

ANALYSIS OF HETEROGENEITY IN MOLECULAR WEIGHT AND SHAPE BY ANALYTICAL ULTRACENTRIFUGATION USING PARALLEL DISTRIBUTED COMPUTING

Borries Demeler,^{*} Emre Brookes,^{*} and Luitgard Nagel-Steger[†]

Contents

1. Introduction	88
2. Methodology	89
3. Job Submission	97
4. Results	100
4.1. Example 1: Simulated five-component system	100
4.2. Example 2: Amyloid aggregation	104
5. Conclusions	109
Acknowledgments	111
References	111

Abstract

A computational approach for fitting sedimentation velocity experiments from an analytical ultracentrifuge in a model-independent fashion is presented. This chapter offers a recipe for obtaining high-resolution information for both the shape and the molecular weight distributions of complex mixtures that are heterogeneous in shape and molecular weight and provides suggestions for experimental design to optimize information content. A combination of three methods is used to find the solution most parsimonious in parameters and to verify the statistical confidence intervals of the determined parameters. A supercomputer implementation with a MySQL database back end is integrated into the UltraScan analysis software. The UltraScan LIMS Web portal is used to perform the calculations through a Web interface. The performance and limitations of the method when employed for the analysis of complex

^{*} Department of Biochemistry, The University of Texas Health Science Center at San Antonio, San Antonio, Texas

[†] Heinrich-Heine-Universität Düsseldorf, Institut für Physikalische Biologie, Düsseldorf, Germany

mixtures are demonstrated using both simulated data and experimental data characterizing amyloid aggregation.

1. INTRODUCTION

Many of today's biomedical research projects studying the molecular basis for cancer and other diseases focus on the understanding of dynamic interactions among molecules implicated in the disease process. Analytical ultracentrifugation (AUC) offers an array of powerful tools to study such interactions. AUC experiments make it possible to observe macromolecules and macromolecular assemblies in *solution*, that is, in a physiological environment unconstrained by crystal packing forces or an electron microscope grid. Systems can be studied under a wide range of concentrations and buffer conditions, and the methods are applicable to a very large range of molecular weights, extending from just a few hundred daltons to systems as large as whole virus particles. The range of applications is further extended by several detection systems, which include absorbance optics for ultraviolet (UV) and visible wavelengths, Rayleigh interference optics, fluorescence intensity optics, turbidity, and schlieren optics. In addition, new detectors are currently being developed, such as multiwavelength absorbance, small angle light scattering, and Raman spectroscopy detectors. AUC experiments, in conjunction with sophisticated numerical analysis, can yield a wealth of information about a wide range of hydrodynamic and thermodynamic properties of the macromolecules under investigation, including molecular weight, association (K_{eq}) and rate (k_{off}) constants, sedimentation coefficients (s), diffusion coefficients (D), and shape factors (f/f_0), as well as partial concentrations of individual solutes. These parameters provide insight into macromolecular organization and function, oligomerization characteristics, conformation, binding stoichiometry, and sample composition. This chapter describes new computational tools and algorithms whose goal it is to identify hydrodynamic parameters with the highest possible resolution.

During an AUC experiment, a sample of interest is dissolved in a buffer, and the solution is placed into a sector-shaped cell and sedimented in a centrifugal force field. The centrifugal force is adjusted according to the size of the molecules under study and can be as high as 260,000 g , permitting resolution of components with a broad molecular weight range. During the experiment, differently sized and shaped components in the sample will sediment away from the rotor center at rates proportional to their molecular weight divided by their shape, creating a moving boundary with a distinct concentration profile that is dependent on s and D of each solute. Depending on optical system, data are collected every few seconds or minutes by monitoring the change in total concentration over the entire radial domain. An example of a multicomponent data set obtained with UV absorption

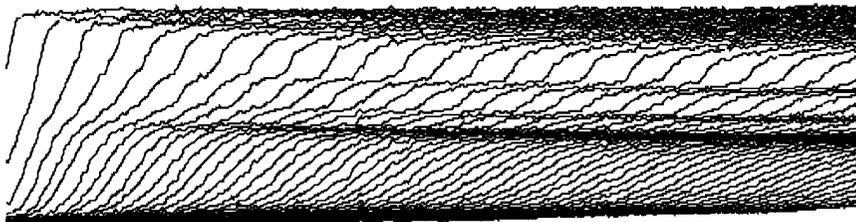


Figure 4.1 Experimental AUC data of a multicomponent mixture. The x axis reflects the radius, and the y axis represents the relative concentration. The direction of sedimentation is from left to right; each trace represents a single time point in the experiment. In this example, multiple components can be distinguished by mere visual inspection of the profile.

optics is shown in Fig. 4.1. Each component in the mixture contributes a partial concentration to the observed concentration profile. In order to analyze these data, one has to identify the contributions of individual components to the overall concentration profile. This process involves modeling the partial concentration, the sedimentation, and the diffusion transport of each solute over time and requires proper accounting of noise contributions. In the case of reacting systems, such as reversibly self- or hetero associating systems, equilibrium and rate constants need to be considered as well. The sedimentation and diffusion transport of each solute is described by a partial differential equation (PDE), the Lamm equation (Lamm, 1929), which can be solved at high resolution using an adaptive space-time finite element approach for either the noninteracting (Cao and Demeler, 2005) or the interacting (Cao and Demeler, 2008) case. Linear combinations of Lamm equation solutions are then used to approximate the entire concentration profile. Finding the correct parameter combinations of partial concentration, s and D , for each solute is accomplished by solving the inverse problem of fitting simulated finite element solutions to experimental data. To this end, we have developed new algorithms that allow us to model such experiments at the highest possible resolution. These algorithms are based on first-principle biophysical descriptions and employ advanced numerical techniques and parallel computing to accomplish this goal. Our developments have been integrated into an open-source software package called UltraScan (Demeler, 2005, 2008), which contains a comprehensive range of tools to help interpret experimental data and derive aforementioned parameters from AUC experiments.

2. METHODOLOGY

The computational task of analyzing experimental data by building a model that best represents the experimental information can be separated into four phases: (1) initialization—determining the appropriate parameter

range to be searched; (2) two-dimensional spectrum analysis (2DSA)—calculating a linear combination of basis functions that covers the parameter space with subsequent linear least-squares fitting, with simultaneous elimination of systematic noise; (3) refinement—parsimonious regularization by genetic algorithm analysis (GA) or optionally nonlinear fitting with a discrete model; and (4) Monte Carlo analysis (MC)—statistical evaluation of the results and attenuation of stochastic noise contributions. In addition, a careful design of the experiment is an important consideration for the success of the experiment. A large range of different approaches exist to evaluate experimental sedimentation data, each with its own advantages and limitations on information content. Of those, model-independent approaches such as the van Holde–Weischet analysis (Demeler and van Holde, 2004) and the dC/dt method (Stafford, 1992) are preferable for the initialization steps; subsequently, direct boundary modeling by finite element solutions of the Lamm equation is desirable because unlike model-independent methods that provide only s -value distributions and, in some cases, partial concentrations, finite element solutions permit simultaneous determination of the sedimentation and diffusion coefficients of each species, as well as determination of partial concentrations. When s and D are available, additional information can be derived, including molecular weight and frictional parameters by applying the Svedberg equation, which relates s and D to the molecular weight of the particle [see Eq. (4.1), where R is the gas constant, T is the temperature, ρ is the density of the solvent, and s , D , M , and \bar{v} are the sedimentation and diffusion coefficients, the molecular weight, and the partial specific volume of the solute, respectively]. Once s and D have been determined, a frictional ratio, f/f_0 , can be calculated as well according to Eq. (4.2). This ratio provides a convenient parameterization for the shape of the molecule. The lower limit of 1.0 can be interpreted as a spherical molecule, whereas values greater than 1 indicate increasing nonglobularity. Values up to 1.3 are common for mostly globular proteins, whereas values between 1.8 and 2.5 are consistent with elongated, denatured, or intrinsically disordered proteins. Values larger than 2.5 can be found for very long molecules such as linear DNA fragments or long fibrils. Solving the inverse problem of fitting a model to the experimental data can be accomplished with a nonlinear least-squares fitting routine to arrive at the best-fit parameter set for the nonlinear parameters to be searched. However, this approach is suitable only for simple cases, where at most one or two solutes are present. The reason for the failure of this approach with a larger number of species is related to the complexity of the error surface, which increases when an increasing number of solutes and parameters is modeled and causes the optimization algorithm often to stall in a local minimum, preventing convergence at the global minimum. This chapter describes three alternative approaches addressing this issue, and each approach provides a complementary description of the solution space.

