Chapter 8
Analytical Ultracentrifugation Data Analysis with UltraScan-III

Borries Demeler and Gary E. Gorbet

Abstract The current status of the UltraScan-III (US3) data analysis software suite is described. An overview of the US3 concepts, software layout, the data workflows and US3 components is presented, followed by a discussion of the analysis methods and their applications. Also described are visualization modules for analysis results, US3's utilities and simulation tools, as well as the collaboration environments for online data and result exchange.

Keywords Analytical ultracentrifugation • UltraScan • Lamm equation modeling • Two-dimensional spectrum analysis • Genetic algorithms • Multi-wavelength • AUC • Supercomputing • Simulations • Sedimentation velocity experiments

8.1 Introduction

UltraScan-III (US3, http://www.ultrascan.uthscsa.edu) is a free and open source, multi-platform software suite designed to provide high-performance and high-throughput data analysis and modeling of hydrodynamic data. The UltraScan-III project grew out of the requirement to support the needs of the analytical ultracentrifugation core facility at the University of Texas Health Science Center in San Antonio. This facility owns multiple analytical ultracentrifugation (AUC) instruments and has several hundred users, many of whom collaborate on joint projects which necessitates sharing data online. As is discussed below, US3 addresses many challenges posed by this large, multi-user environment through the use of a relational MySQL database with a web-based interface, termed the UltraScan Laboratory Information Management System (USLIMS) (Demeler 2009). Foremost, US3 allows researchers to achieve an unsurpassed level of accuracy and resolution in their data analysis. US3 is designed to provide maximum flexibility in formulating a great variety of custom analysis models and optimization

B. Demeler (✉) • G.E. Gorbet
Department of Biochemistry, The University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA
e-mail: demeler@biochem.uthscsa.edu

S. Uchiyama et al. (eds.), *Analytical Ultracentrifugation*,
DOI 10.1007/978-4-431-55985-6_8

approaches, while at the same time offering an intuitive interface that is easy to learn and use. The software currently supports data from the Beckman XLA/I UV/visible absorbance and intensity detector, the Rayleigh interference detector, the Aviv fluorescence detector, and from two recently developed multi-wavelength detectors (Bhattacharyya et al. 2006; Schilling 2014). Both single- and multi-speed experiments are supported. The software is highly configurable so that repetitive actions can be largely automated to speed up routine analysis and accelerate discovery, without compromising rigor, accuracy, and flexibility in the functionality. In recent years, increasing availability of high performance computing (HPC) and network infrastructure has opened up new avenues for biophysical modeling and analysis of sedimentation data. It is now possible to use computationally demanding, parallelized fitting approaches (Demeler and Brookes 2008; Brookes et al. 2006, 2010a; Brookes and Demeler 2007, 2008) based on whole boundary models using an adaptive space-time finite element solution for the underlying flow equation (Cao and Demeler 2005, 2008). The solution built into US3 is able to simulate self- and hetero-associating reactions, including kinetic rate constants (Demeler et al. 2010), supports solvent compressibility, co-sedimenting solutes and gradient formation, as well as concentration dependency of s and D. The parallel methods programmed into US3 provide significantly higher accuracy and resolution than conventional approaches, which are limited by traditional desktop or laptop computers where high-resolution analysis is impractical and time consuming. US3 also allows the user to process many datasets in parallel, greatly improving throughput and time savings (Memon et al. 2013). This is particularly critical for the new multi-wavelength data format, where datasets for several hundred wavelengths must be evaluated from each channel. Additional performance gains are realized from streamlined and automated workflows available through the networked science gateway and offered by the Extreme Science and Discovery Environment (XSEDE, funded by the National Science Foundation in the USA) and from the Partnership for Advanced Computing in Europe (PRACE). These workflows are accessed through efficient grid middleware implementations that allow investigators to distribute jobs to multiple supercomputer clusters simultaneously (Memon et al. 2014). US3 aims to provide a comprehensive and robust analysis environment for all hydrodynamic analysis. In addition to sedimentation data analysis, UltraScan offers the SOMO Solution Modeler with comprehensive facilities for hydrodynamic modeling (Brookes et al. 2010b, c; 2013), further discussed in Chap. 10. The integration of remote HPC resources in UltraScan and the exchange of research data and analysis results are accomplished through the USLIMS, and the Apache Airavata grid middleware that manages the communication with the HPC clusters (Marru et al. 2011). US3 adheres to the OpenAUC data standard (Cölfen et al. 2010), which provides significantly higher storage and I/O efficiency than traditional Beckman formatted ASCII files, increasing data loading and network transfer speed. The OpenAUC standard offers database associations between related data elements which improves accuracy and automation. Below, the most important features of US3 are discussed. US3 is an ongoing project with many collaborators and

contributors. A wiki is available on the UltraScan website that provides many resources for users and developers (UltraScan III).

