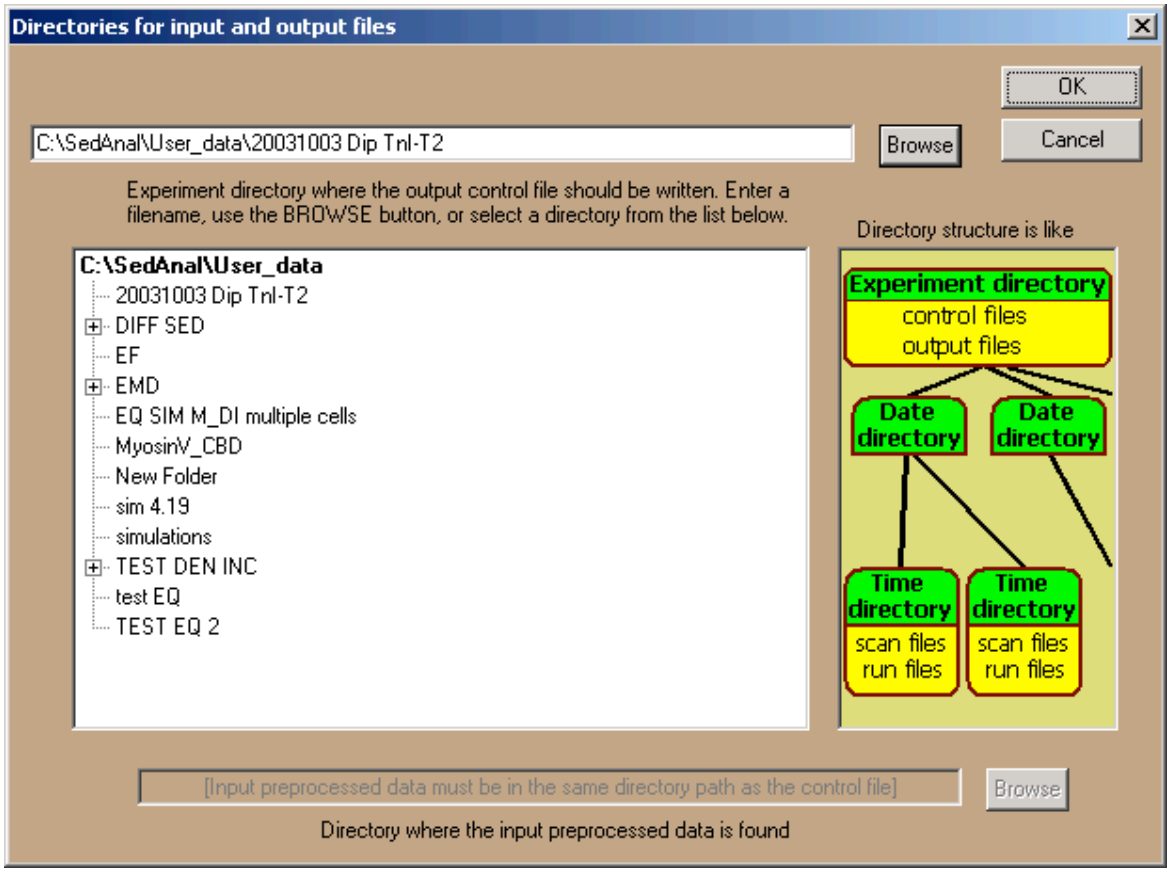


Figure 42. Dialog box for output directory

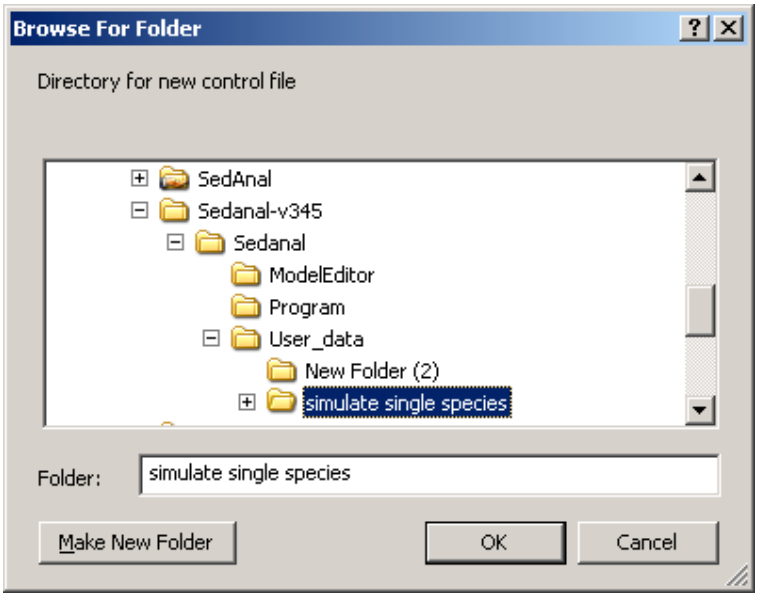


Figure 43. Click on "Browse" and select the experiment directory in which the control file will be stored, here: "simulate single species'

The control file name will appear in the Control file box. It will be named "New_Control_Filename" by default and should be renamed immediately at this point before storing it.

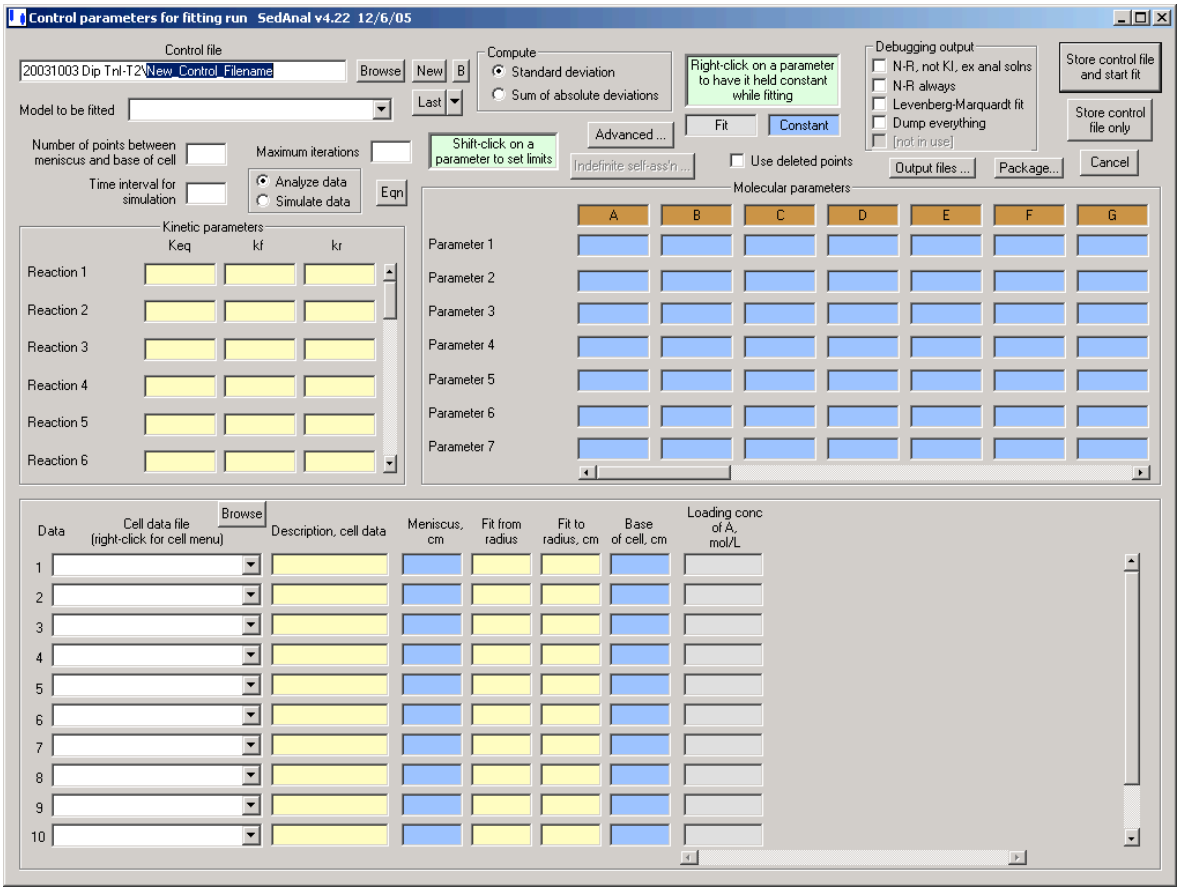


Figure 44. The main control screen for the fitting preprocessed data

2.4.1.2 Choose between Analyze Data and Simulate Data

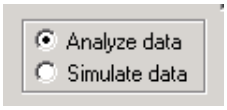


Figure 45. Now indicate whether you are going to "Analyze data" (the default for fitting) or "Simulate data" by clicking the appropriate button, and select the model from the "Model to be fitted" dropdown window.

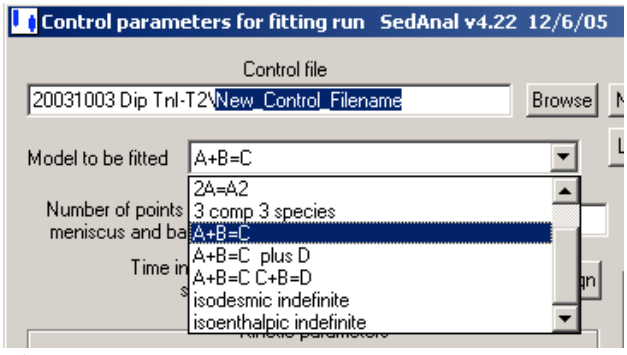


Figure 46.

Now the Control Screen will change to correspond to the model chosen (Figure 47).

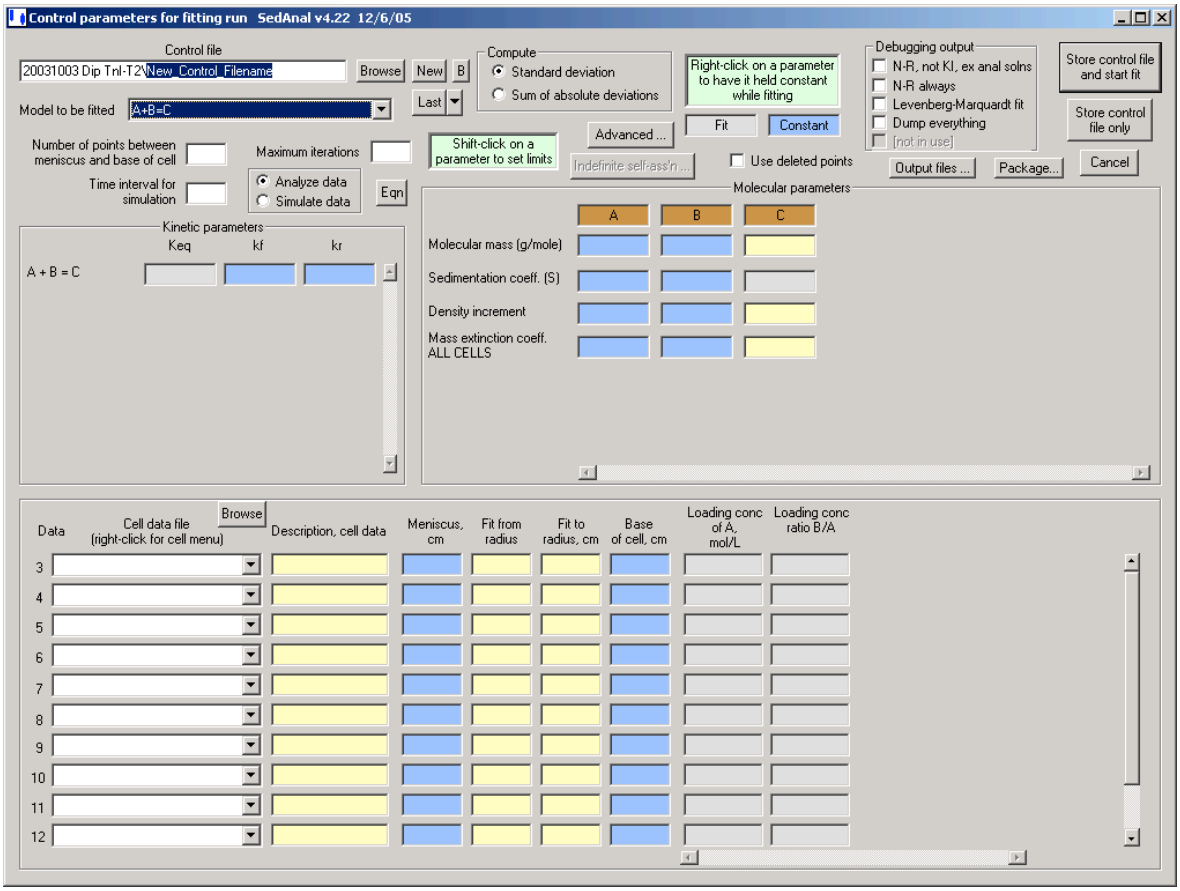


Figure 47. The main control screen for the fitting preprocessed data. As the model is selected the appropriate boxes will appear on the top portion of the control screen.

Loading data files:

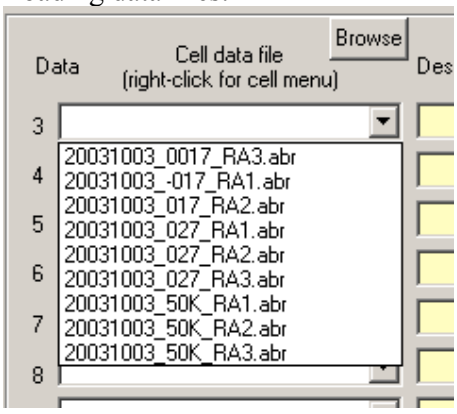


Figure 48. Select the cell data (*.abr) file from the "Cell data" drop-down window. When the Cell data file is selected, the type of run (velocity or equilibrium) will be recognized and the parameter boxes on the control screen updated accordingly. The default screen assumes velocity run until the abr file is loaded.

The control screen is divided into several regions (Figure 47).

TOP LEFT:
In the top left region of the control screen, one supplies a name for the **Control file.** Below that window is a drop-down window from which a **model** may be chosen. Below that are boxes for several parameters used by the Lamm equation solver: The **number of points** to be used between the meniscus and the base of the cell, the **time interval** to be used by the ODE (Ordinary Differential Equation) solver, and the total **number of iterations** to be performed by the non-linear least squares function minimizer.

MIDDLE-LEFT:
In the mid-left region, in the box labeled "**Kinetic Parameters**", kinetic rate constants (k_f, k_r) or equilibrium constant (Keq) will be entered by the user. If an equilibrium constant is chosen, the internal routine uses 0.01 sec^{-1} for the reverse rate constant and fits for the forward rate constant only. The default value for the reverse rate constant may be changed in the **"Advanced"** window. If nothing is entered in the box labeled Keq, one must supply an initial guess for both the forward and reverse rate constants. Other parameters used by the Kinetics ODE solver also may be adjusted in the "Advanced" window.

MIDDLE-RIGHT*:*
In the middle right region, in the box labeled **"Molecular parameters"**, one enters the molecular parameters for each species participating in the reaction scheme.

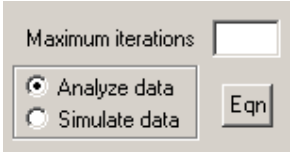
These parameters include molar mass, sedimentation coefficient, and either partial specific volume times density or density increment, and mass scale extinction coefficient.

LOWER-BOX:
The Lower Box allows one to select the data sets that will be fitted either singly (if only one is chosen) or globally, if up to twelve (in this version, v422) are chosen. The data sets are chosen from the drop-down windows on the left and will be found in the **Experiment** folder in which one had preprocessed the data sets under consideration. The abr files live at the “time” directory level.

The best guesses for the loading concentrations in molar units are entered in the boxes at the right end of the data set line of boxes. These should usually be allowed to float (you may think you know what they are but you really don't). One may either float or hold a common value for the ratio of $[B]_o/[A]_o$. In most cases in which a dilution series is being analyzed, this ratio should be the same for all cells. To constrain the ratio to be the same value for all cells, left click on the text above the ratio boxes.

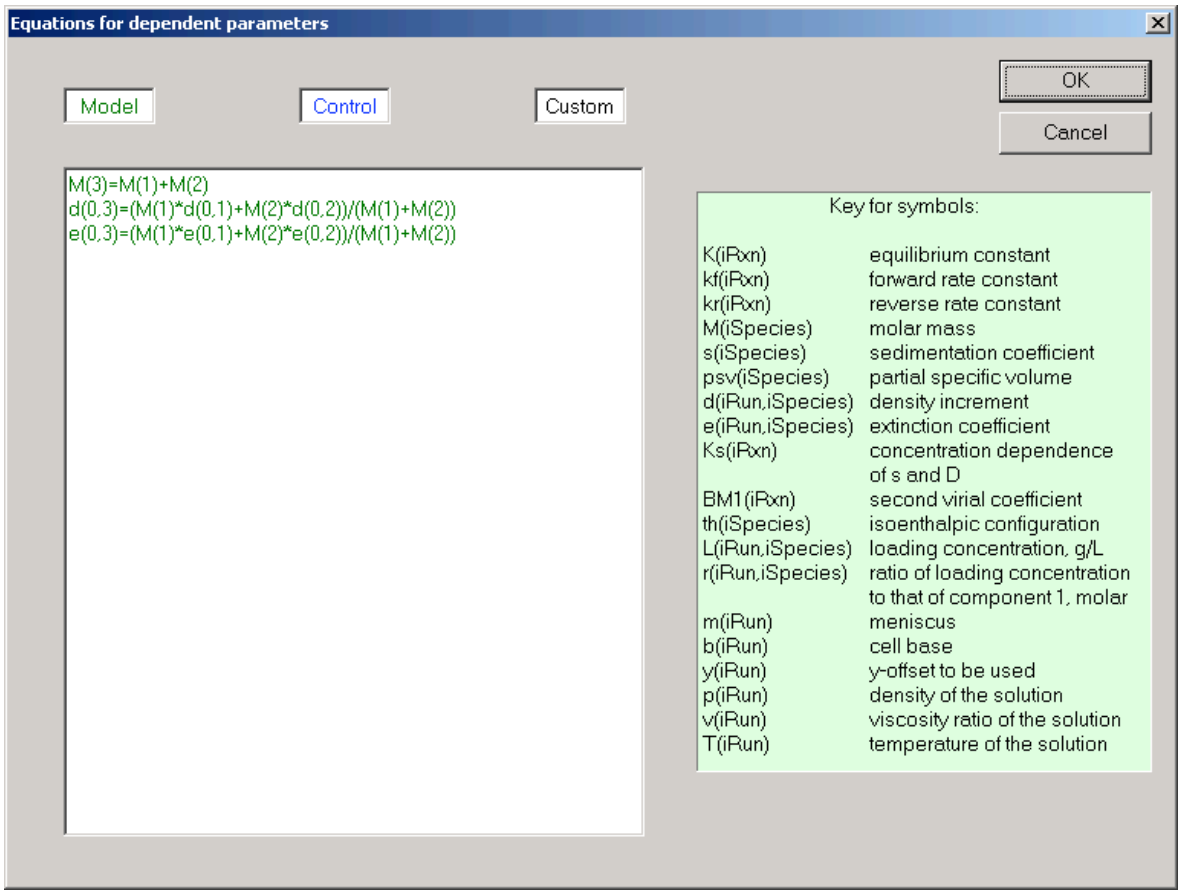
2.4.1.3 The Equation Editor

Fixed relationships between parameters can be established with the Equation Editor.



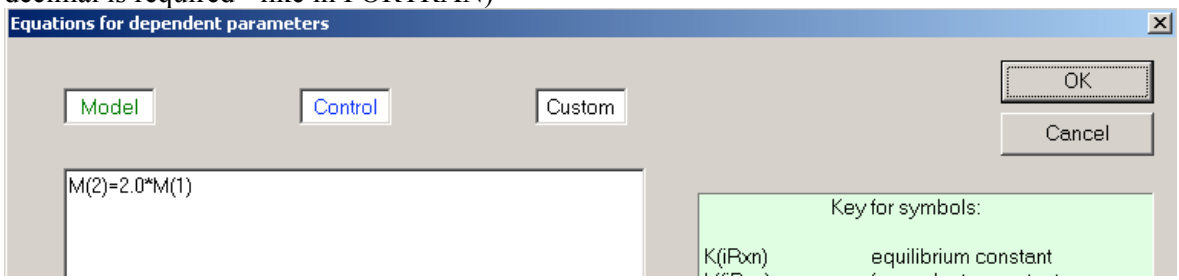
The equation editor window can be opened by clicking on the button labeled **"Eqn"**.

Relationships established by the **ModelEditor** will appear in green. Relationships established by the **control screen** will appear in blue. Custom relationships that are established by the **user** will appear in grey font (see below).



In this example, for the model A+B=C, the equations are generated by the ModelEditor and show the relationships established by clicking on the boxed on the tab for species 3, "C". That is, the molar mass is the sum of those of species A and B. The density increment and the extinction coefficient of "C" are both expressed as mass weighted averages of those of species A and B.

Other relationships between the parameters can be entered manually. For example, consider the case of a 2 component, 2 species system comprising a non-interacting monomer and dimer. One would choose the model "2 comp 2 species" and then enter the relationship relating the molar mass of the two species. (the decimal is required - like in FORTRAN)



Relationships in **Green** are established by the ModelEditor; those in **Blue** by the control screen and those in **Grey** by you, the user. So for the system A+B=C we might have a case in which we had performed an experiment in which we've done a dilution series (4 cells) at a constant ratio of [B]/[A] loading

concentrations and combined those data with another experiment in which we've varied those ratios (2 more cells). Moreover, we've used both absorbance and interference optics for 2 of the cells. That gives us a total of 8 run files to combine in a global fit. Assuming we know the extinction coefficients (both in units of A.U.-mL-mg and of fringes-mL-mg) accurately, we can require that the loading concentrations in the cells in which both optics were used be the same by linking the cells on the control screen. Or we can use the equation editor to establish those constraints if the 2 cells in question were not adjacent to each other in the list of run files on the screen.

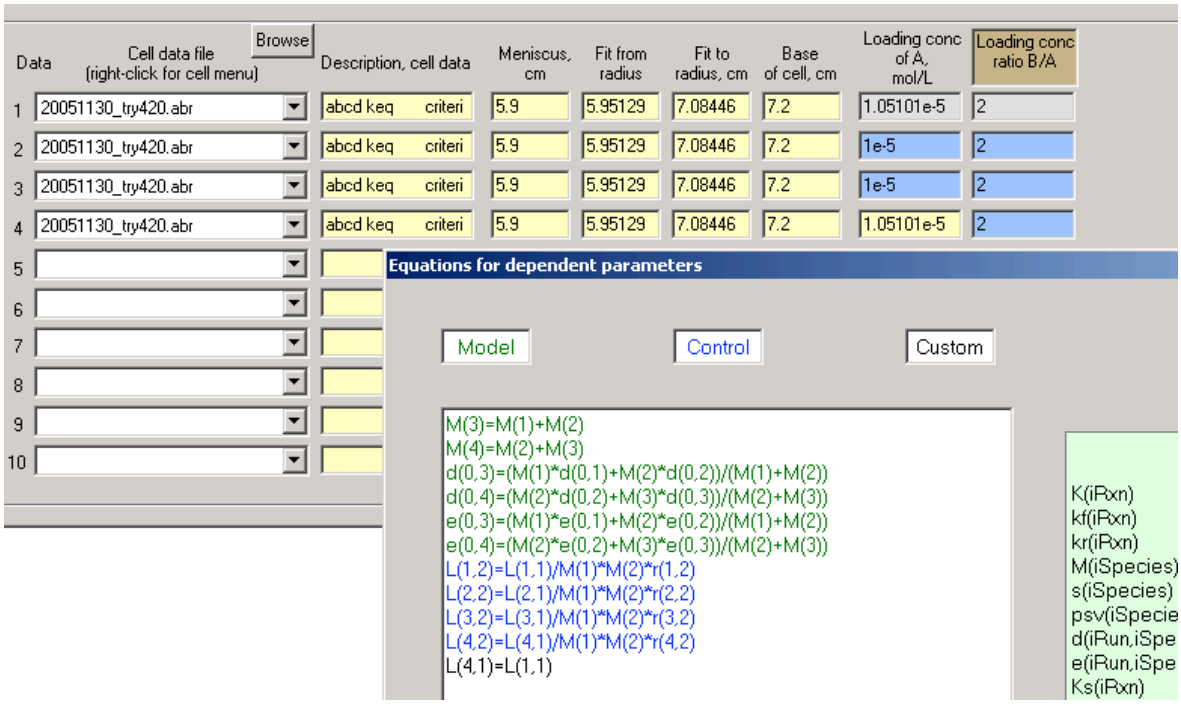


Figure 48B. Using the equation editor to link parameters in different cell data sets.

2.4.1.4 Start the fitting:

To start the fitting after all the initial guesses and other constraints have been entered, the fitting is started by clicking the "**Store control file and start fit**" button.

2.4.1.5 Advanced Parameters Button

2.4.1.5.1 Error Estimation Control

The first set of advanced parameters is under the “Error estimates” tab and allows choice of either a Bootstrap with replacement, a Monte Carlo type of error analysis, or search for confidence limits by computing F-statistics.

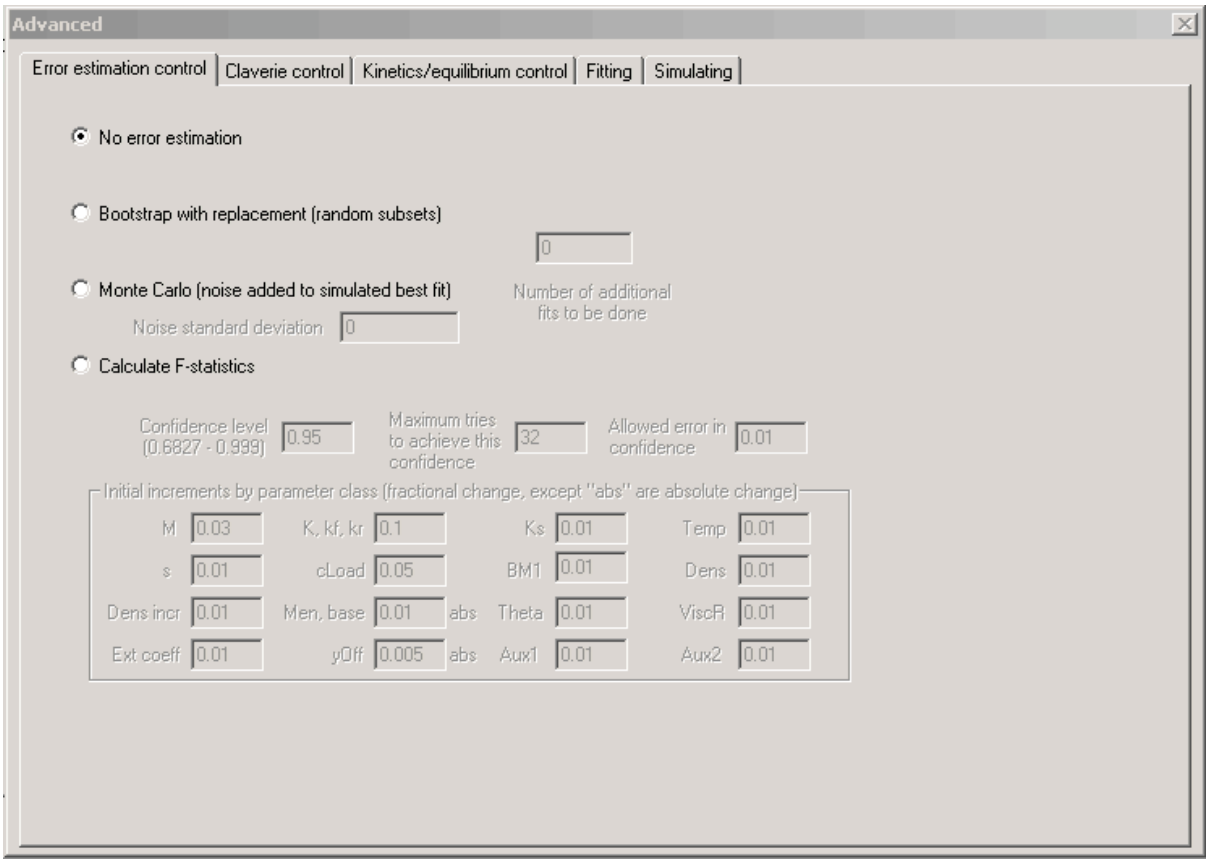
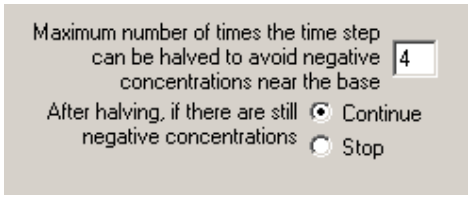


Figure 49. Advanced window pane.

2.4.1.5.2 Lamm Equation Solutions:

The second set of advanced parameters under the “Claverie control” tab allow adjustment of the performance of the finite element procedure for solving the Lamm equation and concerns the adjustable time increment used for each step and the distribution of grid points along the radial axis. The example below says essentially, if the fractional change in concentration does not exceed 0.02 at any point in the cell over that last time interval, then increase the time interval by a factor of 1.02. A smaller rate of change can be set for the first few steps in the box labeled "Sedimentation time step initial fractional change".

The boxes at the bottom of this tab allow termination of the finite element algorithm in the event of numerical instability :viz:



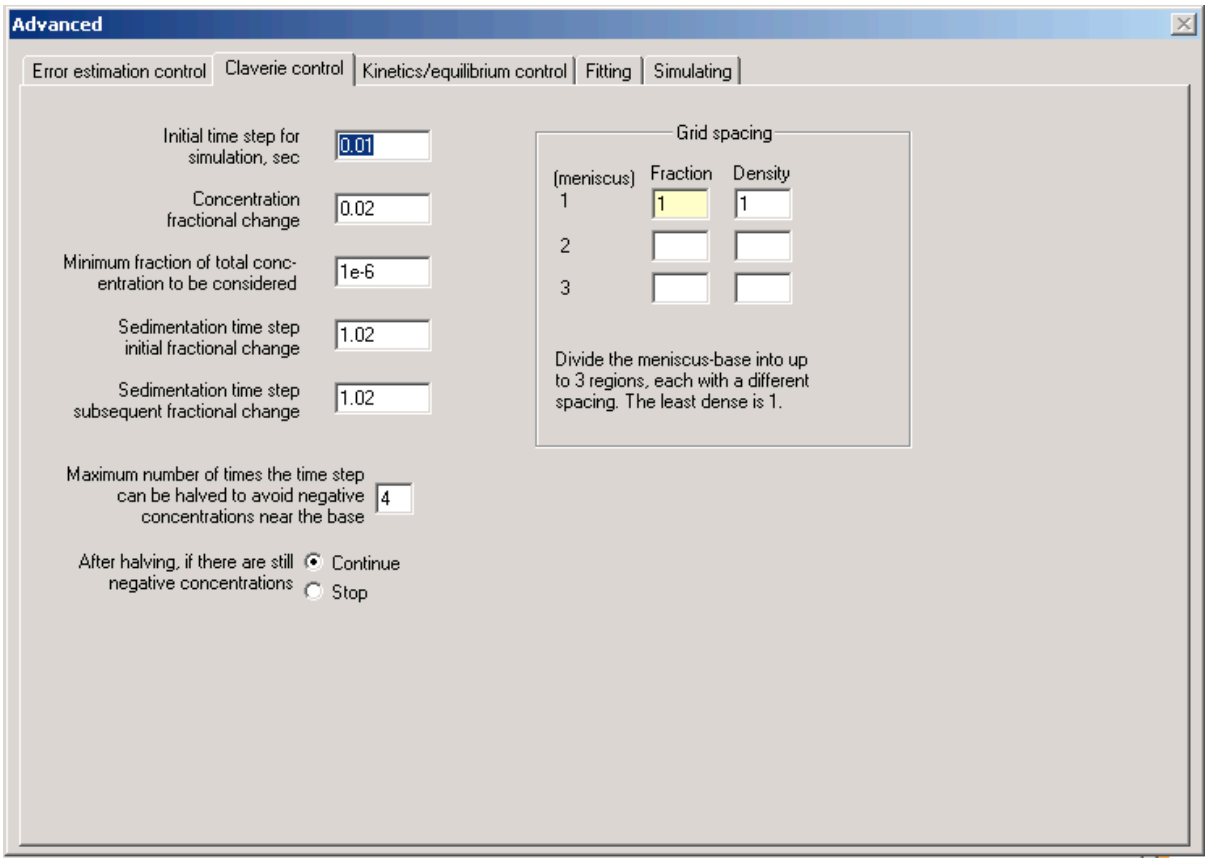
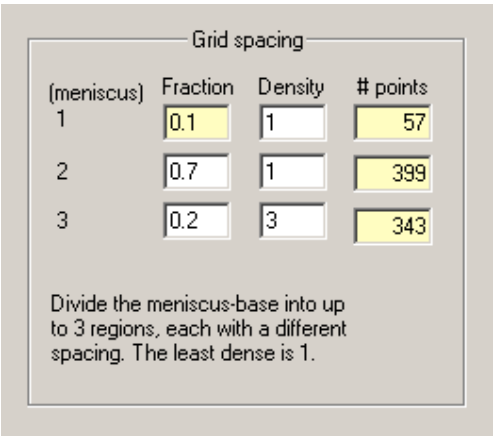


Figure 50. Control of the finite element calculations.

