Application Information

sIA—Solution Interaction Analysis

α-Chymotrypsin as a Model System to Demonstrate the Use of the Data Simulator in the New Beckman XL-A/XL-I Data Analysis Software

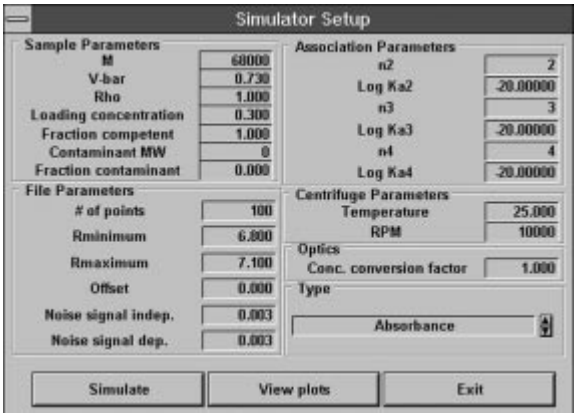
Donald K. McRorie
Beckman Instruments, Inc.

Introduction

The Optima XL-A/XL-I Data Analysis Software from Beckman includes a program in the Utilities section that synthesizes equilibrium data according to parameters set by the user. This article will describe and define the parameters available to the user and the graphics associated with the software. A simple monomer-dimer reversible self-associating system will be used as a model to compare synthesized data with data collected from the analytical ultracen-trifuge. This comparison will show the user how to take advantage of the simulator to design optimal experimental conditions such as initial protein concentrations and rotor speeds according to the characteristics of the association process. Chymotrypsin at pH 4.0 was chosen for this system since it has been well characterized in the literature[1] and repeat analysis shows consistent results with the reported association constants.[2] The simulator will be used to generate data with various contaminants or amounts of monomer without the ability to self-associate (also known as incompetent monomer) to demonstrate the differences in the simulated data from the "ideal" system. Also, it will be apparent how the simulator can be used to set up better experimental conditions for accuracy in modeling data.

Simulator Software Description

In the first screen, the parameters are divided into sections as shown in the following figure:



BECKMAN

Sample parameters describe the solute composition and characteristics of the solute and solvent:

M–molar mass of the monomer in Daltons.

V-bar–partial specific volume of the major solute and contaminant in L/g, independent of the degree of association.

Rho–density of the solvent in g/L.

Loading concentration–loading concentration of the major solute in g/L.

Fraction competent–weight fraction of major solute competent to self-associate.

Contaminant MW–molar mass of contaminant in Daltons.

Fraction contaminant–weight fraction of loading concentration of solute that is contaminant.

File parameters set the data limitations:

# of points–number of data points to used in the calculated array (25-500).

Rminimum–meniscus radius in cm.

Rmaximum–base radius in cm.

Offset–unsedimentable baseline.

Noise, signal-independent–pseudo-noise used to calculate standard deviation independent of solute concentration.

Noise, signal-dependent–pseudo-noise used to calculate standard deviation directly proportional to solute concentration.

Note: the total standard deviation will be the sum of the signal-independent and signal-dependent noise according to the equation: sigma = (signal-independent noise) + (signal-dependent noise)*(total calculated signal).

Association parameters describe self-association between molecules. All associations are monomer-n-mer associations with whole numbers for the stoichiometry and $\log K_a$ for the corresponding association constant. The fact that these are labeled 2, 3, and 4 does not limit the user to these stoichiometries. A $\log K_a$ of -20 will effectively remove an aggregate from the model due to the low concentration relative to monomer:

n_2, n_3, n_4 - stoichiometries of the n2-mer, n3-mer, and n4-mer, respectively.

$\log K_2$, $\log K_3$, $\log K_4$ - base 10 logarithms of the association constants for formation of n2-mer, n3-mer, and n4-mer from monomer in inverse molar units.

Centrifuge parameters are run conditions of the XL-A/XL-I:

Temperature–run temperature in degrees Celsius.

RPM–rotor speed in rpm.

