Published in final edited form as:

Methods. 2011 May ; 54(1): 39–55. doi:10.1016/j.ymeth.2011.01.002.

Evaluating the stoichiometry of macromolecular complexes using multisignal sedimentation velocity

Shae B. Padrick and Chad A. Brautigam*

Department of Biochemistry, The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd. Dallas, TX 75390-8816

Abstract

Gleaning information regarding the molecular physiology of macromolecular complexes requires knowledge of their component stoichiometries. In this work, a relatively new means of analyzing sedimentation velocity (SV) data from the analytical ultracentrifuge is examined in detail. The method depends on collecting concentration profile data simultaneously using multiple signals, like Rayleigh interferometry and UV spectrophotometry. If the cosedimenting components of a complex are spectrally distinguishable, continuous sedimentation-coefficient distributions specific for each component can be calculated to reveal the molar ratio of the complex's components. When combined with the hydrodynamic information available from the SV data, a stoichiometry can be derived. Herein, the spectral properties of sedimenting species are systematically explored to arrive at a predictive test for whether a set of macromolecules can be spectrally resolved in a multisignal SV (MSSV) experiment. Also, a graphical means of experimental design and criteria to judge the success of the spectral discrimination in MSSV are introduced. A detailed example of the analysis of MSSV experiments is offered, and the possibility of deriving equilibrium association constants from MSSV analyses is explored. Finally, successful implementations of MSSV are reviewed.

Keywords

analytical ultracentrifugation; biophysical methods; stoichiometry; multisignal sedimentation velocity; Arp2/3 complex; step by step instructions

1 Introduction

Knowledge of the stoichiometry of the components of macromolecular assemblies is fundamental to understanding their physiology. A host of methods to determine this quantity is available to the modern researcher. Among these methodologies are electron microscopy, X-ray crystallography, NMR spectroscopy, calorimetry, and light scattering. Analytical ultracentrifugation (AUC) is naturally applied to this problem because the data report

*to whom correspondence should be addressed: chad.brautigam@utsouthwestern.edu, phone: +1-214-645-6384, fax: +1-214-645-5383.

²Experience has shown that it is best not to have substantial gaps or overlaps in molar mass- (M -) space when calculating segmented distributions. Given the significantly different f_r 's for the two segments, a 1.5-S gap in s -space was needed to ensure the proper coverage of M -space. As the analysis below demonstrates, there is no deleterious effect consequent to this gap.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

directly on the molar mass (sedimentation equilibrium, SE) or size and shape (sedimentation velocity, SV) of the macromolecular complex.

Five years ago, Schuck and others introduced a new treatment for SV data obtained from multiple signals [1]. These signals can be different wavelengths obtained from the absorbance optical system of the ultracentrifuge and/or interferometric data obtained from the on-board Rayleigh interferometer. Usually, these signals are obtained simultaneously during a single SV experiment. Here, we term this methodology Multi-Signal SV (MSSV). The analysis of MSSV data decomposes the standard $c(s)$ distribution into component distributions called $c_k(s)$, where the k denotes an individual component present in solution. Because individual components can be detected using MSSV, it is possible to analyze the populations of 2–4 cosedimenting components. This method is therefore useful for the analysis of non-interacting species that exhibit little or no hydrodynamic resolution [1]. However, MSSV has found most utility in the analysis of cosedimenting interacting species. Examination of the relative populations of such species, coupled with the complex's hydrodynamic properties, allows the experimenter to determine the complex stoichiometry. This quantity is vital for understanding the functioning of the macromolecular complex, and it is a primary focus of many other analytic methods, e.g. isothermal titration calorimetry.

In this paper, we detail the experimental aspects of the MSSV method. After an introduction to the theoretical basis for the method, we then explore some critical experimental prerequisites that should be met before the experimenter performs an MSSV study. Among these considerations are empirical assessments of sample quality and suitability. Also, we develop pre-analysis processes designed to answer the questions (1) can MSSV work for my macromolecules? and (2) what concentrations of components should I use to maximize the probability of success? We also introduce a binding isotherm that is uniquely available from the MSSV analysis. A detailed description of the analysis of two interacting proteins is presented. Finally, we discuss the implications of the present work and the significant biological impact that MSSV analysis has made thus far.

2 Methods

2.1 Protein Methods

The fusion of glutathione-S-transferase and the VCA domain of WASP (GST-VCA) was constructed and purified as detailed in reference [2]. Arp2/3 complex was isolated from bovine thymus as described [2]. The proteins were dialyzed against a buffer comprising 50 mM KCl, 10 mM imidazole, 1 mM EGTA (Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid), and 1 mM MgCl_2 . The samples were placed in an assembled ultracentrifugation cell that had a charcoal-filled Epon, 1.2-cm-path-length, two-sectored centerpiece sandwiched between sapphire windows. Three samples were placed in an An-60Ti rotor and were centrifuged at 42,000 rpm in an Optima XL-I ultracentrifuge (Beckman-Coulter, Fullerton, CA): GST-VCA alone, Arp2/3 alone, and a mixture of the two proteins. Data were acquired using the absorbance optics (280 nm) and the interference optics.

2.2 Data analysis

All data were analyzed using SEDPHAT version 8.2. Because the refinement of parameters can be dominated by data sets with significantly more data points, a new feature of SEDPHAT was utilized that compensates for this imbalance. This is the “*sqrt(N1/Nx)” checkbox in the Experimental Parameter dialog box. This feature was activated in all analyses presented in this work (except where noted), and a more detailed explanation of its basis and workings is found in [3].

2.3 Simulation of MSSV data for section 4

Data were simulated in SEDFIT and SEDPHAT. First, a mock data set with the proper time and radial resolution was produced using the generate function in SEDFIT. Default values were used for the analysis in SEDFIT, except the meniscus was set to 6 cm and radial resolutions of 0.003 cm and 0.00072 cm were used for absorbance and IF data, respectively. Except where noted, time resolution was one trace per five minutes. Rotor acceleration was modeled using the default settings. Next, the SV data set was created in SEDPHAT. The mock data set was loaded into SEDPHAT, the meniscus was fixed at 6 cm and the sample bottom at 7.2 cm, and baseline fitting was deactivated. For all steps of the simulation, default values for pathlength, (1.2 cm), $\bar{v}$ (0.73 ml/g), buffer density (0.998230 g/ml), buffer viscosity (0.010020 P), and temperature (20°C) were used. The A + B $\leftrightarrow$ AB model was selected. Parameters were entered as follows: Component A: $[A]_{\text{tot}} = 1.9 \mu\text{M}$, molar mass = 200,000 g/mol, $s_A = 8.5 \text{ S}$, $\epsilon_{\text{ABS}}^A = 140,845 \text{ M}^{-1} \cdot \text{cm}^{-1}$ or $\epsilon_{\text{IF}}^A = 550,000 \text{ fringes} \cdot \text{M}^{-1} \cdot \text{cm}^{-1}$, depending on whether absorbance or interference data was being modeled. Component B: $[B]_{\text{tot}} = 4.2 \mu\text{M}$, molar mass = 10,000 g/mol, $s_B = 1.2 \text{ S}$, $\epsilon_{\text{IF}}^B = 27,500 \text{ fringes} \cdot \text{M}^{-1} \cdot \text{cm}^{-1}$ when interference data was being modeled, and ϵ_{ABS}^B was set to 25000, 20000, 16000, 14000, 12000, 11000, 10000, 9000, 8000, 7700, 7500, or 7300 $\text{M}^{-1} \text{ cm}^{-1}$, with the value of ϵ_{ABS}^B varied to generate the range of D_{norm} (see Eq. 9) seen in Figure 2A. The complex AB parameters: $s_{\text{AB}} = 9.2 \text{ S}$, $\log(K_A) = 9$ and $\log(k_{\text{off}}) = -3$. Conservation of mass was switched on. Synthetic noise was added to the simulated SV data and the data was saved. This was performed for each signal used. Finally, the multiple-signal data sets were analyzed a new SEDPHAT session, following steps analogous to steps 20–26 of the Supplemental Protocol (see section 5.1). Absorbance signals were given a noise value of 0.0049 to account for this different number of data points in the IF and absorbance data sets, simulated IF data noise level was left at the default, 0.01 value. The multisignal sedimentation velocity model was chosen, and Marquardt-Levenberg fitting algorithm was chosen. The global analysis was set up with the known signal increments (from the design of different D_{norm} 's), so there was no need to optimize these. Two regions of s -space were defined, one for free excess components, the other for the complex. In each, a spectrum (i.e. a $c_k(s)$ distribution) for each component was calculated. The frictional ratio (f_r) was optimized in each. Initially, a single global run was performed to optimize linear parameters, then a single round of fitting was performed, optimizing only the frictional ratios. Unlike in steps 20–26 of the example, the menisci, TI and RI noise were not fit. Instead the cell bottom and menisci were input from the known parameters from the modeling step (7.2 and 6 cm, respectively). Concentrations of different species sedimenting at different rates were quantified by integrating a range around relevant peaks.

Three-component/three-signal systems were modeled similarly. For simplicity, the simulated SV data for the component system was modeled using the A + B $\leftrightarrow$ AB system described above, with $\log(K_A) = 9$ and $\log(k_{\text{off}}) = -3$. In this case, “B” in the SV data generation is a tight complex of components B and C (as analyzed in the second step of Section 4.1). Sedimentation coefficients were chosen to be: System (#1): $s_A = 8.5 \text{ S}$, $s_B = 2$, $s_{\text{AB}} = 9.2$; System (#2): $s_A = 3.5 \text{ S}$, $s_B = 4.0 \text{ S}$, $s_{\text{AB}} = 6.0 \text{ S}$; System (#3): $s_A = 3.5 \text{ S}$, $s_B = 4.0 \text{ S}$, $s_{\text{AB}} = 6.5 \text{ S}$; System (#4): $s_A = 3.5 \text{ S}$, $s_B = 4.0 \text{ S}$, $s_{\text{AB}} = 6.0 \text{ S}$; System (#5): $s_A = 8.5 \text{ S}$, $s_B = 3.5 \text{ S}$, $s_{\text{AB}} = 9.5 \text{ S}$. Relevant masses and extinction coefficients are given in Table 1, calculated total concentrations (by summing the quantities in the two peaks) are given in Table 2. MSSV analysis was performed as described for the two-signal, two-component system, but using three wavelengths.

3 Theory

3.1 The $c(s)$ distribution

This work is primarily concerned with the modeling of SV data. The concentration profiles obtained over time in an SV experiment may be directly modeled as the integral (or discretized sum) of solutions to the Lamm Equation (LE) scaled by a continuous distribution called $c(s)$ [4,5]. If $a(r,t)$ represents the data acquired from an SV experiment, then

$$a(r, t) \cong \int_{s_{\min}}^{s_{\max}} c(s) \chi(s, D(s), r, t) ds, \quad \text{Equation 1}$$

where s is the sedimentation coefficient and $\chi(s, D(s), r, t)$ is a LE solution that is dependent on $D(s)$, the corresponding diffusion coefficient, r , radius from the center of rotation, and t , the time from the beginning of the experiment. Several important features of this type of analysis must be pointed out. First, the analysis directly fits the SV data, i.e. no modifications to the data like pairwise subtractions of scans are needed. Aiding this is the ability to accurately model the noise structure of the concentration profiles [6]. Another notable feature of this analysis is that the LE solutions are for ideal, non-interacting species. Despite this fact, the diffusional deconvolution afforded by the $c(s)$ analysis allows the accurate description of the mass-transport properties [7,8] and apparent diffusion coefficients for reaction boundaries [9] for interacting systems. Usually, a single f_r is assumed over the entire distribution. However, the distribution can be divided into segments, with each segment having its own f_r . This approach is useful in cases where two or more boundaries are easily identified in the data (as in section 5.1.3, below) [10]. In addition, many frictional ratios can be considered in an analysis that explicitly considers both the s - and f_r -dimensions; this approach results in the so-called “size-and-shape” distribution [11].

3.2 $c_k(s)$ distributions derived from multiple signals

Of course, the amplitude of the measured signal in an SV experiment is dependent on the concentration of the macromolecule and its molar signal increment (or molar extinction coefficient, as appropriate), signified as ε_{λ}^k throughout this text, where λ is the wavelength or signal source of the detection, and k denotes the identity of the component. The Beckman XL-I centrifuge has an on-board spectrophotometer (the “absorbance optics”) and a Rayleigh interferometer (the “interference optics”). The centrifuge’s control software may acquire up to four signals (three absorbance and one interference) for a given experiment. Where the signal is derived from the absorbance optics, λ shall be designated as “ABS x ”, where x is the wavelength of detection in units of nanometers. Where the signal comes from the interference optics, the λ shall be given as “IF”. With this notation in hand, the $c(s)$ notation introduced above may be generalized as:

$$a_{\lambda}(r, t) \cong \sum_{k=1}^K \varepsilon_{\lambda}^k l \int_{s_{\min}}^{s_{\max}} c_k(s) \chi(s, D_k(s), r, t) ds, \quad \text{Equation 2}$$

where $a_{\lambda}(r,t)$ is the signal measured at a given path-length l , K is the number of solutes present, and $c_k(s)$ is a continuous distribution accounting only for component k . Thus, if multiple signals are collected, then multiple $c_k(s)$ distributions may be calculated for the same s -range. Where two solutes cosediment, the molar ratio of the complex may be obtained by integrating the area beneath the $c_k(s)$ peaks in the pertinent s -range. By taking into consideration the hydrodynamic properties of the complex and the molar ratio, a stoichiometry may be derived.