The grid spacing can be controlled as in the following example which shows the cell divided up into three regions spanning 0.1, 0.7 and 0.2 of the distance from the meniscus to the base. The point density and actual number of points used are shown in the designated boxes.



2.4.1.5.3 Kinetic integrator control

The third set of advanced parameters under the "Kinetic integrator control" tab allows one to adjust the performance of the ODE (Ordinary Differential Equation) solver. The current version of SEDANAL uses either the Bulirsch-Stoer algorithm (with Richardson rational polynomial extrapolation) or the Euler extrapolation method to solve the differential equations describing the kinetics. For those cases in which instantaneous equilibrium can be assumed, the equilibrium equations may be solved directly using the Newton-Raphson successive approximation iterative method. One method, kinetic or equilibrium solution, could be faster than the other depending on the model.

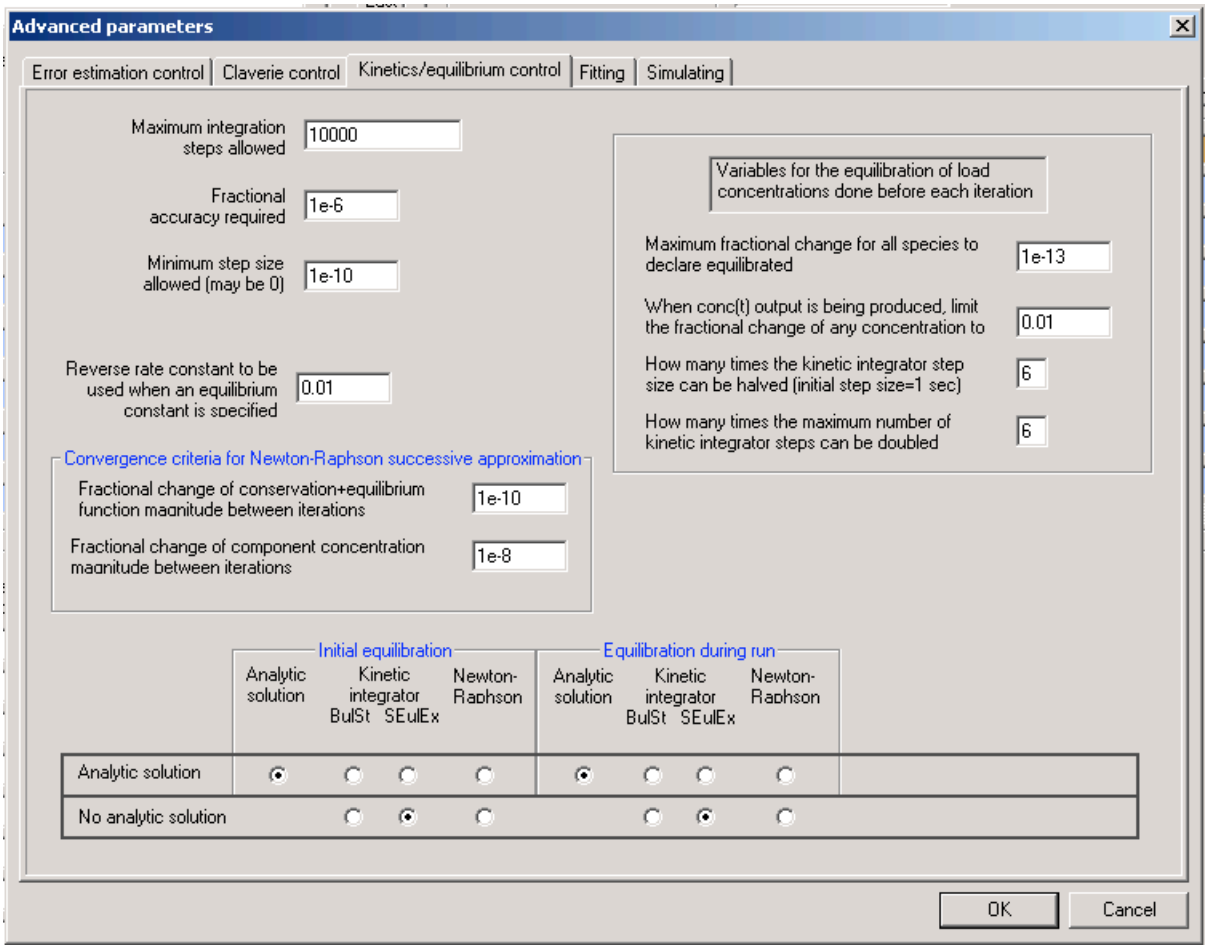


Figure 51. Control of the kinetic integrator.

After each time step of the Lamm equation solution, the updated concentrations of all the species are recomputed according to the kinetic rate constants.

2.4.1.5.4 Fitting Tab

First choose the fitting method, either simplex or Levenberg-Marquardt minimization. For simplex, the mode for setting up the initial simplex is controlled by selecting the forth tab "Fitting". In the upper right-hand area, The top button is to vary only the diagonal elements of the simplex array in the positive direction. The middle selection is experimental – don't use it. The last one (bottom button) varies all the elements by small positive and negative amounts. It seems to work best and is recommended. The

convergence criteria are entered in the bottom set of windows. Whether the fit is stopped, or perturbed and restarted is indicated in the "Perturb fit and re-do" section of this screen. Other settings here affect the display.

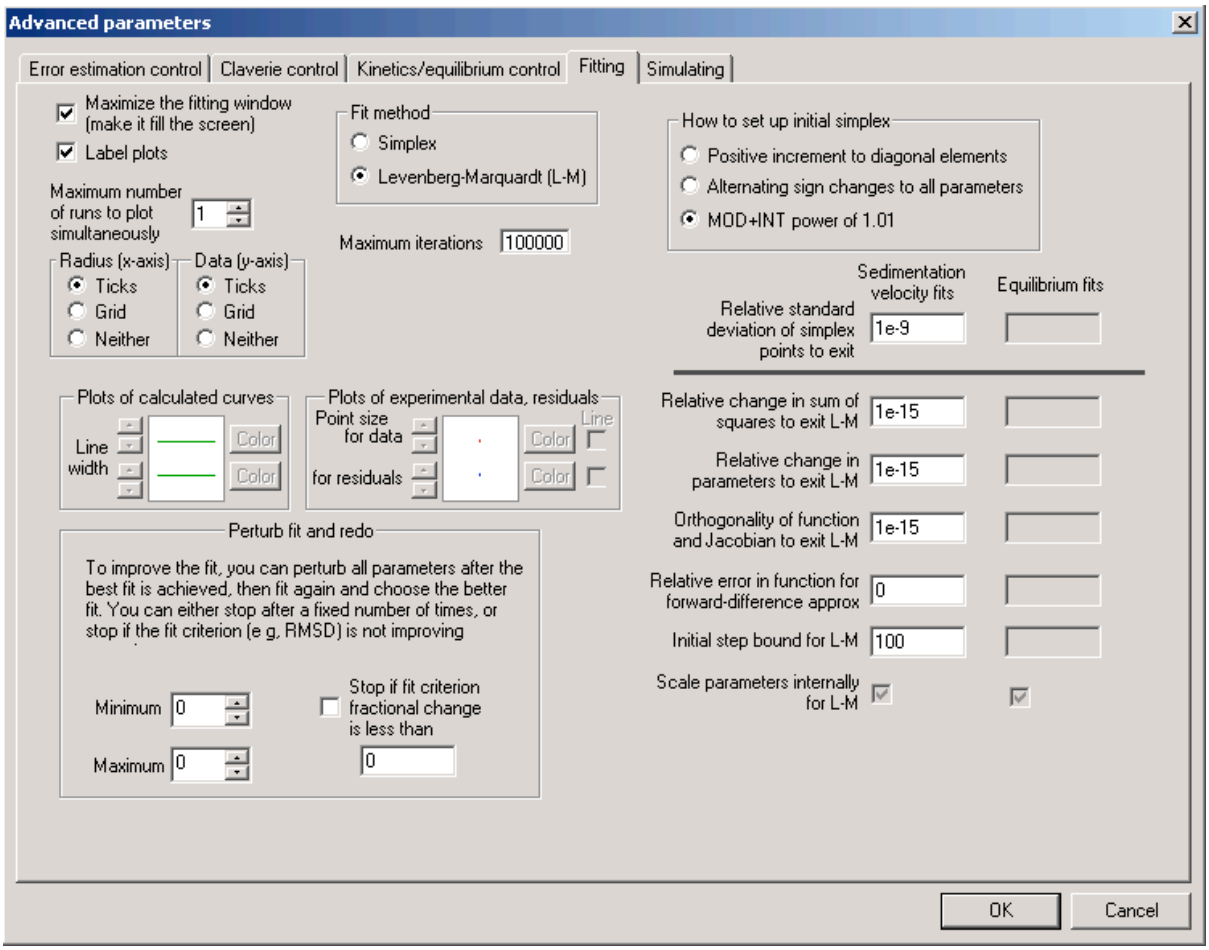


Figure 52. Choosing fitting parameters.

2.4.1.6 Concentration dependence of s and D.

Hydrodynamic concentration dependence of both s and D through the frictional coefficient is expressed through the coefficient k_s defined as follows: $f=f_{(c=0)}(1+k_s c)$. **Thermodynamic** concentration dependence of D is expressed through the second viral coefficient, BM_1, as follows. $D=D_{(c=0)}(1+2BM_1\ c)$. One may either fit for these parameters or hold them constant at preset values.

Therefore, the hydrodynamic concentration dependence for s is given by

$$\mathbf{s(c) = s_{(c=0)}/(1+k_s c)}$$

while the total concentration dependence for D, by

$$\mathbf{D(c) = D_{(c=0)}(1+2BM_1\ c)/(1+k_s c)}$$

2.4.1.7 Selecting data sets to be fitted:

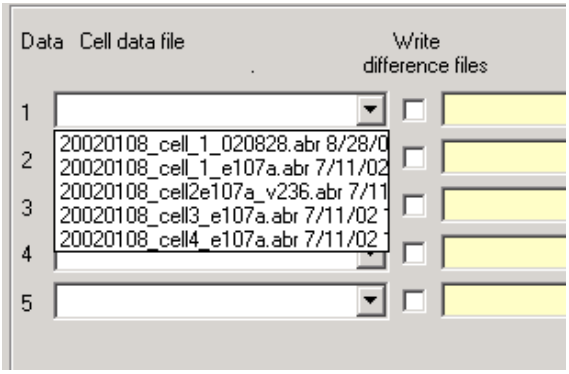


Figure 53. Once the experiment directory has been chosen, all the available preprocessed run files (*.abr) will be visible in the dropdown windows.

When the cells are chosen, one must enter guesses for the loading concentration of each component.

Selected scans within a cell may be fitted. The default is all scans; to change this, right-click on the run file name (lower left), a small window appears that allows you to either choose the scans to fit or to enter a weighting factor for that data set.

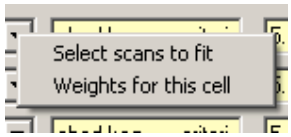


Figure 54A

If "Select scans to fit" is chosen, a window similar to the one for selecting scans to be plotted in the preprocessor will appear. You may select the scans to be used for fitting in that window. The result is written into the control file. There is an option in the lower right of the scan-select dialog box to use the same scans for fitting that had been chosen for plotting in the preprocessor (Figure 54B).

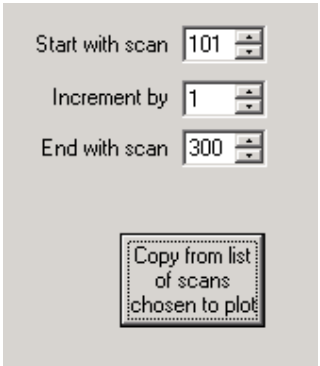


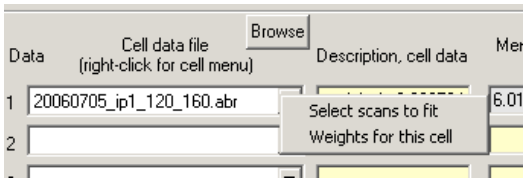
Figure 54B.

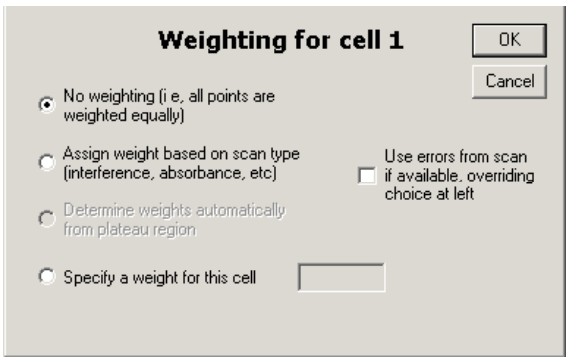
2.4.1.7.1 Removing a data set from the fit.

To remove data set from the fit, highlight the name in the “Cell data file” window (requires two left clicks) and press the “Delete” (not the “Backspace”) key (Figure 53).

2.4.1.7.2 Weighting Factors

To enter weighting factors for a particular data set, right-click on the cell data file name (the *.abr file) and select "weights for this cell"





2.4.1.8 Selecting parameters for the finite element solutions to the Lamm equation.

One must select the grid density, i.e. the number of points between the meniscus and the base, to be used for the fitting. The data are interpolated onto this grid and the finite element solutions to the Lamm equation are generated on this grid. The time interval is the initial time increment used by the finite element procedure. This time increment is adjusted according to the parameters enter under the "Advanced" menu.

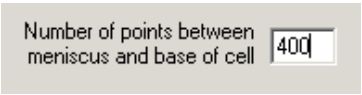


Figure 55.

For initial fits to a new system is recommended to choose 200 points between the meniscus and the base. This will make the exploratory fits relatively fast. The number of points should be increased to 400 or 800 later to refine the fits for the most accurate estimation of fitted parameters.

2.4.1.9 Total Number of iterations.

Criteria for stopping the fitting procedure are inputted on the Advanced "Fitting" tab. To have SEDANAL stop after a specified number of number of iterations, enter that number in the "Maximum iterations" window..

Figure 56. It is recommended that this number be set to a large value like 9999 in order to force SEDANAL use the stopping criterion indicted on the "Fitting" tab.

2.4.1.10 Extinction coefficients: Global fitting with multiple optical systems -

If several optical systems and/or wavelengths have been used, different mass extinction coefficients can be used for each cell. **(Note: the extinction coefficients to be entered must correspond to a 12 mm path length: i.e. one must multiply the usual (mg/mL) UV extinction coefficients by 1.2). Or if 1 or 3 mm centerpieces were used, the value must be multiplied by either 0.1 or 0.333 to correspond to actual path length used.**

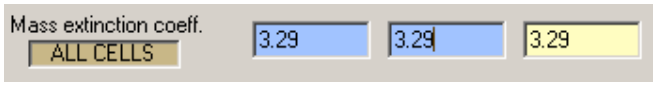


Figure 57. If all cells were run with interference optics, then the extinction coefficients would be entered by choosing the "ALL CELLS" option which is the default:

Figure 58. Left-clicking on the text "ALL CELLS" will allow one to enter different extinction coefficients for species for each cell. Each click advances to the next cell. If interference optics have been used for cell 1, one would input the number of fringes per 1 mg/mL:

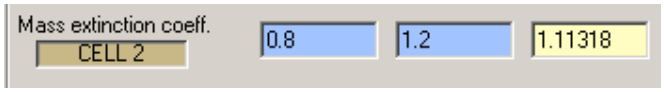


Figure 59. If absorbance at 280 nm had been used, one would click again on the text and enter the mass extinction coefficients of each species at 280 nm: where the last box is computed as a weight average based, in this case, on the model A+B=C. If the data were taken with different optical systems on the same cell, then the loading concentrations for those data will be the same. SEDANAL allows one to constrain the loading concentration to be identical for those data sets.

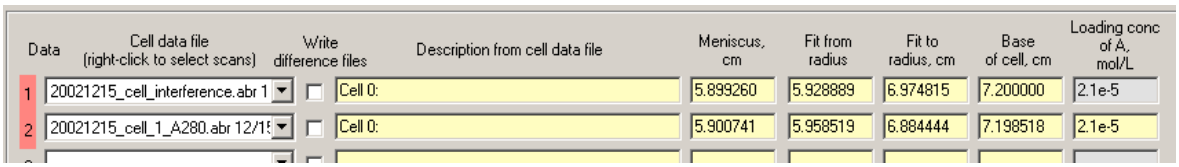


Figure 60. For example, if the data from "cell data files" 1 and 2 were taken from the same cell, their loading concentration will be identical. One may click on the Number "2" on the far left of the row to "slave" that data set to data set # "1".

If cell "3" had been taken at another wavelength on the same cell, then cell data files "2" and "3" would both be "slaved" to cell data file "1" by clicking on the number "3" at the left (not shown).

NOTE: Many of the relationships discussed in the following section can now be established in the Equation editor which is especially useful for cells not adjacent in the list shown in Figure 61 and for more complicated relationships between cells.

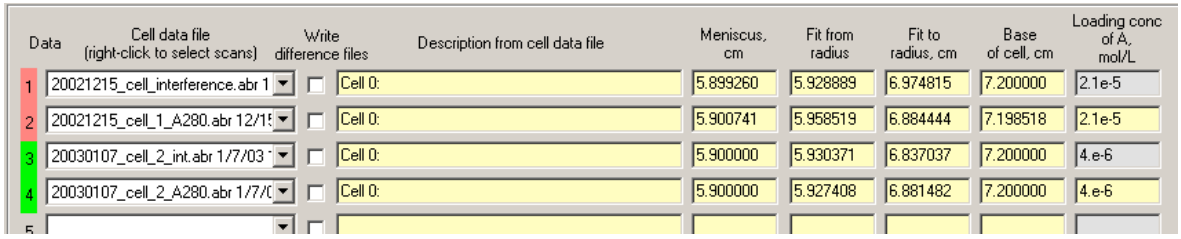


Figure 61. For example, if cell data files "1" and "2" are from one cell and cell data files "3" and "4", then click on the row for cell data file "2" to slave it to cell data file "1". Then click on the row for cell data file "4" to slave it to cell data file "3". Cell data set that are slaved are high lighted in the same color.

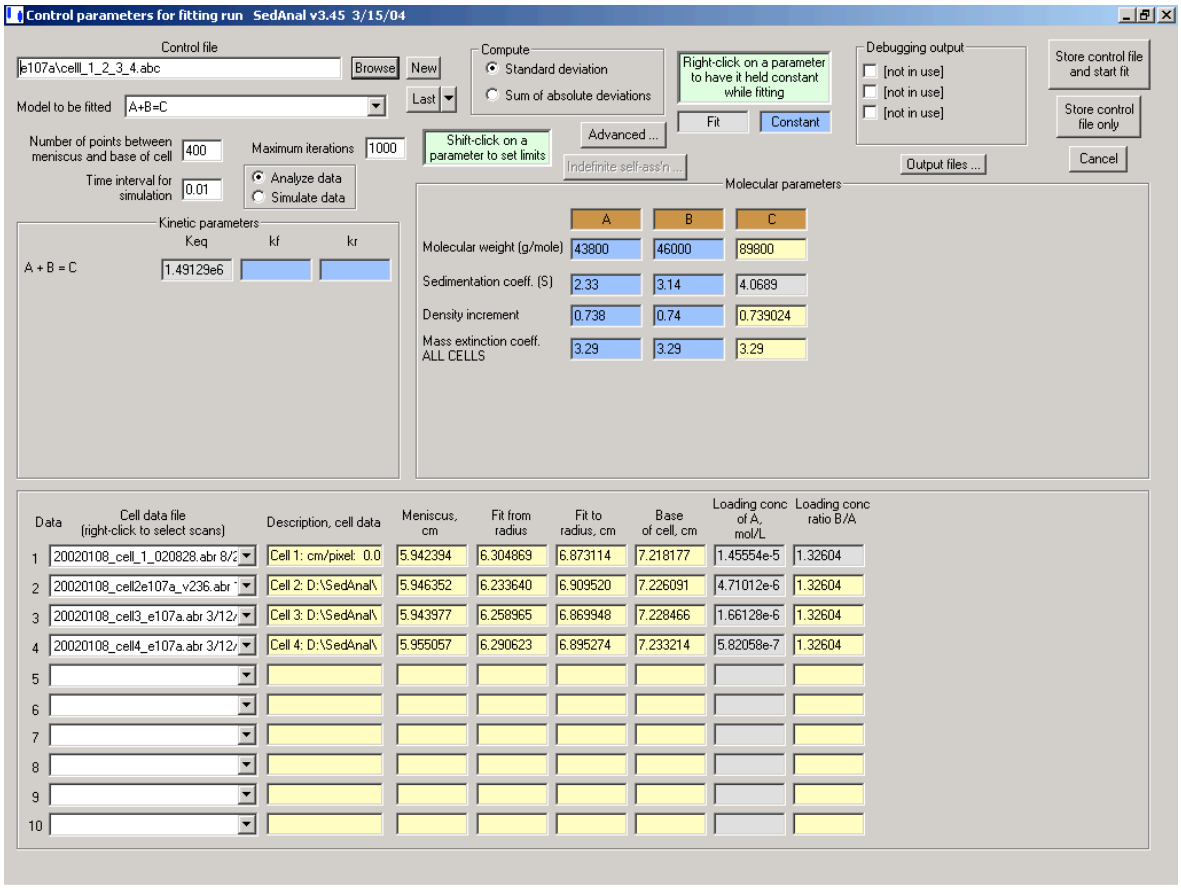


Figure 62. **The main control screen for the fitting preprocessed data with all parameters filled in**.

When all the parameters have been filled in, the screen will look like above.

NOTE: The loading concentrations are expressed in moles per liter. The loading concentration of component 2 (and components 3 and 4) is entered in terms of the molar ratio of B to A: $[B]_o/[A]_o$. For the 3 component system, the last column will be the ratio of $[C]_o/[A]_o$. This will allow one to specify that the ratio be fit as a global value for all runs, and is used in cases in which the runs are part of a dilution series for which these ratios might not be known but which are usually expected to be the same for all cells.

2.4.1.11 Local vs global parameters - changing

To change the ratio from a local to a global parameter left-click on the text "Loading conc ratio B/A" or "Loading conc ratio C/A" above the column of boxes. The first click will fill all the boxes with the value entered into the top box and propagate that value to the other cells and change to background color to yellow to indicate that these are not being fit separately. All cells will use the same ratio in the fitting process.

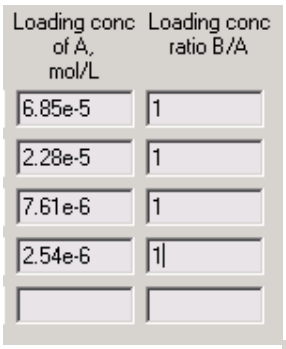


Figure 63. The loading concentration ratio is entered as 1.0 in this example.

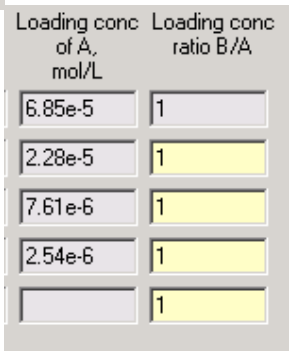


Figure 64. A single left-click on the text "Loading conc ratio B/A" propagates the top 1 into the boxes below with a yellow background to indicate that these values are derived from cell #1 and will be kept equal to and vary with the value for cell #1 as a single global parameter.

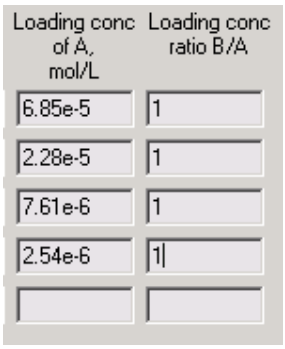


Figure 65. A second click on the text "Loading conc ratio B/A" turns the backgrounds back to gray to indicate that these values will be now be allowed to float independently for all cells:

Figure 66. Right clicking on one of the gray boxes will turn its background blue and now its value would be held constant during the fit, if that were desired:

2.4.1.12 Indefinite Self-association

2.4.1.12.1 Isodesmic case:

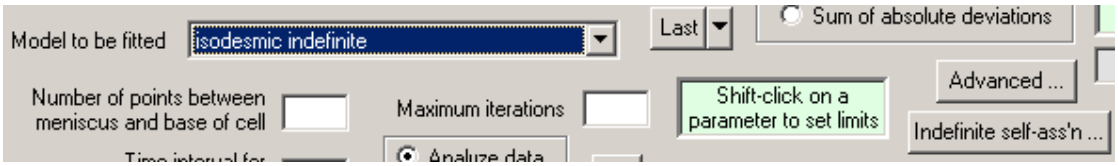


Figure 67. On the control screen, if an isodesmic model has been chosen, a button labeled "Indefinite self-ass'n…" will become active. Click the "Indefinite self-ass'n…" button to enter the coefficients for a polynomial representing the relationship between the sedimentation coefficient of the oligomers and the degree of polymerization (Figure 68).

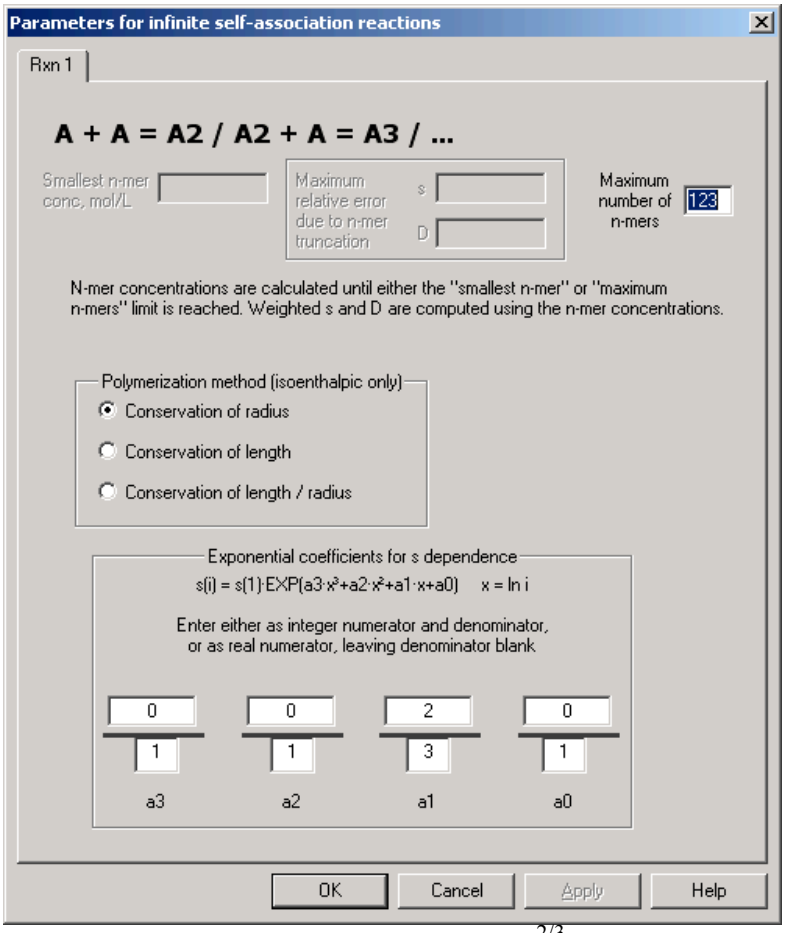


Figure 68. This example is using the s(i-mer) = s(monomer)* $(i)^{2/3}$.

Other coefficients, for end-to-end polymerization for example, can be entered from bead modeling or other theoretical considerations. The coefficients can be entered as rationals in the numerator box or as fractions by specifying both numerator and denominator.

The coefficients, a0, a1, etc… represent the coefficients of the following polynomial

ln(s(i-mer)) = a0 + a1*ln(i) + a2*(ln(i))2 + a3*(ln(i))**3.**

The coefficients are obtained by fitting a polynomial to the values of ln(s(i-mer)) as a function of ln(i). Usually a0 will be set to ln(s(monomer). The values of s(i-mer) can be obtained from the analysis of bead modeling, for example.

2.4.1.12.2 Isoenthalpic indefinite self-association:

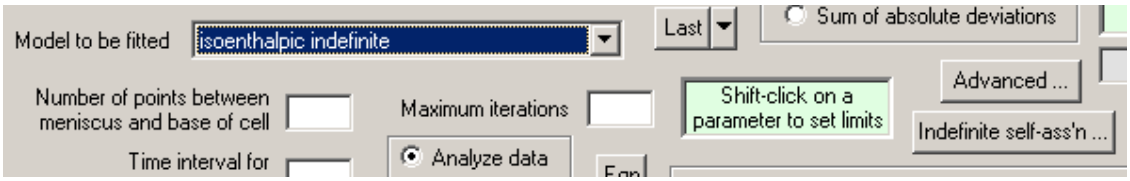


Figure 69. On the control screen, if an isoenthalpic model has been chosen, a button labeled "Indefinite self-ass'n…" will become active:

Click the "Indefinite self-ass'n" button to enter the coefficients for a polynomial representing the relationship between the sedimentation coefficient of the oligomers and the degree of polymerization. Then select the polymerization method according to how the entropy terms will be handled. [For details, see Ronald Chatelier, "Indefinite isoenthalpic self-association of solute molecules", Biophysical Chemistry, 28, 121-128 (1987)]

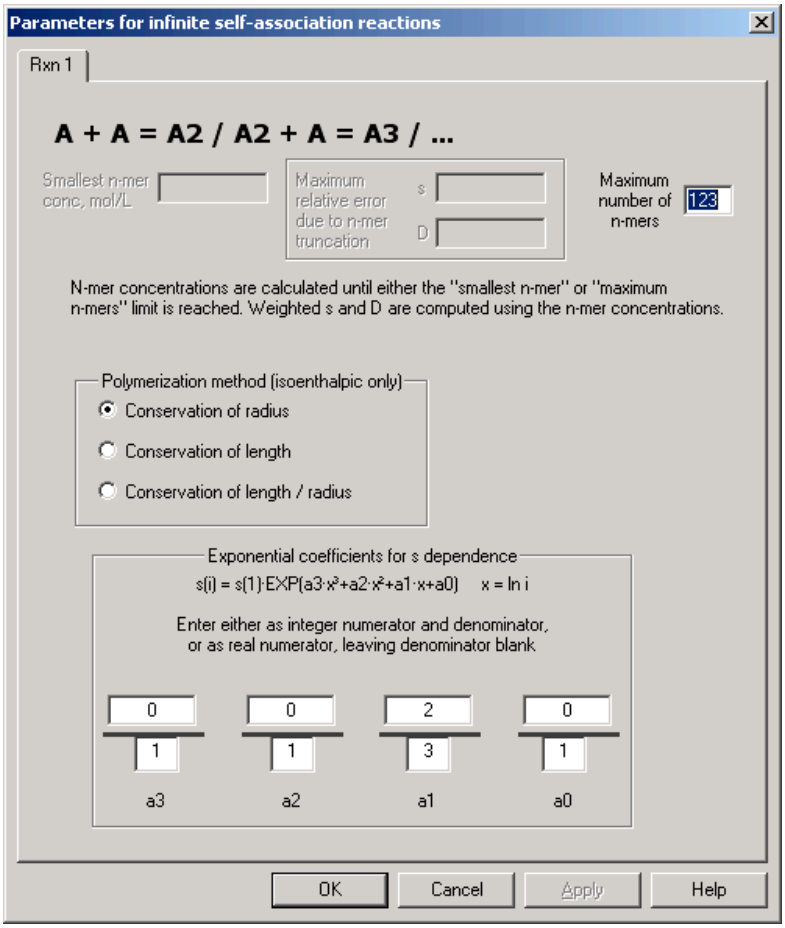


Figure 70. **Entering parameters for indefinite self-association**

2.4.1.13 Constraining the range of parameter values during fitting.

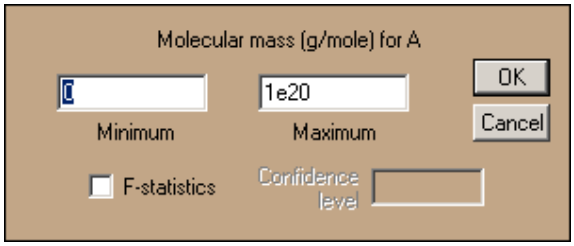


Figure 71. The range of allowable values for a parameter may be set by **shift-left-clicking** on the value in the control screen (remember to check these limits against those determined by the ModelEditor - you cannot override those limits here):

2.4.1.14 Exiting the Control Screen to start fitting

Figure 72. After all the appropriate boxes are filled in, the fit can be started by clicking the button labeled "**Store control file and start fit**" or if you just want to store this control file and maybe open up another one, click the button labeled (you guessed it) "**Store control file only**".