8.2 UltraScan-III Components

The US3 software consists of several local, online, and remote components. A multi-platform desktop binary (Linux/X11, Windows XP/7/8, Macintosh OS-X) is used to import and edit experimental data, visualize results, create analysis reports, and to provide access to many utilities and simulation programs. The user has the option to use US3 for the analysis of data stored locally, or stored in the online database. The latter is required if data are submitted to a remote supercomputer for analysis, the former is intended for situations when Internet is temporarily unavailable, or the user chooses to work without a database back-end. Not all analysis methods are suitable for local analysis, some require supercomputer capability. Computationally demanding routines are all multi-threaded to take advantage of modern multi-core architectures. When data are stored in the database, they are accessible to any authorized user from any Internet location. In addition to functioning as the preferred data storage for the desktop component, this database is also at the core of the online UltraScan Laboratory Information Management System (USLIMS, http://www.uslims3.uthscsa.edu). The user interacts with the USLIMS through a web browser. The USLIMS offers the user remote access to analysis reports and metadata, an administrative interface, and the online submission system for remote supercomputer analysis. The remote analysis is performed by a parallelized MPI routine running the ASTFEM codes (Cao and Demeler 2005, 2008), which is installed on multiple XSEDE resources in the USA, and on Juropa at the Forschungszentrum Jülich, available to European users. Compute cycles in the USA are offered for free to all users through a community account, which is supported through an NSF/XSEDE allocation grant to one of the authors (BD). The final component constitutes the Airavata Science Gateway infrastructure developed at Indiana University. It is responsible for managing all analysis requests from the USLIMS and for sending them to the selected remote supercomputer, and for moving input data and results between supercomputers and the database. All data transfers and communications between database, supercomputer, and Airavata are ssl-encrypted to protect the data from unauthorized access. To facilitate collaboration among users, each desktop installation can be configured to access multiple databases. Database usernames and passwords are stored encrypted on disk and are decrypted with a single master password, specific for each user's installation. The master password is carried in memory as long as the US3 application is open. For the duration of the session, each transaction with a remote database is authenticated by MySQL stored procedures, requiring a one-time sign-on with the master password. To assure data ownership integrity, the US3 database maintains a hierarchy of user levels, depending on the user's role in the database, which in order of decreasing permission level can be *superuser*, *administrator*, *analyst*,

investigator, or *unprivileged*. The desktop version of US3 honors these levels and prevents unauthorized access to data from other users of the database. To facilitate collaboration and exchange of result data through the USLIMS website, the user can choose to share their data with other users registered on the same USLIMS instance, regardless of their user level. In order to access experimental data other than results, the user must have *analyst* or higher permission set. Below this level, data stored in the database cannot be processed with US3, and they can only be viewed in the USLIMS instance.

8.3 UltraScan-III Concepts

An important change in US3 compared to earlier versions is the emphasis on sedimentation velocity (SV) data and the preference of intensity over absorbance data. The main reason sedimentation equilibrium (SE) experiments historically were of interest is the relative simplicity of their analysis. SE models are based on analytical exponential functions instead of finite element solutions of differential equations, and only a single scan at the equilibrium stage of the experiment is needed for each speed or concentration. In addition, equilibrium columns are generally less than 3 mm high, so very little data needs to be modeled, greatly reducing memory requirements. This also means that, compared to SV experiments, only limited information content is available, significantly decreasing the confidence an investigator can have in the results (Demeler et al. 2010). Furthermore, systematic time invariant noise (TI) subtraction is impossible for equilibrium data, since scans are by definition time-invariant. Due to the availability of high-performance computing and high-resolution SV analysis methods in US3, the perceived advantage of simpler models for SE experiments is no longer relevant. As a consequence, the US3 user is encouraged to measure SV experiments instead. Even if an equilibrium experiment is planned, it can still be treated as a velocity experiment by collecting also the data during the approach-to-equilibrium period of the experiment, taking advantage of additional information in the data and using SV analysis methods to interpret the data. Users wishing to analyze legacy equilibrium data are able to do this with the previous version (UltraScan-II), which includes an extensive analysis suite for SE experimental data, and is still available for download from our website. Another important emphasis in US3 is on the replacement of absorbance data (ABS) by intensity data (INT). In ABS, a reference scan is subtracted from the sample scan, thereby convoluting stochastic noise from the reference scan with the stochastic noise from the sample scan. This leads to a $\sqrt{2}$ increase in the stochastic noise signal. Historically, this degradation of experimental data was tolerated because this subtraction also eliminated the majority of TI noise contributions that are present in intensity data, although not completely, since cell windows may have different TI contributions for each channel, and they are not eliminated by reference subtraction, but instead compounded. In US3, efficient algorithms exist to remove both TI and radially invariant (RI) noise contributions from velocity

data (Schuck and Demeler 1999), hence providing a superior dataset and higher confidence in the analysis. Additional advantages of intensity measurements include the fact that the reference sector can be used for a low-concentration sample (<0.5 OD in the Beckman XL-A). Importantly, the design of the experiment is critical for its success. This does not only relate to the optical quality of the data acquired (which is entirely dependent on the maintenance status of the machine), but extends also to the speed of the experiment, which affects the amount of data available for analysis, and the relative sedimentation and diffusion signal contained in the data. Important factors also include sample preparation, buffer selection, and sample concentration. A failure to optimize these parameters is never remedied by applying a sophisticated analysis available in UltraScan. These considerations are further discussed in reference (Demeler 2010). All basic models and optimization algorithms available in US3 provide a general description of SV experiments suitable for most experimental conditions and will always converge to the global minimum. A requirement is that the underlying data behave ideally and are not impacted by systematic instrument errors, and do not exhibit concentration dependent non-ideality or change composition mid-run due to a chemical instability, such as pressure dependence, degradation or time-dependent aggregation processes, or gradient formation. Advanced models are available to handle those special cases but they require additional user input.