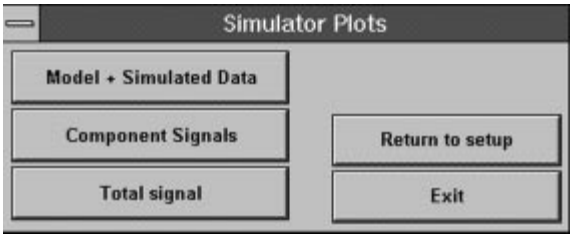
Optics has a concentration conversion factor used to translate signal to weight concentration (assumed equal for major solute and contaminant or if unequal for contaminant, fraction contaminant can be adjusted appropriately). The conversion factor sets a ratio between signal and weight concentration using the extinction coefficient (absorbance) or specific refractive increment (interference) and the necessary instrument variables such as pathlength:

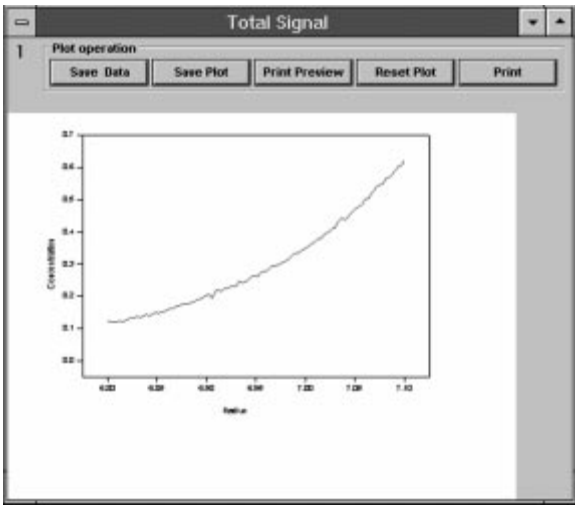
For absorbance gradient, signal = (conversion factor) * (total solute concentration(g/L)).

For interference gradient, signal = (conversion factor) * (total solute concentration(g/L) -total solute concentration at meniscus (g/L)).

Type allows the user to select between absorbance and interference optics by clicking on the up and down arrows of the spinner to the right of the box.

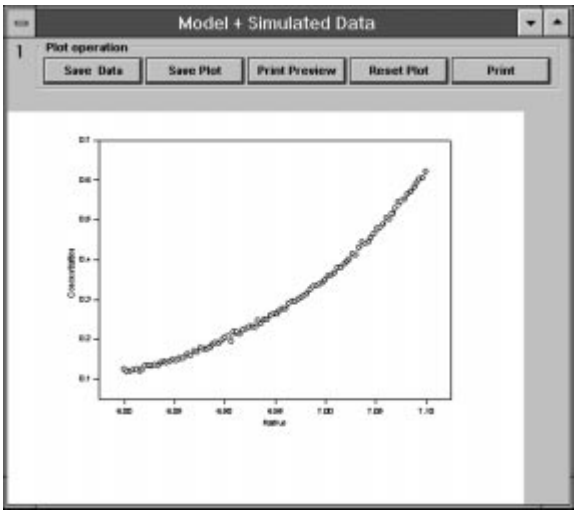
After parameters are set to the user's specification, the SIMULATE button is clicked with the mouse to calculate the gradient. The simulated data are then plotted as a line plot showing the total signal as a function of radius. Several other options are available to the user at this juncture:



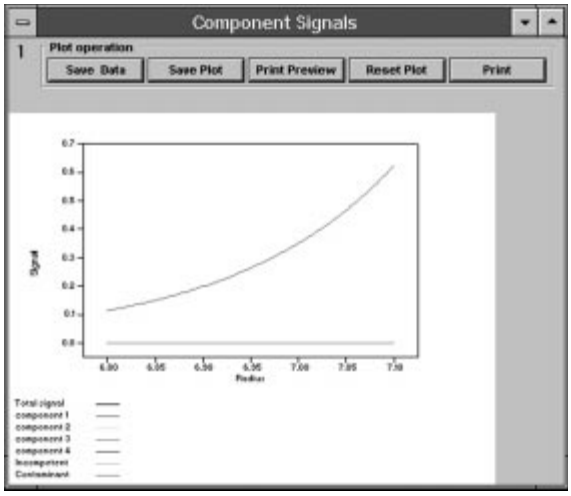


All plots have the same option buttons. SAVE DATA will create an ASCII file with data and a header with run conditions like data from the analytical ultracentrifuge. SAVE PLOT will save the plot shown in Origin format (*.ogg) so that it can be re-