2.4.1.14.1 Fitting Screen

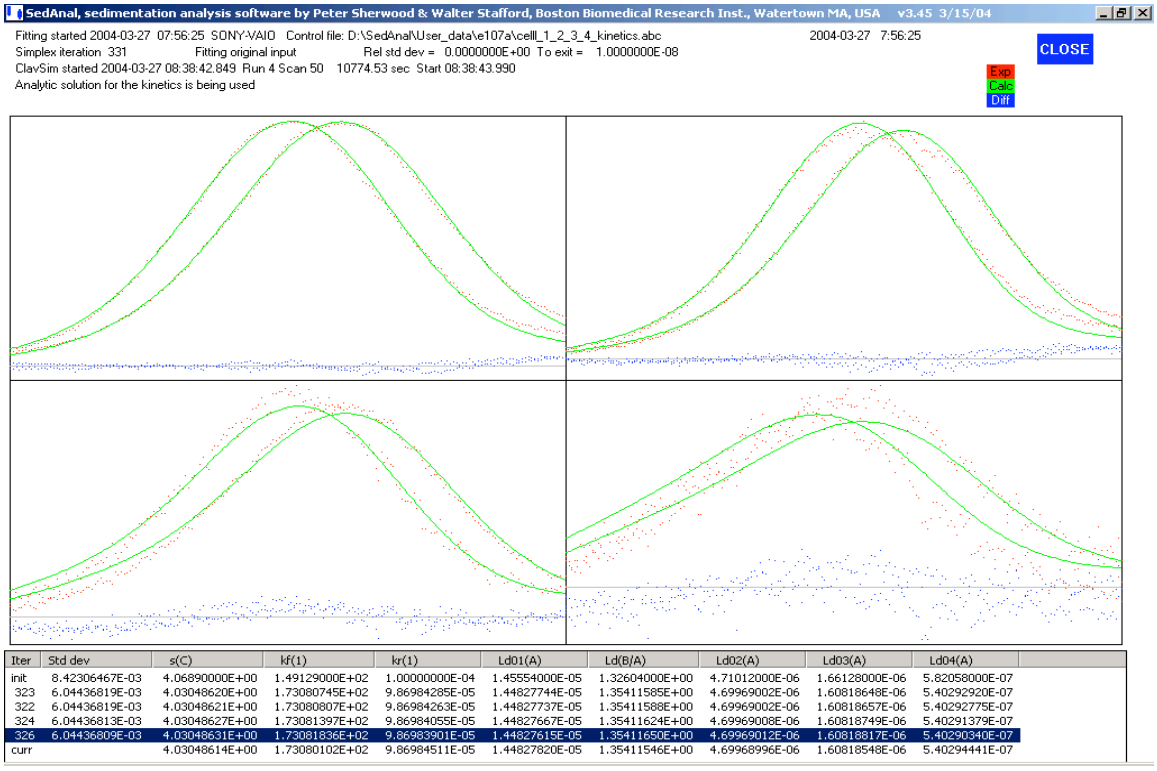


Figure 73. Here is an example of what the fitting screen looks like after a few iterations when performing a 4 cell global fit to the system A+B=C:

During the fitting process, the screen is updated every 10 seconds or each time a new minimum is reached. SEDANAL will print at the bottom of the screen 7 lines of text: (1) column titles, (2) the initial guesses, (3-6) the last 4 best minima, (7) the current guess being processed. The middle region of the screen displays 2 sets of three plots for <u>each</u> cell. The screen above shows 3 cells. In the left panel, the first set of 3 plots is from the first difference curve and the second set is from the last difference curve of the data set. The intermediate difference curves are not displayed; these can be found in the output file (with extension *.min). Within the window for each cell are plotted (1)the difference data being fitted (red dots), (2)the best fit to that difference data (green line) and (3) the residuals for the current best fit (blue dots).

After the fitting process, files are written that contain the data, the fit and the residuals. These files are named "<control_filename_prefix>__0nMin00m.txt" where "n" is the cell number and "m" is the sequence number corresponding to the number of times that control file has been used to start a fit.

The first few lines of a "Min" file look like this:

```
67  12/4/02                    08-Dec-02  23:14:39  SONY-VAIO
Control file: D:\SEDANAL-v2.66\User_data\experiment 1\e107a.abc 2002-12-08  23:14:39
Dataset  1     Minimum 216  Iteration 464
Std dev = 8.8549039430449259E-03 Params: 1.4279486706740931E-05 1.3736775805816817E+00 4.7200407471541204E-06
1.7162984153870679E-06 5.6099736267902298E-07 8.0013037192034372E+05
4.2651157859572259E-13 1.6918920578250105E-02
#  Radius   delta-C(obs)  delta-C(calc)    DEV         [...]
58  6.30782  0.45968E-01   0.30857E-01   -0.15111E-01    [...]
59  6.31423  0.54139E-01   0.32602E-01   -0.21538E-01    [...]
```

The "Min" file header contains information about the fit. Most of the lines are self-explanatory.
The global root mean square deviation for the fit is labeled "Std dev =" The fitted parameters corresponding to the last minimum are listed after the word "Params:" These will be labeled more clearly in a future version.

After the first two columns, "#" and "Radius", the min file contains, in groups of three columns, the entire set of difference curves from the experimental data, "delta-C(obs)", as well as all the difference curves generated by the fitting procedure, "delta-C(calc)" and then the deviations between the data and the fit for each difference curve, "DEV".

"#" is the point number
"Radius" is the radius
"delta-C(obs)" is the first difference curve from the experimental data,
"delta-C(calc)" is the first difference curve generated for the same times
"DEV" is the residual at that point for the first difference curve

From this point on, the last three columns are repeated for each difference curve:
delta-C(obs) delta-C(calc) DEV delta-C(obs) delta-C(calc) DEV …etc…

These data can be imported into the plotting program of your choice.

2.4.1.14.2 Screen Dumps

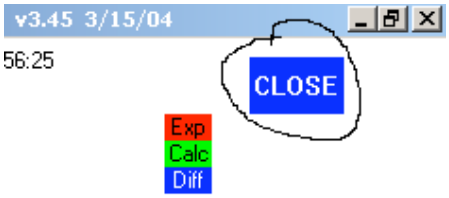
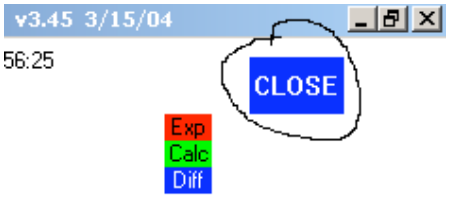


Figure 74. The screen can either be dumped directly to a printer at any time during the fitting process by right clicking on the "Stop/Close" button: Doing so will cause a printer dialog window to open. The screen can be dumped as a bit map to a file by shift-right-clicking on the Stop/Close button. This is done

silently and no dialog window will appear. The bit map will be found in a bmp file with same name and sequence number as the report file.

2.4.1.14.3 Toggling Plots During Fitting

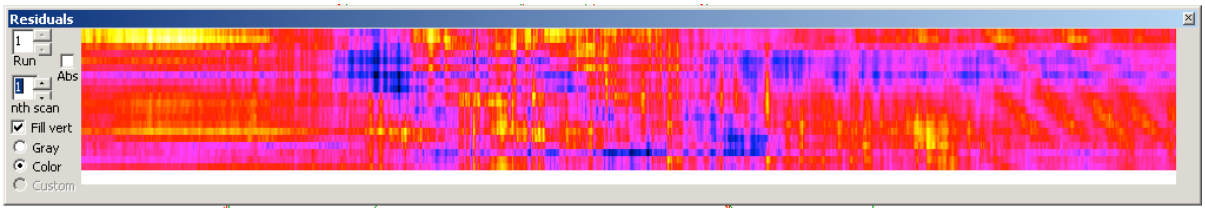


As long we have this little picture handy … One can turn on or off any of the plots on the fitting screen by clicking the little red, green or blue boxes sporting the text "Exp", "Calc" or "Diff" corresponding to the plot of the experimental data, the best fit, or the residuals, respectively.

2.4.1.14.4 Displaying the Residuals Bitmap

Residuals can be displayed in either grey scale or on a color scale by clicking on the "R" button on the upper right hand corner of the fitting window:

The residuals window shows the difference between observed and calculated difference curves (the residuals). the colors can be scaled either relative to the minimum and maximum values or to absolute minimum and maximum user specified values by clicking on the "abs" button.



The plot in the residuals window has x = the radial point, y = the scan number, and the color = the residual value. Only one run is shown at a time. The range of x is the user-selected range to fit. The color mapping is determined in one of three different ways, and scaled to the range of data, z, in the run:

Gray scale: *intensity* = 255 $\left((z - z_{min}) / (z_{max} - z_{min})\right)$, so z_{min} gives black and z_{max} gives white.

Color: A 16-bit look-up table is used (65,535 colors) to convert $p = 2^{16}\left((z - z_{min}) / (z_{max} - z_{min})\right)$ to a color. The table is shown below. Again, z_{min} gives black and z_{max} gives white.

Custom: Same as color, but the look-up table is user-specified, read in from the file `ResidualColors.txt`. The format of `ResidualColors.txt` is

```
// Optional comments
0 0 0 0
1 0 1 0
2 1 0 0
...
65534
65535
```

Each line is *p Red Green Blue*, where the color values are an 8-bit intensity 0–255.

Horizontal and/or vertical scroll bars appear if the plot window is not large enough to accommodate the data (current size is 930 radial points and 132 scan pairs). Clicking the scroll bar arrows moves one pixel (i e, one radial point or one scan pair), while clicking the bar outside the thumb moves half the size of the plot window.

The run to be shown is selected with a thumbwheel. Only a single run's residuals are visible.

Normally, every scan is displayed (one scan per row of pixels). This can be changed with the nth scan thumbwheel to show only every n^{th} scan in adjacent rows.

2.4.1.15 Fitting to Sedimentation Equilibrium Data:

2.4.1.15.1 Loading the data set - load equilibrium data sets before choosing the model

Open the control screen and click "New' and set the path and create a new control file. Select one of the cell data files. (Selecting a cell data file before a model is selected is not mandatory but is less confusing than the other way around). Selecting the cell data file tells the Fitter whether this is SedEq or SedVel data and arranges the boxes appropriately when the model is chosen (Fig. 75).

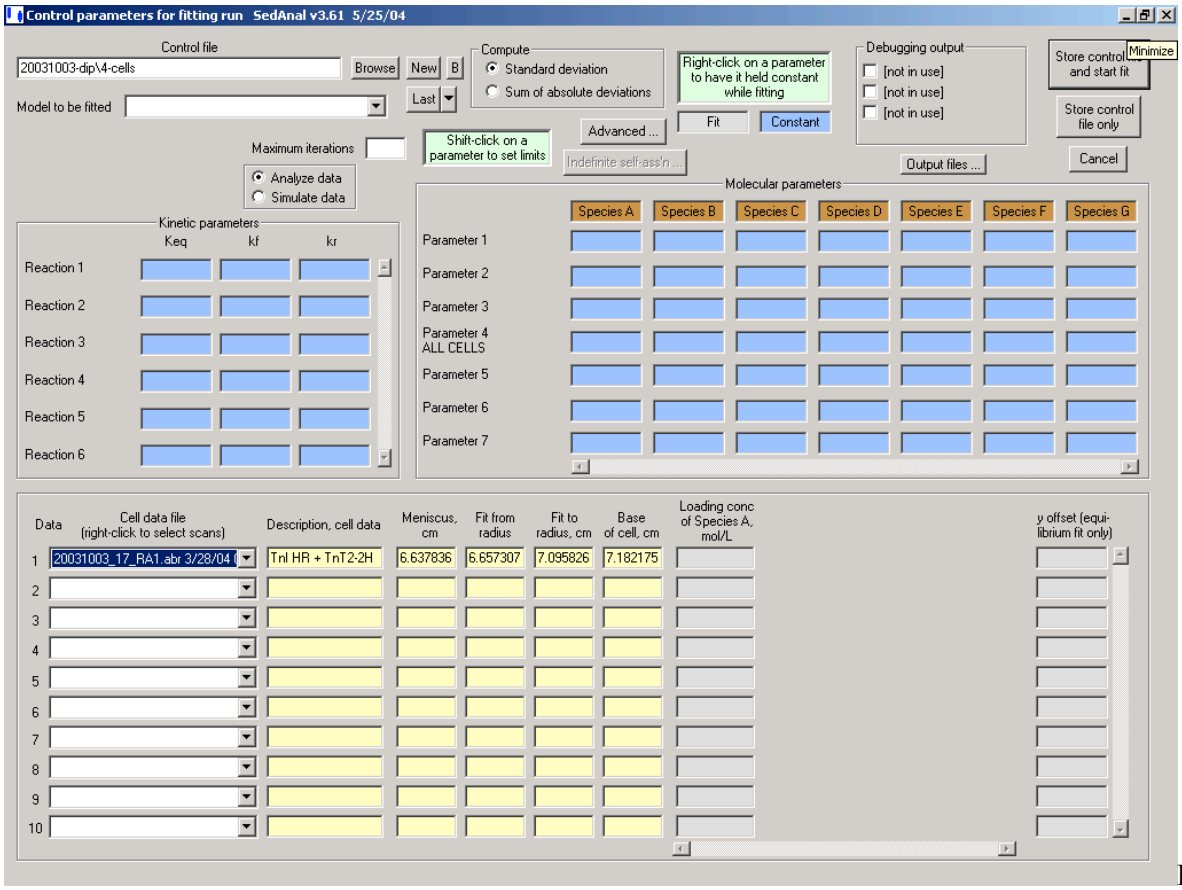


Figure 75. The main control screen for the fitting preprocessed data

2.4.1.15.2 Choose a model

Choose a cell data file first (Fig. 75) and then choose a model (in this case for a single species) and now the control screen changes to the appropriate sedimentation equilibrium model (Fig. 76).

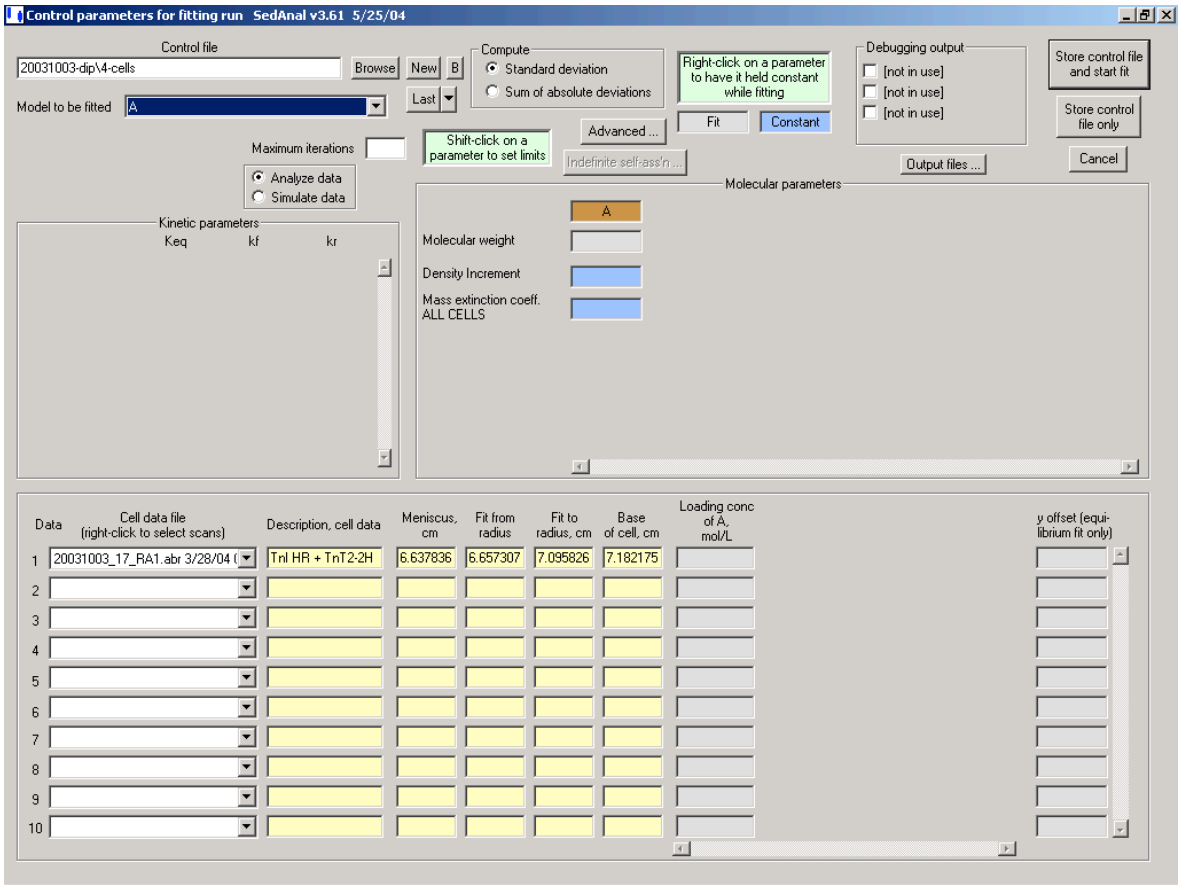


Figure 76. The main control screen for the fitting of preprocessed data for a sedimentation equilibrium run.

For an A+B=C model the screen would change to (Fig. 77). In addition to the to molar mass, density increments, extinction coefficients, an additional "local" cell parameter, y-offset, has been added to account for vertical offsets in the data. It is absolutely necessary to include this parameter in fit of interference data since the fringe displacements are known only to within an additive constant. Often it is necessary to include a y-offset to fit absorbance data if there are buffer mismatches or optical system calibration issues.

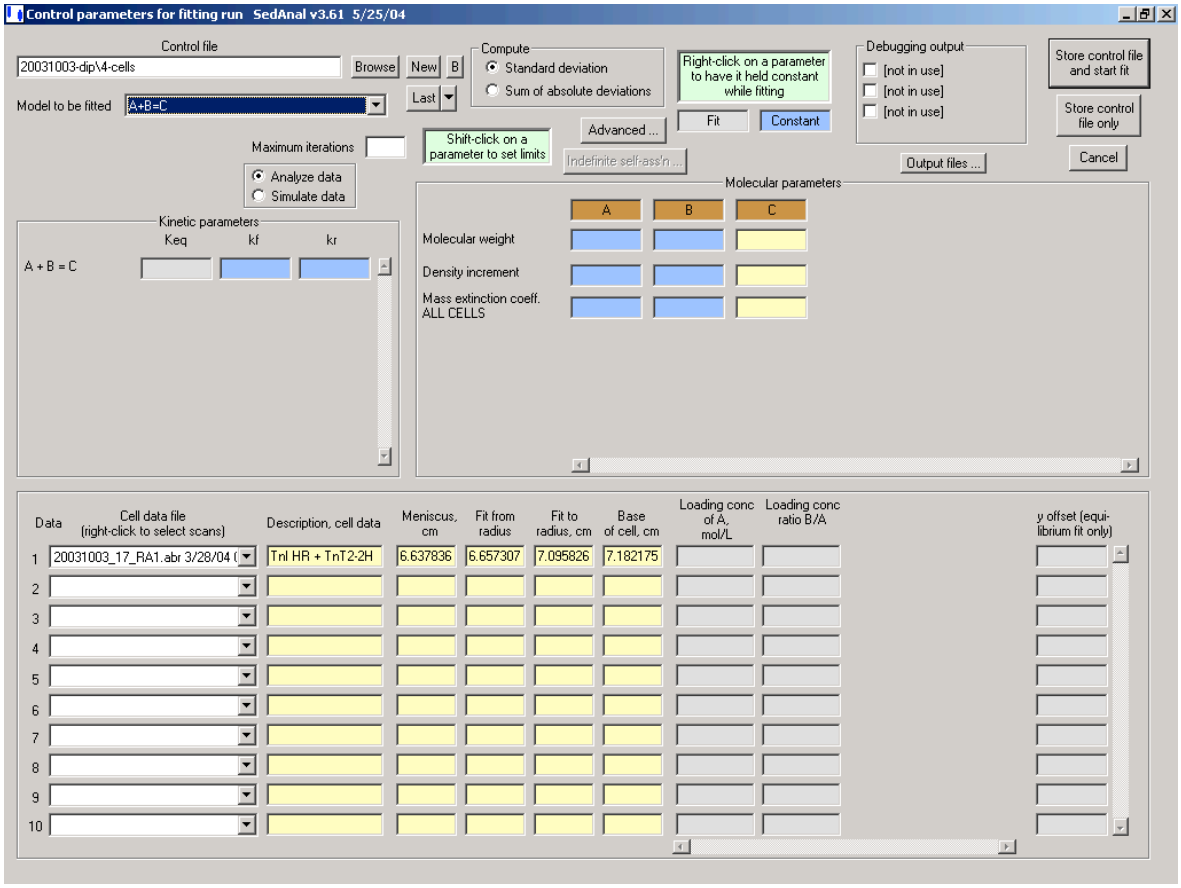


Figure 77. The main control screen for the fitting of preprocessed data for sedimentation equilibrium run after changing the model from a single species to A + B =C.

2.4.1.15.3 Enter initial guesses and other parameters

After filling in guesses for the parameters for a single species fit, the screen looks like the one below (Fig. 78).

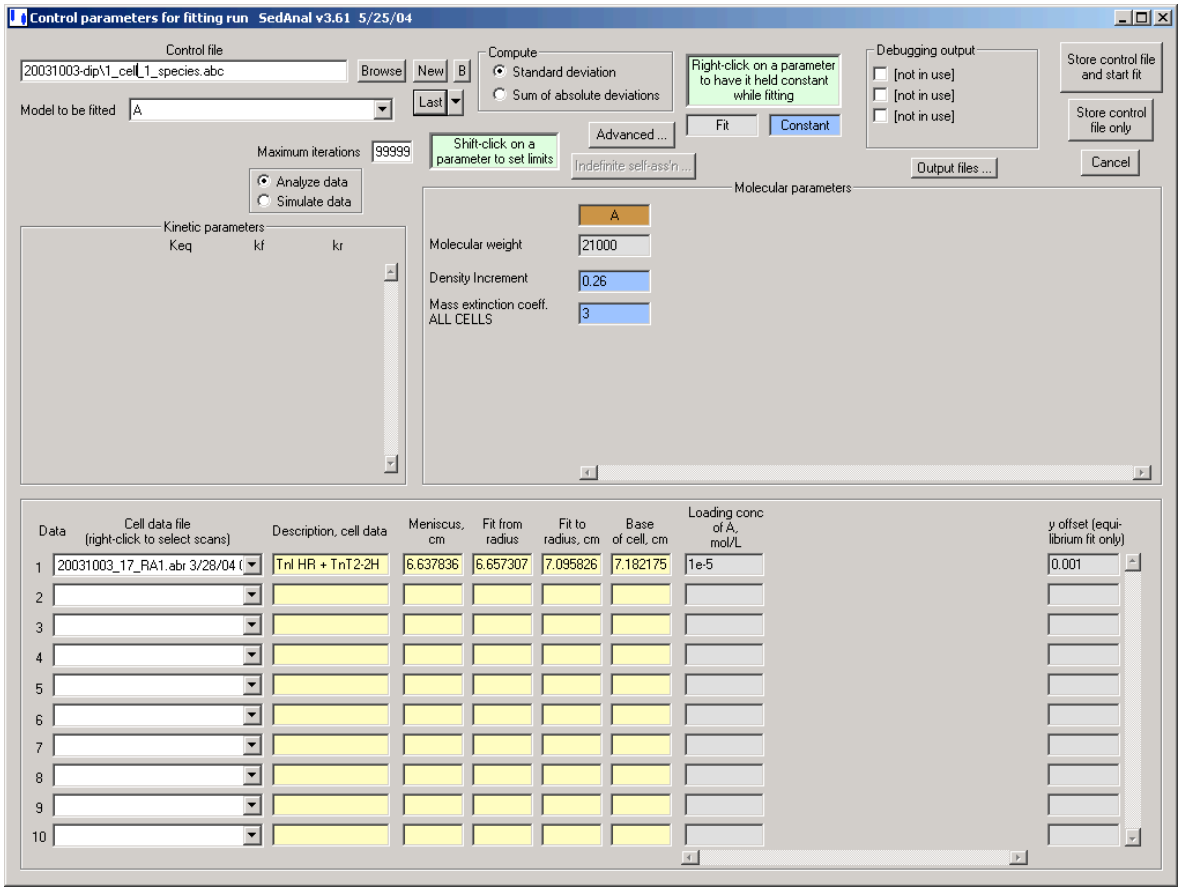
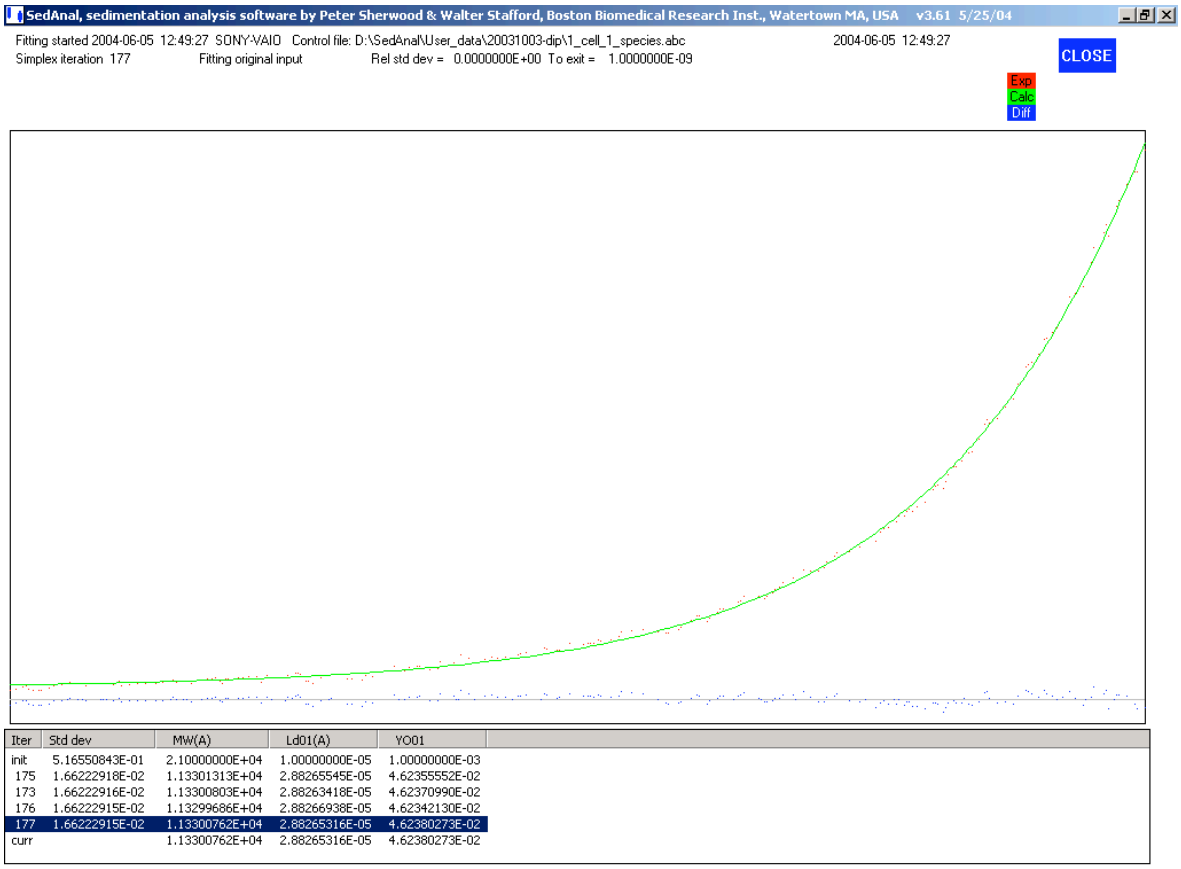


Figure 78. The main control screen for the fitting of preprocessed data for a sedimentation equilibrium run.

2.4.1.15.4 Start the fit

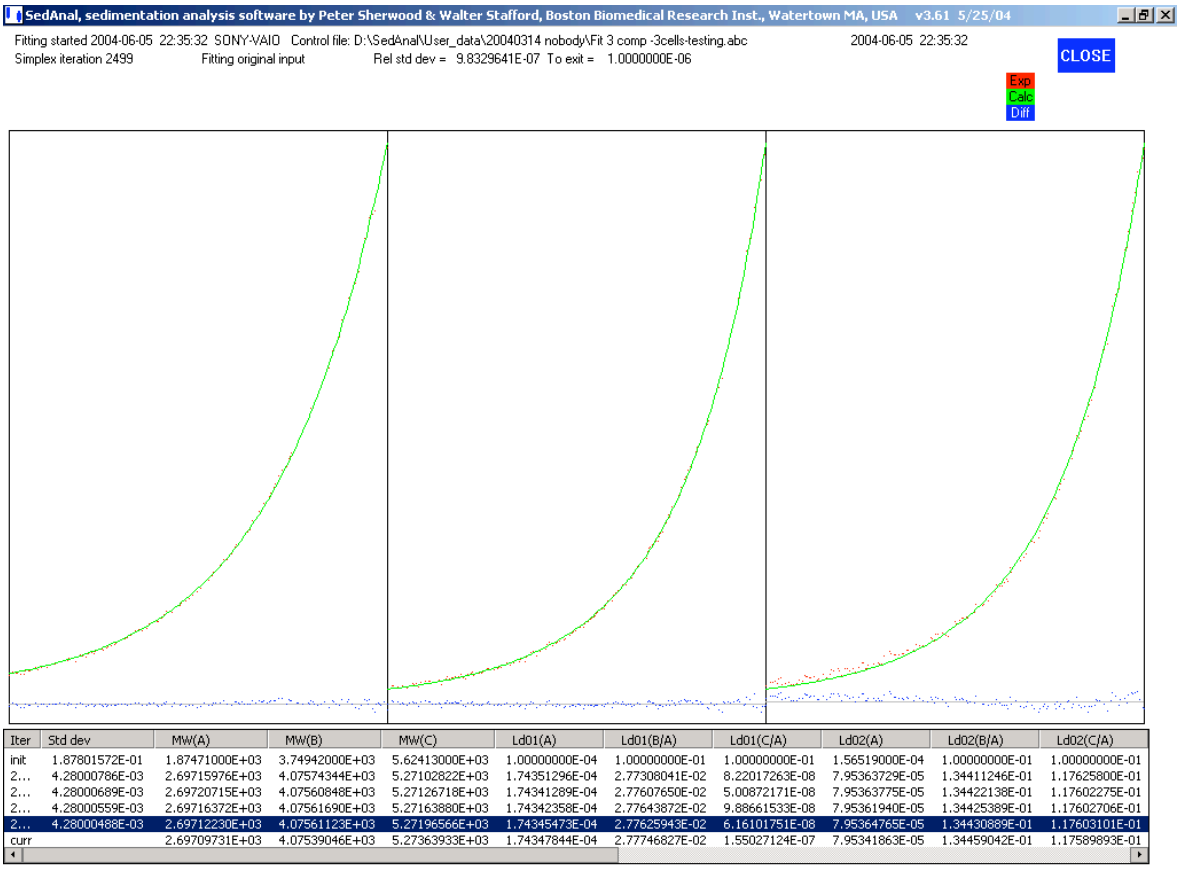
Now, after pressing “Store … and start fit”, we get to the following fitting screen at convergence (Fig. 79).



Iter	Std dev	MW(A)	Ld01(A)	YO01
init	5.16550843E-01	2.10000000E+04	1.00000000E-05	1.00000000E-03
175	1.66222918E-02	1.13301313E+04	2.88265545E-05	4.62355552E-02
173	1.66222916E-02	1.13300803E+04	2.88263418E-05	4.62370990E-02
176	1.66222915E-02	1.13299686E+04	2.88266938E-05	4.62342130E-02
177	1.66222915E-02	1.13300762E+04	2.88265316E-05	4.62380273E-02
curr		1.13300762E+04	2.88265316E-05	4.62380273E-02

Figure 79. Fit to equilibrium data after convergence.

For a multi-cell fit of a 3 component model to 3 data sets spanning 3 loading concentrations at one speed we might see (Fig. 80).