These approaches can be linked to provide an optimal description of data. Before detailing our approach, we should mention the importance of high-quality data. No amount of sophisticated analysis can compensate for poor quality of primary data and all efforts should be taken to eliminate unnecessary noise from data. The precision of parameter estimation is inversely correlated with the experimental noise present in primary data. It is therefore important that systematic noise contributions resulting from instrument flaws are accounted for and that stochastic noise contributions are attenuated. It has been shown previously that systematic noise contributions such as time- and radially invariant noise can be eliminated effectively using algebraic means (Schuck and Demeler, 1999) and that the effect of stochastic noise contributions can be reduced using MC methods (Demeler and Brookes, 2008). Experimental design considerations can further improve noise characteristics, for example, by using intensity measurements instead of absorbance measurements, stochastic noise is reduced by a factor of $\sqrt{2}$ by not subtracting the reference signal. This subtraction leads to the convolution of two stochastic noise vectors and an increase in the stochastic noise:

$$\frac{s}{D} = \frac{M(1 - \bar{v}\rho)}{RT} \quad (4.1)$$

$$\frac{f}{f_0} = \frac{RT}{3D\eta(6N^2\pi^2M\bar{v})^{\frac{1}{3}}} \quad (4.2)$$

Step 1. Selecting the appropriate parameter space. The complexity of the evaluation can be reduced if a subspace of reasonable parameter values can be obtained through model-independent approaches. The enhanced van Holde–Weischet method (Demeler and van Holde, 2004) is ideally suited for this purpose, as it provides diffusion-corrected s -value distributions from sedimentation velocity experiments. The diffusion coefficient, which is required for the solution of the Lamm equation, can be initially estimated by parameterizing the shape function using the frictional ratio, f/f_0 , which is a measure of the globularity of a particle [Eq. (4.3), where N is Avogadro's number, k is the frictional ratio f/f_0 , η is the viscosity of the solvent, and all other symbols are the same as in Eq. (4.1)]. A reasonable assumption can be made that the shape of the particle ranges somewhere between spherical ($f/f_0 = 1.0$) and rod shaped ($f/f_0 \leq 4.0$) for most solutes. Given the limits from the s -value range determined with either the van Holde–Weischet or the dC/dt method, and the assumption on particle shape, it is now possible to define the limits of a two-dimensional parameter space over s and f/f_0 :

$$D = RT \left[N18\pi(k\eta)^{3/2} \left(\frac{s\bar{v}}{2(1 - \bar{v}\rho)} \right)^{1/2} \right]^{-1} \quad (4.3)$$

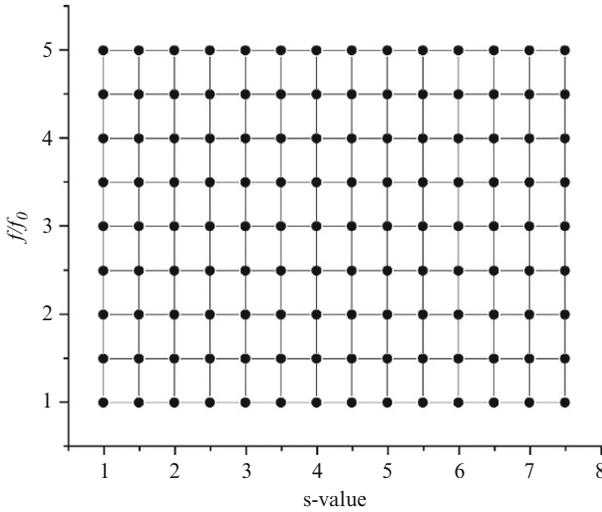


Figure 4.2 Two-dimensional grid over s and f/f_0 . Each node point represents a term in the linear model, whose amplitude is determined by the 2DSA analysis through least-squares fit using NNLS.

Step 2. Evaluating the basis functions over all parameters and eliminating systematic noise. A two-dimensional grid with discrete s and f/f_0 parameter combinations covering the s -value range and the f/f_0 range is constructed. Such a grid is shown in Fig. 4.2. Here, each grid point represents a complete finite element simulation over both space and time, as well as a solute in the delimited parameter space. The solutes at each grid point are simulated at unity concentration and their contributions are summed to obtain the final concentration profile. The amplitudes of each term in the sum represents the relative contribution of each solute. The fitting problem is thus reduced from a nonlinear fitting problem to a linear least-squares approach [Eq. (4.4)] that only requires determination of the amplitudes of each grid point. In this approach, \mathbf{M} is the model [Equation (4.5)], which is compared to experimental data \mathbf{b} over all time and radius points r and t . $c_{l,m}$ represents the amplitude of each grid point in Fig. 4.2, and L is the solution of the Lamm equation for a single, nonreacting solute (Brookes and Demeler, 2006; Cao and Demeler, 2005). The fit is accomplished with the nonnegatively constrained least-squares fitting algorithm called “NNLS” (Lawson and Hanson, 1974). This approach will result in positive amplitudes only, or zero, if a component is not present in experimental data. There are several considerations to be made relating to computer memory requirements. The finite element solution at each grid point in Fig. 4.2 requires approximately 800–1000 radial points and 50–500 time points, depending on the experiment. It is clear that fitting a high-resolution grid will require excessive

amounts of memory. To circumvent this significant problem, we employed a divide-and-conquer approach, termed the multistage two-dimensional spectrum analysis, which repeatedly moves the initial low-resolution grid by small increments Δs and $\Delta f/f_0$ until the entire parameter space is covered. This approach accomplishes a higher resolution analysis without overwhelming available memory and has been shown to provide a serial speedup over standard NNLS and parallelizes with no excess computation and negligible communications overhead (Brookes *et al.*, 2006). Scalability testing of 2DSA has shown a linear speedup from 4 to 512 processors. We formulated an equation for processor utilization (Brookes *et al.*, 2006), which predicts the optimal number of processors to guarantee a minimum processor utilization for a specific problem size. We currently use this method of computing the number of processors required for a grid job submission by targeting 80% processor utilization for problems without time-invariant noise and 70% processor utilization for better speedup on the more computationally intensive time-invariant noise calculation. As a consequence, our implementation achieves the best quality of results without wasting computational resources on a large cluster. Since the solution is sparse, only a few parameters are returned from a coarse grid. Typically, we apply 100–300 grid movings of a 10×10 grid to obtain a resolution that is commensurate with the resolution of the analytical ultracentrifuge. Fewer grid movings can be used if the van Holde–Weischet analysis reports a narrow range. Solutes with positive amplitudes from different grids are then unioned with each other to form new grids with a maximum number of solutes equivalent to that of a single initial grid (generally less than 100 solutes). Each unioned grid represents a single stage in the multistage process and is refitted until all grids have been successively unioned into a single grid. An iterative variation of this algorithm unions the final grid with all initial grids and repeats the 2DSA analysis until the solution is converged. A stable solution can be reached for cases where time- and radially invariant noise is not considered, which is equivalent to performing the solution containing all grids in a single iteration. For cases where invariant noise components are calculated, three to four iterations generally converge to a mostly stable solution. Our approach is parallelized with MPI on the level of each grid calculation, and the relevant UltraScan modules have been deployed on the NSF TeraGrid and on TIGRE [the Texas Internet Grid for Research and Education project is a computational grid that integrates computing systems, storage systems and databases, visualization laboratories, displays, instruments, and sensors across Texas (<http://www.hipcat.net/Projects/tigre>)] sites (Vadapalli *et al.*, 2007). Simultaneously to the parameter determination, we use the 2DSA to algebraically extract the systematic noise contributions that result from imperfections in the optical system or instrument (Schuck and Demeler, 1999). After convergence, the vector of time-invariant noise contributions is subtracted from experimental data,

yielding a data set only perturbed by Gaussian random noise. At that point, a subset of the initial two-dimensional parameter space is obtained, and relative concentrations for all solutes have been determined. Although the solution provides a good qualitative representation of experimental data, the solution cannot be regarded as unique at this point because it is overdetermined and subject to degeneracy. In this solution, low-frequency false positives are common—they result from the random noise present in experimental data and because the true solutes are not necessarily aligned with the fitted grid. Additional processing is required to further refine the 2DSA solution:

$$\text{Min} \sum_{i=1}^r \sum_{j=1}^t [M_{ij} - b_{ij}]^2 \quad (4.4)$$

$$\mathbf{M} = \sum_{l=s_{\min}}^{s_{\max}} \sum_{m=f_{\min}}^{f_{\max}} c_{l,m} L(s_l, D(s_l, k_m)). \quad (4.5)$$