Constraints on the stoichiometry of components may be imposed in order to analyze the data. For example, the number of components k in complex κ can be designated S_{κ}^k . The new signal increment $\varepsilon_{\lambda}^{\kappa}$ is thus defined:

$$\varepsilon_{\lambda}^{\kappa} = \sum_k S_{\kappa}^k \varepsilon_{\lambda}^k. \quad \text{Equation 3}$$

Eq. 2 then becomes

$$a_{\lambda}(r, t) \cong \sum_j \sum_{\kappa_j, k} \varepsilon_{\lambda}^{\kappa_j} l \int_{s_{\min, j}}^{s_{\max, j}} c_{\kappa}^{(j)}(s) \chi(s, D_{\kappa_j}(s), r, t) ds, \quad \text{Equation 4}$$

where the distribution is divided into j segments, with each segment possibly reporting on a different stoichiometry.

3.3 Goodness-of-fit statistics in SEDPHAT

The goodness-of-fit statistic used by SEDPHAT is the global reduced χ^2 (χ_r^2). Once the fit has converged, the χ_r^2 is designated χ_b^2 (i.e. the “best” fit). For the purposes of finding error intervals or for checking the statistical validity of an applied constraint, it is common to change a fitted parameter (or apply a constraint) and to monitor the change in the quality of the fit. Previously, we have defined two criteria to judge the change [3]. If the perturbation worsens the quality of the fit by $> 1 \sigma$, then the change is deemed statistically distinguishable from the best fit. If the change worsens the fit by 2σ , it is rejected as being unlikely to be valid. The χ_r^2 values obtained during these fitting sessions are called “test” χ^2 ’s, or χ_t^2 . Thus, it is necessary to define the two “critical” values of χ^2 that lead to the two conclusions introduced above. SEDPHAT has a built-in F-statistics calculator that determines critical χ^2 values using the formula

$$\chi_{c, n\sigma}^2 = \chi_b^2 F_{\mu, \nu}^{\alpha}, \quad \text{Equation 5}$$

where n is either 1 or 2, depending on whether the $1(\sigma)$ or $2(\sigma)$ value is desired, respectively, and $F_{\mu, \nu}^{\alpha}$ is the $(1-\alpha)$ one-sided F statistic with $\alpha = 0.683$ ($n = 1$) or $\alpha = 0.95$ ($n = 2$) and $\mu = \nu$ = degrees of freedom = number of analyzed data points.

3.4 An MSSV-based population isotherm

The MSSV population isotherm (see section 5.2) was fitted using the curve-fitting routines available in SigmaPlot version 11. In short, a protein (A) interacting with a small-molecule ligand (B) forming a 1:1 AB complex was simulated using the parameters discussed below. An analytical expression for [AB] given the known total concentrations of A ($[A]_{\text{tot}}$) and B ($[B]_{\text{tot}}$) was fitted:

$$[\text{AB}] = \frac{\left([A]_{\text{tot}} + [B]_{\text{tot}} + \frac{1}{K_A}\right) - \sqrt{\left(-[A]_{\text{tot}} - [B]_{\text{tot}} - \frac{1}{K_A}\right)^2 - 4[A]_{\text{tot}}[B]_{\text{tot}}}}{2} \quad \text{Equation 6}$$

where K_A is the association constant of AB interaction. The Marquardt-Levenberg fitting algorithm was used to fit this nonlinear equation. In the simulated titration, $[A]_{\text{tot}}$ was invariant; thus, its value was supplied to the algorithm as a constant.

4 Design and evaluation of MSSV experiments

In designing MSSV experiments several parameters must be simultaneously considered. The first is the choice of signals to use in the multisignal experiment. The choice of signals will directly influence the ability to reliably decompose the data into component distributions. In section 4.1, we discuss how to assess the likelihood of success from parameters estimable outside of the centrifuge. The second challenge of MSSV experimental design is balancing the concentrations of all the species added to optimize signal quality. This requires balancing of several factors simultaneously, and can be done graphically as described in section 4.2. Finally, a significant effort should be put into assessing the quality of any MSSV analysis. In many cases, simple goodness-of-fit (i.e. χ^2 -based) post-analysis tests completely fail to assess the quality of the analysis. In section 4.3 we summarize the quality control checks that should be applied to MSSV analysis.

4.1 Can the signals be decomposed into $c_k(s)$ distributions?

Although performed as a global analysis, the multisignal method relies on the observed signal intensities arising from linear combinations of the components' signals. It should be noted that this assumes that there be no significant hypo- or hyperchromicity. This assumption can be tested by comparing the absorbance of a mixture of the components to the sum of the absorbances of the individual components. Within this constraint, a series of concentrations and a series of signal increments uniquely determine the expected signals, following a Beer-Lambert relation (and similar relation when using interference optics). When there are as many signals as components, this process is typically invertible, allowing concentrations of all components to be deduced from the signal intensities using Eq. 7.

$$lE\vec{C}=\vec{O} \quad \text{Equation 7}$$

$$l * \begin{bmatrix} \epsilon_{IF}^A & \epsilon_{IF}^B \\ \epsilon_{ABS}^A & \epsilon_{ABS}^B \end{bmatrix} * \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \Delta J \\ OD_{ABS} \end{bmatrix} \quad \text{Equation 8}$$

In Eq. 7, the concentrations of a series of components are collected into the vector C , the extinction coefficients for the components are the elements of the matrix E (which is scaled by the path-length l), and the resulting optical signals are in the vector O . This is shown explicitly for the two components and two signals case in Eq. 8, where the signal increments are as described above, A and B are the concentrations of the two species, and ΔJ and OD_{ABS} are the interferometric and absorbance signals, respectively. The matrix E also appears in the "Extinction Box" of the global parameters window for the MSSV analysis in SEDPHAT, e.g. in step 21 of the Supplemental Protocol.

This process is reversible; if both sides of Eq. 7 are divided by l , and then multiplied by the matrix inverse of E (E^{-1}), we find that the optical signals uniquely determine the component concentrations. This assumes E^{-1} can be found, which will generally be true when the determinant of E is not zero. In the absence of noise, (and with an invertible signal increment matrix E), we can uniquely determine the concentrations which correspond to a collection of interference and absorbance signals.

The addition of experimental noise clouds this process. This is illustrated in Figure 1. Eq. 7 was used to find the absorbance at two wavelengths of a 1:1 mixture of two species. Next, normally distributed, synthetic noise was added to the absorbances. The signals were between an optical density of 0.4 and 0.9 in all cases, and the added noise was centered on zero with a standard deviation of 0.003, and thus is less than 1% noise. The resulting absorbances are then decomposed into concentrations by left multiplying by E^{-1} , and a ratio of concentrations is found. The process was repeated 10000 times and the results are shown as a normalized histogram of the A–B ratios. This was performed for two different E matrices. In one case (the dashed curve), small experimental noise leads to small errors on the estimated concentrations. In a second case (the solid curve), small experimental errors lead to large errors in the estimated concentrations. The latter is an example of poor spectral discrimination.

The determinant of the matrix E serves as the basis for our quantification of spectral discrimination, which we term D_{norm} . One way of thinking of the determinant of a two by two matrix is that it gives the area of a parallelogram enclosed by the vectors making up the columns of the matrix. An analogous thinking for higher order square matrices also applies, but the determinant's value is the volume or hypervolume enclosed by a parallelepiped (an n -dimensional generalization of the parallelogram) whose edges are the column vectors of the matrix. This volume includes information about the organization and magnitude of the vectors, and thus the absolute value of the determinant is divided by the product of the magnitude of the vectors to give a parameter that scales with the overall ‘orthogonality’ of the vector system, the normalized determinant or D_{norm} (Eq. 9).

$$D_{norm} = \frac{\|\det(E)\|}{\prod_k \|\vec{E}_k\|} \quad \text{Equation 9}$$

$$D_{norm} = \frac{\|(\epsilon_{IF}^A \epsilon_{ABS}^B - \epsilon_{IF}^B \epsilon_{ABS}^A)\|}{\sqrt{\epsilon_{IF}^A{}^2 + \epsilon_{ABS}^A{}^2} \sqrt{\epsilon_{IF}^B{}^2 + \epsilon_{ABS}^B{}^2}} \quad \text{Equation 10}$$

In the equation for ‘ k ’-dimensional version of D_{norm} (Eq. 9), vertical double bars denote magnitude of a scalar or vector, $\det(E)$ is the determinant of E , and $\vec{E}_k$ is the k^{th} column vector of the matrix E . In the two-component version of D_{norm} (Eq. 10), the calculation is expanded to the level of extinction coefficients (or signal increments). A similar, but less compact expression can be generated for higher-dimension systems as desired.

D_{norm} can adopt values between zero and one. As all the signal increments are positive, one limit of D_{norm} ($D_{norm} = 1$) corresponds to the case of signals that are completely representative of only one component each (complete spectral discrimination). In this case, MSSV decomposition is similar to calculating separate $c(s)$ distributions for each signal. In the two component system, a $D_{norm} = 0$ describes parallel signal increment vectors for the two species, where there is no spectral discrimination. Values of D_{norm} larger than 0 describe signal increment vectors that are increasingly further apart. Using the parallelogram area formalism, D_{norm} is approximately the fraction of observable optical signals that can actually be sampled by positive component concentrations. Small, but non-zero values of D_{norm} can only sample a narrow wedge of the possible optical signals from two observables, and spectral discrimination will be poor. Further, when more than two signals are used, co-

linearity alone is an insufficient test for poor spectral discrimination. There are situations where three vectors are not co-linear, but are still degenerate; for three signals/three components, this corresponds to the three vectors falling in a plane. D_{norm} will be zero in these cases, and will scale towards one as a large fraction of the three-optical-signal space is accessible. Noise essentially shifts the position in 'optical signal' space slightly, and when the signal increment vectors are only slightly divergent, they sample only a small fraction of the optical signal space, and a small shift can result in large errors.

In the simple two-signal decomposition example described in Figure 1, the signal-increment matrix giving rise to the narrower distribution of determined ratios (the dashed curve) has a D_{norm} of ~ 0.7 , while the D_{norm} of the signal increment matrix of the broader distribution (the solid curve) is 0.07. In both the above cases, the same magnitude of noise was used, less than 1% of total signal. This limited amount of noise only modestly affects the system with $D_{norm} \sim 0.7$; no examples of the ratio were found to be less than 0.9 or greater than 1.1. In contrast, for the system where D_{norm} is ~ 0.07 , about half of ratios are less than 0.9 or greater than 1.1. Decomposing multi-signal sedimentation velocity experiments into $c_k(s)$ distributions is a related process, but it is a global analysis of more than ten thousand data points. We explored the effect of D_{norm} on the success rate for decomposition of MSSV data into $c_k(s)$ distributions using simulation of MSSV experiments in SEDPHAT.

This process had two stages. First, we modeled the formation of a heterodimer in an MSSV experiment, detected using interference optics and a single absorbance wavelength. One species was heavy and fast sedimenting (200 kDa, 8.5 S), the other light and slow sedimenting (10 kDa, 1.2 S), such that the complex is very well resolved from the light component in sedimentation analysis. To mimic the situation where the stoichiometry is not known prior to the experiment, more than a two-fold excess of the light component was included. We modeled a tight complex (1 nM) with an intermediate off rate (0.001 s^{-1}). This modeling exercise was performed by using the $A + B \leftrightarrow AB$ hetero-association model in SEDPHAT. The process was repeated for the IF and absorbance signal separately and synthetic noise (standard deviation of 0.005 OD) was added. Extinction coefficients were chosen to produce a range of D_{norm} values. D_{norm} values were varied by keeping both of the interference signal increments and the absorbance signal increment for the heavy component fixed, while varying the absorbance signal increment for the light component.