8.4 The UltraScan Analysis Workflow

8.4.1 Overview

An overview of the general workflow in US3 is illustrated in Fig. 8.1. In step 1, the user enters a project request together with related information such as the solution details, images for absorbance scans, gel pictures, and experimental designs. The solution details describe the analytes, their partial specific volumes and extinction coefficients, and buffer components, which US3 uses to predict the viscosity and density of the solution. Buffers also specify the pH and compressibility of the solution. Next, experimental data are acquired and imported into the database and associated with the solution information and other ancillary metadata (step 2). In step 3, these data are retrieved to a PC where the data are edited. The edit profiles are stored in the database (step 4) and analyzed locally or on a supercomputer to obtain Lamm equation models (step 5). Next, additional data analysis can be performed locally, and all models and results are visualized. All analysis results and visualizations are deposited in the database (step 6) where a report can later be retrieved from the USLIMS website by the user (step 7). A detailed and updated flowchart for the analysis of SV experiments is available on our website: http://www.ultrascan3.uthscsa.edu/sed-veloc-flowchart.php

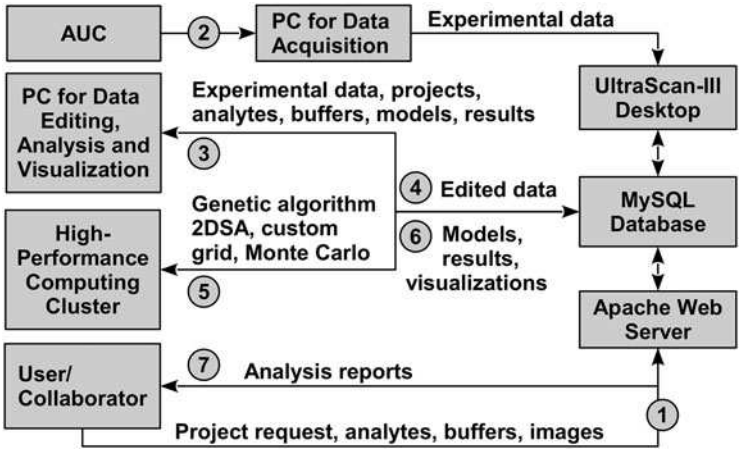

Fig. 8.1 UltraScan-III data flow. The order of steps performed is indicated by numbers (see text)

8.4.2 Importing Experimental Data

Before data can be analyzed with US3 modules, the data must be converted into the binary OpenAUC standard data format (Cölfen et al. 2010). Experimental data from all supported detectors and instruments can be converted by US3. The OpenAUC format is very efficient, scaling the precision of the data type to the accuracy of the detector that was used for data acquisition. In this step, data belonging to an experiment are separated into a unique cell, channel, and wavelength combination. Each combination is referred to as a *triple.* For example, triple 3/B/280 is a dataset containing all scans from the sample channel of cell 3, acquired at 280 nm. Channels are assigned letters A-H, supporting up to eight channels per cell, with "A" referring to the innermost reference channel, and "H" the outermost sample channel. For a multi-wavelength experiment, the number of triples for each channel equals the number of wavelengths acquired (see Fig. 8.2). Additional data relationships are then established: An experiment is first associated with an investigator (the data owner), an instrument operator, the laboratory, instrument and optical system used for acquisition, an experimental project description, and a rotor and rotor calibration (discussed below), the date of the experiment, average temperature, comments, and a run protocol. For each triple, a centerpiece also needs to be selected. The centerpiece geometry, together with the rotor stretching factor (provided from a stored rotor calibration), is later used to calculate a precise position for the bottom of any cell, which is needed for one of the boundary conditions in the solution of the Lamm equation. Geometries of all common centerpieces and previously measured stretching factors for rotors available commercially have been calibrated and entered into US3 for reference, but users can upload their own calibrations. For each triple, a solution must also be defined. A solution is composed of one or more analytes and a buffer, and both are entered by the user. Where possible, partial specific