Step 3. Parameter refinement and regularization. We found that parameter refinement of values obtained with the 2DSA is best achieved with a genetic algorithm implementation (Brookes and Demeler, 2006). We can adapt Occam’s razor for our problem, which can be stated as follows: the most parsimonious solution capable of producing nearly the same residual mean square deviation (RMSD) is the preferred solution. Because the solution obtained in Step 2 is overdetermined and not unique we need to determine the most parsimonious solution. Implementation of the genetic algorithm analysis is described in detail in Brookes and Demeler (2006). Briefly, an evolutionary process is used to optimize the solution from a pool of multiple, individual solutions. The individual solutions are derived from the 2DSA analysis. Here, each new solution contains exactly one solute for each solute determined previously in the 2DSA analysis. Each solute is initialized randomly with an s and f/f_0 pair drawn from a symmetrical, rectangular region defined by a small, user-selected Δs and $\Delta f/f_0$ range surrounding each solute determined in the 2DSA analysis. If two regions overlap, one or two new ranges are defined in which a new, additional solute is placed (see Fig. 4.3A). All new solutions are then allowed to evolve to the best-fitting parameter combination according to the rules defined in the GA. Random number generators perform mutation, crossover, deletion, and insertion operations on the parameter combinations in order to generate new individuals for the next generation. The overriding selection criterion is the RMSD of a particular solution. Several rules favor survival of the fittest solution. To guard against loss of parameter diversity, we coevolve multiple demes, each consisting typically of 100 individual solutions, and permit only limited parameter migration between demes. Regularization is

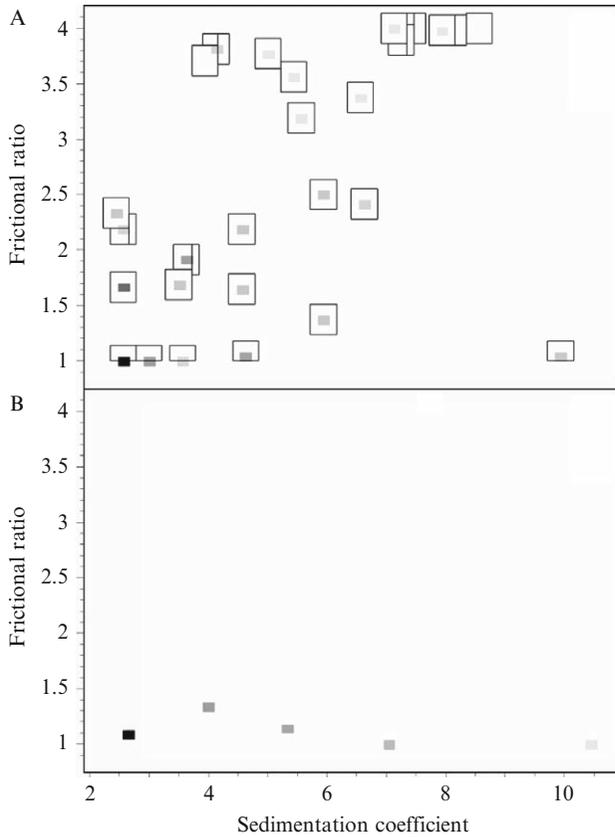


Figure 4.3 GA initialization (A) and GA analysis result (B). (A) Initializing regions formed around each solute determined in the 2DSA analysis. Each solute is shown by a rectangle drawn in a gray level corresponding to the partial concentration of the solute. Regions are clipped at a frictional ratio equal to unity. Overlapping regions are subdivided and a new GA solute is placed into the subdivided region. (B) GA analysis results from the initialization shown in A, indicating not only a successful parsimonious regularization, but also a more consistent tendency of frictional ratios.

achieved by penalizing the fitness of a particular solution in direct proportion to the number of solutes represented in this solution (Brookes and Demeler, 2007). In contrast to Tikhonov and maximum entropy regularization, we term this approach “parsimonious regularization” because of its ability to find the smallest number of necessary, discrete solutes to describe a given experimental system (see Fig. 4.3B). MPI is used to parallelize on the level of the deme calculation, with each deme occupying a different processor. This approach is embarrassingly parallel and scales linearly on all platforms. Individual generations progress asynchronously to prevent processor idling. Furthermore, we can choose to terminate evolution early

when no further improvement in solutions is occurring. Importantly, we note that the GA analysis without 2DSA initialization will achieve the same result, although at much higher computational cost. Our strategy of linking 2DSA with GA has reduced the overall computational task by over 200-fold for the inverse problem of AUC experimental data fitting by drastically reducing the parameter space to interesting regions only and thus simplifying the parameter search.

Step 4. Statistical evaluation of the results. After Step 3, a discrete, parsimonious solution is obtained that represents the best fit to experimental data without any sacrifice in goodness of fit. Any uncertainty associated with the determined parameters is solely a result of Gaussian noise in the experimental observations. In order to determine the confidence interval on each parameter, a Monte Carlo analysis is performed. In this approach, synthetic noise of the same quality as observed in original experimental data is added to the best-fit solution from Step 3, and new data are fitted again with the procedure outlined in Step 3. This is typically repeated 50–100 times or until a statistically relevant description of the solution space is achieved. Each time, a slightly different result is generated, resulting in a distribution of parameter values for each parameter. These distributions can now be used to derive standard deviations for all parameters determined in Step 3, and a statistically reliable confidence interval can be assigned to the s and D values, molecular weights, frictional ratios, and partial concentrations from each solute identified in the previous steps. Monte Carlo analysis also achieves a second goal: When a stochastic signal is added to a solution, the amplification of the noise signal proceeds with a factor of $\sqrt{2}$ while the intrinsic signal of the sample contained in the analyzed system is amplified linearly. As a result, the Monte Carlo analysis attenuates the effect of stochastic noise on the solution, which can be quite apparent when the 2DSA analysis is performed at high resolution, where the analysis often finds false positives at the high end of the frictional ratio spectrum. An additional improvement in results can be obtained if signals from a high-speed and low-speed measurement are combined and the same sample is analyzed in a global fashion by combining multiple speeds in the GA analysis. In the high-speed experiment, a maximum signal from the sedimentation coefficient is obtained due to the large centrifugal force. However, sedimentation is rapid and the time allowed for diffusional boundary spreading is minimal. As a consequence, sedimentation coefficients in a high-speed experiment can be determined with high precision, whereas diffusion coefficients are often unreliable. Because of the relationships among s , D , and molecular weight shown in Eqs. (4.1) and (4.2), any lack of precision in the diffusion coefficient translates into a lack of precision of molecular weight and shape. By measuring the same sample also under low-speed conditions, sedimentation is much slower, and the sample has sufficient time to diffuse before being pelleted at the bottom of the cell. This provides a much improved

signal on the diffusion transport and, as a consequence, on the accuracy with which the shape of the solute can be determined. Constraining s to the 95% confidence region of the sedimentation coefficient from each solute determined in the high-speed experiment, it is now possible to converge only on the diffusion coefficients using low-speed data.

3. JOB SUBMISSION

Our methods have been implemented on a parallel computing platform in the UltraScan software package (Demeler, 2008). We have developed modules that allow submission to compute resources. These resources are local or remote clusters and grid-based supercomputers. High-performance computers are generally dedicated to specific jobs and operate in a batch mode maintained by a queue mechanism. We use the Globus-based TIGRE software stack (Vadapalli *et al.*, 2007), a grid middleware environment developed by HIPCAT (Consortium for High Performance Computing across Texas, <http://www.hipcat.net>), to communicate with the various compute resources. The 2DSA and GA processes are submitted by the user through a Web interface to the grid. Monte Carlo analysis can be added to each method. Analysis is performed on a target cluster and when the job completes, the results are e-mailed back to the researcher. Our submission scheme is shown in Fig. 4.4. Submission is performed from a Web page shown in Fig. 4.5. When the user submits a job, the Web server sends the user's request to `us_gridpipe`, which is a *named pipe*. This is a special type of file that simply holds written data until they are read. The PERL (Wall *et al.*, 2000) script `us_gridpipe.pl` *daemon*, a program that is always running in the background, reads from `us_gridpipe`, manages a global list of jobs, and controls startup of TIGRE jobs. Upon receipt of the researcher's job request, the `us_gridpipe.pl` daemon will first execute `us_gridcontrol`, a C++ program, to collect experimental data from the LIMS database in preparation for job execution. When `us_gridcontrol` completes, it informs the `us_gridpipe.pl` daemon via the named pipe that all of the experimental data have been extracted from the database and placed into a file on the disk. The `us_gridpipe.pl` daemon inserts the job request into its list of TIGRE jobs and begins execution of the PERL script `us_tigre_job.pl`, which controls the job execution. TIGRE resources are shared and it is important to select the number of processors carefully. The authors have developed a formula to compute the optimal number of processors to achieve a specific processor utilization (Brookes *et al.*, 2006) and this computation is performed for TIGRE jobs. Once the number of processors is known, `us_tigre_job.pl` sends experimental data and user's Web request parameters to the user-selected cluster and submits the job to