Next, a multisignal sedimentation velocity analysis was performed on the simulated data as described in section 5.1. Critically, the extinction coefficients were known, thus only the mixture was analyzed (see section 5.1.3). This analysis was repeated five times for each D_{norm} , with different synthetic noise each time. The quantities of light and heavy components were determined by integrating the $c_k(s)$ distributions in the range of the heavy (9.2 S) complex peak and the slowly sedimenting (1.2 S) peak resulting from excess light component. No additional peaks were observed for any simulated experiment. From these, conservation of mass and heavy-to-light ratio in the complex peak (~ 9.2 S) were calculated. Because we knew the actual ratio (very close to 1:1 in this case) and the input mass, we could score each simulation for these quantities. We scored the result in two ways. First, the average and standard deviation for the ratio measurements were calculated for the five repeats (Figure 2A). Next, each simulation was scored for conservation of mass (scored as successful when the total mass of each component was within 90 to 110 % of the input mass) and ratio (scored as successful when the heavy-over-light material ratio in the complex peak was between 0.9 and 1.1 for these 1:1 complexes). Finally, the simulations were scored for the heavy species appearing at the light species sedimentation coefficient. When more than 15% of the signal at this sedimentation coefficient stemmed from the wrong species, the simulation was scored as a failure. For the two-component cases

discussed first, there were only a few examples of the last mode of failure that had not also failed in the conservation of mass analysis.

The simulation results showed that as D_{norm} decreases from large values (> 0.2) to small values, the average ratio found abruptly increases (due to less light component found in the complex) for D_{norm} values below 0.06 (Fig. 2A). This observation was due to insufficient spectral discrimination alone, as the simulations use known extinction coefficients and did not fit many of the features usually fit in MSSV analysis. An advantage of the simulation approach was that features of the analysis that might ordinarily be blamed on sample heterogeneity can clearly be attributed to poor spectral discrimination. For example, if examination of the $c_A(s)$ distribution finds a significant amount of the heavy component (A) appearing with the sedimentation coefficient of the light species (B), one might be tempted to blame a degradation product or related sample artifact. However, given that these data were synthetic and that we understand the system completely, we could unequivocally say that the defect was not due to A actually sedimenting at 1.2 S for some reason, but due to poor spectral discrimination. Usefully, in the simulations, when a small quantity of the heavy component appeared at 1.2 S, the presence of its signal reduced the amount of the light component at 1.2 S, resulting in the analysis failing on the grounds of conservation of mass. Interestingly, conservation of mass fails in many cases where the molar ratio is approximately correct (Fig. 2B). For all the cases examined in Figure 2, there was only one example of both components being within 90% – 110% of input concentrations, and the ratio present in the complex (9.2 S peak) being outside of 0.9 – 1.1 (see note in the legend for Figure 2). This result tells us that conservation of mass is a critical criteria for evaluating the success of an MSSV experiment.

Use of conservation of mass as a criterion for successful decomposition is important because the use of χ_r^2 does not give any information on the success of a MSSV decomposition. When spectral resolution was poor, small errors were fit well with the wrong species. Thus, there was no difference between χ_r^2 of decompositions that succeed in conservation of mass, ratio in the complex, or absence of heavy component at 1.2 S, and those that fail. There was also no systematic trend in the residuals of either the successful or failed decompositions. The range of χ_r^2 found for all the simulations in Figure 2A was very narrow, which was dictated by the weighting of the two data sets (see section 2.2), and the noise introduced into the data sets. By comparing the values of χ_r^2 for trials with a successful ratio determination and those with an unsuccessful ratio determination, we found that there is almost complete overlap in χ_r^2 values. Positively scored trials have a χ_r^2 range of 0.3944 to 0.4118 and the negatively scored trials have a range of 0.3993 to 0.4117. Dividing the pool of results according to the success or failure of conservation of mass gave the same ranges. Thus, we conclude that the experimenter will be able to deduce very little from the quality of the fit to MSSV data on its own, demonstrating the importance of alternative criteria such as the conservation of mass in assessing the quality of the decomposition.

We asked what effect the magnitude of the added synthetic noise had on the marginal case of $D_{norm} = 0.032$ (Figure 2C). Above, the data were simulated with optical and interference (in units of fringes) noise levels of 0.005 (data from Figure 2B). We therefore explored identical systems, but with noise levels of 0.004 and 0.008. Simulations with decreased noise gave a marginal improvement in the average component ratio, but no practical improvement in the number of cases that pass the conservation of mass criteria and find the correct ratio. In contrast, for a set of simulations with the noise level increased to 0.008, none of the simulations passed the conservation of mass criteria and only one found an appropriate ratio of components in the complex. Typically, the user only has a limited level of control over noise levels (lamp cleanliness, sample purity, experimental design etc.), and

these are all reasonable levels of noise. This implies that the success or failure of an MSSV analysis may depend on collecting an exceptionally good data set for marginal D_{norm} 's. Of note, a less marginal case, with a D_{norm} of 0.066, found no practical change in conservation of mass or ratio determination performance when the optical and interference (in units of fringes) noise was increased from 0.005 to 0.010. Thus, the likelihood of a successful MSSV decomposition can be greatly improved by even a modest improvement of D_{norm} .

Is there an aspect of experimental design that can affect the likelihood of success in a positive way for marginal cases? One aspect that the user has control over is the quantity of data collected during acquisition. Thus, we repeated the simulation with D_{norm} of 0.032 and using acquisitions of sample absorbance at one – ten minute intervals (Figure 2D). It was found that the number of successful determinations of ratio, and number of simulations that scored positively for conservation of mass increased to five-of-five as the acquisition frequency increased. A five-minute acquisition interval (reasonable for actual experiments) gave intermediate results, while a two-minute acquisition interval found five of five simulations with good conservation of mass and good ratio determination. However, using an eight-minute acquisition interval gave several positively scored ratios, but no simulations which scored positively for conservation of mass. Clearly, the lower sampling frequencies hurt the analysis. These lower frequencies are sometimes needed for reasons of number of samples (e.g. if seven samples are acquired simultaneously). Admittedly, the user has limited options for achieving a faster acquisition rate, but for cases of marginal D_{norm} values, there are two remedies suggested by this simulation. The experimenter may either collect fewer samples such that a faster sampling rate may be achieved, or slow down the centrifugation speed such that sedimentation is slower and more data can be acquired over a longer period of time.

A last series of simulations asked what values of D_{norm} are useful in a three-component, three-signal experiment. Because of the number of parameters involved, a systematic exploration of D_{norm} is impractical. Instead, we chose several likely scenarios and simulated those. These are listed in Table 1. First, we considered a case (#1) where three proteins are used, in which two of the proteins have been labeled with extrinsic chromophores, imagined to be fluorescein and rhodamine derivatives in this case, and where absorbance wavelength 1 is 496 nm and absorbance wavelength 2 is 555 nm. Three unlabeled proteins, cases (#2) – (#4), were considered at imagined wavelengths of 280 nm (absorbance wavelength 1) and 250 nm (absorbance wavelength 2). Finally, we consider the situation in which a third labeled protein is added to a two-component system with reasonable D_{norm} . All three unlabeled protein cases had D_{norm} 's of less than 0.01, which is low compared to those found in the two-component, two-signal case. Given that two of these appear to be useful systems (see below), this implies that the minimum D_{norm} needed for a reasonable quality fit of data in an N dimensional system will need to be empirically determined. Here we begin the process for a three-component, three-signal system.

For analyzing modeled three-component systems, system (#1) performed exceptionally well. D_{norm} for this system was roughly 0.7 and this led to easy and reliable decomposition into components. In every case, the correct concentration of each protein was found in the complex and conservation of mass was very good and there was little evidence of spectral contamination in the peak at 4 S (Table 2).

Signals from three unlabeled proteins detected using interference, absorbance at 280 nm and absorbance at 250 nm, could also be decomposed when there was reasonable spectral discrimination. Spectral discrimination will come from substantial differences in the content of UV absorbing amino acids relative to the mass of the components, or the presence of cofactors. When the amino acid content is quite different for three interacting proteins, an

unlabeled three-component experiment may be feasible and D_{norm} should be calculated. In practice, one should calculate D_{norm} from observed absorbances at 280 nm and 250 nm from independently determined protein concentrations (e.g. using the method of Pace [12], or using Rayleigh interferometry to determine protein concentration), as estimation of the native protein extinction coefficient at 250 nm is not straightforward. The choice of 250 nm as our third signal may also not always be the correct choice; comparison of absorbance spectra normalized for protein mass for the three independent proteins may suggest a different wavelength to use.

The three-component cases designated (#2) – (#4) correspond to plausible extinction coefficients for unlabeled proteins that were chosen to produce different D_{norm} 's. Case (#2) corresponds to a favorable case, and is one where one protein has a fairly typical content of tryptophan, tyrosine and phenylalanine (the 50 kDa protein), one has substantially more tryptophans per unit mass (the 65 kDa protein) and one (the 20 kDa) has less tryptophan, but was relatively rich in tyrosine and phenylalanine, giving it a higher ϵ_{ABS250} to ϵ_{IF} ratio than the other proteins. The resulting D_{norm} was ~ 0.0057 . A simple ratio of component concentrations in the ABC complex is not a sensible criterion in this case, as there are three components, so instead we checked for concentrations within 90% – 110% of expected component concentration. Three of five trials found the correct component concentrations in the complex, and the remaining two were within 80% – 120%. Conservation of mass, scoring as successful when total concentration was within 90%–110% of input concentration, held up for all three components, and there was a minimum of species A in the 4 S peak, which would have indicated spectral contamination. In case (#3), where D_{norm} was ~ 0.0027 , the system actually performed slightly better at decomposing the system into component concentrations, although there was substantially more spectral contamination (Table 2). In test case (#4), D_{norm} was only 0.0016, and the analysis found the wrong concentration of species C in the complex for five of five trials. Alarmingly, mass is apparently conserved and spectral contamination is less than 15% in all five cases, such that our scoring criteria from the two component system does not catch these problems (Table 2). This suggests that for three-component systems, either low D_{norm} systems should be avoided, or additional post analysis quality control tests need to be developed.

Our last test system, (#5), was composed of two unlabeled proteins which together have a very good two-component, two-signal D_{norm} , and a third protein that is labeled with a chromophore such as fluorescein. Absorbance wavelengths are imagined to be ~ 490 nm for the fluorescein signal, and 280 nm for the tryptophan signal. The value of ϵ_{ABS280} for the labeled component was chosen as higher than expected for a protein of this mass, to reflect the UV absorbance of the label. This system had a D_{norm} of 0.0629, and the system performed very well. Five of five trials passed both conservation of mass tests and signal contamination tests, and found the correct concentrations in the complex peak (Table 2).

In summary, two-component systems can be reliably decomposed for D_{norm} 's of 0.065 and greater. At D_{norm} values near 0.03, decomposition is still possible, but is less reliable. When a system is found to have a D_{norm} of 0.03, it may be worth attempting an experiment, but several optimizations may need to be considered, such as collecting fewer samples at a time or slowing the rotor speed. For systems with D_{norm} 's of less than 0.03, a labeling scheme should be considered. D_{norm} can be used to evaluate spectral discrimination in three-signal systems as well. As with two-component systems, introducing a chromophore can result in very good spectral discrimination and will produced the most reliable decompositions. However, proteins frequently have different ratios of ϵ_{ABS280} to ϵ_{ABS250} to ϵ_{IF} , and plausible ranges of these signal increments can yield a system with sufficient spectral resolution to resolve three components. Given the difficulty of labeling some systems, it is certainly worth calculating D_{norm} for a three-protein system and considering such an experiment. If

three signals that provide a satisfactory D_{norm} cannot be found, but two of the proteins have a very good two-component, two-signal D_{norm} (e.g. greater than 0.06), labeling only the third protein with an exogenous chromophore is a good option. It is worth noting that, by their nature, the simulations performed here lack much of the uncertainty present in actual experimental data (e.g. uncertainty in the extinction coefficients, more systematic noise etc.), thus, performing experiments right at the edge of feasibility should be approached with caution.

4.2 Choice of component concentrations

The second challenge in designing an MSSV experiment is to choose the concentrations of the components used. This feat requires balancing both the dynamic range of absorbance measurements and the concentration needs of the interaction system studied. Below, we develop a graphical treatment that works well for most MSSV experimental design situations.

First, we considered an MSSV experiment with high affinity and slow off rates. In this case we needed only consider the absorption limits for each signal. Assuming that each component has some absorbance at each absorbance wavelength, we needed to include enough materials to generate a high-confidence absorption measurement, but low enough not to saturate the system. For MSSV analysis, we suggest minimum optical density of 0.1 and maximum optical density of 0.8. This upper limit was intentionally conservative; it included a roughly optical density margin of 0.2 for errors in protein concentration or estimates of extinction coefficient. Also, our experience with MSSV analysis for samples with optical densities greater than one has not been as good as at optical densities of less than one, which may stem from somewhat higher noise levels for such samples. For the interference optics there will not be a practical upper limit, which removes a constraint. We avoided total protein concentrations much in excess of 1 mg/mL to avoid hydrodynamic nonideality, which could have been included as a separate constraint. These constraints were cast as inequalities of the form of Eqs. 11, 12 and 13.

$$\epsilon_{ABS}^A l[A]_{tot} + \epsilon_{ABS}^B l[B]_{tot} \geq 0.1 \quad \text{Equation 11}$$

$$\epsilon_{ABS}^A l[A]_{tot} + \epsilon_{ABS}^B l[B]_{tot} \leq 0.8 \quad \text{Equation 12}$$

$$\epsilon_{IF}^A l[A]_{tot} + \epsilon_{IF}^B l[B]_{tot} \geq 0.1 \quad \text{Equation 13}$$

Where l is the path length (typically 1.2 cm), ϵ_{ABS}^A is the molar signal increment for the absorbance signal (also referred to as the molar extinction coefficient) of species A. $[A]_{tot}$ and $[B]_{tot}$ are the total input molar concentrations of species A and B, respectively.