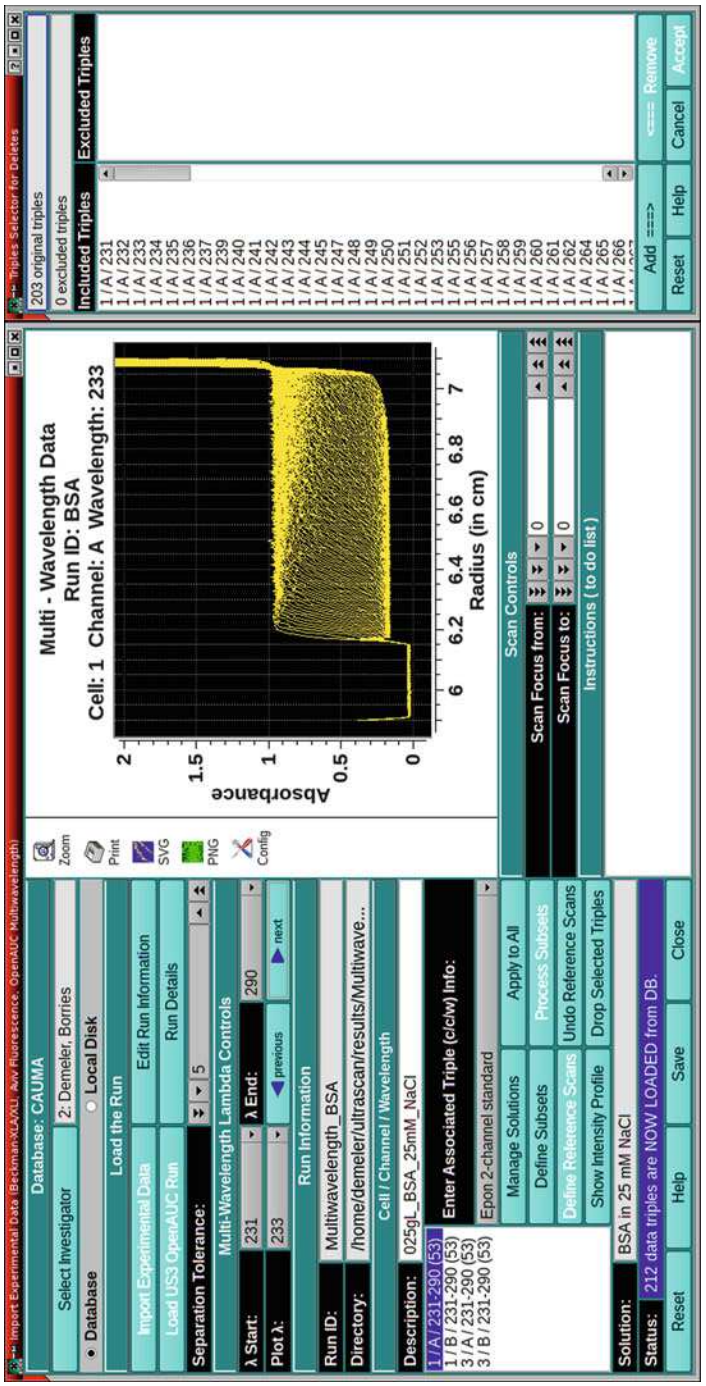

Fig. 8.2 *Left*: Import data dialog with a loaded multi-wavelength experiment for BSA with four channels and 53 wavelengths each. *Right*: The user can exclude selected triples from the data

volumes, molecular weight, and extinction coefficients are estimated automatically by US3 from sequence, or they can be specified by the user. Buffers are composed of buffer components whose density and viscosity increments are directly provided by US3, analogous to the Sednterp software (Laue et al. 1992). Once defined, a solution provides density, viscosity, absorption spectra, extinction coefficient(s), and estimates for the partial specific volumes of the analytes. This information is used in all analyses implemented in US3 to correct all results automatically to standard conditions (water at 20 °C). For temperature correction, US3 assumes an aqueous solution unless a manual correction is specified. This way, experiments performed under different solution conditions can be compared directly, and global fitting of sedimentation data is greatly facilitated (e.g., for organic solvents). Newer multi-wavelength analytical ultracentrifugation (AUC) instruments can produce a time state file, which records the temperature, rotor speed, vacuum, and other system diagnostics throughout the run at short time increments (∼1 s). This information is also stored in binary format and uploaded to the database and provides details about the speed profile, which can be used to more accurately simulate multiple speed steps and acceleration during the run, as well as additional run diagnostics. Finally, an XML file is produced which stores all metadata for the experiment. For multi-speed experiments, special care needs to be taken to correct for the shift in meniscus and bottom of cell position due to different rotor stretching. This is handled automatically by the rotor calibration routine discussed below. In addition, due to the radial shift, TI noise must be determined for each speed separately. Once uploaded, any subsequent analysis of these data is now inextricably linked to the metadata associated with the experiment. While changes to the primary data are possible, such changes would invalidate any derived results and violate their integrity. Therefore, US3 will enforce a deletion of all analysis results if any changes are made to the primary data. Importantly, by storing all associated run details with the experimental data in a relational database at the time of data acquisition, important details about the experiment are associated with the experiment and the user can easily retrieve them later during manuscript preparation. Especially for multi-user facilities, this practice greatly reduces errors and automates the analysis process.