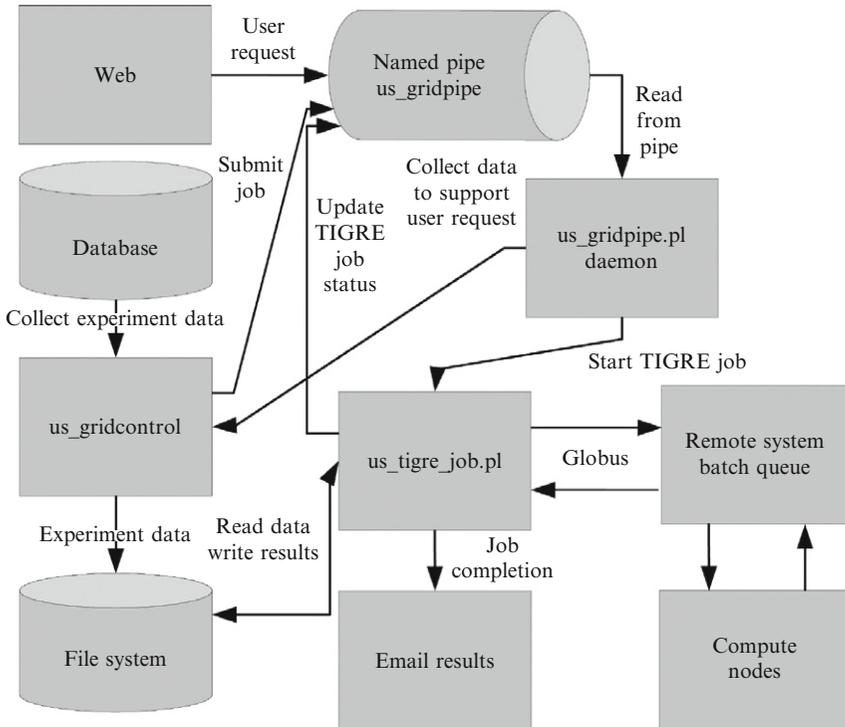


Figure 4.4 A flow diagram of the mechanism for submission of parallel analysis jobs from the USLIMS Web portal. The mechanism links Web server, database, local and remote file systems, e-mail, and remote batch queues operating on multiple supercomputers. The function of each component is explained in the text.

the cluster's queue. The execution scheduling is then controlled by the target cluster utilizing cluster-dependent queue control software. The cluster's queue control mechanism may be PBS (Portable Batch System), LSF (Load Sharing Facility), or some other queue control software. The Globus component of the TIGRE middleware hides the differences between the cluster's queue control specifics by providing a uniform application interface. The TIGRE job is monitored until completion. Upon completion, `us_tigre_job.pl` retrieves result data from the target cluster, e-mails the results to the researcher, and informs the `us_gridpipe.pl` daemon that the TIGRE job is finished. The `us_tigre_job.pl` script also collects all run-time statistics and stores them in a database. At this point, `us_gridpipe.pl` deletes the completed job from its job list and `us_tigre_job.pl` exits. The `us_gridpipe.pl` daemon also accepts requests to obtain information about the job list, which is available for viewing directly from the Web interface. The user imports the e-mailed results into the UltraScan software where the model can be visualized in 2D and 3D by a C++ GUI module of UltraScan,

UltraScan Public LIMS Portal - Mozilla Firefox

Center for Analytical Ultracentrifugation of Macromolecular Assemblies (CAUMA)
A Component Laboratory of the UTHSCSA Center for Macromolecular Interactions

Home Background DB Access HPC Access Resources User Area Related

Queue Email: demeler@biochem.uthscsa.edu [edit] | User Area

Initialize 2DSA Parameters

S-Value Setup

1	S-Value Minimum
10	S-Value Maximum
10	S-Value Resolution

f/f0 Setup

1	f/f0 minimum
4	f/f0 maximum
10	f/f0 resolution

Uniform Grid Repetitions Setup

6	Uniform Grid Repetitions
---	--------------------------

Monte Carlo Iterations

Value: 0 Minimum: 0 Maximum: 100

Fit Time Invariant Noise

On
 Off

Show Advanced Options

Modify Queue Change Analysis Selection Finish

Select TIGRE Cluster

Cluster	Status	High Priority Queue running/queued	Low Priority Queue running/queued
<input type="radio"/> lonestar.tacc.utexas.edu	up	*	*
<input type="radio"/> eldorado.acrl.uh.edu	down	*	*
<input type="radio"/> antaeus.hpc.ttu.edu	up	*	*
<input type="radio"/> ccsmos.tamu.edu	down	*	*
<input type="radio"/> bcf.uthscsa.edu	up	0/0	0/0
<input type="radio"/> laredo.uthscsa.edu	up	0/0	0/0
<input type="radio"/> meta	up	TESTING	*

Submit via TIGRE

Figure 4.5 USLIMS job submission Web page for 2DSA analyses. The user selects analysis parameters, such as s -value range, f/f_0 range, number of grid movings, and Monte Carlo iterations, and selects systematic noise correction options. Under advanced options, regularization, radially invariant noise, iterative fitting, and meniscus fitting can be selected. At the bottom, a list of clusters is offered for submission of a job, and relevant system status information is provided.

and residuals and systematic noise contributions can be displayed and processed (see Fig. 4.6). All required parallelization modules are available for the Linux operating system and can be downloaded for free from the UltraScan Web site (<http://www.ultrascan.uthscsa.edu>).

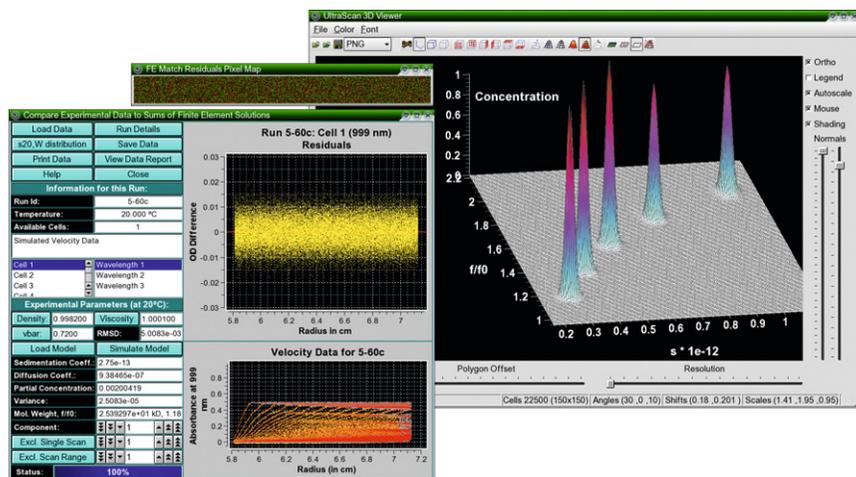


Figure 4.6 Visualization tools from the UltraScan GUI used to display models obtained in the supercomputer calculation, showing residual bitmap, residuals, experimental data and model overlay, and 3D solute distribution (f/f_0 vs s). The model shown here represents the global GA analysis of Example 1.

4. RESULTS

Using simulated and experimental example systems, we demonstrate here the capability of our method to resolve heterogeneity in mass and shape for complex systems and then explain additional insights gained from these improved methods. The first example shows a simulated data set with noise equivalent to noise produced in a well-maintained Beckman Optima XL-A instrument. Using simulated data allows us to determine the reliability of the method by comparing the fitting results with known target values used for the simulation. The second example explores the ability of our methodology to characterize the heterogeneity observed in the aggregation of amyloid-forming proteins and use it to detect changes in shape and mass induced by a ligand thought to interrupt the formation of larger amyloid aggregates.