Typically, experiments to determine stoichiometry will be performed in the presence of an excess of one component. An additional constraint is that the complex whose stoichiometry is being determined should absorb sufficiently to measure confidently in each signal (conservatively, at least an optical density of 0.1 or interference signal of 0.1 fringes). If we consider the case where component A is in excess, then the concentration of the complex will be limited by $[B]_{tot}$. If we assume a minimum stoichiometry (1:1 in most cases) then we can determine what the complex signal increment is. By casting the complex absorbance as

an inequality with minimum complex absorbance of 0.1 or interference signal of 0.1 fringes, we generated an additional constraint:

$$(\varepsilon_A^A + \varepsilon_A^B)I[B]_{\text{tot}} \geq 0.1 \quad \text{Equation 14}$$

A similar constraint is generated by considering the case where component B is in excess:

$$(\varepsilon_A^A + \varepsilon_A^B)I[A]_{\text{tot}} \geq 0.1 \quad \text{Equation 15}$$

Finally, we plotted the inequalities of Eq. 11 – 15 on a single $[A]_{\text{tot}}$ by $[B]_{\text{tot}}$ field (Figure 3A). The region where all the inequalities are satisfied was the set of concentrations useful for MSSV experiments (triangular gray region in Figure 3A).

A somewhat more complex situation occurs when the affinity is weaker, or the off-rate is fast relative to the timescale of the MSSV experiment (which occurs quite frequently). In this case, two additional constraints are added. First, a large excess of one species over the other is used. As the stoichiometry is not known ahead of time, it is warranted to keep the relative concentration of ligand over receptor (here $[B]_{\text{tot}}$ over $[A]_{\text{tot}}$) above some reasonable upper limit on the stoichiometry, five or ten fold is often sensible. This also solves a problem of $c(s)$ treatments for dynamic interactions [1]. This constraint can be expressed as an inequality:

$$[B]_{\text{tot}} \geq n[A]_{\text{tot}} \quad \text{Equation 16}$$

where $[A]_{\text{tot}}$ dictates the minimum $[B]_{\text{tot}}$, which is ‘ n ’ fold over $[A]_{\text{tot}}$ 1. The final constraint comes from the need to have complex in a degree of saturation, usually one species (A here) can be thought of as the receptor and the other (B here) as the ligand. Complete saturation of the receptor with ligand is of course impractical, but the experiment can be designed to achieve a minimum degree of saturation.

$$[B]_{\text{tot}} \geq f_{\text{sat}}[A]_{\text{tot}} + \frac{f_{\text{sat}}K_D}{1 - f_{\text{sat}}} \quad \text{Equation 17}$$

In Eq. 17, f_{sat} is the minimum fraction of A bound to B, and K_D is the dissociation constant. An estimate for weakest reasonable K_D is acceptable here. By assuming a one to one stoichiometry, and asking what concentration of B is required to achieve a minimum degree of saturation of A, f_{sat} , given a value for the dissociation constant (K_D), the relationship can be arrived at through algebraic rearrangement. The assumed stoichiometry should not be problematic as long as the large excess constraint (Eq. 16) is imposed simultaneously. For cases where off-rate is fast (i.e. $k_{\text{off}} < 10^{-2} \text{ s}^{-1}$), effective particle theory may provide an improved version of this constraint [13].

To determine the set of concentrations that may be used for MSSV experiments, we plot the inequalities Eqs. 11 – 17 on a single $[A]_{\text{tot}}$ by $[B]_{\text{tot}}$ field (Figure 3B). Here we consider a situation of B being in excess over A, but an analogous construction can be made for A in

¹This assumes a receptor-ligand relationship is sensible to impose. When an assembly forms with a complex stoichiometry (e.g. 4A:8B, but A is monomeric on its own), having a large excess may not be sensible. However, many systems that exhibit complex stoichiometry will also be cooperative, and thus the results may still be fine.

excess over B. The constraints define an irregular area of allowable concentrations in this example (quadrilateral gray region in Figure 3B). In other systems, triangular and pentagonal regions may also be found.

The same experimental design approach also works for three or more components as well. In principle, one would need to use these constraints to generate planes instead of lines, and to examine the resulting allowed volume (or hypervolume) for reasonable component concentrations. In practice, only two-dimensional projections need be considered in many cases, and the remaining cases can operate using two-dimensional projections followed by testing against the absorbance upper limits (using an equation similar to Eq. 12).

4.3 Quality checks for MSSV experiments

Following the analysis of an MSSV experiment, it is important to impose a series of quality checks on the analysis. There are several reasons for these checks, not the least of which are situations of insufficient spectral resolution, in which case problems may only become evident once all of the quality checks have been used. We suggest a four-stage quality-control process.

4.3.1 Concentration correspondence—Do the calculated concentrations correspond to the input concentrations? As discussed above, in many cases poor spectral resolution results in violations of conservation of mass. In general, pipetting error will not exceed 10%, and it is worth some care to ensure that experiments are designed to minimize pipetting error. Concentration estimates based on calculated UV absorbances for native proteins can be inaccurate by more than 10% [12], thus some effort may need to go into calculating the actual concentration present. In most cases, the IF optics provide an excellent alternative, and the input concentrations measured for the individual components, as part of the signal increment optimization process, should serve as the reference value for input concentration. More difficult cases may require alternative means of estimating concentrations [3].

4.3.2 Rationality of the $c_k(s)$ distribution—Do peaks appear in unexpected places in the $c_k(s)$ distribution? For well-resolved species, it is not uncommon for a small amount of one species appear to cosediment with another uncomplexed species. We term this phenomenon “spectral contamination,” and it occurs even in simulated MSSV analyses where it is known that there is no actual cosedimentation of both species at the uncomplexed sedimentation coefficient. Instead, this results from poor spectral discrimination. Small amounts of this spectral contamination are common and will have little effect on the interpretation of the system. Ideally, spectral contamination should account for no more than 15% of any signal. Due to the possibility of very different signal increment values for different components, it generally makes sense to consider spectral contamination in terms of the absorbance or IF signal, not concentration. In other words, use the concentration sedimenting anomalously, multiplied by the signal increment to calculate the fraction of spectral contamination. More than 15% spectral contamination in any one species indicates poor spectral resolution, and stoichiometries calculated from such an analysis will be unreliable.

4.3.3 Rationality of the computed molar ratio—Does the molar ratio make sense? The experimenter should ask several questions about a stoichiometry determined by MSSV analysis at this point. The molar ratio only gives relative stoichiometry. Is the ratio near a unit value (e.g. a two to one ratio)? Would a complex with that ratio potentially sediment with the sedimentation coefficient and frictional ratio found in a $c(s)$ analysis of the same data? Does the stoichiometry make sense biologically? Are there nearby alternative stoichiometries that should be considered?

4.3.4 Statistical tests—Do χ^2_c tests reject or accept sensible stoichiometries? The MSSV analysis affords the experimenter the opportunity to test the apparent stoichiometry for statistical validity (see sections 3.3 and 5.1.4; also ref. [3]). In particular, does an apparent stoichiometry fit the data well when imposed rigidly? In cases of excellent spectral resolution, imposition of a unit stoichiometry suggested by an MSSV analysis (e.g. a calculated ratio of 1.8 suggests a true stoichiometry of 2) may make the fit statistically worse. If this is the case, consider what evidence supports saturation of the receptor protein, or an incompetent fraction of species sedimenting similarly. Alternatively, is there a chance that there is some hypo- or hyper-chromicity? Are alternative stoichiometries rejected by such an analysis? In cases of poor spectral resolution, the MSSV suggested stoichiometry may fit the data very well, but alternative stoichiometries may also fit the data well. In these cases, incorrect alternative stoichiometries that fit the data by χ^2_c tests may violate conservation of mass. In this way, marginal spectral resolution may still afford clear determination of stoichiometry in some cases.

5 Results

5.1 GST-VCA binding to Arp2/3: An example of MSSV analysis

The cytoskeletal protein actin cycles between a soluble monomeric form and a filamentous form, driven by the hydrolysis of ATP [14]. The dynamics of actin converting between these states are important for generating protrusive forces in many cellular processes, including motility and vesicle trafficking. Actin polymerization dynamics are largely regulated at the level of nucleation, which is a kinetically limiting step. The heteroheptameric Arp2/3 complex nucleates new actin filaments from the sides of pre-existing actin filaments, generating a branched network [15]. This actin nucleation activity of Arp2/3 complex is basally inhibited, but is stimulated by nucleation promotion factors; one of which is the Wiskott-Aldrich Syndrome protein (WASP) [15–17]. A C-terminal domain of WASP, termed the “VCA,” is responsible for WASP binding to, and activating, Arp2/3 [18]. Arp2/3 complex is asymmetric, and thus it was assumed that only a single VCA peptide bound and activated the complex. Higgs and Pollard first noted that a GST-VCA fusion more potently stimulated Arp2/3-mediated nucleation [17]. Padrick *et al.* followed this result by demonstrating that the dimerized VCA binds to Arp2/3 with ~100-fold greater affinity than does monomeric VCA [2]. Padrick *et al.* used various means to accomplish the tethering of two VCA domains; among them was fusing the VCA domain to the C-terminus of glutathione S-transferase (GST), which is dimeric.

The higher affinity of dimeric VCA for Arp2/3 suggested that there are two VCA-binding sites on the Arp2/3 complex [2]. Thus, the question arose: how many GST-VCA dimers bind to Arp2/3? Alternatively, how many Arp2/3 complexes bind to a GST-VCA dimer? Previously, Padrick *et al.* answered this question by examining the hydrodynamic properties of the GST-VCA:Arp2/3 complex. They concluded that the complex was a 1:1 mixture of GST-VCA dimer and Arp2/3. More recently, it has been shown that two labeled VCA monomers can bind Arp2/3 independently as well [3]. In the example that follows, MSSV is employed to answer the same question.

Before undertaking the experiment, we asked whether MSSV analysis was likely to work with GST-VCA and Arp2/3. The principle concern was of unlabeled protein spectral discrimination. From the primary structure of the proteins we calculated their molar mass and thus their interference signal increments, 193,089 fringes·M⁻¹·cm⁻¹ for GST-VCA and 615,516 fringes·M⁻¹·cm⁻¹ for Arp2/3. Additionally, an estimate of $\epsilon_{\text{ABS}280}$ was made based on tyrosine and tryptophan content in SEDNTERP, 96,000 M⁻¹·cm⁻¹ and 230,000 M⁻¹·cm⁻¹ for GST-VCA and Arp2/3, respectively. D_{norm} calculated for these signal

increments (using Eq. 10) was slightly larger than 0.1, which, as discussed in section 4.1, is expected to yield excellent signal discrimination.

Three experiments were performed. One centrifugal cell (Cell 1) contained only GST-VCA, one (Cell 2) contained Arp2/3 only, and a third (Cell 3) contained a mixture of both proteins, with GST-VCA present in roughly a ten-fold molar excess. Both IF and A_{280} data were acquired from these experiments. The experiments were carried out in a single centrifugal run at 20° C, with a rotor speed of 42,000 rpm. Further experimental details can be found in Methods. The analysis presented immediately below is extensively documented pictorially in a pdf file that is available as Supplementary Information to this paper. In general, there are three steps taken to accomplish this analysis: (1) refinement of ϵ_{ABS280} for GST-VCA alone, (2) refinement of ϵ_{ABS280} for Arp2/3 alone, and (3) the MSSV analysis of the mixture. The first two steps are necessary to put the extinction information on the same scale as that of the on-board detection systems. In only one instance have we found it advisable to ignore one of the first two steps: the interaction of Arp2/3 with a chromophore-labeled VCA [3]. In that example, one of the monitored signals was A_{496} . Because unlabeled Arp2/3 was known to have no signal at that wavelength, $\epsilon_{\text{ABS496}}^{\text{Arp2/3}}$ was known to be zero, and thus it was unnecessary to refine it.