Iter	Std dev	MW(A)	MW(B)	MW(C)	Ld01(A)	Ld01(B/A)	Ld01(C/A)	Ld02(A)	Ld02(B/A)	Ld02(C/A)
init	1.87801572E-01	1.87471000E+03	3.74942000E+03	5.62413000E+03	1.00000000E-04	1.00000000E-01	1.00000000E-01	1.56519000E-04	1.00000000E-01	1.00000000E-01
2...	4.28000786E-03	2.69715976E+03	4.07574344E+03	5.27102822E+03	1.74351296E-04	2.77308041E-02	8.22017263E-08	7.95363729E-05	1.34411246E-01	1.17625800E-01
2...	4.28000689E-03	2.69720715E+03	4.07560848E+03	5.27126718E+03	1.74341289E-04	2.77607650E-02	5.00872171E-08	7.95363775E-05	1.34422138E-01	1.17602275E-01
2...	4.28000559E-03	2.69716372E+03	4.07561690E+03	5.27163880E+03	1.74342358E-04	2.77643872E-02	9.88661533E-08	7.95361940E-05	1.34425389E-01	1.17602706E-01
2...	4.28000488E-03	2.69712230E+03	4.07561123E+03	5.27196566E+03	1.74345473E-04	2.77625943E-02	6.16101751E-08	7.95364765E-05	1.34430889E-01	1.17603101E-01
curr		2.69709731E+03	4.07539046E+03	5.27363933E+03	1.74347844E-04	2.77746827E-02	1.55027124E-07	7.95341863E-05	1.34459042E-01	1.17589893E-01

Figure 80. Three-cell global fit to sedimentation equilibrium data

2.4.1.15.5 Global fit to a dilution series with <u>linked cells</u>

Global fit to 9 data sets from a run at 3 speeds and 3 loading concentrations to the Model A+B=C: The control screen was set up as follows with linked cells (note colored bars on the left side) and, since this is a dilution series, with fixed molar ratios of $[B]_o/[A]_o$

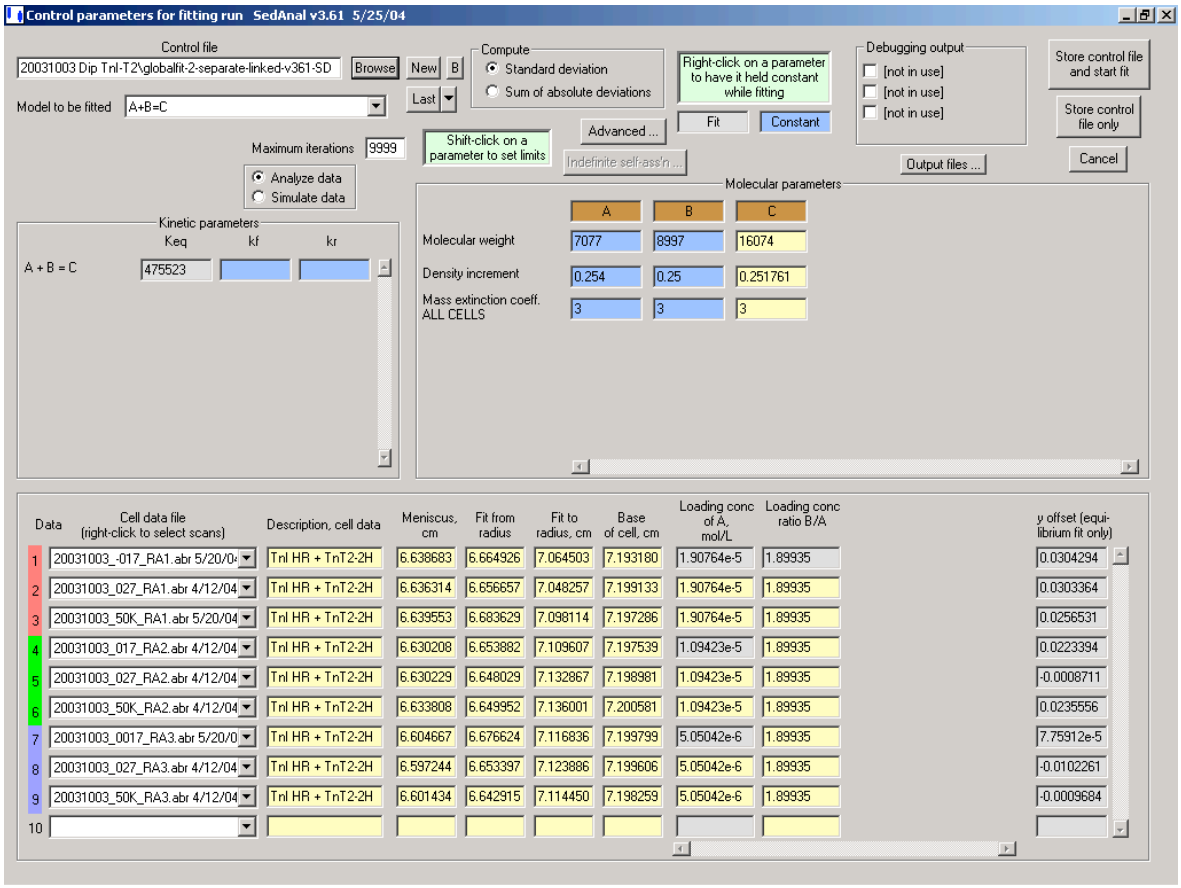
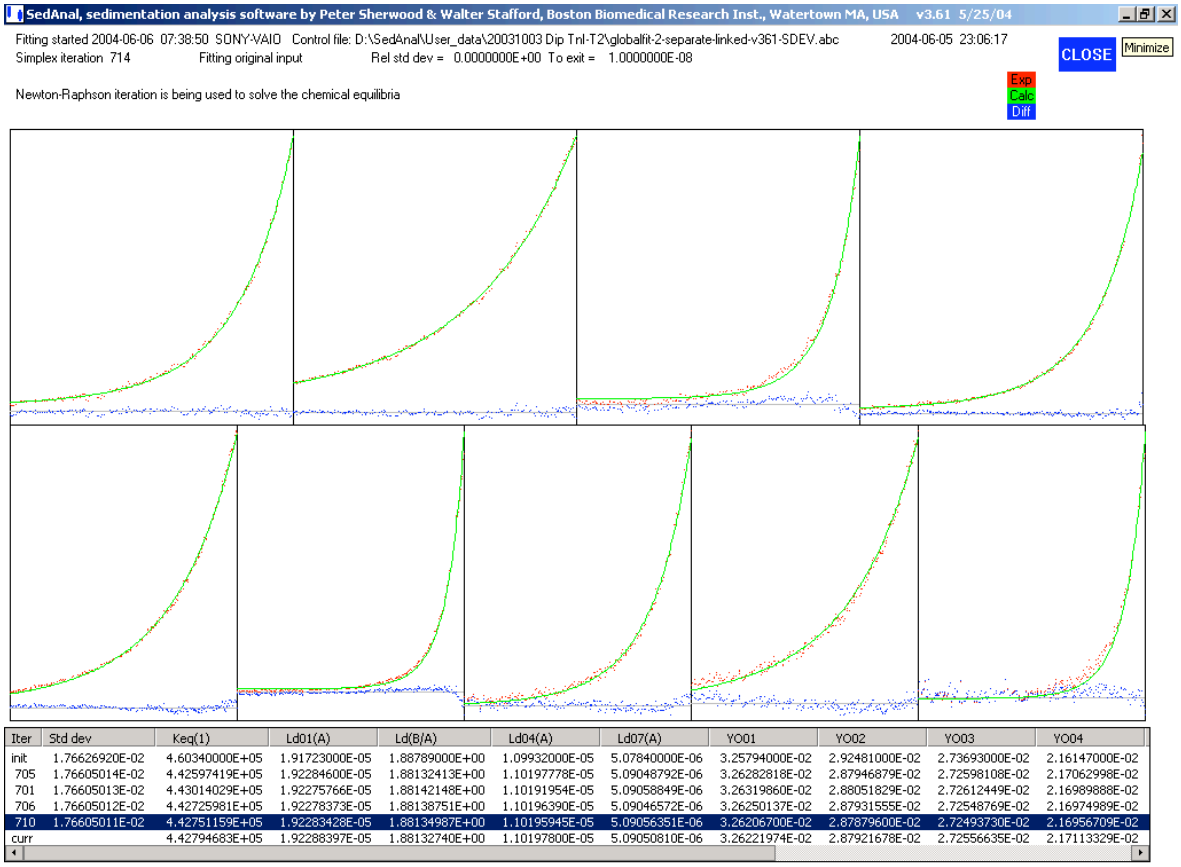


Figure 81. The main control screen for the fitting preprocessed data for sedimentation equilibrium run.

This constrained fit (i.e. requiring the cell loading concentration to be the same for the same cell at different speeds) resulted in a fairly good fit but having some systematic errors in a few of the data sets.



Iter	Std dev	Keq(1)	Ld01(A)	Ld(B/A)	Ld04(A)	Ld07(A)	YO01	YO02	YO03	YO04
init	1.76626920E-02	4.60340000E+05	1.91723000E-05	1.88789000E+00	1.09932000E-05	5.07840000E-06	3.25794000E-02	2.92481000E-02	2.73693000E-02	2.16147000E-02
705	1.76605014E-02	4.42597419E+05	1.92284600E-05	1.88132413E+00	1.10197778E-05	5.09048792E-06	3.26282818E-02	2.87946879E-02	2.72598108E-02	2.17062998E-02
701	1.76605013E-02	4.43014029E+05	1.92275766E-05	1.88142148E+00	1.10191954E-05	5.09058849E-06	3.26319860E-02	2.88051829E-02	2.72612449E-02	2.16989888E-02
706	1.76605012E-02	4.42725981E+05	1.92278373E-05	1.88138751E+00	1.10196390E-05	5.09046572E-06	3.26250137E-02	2.87931555E-02	2.72548769E-02	2.16974989E-02
710	1.76605011E-02	4.42751159E+05	1.92283428E-05	1.88134987E+00	1.10195945E-05	5.09056351E-06	3.26206700E-02	2.87879600E-02	2.72493730E-02	2.16956709E-02
curr		4.42794683E+05	1.92288397E-05	1.88132740E+00	1.10197800E-05	5.09050810E-06	3.26221974E-02	2.87921678E-02	2.72556635E-02	2.17113329E-02

Figure 82.

Since the molar masses are known for the two components, only the loading concentrations, y–offsets and global equilibrium constant were fit.

2.4.1.15.6 Global fit with <u>un-linked cells</u>

Now not linking the cells (i.e. not linking the cell loading concentrations) gives a better fit. This because the cell loading concentration depends quite strongly on having estimated the correct position of base of the cell. In this fit we are still holding the molar ratio of the loading concentrations of B/A.

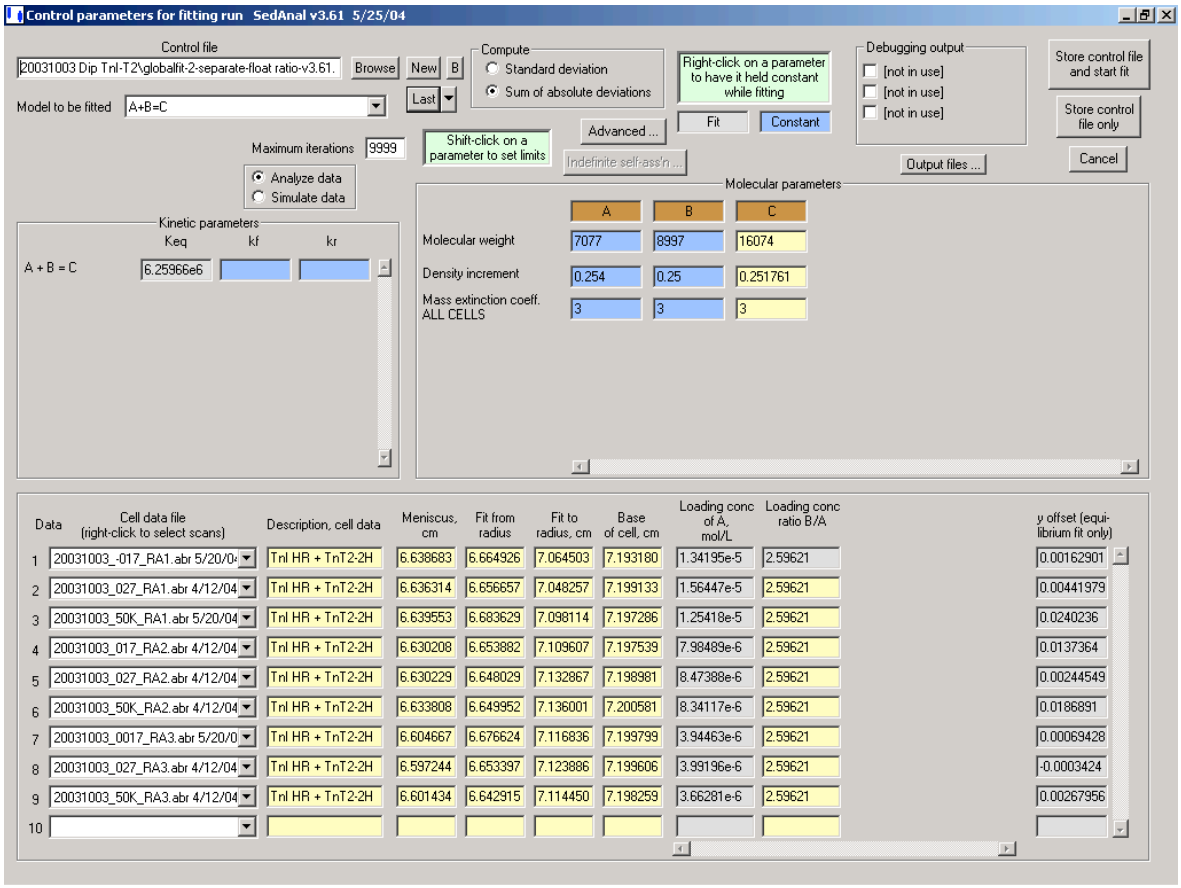
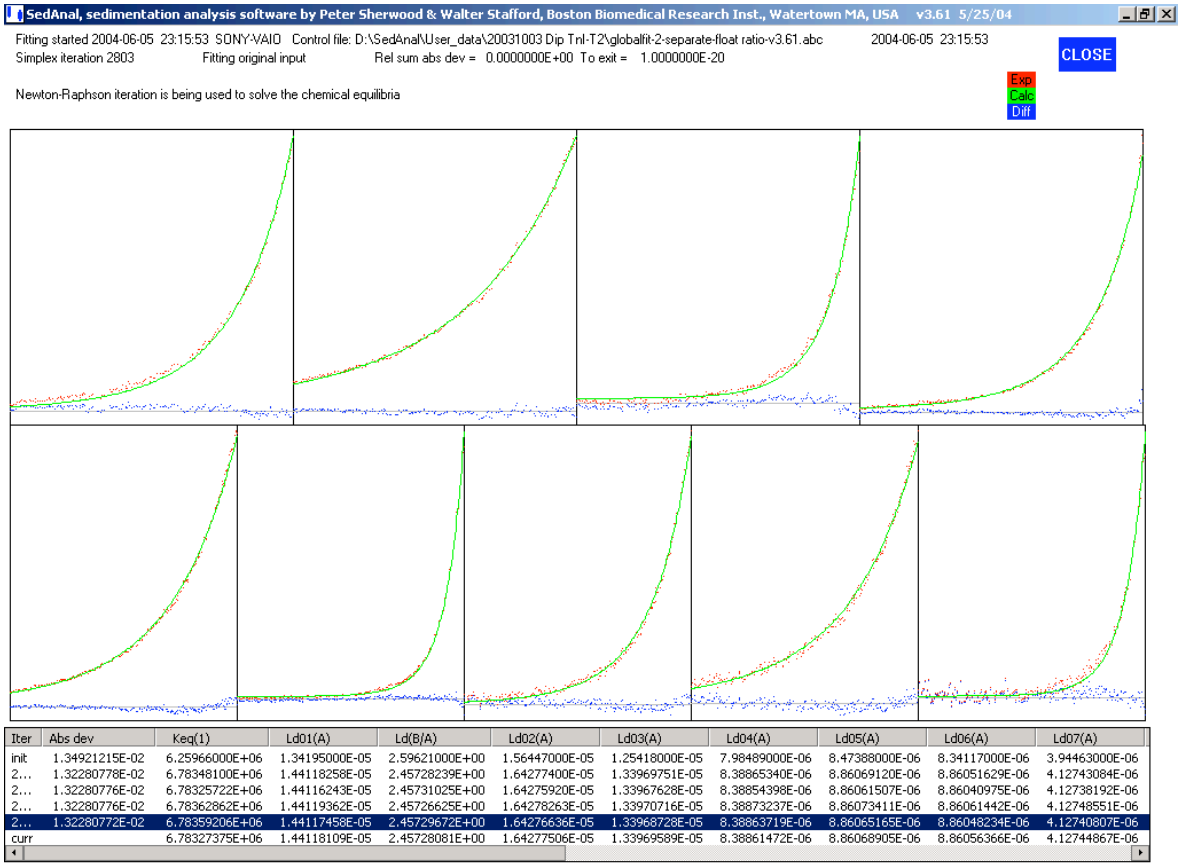


Figure 83. The main control screen for the fitting preprocessed data for sedimentation equilibrium run.



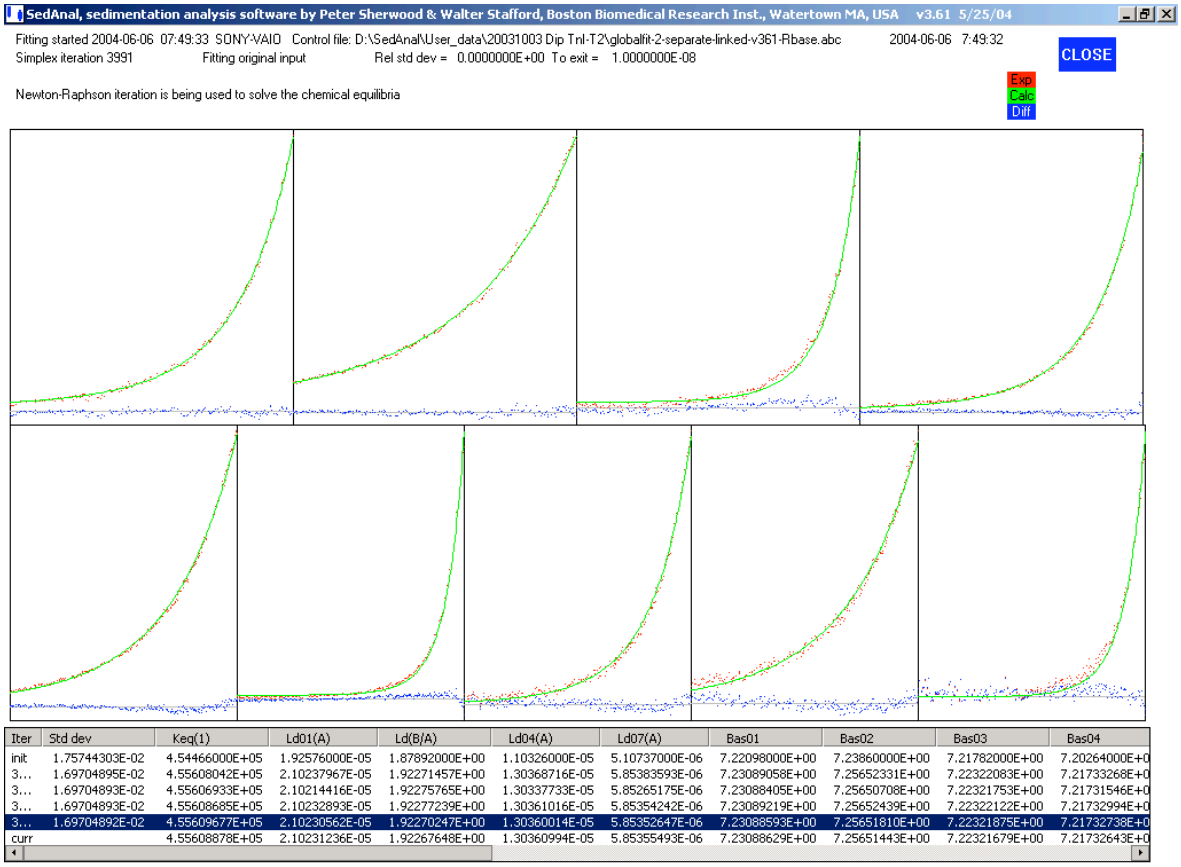
Iter	Abs dev	Keq(1)	Ld01(A)	Ld(B/A)	Ld02(A)	Ld03(A)	Ld04(A)	Ld05(A)	Ld06(A)	Ld07(A)
init	1.34921215E-02	6.25966000E+06	1.34195000E-05	2.59621000E+00	1.56447000E-05	1.25418000E-05	7.98489000E-06	8.47388000E-06	8.34117000E-06	3.94463000E-06
2...	1.32280778E-02	6.78348100E+06	1.44118258E-05	2.45728239E+00	1.64277400E-05	1.33969751E-05	8.38865340E-06	8.86069120E-06	8.86051629E-06	4.12743084E-06
2...	1.32280776E-02	6.78325722E+06	1.44116243E-05	2.45731025E+00	1.64275920E-05	1.33967628E-05	8.38854398E-06	8.86061507E-06	8.86040975E-06	4.12738192E-06
2...	1.32280776E-02	6.78362862E+06	1.44119362E-05	2.45726625E+00	1.64278263E-05	1.33970716E-05	8.38873237E-06	8.86073411E-06	8.86061442E-06	4.12748551E-06
2...	1.32280772E-02	6.78359206E+06	1.44117458E-05	2.45729672E+00	1.64276636E-05	1.33968728E-05	8.38863719E-06	8.86065165E-06	8.86048234E-06	4.12740807E-06
curr		6.78327375E+06	1.44118109E-05	2.45728081E+00	1.64277506E-05	1.33969589E-05	8.38861472E-06	8.86068905E-06	8.86056366E-06	4.12744867E-06

Figure 84.

Returning to the case of linked cells but allowing the radius of the base of each cell to be a fitting parameter: we get a better linked-cell fit than the first one but still not as good letting all the loading concentrations float.

	Data	Cell data file (right-click to select scans)	Description, cell data	Meniscus, cm	Fit from radius	Fit to radius, cm	Base of cell, cm	Loading conc of A, mol/L	Loading conc ratio B/A	y offset, fringes
1		20031003_-017_RA1.abr 5/20/04	TnI HR + TnT2-2H	6.469880	6.664926	7.064500	7.220980	1.92578e-5	1.87892	0.03
2		20031003_027_RA1.abr 4/12/04	TnI HR + TnT2-2H	6.441330	6.656657	7.048267	7.238600	1.92578e-5	1.87892	0.02
3		20031003_50K_RA1.abr 5/20/04	TnI HR + TnT2-2H	6.324220	6.683629	7.098134	7.217820	1.92578e-5	1.87892	0.02
4		20031003_017_RA2.abr 4/12/04	TnI HR + TnT2-2H	6.630210	6.653882	7.109607	7.202640	1.10326e-5	1.87892	0.02
5		20031003_027_RA2.abr 4/12/04	TnI HR + TnT2-2H	6.630230	6.648029	7.132657	7.202210	1.10326e-5	1.87892	-0.0
6		20031003_50K_RA2.abr 4/12/04	TnI HR + TnT2-2H	6.633810	6.649952	7.136891	7.198950	1.10326e-5	1.87892	0.02
7		20031003_0017_RA3.abr 5/20/0	TnI HR + TnT2-2H	6.604670	6.676624	7.116886	7.200400	5.10737e-6	1.87892	7.64
8		20031003_027_RA3.abr 4/12/04	TnI HR + TnT2-2H	6.597240	6.653397	7.123886	7.196070	5.10737e-6	1.87892	-0.0
9		20031003_50K_RA3.abr 4/12/04	TnI HR + TnT2-2H	6.601430	6.642915	7.114650	7.198580	5.10737e-6	1.87892	-0.0
10										

Figure 85.



Iter	Std dev	Keq(1)	Ld01(A)	Ld(B/A)	Ld04(A)	Ld07(A)	Bas01	Bas02	Bas03	Bas04
init	1.75744303E-02	4.54466000E+05	1.92576000E-05	1.87892000E+00	1.10326000E-05	5.10737000E-06	7.22098000E+00	7.23860000E+00	7.21782000E+00	7.20264000E+0
3...	1.69704895E-02	4.55608042E+05	2.10237967E-05	1.92271457E+00	1.30368716E-05	5.85383593E-06	7.23089058E+00	7.25652331E+00	7.22322083E+00	7.21733268E+0
3...	1.69704893E-02	4.55606933E+05	2.10214416E-05	1.92275765E+00	1.30337733E-05	5.85265175E-06	7.23088405E+00	7.25650708E+00	7.22321753E+00	7.21731546E+0
3...	1.69704893E-02	4.55608685E+05	2.10232893E-05	1.92277239E+00	1.30361016E-05	5.85354242E-06	7.23089219E+00	7.25652439E+00	7.22322122E+00	7.21732994E+0
3...	1.69704892E-02	4.55609677E+05	2.10230562E-05	1.92270247E+00	1.30360014E-05	5.85352647E-06	7.23088593E+00	7.25651810E+00	7.22321875E+00	7.21732738E+0
curr		4.55608878E+05	2.10231236E-05	1.92267648E+00	1.30360994E-05	5.85355493E-06	7.23088629E+00	7.25651443E+00	7.22321679E+00	7.21732643E+0

Figure 86.

Since a better fit was obtained without linking the cells, it may be that some material had pelleted or gelled at the bottom at the higher speeds.

Allowing separate ratios of B/A for each cell gives an even better fit in the unlinked case.

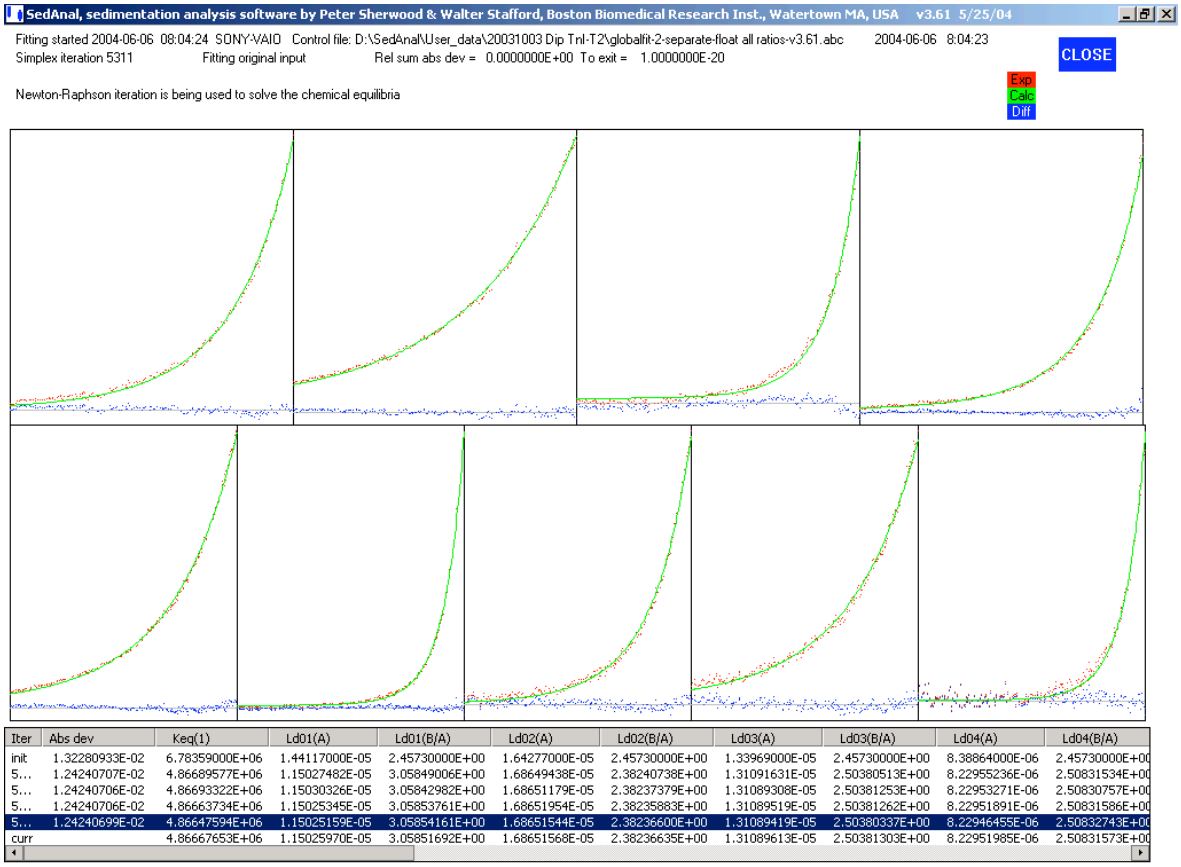


Figure 87.

Note that although this fit is better, the equilibrium constant found in this case is about 1 order of magnitude higher than that value found with the more constrained fits. Why?

2.4.1.16 Error Analysis Boot-strap with Replacement, Monte Carlo and F-statistics

2.4.1.16.1 Error Estimation: Parameter standard deviations.

2.4.1.16.1.1 Bootstrap with replacement

SEDANAL will carry out a specified number of so called "**boot-strap with replacement**" or fitting operations to estimate the standard deviations of the estimates of the fitted parameter values. Essentially the way it works is that the original data, consisting of N data points, are randomly sampled, with replacement, N times to produce a new data set that is then fit using the best fit parameters values as the starting guesses. Each fit is carried out until convergence is reached and the new fitted parameters are written into a table. After a specified number of boot-strap fits, the standard deviation of each parameter is computed. In the case of SEDANAL, for which the fitting times can be very long, it is practical to carry out only a limited number of boot-strap fits. Ideally one would prefer to repeat the boot-strap operation 500 to 1000 times and generate a distribution function for each parameter from which true confidence limits can be computed. However, it is feasible to carry out only a limited number of boot-strap fits and so an abbreviated boot-strap procedure has been implemented to compute only the standard deviation of each parameter. The ability to compute a standard deviation from a smaller number of samples of a potentially non-normal distribution relies on the Central Limit Theorem which basically states that the distribution of repeated estimates of the parameters will be normally distributed no matter what the shape of the actual distribution of the parameters. This will be an approximation that will give one a reasonably good idea of the degree to which any given parameter is determined by the data being fitted.

Figure 88. To specify the number of boot-strap operations to be performed, one enters the number in the boot-strap tab under the **Advanced** button of the Control Screen.

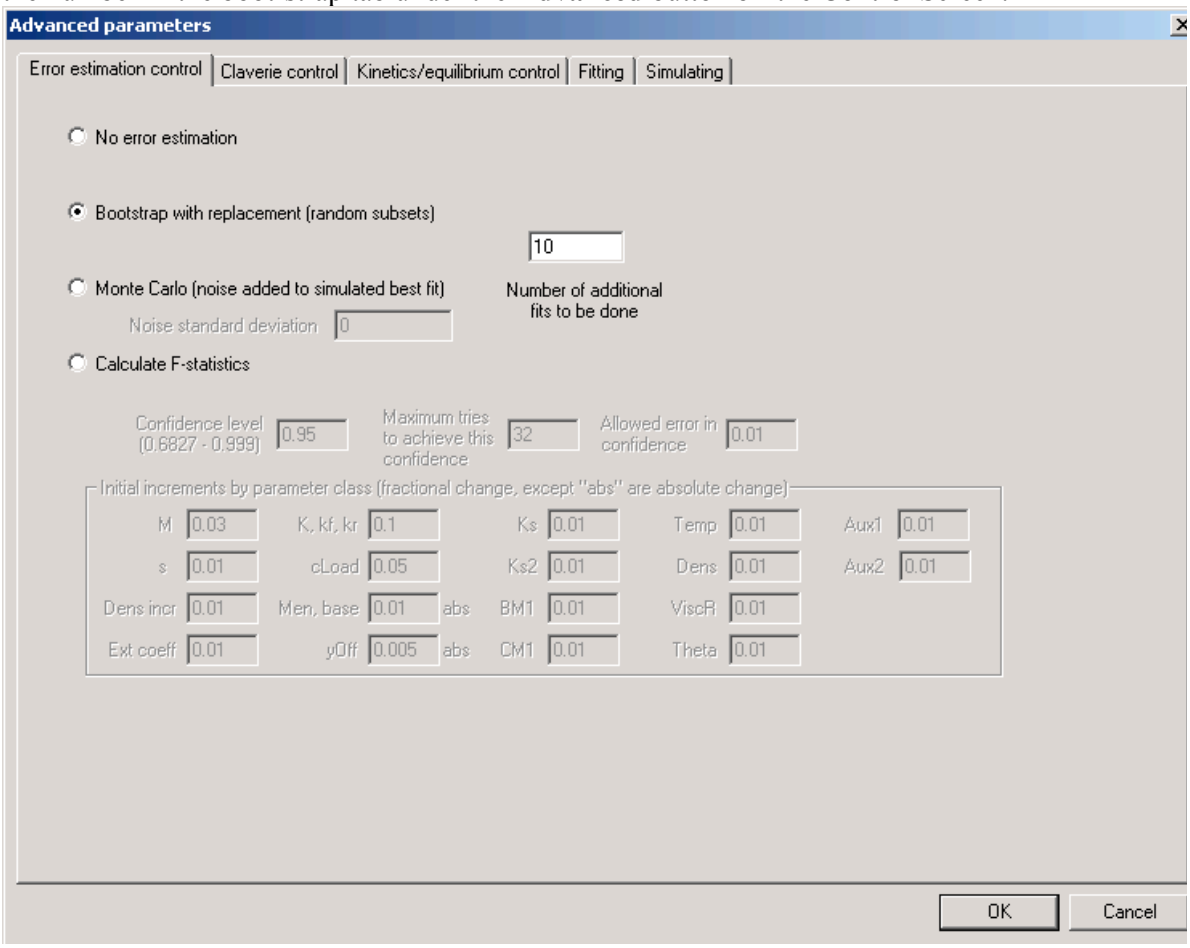


Figure 101. Error estimation control

2.4.1.16.1.2 Monte Carlo Analysis

The **Monte Carlo analysis** works similarly except that the best fit model is simulated first and then multiple different noise sets are added and the simulated data are refit each time.