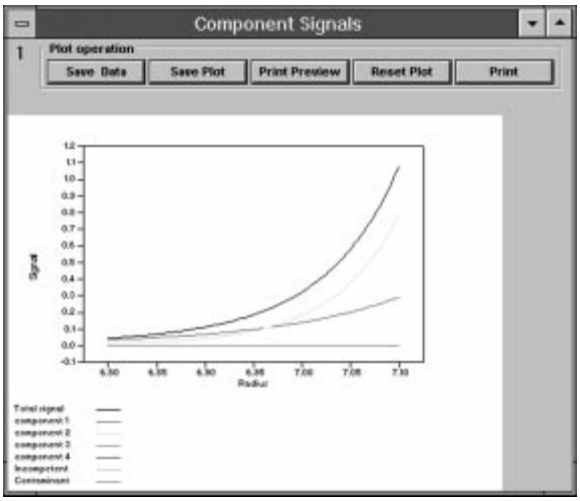
called from within the Origin program as seen on the screen with or without user modifications. PRINT PREVIEW is to show the user the print font and RESET PLOT returns to the larger, more readable screen font. PRINT will print out the plot as seen in print preview. Another plot available to the user is MODEL + SIMULATED DATA where the simulated data are shown as a scatter plot around a line plot showing data with the same parameters and no noise function added:



The final plot shows the component signals plus a total signal. In the case of a single species as demonstrated here, component 1 overlaps the total signal and the two are indistinguishable:



In this plot, however, if there is an association constant set for a self-associating monomer-dimer system, the plots of two individual component signals are seen along with the total signal:



Up to four components of a self-association can be shown as well as any monomer incompetent to associate and any contaminants described by the parameters in the setup window. To remove plots of any of the species shown, *e.g.*, the components with no concentration contributing to the total signal, double click the mouse on the "1" in the upper left of the screen and transfer the data to the window not to be plotted (see Origin manual for more detail).

Model System

α-Chymotrypsin (lot # 37K093) purchased from Worthington Biochemical Corp. was evaluated by sedimentation equilibrium using the Optima XL-I analytical ultracentrifuge equipped with both absorbance and interference optics. The enzyme (85% protein), without further purification, was run at three weight concentrations (0.6, 0.4, and 0.2 mg/mL) at 20,000 rpm and 20°C in 10 mM NaOAc, 0.2 M NaCl, pH 4.0, until an equilibrium concentration distribution had been attained. The calculated final protein concentrations of 0.51, 0.34, and 0.17 mg/mL, taking into account 15% non-absorbing material in the sample, were consistent with the absorbance readings before the experiment. Absorbance at 280 nm and interference data were collected on the six-channel centerpiece. For analysis, a buffer density of 1.007 g/L was calculated according to the method of Laue[3] using the program "Sednterp" which is available as shareware.[4] The literature value of 0.736 L/g was used for the partial specific volume.[5] The sedimentation equilibrium data were evaluated using a non-linear least-squares algorithm[6] in the XL-A/XL-I Data Analysis Software from Beckman and results compared with simulated data created with the Beckman software using the fit parameters.

Experimental Results

Chymotrypsin is a well-characterized monomer-dimer species at pH 4.0.[7, 2] Fitting the absorbance and interference data to a monomer-dimer reversible association, floating the variables C_o (the signal concentration at the reference radius), offset (any unsedimentable baseline), and K_{a2} (the association constant of the monomer-dimer equilibrium in reciprocal signal units), gave the results in Figure 1. The resulting K_{a2} value of 4.419 is in A^{-1} units. To convert this value to M^{-1} units, the following equation is used[7-9]:

$$K_{conc.} = K_{abs.} \frac{(\varepsilon \cdot \ell)^{n-1}}{n}$$

$K_{conc.}$ is the association constant in molar concentration terms depending on the stoichiometry, $K_{abs.}$ is the absorbance association constant from the fit, ε is the molar extinction coefficient, ℓ is the pathlength of the centerpiece in cm, and n is the stoichiometry of the larger association species. Using this equation with the molar extinction coefficient of 44,064 L/mol cm, pathlength of 1.2 cm, and stoichiometry of 2, the association constant is 116,831 M^{-1}. For interference, the resulting association constant from the best fit nonlinear regression was 5.344 $fringe^{-1}$ (data not shown). The conversion equation here is different since the fringe displacement is proportional to weight concentration:

$$K_{conc.} = K_{fringe}\left(\frac{\frac{dn}{dc} \cdot \ell}{\lambda}\right)^{n-1} \frac{(M_1)^{n-1}}{n}$$

$K_{conc.}$ is the association constant in molar concentration terms, K_{fringe} is the signal association constant, dn/dc is the specific refractive increment (L/g), ℓ is the centerpiece pathlength (cm), λ is the lightsource wavelength (cm), M_1 is the monomer molecular weight and n the maximum stoichiometry of the association. Most proteins have a consistent value of 3.31 fringes/g/L ($dn/dc * \ell/\lambda$) for the XL-I assuming a wavelength of 675 nm for the laser diode, a 1.2 cm pathlength, and a specific refractive increment of 0.186 mL/g. Some lasers will vary up to 5 nm and specific refractive increment can vary for solute depending on modifications such as glycosylation. In these cases the appropriate modifications to the calculation would have to be made. For chymotrypsin, the K_{fringe} converts to a $K_{conc.}$ of 191,037 M^{-1}. While this value may be different from the absorbance data, the user should keep in mind it is the log K_a that is the thermodynamically significant term. The log K_{a2} for the absorbance and interference data are 5.086 and 5.281, respectively. This difference is only about 4%.