5.1.1 Refinement of ϵ_{ABS280} for GST-VCA—The purpose of this step is to refine $\epsilon_{\text{ABS280}}^{\text{GST-VCA}}$. To accomplish this optimization, two SV data sets, one of IF data and one of A_{280} data, were globally modeled with a single $c_{\text{GST-VCA}}(s)$ distribution. In order for this distribution to model the data well, the extinction information for the protein must be correct for both signals (see Eq. 4). The $\epsilon_{\text{IF}}^{\text{GST-VCA}}$ was treated as a known, reliable quantity; for most proteins, ϵ_{IF} for the interferometer in our instrument (laser wavelength = 675 nm) can be represented well by multiplying the molecular weight by 2.75 [3,19]; that approach was taken here. By constraining the analysis such that a single distribution must model both data sets, $\epsilon_{\text{ABS280}}^{\text{GST-VCA}}$ can be refined to its optimal value given the data.

First, the concentration profile data (“scans”) from Cell 1 were loaded into memory. This action was organized into two “experiments” in SEDPHAT. The first experiment was defined as scans 1–101 of the IF data (every other scan was loaded). Experiment 2 was the same scan range for the A_{280} data. This span of scans represented about 7 hours of centrifugation. The experimental parameters were input into the “Experimental Parameters” dialogs as shown in Table 3. There are a few items worth expanding on here. The $\bar{v}$ of all experiments in sections 5.1.1, 5.1.2, and 5.1.3 does not have a strong bearing on the goal of this analysis, i.e. the molar ratio of the GST-VCA/Arp2/3 complex. It was thus allowed to remain at its default value in SEDPHAT, 0.73 cm³/g. In some cases, it may be necessary to estimate the mass of the complex in order to assess the stoichiometry. In such cases, the estimated $\bar{v}$ of the complex should be input. Also, for the A_{280} data, the checkbox (labeled “*sqrt(N1/Nx)”) near to the “noise” entry box in the Experimental Parameters dialog was checked. This expedient compensated for the difference in the number of data points between the two experiments (see section 2.2 and ref. [3]). The menisci were chosen graphically using optical artifacts of the data acquisition as guides. The cell bottoms were chosen similarly. The fitting limits were chosen such that the data considered no optical artifacts and little to no back-diffusion.

At this point, the “Multi-Wavelength Discrete/Continuous Distribution Analysis” was selected in the Model menu. This model allows the user to fit the data to multiple discrete species and/or continuous distributions. In this case, a single continuous distribution, with s -values ranging from 0.2–10 S was selected in the Global Parameters dialog. Because only

GST-VCA was present, this distribution was considered to be a $c_{\text{GST-VCA}}(s)$ distribution. Because two signals were present, up to two macromolecules (or “chromophores,” in the program’s notation) could be analyzed, but only a single chromophore was analyzed in this section. GST-VCA was defined in the program as “chromophore #1”. In the Global Parameters dialog, the known extinction information concerning GST-VCA was entered.

For GST-VCA, $\epsilon_{\text{IF}}^{\text{GST-VCA}}$ was 193,089 fringes·M⁻¹·cm⁻¹. This value was treated as a known, reliable, and fixed quantity. For $\epsilon_{\text{ABS280}}^{\text{GST-VCA}}$, an initial guess of 96,000 M⁻¹·cm⁻¹ was calculated using SEDNTERP. The refined value of this parameter was the goal of this analysis, so it was allowed to refine by activating the checkbox next to its input box. This and all subsequent distributions in this work were regularized using the Tikhonov-Phillips method at a confidence level of 0.7 [4]. Further, the resultant distributions were normalized such that $\int_{s_{\min}}^{s_{\max}} c_k(s) ds = [k]$ between $s_{\min}$ and $s_{\max}$.

The parameters were allowed to globally refine. SEDPHAT can utilize one of three minimization algorithms: Simplex, Marquardt-Levenberg (ML), and Simulated Annealing. Simplex is the default in SEDPHAT, and this algorithm was used in this initial fitting session. The only nonlinear parameter refining was $\epsilon_{\text{ABS280}}^{\text{GST-VCA}}$. The others, namely the two sample menisci and the f_r , were fixed. This initial refinement strategy allows $\epsilon_{\text{ABS280}}^{\text{GST-VCA}}$ to quickly and efficiently refine to a value that is probably close to its final value. After this initial refinement, the other nonlinear parameters named above were allowed to refine in a subsequent fitting session using the ML algorithm.

After that fitting session had converged, the attributes and quality of the fit were examined (Fig. 4). The $c_{\text{GST-VCA}}(s)$ distribution (Fig. 4C) had two main peaks; one at 3.7 S, the other at 5.3 S. Given the purity of the sample (not shown), it seems likely that GST-VCA at this concentration was mostly dimeric, but a detectable population of higher-order oligomer was also present. The fits (lines in the upper parts of Figs. 4A & 4B) matched the data (circles) well, and the residuals (lower parts of Figs. 4A & 4B) did not exhibit any serious systematicity. The local root-mean-square deviations between the data and the fit were 0.008 fringes for the IF data set and 0.005 AU for the A_{280} data set; these values are close to instrumental noise for the centrifuge used in these experiments. The fit to the data was therefore deemed to be good. The final refined values of f_r and $\epsilon_{\text{ABS280}}^{\text{GST-VCA}}$ were 1.781 and 92,416.2 M⁻¹·cm⁻¹, respectively. The fitting session was saved.

As noted above, the result of integrating a $c_k(s)$ distribution over a range of s -values is the molar concentration of the component k . It is usually of interest to integrate a distribution as obtained above to see if it meets the experimenter’s expectations. Integrating the entire $c_{\text{GST-VCA}}(s)$ distribution yields $[\text{GST-VCA}] = 4.09 \mu\text{M}$, which is about what was expected. Thus, the quality-check criterion established in section 4.3.1 is met for GST-VCA.

5.1.2 Establishing the ϵ_{ABS280} for Arp2/3—The approach here is essentially the same as in section 5.1.1. The data (Figs. 5A & 5B) were loaded into SEDPHAT as before, with some exceptions. Arp2/3, being much larger than GST-VCA, sediments more quickly. Thus, only scans 1–50 (all scans in this range) from both the IF and A_{280} SV data were loaded into the program. These scans represent about 3.5 hours of sedimentation.

For this data set, there was again only one “chromophore,” i.e. Arp2/3. Therefore, we designated Arp2/3 as “chromophore #2” in the program. The calculated value of $\epsilon_{\text{IF}}^{\text{Arp2/3}}$ (615,516 fringes·M⁻¹·cm⁻¹) was taken as known, reliable, and fixed in this analysis, and the goal was to refine the value of $\epsilon_{\text{ABS280}}^{\text{Arp2/3}}$. The starting value of $\epsilon_{\text{ABS280}}^{\text{Arp2/3}}$ was estimated to be about

230,000 M⁻¹·cm⁻¹, according to SEDNTERP. Of course, the checkbox allowing refinement of $\epsilon_{\text{ABS280}}^{\text{Arp2/3}}$ was activated. The distribution used to analyze these data used 50 *s*-values between 0.2 and 15 S.

The analysis proceeded almost exactly as that presented in section 5.1.1. The final refined value of $\epsilon_{\text{ABS280}}^{\text{Arp2/3}}$ was 244,420 M⁻¹·cm⁻¹, and that of f_r was 1.618. The result of this fitting session is shown in Fig. 5. By integrating the $c_{\text{Arp2/3}}(s)$ distribution (Fig. 5C), its total concentration, 1.49 μM, was calculated. Again, this result met with our expectations. The fitting session was saved.

5.1.3 Analysis of the mixture—The next step was the multisignal analysis of the mixture of GST-VCA and Arp2/3. All of the extinction information for both proteins was now known; it was either calculated or refined in sections 5.1.1 or 5.1.2. The task at hand was to globally model two data sets that were simultaneously acquired and that contained a mixture of GST-VCA and Arp2/3. The approach was to analyze the data with two completely overlapping distributions: one representing the concentration of GST-VCA ($c_{\text{GST-VCA}}(s)$), and the other representing the concentration of Arp2/3 ($c_{\text{Arp2/3}}(s)$). This decomposition of the data can take place in cases where D_{norm} is sufficiently large (see section 4.1), as it is here. The D_{norm} of the calculated and refined signal increments was 0.068, slightly lower than estimated above, but still likely to result in good spectral discrimination; see section 4.1.

Using the above calculated concentrations, we estimated that the total concentration of GST-VCA ($[\text{GST-VCA}]_{\text{tot}}$) should be 3 μM and that $[\text{Arp2/3}]_{\text{tot}}$ should be 0.4 μM in the mixture. The stoichiometry of the complex between the proteins, based on mass considerations, was estimated to be 1:1[2]. Here, we use the MSSV approach to test the hypothesis that the complex of Arp2/3 and GST-VCA has a molar ratio of 1:1. Such a result would substantially strengthen the supposition of a 1:1 stoichiometry.

The first step in the analysis was to load into SEDPHAT the SV scans that monitored the sedimentation of the mixture. For both the IF and A_{280} data, scans 1–101 were loaded into SEDPHAT as Experiment #1 and Experiment #2, respectively. After choosing the multisignal model, the extinction properties were entered. In keeping with the precedents established in sections 5.1.1 and 5.1.2, GST-VCA was designated as “chromophore #1” and Arp2/3 as “chromophore #2.” The previously calculated values for $\epsilon_{\text{IF}}^{\text{GST-VCA}}$ and $\epsilon_{\text{IF}}^{\text{Arp2/3}}$ were input, as were the refined values for $\epsilon_{\text{ABS280}}^{\text{GST-VCA}}$ and $\epsilon_{\text{ABS280}}^{\text{Arp2/3}}$. Of course, all of the extinction information is now treated as known and fixed; it is inappropriate to refine it at this stage of the analysis.

In this case, two segments of sedimentation coefficient space were used to model the SV data. Segment 1 ranged from 0.2 to 5.3 S, and Segment 2 spanned 6.8 to 15 S². The reasons behind this expedient were twofold. First, the two segments allowed two different f_r 's to be refined, one for each segment. Because GST-VCA and Arp2/3 had significantly differing f_r 's, this approach avoids forcing both molecules to have identical f_r 's. Indeed, the f_r 's of the two segments were initialized at 1.8 (Segment 1) and 1.6 (Segment 2), values that are close to those refined for GST-VCA alone and Arp2/3 alone, respectively. Second, the two separate segments allow for stoichiometric constraints to be placed on the *s*-range where the complex was expected and not on the *s*-range where GST-VCA alone sedimented. Importantly, each segment had two $c_k(s)$ distributions, as described earlier in this section. For this step, the f_r 's and the sample menisci were allowed to refine. Also, experience has demonstrated that it is best to choose the ML optimization algorithm for this stage of the analysis.

The program converged on a fit that was deemed to be good (Fig. 6). Clearly, GST-VCA and Arp2/3 were cosedimenting at an s -value of about 10.4 S. The local r.m.s.d.'s were 0.006 fringes for the IF data and 0.004 AU for the A_{280} data, and the residuals lacked significant systematicity. The final value of χ_r^2 , i.e. χ_b^2 , was 0.3061391; this value was noted for later use.

5.1.4 Assessing the result and its quality—At this point in the analysis, two of the quality criteria enumerated in section 4.3 can be examined (see sections 4.3.1 and 4.3.2). First, both $c_k(s)$ distributions were integrated over their entirety to find $[\text{GST-VCA}]_{\text{tot}}$ and $[\text{Arp2/3}]_{\text{tot}}$. Here, the integrated $[\text{GST-VCA}]_{\text{tot}} = 3 \mu\text{M}$, and the integrated $[\text{Arp2/3}]_{\text{tot}} = 0.4 \mu\text{M}$. These integrated values were in excellent agreement with those expected from the amount of material pipetted (section 5.1.3). Thus, the criterion of section 4.3.1 was passed. Also, an examination of the distribution (Fig. 6C) showed that no peaks were observed for either distribution in unexpected or unphysical regions of s -space. The criterion of section 4.3.2 was also therefore met.

The goal of the analysis is to glean the molar ratio of the GST-VCA:Arp2/3 complex. Toward this end, the distributions were integrated in the s -range from 9.3 to 11.2 S, where cosedimentation was evident (Fig. 6). The resulting concentrations were $[\text{GST-VCA}] = 0.36 \mu\text{M}$ and $[\text{Arp2/3}] = 0.39 \mu\text{M}$. The molar ratio of GST-VCA to Arp2/3 is therefore 0.92 to 1. This result conforms well to our expectations and is rational. Therefore, it meets the quality criterion put forth in section 4.3.3.

Finally, using F-statistics, it is straightforward to test possible molar ratios of the cosedimenting complex. To this end, $\chi_{c,1\sigma}^2$ and $\chi_{c,2\sigma}^2$ were calculated (see Eq. 5). They were $\chi_{c,1\sigma}^2 = 0.306843$ and $\chi_{c,2\sigma}^2 = 0.308574$. Then, molar-ratio constraints were added, followed by fitting sessions that resulted in new values of $\chi^2(\chi_i^2)$. The interpretations of relative values of χ_i^2 , $\chi_{c,1\sigma}^2$, and $\chi_{c,2\sigma}^2$ are discussed in section 3.3.