For a Monte Carlo analysis, the Monte Carlo button is clicked the number of simulations is entered and the standard deviation of the normally distributed random noise that will be added to the data is entered:

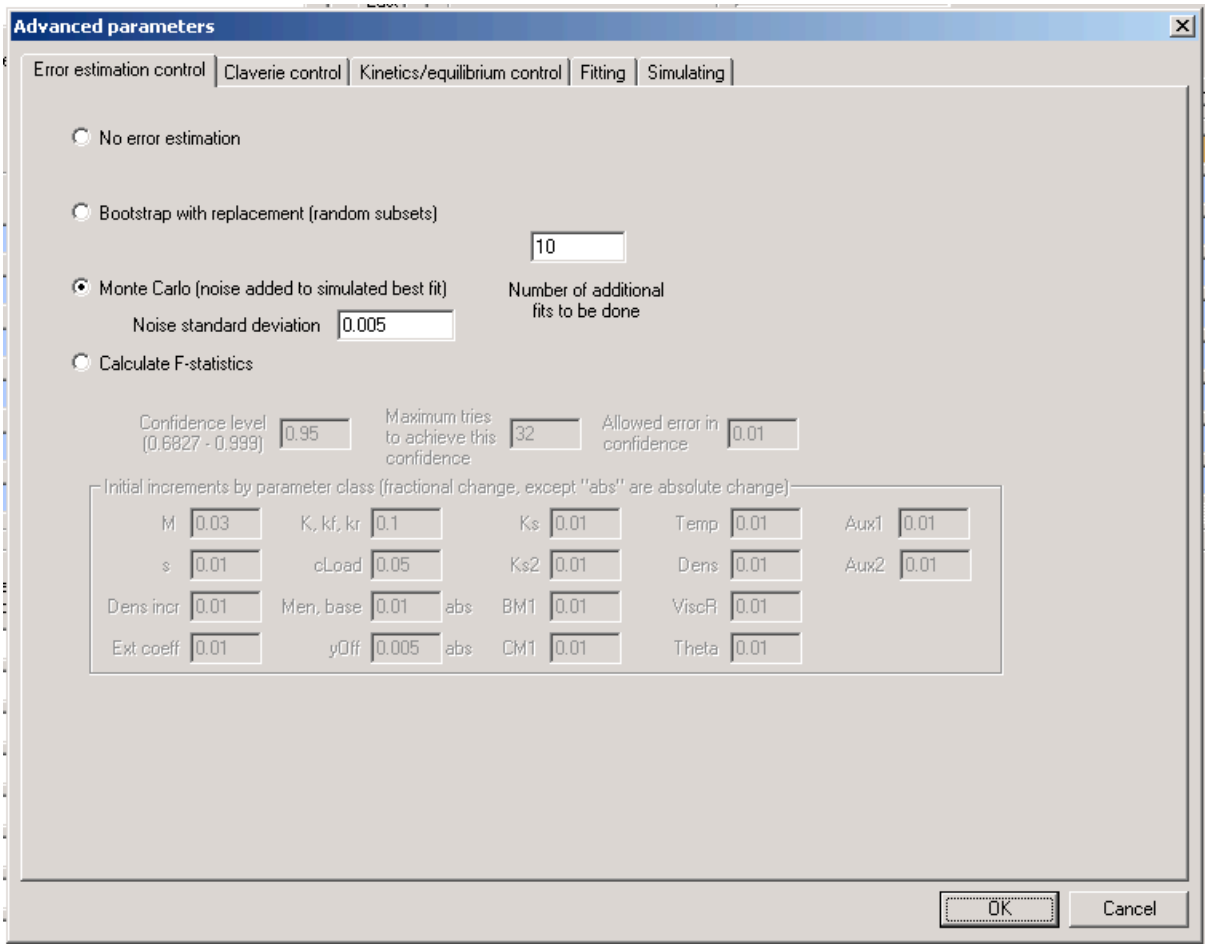


Figure 89. Error estimation - Monte Carlo

N.B.: Monte Carlo analysis is also useful when designing experiments. One can simulate particular cases using the values of the expected parameters to see the effect of noise on the ability to obtain reliable values of the fitted parameters.

The convergence level of each fit, including each boot-strap fit, can be set under the “Simplex control” tab under the Advanced button.

2.4.1.16.2 Error Estimation: Parameter Confidence Limits

2.4.1.16.2.1 F-statistics.

2.4.1.16.2.1.1 General F-stat preferences

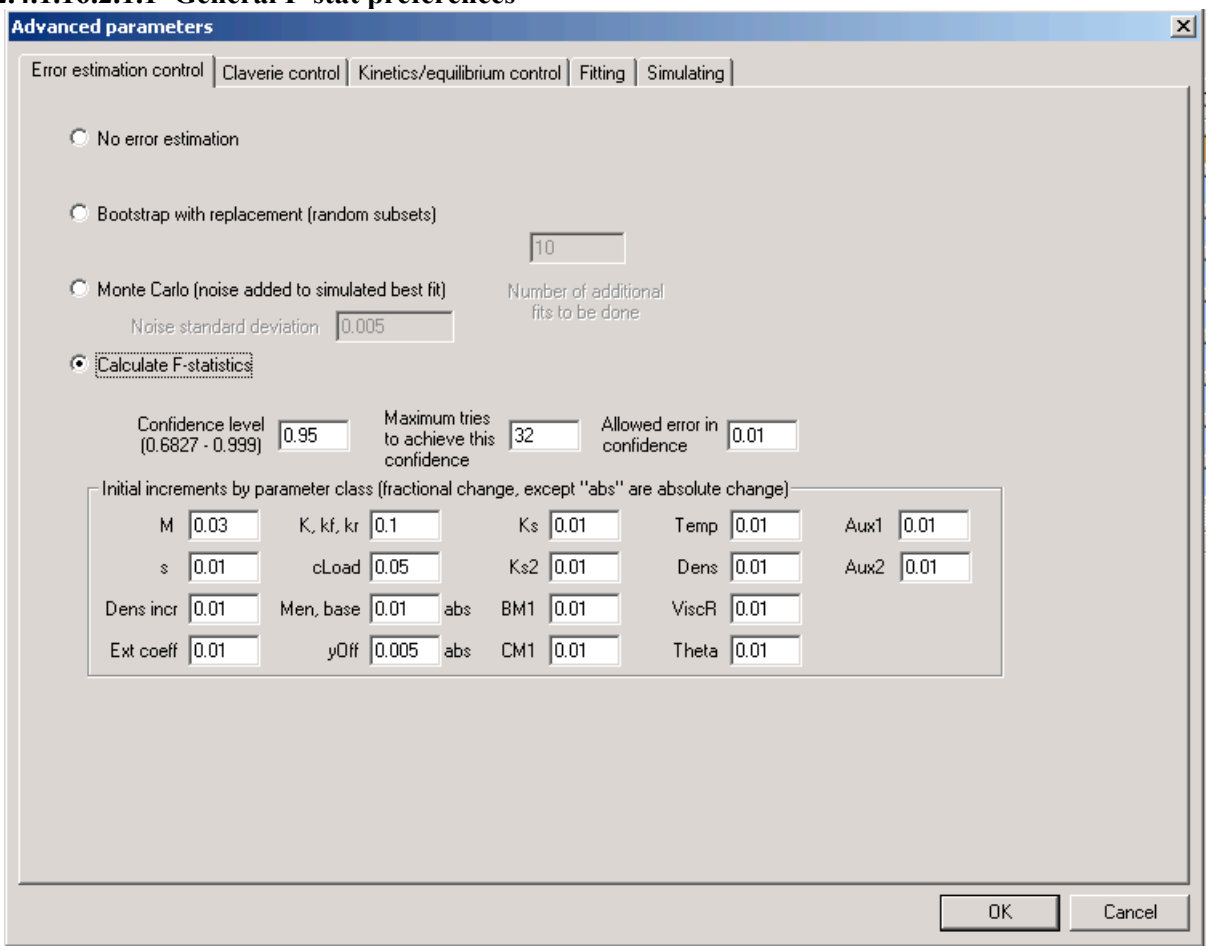


Figure 90. Error estimation: F-statistics.

To turn on the ability to compute F-Statistics for **any** parameter, under the advanced menu, error estimation tab, click the "Calculate F-statistics" button. Indicate the desired default confidence level, the maximum number of steps to take in each direction and the tolerance, i.e. allowed error in the confidence level, to stop the search of parameter space. In the boxes along the bottom of this window, you can tailor the size of the search steps to be taken for each type of variable.

2.4.1.16.2.1.2 Turning on F-stats for a particular parameter

To indicate that F-statistics are to be calculated for a **particular** parameter, left-shift-click on the parameter window on the control screen. A window will appear; check the F-statistics box and enter the desired confidence level for that parameter.

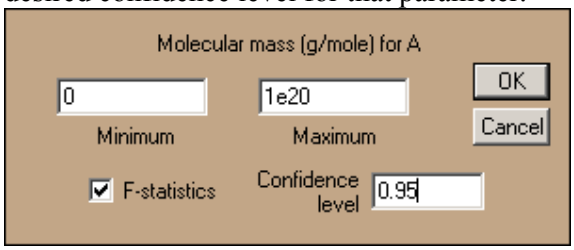


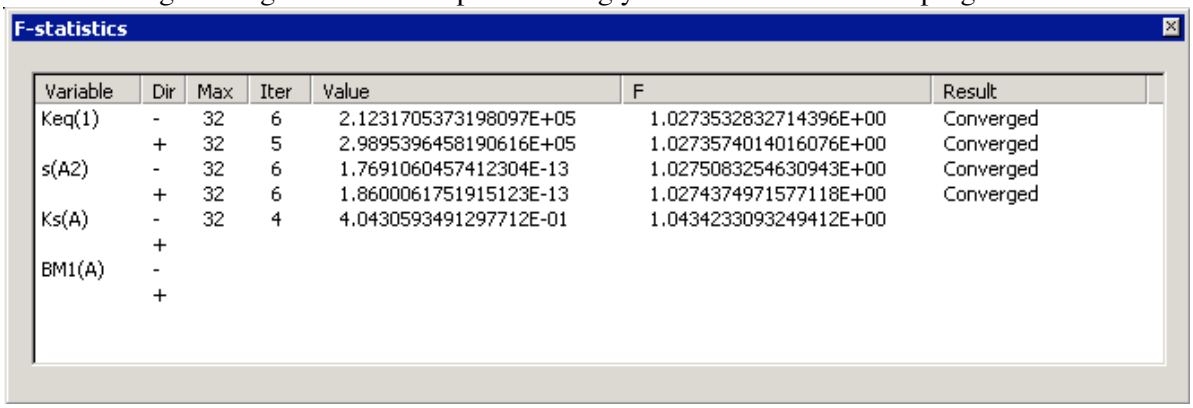
Figure 91. Turning on F-stats and choosing the CL for an individual parameter

Confidence intervals can be determined by computing F-statistics in the following way (cf. M.L. Johnston and M. Straume, Meth. Enz. 1994). After the main fitting has finished, confidence intervals of any desired magnitude can be computed for any of the parameters by shift-left-clicking on the parameter box on the control screen, checking the "F-statistics" box and inputting the confidence level (CL), e.g. 0.95. F-Statistics are computed by stepping each parameter, holding it constant and refitting the data set. The new rmsd is used to compute F = [rmsd(new fit)/rmsd(minimum)]^2. When F increases to the value corresponding to the CL and number of degrees of freedom, that value of the parameter is the 95% confidence level for the parameter. The parameter is varied in both directions to yield the (usually) asymmetric confidence intervals. When a fit has finished, SEDANAL will compute the confidence intervals at the level indicated (95% in this case) for those variables for which F-Stats have been enabled. F-Stats must also be enabled in the Error Estimation control tab under the "Advanced ..." button on the control screen or under the "Control Extended tab" on the preferences screen that can be accessed from the Main Menu.

2.4.1.16.2.1.3 Displaying F-stat progress

To display the F-stat progress, click on the "F" button on the upper right corner of the fitting window:

The following floating window will open allowing you to monitor the f-stat progress.



Variable	Dir	Max	Iter	Value	F	Result
Keq(1)	-	32	6	2.1231705373198097E+05	1.0273532832714396E+00	Converged
	+	32	5	2.9895396458190616E+05	1.0273574014016076E+00	Converged
s(A2)	-	32	6	1.7691060457412304E-13	1.0275083254630943E+00	Converged
	+	32	6	1.8600061751915123E-13	1.0274374971577118E+00	Converged
Ks(A)	-	32	4	4.0430593491297712E-01	1.0434233093249412E+00	
	+					
BM1(A)	-					
	+					

Figure 91b. F-statistics monitoring window.

2.5 DCDT and WDA

Time derivative analysis, g(s*), and multi-speed wide distribution (WDA) analysis can be carried out by selecting the “dc/dt” button on the main menu figure 17). The “Concentration profile time-derivative analysis” window will then appear.

2.5.1 DCDT

A run file is loaded by clicking in the “Experiment” window to specify which experiment is to be analyzed followed by clicking in the “Cell data” window to select a particular cell data (.abr) file. When the cell data file is selected, all the data in the “abr” file will be presented. Usually the run will span a large time interval and the data plotted may look rather strange. The scans to be analyzed are selected in the same

way they are in the Fitting control screen, by right clicking on the Cell data file window. The numbers of scans to be used should be collected consistent with the molar mass of the species present. If too many scans are included, the g(s*) peaks will be artificially broadened. After the desired scans have been selected, the plots should look similar to the following screen dump (Fig. 92).

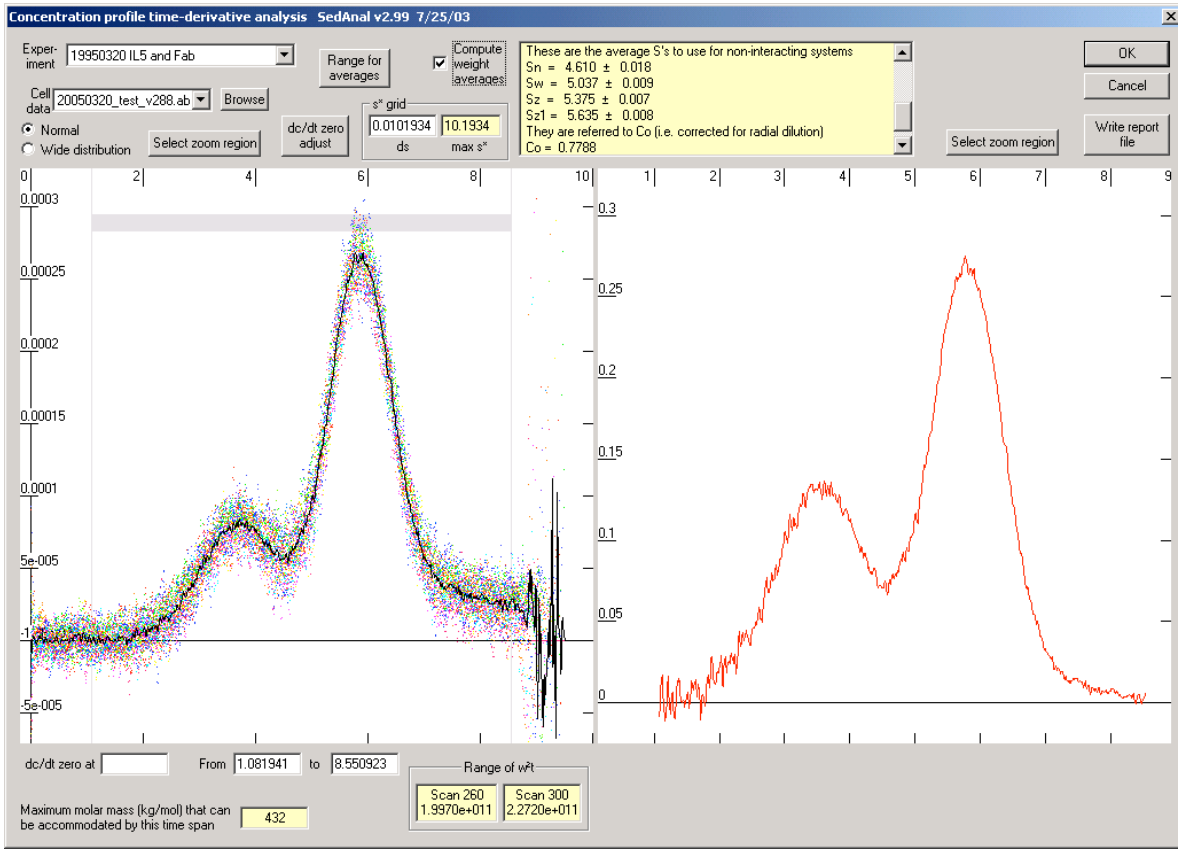


Figure 92.

The left window contains a plot of "dc/dt vs s*" for every scan pair used in the analysis. A range of data can be selected in the left window to remove undesirable stray data points that might interfere with the analysis. The range chosen at this point will be used for the final "g(s*) vs s*" plot displayed in the right window, as well as in the computation of the various averages of the sedimentation coefficient, which will be displayed in the top window after the "Compute weight averages" box is checked.

The s*-axis scaling is selected in the s*-grid window. There are a thousand s* points plotted across the window. For example, if a value of 0.01 is selected for the s*-grid spacing, the value of s* will span from 0-10S. After the value of s*-grid has been edited, you must right click the Cell data window and then click on the OK button to refresh the screen. Left clicking on the Cell data window will reload the entire data set.

The plots may be re-scaled by clicking on the "**Select zoom region**" button and using the mouse to select a rectangular region to be re-plotted. The previous zoom levels are stored in a stack. A previous zoom level can be regained by typing CTRL-Z and a later zoom level, by typing CTRL-Y .

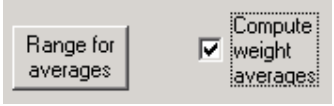


Figure 92a. The weight average sedimentation coefficient (among others) may be computed by selecting a range over which to do the average by first

clicking on the button labeled “Range for averages”, selecting a range on the dc/dt plot by clicking and dragging and then clicking on the button labeled “Compute weight averages”.

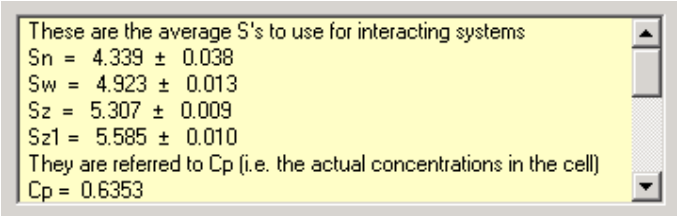


Figure 90. The corresponding averages appear highlighted in yellow in the adjacent box.

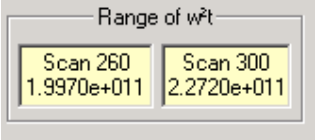


Figure 92b. The temporal range of the data is displayed in a window at the bottom of the screen.

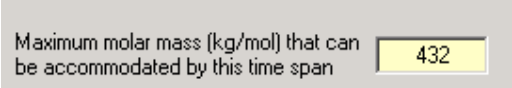
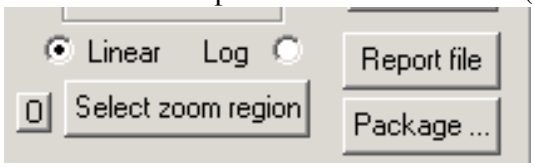


Figure 92c. Also displayed is the maximum molar mass of a macromolecule whose diffusion coefficient could be reliably determined from the data set (i.e. the set of scans) chosen by fitting the g(s*) vs. s*pattern to a gaussian.

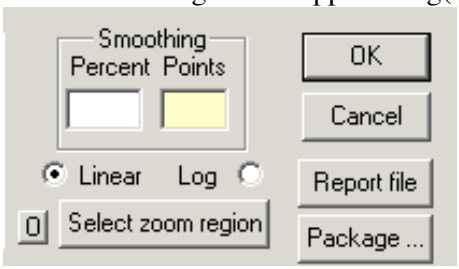
2.5.1.1 Switching between g(s*) vs s* and s* •g(s*) vs ln(s*)

It is possible to change the y and x-axis scales. On the s* •g(s*) vs ln(s*) scale, the vertical off-set is a constant and independent of the value of ln(s*).

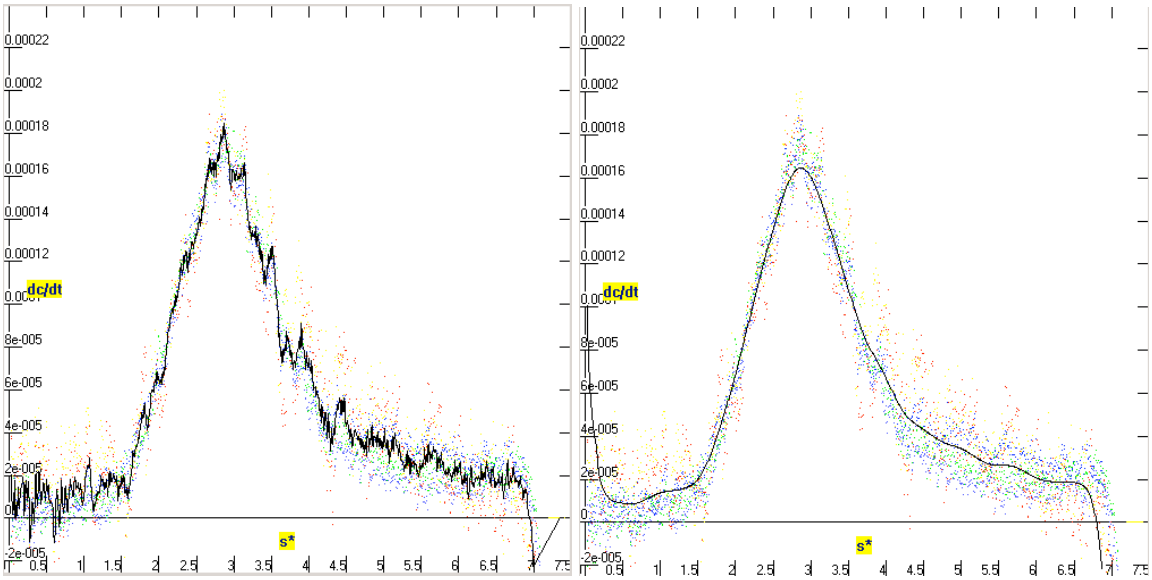


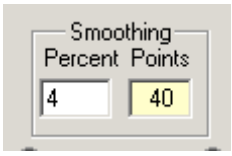
2.5.1.2 Smoothing of g(s*) vs s* for presntation purposes

Some smoothing can be applied to g(s*) curves.



With no smoothing a plot might look noisy; like:





Adding 4% (i.e. using 4% of the total data span in the sliding fit) smoothing gives the plot on the right.

2.5.2 Wide Distribution Analysis (WDA)

Data from a multi-speed experiment must be processed as described above. Before loading the run file (.abr file), click on the "Wide distribution" button (Fig. 93).

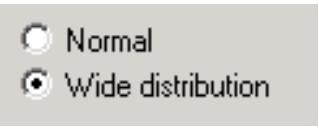


Figure 93. Dialog box

The following window (Fig. 94) will appear.

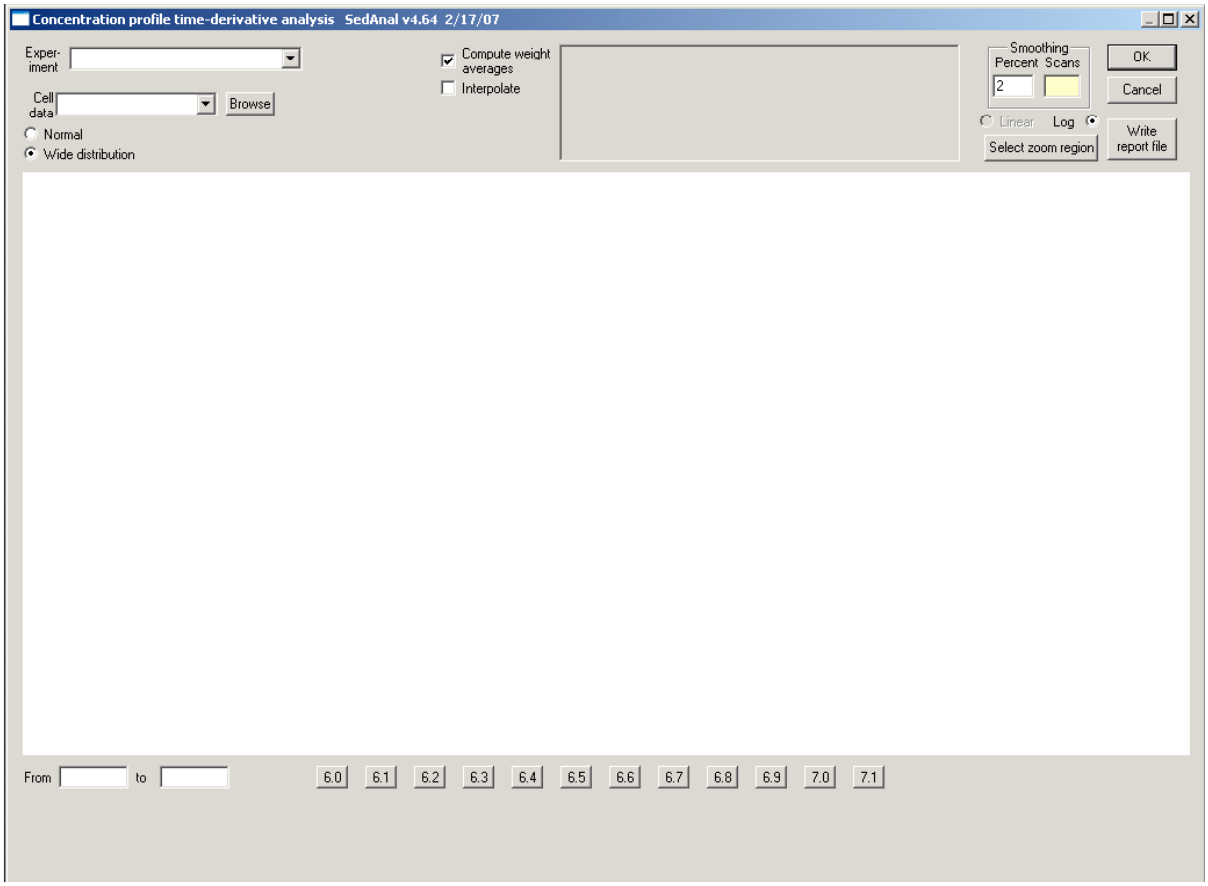


Figure 94. The control box for performing wide-distribution analysis.

After choosing a run file, and selecting a smoothing range (see below), a plot will appear from data taken at 6.5 cm (Fig. 95).

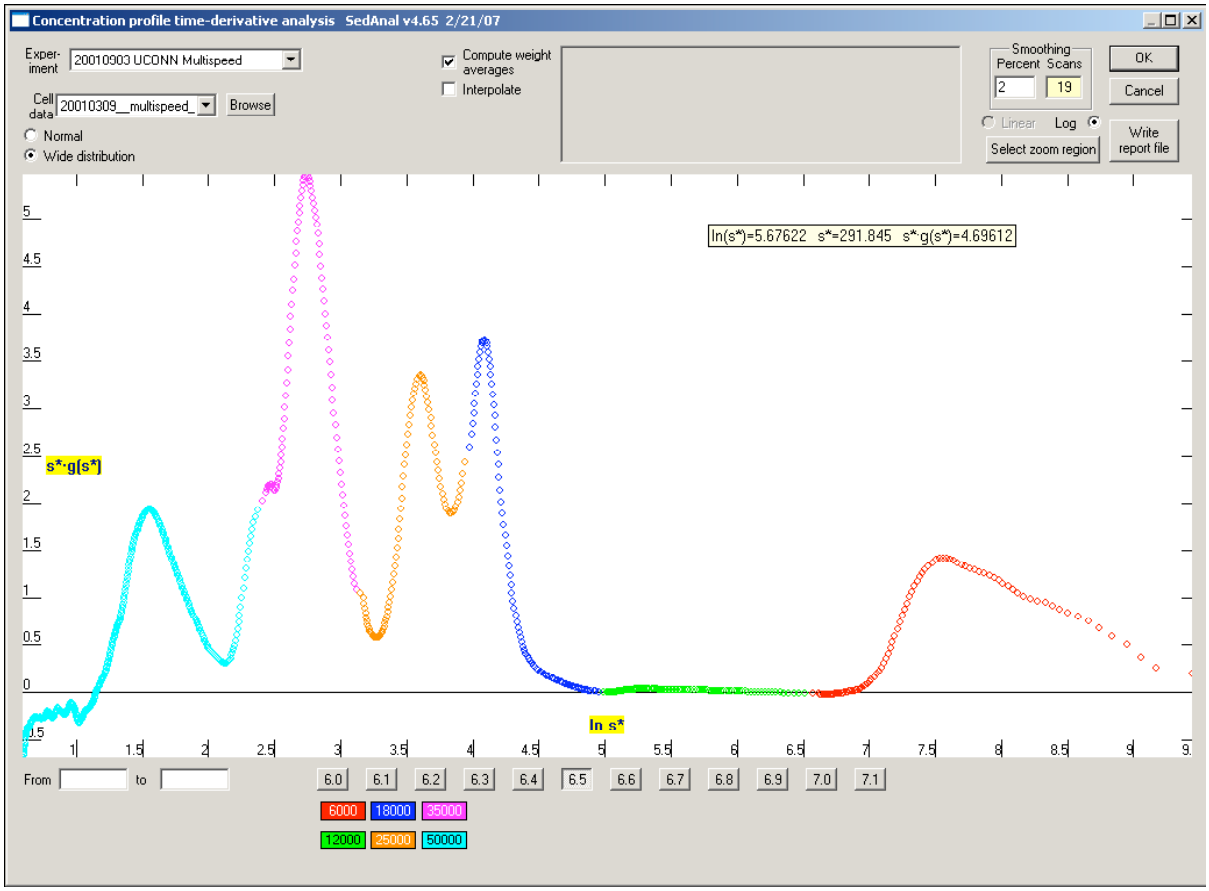


Figure 95. Results of WDA analysis.

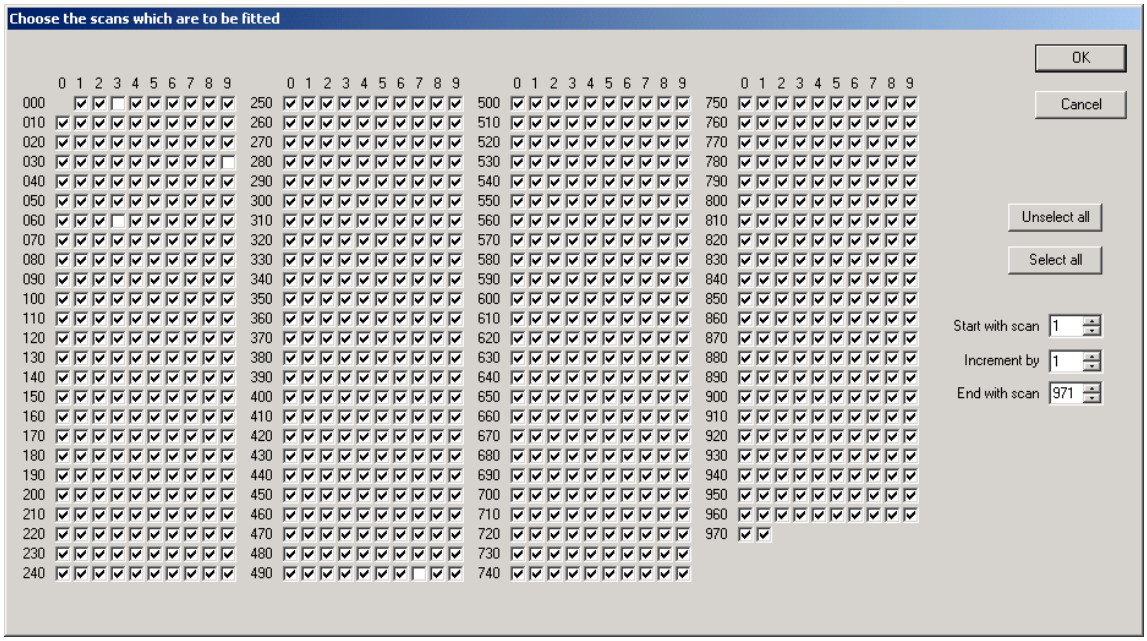


Figure 96. Control box for picking bad scans in WDA.

2.5.2.1 Bad Scans

Click on the numbers of any bad scans (Figure 96) (There are 4 in this example: #'s 3, 39, 63, & 497) to eliminate them from the analysis. Bad scans can be identified in the preprocessor.

2.5.2.2 Smoothing after numerical differentiation

Some smoothing must be preformed on the data since the derivative is computed by taking simple first differences of the c(r,t) vs s*(r,t) data. Smoothing is performed with three passes of a moving box-car filter using a window (i.e. number of points) expressed as a percentage of the total span of the data. This algorithm was chosen because it has very small leakage of high frequency components common to many other smoothing algorithms and it's frequency response can be easily controlled by varying the size of the window (Stafford, 1994).

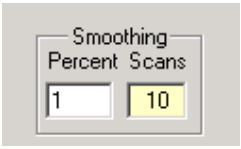


Figure 97. Control dialog box for choosing the size of the smoothing window is expressed as a percentage of the entire span of the data.
Warning: after entering the percent (i.e. "1" in this case) press enter to have it take effect.

The example given above (Figs. 95 and 98) is for a mixture of hemacyanin and polystyrene beads. The data are plotted as a differential distribution function, dc/dln(s*) vs ln(s*), on a logarithmic scale of s*.