8.4.3 Editing Experimental Data

The next step requires data to be edited. During this process, experimental data are prepared for analysis. An edit profile (stored in XML) is generated, which identifies the meniscus position, the radial data range, any excluded scans, an estimated plateau position needed for analysis methods such as time derivative and second moment, and allows automatic removal of spikes in the data resulting from failed lamp flashes. Each triple can be associated with one or more edit profiles. Multiple profiles are possible to allow investigators to easily evaluate the effect of different editing strategies (for example, exclusion of different scans, selecting

different radial ranges), which can be of interest to diagnose whether samples show time dependent changes during the run, are sensitive to pressure effects, or contain aggregates visible only in early scans of the experiment. After defining an edit profile, a dataset can be analyzed with any analysis available in US3, including remote supercomputer analysis.

8.4.4 Data Refinement

After editing, the user is tasked with the removal of TI and RI noise that may contribute to the experimental data and to find an optimal meniscus position. Intensity and interference data generally contain noticeable TI and RI noise components, while fluorescence and absorbance data contain less. To remove TI and RI noise, data need to be simultaneously modeled with the intrinsic sedimentation and diffusion transport. By using a degenerate, high-resolution two-dimensional model over all possible s and D coefficients present in the experimental data, the intrinsic sedimentation and diffusion transport will be optimally represented, resulting in an uncorrelated noise determination which can then be subtracted from the experimental data (Schuck and Demeler 1999). This process is accomplished in three steps: In the first step, the sedimentation coefficient range is estimated either from an enhanced van Holde–Weischet analysis (Demeler and van Holde 2004), or, for cases where the data contain a significant amount of time-invariant noise, with the time derivative method (Stafford 1992). A single two-dimensional spectrum analysis (2DSA) (Brookes et al. 2010a) with TI noise removal is then fitted over the determined range. The initial fit not only removes a first-order estimation of the TI noise but also baseline offsets common for all scans. In the next step, the 2DSA is iterated with typically 10–30 meniscus positions in the vicinity of the graphically determined meniscus position during editing. At the same time, RI and TI noise contributions are re-fitted to obtain less correlated noise components. This results in an optimally noise-corrected fit for each fitted meniscus position. Next, the root mean square deviations (RMSD) from each fit is plotted against the meniscus position and fitted to a second-order polynomial (see Fig. 8.3). The lowest RMSD position is used to update the meniscus position in the associated edit profile. Depending on the resolution of the meniscus fit, this position does not necessarily correspond to a position previously fitted. To obtain an uncorrelated TI, RI noise profile for the new position, a final 2DSA with TI and RI noise analysis is performed. This last analysis is performed with a maximum of ten iterative refinement steps to obtain the most optimal solution. The TI, RI noise vectors obtained in this last fit are then subtracted from the data in any subsequent data analysis, allowing the user to omit further noise analysis. By maintaining the instrument well, and using intensity mode for UV/visible data collection, the remaining stochastic noise should then be minimal and random. While the noise vectors from the last step are to be preferred, all noise vectors from previous steps are stored in the database and can be evaluated and applied instead, if desired.

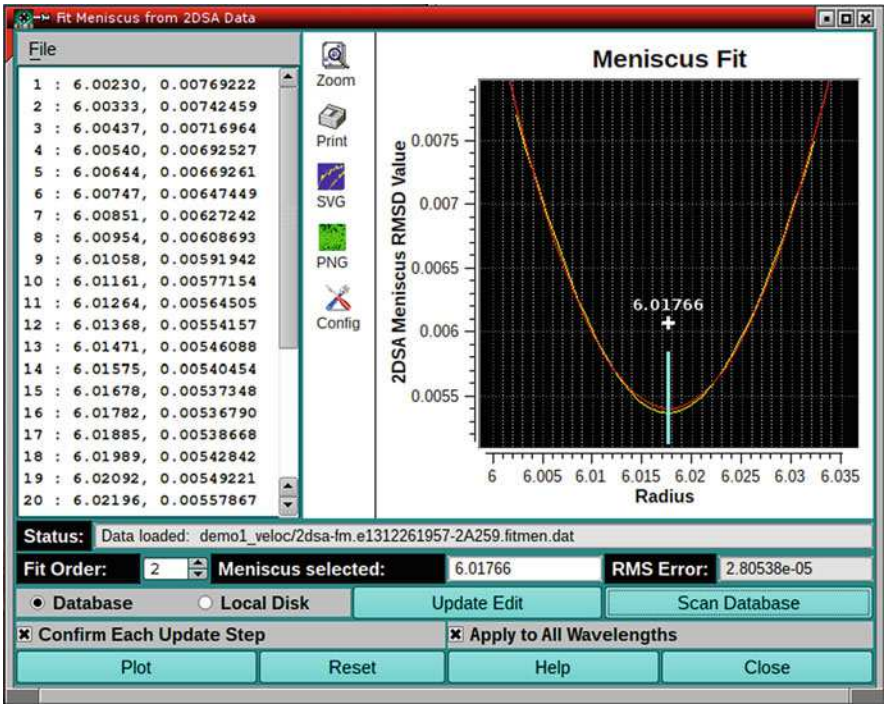

Fig. 8.3 US3 dialog for the meniscus fitter. RMSD values from 2DSA Model files are plotted against meniscus position to identify the optimal meniscus position at the lowest RMSD