Various molar-ratio constraints were applied to the second “segment” of s -space, i.e. the region from 6.8 to 15 S. First, the data were modeled with a 1:1 molar ratio of GST-VCA and Arp2/3. That is, a single $c_{\text{GST-VCA:Arp2/3}}(s)$ distribution was used to model the data in this region of s -space. After fitting, the χ_i^2 was 0.3063701. Thus, constraining the molar ratio here did not deleteriously affect the quality of the fit ($\chi_i^2 < \chi_{c,1\sigma}^2$). With this result, our fourth criterion (section 4.3.4) was met; a sensible molar-ratio constraint was not rejected by the statistical test.

A stricter statistical test would be to probe other possible molar ratios to see if they are also acceptable by our criteria. To that end, alternative molar ratios were studied. It is possible that two GST-VCA dimers could bind to a single molecule of Arp2/3 [2,3]. Therefore, Segment #2 of s -space was constrained to have a molar ratio of two GST-VCA dimers for every one Arp2/3. The resulting χ_i^2 was 0.3096959, well above $\chi_{c,2\sigma}^2$. The 2:1 molar ratio may thus be safely rejected.

Also tested was a 1:2 molar ratio of GST-VCA and Arp2/3. This ratio is extremely unlikely, given our hydrodynamic results—a complex containing two Arp2/3 complexes would necessarily have a much higher sedimentation coefficient or f_r than was observed. Nonetheless, for illustrative purposes, this molar ratio was probed as above. This study yielded a χ_i^2 of 0.307723, presenting the interesting case of $\chi_{c,2\sigma}^2 > \chi_i^2 > \chi_{c,1\sigma}^2$. The tested molar ratio was statistically worse than the best, unconstrained fit, but it did not meet our

criterion for outright rejection. In this case, other evidence, i.e. the unphysical nature of a 1:2 species with a sedimentation coefficient of 10.4 S, can be used to reject this molar ratio and stoichiometry.

Given that all of the criteria stated in section 4.3 have been met, it was concluded that the data are consistent with a 1:1 molar ratio of GST-VCA and Arp2/3 sedimenting at about 10.4 S. Coupled with the mass information obtained from this system, a 1:1 stoichiometry of the complex is strongly supported by the data. Although a 2:1 complex is possible, it is rejected by this MSSV analysis.

5.2 An MSSV binding isotherm

As detailed below in section 6.1, the successful implementations of MSSV have until now involved systems in which the binding of two macromolecules results in a complex that has a significantly different sedimentation coefficient. Simplistically, in an $A + B \leftrightarrow AB$ system, the molecules have the sedimentation coefficients s_A , s_B , and s_{AB} , respectively. The systems thus far examined have the properties $s_A \neq s_B \neq s_{AB}$, $s_{AB} > s_A$, and $s_{AB} > s_B$. However, it is easy to imagine a scenario in which $s_A \gg s_B$. In such a case, it is possible that $s_A \approx s_{AB}$. The standard methods of analyzing SV data to obtain the association constant of two interacting solutes are Lamm-Equation fitting with explicit treatment of reaction kinetics [20,21] and $c(s)$ -based isotherm analysis [5,7,10,22]. Although both of these methods are capable of analyzing such a system, the lack of a difference between s_A and s_{AB} reduces the information content of the experimental data. However, if the D_{norm} of the two interacting solutes is large enough (see section 4.1), MSSV could be used to monitor the concentrations of free A, free B, and the AB complex. From these data, population isotherms could be constructed that contain information on K_A .

Recently, we encountered an interacting system of proteins similar to that described above. The human mitochondrial enzyme dihydrolipoamide dehydrogenase (hE3) is secured to a multienzyme complex (pyruvate dehydrogenase complex) as a result of its interactions with a small binding domain, called E3-binding domain (hE3BD). hE3 has an $s_{20,w}$ -value of 5.6 S, while that of an hE3BD-containing protein construct (XDD1) is about 1.4 S. When mixed, the two proteins form a 1:1 5.8-S complex [23]. An isotherm based on the weighted average of the sedimentation of the entire mixture (s_w) can be used to estimate K_A in such a system. However, additional information is available from the populations of the species [10]; unfortunately, these populations are not readily available from a conventional $c(s)$ -based analysis due to the inevitably poor resolution between s_{hE3} and $s_{hE3:XDD1}$.

To explore the utility of an MSSV-based population isotherm, a system similar to that described above was simulated. A mutant form of XDD1 (K160A) was chosen for simulation because this protein has an association constant ($K_A \approx 454,500 \text{ M}^{-1}$; $K_D \approx 2.2 \text{ }\mu\text{M}$) that should be accessible to this type of analysis. Thus, a 100,000 g/mol protein (A) with an s -value of 5.6 S binding to a 19,000 g/mol protein (B) having an s -value of 1.4 S was simulated. The extinction information for this simulated system is found in Table 4. These values were approximated for hE3-like and XDD1-like proteins detected using IF and A_{280} . The D_{norm} resulting from this simulation is therefore 0.26; such a value should afford excellent spectral discrimination (section 4.1). Although the magnitude of D_{norm} may seem high here, it is close to that observed (0.26) in an actual experiment with hE3 and XDD1 [23]. The rotor speed of the simulation was 50,000 rpm, and data were acquired once every 10 minutes for 800 minutes. The K_A of the simulated association was approximately 457,000 M^{-1} (i.e. $\log(K_A) \equiv 5.66$; $K_D \approx 2.19 \text{ }\mu\text{M}$). The simulated concentration of protein A was held constant at 1 μM , and that of protein B was varied from 0.5 to 16 μM (six samples were simulated). The k_{off} of the association was defined to be slow (10^{-6} s^{-1}).

The data were analyzed using MSSV, essentially as described in section 5.1. The [B] that cosedimented with A was treated as identical to [AB]. This approach allowed the construction of a population isotherm that predicted [AB] given the total concentration of A ($[A]_{\text{tot}}$; held constant here) and $[B]_{\text{tot}}$ (see Eq. 6). Figure 7A displays the values for [AB] obtained from the MSSV analysis (circles) and the fit to those data (line). The fitted value of K_A was $445,200 \text{ M}^{-1}$ ($K_D \approx 2.25 \text{ }\mu\text{M}$), which was only 2.6% lower than the simulated value. Thus, in this realistic hypothetical system, the proposed MSSV-based population isotherm performs well.

As mentioned above, the s_w -based isotherm could also be used to analyze this hypothetical system. Thus, the question naturally arose as to whether the MSSV population isotherm and an s_w isotherm could be globally analyzed to arrive at an accurate estimate for K_A . Such a global analysis is shown in Fig. 7B. The fitted value for K_A in this case was $441,500 \text{ M}^{-1}$ ($K_D \approx 2.26 \text{ }\mu\text{M}$), 3.4% lower than the simulated value. We concluded that the MSSV population isotherm, alone or in conjunction with an s_w isotherm, can arrive at an accurate estimate for K_A when implemented as above. Of course, the free concentrations of A and B are either directly available or easily calculable from the MSSV analysis. Thus, their concentrations could also be included in the MSSV population isotherm. We have not yet thoroughly explored this possibility.

Although the simulation shown here had a slow k_{off} , another simulation with identical parameters except for instantaneous interaction kinetics performed similarly (fitted $K_A = 451,200 \text{ M}^{-1}$; $K_D \approx 2.22 \text{ }\mu\text{M}$). An interacting system with instantaneous interaction kinetics may be characterized with effective particle theory (EPT) [13]. It has been observed that rapidly reversible protein interactions of the type described above usually result in two boundaries. One boundary, containing only free A or free B, is the “slow” or “undisturbed” boundary. The other, termed the “fast”, “moving”, or “reaction” boundary, contains free A, free B, and the AB complex. EPT envisions that the reaction boundary of such a system as a single effective particle with a single sedimentation coefficient that may be easily calculated given the parameters of the system. EPT also makes predictions governing the “stoichiometry” (as defined in [13]) of the components comprising the reaction boundary, which may be fit to an isotherm. Because stoichiometric information is available from an MSSV experiment, application of EPT to the current analysis may be advantageous. Following on the definition of [AB] given above, we define the stoichiometry of the reaction boundary as $[AB]/[A]_{\text{tot}}$ ³. Studying this stoichiometry as a function of $[B]_{\text{tot}}$ allows the construction of an EPT/MSSV stoichiometry isotherm that can be fitted for the term K_A . Such an analysis was performed (not shown), and the resulting K_A was $428,400 \text{ M}^{-1}$ ($K_D \approx 2.33 \text{ }\mu\text{M}$). When the assumption of instantaneous interaction kinetics is violated, the EPT/MSSV stoichiometry isotherm performs nearly as well (K_A and K_D were $422,700 \text{ M}^{-1}$ and $2.36 \text{ }\mu\text{M}$, respectively). We therefore find that the kinetics of the interaction play very little role in the performance of the MSSV population isotherm or the EPT/MSSV stoichiometry isotherm. As long as the assumptions of the simulation (i.e. $s_A \gg s_B$, $s_A \approx s_{AB}$, resolvable D_{norm}) are met, an accurate estimate of K_A can be derived from the MSSV population isotherm presented here.

Theoretically, it is also possible to study the interaction of a protein and a small-molecule ligand using the MSSV population isotherm. However, there are a few factors that may limit the usefulness of such an approach. For example, the extinction properties of a small molecule may change significantly upon interacting with a macromolecule. Such a change,

³The parameters of this simulated system make it extremely unlikely that molecule A would form the undisturbed boundary. Only concentration regimes with a very large (i.e. > 50-fold) excesses of A would exhibit the boundary of free A as the undisturbed boundary. Thus, it may safely be assumed that *all* of A is in the reaction boundary. See reference [13].

if significant, would be deleterious to the MSSV analysis, which depends on invariant extinction coefficients and signal increments. Further, in simulating systems with small molecules and fast interaction kinetics, we observed an interesting phenomenon that degrades the quality of the MSSV isotherm. Protein/ligand complexes that approach the bottom of the centrifugation cell dissociate on the time-scale of sedimentation. The freed small molecules diffuse centripetally, encountering a region of radial space that has a relatively low protein concentration. With little protein to interact with the ligand, a significant gradient of back-diffused ligand can form, and the $c_k(s)$ distribution does not account for it. Thus, there are systematic differences between the data and the fit (Fig. 8). These differences result in inaccurate concentration determinations for the protein-ligand complex, leading to inaccurate K_A determinations (a 25% error was observed in the simulation that resulted in Fig. 8). As a result, MSSV isotherms for macromolecules interacting with small-molecule ligands should be approached with caution.

6 Discussion

6.1 Successful implementations

In their work introducing MSSV, Balbo *et al.* [1] described the determination of the stoichiometries two protein complexes. These systems were chosen because their stoichiometries were known. First, the interaction of peptides from SLP-76 and PLC γ 1 were studied. These two molecules interact in activated T-cells (see ref. [24] and references therein). The stoichiometry, obtained from ITC, was known to be 1:1. Application of MSSV using one absorbance wavelength and the interference optical system resulted in the determination of the correct complex stoichiometry. In their other example, two molecules of the small protein HEL were known to interact with one of a monoclonal antibody called “D1.3” [25]. The same strategy was used, i.e. data obtained using a single absorbance wavelength was combined with data simultaneously collected interferometrically to obtain $c_k(s)$ distributions demonstrating a 2:1 interaction of HEL and D1.3. Thus, the two proof-of-principle experiments worked well.

MSSV has been used to study the interaction of protein complexes that assemble at sites of receptor activation in T cells [26,27]. The adaptor protein GRB2 has both SH2 and SH3 protein-interaction domains. It binds to phosphorylated T-cell receptors via its SH2 domain. Another adaptor protein, called LAT, can also interact with GRB2 via the latter’s SH2 domain. Previous research had shown that LAT contains multiple binding sites for GRB2 [24]. Further, the SH3 domains of GRB2 allow it to bind to a protein called SOS1. Thus, there is a possibility for a very complex network of proteins at sites of receptor activation. Houtman *et al.* [27] used MSSV to confirm their calorimetric observation that two molecules of GRB2 can bind to a single SOS1 molecule (only the proline-rich domain of SOS1 was used). Also using MSSV, they found that two molecules of GRB2 bind to one LAT peptide containing the GRB2-binding sites. In a series of three-signal experiments, it was demonstrated that GRB2, a LAT peptide, and a SOS1 peptide can all exist in a single complex containing two GRB2’s, two LAT peptides, and one SOS1 peptide. Differently phosphorylated LAT peptides apparently engender the formation of higher-order, concentration-dependent complexes containing all three proteins.