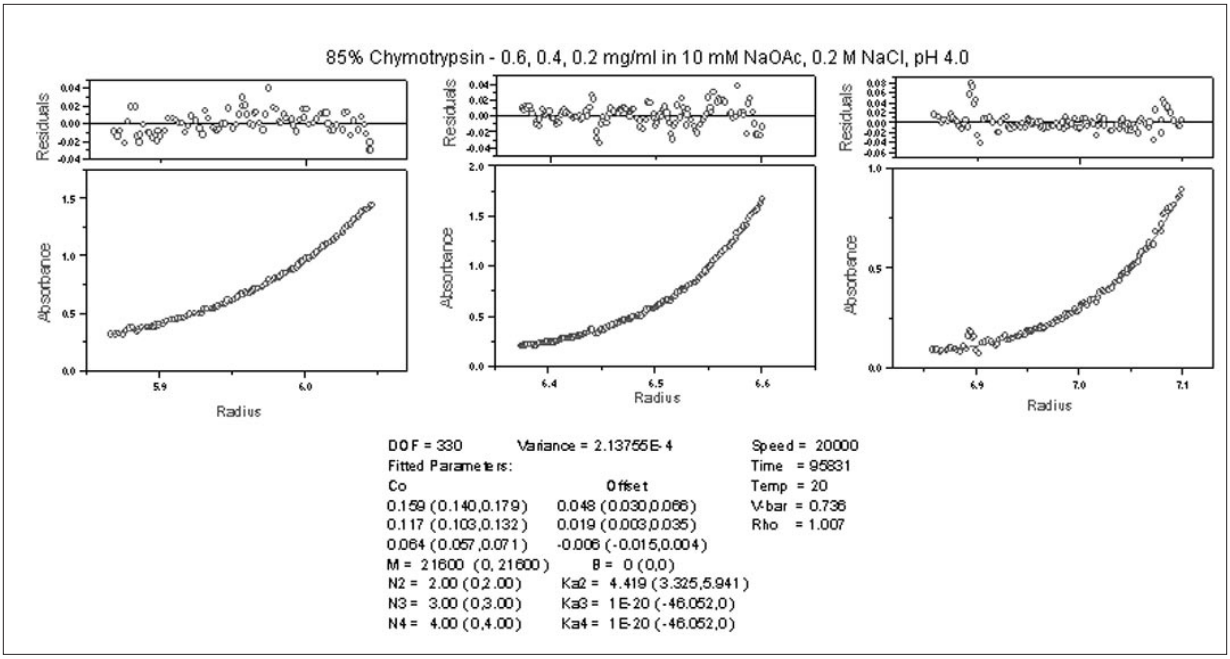
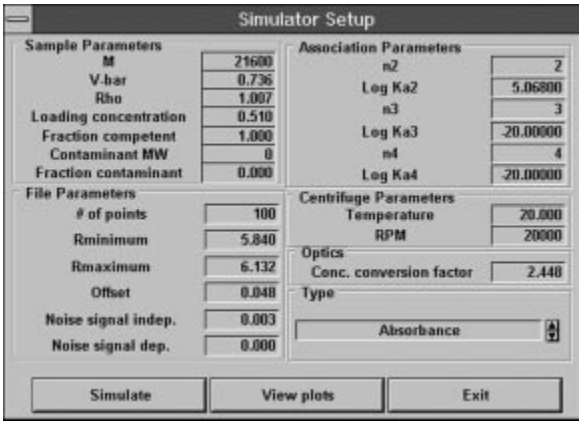


Figure 1. *Results of three concentrations fit simultaneously to a monomer-dimer equilibrium with the variables of C_o, offset, and K_{a2}. The monomer molecular weight was fixed at 21,600 Da, with a partial specific volume of 0.736 L/g and solution density of 1.007 g/L. Confidence intervals are calculated at 95% for varied parameters.*