8.4.5 Remote Supercomputer Analysis

The steps described in Sect. 8.4.4 can either be performed with the desktop version of the 2DSA analysis using a local computer (Fig. 8.4) or by submission through the USLIMS system to a remote compute cluster. In the latter case, the user will log into their Apache web account on their institutional USLIMS instance and select their edited data from the LIMS3 database. The data is then submitted to a remote cluster for analysis (see Fig. 8.5). This process is handled by the Apache Airavata middleware (Marru et al. 2011), currently operating out of Indiana University. A record for each submission is created in the GFAC database, which allows multiple users to submit multiple jobs synchronously from different instances. The Airavata middleware then stages the job(s) on the requested resource's local queuing system and registers the job as started in the generic factory (GFAC) database. A daemon (gridctrl.php) continually monitors the contents of the GFAC database and updates the USLIMS queue viewer where the user can track progress. Once the job starts, status information is sent via UDP to a second daemon (listen.php), which updates each running job in the queue viewer with status details. Once completed, the resulting models are deposited in the LIMS3 database, and the job is marked as

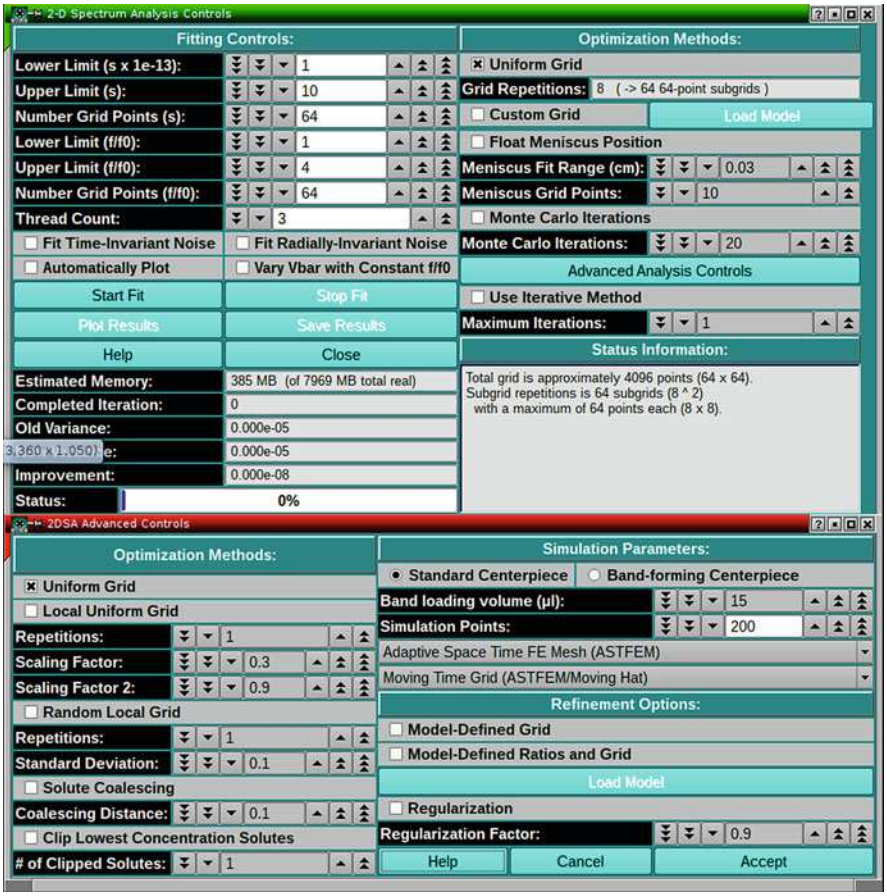

Fig. 8.4 *Top*: 2DSA analysis control window for the UltraScan-III desktop version. *Bottom*: Advanced controls. The desktop version is multi-threaded to support multi-core architectures

completed in the GFAC database and deleted by the grid control daemon. This daemon also notifies the user per e-mail of job completion and updates the queue viewer. Completed jobs are permanently stored in the LIMS database, where the results can be accessed both by the US3 desktop version and through the LIMS system.

8.4.6 Advanced Data Analysis

Once refinement is completed, all subsequent analysis can be performed without additional noise processing or optimization of the boundary conditions. This