In other work regarding the complex protein assemblies formed in activated T-cells, Barda-Saad *et al.* [26] used MSSV to examine the interactions between the proteins SLP-76 (which is phosphorylated after T-cell activation), Nck (an adapter protein), and VAV1 (a guanine-nucleotide exchange factor). Because complexes containing all three proteins were studied, two different chromophores were incorporated into VAV1 and SLP-76. An elegant transpeptidase-based method was used to stoichiometrically label the VAV1 protein with a fluorescein derivative, and SLP-76 peptides were labeled with a rhodamine derivative. With

the excellent spectral resolution that these materials impart on the three proteins, Barda-Saad *et al.* were able to resolve VAV1 and Nck complexes with doubly and triply phosphorylated SLP-76 peptides. Most notably, they identified a complex with molar ratio of 2:2:1 (VAV1:Nck1:SLP-76). These authors went on to study these interactions using *in vivo* fluorescence techniques, and their results are compatible with the posited formation of a complex between the three proteins.

The bacterium *Treponema pallidum* expresses a lipoprotein in its periplasm [28] called Tp34. In earlier work [29,30], a protein with similar characteristics had been shown to bind to the human mucosal iron-sequestering protein lactoferrin (hLF). Dekka *et al.* confirmed that Tp34 is binding to hLF using ITC [28]. However, the thermogram obtained from the titration had an unusual “peninsular” appearance, indicating two binding events with different K_A 's and different ΔH 's. This unexpected stoichiometry was confirmed using MSSV. The SV experiment clearly demonstrated that two Tp34's can bind to a single molecule of hLF. This stoichiometry holds true even in the presence of Zn^{2+} , which causes Tp34 to dimerize in solution [3,28].

As introduced in section 5.2, hE3 is held in the multienzyme pyruvate dehydrogenase complex (PDC) by specific, non-covalent interactions with hE3BD. In separate reports, Ciszak *et al.* [31] and Brautigam *et al.* [32] described the crystal structures of hE3 bound to hE3BD. The domain bound to the dimeric hE3 near to its dyad axis. Thus, even though there are conceivably two hE3BD-binding sites on hE3, the binding of a single hE3BD sterically excludes the binding of a second. Brautigam *et al.* confirmed this hypothesis using ITC. However, this reported stoichiometry was challenged in the literature [33]. Therefore, in a later report, Brautigam *et al.* [34] used MSSV to confirm again the reported stoichiometry. Only 1:1 complexes of hE3 and various E3BD-containing protein constructs could be observed, even when the E3BD constructs were present in large molar excesses. This latter report also demonstrated that roughly 20 hE3 molecules can bind to the PDC core (which comprises pyruvate dehydrogenase (hE1) and E3-binding protein). This result and others in that report strongly suggest that there are 20 molecules of E3-binding protein in the PDC core. This result has implications for the overall stoichiometry of the core, which is posited by Brautigam *et al.* to be 40 molecules hE1 to 20 molecules hE3[34].

MSSV has been successfully applied to determine the stoichiometry of plasminogen activator inhibitor-1 (PAI-1) to vitronectin [35]. PAI-1 is a serpin-type protease inhibitor, and its binding to vitronectin stabilizes the former in its active conformation. Vitronectin is a multifunctional glycoprotein associated with the extracellular matrix. These two proteins are normally associated in a 1:1 complex [36]; however, during certain pathological states, the concentration of PAI-1 increases. Minor *et al.* postulated that this disease state could effect the formation of 2:1 PAI-1: vitronectin complexes. They detected complexes with the apparent stoichiometries of 2:1 and 4:2. Having determined that such complexes can form, they went on to discuss the mechanism of complex formation and oligomerization.

The low-density-lipoprotein-receptor-associated protein (LRP) contains 31 complement-type repeats (CR's). These ~40 amino-acid motifs harbor disulfide bonds and a calcium-binding site. These features, coupled with the CR's propensity to bind protein ligands, apparently necessitate the presence of a chaperone called receptor-associated protein (RAP). Jensen *et al.* demonstrated with MSSV and SV that RAP binds at approximately two polypeptides containing either 2 or 3 CR motifs [37]. These results, coupled with ITC and fluorescence data, enabled Jensen *et al.* to postulate a model for the interaction of RAP with the multiple CR motifs present in LRP.

In section 5.1, we introduced the VCA-Arp2/3 system. The interaction of monomeric VCA protein constructs and the Arp2/3 complex has recently been characterized using MSSV. Padrick *et al.* showed that two VCA-containing proteins can bind to one Arp2/3 complex [3]. This observation is compatible with earlier results showing that a dimeric VCA had enhanced affinity for Arp2/3 [2]. It was also demonstrated with MSSV that the NtA domain of cortactin effectively competes with VCA at only one of the two VCA-binding sites on Arp2/3. This conclusion was reached by monitoring the stoichiometry of VCA binding to Arp2/3 as a function of the concentration of NtA added.

7 Conclusions

In this work, new pre-analysis tests of the feasibility of MSSV for given macromolecules were presented. It is suggested that the researcher interested in utilizing MSSV to determine the stoichiometry of an interaction perform these tests before embarking on the experiment. Criteria for the goodness of the spectral discrimination between two or more components were introduced. Experience has demonstrated that the lack of careful planning of an MSSV experiment can result in its failure. Also, criteria for the evaluation of a completed MSSV analysis were introduced. In addition to the determination of stoichiometry, simulation shows that MSSV can be used to assay the populations of interacting materials, suggesting that binding isotherms based upon these population measurements could be constructed and fruitfully analyzed. Finally, there are now numerous cases in which MSSV has successfully yielded information concerning stoichiometry, and this method holds the promise of significant future utility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Michael K. Rosen and Sanjay C. Panchal for the Arp2/3 and GST-VCA analytical ultracentrifugation data. This work was supported by a grant from the National Institutes of Health (NIH) (R01-GM56322) to Dr. Michael K. Rosen. S.B.P. was supported by a fellowship from the NIH (1F32-GM06917902).

Abbreviations

Arp2/3	actin related protein 2 – actin related protein 3 complex
AUC	analytical ultracentrifugation
LE	Lamm equation
MSSV	multisignal sedimentation velocity
OD	optical density
SE	sedimentation equilibrium
SV	sedimentation velocity analytical ultracentrifugation
VCA	verprolin homology – central region – acidic region
f_r	frictional ratio
r.m.s.d	root-mean-square deviation

References

1. Balbo A, Minor KH, Velikovsky CA, Mariuzza RA, Peterson CB, Schuck P. Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation. *Proc Natl Acad Sci (USA)*. 2005; 102:81–86. [PubMed: 15613487]
2. Padrick SB, Cheng HC, Ismail AM, Panchal SC, Doolittle LK, Kim S, Skehan BM, Umetani J, Brautigam CA, Leong JM, Rosen MK. Hierarchical regulation of WASP/WAVE proteins. *Mol Cell*. 2008; 32:426–438. [PubMed: 18995840]
3. Padrick SB, Deka RK, Chuang JL, Wynn RM, Chuang DT, Norgard MV, Rosen MK, Brautigam CA. Determination of protein complex stoichiometry through multisignal sedimentation velocity experiments. *Anal Biochem*. 2010; 407:89–103. [PubMed: 20667444]
4. Schuck P. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical J*. 2000; 78:1606–1619.
5. Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophysical J*. 2002; 82:1096–1111.
6. Schuck P, Demeler B. Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. *Biophysical J*. 1999; 76:2288–2296.
7. Dam J, Schuck P. Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions $c(s)$ and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophysical J*. 2005; 89:651–666.
8. Schuck P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal Biochem*. 2003; 320:104–124. [PubMed: 12895474]
9. Schuck P. Diffusion of the reaction boundary of rapidly interacting macromolecules in sedimentation velocity. *Biophysical J*. 2010; 98:2741–2751.
10. Brown, PH.; Balbo, A.; Schuck, P. *Current Protocols in Immunology*. John Wiley & Sons; 2008. Characterizing protein-protein interactions by sedimentation velocity analytical ultracentrifugation; p. 18.15.1-18.15.39.
11. Brown PH, Schuck P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophysical J*. 2006; 90:4651–4661.
12. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci*. 1995; 4:2411–2423. [PubMed: 8563639]
13. Schuck P. Sedimentation patterns of rapidly reversible protein interactions. *Biophysical J*. 2010; 98:2005–2013.
14. Pollard TD. Regulation of actin filament assembly by Arp2/3 complex and formins. *Annu Rev Biophys Biomol Struct*. 2007; 36:451–77. [PubMed: 17477841]
15. Welch MD, Mullins RD. Cellular control of actin nucleation. *Annu Rev Cell Dev Biol*. 2002; 18:247–88. [PubMed: 12142287]
16. Higgs HN, Blanchoin L, Pollard TD. Influence of the C terminus of Wiskott-Aldrich syndrome protein (WASp) and the Arp2/3 complex on actin polymerization. *Biochemistry*. 1999; 38:15212–22. [PubMed: 10563804]
17. Higgs HN, Pollard TD. Activation by Cdc42 and PIP(2) of Wiskott-Aldrich syndrome protein (WASp) stimulates actin nucleation by Arp2/3 complex. *J Cell Biol*. 2000; 150:1311–20. [PubMed: 10995437]
18. Marchand JB, Kaiser DA, Pollard TD, Higgs HN. Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. *Nat Cell Biol*. 2001; 3:76–82. [PubMed: 11146629]
19. Cole, JL.; Lary, JW.; Moody, TP.; Laue, TM. Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. In: Correia, JJ.; Detrich, HWI., editors. *Biophysical Tools for Biologists. Volume One: In Vitro Techniques*. Academic Press; 2008. p. 143-179.
20. Dam J, Velikovsky CA, Mariuzza RA, Urbanke C, Schuck P. Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions $c(s)$. *Biophysical J*. 2005; 89:619–634.
21. Brautigam CA. Using Lamm-equation modeling of sedimentation velocity data to determine the kinetic and thermodynamic properties of macromolecular interactions. *Methods this volume*. 2011

22. Zhao H. This volume.
23. Brautigam CA, Wynn RM, Chuang JL, Chuang DT. Subunit and catalytic component stoichiometries of an *in vitro* reconstituted human pyruvate dehydrogenase complex. *J Biol Chem*. 2009; 284:13086–13098. [PubMed: 19240034]
24. Samelson LE. Signal transduction mediated by the T Cell antigen receptor: the role of adapter proteins. *Annu Rev Immunol*. 2002; 20:371–394. [PubMed: 11861607]
25. Sundberg EJ, Mariuzza RA. Molecular recognition in antibody-antigen complexes. *Adv Protein Chem*. 2002; 61:119–160. [PubMed: 12461823]
26. Barda-Saad M, Shirasu N, Pauker MH, Hasan N, Perl O, Balbo A, Yamaguchi H, Houtman JCD, Appella E, Schuck P, Samelson LE. Cooperative interactions at the SLP-76 complex are critical for actin polymerization. *EMBO J*. 2010; 29 in press.
27. Houtman JCD, Yamaguchi H, Barda-Saad M, Braiman A, Bowden B, Appella E, Schuck P, Samelson LE. Oligomerization of signaling complexes by the multipoint binding of GRB2 to both LAT and SOS1. *Nat Struct Molec Biol*. 2006; 13:798–805. [PubMed: 16906159]
28. Deka RK, Brautigam CA, Tomson FL, Lumpkins SB, Tomchick DR, Machius M, Norgard MV. Crystal structure of the Tp34 (TP0971) lipoprotein of *Treponema pallidum*: implications of its metal-bound state and affinity for human lactoferrin. *J Biol Chem*. 2007; 282:5944–5958. [PubMed: 17192261]
29. Alderete JF, Peterson KM, Baseman JB. Affinities of *Treponema pallidum* for human lactoferrin and transferrin. *Genitourinary Medicine*. 1988; 64:359–363. [PubMed: 3066739]
30. Staggs TM, Greer MK, Baseman JB, Holt SC, Tryon VV. Identification of lactoferrin-binding proteins from *Treponema pallidum* and *Treponema denticola*. *Mol Microbiol*. 1994; 12:613–619. [PubMed: 7934885]
31. Ciszak EM, Makal A, Hong YS, Vettaikorumakankauv AK, Korotchikina LG, Patel MS. How dihydrolipoamide dehydrogenase-binding protein binds dihydrolipoamide dehydrogenase in the human pyruvate dehydrogenase complex. *J Biol Chem*. 2006; 281:648–655. [PubMed: 16263718]
32. Brautigam CA, Wynn RM, Chuang JL, Machius M, Tomchick DR, Chuang DT. Structural insight into interactions between dihydrolipoamide dehydrogenase (E3) and E3 binding protein of human pyruvate dehydrogenase complex. *Structure*. 2006; 14:611–621. [PubMed: 16442803]
33. Smolle M, Prior AE, Brown AE, Cooper A, Byron O, Lindsay JG. A new level of architectural complexity in the human pyruvate dehydrogenase complex. *J Biol Chem*. 2006; 281:19772–19780. [PubMed: 16679318]
34. Brautigam CA, Wynn RM, Chuang JL, Chuang DT. Subunit and catalytic component stoichiometries of an *in vitro* reconstituted human pyruvate dehydrogenase complex. *J Biol Chem*. 2009; 284:13086–13098. [PubMed: 19240034]
35. Minor KH, Schar CR, Blouse GE, Shore JD, Lawrence DA, Schuck P, Peterson CB. A mechanism for assembly of complexes of vitronectin and plasminogen activator inhibitor-1 from sedimentation velocity analysis. *J Biol Chem*. 2005; 280:28711–28720. [PubMed: 15905170]
36. Declerck PJ, De Mol M, Alessi MC, Baudner SS, Paques EP, Perissner KT, Müller-Berghaus G, Collen D. Purification and characterization of a plasminogen activator inhibitor 1 binding protein from human plasma: identification as a multimeric form of S protein (vitronectin). *J Biol Chem*. 1988; 263:15454–15461. [PubMed: 2459123]
37. Jensen JK, Dolmer K, Schar C, Gettins PGW. Receptor-associated protein (RAP) has two high-affinity binding sites for the low-density lipoprotein receptor-related protein (LRP): consequences for the chaperone functions of RAP. *Biochem J*. 2009; 421:273–282. [PubMed: 19397492]

Figure 1. Linear decomposition of optical signals with noise

To demonstrate the effects of noise on decomposition of absorbance signals into concentrations, we calculated the ratio of two components found when 5 μM of each component was converted to absorbance signal using the indicated extinction-coefficient matrices (E , indicated in panel) using Eq. 8. Synthetic, normally distributed noise with a standard deviation of 0.003 was added, and the process reversed to produce concentrations again. These concentrations were converted to an A-over-B ratio. Shown are histograms of results when the process was repeated 10,000 times each for the two different E matrices. The histograms were normalized to give an area of 1.