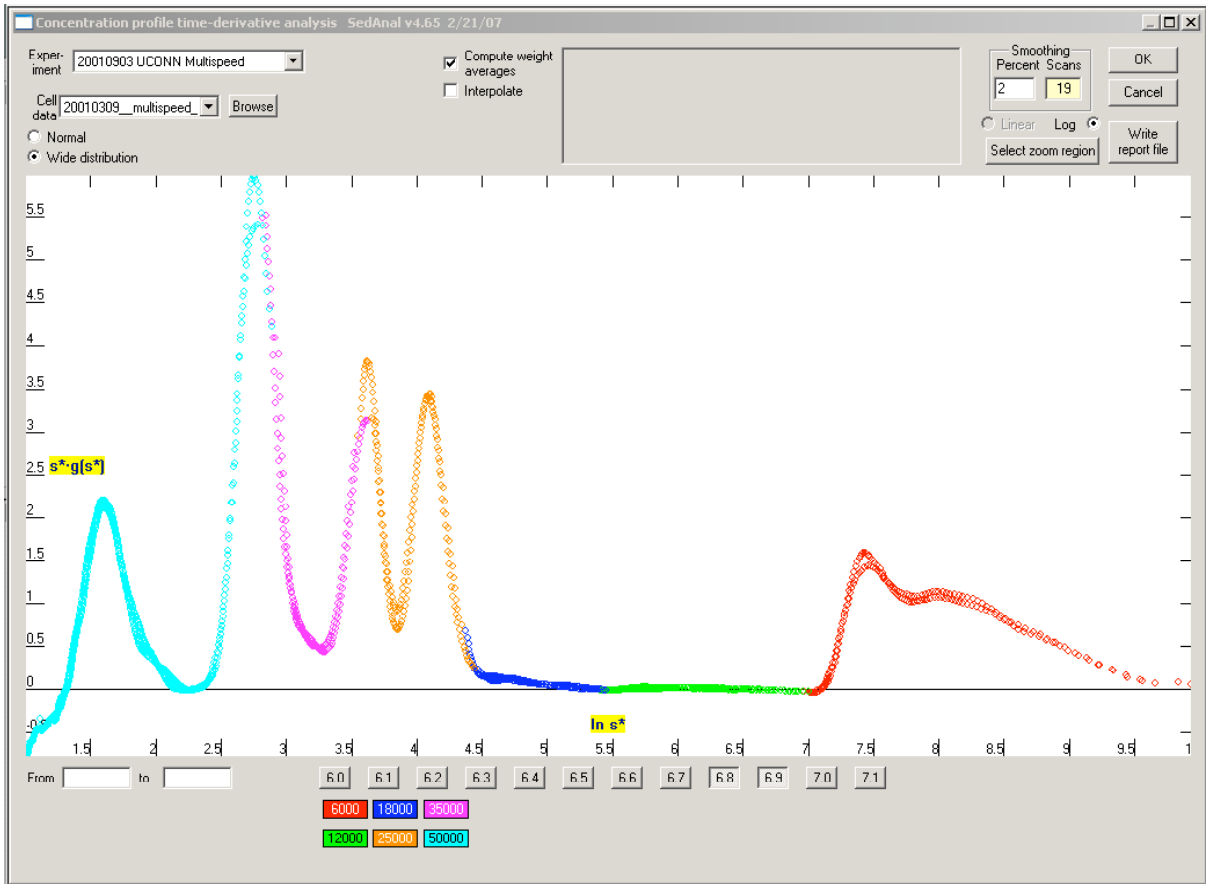


Figure 98. WDA for the mixture of hemacyanin with polystyrene beads

Here, plots from radii 6.8 and 6.9cm are superimposed.

Plots from higher radii will have higher resolution than those from a lower radius.

The range of these data is from 2.7 S to over 22,000 S

The color of the points in the plots corresponds to the speed at which the data were acquired. The color code is given below the plot.

2.6 Preferences

The Preferences screen allows one to choose the names of the directories used by SEDANAL, the default number of scans to show in the initial plot in the Preprocessor, and the maximum number of control files (*.abc) that are shown in the "Last" drop-down list on the Control Screen (Fig. 99). Also, various default and other parameters can be set for other SEDANAL functions like DCDT, BIOSPIN, the control screen, etc…

2.6.1 General Preferences

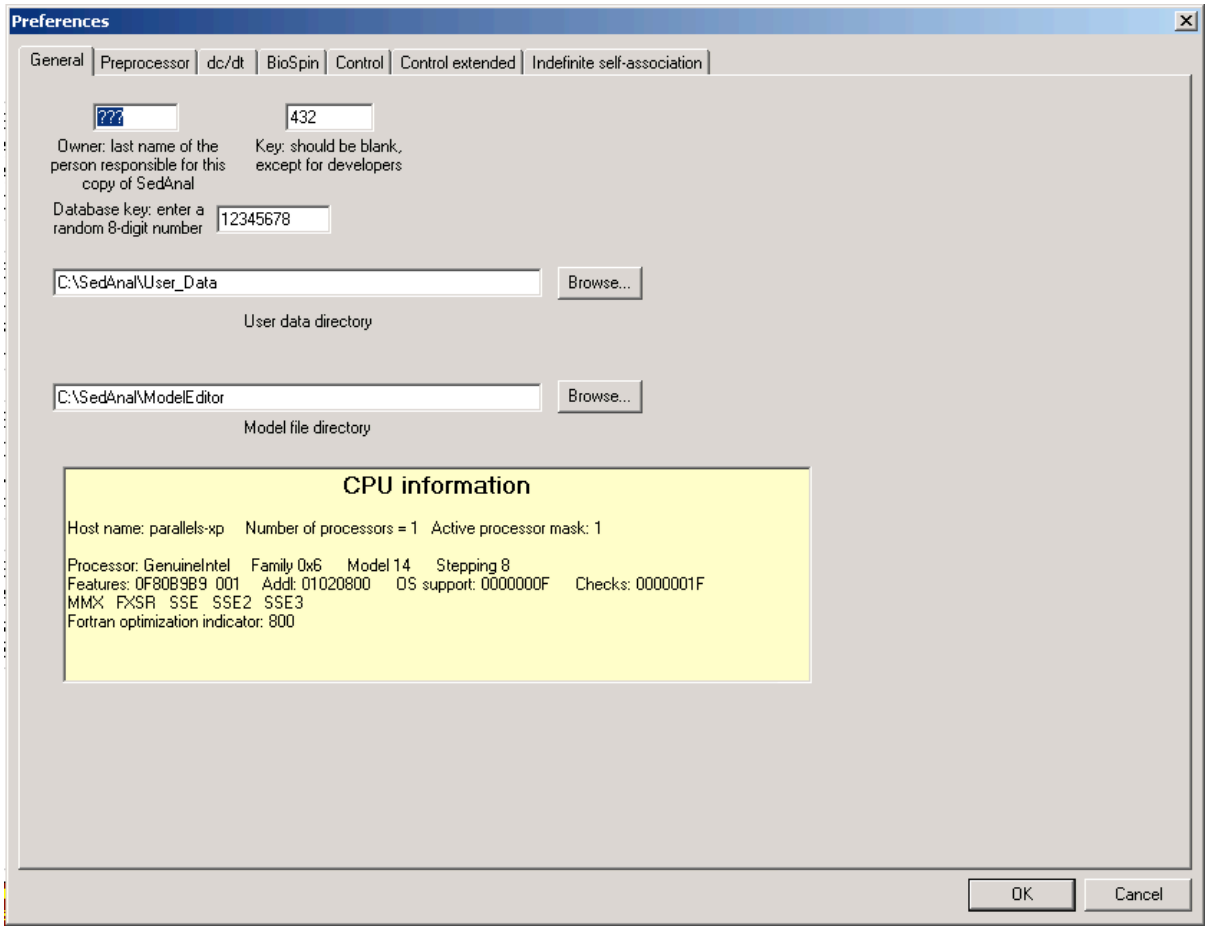


Figure 99A. Setting user data's and model editor files' paths.

2.6.2 Preprocessor preferences

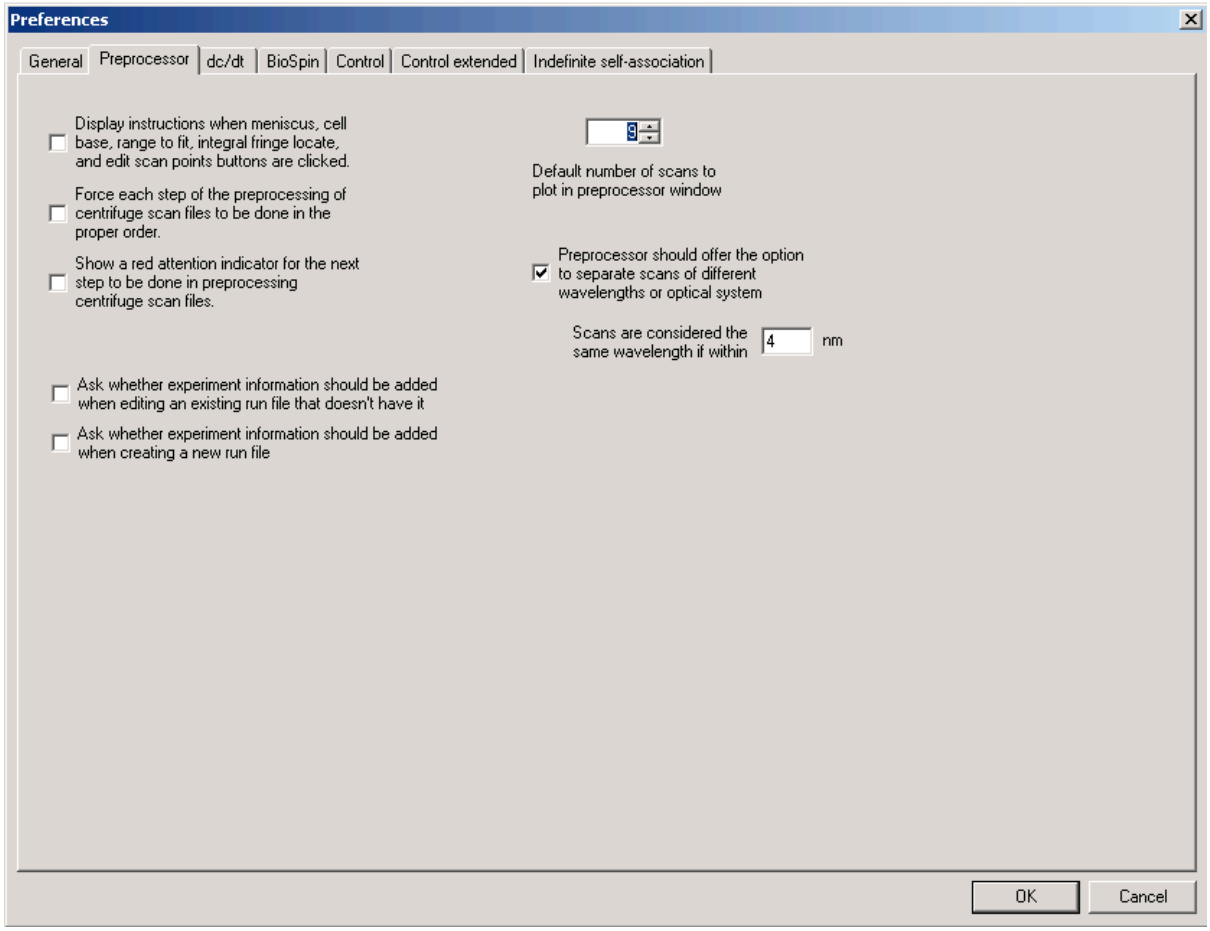


Figure 99B.

2.6.3 DCDT preferences

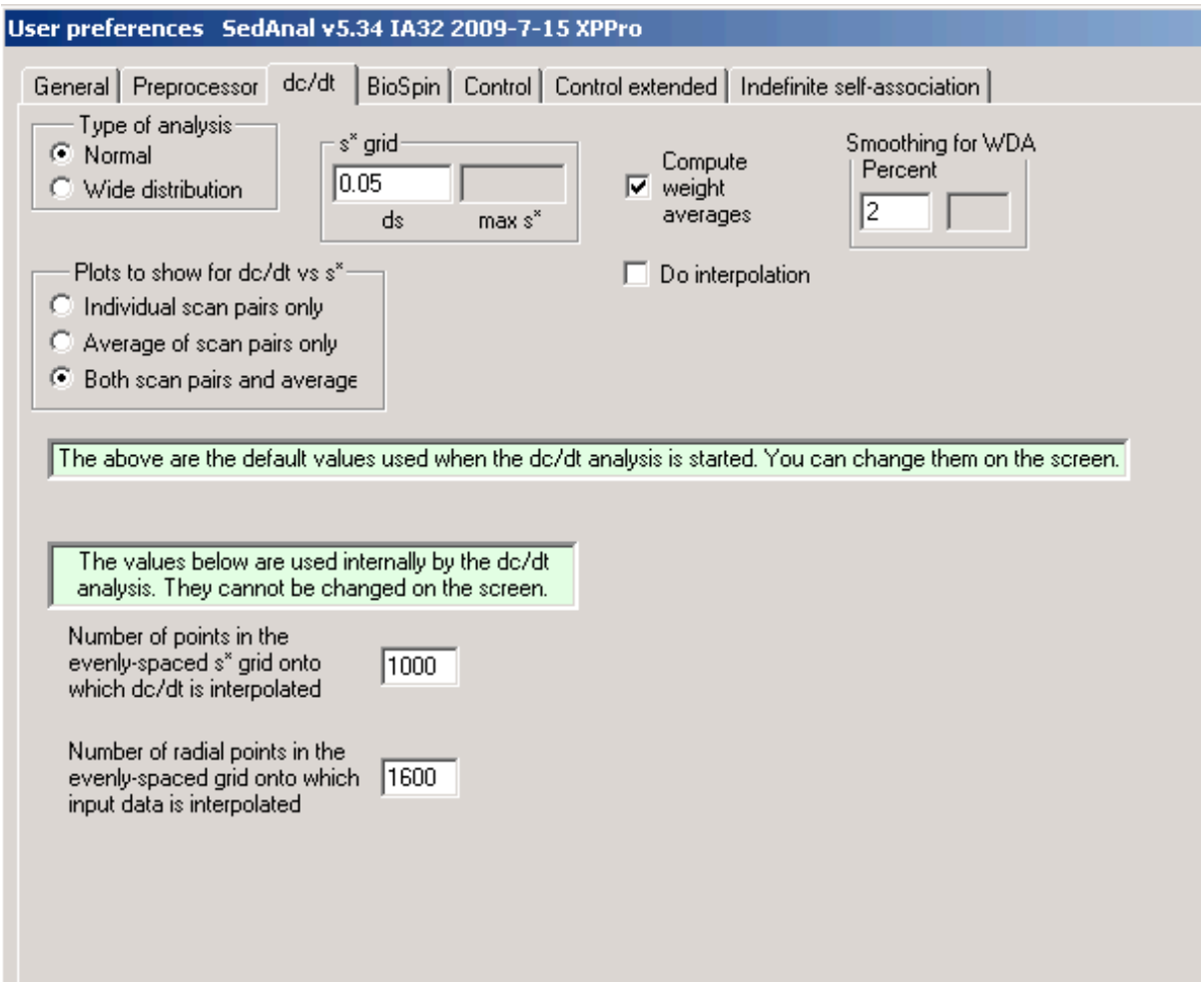


Figure 99C. Setting parameters for dc/dt. It is recommended that a default value of smoothing percent = 5 be used for the initial look at WDA data,

BIOSPIN preferences

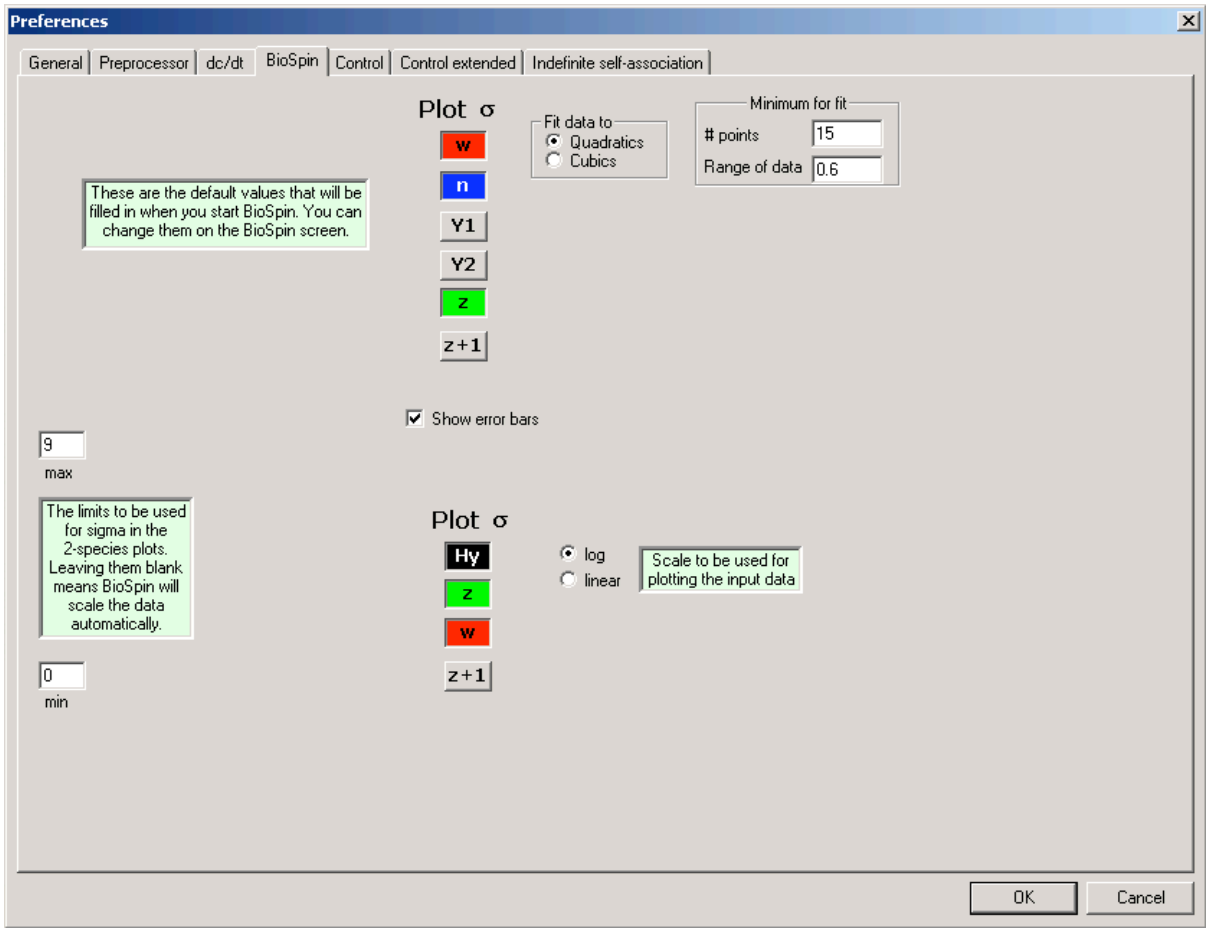


Figure 99D Preferred settings for BIOSPIN

2.6.4 Weighting Factor preferences - Control Tab

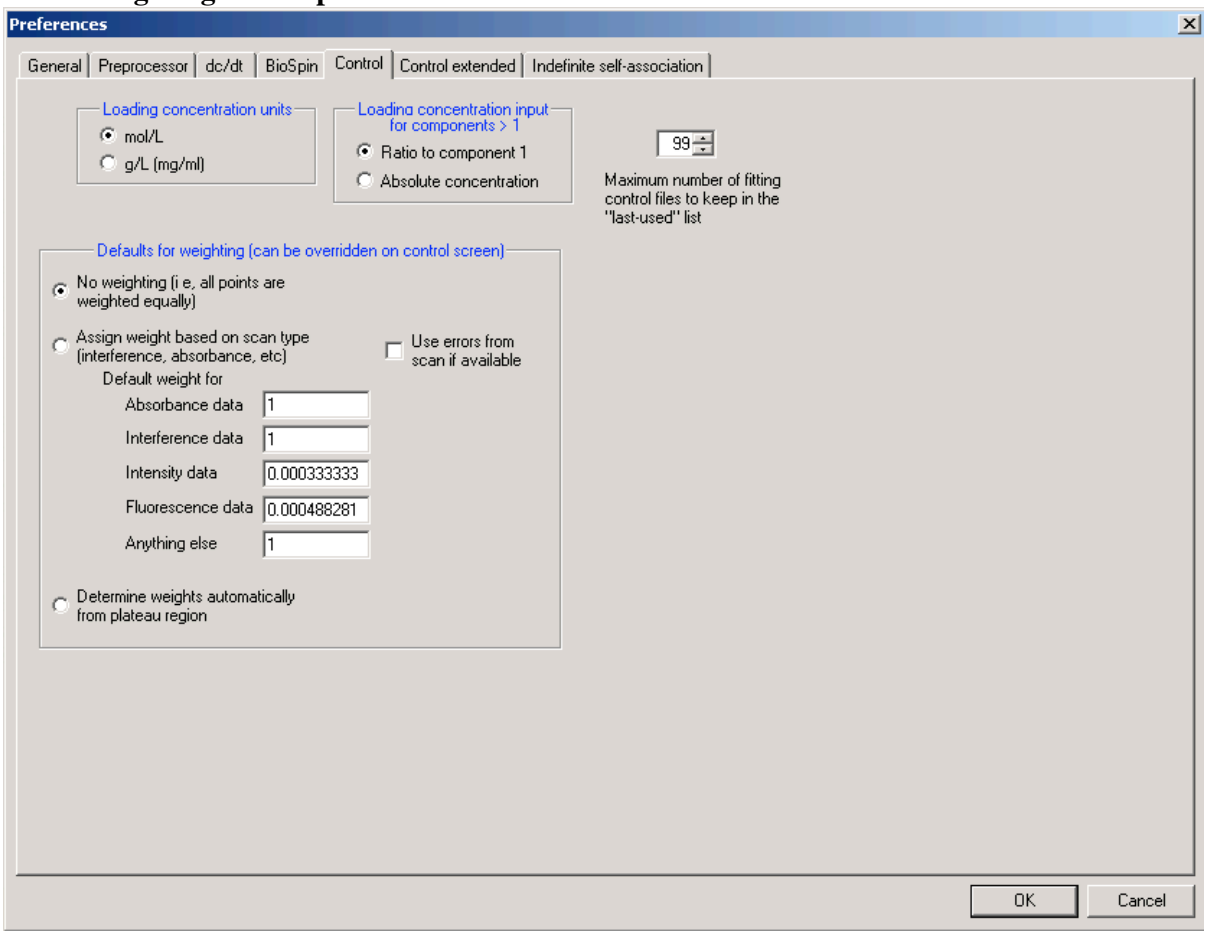


Figure 99E. Parameters for the control screen: setting defaults for use of weighting factors.

The use of weighting factors and user settable values for weighting factors can be set for each cell data file by right clicking on the cell data file name on the control screen:

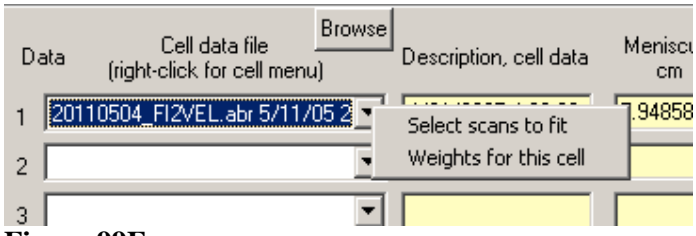


Figure 99F.

Selecting “weights for this cell” brings up the following window

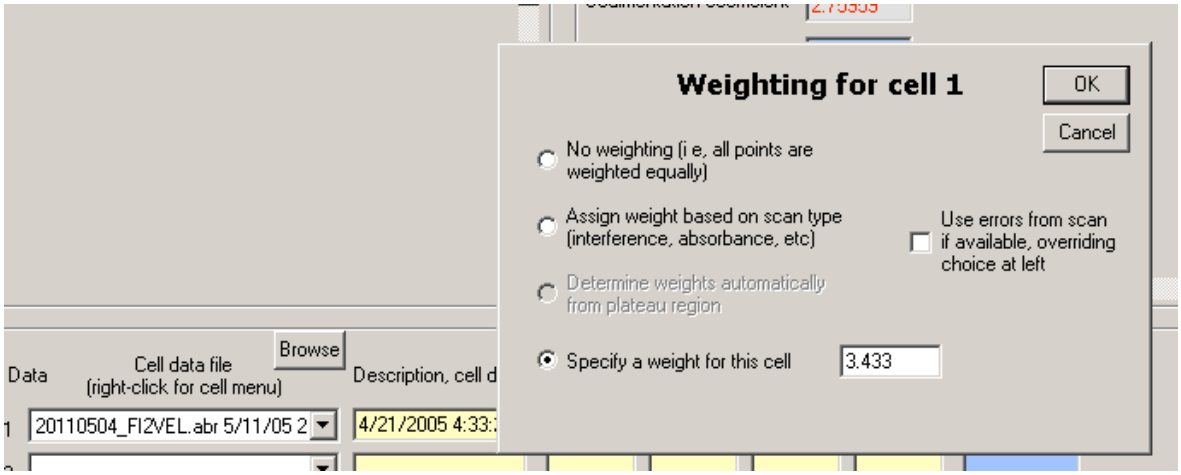


Figure 99G.

If data sets from different optical systems are to be combined in a global fit, they must be weighted according to the magnitude of the data and the noise on the data. If the inverse of the variance of the data is used as the weighting factor, then those data from different data sets can be combined. In the case shown here, using fluorescence data from the XL-F machine at the University of New Hampshire, the noise (i.e. standard deviation of the noise) on the data was +\- 0.5397 fluorescence units. The inverse square of that is 3.433 and so that value is used as a weighting factor for that data set. If these data had been combined with absorbance data whose noise is about +/- 0.006, a value of 27778 (=1.0/(0.006*0.006) would be entered for that data set. When weighting factors derived this way are used, the **reduced chi-squared** values instead of sums of squares are minimized.

NOTE:: **SEDANAL displays on the fitting screen the square root of the reduced chi-squared values instead of the rms deviation when inverse variances are used as weighting factors.**

2.6.5 Error estimation control: Advanced, Control Extended tab

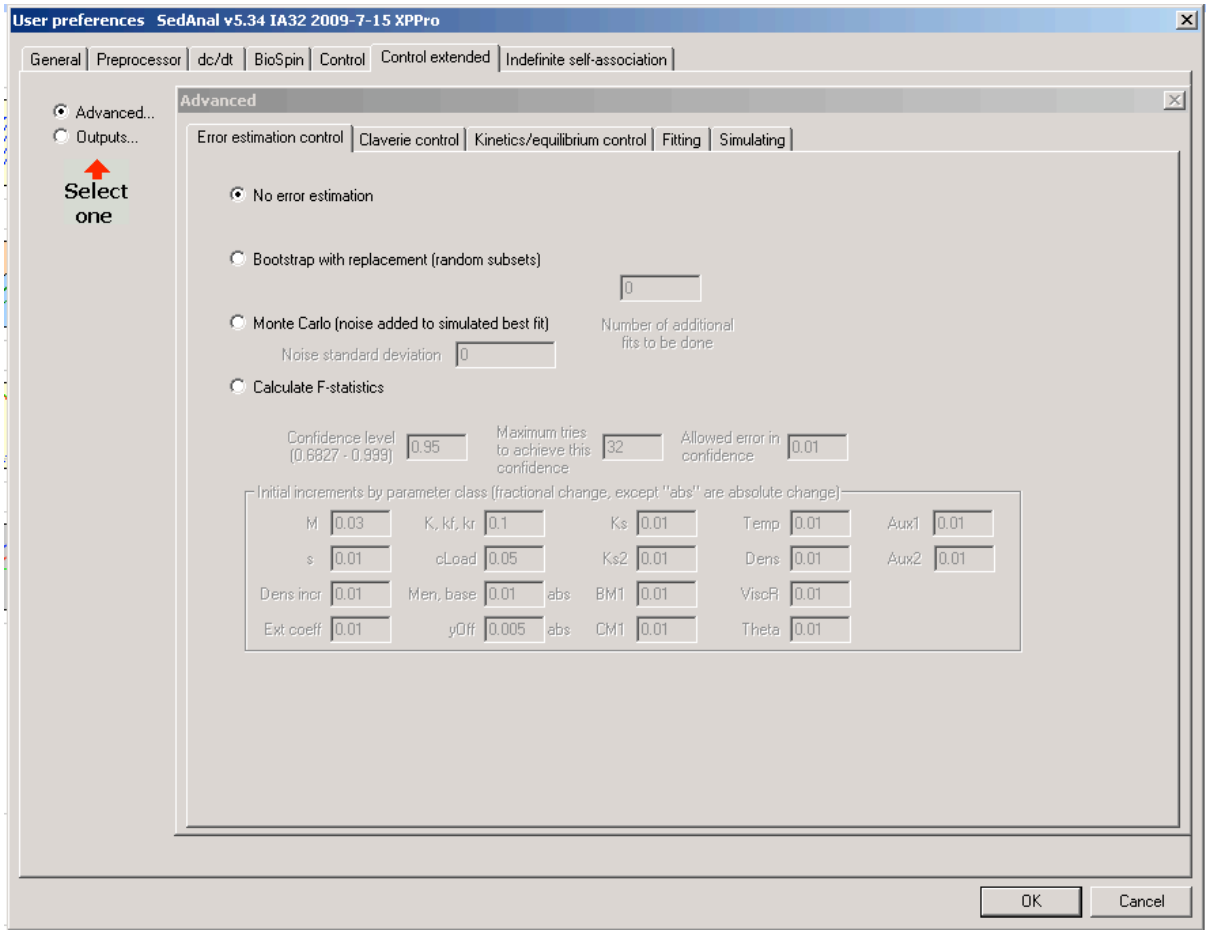


Figure 99H. "Control extended" parameters allow setting of the default parameters found under the "Advanced…" button on the control screen.

2.6.6 Reports: Output, Control Extended tab

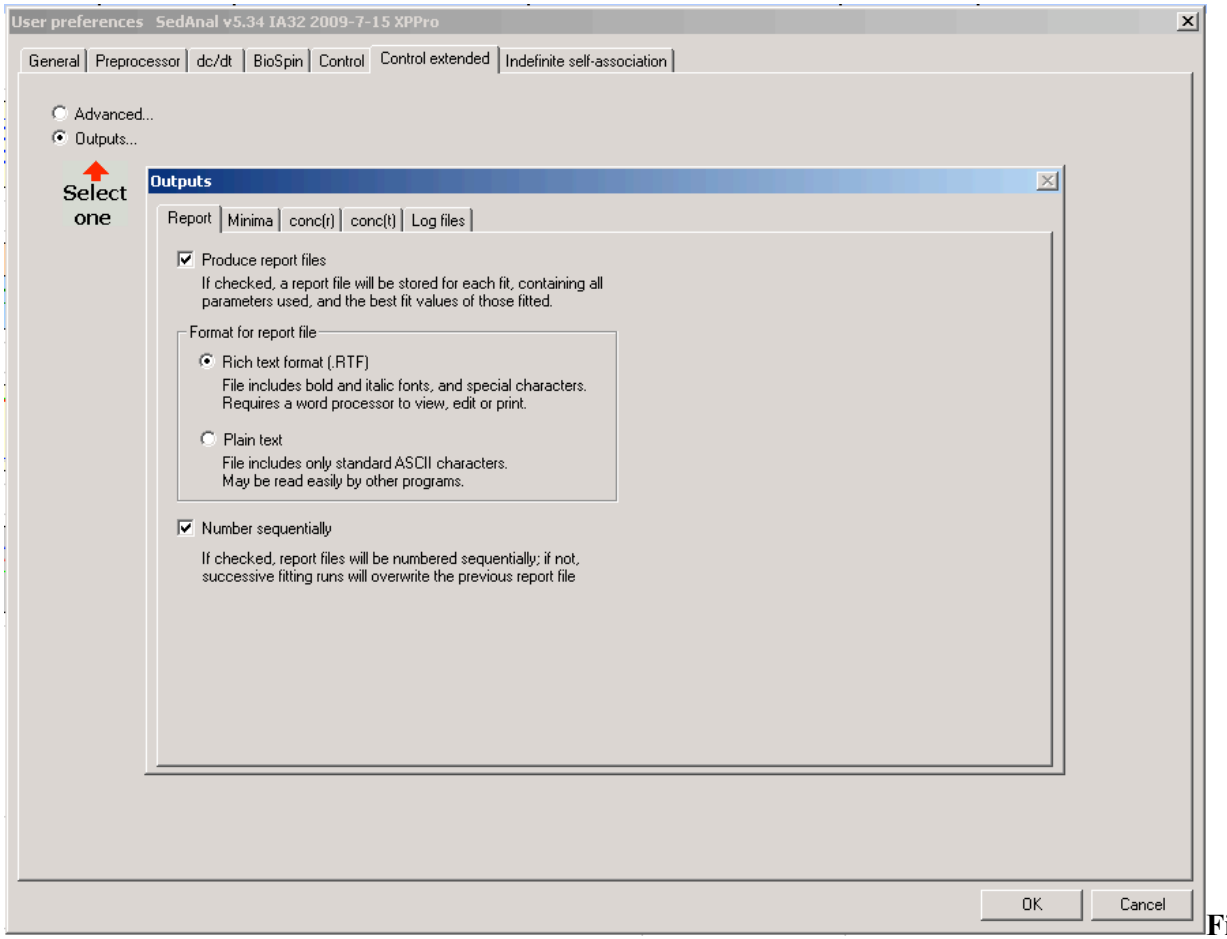


Figure 99I Control extended parameters also allow setting of the default parameters found under the "Outputs…" button on the control screen.

2.6.7 The other "Output …" tabs

2.6.7.1 Minima

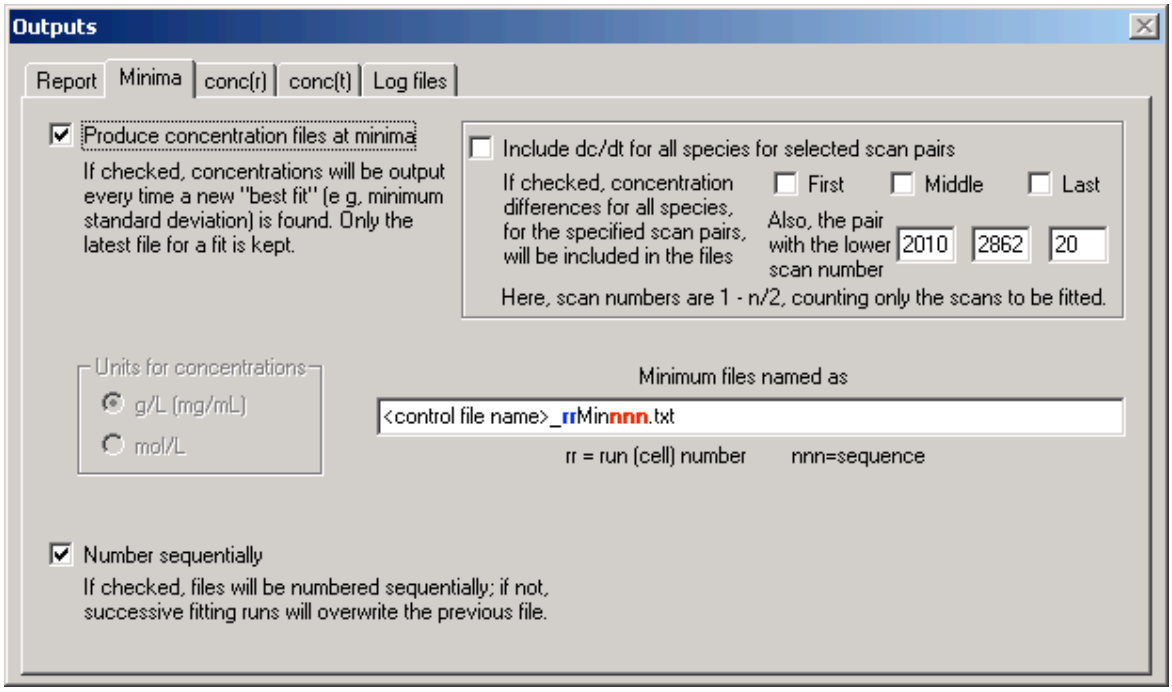


Figure 100.