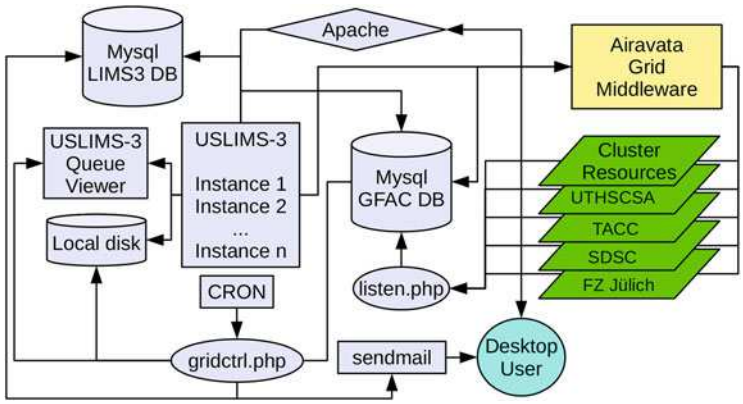

Fig. 8.5 Data flow during supercomputer analysis

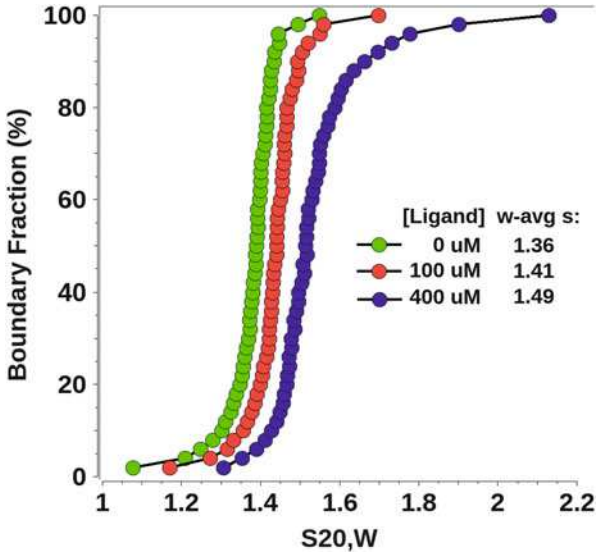

Fig. 8.6 Diffusion-corrected van Holde–Weischet integral s-value distributions provide very sensitive comparisons. In this case, multiple ligand concentrations are measured for a weakly interacting system

reduces calculation time for all subsequent analysis. A range of additional analysis methods are available in US3 which can now be applied. For example, to compare diffusion-corrected sedimentation profiles for multiple samples, an enhanced van Holde–Weischet analysis (Demeler and van Holde 2004) (vHW) can be performed. Combined diffusion-corrected integral G(s) distributions available from the vHW analysis provide a very sensitive way to compare the sedimentation profiles from multiple datasets, even when the changes are very slight (see Fig. 8.6). For

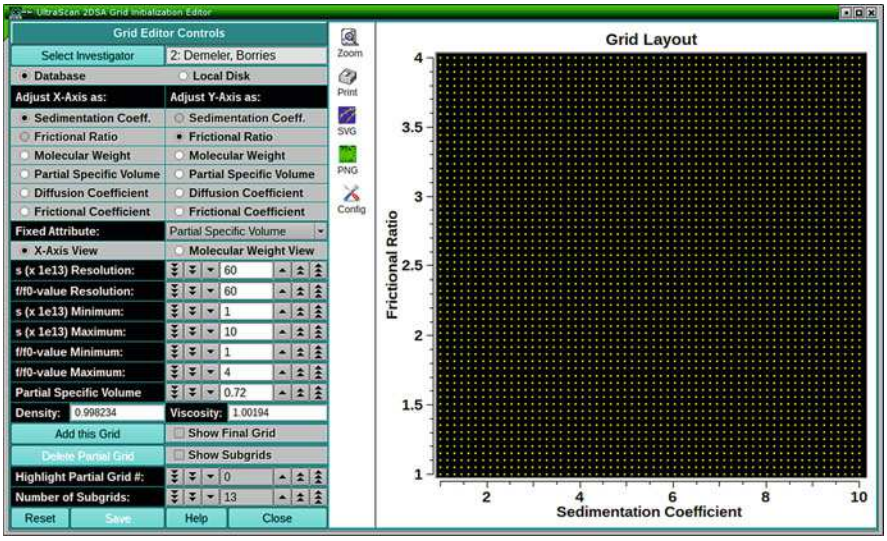

Fig. 8.7 The custom 2DSA grid initialization editor is used to define custom grids that can be fitted with the 2DSA analysis

heterogeneous samples, a 2DSA-Monte Carlo analysis is used to obtain molecular weight and anisotropy distributions, emphasizing intrinsic sedimentation signals while simultaneously attenuating the stochastic noise contributions. For pauci-disperse systems, a parsimonious regularization using the genetic algorithm (GA) optimization method can be used to eliminate non-essential species from the solution without degrading the quality of the fit by applying Occam's razor (Brookes and Demeler 2007). The GA result can be further refined with a Monte Carlo analysis to provide statistical evaluations for all parameters fitted for each identified species and to test the reliability of the fit. A recent addition to US3 is the Custom Grid (CG) method (Fig. 8.7), which allows the user to define the two-dimensional grid analyzed by the 2DSA analysis in terms of any two hydrodynamic parameters that define the sedimentation and diffusion process: s, D, anisotropy, molar mass, partial specific volume, and frictional coefficient when a third parameter is available from an independent measurement. This approach provides great flexibility and allows mixed grids with different parametrizations to be combined. For example, when fitting a DNA–protein associating system, free DNA, free protein, and DNA/protein complex each have a different partial specific volume that can be accommodated by individual custom subgrids to more accurately describe the molar mass distributions present in a mixture. Should molar mass be available from sequence or a mass spectrometry experiment, it can be fixed in the CG analysis and anisotropy and partial specific volume can be fitted (Demeler et al. 2014). Likewise, when anisotropy is available from electron microscopy or crystal structure, it can be held fixed, and heterogeneity in partial specific volume and molar mass can be fitted with the CG analysis. A special case where oligomerization leads to a predictable