Figure 2. Effects of D_{norm} on MSSV analysis

Results from fitting simulated hetero-association MSSV data. (A) Using common conditions, a single signal increment was varied to produce systems with the indicated range of D_{norm} values. Simulations were repeated five times with distinct synthetic noise added (optical signal noise of 0.005 rmsd) for each system. (A) ratio of the two components present in the integrated complex sedimentation peak in $c_k(s)$ analysis (9.2 S peak). The average ratio (top) and standard deviation (bottom) for the each set of five repeats are shown. (B) Assessment of the fitting results in (A) by qualitative scoring. Conservation of mass was scored by integrating the entire $c_k(s)$ distribution for both components and comparing to the input component concentration. Integrated concentrations between 90% and 110% are scored as positive. Complex component ratios were compared to the expected value (1) and those between 0.9 and 1.1 were scored as a positive result. Bars show the number of positive results from five trials for conservation of mass (black bars) and ratio (gray bars). (C and D) Summary of simulated MSSV analysis similar to that in panel B, but for D_{norm} of 0.032 or 0.066 at different noise levels (C) or for D_{norm} of 0.032 and different acquisition frequencies (D). A pound symbol, #, indicates one analysis with more than 15% spectral contamination. The asterisk indicates the one example of an analysis which found the conservation of mass to be satisfied, spectral contamination acceptable, but the component ratio in the complex to be outside of the range of 0.9 – 1.1. In this case, conservation of mass was only barely satisfied 90.01 % of the light component was found, but the ratio determined was 1.13.

Figure 3. Choice of component concentrations for MSSV experiments

Acceptable concentrations for an example MSSV analysis were determined by analysis of concentration constraints. Constraints are cast as inequalities and plotted on a single $[A]_{\text{tot}}$ versus $[B]_{\text{tot}}$ field. (A) The inequalities given in Eq. 11–15 are overlaid. Heavy lines are the limits of the inequalities, and the labels (Eq. 11 – Eq. 15) represent the equations in the text that give rise to them, respectively. Small arrows indicate the direction that satisfies the inequality for each constraint. Gray area is the region where all inequalities are satisfied. (B) Similar to panel A, but using the inequalities given in Eq. 11–17. Parameter values used in the inequalities: $l = 1.2$ cm,

$\epsilon_{\text{ABS}}^A = 140,845 \text{ M}^{-1} \text{ cm}^{-1}$, $\epsilon_{\text{ABS}}^B = 5,000 \text{ M}^{-1} \text{ cm}^{-2}$, $\epsilon_{\text{IF}}^A = 550,000 \text{ fringes M}^{-1} \text{ cm}^{-2}$, $\epsilon_{\text{IF}}^B = 27,500 \text{ fringes M}^{-1} \text{ cm}^{-2}$, $K_D = 500 \text{ nM}$, $f_{\text{sat}} = 0.98$.

Figure 4. Multisignal fit for GST-VCA alone

(A) The fit to the IF data. In the top part, the circles represent the individual data points obtained from a centrifugation experiment containing GST-VCA alone. TI and RI noise were subtracted prior to plotting. Only every 6th data point is shown for clarity, and only every 3rd scan used in the analysis is shown for the same reason. The lines represent the fit to these data using the $c_{\text{GST-VCA}}(s)$ distribution (part C) to scale solutions to the LE, as shown in Eq. 2. The lower part is the residuals between the data and the fit. This figure establishes precedents that will be used in all figures in this paper presenting SV data. The only difference in subsequent figures is that all absorbance data will be presented with every 3rd data point instead of every 6th. (B) Data, fit, and residuals to the A_{280} data for GST-VCA alone. (C) The $c_{\text{GST-VCA}}(s)$ distribution (solid line) used to globally fit the data in parts A and B.

Figure 5. Multisignal fit for Arp2/3 alone

(A) The data, fit, and residuals for the IF data. (B) The data, fit and residuals for the A_{280} data.

(C) The $c_{\text{Arp2/3}}(s)$ distribution (dashed line) used to model the data presented in parts A and B.

Figure 6. Multisignal fit for the mixture of GST-VCA and Arp2/3

(A) The data, fit, and residuals for the IF data. (B) The data, fit and residuals for the A_{280} data. Panels (A) and (B) show only every sixth scan, as 101 scans were fit in this analysis. (C) The two distributions used to model the data in parts A and B. The solid line is the $c_{\text{GST-VCA}}(s)$ distribution, and the dashed line is the $c_{\text{Arp2/3}}(s)$ distribution. Note the cosedimentation at ~ 10.4 S. An integration of the peak at 10.4 S yields the concentration of the cosedimenting components. They are approximately equimolar (see text).

Figure 7. MSSV-based population isotherms

(A) The MSSV population isotherm analyzed alone. The circles represent the individual [AB] results obtained from MSSV analysis. The [B] cosedimenting with A was taken as [AB]. The line represents the fit to these data; only one parameter is fitted: K_A . (B) The global analysis of an MSSV population isotherm and the s_w isotherm. The circles and solid line represent the same quantities described in part A. The open triangles are the values of s_w from the analysis of the IF data alone in the simulated system, and the dashed line is the fit to these data. Both the population data and the s_w data were analyzed globally; the data sets shared a single fitted parameter: K_A .

Figure 8. Back-diffusion causing fitting problems in a quickly reacting system

For this simulation, molecule A was a protein with a molar mass of approximately 88,000 g/mol, a sedimentation coefficient of 5.5 S, and a f_r of 1.3. Molecule B was a small molecule of apparent molar mass 350 g/mol and an apparent sedimentation coefficient of 0.2 S. The AB complex had a sedimentation coefficient of 5.5 S. The two molecules associated with a K_D of 200 nM, and the k_{off} of the interaction was 10^{-1} s^{-1} . The rotor speed of the simulation was 50,000 rpm; about 8.3 hours of sedimentation was simulated. The protein absorbed in the UV range, but not in the visible (VIS) range, while the opposite was true of the small molecule. The molar extinction properties were:

$\epsilon_{UV}^A = 40,000 \text{ M}^{-1} \cdot \text{cm}^{-1}$, $\epsilon_{VIS}^B = 30,000 \text{ M}^{-1} \cdot \text{cm}^{-1}$, $\epsilon_{VIS}^A = \epsilon_{UV}^B = 0 \text{ M}^{-1} \cdot \text{cm}^{-1}$. The simulated VIS data (circles) and the fit thereto (lines) using a $c_k(s)$ distribution are shown in the upper part for the situation where $[A] = 6 \text{ } \mu\text{M}$ and $[B] = 10 \text{ } \mu\text{M}$. The lower part shows the clearly systematic residuals, and the fit is poor for late scans.

Table 1

Three-component systems modeled

Test case	D_{norm}	Species	Mass (kg/mol)	ε_F (fringes $M^{-1}cm^{-1}$)	ε_{ABS1} ($M^{-1}cm^{-1}$)	ε_{ABS2} ($M^{-1}cm^{-1}$)
#1	0.6989	A	200	550,000	0	0
		B	10	27,500	70,000	0
		C	25	68,750	15,000	80,000
#2	0.0057	A	65	178,750	45,510	13,933.5
		B	50	137,500	15,640	5,827
		C	20	55,000	12,790	6,392.5
#3	0.0027	A	65	178,750	42,660	14,499
		B	50	137,500	24,160	11,791
		C	20	55,000	17,060	6,821
#4	0.0016	A	65	178,750	69,670	25,724.5
		B	50	137,500	29,860	10,660
		C	20	55,000	17,060	6,821
#5	0.0629	A	200	550,000	140,845	0
		B	20	55,000	18,480	0
		C	12	33,000	26,040	70,000

Table 2

Results for three-component, three-signal MSSV simulations

Test case	Trial	Concentration sedimenting with free BC sedimentation coefficient			Concentration sedimenting with ABC sedimentation coefficient		
		Species A (μM)	Species B (μM)	Species C (μM)	Species A (μM)	Species B (μM)	Species C (μM)
#1	Expected	0	3	3	2	2	2
	Trial 1	0.000	3.005	3.005	2.000	1.998	1.998
	Trial 2	0.000	3.006	3.003	2.000	1.997	1.999
	Trial 3	0.000	3.008	3.004	2.000	1.994	1.999
	Trial 4	0.000	3.009	3.003	2.000	1.995	1.999
	Trial 5	0.000	3.006	3.005	2.000	1.999	1.998
#2	Expected	0	7	7	5	5	5
	Trial 1	0.109	6.753	6.463	4.942	5.241	5.501
	Trial 2	0.054	6.756	6.676	4.986	5.233	5.320
	Trial 3	0.029	6.744	6.771	5.015	5.257	5.189
	Trial 4	0.143	6.789	6.247	4.945	5.175	5.643
	Trial 5	0.132	6.823	6.422	4.926	5.175	5.494
#3	Expected	0	6	6	4	4	4
	Trial 1	0.207	5.785	5.598	3.866	4.160	4.341
	Trial 2	0.158	5.814	5.654	3.886	4.161	4.306
	Trial 3	0.145	5.807	5.710	3.880	4.180	4.282
	Trial 4	0.141	5.813	5.705	3.882	4.174	4.291
	Trial 5	0.182	5.784	5.661	3.877	4.175	4.286
#4	Expected	0	5.2	5.2	2.8	2.8	2.8
	Trial 1	0.149	5.285	4.304	2.735	2.783	3.281
	Trial 2	0.215	5.351	3.919	2.688	2.749	3.534
	Trial 3	0.247	5.369	3.770	2.661	2.731	3.657
	Trial 4	0.220	5.364	3.866	2.662	2.703	3.730
	Trial 5	0.256	5.412	3.635	2.639	2.665	3.897
#5	Expected	0	4.5	4.5	2.5	2.5	2.5

Test case	Trial	Concentration sedimenting with free BC sedimentation coefficient			Concentration sedimenting with ABC sedimentation coefficient		
		Species A (μM)	Species B (μM)	Species C (μM)	Species A (μM)	Species B (μM)	Species C (μM)
	Trial 1	0.000	4.506	4.503	2.499	2.505	2.498
	Trial 2	0.002	4.481	4.504	2.499	2.510	2.500
	Trial 3	0.000	4.505	4.502	2.500	2.494	2.500
	Trial 4	0.001	4.495	4.505	2.501	2.480	2.499
	Trial 5	0.002	4.488	4.503	2.501	2.486	2.499

Table 3

Starting Experimental Parameter Values for All Three SEDPHAT Sessions

	Experiment 1 (IF)	Experiment 2 (ABS280)
Parameter	Value	
*sqrt(N1/Nx)	unchecked (default)	checked
vbar	0.73	0.73
density	1.00079	1.00079
viscosity	0.010024	0.010024

Table 4

Extinction and signal-increment information for the simulation of Section 5.2

Protein	ϵ_{IF} (fringes $\cdot\text{M}^{-1}\cdot\text{cm}^{-1}$)	ϵ_{ABS280} ($\text{M}^{-1}\cdot\text{cm}^{-1}$)
A (hE3-like)	275,000	140,000
B (XDD1-like)	52,250	9,500