4.1. Example 1: Simulated five-component system

This example presents the result of a four-step analysis on a simulated aggregating five-component system of a 25-kDa monomeric protein that is oligomerizing irreversibly up to a hexadecameric association state. During oligomerization, the protein changes frictional ratio. The target solute properties are listed in [Table 4.1](#), the molecular masses of the simulated

Table 4.1 Monte Carlo results from a global genetic algorithm optimization using multispeed data^a

Solute	Molecular mass (kDa)	Partial concentration	Frictional ratio, f/f_0
1	25.08 (24.75, 25.21) [25]	0.0994 (0.0980, 0.101) [0.1]	1.19 (1.187, 1.206) [1.2]
2	49.54 (49.08, 50.58) [50]	0.100 (0.987, 0.102) [0.1]	1.39 (1.386, 1.408) [1.4]
3	100.7 (99.28, 102.3) [100]	0.102 (0.0983, 0.102) [0.1]	1.61 (1.59, 1.62) [1.6]
4	204.5 (196.8, 207.4) [200]	0.0992 (0.0981, 0.100) [0.1]	1.83 (1.78, 1.84) [1.8]
5	399.1 (387.7, 409.3) [400]	0.100 (0.0998, 0.101) [0.1]	2.00 (1.96, 2.03) [2.0]

^a Results demonstrate remarkable agreement with the original target model. Parentheses: 95% confidence intervals; square brackets: target value. All values are rounded off to three or four significant digits. In all cases, target values fall within the 95% confidence intervals of the predicted values.

solutes are 25, 50, 100, 200, and 400 kDa, and the corresponding frictional ratios are 1.2, 1.4, 1.6, 1.8, and 2.0, simulating an end-to-end aggregation event. Initially, the system is simulated at 60 krpm with 60 scans equally spaced over 4 h and is simulated over a 1.4-cm column length. The same system is then simulated at 20 krpm for 30 scans equally spaced over 23.3 h. Resulting data are fitted to ASTFEM solutions of the Lamm equation (Cao and Demeler, 2005) using first the 2DSA (see Fig. 4.7), which was initialized with the enhanced van Holde–Weischet analysis (Demeler and van Holde, 2004) (suggesting a fitting range of 2–12 s, data not shown), and frictional ratios were selected to range between 1 and 3. The 2DSA analysis for both experiments resulted in a solute distribution where the concentration signals roughly mapped out the region of the target values, indicating a system with 24 (60 krpm) and 21 (20 krpm) discrete solutes at different concentrations ranging between 2.5 and 11.5 s and frictional ratios ranging between 1 and 3. The four- to fivefold excess of the number of solutes observed in the 2DSA analysis is a consequence of low-concentration stochastic noise contributions and from the possible lack of alignment of the target values with the discrete grid. In addition, the fitting system is overdetermined, and hence a unique solution is unlikely. Results for high- and low-speed data differed as follows. First, for 60-krpm data, results indicated a large spread of frictional values, whereas 20-krpm data showed a much more narrow frictional range. This can be attributed to the lack of diffusion signal in the high-speed experiment, which is needed to get

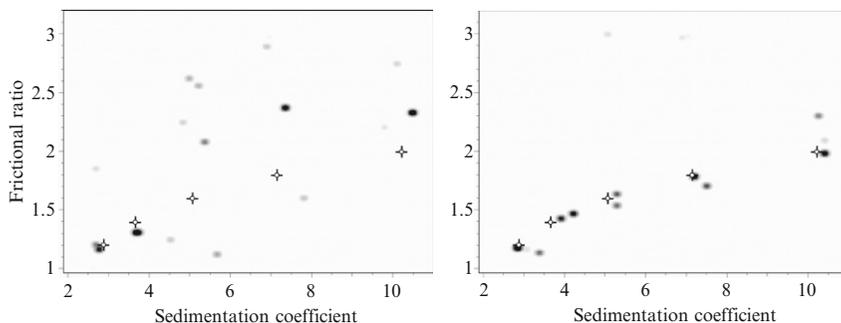


Figure 4.7 2DSA analysis of 60-krpm data (left) and 20-krpm data (right). High-speed data show more frictional ratio variation than low-speed data due to lack of a diffusion signal. Because of stochastic noise and because solutes do not necessarily align with target values, false positives are prevalent. Crosses indicate known target values, and gray levels indicate relative concentration (black: most concentrated).

accurate shape information. However, for high-speed data, a better separation resulted in more accurate determination of partial concentrations, as individual solutes were better separated and diffusing boundaries did not overlap as much. In the next step, output from the 2DSA analysis is used to initialize a GA analysis in order to obtain a parsimonious regularization (Brookes and Demeler, 2007). Results of the GA analysis are shown in Fig. 4.8. Most notably, both low- and high-speed experiments are able to correctly identify the number of target components by employing the parsimonious regularization [reducing the number of solutes from 24 (60 krpm) or 21 (20 krpm) solutes in the 2DSA to 5 in the GA]. Furthermore, the same observations regarding frictional ratio range and partial concentration made in the 2DSA analysis again apply in the GA analysis. Because the GA analysis is not restricted to a fixed grid, which may not necessarily align with the true target values, the GA analysis also comes much closer to the target values without any increase in RMSD, despite fewer parameters. Any deviation of results from the target values at this point is caused by stochastic noise in data (assuming all systematic noise sources have been eliminated). Hence, a Monte Carlo analysis will be able to map out the range of possible parameter values. To improve the results further, the signal from more precise s values and partial concentration derived from the high-speed experiment can be combined with the improved diffusion signal obtained in the low-speed experiment by globally fitting both speeds to a single model using the GA analysis. In such a global fit, sedimentation coefficients and partial concentrations will be constrained by the high-speed experiment, whereas the diffusion signal from the low-speed experiment constrains the frictional ratio range, generating optimal results for a system heterogeneous in shape and molecular weight. Results

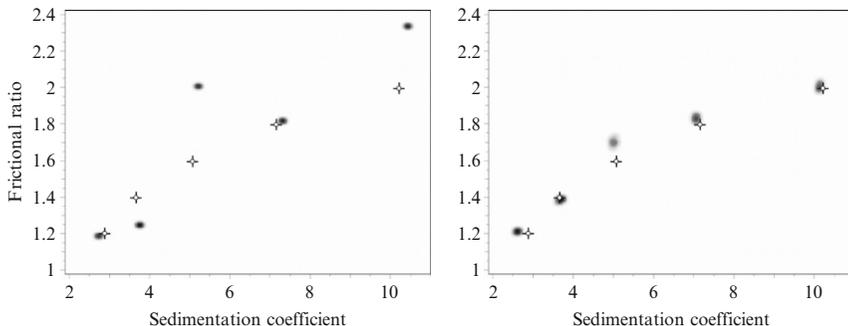


Figure 4.8 GA analysis of 60-krpm data (left) and 20-krpm data (right). As in the 2DSA, high-speed data show more frictional ratio variation than low-speed data due to lack of a diffusion signal, but reproduce the partial concentration better due to better separation. Because of stochastic noise, solutes do not necessarily align with target values, but for both speeds, parsimonious regularization achieved the correct number of total solutes. Crosses indicate known target values, and gray levels indicate relative concentration (black: most concentrated).

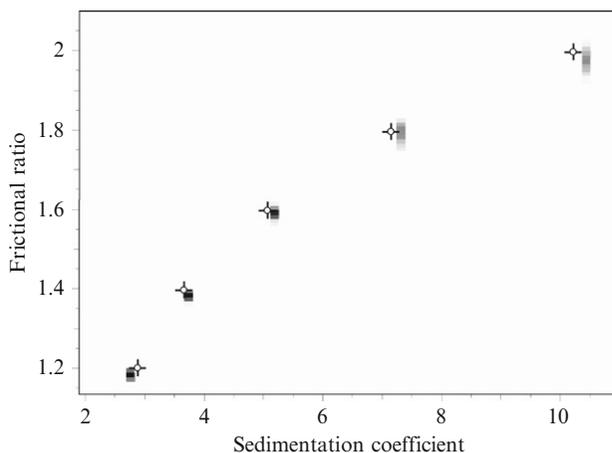


Figure 4.9 Global GA-Monte Carlo analysis for 60- and 20-krpm data. Crosses indicate the target solute position. As can be seen from this plot, target values are reproduced very closely and even partial concentrations are well matched. Gray levels indicate partial concentration (black: highest concentration).

for a global GA-Monte Carlo analysis of 60- and 20-krpm data are shown in Fig. 4.9. As can be seen here, a very faithful reproduction of the original target values is achieved in the global fit, and the combination of low- and high-speed information leads to a better definition of the shape function, as well as the partial concentration profile. For each solute, the final partial concentrations, molecular weights, and frictional ratios, as well as each

parameter's 95% confidence interval, are summarized in [Table 4.1](#). For the global multispeed Monte Carlo analysis, we find that all parameters are determined correctly within the 95% confidence interval.