Synthesized Results

The simulator will now be used to generate data matching the best fit parameters from the absorbance data to demonstrate its use in setting up experimental conditions. The starting absorbances used were determined at the beginning of the experiment and agree well with that expected for 85% chymotrypsin at 0.6, 0.4, and 0.2 mg/mL (1.248, 0.832, and 0.416 O.D. corresponding to 0.51, 0.34 and 0.17 mg/mL protein). The expected values would be 1.25, 0.82 and 0.40 using an $E^{0.1\%}$ of 2.04 L/g-cm and pathlength of 1.2 cm. An example of the setup window for the chymotrypsin data is shown in the following figure:



The monomer molar mass (M) for chymotrypsin is 21,600 Da, partial specific volume (V-bar) is 0.736 L/g, and solvent density (Rho) is 1.007 g/L. The loading concentration for the first file is 0.51 g/L. The minimum and maximum radii were determined from the fringe display screen from the meniscus to the bottom of the cell at the rotor speed for data acquisition. Offset and log K_{a2} were used from the fit analysis (Figure 1). The K_{a2} is in M^{-1} units. Centrifuge parameters are 20°C and 20,000 rpm. The concentration conversion factor converts solute concentration (g/L) to signal units. For absorbance:

signal = conc. conversion factor * total solute concentration (g/L) = $E^{0.1\%} * \ell *$ total solute concentration (g/L)

For chymotrypsin, $E^{0.1\%}$ is 2.04 L/g-cm and, for a 1.2 cm pathlength, the concentration conversion factor is 2.448 L/g. For interference:

signal = conc. conversion factor * (total solute concentration - total concentration at meniscus) = (dn/dc * ℓ / λ) * (total solute concentration - total solute concentration at meniscus)

For chymotrypsin, the specific refractive increment (dn/dc) is 0.187 mL/g, ℓ is 1.2 cm, and the wavelength 675 nm, resulting in a concentration conversion factor of 3.32 fringes/g/L with a depleted meniscus, essentially equal to the 3.31 fringes/g/L expected for unmodified proteins.

Simulated data for the three protein concentrations, 0.51, 0.34, and 0.17 g/L, created as described in the setup screen, are shown in Figure 2 overlaid with the actual data from the experiment. The differences between the simulated data and the experimental data can arise from several factors. Part of the difference could be from the 15% non-protein material in the sample. The large number of variables, including C_o, offset, K_{a2}, concentration conversion factor, and selected radii, as well as the algorithm simulating the data, can give slight errors from the experimental data if any of these parameters are not accurate. This is especially the case in selecting accurate extinction coefficients and cell radii. A good test is to check the differences in a fit analysis between the simulated data and the experimental data. In this case, the simulated chymo-

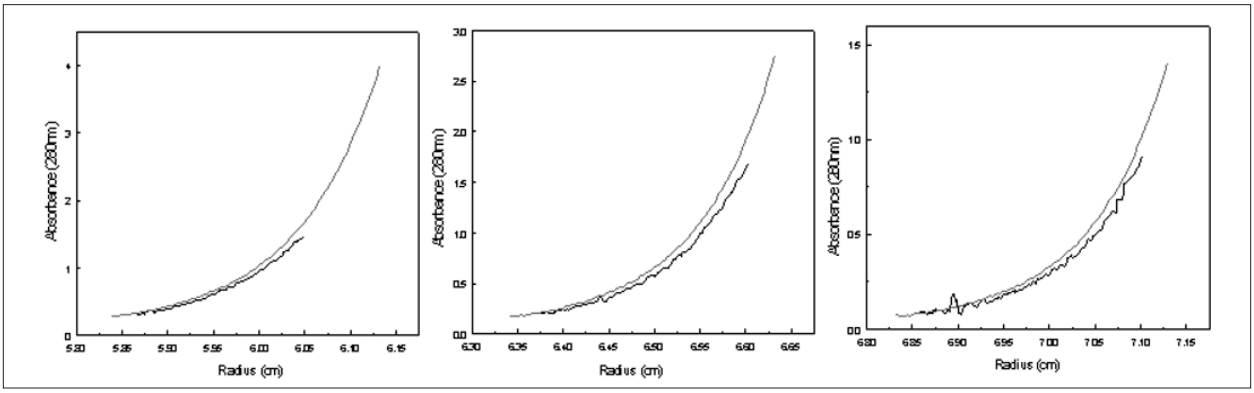


Figure 2. *Simulated data for three concentrations (0.51, 0.34, and 0.17 g/L) of simulated chymotrypsin data overlaid with experimental data. The simulated data have a greater radial range, showing the higher absorbance values not obtainable experimentally due to stray light.*

trypsin data showed less than 5% difference from the experimental data on all fitted parameters. These small differences will have no effect on experimental conditions selected on the basis of the simulator.