2.6.7.2 Concentration as a function of radius

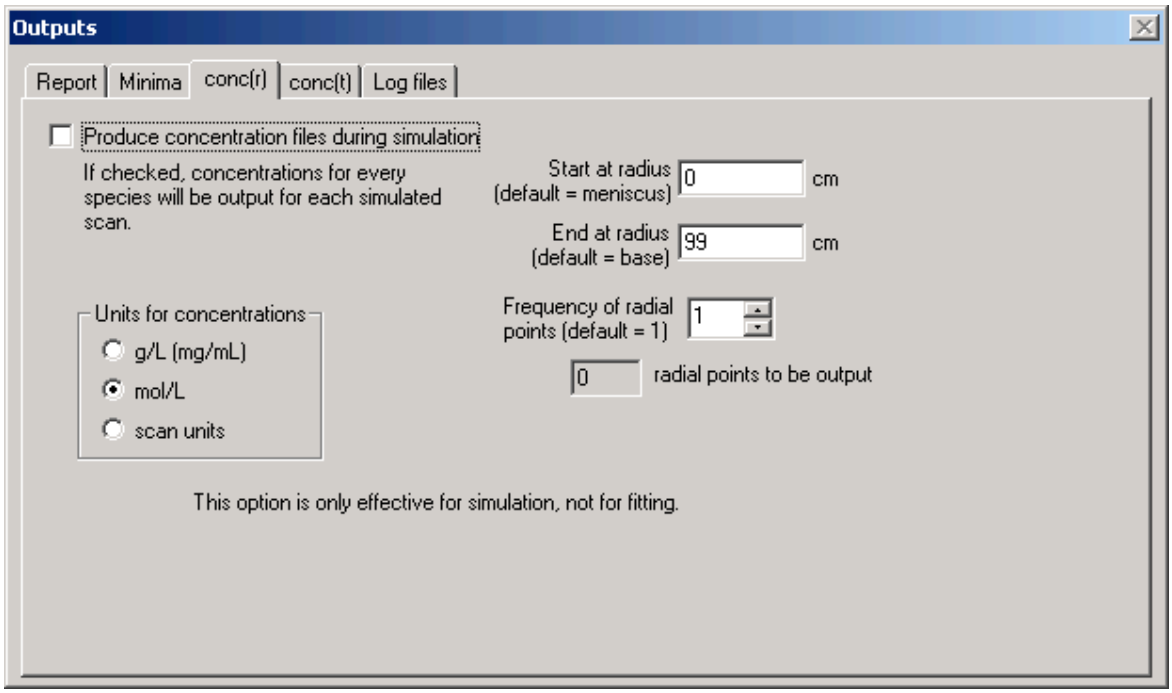


Figure 101.

2.6.7.3 Concentration as a function of time for the initial equilibration step

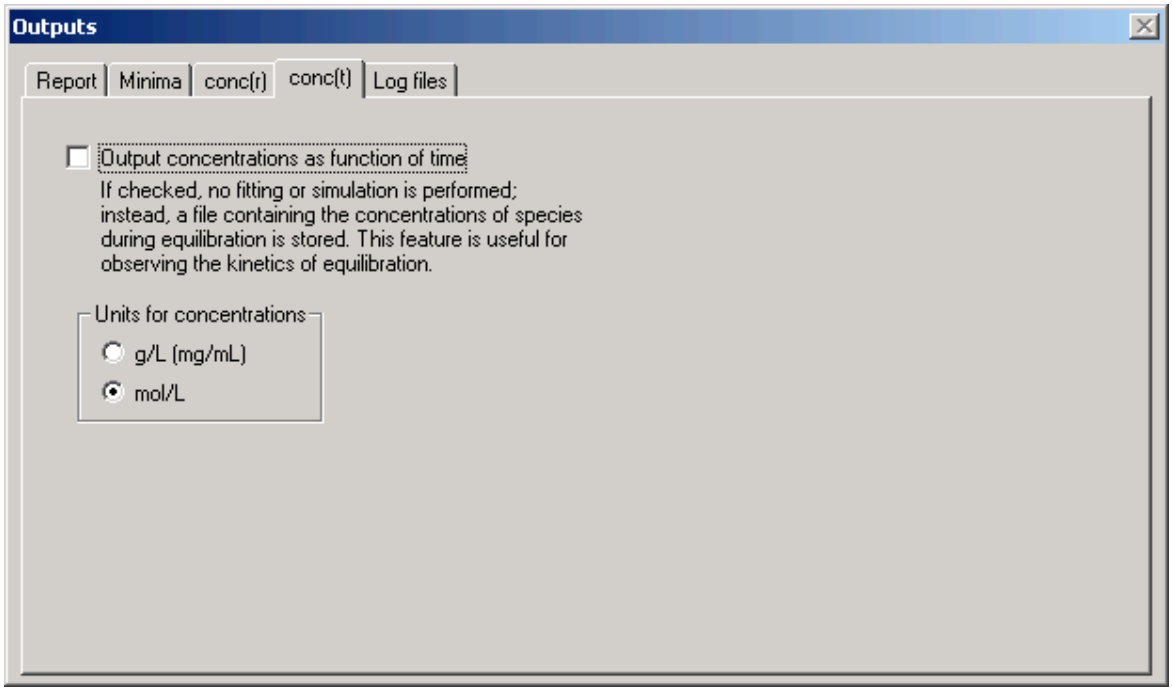


Figure 102.

2.6.7.4 Disposition of various log files

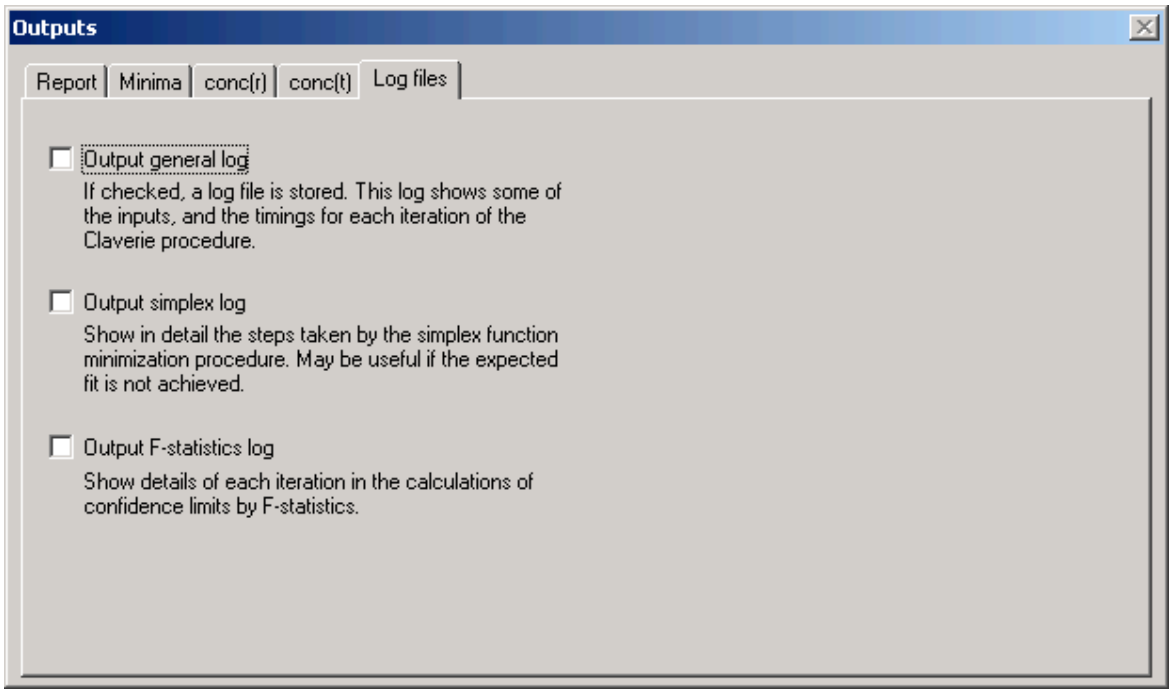


Figure 103.

2.6.7.5 Preferences for indefinite self-association models

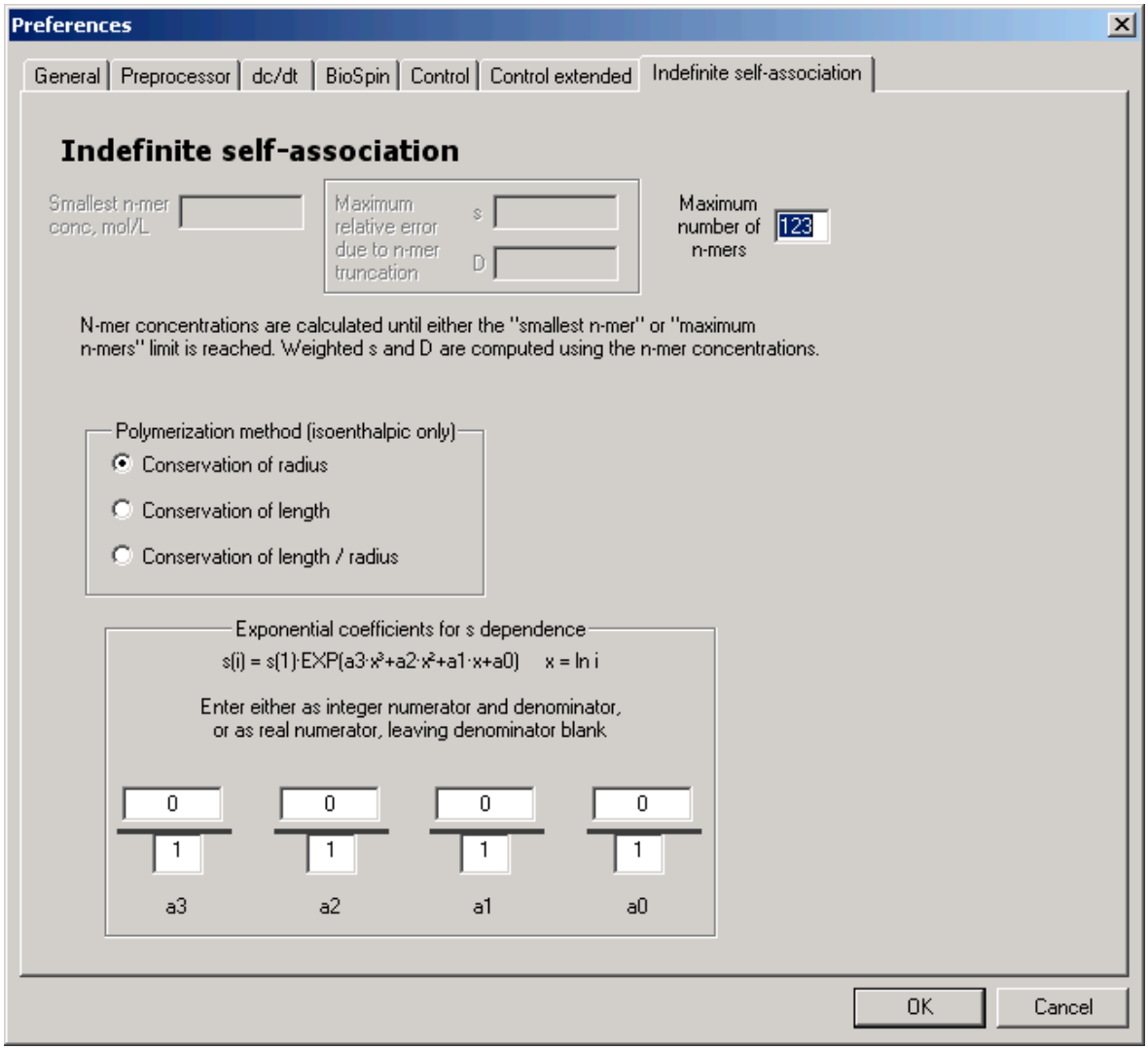


Figure 104. Setting the default parameters for indefinite self-association reactions.

2.6.8 Fitting preferences

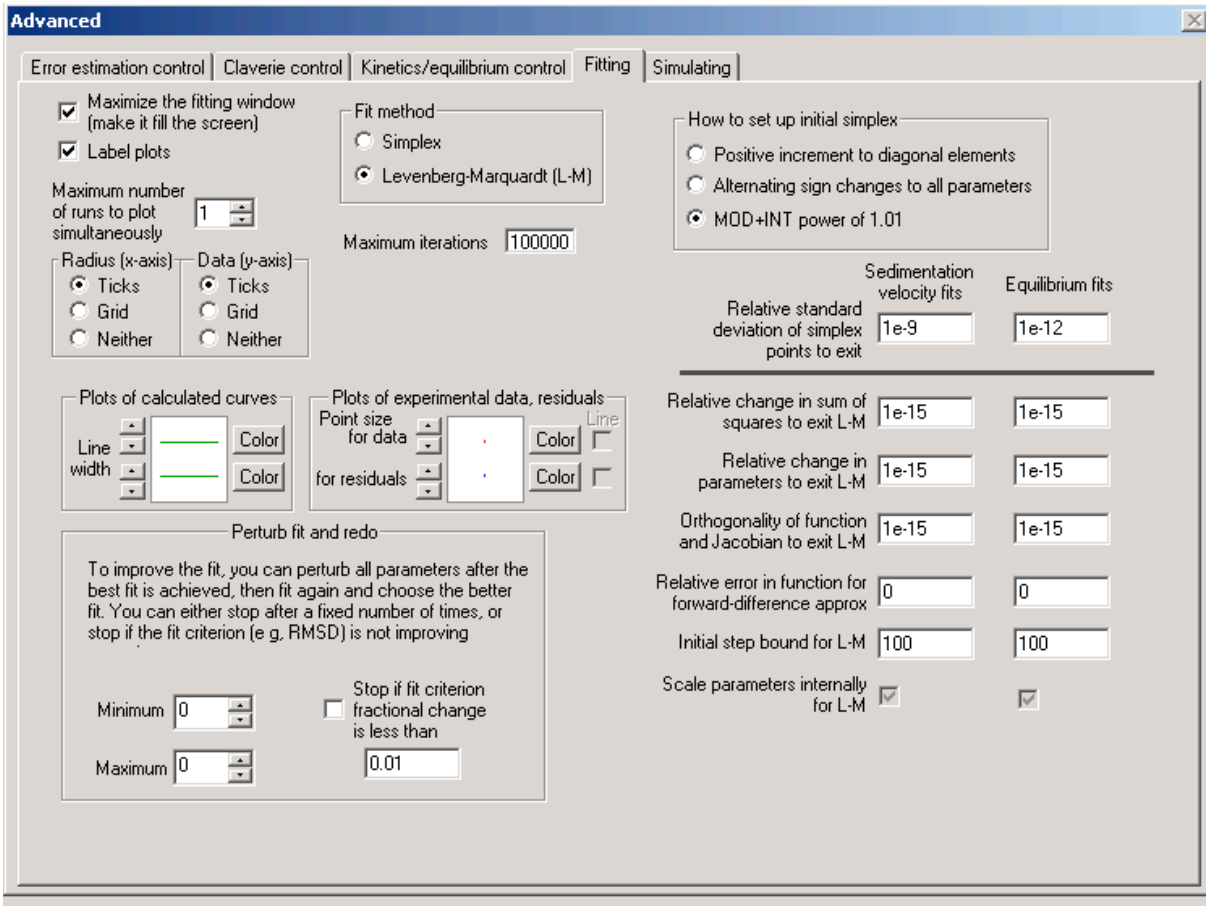


Figure 105. Fitting preferences.

[the "MOD+INT power of 1.01" button causes each element of the initial simplex to be filled with a different number. The "Positive increment …" button alters only the diagonal elements of the initial simplex. Don't use the "Alternating sign …" button; we don't remember what is does.

2.6.9 Simulation Preferences

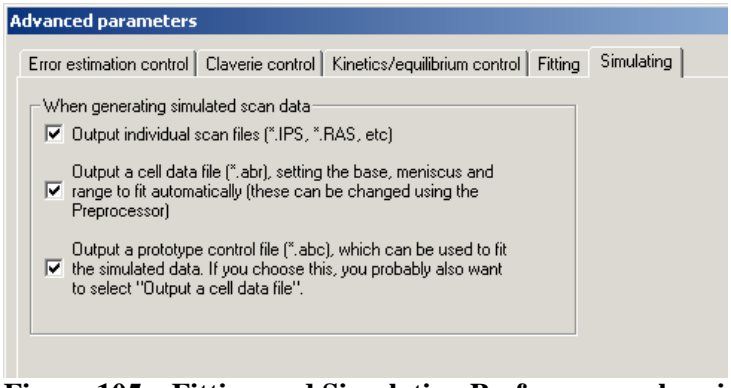


Figure 105a. Fitting and Simulating Preferences - choosing output files

2.7 BIOSPIN

BIOSPIN is the program originally developed by Dennis Roark and David Yphantis to compute point-by-point molar mass averages as a function of the local cell concentration from sedimentation equilibrium data (D.E. Roark & D.A. Yphantis, 1969). It has been implemented as a subprogram of SEDANAL. The reader is referred to the original literature for further information. In addition to the point-wise number, weight, Z- and Z+1 molar mass averages, it also computes several of the so-called “charge-independent” molar mass averages that are independent of the second and higher virial coefficients: the Y1, Y2, Y3, etc… averages. (Yphantis & Roark, 1972). The charge independent averages are useful for analyzing systems under non-ideal conditions.

2.8 SIMULATING DATA

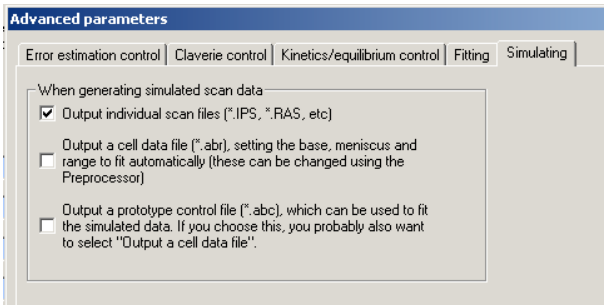
To simulate XLA/I data, select "Fit preprocessed data" from the Main Menu. When the Control Screen appears, click the "new" button and either create a new "experiment" folder or select an existing “experiment” folder in which the simulated data files will be stored. Then click the button labeled "Simulate data"; the screen will change to the simulator screen. Now select a model from the drop-down "Model to be fitted" list. Enter the number of points to be used for the grid between meniscus and base. Enter the molecular parameters in the "Molecular parameters" boxes.

In the bottom panel, select the optical system; indicate whether it is a velocity or equilibrium experiment: enter a comment; specify the meniscus, speed, base and loading concentration. For a velocity run enter the total number of scans and the time between scans. Now enter the magnitude of normally distributed random noise to add to the simulated data.

The window labeled "Bottom time, sec" will display the total time for the slowest species to reach the bottom of the cell. You can adjust the total number of scans and time between scans as you desire.

When all the boxes have been filled, click "Store control file and start fit" to start the simulation and to save the simulation control file. The simulated data files will have names of the form 000nnn.IPs or 00nnn.RAs and will be in XLA/I standard format. The simulated data, *.IPs or *.RAs files, will be stored in standard time and date subdirectories of the experiment directory which was selected when you clicked on the "New" button. The time and date used for the path name will be the same as the time and date of the simulation run. The simulated XLA/I data files are recognized and can be processed like any other XLA/I data files by SEDANAL or other programs.

Simulation can generate a cell data (run, or .abr) file, as well as the usual scan files. The run file's name is YYYYMMDD_simHHMMSS.abr, and it has the meniscus and cell base from the simulation, and the range to fit is (meniscus+0.05) - (base-0.1) (i e, 5.95-7.1 for m=5.9 and b=7.2). It can also write a prototype control file that can be used to fit the simulated data. To turn on either of these features go to the "Advanced …" button and click on the "Simulating" tab and select the appropriate buttons. A prototype fitting control file can also be produced by clicking the appropriate button below:



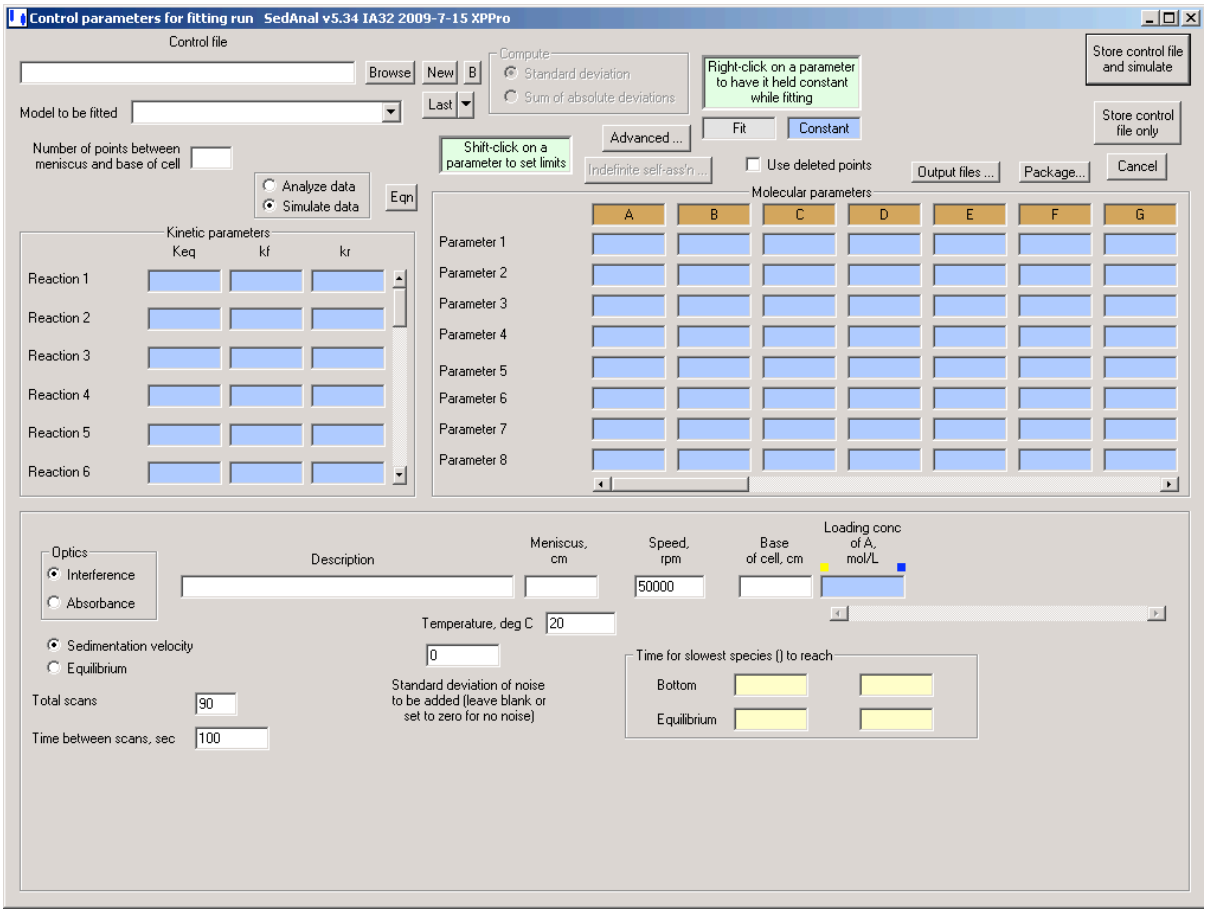


Figure 106. Control screen after "Simulate data" has been selected.

2.9 Other features:

2.9.1 Output of initial reaction time course.

(See section 2.10 for a more user friendly and comprehensive way of computing the kinetics.)
Kinetics: To write a file of the time course of the initial equilibration of a particular model, click on the "Output files button and select the "conc(t)" tab and check the "Output concentrations as a function of time" box. Then select the units for the concentrations

Figure 107. Output files

Figure 108. Outputting the time course of the initial equilibration step

When the "Store control file and …" button is clicked, only the kinetics of the model are calculated and a file named after the abc and report files is written. No fitting will be done in this case.

If the abc file is named xxxxxx.abc, the report file will be named xxxxxx_Report001.rtf and the kinetic time course will be found in a filed named: xxxxxx _Equilibration.txt. Remember **to unclick the "Output concentrations as a function of time" box when you are done.**

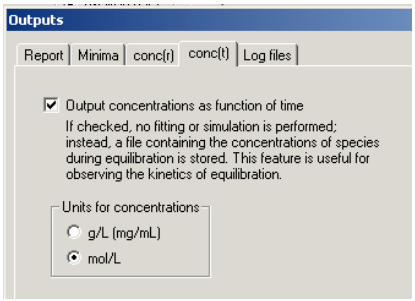


Figure 108.

2.10 Kinetics Simulator

A general purpose kinetics simulator has been added to SEDANAL and is accessed from the Main Menu by clicking the Kinetics button. The kinetics simulator can handle any model that can be represented in the ModelEditor. When the kinetic simulator screen is first opened, a blank screen showing the possible combinations of 28 species and 27 reactions. One should click on "New" to create a new kinetics control file, or select a previously used control file from either the drop down menu or by browsing for it.

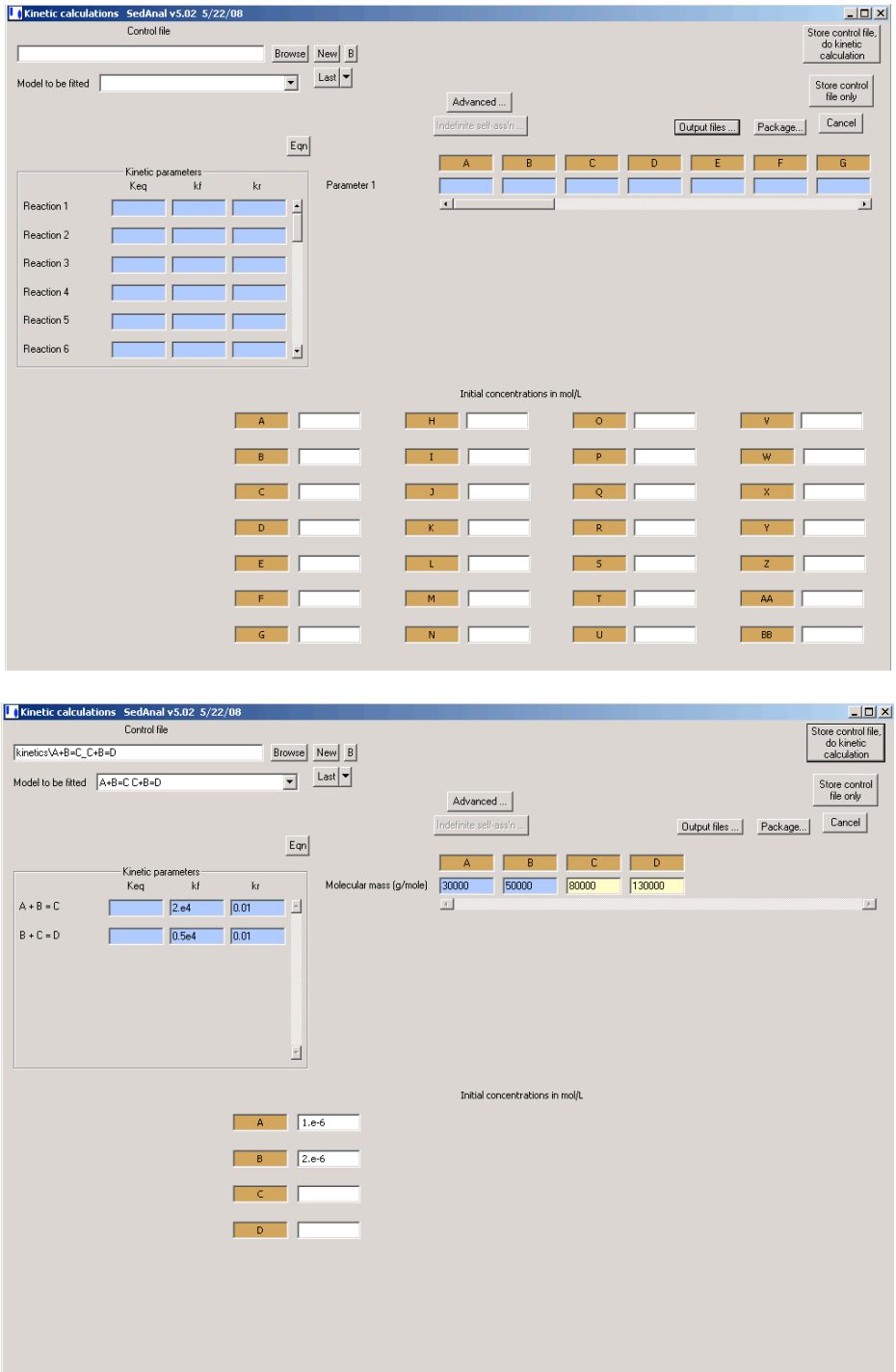


Figure 109 . Kinetics control screen for the model A+B=C; C+B=D. Molar masses of A and B are 20,000 and 50,000 g/mol, respectively. The forward and reverse rate constants for the first and second reactions are 2 x 10^4 $M^{-1}s^{-1}$, 0.01 s^{-1}, and 0.5 x 10^4 $M^{-1}s^{-1}$, 0.01 s^{-1}, respectively. The initial concentrations of A and B are 1 x 10^{-6} and 2 x 10^{-6}M, respectively.

Clicking "Store control file and do kinetic calculation" results in the following display.

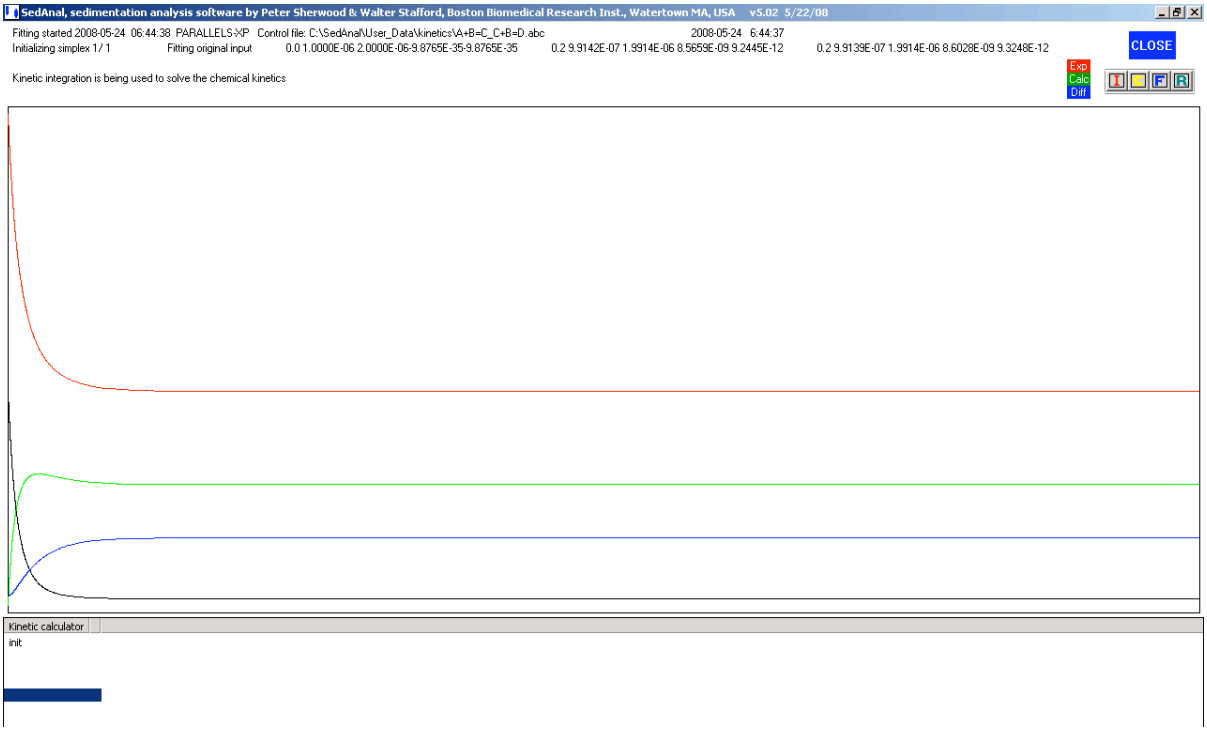


Figure 110. Screen dump of kinetic simulation

These data have been written to a file whose file name will have the following form:

New_Control_Filename_Equilibration.txt.

The text "New_Control_Filename" will be replaced by the name you gave the control file after you pressed "New" or changed the control file name after loading an exiting abc file with either the "Last" or "Browse" buttons.