4.2. Example 2: Amyloid aggregation

This example presents sedimentation studies on the Alzheimer's disease-related amyloid- β peptide. Amyloid formation ([Iqbal *et al.*, 2001](#)) is under intense scrutiny because of the role it appears to play in common neurodegenerative diseases such as Alzheimer's and Parkinson's disease ([Selkoe, 2003](#)). Amyloidogenic protein aggregates are characterized by a cross β -sheet structure and the formation of fibrils. These fibrils ([Hardy and Higgins, 1992](#)) and, to an increasing extent, also oligomeric species ([Barghorn *et al.*, 2005](#); [Klein *et al.*, 2004](#)) are discussed as potential causes for disease formation. Current effort focuses on the characterization of the mechanism causing pathologic protein aggregation. In the case of Alzheimer's disease, the aggregating entity leading to formation of the characteristic amyloid plaques detectable postmortem in the brain tissue is a proteolytic fragment of the amyloid precursor protein. Aggregation of this 39 to 42 amino acid long fragment leads to the formation of fibrils with a typical width of about 10 nm ([Holm Nielsen *et al.*, 1999](#)) and up to several hundred nanometers in length. In addition to fibrillar structures, typically obtained as an end point in incubation studies of disease-associated amyloidogenic proteins, globular or amorphous oligomeric aggregates, ranging from dimers to several hundred-mers, are being considered as entities involved in the disease process ([Rochet and Lansbury, 2000](#)). Oligomeric species that differ in size and shape and exhibit neurotoxic properties have been characterized ([Klein *et al.*, 2004](#)). Several therapeutic strategies, as well as diagnostic methods, try to target the protein aggregates directly ([Spencer *et al.*, 2007](#)). In this context, the need for methods suitable to monitor the aggregate size and shape distributions of protein solutions is evident, and our methods promise to provide insight into the aggregation mechanism by following mass and shape changes.

4.2.1. Sample preparation

The amyloid- β peptide (1–42) is from Bachem (Bubendorf, Switzerland) and is dissolved in 2 mM NaOH ([Fezoui *et al.*, 2000](#)). After freeze-drying, aliquots are stored at -70° until use. The amyloid- β peptide (1–42) labeled on the N terminus with fluorescent dye Oregon Green (A β (1–42)-OG) is synthesized by P. Henklein (Charite, Berlin, Germany). It is dissolved in anhydrous dimethyl sulfoxide (DMSO) and stored in aliquots at -70° . The inhibitor ligand ([Kirsten and Schrader, 1997](#); [Rzepecki, *et al.*, 2004](#)) is synthesized by the group of T. Schrader (Organic Chemistry Department, University Duisburg-Essen, Germany). A 5 mM stock solution is stored at 4°

in 100% DMSO. In order to reduce aggregate formation, protein concentrations are kept as low as possible, and low salt conditions are chosen. Aggregation mixtures containing 17.5 μM A β (1–42) and 3.5 μM A β (1–42)-OG are dissolved in 10 mM sodium phosphate, pH 7.4, and 4% anhydrous DMSO. For inhibitor studies, 200 μM inhibitor is added to the sample. Prior to sedimentation velocity centrifugation the solutions are incubated and slowly agitated for 5 days at room temperature. Sample treatment as described here minimizes loss due to the formation of large insoluble aggregates during sedimentation experiments.

4.2.2. Analytical ultracentrifugation

Sedimentation velocity experiments are performed in a Beckman Optima XL-A using the four-hole AN-60 Ti rotor. The 300- to 400- μl sample is filled into standard double sector aluminum center pieces using both sectors as sample sectors. Radial scans are taken in intensity mode at a resolution of 0.002 cm. All samples are measured at 493 nm to observe end-labeled Oregon Green and to avoid background absorbance from the aggregation inhibitor, which absorbs strongly in the ultraviolet region. The partial specific volume of the A β (1–42) peptide ($\bar{v} = 0.7377\text{cm}^3/\text{g}$) is calculated on the basis of its amino acid content by a routine implemented in Ultra-Scan. Experimental intensity data are time-invariant noise corrected using the 2DSA analysis. The van Holde–Weischet analysis is used to initialize the s -value range in the 2DSA from 1 to 150 S. The frictional ratio range is initialized between 1 and 10. 2DSA analyses are performed with 24 grid movements with a 10-point resolution in both dimensions, resulting in a final s -value resolution of 0.625 S and 0.042 f/f_0 units. The 2DSA results are used to initialize the GA analysis, and parsimoniously regularized GA distributions are used to initialize the GA Monte Carlo analysis.

4.2.3. Transmission electron microscopy (TEM)

Transmission electron microscopy experiments were performed by W. Meyer-Zaika in the inorganic chemistry department of the University Duisburg-Essen, Germany, with a Phillips CM 200 FEG instrument. After absorbing to the holey carbon film-coated copper grids (Plano, Wetzlar, Germany), the samples are stained negatively with a 2% (w/v) ammonium molybdate solution.

4.2.4. Experimental results

Based on the loading concentration measured before the experiments, and the first scan's plateau concentration, we concluded that approximately 10% of the dye-labeled A β (1–42) peptide was lost during the acceleration phase. This suggests that about 90% end-labeled A β (1–42) remained soluble and that our sedimentation velocity analysis will provide a representative picture for most, but not all, of the sample. The A β (1–42) peptide incubated at the

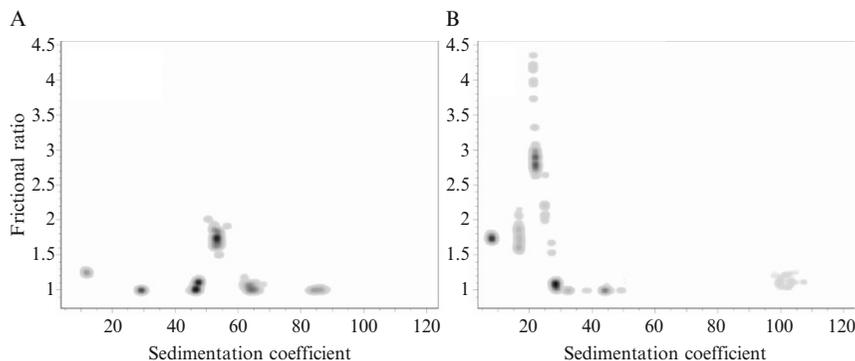


Figure 4.10 GA–Monte Carlo analysis of 21 μM $\text{A}\beta$ (1–42)/ $\text{A}\beta$ (1–42)–OG in 10 mM sodium phosphate buffer, pH 7.4, 4% DMSO. The sample was incubated slowly and agitated for 5 days at room temperature prior to centrifugation. (A) Without inhibitor. (B) 200 μM inhibitor added. A noticeable decrease in the concentration of larger, globular structures is observed when an inhibitor is added, favoring smaller species with larger frictional coefficients.

described conditions reproducibly showed sedimentation boundaries, which indicated at least three different classes of molecular species with the majority sedimenting with an s value around 50 S, less than 20% sedimented with s values below 30 S, and a minor amount of material did not sediment and simply contributed to the baseline at the applied forces and is presumably monomer. A rotor speed of 20 krpm seemed to be optimal for obtaining a sedimentation signal for most species present in the mixture. The 2DSA resulted in an s -value range between 1 and 150 S with an RMSD of 0.003 and a total concentration of 0.15 OD for the control experiment, and 0.16 OD for the experiment where inhibitor was added. Parsimonious regularization with the GA analysis reduced the total number of detected species by a factor of 5. The GA results were used to initialize a GA/Monte Carlo analysis, and the GA/Monte Carlo results are shown in Fig. 4.10A for $\text{A}\beta$ (1–42) without inhibitor (control) and in Fig. 4.10B for $\text{A}\beta$ (1–42) with inhibitor. In the control sample all species appear to be mostly globular, consistent with the globular and slightly nonglobular structures observed in TEM images shown in Figs. 4.11 and 4.12. This sample also contained a species sedimenting with 53.0 S (+1.47 S/–1.3 S) and a frictional ratio of 1.76 (+0.15/–0.18), representing about 35% of the material. This species appears to be more elongated, consistent with some shorter fibrils seen in Fig. 4.12, similar to fibrils reported in the literature (Antzutkin *et al.*, 2002). When simulated as a long rod, this species is consistent with an axial ratio of 16.2, a diameter of 7.4 nm, a length of 120 nm, and a molecular mass of about 4.4×10^6 Da. Networks of large 10-nm amyloid fibrils also seen in Figs. 4.11 and 4.12 and, to a much lesser

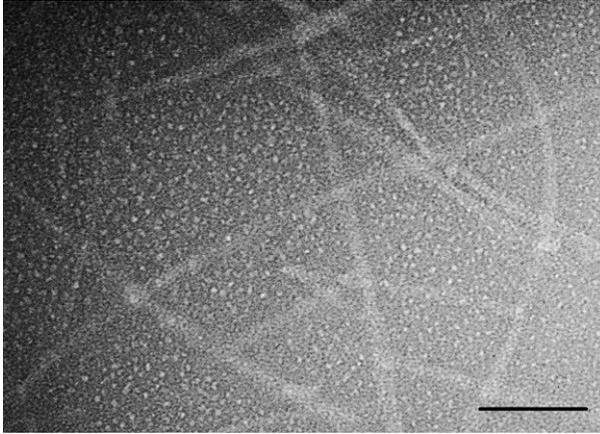


Figure 4.11 Electron micrograph of $A\beta$ (1–42) fibrils seen together with small globular oligomers after 5 days of incubation agitated slowly at room temperature. Bar: 50 nm.