The component graphs in the simulator can be used to examine experimental conditions and the effect on the equilibrium distribution of the components to help the user optimize the conditions for modeling data in an analysis. For example, since the model fits to the monomer concentration at the reference radius and by default this is the first data point, a significant concentration of monomer should be at this radial position. Also, speed and concentration should significantly change the monomer to n-mer concentration ratios across the cell. With the chymotrypsin data, Figure 3 shows differences in the ratios resulting from simply changing the concentration. These graphs show that for a single speed, 20,000 rpm was a good choice due to the variability of the component concentrations. Also, a significant concentration of monomer remains at the first data point (the default for the reference radius) in all plots for the analysis.

The user could then use the simulator to show similar effects as a function of speed to help select optimum speed variations in the experiment (Figure 4). The data show that for the 0.51 mg/mL concentration, a speed of 18,000-20,000 rpm would be maximum before losing a significant portion of the data due to excessive absorbance (2.0-2.5 O.D. at 280 nm). At these higher absorbances, the signal is not proportional to concentration. Examining the data in the overlay plot (Figure 2) showed that the

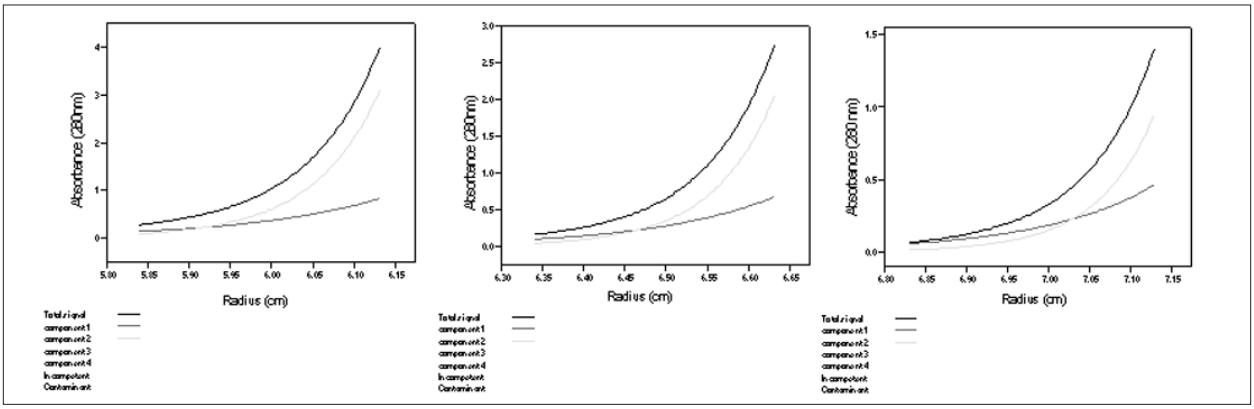


Figure 3*. Component plots for three concentrations (0.51, 0.34, and 0.17 mg/mL) of chymotrypsin from the simulator. Shown are the monomer (component 1), dimer (component 2), and total absorbance values using the parameters from the experimental data and radii from the fringe display screen. Two other components from the self-association as well as monomer that cannot associate and contaminant could also be plotted if included in the simulation.*