These simulated data can be read into your favorite plotting program to produce a plot something like this:

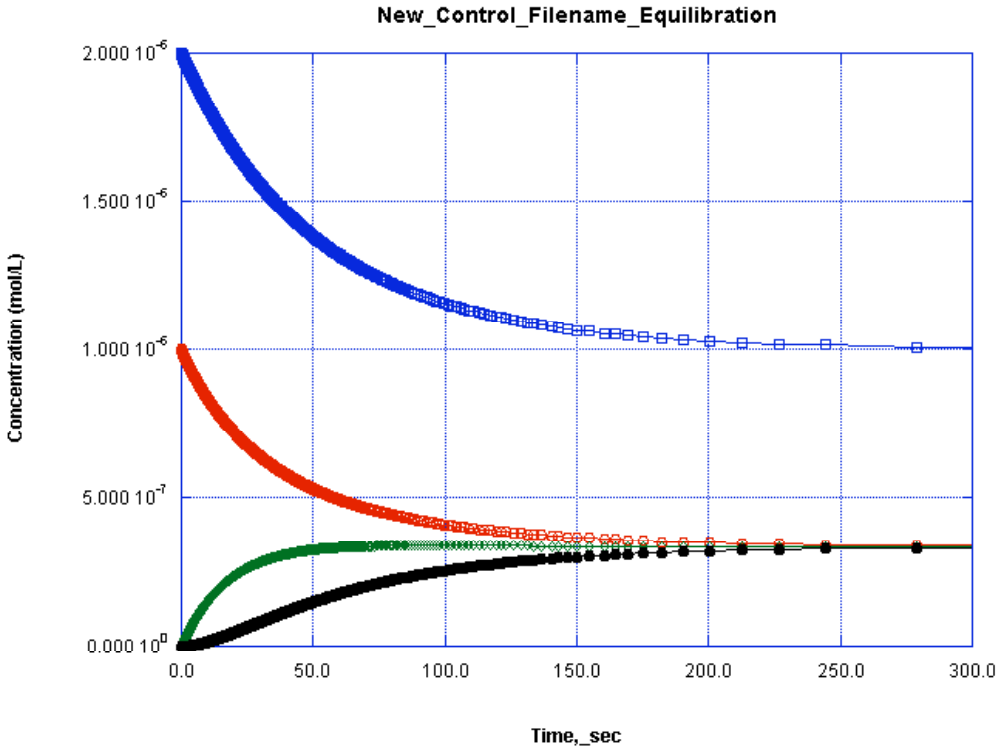


Figure 111. Plot of molar concentrations as a function of time for the system A+B=C; C+B=D. The forward and reverse rate constants for both the first and second reactions are 1 x 10^4 $M^{-1}s^{-1}$, 0.01 s^{-1}, respectively. The initial concentrations of A and B are 1 x 10^{-6} and 2 x 10^{-6}, respectively.

Here is another example with the same equilibrium constants but the second reaction forward and reverse rate constants are one tenth those the previous example.

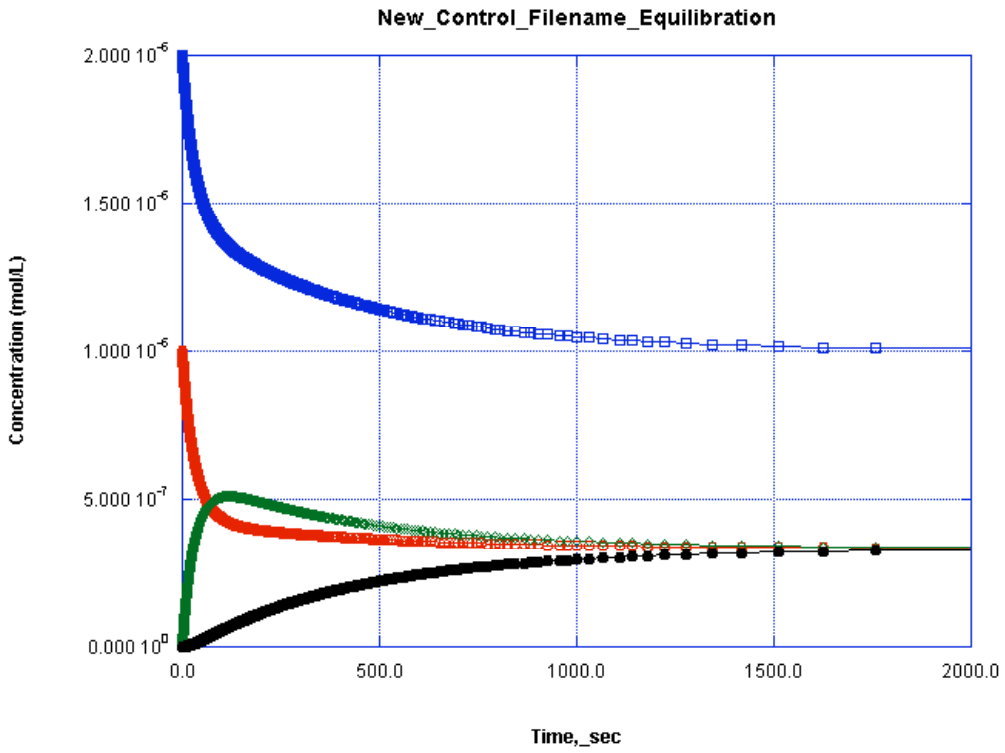


Figure 112. Another plot with same equilibrium constants but forward and reverse rate constants for the second step are 1/10 those of the first example above. The **forward and reverse rate constants for the first and second reactions are 1 x 10^4 $M^{-1}s^{-1}$, 0.01 s^{-1}, and 1 x 10^3 $M^{-1}s^{-1}$, 0.001 s^{-1}, respectively. The initial concentrations of A and B are 1 x 10^{-6} and 2 x 10^{-6}, respectively.**

2.11 Scripting sedimentation fitting and simulation

The script functionality is rudimentary at present, to give us a chance to experiment with it and find out what is useful and what isn't. Not all the error-checking or UI features are present.

The script files must be in the User_data directory, and have an extension of .abs. Script files are created and modified with a text editor.

2.11.1 Scripting commands:

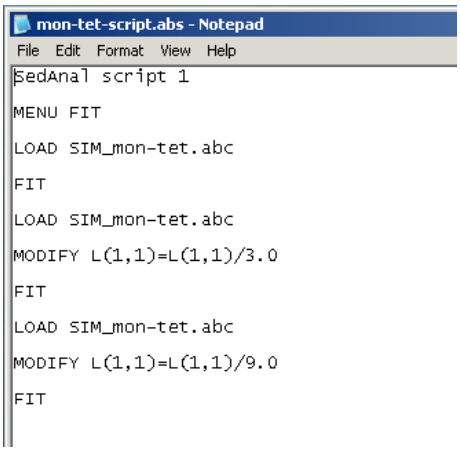
The initial line of the script must be exactly

```
SedAnal script 1
```

The following commands may be included in a script:

Command	Meaning	Example
`MENU FIT`	Equivalent to clicking Fit preprocessed data on the main menu.	`MENU FIT`
`LOAD` *controlFile*	Specify a control file to be loaded; *controlFile* can be either the name of a control file (see below), or `$LAST`.	`LOAD MyExperiment\ABCD_Fit`
`MODIFY` *expression*	Change the value for a kinetic, molecular, or cell parameter from the value in the control file, and as previously modified. The symbols and format of the expression are the same as in the equation editor.	`MODIFY K(1)=K(1)/10`
`FIT`	Equivalent to clicking Store control file and start fit on the control screen. See note below.	`FIT`
blank lines	These lines are ignored.	[this line intentionally left blank]
comment lines	These lines are ignored, except for checking spelling, grammar, and clarity of style.	`// This is a comment`

A typical script file might look like:



This particular script was used to simulate a dilution series of a monomer-tetramer system creating three abr files for 1 to 3 serial dilutions of the monomer tetramer system that was specified in the "SIM_mon-tet.abc" file. This script will load the original abc file and perform the simulation; then reload it and simulate again with the loading concentration divided by 3; and then repeat with the loading concentrations divided by 9. This is very useful for generating large data sets for exploratory studies of interacting systems.

2.11.2 Specifying control files in scripts

The control file is identified by the following rules:

1) If it's just a name, the complete path is completed by trying the directory of the most recent control file loaded, then the most recent control file stored for this instance of SedAnal, then the most recent control file stored for any instance of SedAnal. The idea is that SedAnal tries to guess what you mean if you just say, e g, `ABCD_Fit`.

2) If it is a directory followed by a name (e g, `MyExperiment\ABCD_Fit`), the directory is assumed to be in the user data directory.

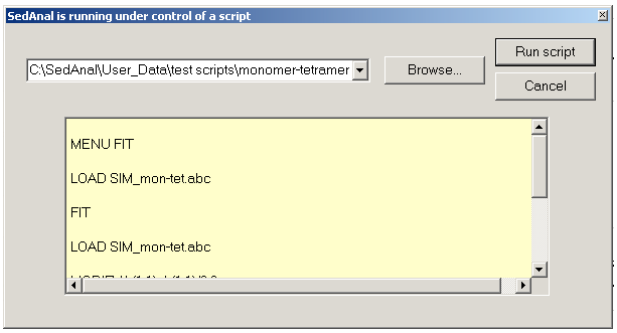
3) Anything else is a complete path (e g, `C:\SedAnal\User_data\MyExperiment\ABCD_Fit`).

4) If the control file name does not end in `.abc`, the extension is added. Thus, `ABCD_Fit` and `ABCD_Fit.abc` mean the same.

2.11.3 Storing scripting control files

To avoid modifying the original control file, when you have `FIT` in a script, the control file is stored with the suffix "`_script001`". For example, if the script in the example column above is run, the name of the control file with K_1 a factor of 10 smaller will be `ABCD_Fit_script001.abc`, and `ABCD_Fit.abc` will not be changed.

When a script file has been loaded, the following screen will appear:



The script is initiated by clicking on "Run script".

2.12 Help

An embryonic help system has been conceived and is growing … It can be accessed from the main menu. Well, actually it does nothing just now… Read this manual.

2.13 Exit

Self-explanatory … bye…

3 Background Theory

The curve fitting portion of SEDANAL is designed to analyze sedimentation velocity data from interacting systems composed of multiple macromolecular components.

First, a word about **components**, **species** and **reactions**:

In the context of this discussion, a component is an electroneutral macromolecule or several molecules related by chemical reaction that can be added or removed, at least conceptually, from a solution independently of other molecules. A component may be composed of several macromolecular species. Species interact through chemical reactions. In general, the number of components is equal to the number of species minus the number of chemical reactions. At constant temperature and pressure (Actually, we are assuming here that the solution is incompressible), the number of degrees of freedom is equal to the number of components. For any given values of the equilibrium constants, the composition of the solution (i.e. the amount of each species) is determined by the total concentration of each component at each position in the centrifuge cell.

An example of a single component, multi-species system is a monomer-dimer, rapidly reversible self-association. The monomer and dimer are in equilibrium with each other and, therefore, cannot be added to or removed from the solution independently. This is why they cannot be separated on a gel filtration column, for example. As soon as some dimer and monomer become separated either the monomer reassociates to form more dimer or the dimer dissociates to form more monomer. At any given temperature and pressure, the amount of monomer and dimer present is determined completely by the total concentration of the component and the equilibrium constant for the dimerization reaction. This is a one component system composed of two species with one chemical reaction between them. If this were a monomer-dimer-tetramer system, it would be a one component, three species system with two chemical reactions between them. It would have one degree of freedom such that the amounts of monomer, dimer and tetramer at each point in the boundary would be uniquely determined by the total macromolecular concentration at that point.

Similarly, consider a system composed of two components, A and B that interact to form a complex, C. Let C interact with another molecule of B to form D. This is a two component system which is composed of four species that are related by 2 chemical reactions. The system has two degrees of freedom, the total concentrations of A and B, respectively.

$$A + B = C \qquad K1=k_{1f}/k_{1r}$$
$$C + B = D \qquad K2=k_{2f}/k_{2r}$$

However, if D were composed of 1 mole of A and 2 moles of B but was not in equilibrium with C and B, it would be a third component. This system would have 3 degrees of freedom, the total concentrations of A, B that participate in the reactions and the total concentration of D. In principal D could be removed from the system by gel filtration without being reformed from C and B. D might be a covalently cross-linked aggregate, for example. So …

Number-of-components = number-of-species - number-of-reactions

3.1 Sedimentation Velocity Theory:

3.1.1 Concentration time-difference curves.

Consider the signal obtained from the centrifuge. Whether it is fringe displacements, absorbance or fluorescence, it is a function of both time and radius, call it S(r,t). After S(r,t) has been preprocessed to

remove optical jitter and integral fringe shifts, it is composed of the contribution from the true concentration distribution, C(r,t), which is also a function of time and radius, as well as a background optical systematic component, B(r), that is time independent and a function of radius only. The signal, S(r,t), also has stochastic noise included with it. We have

$$S(r,t) = \alpha C(r,t) + B(r) + noise$$

where α is the conversion factor (also known as the extinction coefficient) between concentration (g/L) and either fringes, absorbance or fluorescence.

We can remove completely the time independent background component, B(r), of the signal by subtracting any two experimental curves, say at times t_1 and t_2.

$$S(r,t_2) = \alpha C(r,t_2) + B(r) + noise$$
$$S(r,t_1) = \alpha C(r,t_1) + B(r) + noise$$
$$- -$$
$$\Delta S(r,t_1,t_2) = \Delta\alpha C(r,t_1,t_2) + 0 + \sqrt{2}\, noise$$

The time difference curves, $\Delta S(r,t_1,t_2)$, are proportional to the concentration difference curves, $\Delta C(r,t_1,t_2)$ but have no time independent systematic error. The irregular, time dependent, systematic, error from the optics (often referred to as jitter) will have been removed at the preprocessing stage. The data to be fitted, $\Delta S(r,t_1,t_2)$, have only stochastic errors which makes them suitable for least squares fitting.

SEDANAL fits to the time difference curves $\Delta S(r,t_1,t_2)$ for any set of parameter guesses by generating concentration curves, $C(r,t_1)$ and $C(r,t_2)$, corresponding to times t_1 and t_2, and subtracting them to form $\alpha\Delta C(r,t_1,t_2)$. The root mean square residual is computed as the triple sum over all the points, scans and cells.

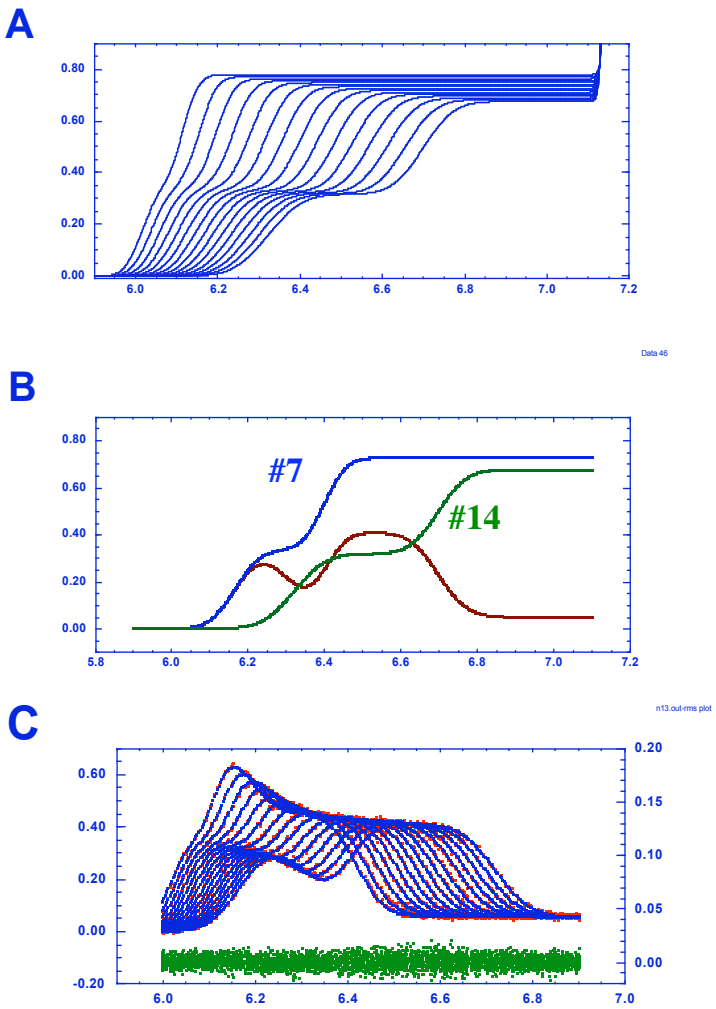
$$\text{rmsd} = \left[\frac{\sum_{k=1}^{L} \left\{ \sum_{j=1}^{M} \left(\sum_{i=1}^{N} \left(\Delta S_{k,j}(r_i,t_j,t_{j+M}) - \alpha\Delta C_{k,j}(r_i,t_j,t_{j+M}) \right)^2 \right) \right\}}{LMN} \right]^{1/2}$$

where k is the cell number, L is the number of cells, j is the difference curve index, M is the number of difference curves and 2M is the total number of scans, j and j+M are the indices of the scans being subtracted, N is the number of radial points in each difference curve, and i is the radial point index.

Choose a series of scans to be analyzed (e.g. 14 scans) with the Preprocessor

Subtract scans in pairs to generate Δc curves to eliminate time independent background (e.g. C_8-C_1, C_9-C_2...)

Fit to the Δc curves
(from which time independent systematic errors have been eliminated by the subtraction)



The figure above schematically outlines the basic procedure used by SEDANAL to fit the time difference data, Δc vs. radius.

The rmsd is minimized with respect to the model parameters using the simplex directed search method of Nelder and Mead (1965)

The numerical solutions to the Lamm equation are generated using the finite element method of Claverie, (1975, 1976).

The general fitting approach using numerical solutions to Lamm equation and the Claverie procedure is based on the method of Todd and Haschemeyer, (1981)

The kinetic differential equations are solved using the Bulirsch-Stoer algorithm with Richardson extrapolation. (Numerical Recipes in FORTRAN)

Bulirsch, R. and Stoer, J. §2.2 in Introduction to Numerical Analysis. New York: Springer-Verlag, 1991.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. ``Richardson Extrapolation and the Bulirsch-Stoer Method." §16.4 in Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed. Cambridge, England: Cambridge University Press, pp. 718-725, 1992.

3.2 Sedimentation Equilibrium Theory

3.2.1 Ideal Case:

For a thermodynamically ideal, incompressible system we have for each species that

$$c(r) = c(ref)\exp\left(\sigma\left(\xi - \xi_{ref}\right)\right)$$

where σ is defined as $\sigma = M\left(1-\nu\rho\right)\omega^2/RT$, $\xi=r^2/2$ and $\xi_{ref}=r^2_{ref}/2$ where r_{ref} is an arbitrary reference radius usually chosen as the first data point. The value of c_{ref} can be related to the loading concentration by invoking conservation of mass and noting that the concentration at the beginning of the run is uniform and equal to the loading concentration, C_o. The total amount of macromolecule, T, in the cell is given by multiplying the concentration by the volume:

$$T = h\theta\int_{\xi_m}^{\xi_b} c_o d\xi = hc_o\left(\xi_b - \xi_m\right)$$

since this quantity doesn't change with time the following must also be true at equilibrium:

$$T = h\theta\int_{\xi_m}^{\xi_b} c\left(\xi\right)d\xi = h\theta\int_{\xi_m}^{\xi_b} c_{ref}\exp\left(\sigma\left(\xi - \xi_{ref}\right)\right)d\xi$$

$$T = \frac{h\theta c_{ref}}{\sigma}\left[\exp\left(\sigma\left(\xi_b - \xi_{ref}\right)\right) - \exp\left(\sigma\left(\xi_m - \xi_{ref}\right)\right)\right]$$

Now we can equate the two relationships and write:

$$c_o\left(\xi_b - \xi_m\right) = \frac{c_{ref}}{\sigma}\left[\exp\left(\sigma\left(\xi_b - \xi_{ref}\right)\right) - \exp\left(\sigma\left(\xi_m - \xi_{ref}\right)\right)\right]$$

So, C_{ref} can be related to C_o if the position of the meniscus and the base of the cell can be determined with satisfactory accuracy.

$$c_{ref} = \frac{c_o\sigma\left(\xi_b - \xi_m\right)}{\left[\exp\left(\sigma\left(\xi_b - \xi_{ref}\right)\right) - \exp\left(\sigma\left(\xi_m - \xi_{ref}\right)\right)\right]}$$

For a reversibly, non-interacting system, each species' loading concentration can be related its value of c_{ref} in this way, independently of the other species. (In this case, each species is also an independent component.)

For reversibly interacting systems, on the other hand, the law of mass action has be taken into account. This leads to somewhat more complicated, but still easily solvable relationships between the loading concentrations and the values of c_{ref} for each species in equilibrium.

We will consider two reversible systems as examples: a monomer-dimer and a simple one-to-one hetero-association.

3.2.1.1 Monomer-dimer:

$$2A = A_2 \qquad k_{1,2} = \frac{[c_2]}{[c]^2}$$

$$T = h\theta\int_{\xi_m}^{\xi_b} c(\xi)\,d\xi = h\theta\int_{\xi_m}^{\xi_b} c_{1,ref}\exp\left(\sigma_1\left(\xi-\xi_{ref}\right)\right) + c_{2,ref}\exp\left(2\sigma_1\left(\xi-\xi_{ref}\right)\right)d\xi$$

$$c_o\left(\xi_b-\xi_m\right) = \frac{c_{1,ref}}{\sigma_1}\left[\exp\left(\sigma\left(\xi_b-\xi_{ref}\right)\right)-\exp\left(\sigma\left(\xi_m-\xi_{ref}\right)\right)\right] + \frac{c_{2,ref}}{2\sigma_1}\left[\exp\left(2\sigma_1\left(\xi_b-\xi_{ref}\right)\right)-\exp\left(2\sigma_1\left(\xi_m-\xi_{ref}\right)\right)\right]$$

Now, invoking the Law of Mass Action, we can write:

$$c_o\left(\xi_b-\xi_m\right) = \frac{c_{1,ref}}{\sigma_1}\left[\exp\left(\sigma\left(\xi_b-\xi_{ref}\right)\right)-\exp\left(\sigma\left(\xi_m-\xi_{ref}\right)\right)\right] + \frac{k_{1,2}c_{1,ref}^2}{2\sigma_1}\left[\exp\left(2\sigma_1\left(\xi_b-\xi_{ref}\right)\right)-\exp\left(2\sigma_1\left(\xi_m-\xi_{ref}\right)\right)\right]$$

This can be solved easily for $c_{1,ref}$ with the quadratic equation. For higher oligomers and polydisperse self-association, the equations must be solved by successive approximation.

3.2.1.2 Hetero-Association:

Now for the hetero-association,

$$A + B = AB \qquad K_{AB} = \frac{c_{AB}}{c_A c_B}$$

Conservation of mass for each component, A and B, requires

$$c_{A,o} = \frac{c_{A,ref}}{\left(\xi_b-\xi_m\right)\sigma_A}\left[\exp\left(\sigma_A\left(\xi_b-\xi_{ref}\right)\right)-\exp\left(\sigma_A\left(\xi_m-\xi_{ref}\right)\right)\right] +$$
$$\left[\frac{M_A}{M_{AB}}\right]\frac{k_{AB}c_{A,ref}c_{B,ref}}{\left(\xi_b-\xi_m\right)\left(\sigma_A+\sigma_B\right)}\left[\exp\left(\left(\sigma_A+\sigma_B\right)\left(\xi_b-\xi_{ref}\right)\right)-\exp\left(\left(\sigma_A+\sigma_B\right)\left(\xi_m-\xi_{ref}\right)\right)\right]$$

$$c_{B,o} = \frac{c_{B,ref}}{\left(\xi_b-\xi_m\right)\sigma_B}\left[\exp\left(\sigma_B\left(\xi_b-\xi_{ref}\right)\right)-\exp\left(\sigma_B\left(\xi_m-\xi_{ref}\right)\right)\right] +$$
$$\left[\frac{M_B}{M_{AB}}\right]\frac{k_{AB}c_{A,ref}c_{B,ref}}{\left(\xi_b-\xi_m\right)\left(\sigma_A+\sigma_B\right)}\left[\exp\left(\left(\sigma_A+\sigma_B\right)\left(\xi_b-\xi_{ref}\right)\right)-\exp\left(\left(\sigma_A+\sigma_B\right)\left(\xi_m-\xi_{ref}\right)\right)\right]$$

This pair of equations can be solved for $c_{A,ref}$ and $c_{B,ref}$, given values of $c_{A,o}$ and $c_{B,o}$, easily by rearrangement into an analytical expression. More complicated stoichiometries can be solved easily by successive approximation (The Newton-Raphson method works nicely in the general case). Success in fitting multi-component systems globally to multiple data sets requires accurate knowledge of the meniscus and base positions in order to estimate the correct values of c_{ref} from the conservation of mass relationships.

3.2.2 Non-ideal Case:

For a thermodynamically non-ideal incompressible system we have :

Addition 2010-2-12

$$c_i(r) = c_i(r_m)\, e^{\sigma_i \xi - \tau} \qquad i = 1, \ldots, n_s$$

$$\tau = \sum_{k=1}^{n_s} \left(2BM1_k \cdot c_k + \frac{3}{2} CM1_k \cdot c_k^2 \right)$$

These $n_s + 1$ eqns are solved simultaneously for c_i and τ, at each r.

$$c_i(r) = c_i(r_m)\, \exp\left(\sigma_i \xi - \tau\right)$$

or

$$f_i \equiv c_i(r_m)\, e^{\sigma_i \xi} \cdot \exp\left(-\tau\right) - c_i = 0 \qquad i = 1, \ldots, n_s$$

We need the Jacobian of $\mathbf{f}$:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{c}} = \begin{pmatrix} \frac{\partial f_1}{\partial c_1} & \frac{\partial f_1}{\partial c_2} & \cdots & \frac{\partial f_1}{\partial c_{n_s}} \\ \frac{\partial f_2}{\partial c_1} & \frac{\partial f_2}{\partial c_2} & \cdots & \frac{\partial f_2}{\partial c_{n_s}} \\ & \vdots & \ddots & \\ \frac{\partial f_{n_s}}{\partial c_1} & \frac{\partial \dot{f}_{n_s}}{\partial c_2} & \cdots & \frac{\partial f_{n_s}}{\partial c_{n_s}} \end{pmatrix}$$

$$\frac{\partial f_i}{\partial c_j} = c_i(r_m)\, e^{\sigma_i \xi} \cdot \exp\left(-\tau\right) \cdot \left(-2BM1_j - 3CM1_j \cdot c_j\right) - \delta_{ij}$$

using $\dfrac{\partial \tau}{\partial c_j} = 2BM1_j + 3CM1_j \cdot c_j$, and δ_{ij} is the Kronecker delta (1 if $i = j$, else 0).

In the program SedAnal, the variables corresponding to the expressions above are

`a(i) + b(i) * rad**2`	$\sigma_i \xi$	`EqRunSim`, etc
`concEq(i)`	$c_i(r_m)$	`EqRunSim`, etc
`s`	$c_i(r_m)\, e^{\sigma_i \xi}$	`funcNonIdeal`
`nonIdealExp`	$\exp\left(-\sum_{k=1}^{n_s} 2BM1_k \cdot c_k + \frac{3}{2} CM1_k \cdot c_k^2\right)$	`JacobianNonIdeal`

---

4 APPENDIX A: CHANGE LOG.

See http://sedanal.bbri.org/sedanal_change_log.html

5 INDEX

(1+2BM$_1$ c), 49
(1+k_sc), 49
abbreviated boot-strap procedure, 74
absorbance, 5, 10, 22, 108, 109
Advanced Parameters, 46
aggregate, 14, 108
air-air space, 22, 25
alignment, 25
ALL CELLS, 52
Analyze data, 6, 41
antigen-antibody, 13
association constant, 12
bad scans, 23, 24, 30, 84
base, 6, 22, 28, 32, 43, 51, 98
best fit, 59
BIOSPIN, 86, 89, 98
blue, 10, 54, 59
BM1, 9, 15, 16, 17, 49
boot-strap, 74, 75, 76
boot-strap with replacement, 74
BSA, 7
Bulirsch-Stoer, 48, 110
Claverie, 110
complex, 12, 108
component, 9, 10, 11, 12, 13, 14, 44, 50, 54, 68, 108, 109, 111, 112
concentration dependence, 9, 15, 49
concentration distribution, 109
confidence level, 77, 78
control file, 6, 39, 40, 50, 59, 98
Control Screen, 6, 9, 10, 11, 14, 18, 19, 23, 24, 39, 41, 42, 43, 44, 53, 56, 57, 58, 59, 62, 63, 64, 65, 66, 69, 71, 75, 77, 79, 86, 90, 92, 93, 98
DCDT, 6, 24, 34, 78
degrees of freedom, 108
density increment, 9, 43
deviations, 60
difference curve, 59, 60, 109
difference curves, 5, 59, 60, 108, 109
differential equations, 48, 110
diffusion coefficient, 15
dimer, 108
directory structure, 6, 7
dummy file, 24
Equation Editor, 18, 43
equilibrium, 12, 13, 43, 108
equilibrium constant, 12, 13, 43, 108
equilibrium constants, 108
Error Estimation Control, 46
expand, 32
experiment, 7, 22, 39, 40, 50, 98
experiment folder, 7, 22, 98
extinction coefficient, 9, 13, 43, 109
finite element solutions, 51
Fit Preprocessed Data, 39
Fitting preferences, 97
Fitting Tab, 48
fluorescence data, 22, 91
forward rate, 12, 43
forward rate constant, 12, 43
fractional change, 46
frictional coefficient, 9, 15, 49
fringe displacements, 108
fringes, 9, 10, 22, 31, 109
F-stat progress, 78
F-statistics, 46, 74, 77, 78
F-statistics monitoring window, 78
gel filtration, 108
Global fitting, 52
global parameter, 54
Gralén, 49
gray, 10, 54
grid density, 51
grid spacing, variable, 47
Haschemeyer, Rudy, 110
hinge point, 22
ideal, 9, 11
Indefinite self-association, 17
initial equilibration step, 100
initial guesses, 6, 59
integer fringe jumps, 25
integral fringe adjustment, 27
integral fringe shifts, 27, 31, 109
intensity data, 22
interacting systems, 108
interference, 5, 10, 22, 25
interpolated, 51
isodesmic, 17, 56
isoenthalpic, 17, 18, 57
iterations, 43, 51
jitter, 22, 31, 109
Jitter adjust, 25
kinetic, 43, 48, 110
Kinetic integrator control, 48
kinetics, 5, 43, 48, 100
Kinetics Simulator, 6, 100
Ks, 15, 16, 17
Lamm, 5, 43, 46, 48, 51, 110
lease squares, 31
least squares fitting, 109
Levenberg-Marquardt, 48
loading concentration, 43, 50, 54, 98
Locate meniscus, 32
Main Menu, 21, 39, 98
meniscus, 5, 6, 22, 27, 28, 31, 32, 43, 51, 98
min, 60
model, 6, 9, 10, 11, 14, 32, 41, 42, 43, 98, 110
Model Editor, 14
Model to be fitted, 41, 98
ModelEditor, 6, 7, 9, 11
molar mass, 43
molecular weight, 9, 10
monomer, 108
monomer-dimer, 108
Monte Carlo, 46, 74, 75, 76
multiple optical systems, 52
non-ideality, 9, 16, 17
Non-interacting systems, 18
non-linear least squares, 43
Numerical Recipes in FORTRAN, 110

ODE, 43, 48
optical system, 10, 98
Ordinary Differential Equation. See ODE, See ODE
output file, 59
parameter, 9, 11, 14, 26, 109
parameters, 5, 6, 9, 10, 13, 15, 43, 46, 48, 49, 51, 53, 60, 98, 110
partial specific volume, 9, 13, 43
Perturb fit and re-do, 49
plateau, 22, 31
Preferences, 6, 7, 86, 87
preprocessed data, 6, 32, 98
preprocessor, 22, 50
ProteomeLab, 5
pseudo-absorbance, 22
range, 6, 11, 14, 22, 28, 32
rapidly reversible, 108
rate constants, 43, 48
reaction, 6, 9, 13, 43, 108
remove data set, 50
rescaled, 29
residuals, 5, 59, 60
reverse rate, 12, 43
reverse rate constant, 12, 43
Richardson extrapolation, 110
Richardson rational polynomial extrapolation, 48
rms deviation, 14
rmsd, 5, 110
Roark, 98
root mean square deviation, 60
root mean square residual, 109
run files, 7, 50
Run script, 107
scans, 29, 31, 50, 98, 109
scans to be fitted, 29
Screen Dumps, 60
script, 105, 106, 107
Scripting, 105
SEDANAL, 5, 6, 7, 10, 21, 48, 59, 98, 108, 109, 110
sedimentation coefficient, 9, 10, 11, 12, 43
sedimentation equilibrium, 5, 64, 65, 66, 69, 71, 98
Select scans to be plotted, 29
self association, 9, 17, 57, 96, 108, 112
Simulate data, 6, 41
simulating data, 98
slaved cell data sets, 52, 53
species, 9, 10, 11, 12, 13, 14, 16, 17, 43, 48, 98, 108
stochastic, 109
store, 7, 59, 98
systematic error, 6, 109
tetramer, 108
time course, 100
time increment, 6, 46, 51
time independent, 109
Todd, Peter, 110
user_suffix, 32
virial coefficient, 9, 15
wavelengths, 10
WDA, 6, 34, 78
weighting factor, 50, 91
Weighting Factors, 51
Wide Distribution Analysis, 34, 82
Write adjusted data file, 32
Yphantis, 38, 98
Zoom, 31, 79