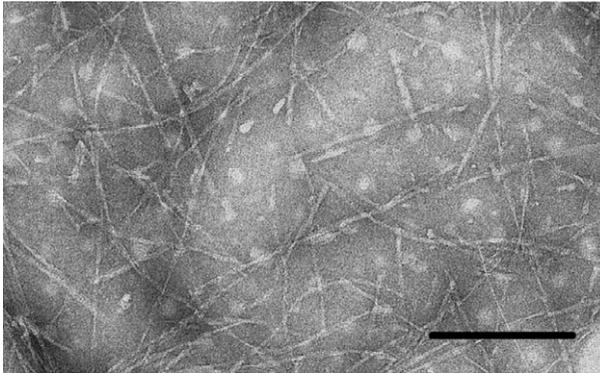


Figure 4.12 Electron micrograph of $A\beta$ (1–42) fibrils seen together with larger globular and elongated oligomers after 5 days of incubation agitated slowly at room temperature. Bar: 200 nm.

degree, also in [Fig. 4.13](#) are most likely insoluble and part of the fraction lost during the acceleration phase. The sedimentation experiment of the sample containing the inhibitor showed a significant decrease of the large globular structures, which is consistent with the absence of such particles in the corresponding TEM image shown in [Fig. 4.13](#). The inhibitor-containing sample also displayed a much reduced presence of large fibril networks, suggesting that the inhibitor may successfully degrade both globular aggregates and also fibril networks. Most notably, the GA Monte Carlo analysis suggested that the majority of the material sedimented much slower when inhibitor was added to the sample, resulting in a decrease of the

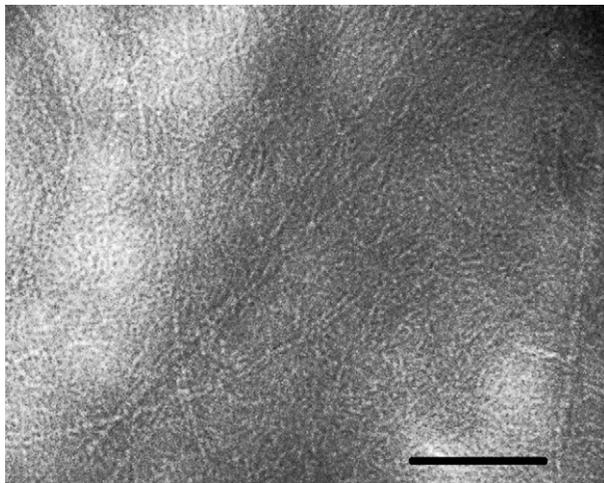


Figure 4.13 Electron micrograph of 21 μM $\text{A}\beta$ (1–42)/ $\text{A}\beta$ (1–42)-OG mixed with 200 μM inhibitor after 5 days of incubation agitated slowly at room temperature. Bar: 100 nm.

weight-average sedimentation coefficient from 52.0 to 25.5 S and a concomitant increase in the weight average frictional coefficient from 1.29 to 1.93 when compared to the control experiment. The major species (33% sedimenting at 21.8 S) detected in the sample containing the inhibitor displayed a significantly increased frictional ratio with a large standard deviation (2.79, +1.11/−0.63), which is strongly indicative of a large range of fibril structures with varying lengths. Such structures were not measured in the control experiment (Fig. 4.10A). Again, this observation is closely matched by the corresponding TEM experiment, which shows a high concentration of varying length fibrils that are quite short in comparison to the large fibril networks observed in Figs. 4.11 and 4.12. Instead, Fig. 4.13 revealed the presence of high numbers of thin filaments (width below 5 nm). Because of their high abundance and their dimensions close to the resolution limit of TEM it was impossible to give reasonable estimates about their length. Nevertheless, a process leading to thinner filaments, for example, by preventing lateral association, should result in larger frictional ratios and increased axial ratios. For the control experiment, we also modeled the globular solutes seen in Figs. 4.11 and 4.12. Small spherical particles seen in the TEM experiment (Figs. 4.11 and 4.12) suggest diameters ranging from 4 to 20 nm. Such particles are consistent with the globular solutes found in the velocity experiment shown in Fig. 4.10A. A summary of the globular particles is shown in Table 4.2. The literature reports micellar amyloid β (Sabatè and Estelrich, 2005) as well as other globular aggregates (Hepler *et al.*, 2006; Lambert *et al.*, 1998) of the peptide,

Table 4.2 Globular species detected in sedimentation velocity experiment and their predicted diameter^a

<i>s</i> value	f/f_0	Relative concentration (%)	Calculated diameter (nm)	Approximate molecular mass (Da)
11.5 S	1.25	4.20	4.3	2.7×10^5
28.7 S	1.00	7.90	12.0	7.5×10^5
46.4 S	1.00	27.80	15.5	1.5×10^6
65.2 S	1.00	17.90	18.2	2.6×10^6
84.0 S	1.00	7.39	20.7	3.8×10^6

^a Predicted sizes correlate well with the measured particles in TEM images.

which also supports the detection of these particles by sedimentation velocity centrifugation. The very small globular particles visible in Fig. 4.11 have an estimated diameter of 3 to 4 nm.

5. CONCLUSIONS

We have developed new procedures that allow investigators to analyze AUC experiments with unsurpassed precision. We have shown that by applying computationally advanced tools we can improve AUC data analysis by a significant measure and provide information content and reliability of the results that exceed, by far, the information gleaned from traditional methods (Brown and Schuck, 2006; Demeler and Saber, 1998; Schuck, 2000; Stafford, 1992; van Holde and Weischet, 1978). The investigator can now reliably characterize mixtures of solutions that are simultaneously heterogeneous in molecular weight and also in shape. In addition, our methods make it possible to measure the statistical relevance of the results. We have developed a grid-based implementation of all necessary tools and have built the necessary environment to conveniently use these tools through user-friendly Web interfaces and to distribute compute jobs to remote supercomputers using the Globus grid technology (Foster, 2005). Example 1 demonstrated the ability of our method to extract very detailed molecular weight and shape information from sedimentation velocity experiments containing stochastic noise equivalent to that observed in a Beckman Optima XL-A by accurately resolving up to five different species ranging between 25 and 400 kDa and covering frictional ratios ranging between 1.2 and 2.0. Because these results are obtained when optimal conditions are present, we would like to stress the importance of high-quality data. We note that these results can only be obtained when data are free of

systematic error sources, which need to be carefully controlled or eliminated. While our software allows for the correction of time- and radially invariant noise contributions, other systematic error sources may still be present that could significantly impact the accuracy of the results obtained. Potential distortions of data could result from concentration-dependent nonideality terms that are not considered in the model, from nonlinearities in the optics, refractive artifacts, including Wiener skewing, temperature, wavelength, and rotor speed variations, as well as buffer gradients that may be changing density and viscosity as a function of radius, and incorrect speed selections for the execution of the experiment. Other factors affecting accuracy can be controlled by the experimentalist using special care during the design of the experiment to assure that optimal conditions exist during the experiment, such as (1) cleaning the lamp before the experiment to assure maximum intensity; (2) using only carefully aligned cells (we use the Spin CAT cell alignment tool, available from Spin Analytical, P.O. Box 865, Durham, NH 03824-0865); (3) avoiding the use of worn out or deformed cell housings or scratched/damaged centerpieces and windows; (4) making sure that the optical system is aligned properly; (5) measuring at 230 nm, where light intensity is maximal, if buffer conditions permit the use of low UV wavelengths; and (6) making sure that optical density levels do not exceed the linearity of the optics, which is a function of light intensity that varies with wavelength of the light used. We also recommend using intensity measurements instead of absorbance measurements, which avoids the subtraction of two stochastic noise vectors, and hence avoids an increase of stochastic noise by a factor of $\sqrt{2}$; we also recommend avoiding buffer components that either contribute to the background absorbance or change the absorbance or refractive index during a run (certain reductants are prone to exhibit this problem when they change oxidation state midrun). Results obtained from the methods described here, when applied to the experimental amyloid aggregation system, provided new information on several levels and the presented AUC analysis methods proved to be a valuable tool in characterizing the aggregation process of amyloidogenic proteins. Particularly with respect to the evaluation of aggregation modulating compounds the method will be of great importance. Due to the improved data evaluation method, differently shaped and sized particles could be detected in one experiment, which could be validated by electron microscopic images performed with aliquots of the identical samples. A frictional ratio of 2.8 associated with the dominant species with 21.7 S in samples where the A β (1–42) peptide was incubated for 5 days together with the inhibitor suggested an increased axial ratio of the underlying particles. This corresponds well to the TEM experiments of the same sample, which show a large number of thin filaments, which are not present in controls of A β (1–42) or in the ligand alone. Also, based on the sedimentation velocity results we could determine the relative amount of globular aggregates present in the