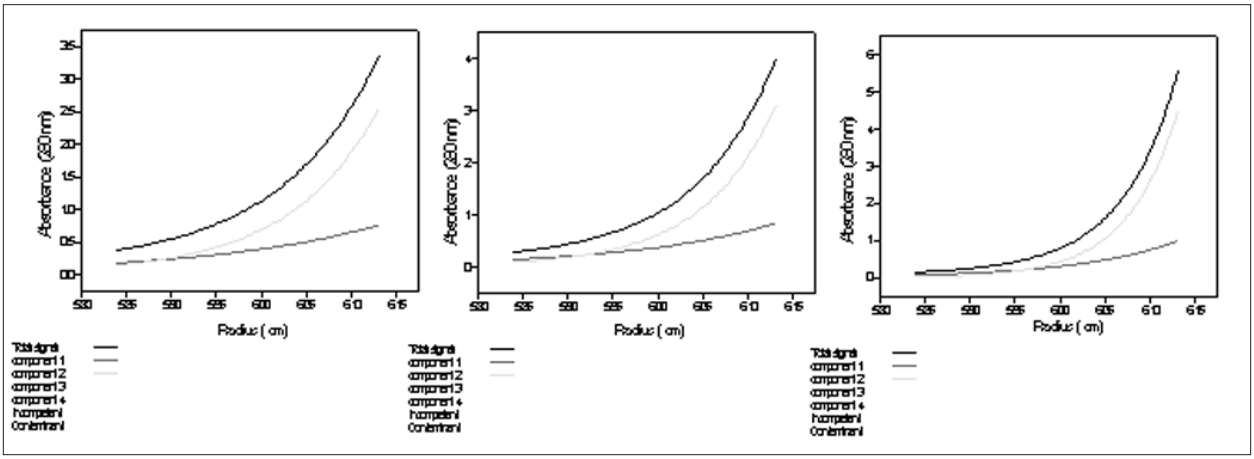


Figure 4. *Component distribution of 0.51 mg/mL chymotrypsin at 18,000, 20,000, and 24,000 rpm from the simulator. Shown are the monomer (component 1), dimer (component 2), and total absorbance values.*

experimental data had to be limited to a much smaller radial range due to this problem at the higher absorbances for the 0.51 mg/mL sample. Also, the meniscus became depleted of monomer at the higher speeds, prompting the user to select a different reference radius in the data analysis. The higher concentration would not be a problem with the interference optics and in some cases the user may choose to deplete the meniscus of one or more of the components.

Concentration and speed are the two variables that most affect the simulated data. The presence of incompetent monomer or contaminant will not alter the total concentration unless either is present in very high concentration or has a large difference in molecular weight from the major species present (Figure 5).

The major use of the data simulator is to help the user decide on better experimental conditions and parameters for data analysis. But, with the inclusion of the incompetent monomer and contaminant species, it may also be used as an indicator, when analyzing data, of other species present in a self-associating system. Chymotrypsin has been used as a model system to demonstrate component contributions to total concentration with varied experimental conditions such as total loading concentration and rotor speed to demonstrate the use of the simulator for experimental design. In the case of chymotrypsin, it can be demonstrated that the 20,000 rpm is an optimum speed for the range of concentrations used to get a distribution of monomer and dimer. Additional data at other rotor speeds and loading concentrations could also prove helpful if a more complex association is present. Users are encouraged to vary all parameters to see effects on the equilibrium profiles so that they may begin to recognize them when encountered in analyzing data.