inhibitor-free control sample and estimate the size and molecular weight of these species. Such species were also detected in the TEM images. Since electron microscope images show only a minute section of a sample and presume that all species absorb equally well to the grid, analytical ultracentrifugation and our analysis methods can nicely complement TEM and add important information to the study of amyloids by more comprehensively representing the soluble portion of the sample and by describing the relative ratio in the mass and shape of the soluble particles. We suggest that further velocity experiments performed at lower speeds could potentially derive additional information about the presence of fibril networks that appeared to have sedimented during acceleration and were therefore not detected in our experiments and plan to investigate this further.

ACKNOWLEDGMENTS

We thank Jeremy Mann and the UTHSCSA Bioinformatics Core Facility and the Texas Advanced Computing Center (UT Austin) for computational support. We gratefully acknowledge support from the National Science Foundation through Teragrid Grant TGMCB060019T, as well as the National Institutes of Health through Grant RR022200 (both to BD). L.N.S. was supported by a grant of the Volkswagenstiftung (I/82 649) to Dieter Willbold, Head of the Institute for Physical Biology of the Heinrich-Heine-University.

REFERENCES

- Antzutkin, O. N., Leapman, R. D., Balbach, J. J., and Tycko, R. (2002). Supramolecular structural constraints on Alzheimer's β -amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry* **41**, 15436–15450.
- Barghorn, S., Nimmrich, V., Striebinger, A., Krantz, C., Keller, P., Janson, B., Bahr, M., Schmidt, M., Bitner, R. S., Harlan, J., Barlow, E., Ebert, U., *et al.* (2005). Globular amyloid β -peptide oligomer: A homogenous and stable neuropathological protein in Alzheimer's disease. *J. Neurochem.* **95**, 834–847.
- Brookes, E., Boppana, R. V., and Demeler, B. (2006). Computing large sparse multivariate optimization problems with an application in biophysics. Supercomputing '06. *ACM* 0-7695-2700-0/06.
- Brookes, E., and Demeler, B. (2006). Genetic algorithm optimization for obtaining accurate molecular weight distributions from sedimentation velocity experiments. "Analytical Ultracentrifugation VIII" (C. Wandrey and H. Cölfen, eds.). Springer. *Progr. Colloid Polym. Sci.* **131**, 78–82.
- Brookes, E., and Demeler, B. (2007). Parsimonious regularization using genetic algorithms applied to the analysis of analytical ultracentrifugation experiments. *ACM GECCO Proc.* 978-1-59593-697-4/07/0007.
- Brown, P. H., and Schuck, P. (2006). Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys. J.* **90**, 4651–4661.
- Cao, W., and Demeler, B. (2005). Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation. *Biophys. J.* **89**, 1589–1602.

- Cao, W., and Demeler, B. (2008). Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution for multi-component reacting systems. *Biophys. J.* **5**, 54–65.
- Demeler, B. (2005). UltraScan: A comprehensive data analysis software package for analytical ultracentrifugation experiments. In “Modern Analytical Ultracentrifugation: Techniques and Methods” (D. J. Scott, S. E. Harding, and A. J. Rowe, eds.), pp. 210–229. Royal Society of Chemistry, United Kingdom.
- Demeler, B. (2008). UltraScan: A comprehensive data analysis software for analytical ultracentrifugation experiments. <http://www.ultrascan.uthscsa.edu>.
- Demeler, B., and Brookes, E. (2008). Monte Carlo analysis of sedimentation experiments. *Colloid Polym. Sci.* **286**, 129–137.
- Demeler, B., and Saber, H. (1998). Determination of molecular parameters by fitting sedimentation data to finite element solutions of the Lamm equation. *Biophys. J.* **74**, 444–454.
- Demeler, B., and van Holde, K. E. (2004). Sedimentation velocity analysis of highly heterogeneous systems. *Anal. Biochem.* **335**, 279–288.
- Fezoui, Y., Hartley, D. M., Harper, J. D., Khurana, R., Walsh, D. M., Condrón, M. M., Selkoe, D. J., Lansbury, P. T. Jr., Fink, A. L., and Teplow, D. B. (2000). An improved method of preparing the amyloid beta-protein for fibrillogenesis and neurotoxicity experiments. *Amyloid* **7**, 166–178.
- Foster, I. (2005). Globus Toolkit Version 4: Software for Service-Oriented Systems. IFIP International Conference on Network and Parallel Computing, Springer-Verlag. LNCS **3779**, 2–13.
- Hardy, J. A., and Higgins, G. A. (1992). Alzheimer’s disease: The amyloid cascade hypothesis. *Science* **256**, 184–185.
- Heppler, R. W., Grimm, K. M., Nahas, D. D., Breese, R., Dodson, E. C., Acton, P., Keller, P. M., Yeager, M., Wang, H., Shughrue, P., Kinney, G., and Joyce, J. G. (2006). Solution state characterization of amyloid β -derived diffusible ligands. *Biochemistry* **45**, 15157–15167.
- Holm Nielsen, E., Nybo, M., and Svehag, S. E. (1999). Electron microscopy of prefibrillar structures and amyloid fibrils. *Methods Enzymol.* **309**, 491–496.
- Iqbal, Z., Sisosdia, S. S., and Winblad, B. (2001). In “Alzheimer’s Disease and Related Disorders: Advances in Etiology, Pathogenesis and Therapeutics.” Wiley, London.
- Kirsten, C., and Schrader, T. (1997). Intermolecular β -sheet stabilization with aminopyrazoles. *J. Am. Chem. Soc.* **119**, 12061–12068.
- Klein, W. L., Stine, W. B. Jr., and Teplow, D. B. (2004). Small assemblies of unmodified amyloid β -protein are the proximate neurotoxin in Alzheimer’s disease. *Neurobiol. Aging* **25**, 569–580.
- Lamm, O. (1929). Die Differentialgleichung der Ultrazentrifugierung. *Ark. Mat. Astron. Fys.* **21B**, 1–4.
- Lawson, C. L., and Hanson, R. J. (1974). “Solving Least Squares Problems.” Prentice-Hall, Englewood Cliffs, NJ.
- Rochet, J. C., and Lansbury, P. T. (2000). Amyloid fibrillogenesis: Themes and variations. *Curr. Opin. Struct. Biol.* **10**, 60–68.
- Rzepecki, P., Nagel-Steger, L., Feuerstein, S., Linne, U., Molt, O., Zadnard, R., Aschermann, K., Wehner, M., Schrader, T., and Riesner, D. (2004). Prevention of Alzheimer’s associated A β aggregation by rationally designed nonpeptidic β -sheet ligands. *J. Biol. Chem.* **279**, 47497–47505.
- Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619.
- Schuck, P., and Demeler, B. (1999). Direct sedimentation boundary analysis of interference optical data in analytical ultracentrifugation. *Biophys. J.* **76**, 2288–2296.

- Selkoe, D. J. (2003). Folding proteins in fatal ways. *Nature* **426**, 900–904.
- Spencer, B., Rockenstein, E., Crews, L., Marr, R., and Masliah, E. (2007). Novel strategies for Alzheimer's disease treatment. *Expert Opin. Biol. Ther.* **7**, 1853–1867.
- Stafford, W. (1992). Boundary analysis in sedimentation transport experiments: A procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Anal. Biochem.* **203**, 295–301.
- Vadapalli, R. K., Sill, A., Dooley, R., Murray, M., Luo, P., Kim, T., Huang, M., Thyagaraja, K., and Chaffin, D. (2007). Demonstration of TIGRE environment for Grid enabled/suitable applications. Grid2007, Austin, TX. <http://www.grid2007.org/demo-vadapalli.pdf>.
- van Holde, K. E., and Weischet, W. O. (1978). Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers* **17**, 1387–1403.
- Wall, L., Christiansen, T., and Orwant, J. (2000). In "Programming PERL," 3rd Ed.