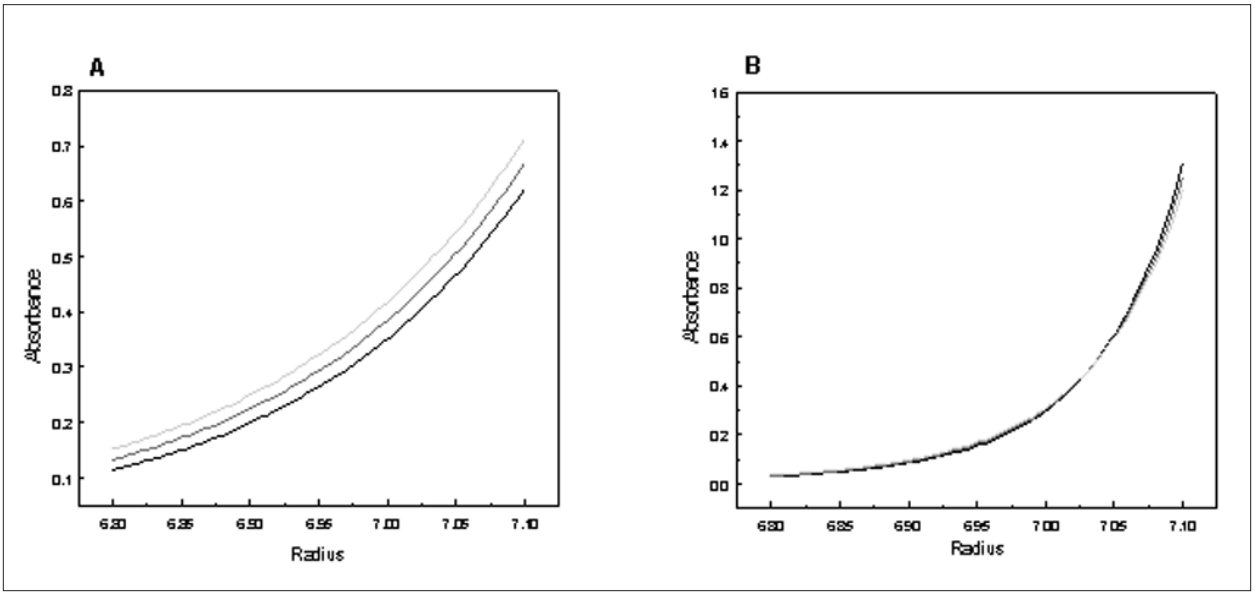


Figure 5. *Model systems comparing 0, 10%, and 20% A) contaminant and B) incompetent monomer. A) A single ideal species of 68,000 Da molar mass is shown with the increasing ratio of contaminant of 34,000 Da molar mass. The increase in concentration of contaminant results in an increase in signal across the radius. B) A monomer-dimer system with monomer molar mass of 34,000 and log K_{a2} is shown with increasing amounts of incompetent monomer. Increasing amounts of incompetent monomer shows opposite effects at the radial extremes. At the inner radii, there is an increase in total absorbance and, at the outer radii, a decrease in total signal with increasing concentrations of incompetent monomer. An inflection point, in this example, is seen at about 7.03 cm.*

References

1. Teller, D. C. Characterization of proteins by sedimentation equilibrium in the analytical ultracentrifuge. *Methods in Enzymology, Vol. 27*, pp. 346-441. Editors-in-chief: S. P. Colowick and N. O. Kaplan. New York: Academic Press, 1973.
2. Voelker, P. and McRorie, D. α–Chymotrypsin: Characterization of a Self-Associating System in the Analytical Ultracentrifuge. *Technical Information Bulletin T-1782A* (1994)
3. Laue, T. M., Shah, B. D., Ridgeway, T. M., Pelltier, S. L. Computer-aided interpretation of analytical sedimentation data for proteins. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, pp. 90-125. Edited by S. E. Harding, A. J. Rowe, and J. C. Horton. Cambridge: Royal Society of Chemistry, 1992.
4. Voelker, P. *Exploration: Techniques and Answers for Solution Interaction Analysis on the Optima™ XL-A*, vol. 2, number 1 (August 1995)
5. *Handbook of Biochemistry, Selected Data for Molecular Biology*, pp. C1-C98. 2nd Ed. Cleveland, OH: Chemical Rubber Co., 1970.
6. Johnson, M. L., Correia, J. J., Yphantis, D. A., Halvorson, H. R. Analysis of data from the analytrical ultracentrifuge by nonlinear least-squares techniques. *Biophys J. 36*, 575-588 (1981)
7. McRorie, D. K., Voelker, P. *Self-Associating Systems in the Analytical Ultracentrifuge*. Fullerton, CA: Beckman Instruments, Inc., 1993.
8. Becerra, S. P., Kumar, A., Lewis, M. S., Widen, S. G., Abbotts, J., Karawya, E. M., Huges, S. H., Shiloach, J., Wilson, S. H. Protein-protein interactions of HIV-1 reverse transcriptase: implication of central and C-terminal regions in subunit binding. Appendix: Ultracentrifugal analysis of mixed associations, by M. S. Lewis. *Biochemistry 30*, 11707-11719 (1991)
9. Ross, P. D., Howard, F. B., Lewis, M. S. Thermodynamics of antiparallel hairpin-double helix equilibria in DNA oligonucleotides from equilibrium ultracentrifugation. *Biochemistry 30*, 6269-6275 (1991)

Beckman Instruments, Inc. • 2500 N. Harbor Boulevard, P.O. Box 3100 • Fullerton, California 92834-3100
Sales: 1-800-742-2345 • Service: 1-800-551-1150 • Internet: http://www.beckman.com • Telex: 678413 • Fax: 1-800-643-4366

Worldwide Offices: Africa, Middle East, Eastern Europe (Switzerland) (41) 22 994 07 07. **Australia** (61) 2 9844 6000. **Austria** (43) 1 7292164. **Canada** (905) 819-1234. **China** (8610) 6527 9344-9. **France** (33) 1 43 01 70 00. **Germany** (49) 89-35 870-0. **Hong Kong, PRC** (852) 2814 7431. **Italy** (39) 2-953921. **Japan** (03) 5352-2820. **Mexico** (525) 559-1635. **Netherlands** 0297-230630. **Poland** (48) 22 408822, 408833. **Singapore** (65) 339 3633. **South Africa** (27) 11-805-2014/5. **Spain** (1) 358-00-51. **Sweden** (08) 98-53-20. **Switzerland** (41) 22 994 07 46. **Taiwan** (886) 02 378-3456. **U.K.** (01494) 441181. **U.S.A.** 1-800-742-2345.

SBU9/97131-CB-10

Printed in U.S.A. on recycled paper.