anisotropy change can be addressed well by the parametrically constrained spectrum analysis (PCSA (Gorbet et al. 2014)). It allows the user to find the best functional parametrization for the two-dimensional parameter space to constrain the solution to a uni-valued function where only a single frictional ratio matches a single sedimentation coefficient. Arbitrary functional forms can be defined in this method to accommodate any distribution function. GA optimization can be initialized with a manual model, or with the results from any 2DSA, Monte Carlo, CG or PCSA analysis. A second-moment analysis is also available in US3 to provide a diagnostic for samples that are not at chemical equilibrium and change sedimentation behavior throughout the experiment. The second moment analysis reports a weight average sedimentation coefficient for each scan, independent of time.

8.4.7 Global Analysis

US3 offers true global fitting where multiple SV datasets can be fitted simultaneously to a single model with either the 2DSA or GA analysis. Due to the large memory requirements of fitting combined datasets, global fitting is only available on remote supercomputers. In either method, the assumption is made that multiple experiments represent the same sample. The samples can be measured either at the same or different speeds and can be at multiple concentrations. The underlying assumption is that they contain the same set of solutes but not necessarily at the same concentration. By fitting multiple datasets globally, additional signal can be obtained. For example, if a sample is measured at a slow speed, diffusion signal is favored, while a fast rotor speed improves resolution of the sedimentation information. By globally fitting both to a single model, the optimized signal from both experiments for either transport process is combined to provide a more reliable model. The number of models generated from a global fit is $2n + 1$, where n is the number of datasets. The first model represents the best fit global model for all datasets. For each dataset, two more models are generated: The first model contains the same set of solutes found in the global model with identical ratios for each solute maintained from the best-fit global model, but scaled to the total concentration of each dataset. The second model contains the same solutes but re-adjusted in partial concentration to optimally match each dataset. The latter model may have one or more solutes from the global model set to zero concentration. This situation could arise when a reversibly self-associating system was measured at multiple concentrations, and the ratio of the oligomers changes as a function of solute concentration, or aggregates appear in a high concentration sample. The comparison of RMSD for each model therefore serves as a reliable diagnostic for the absence or presence of reversible association. If the composition does not change as a function of concentration, both models will produce similar RMSD values, and mass action is absent. For samples with appreciable mass action occurring, only the model with the adjusted ratios of concentrations will fit well. For samples that

are true replicates of the same sample, at the same concentration, all three types of models will have similar RMSD values.

8.4.8 Models

The core analysis in US3 is based on modeling SV data using Lamm equations solved by the finite element method. One set of optimization methods in US3 attempts to find the best fitting model consisting of linear combinations of Lamm equation solutions representing any solutes present in the experimental data. Each solute has a set of properties, which include partial specific volume, partial concentration, $s_{20,W}$ and $D_{20,W}$ coefficients, concentration dependency factors for s and D, and extinction coefficients. A second set of optimization methods uses GA to fit models containing two or more solutes as well as reaction terms, such as equilibrium constants, rate constants, and stoichiometry (Demeler et al. 2010). A model also contains details about co-sedimenting solutes, such as gradient forming materials which affect flow of any other solutes as a function of time and radius. From the three hydrodynamic parameters stored for each solute, other parameters can be derived, such as molar mass, anisotropy or frictional ratio, and frictional coefficients. This model structure is used universally in US3 to communicate information between all analysis and visualization modules, including: (1) output from any optimization method based on finite element modeling. Such a model will always contain a reference to the method that was used to determine the model, as well as a root mean square deviation (RMSD) of the fit; (2) user-defined models, either for simulation or to fit the model's parameters directly; (3) input models for the refinement or initialization of a subsequent method, and as a custom grid; (4) a model structure can be visualized with several modules in UltraScan; (5) multiple models can be combined to create global models used for global analysis. Thus, a model serves as a well-defined structure for different methods and modules in US3.

8.4.9 Visualization

US3 contains several powerful model visualization capabilities. First, the user can simulate any model using the US3 simulation routine, either by defining a custom instrument setting or by comparison of an experimental data set with a fitted model. In the latter case, all settings from the experiment are inherited, and an overlay is generated. If time- or radially invariant noise is part of the model, it can be subtracted from the dataset and the experimental and simulated data can be shown overlaid, and all hydrodynamic parameters from the model can be plotted in two dimensions (Fig. 8.8, upper left). The distribution of solutes and their partial concentrations can be visualized in three dimensions by using pseudo-three-dimensional plots or true three-dimensional plots as shown in Fig. 8.9. Here, the user has the option

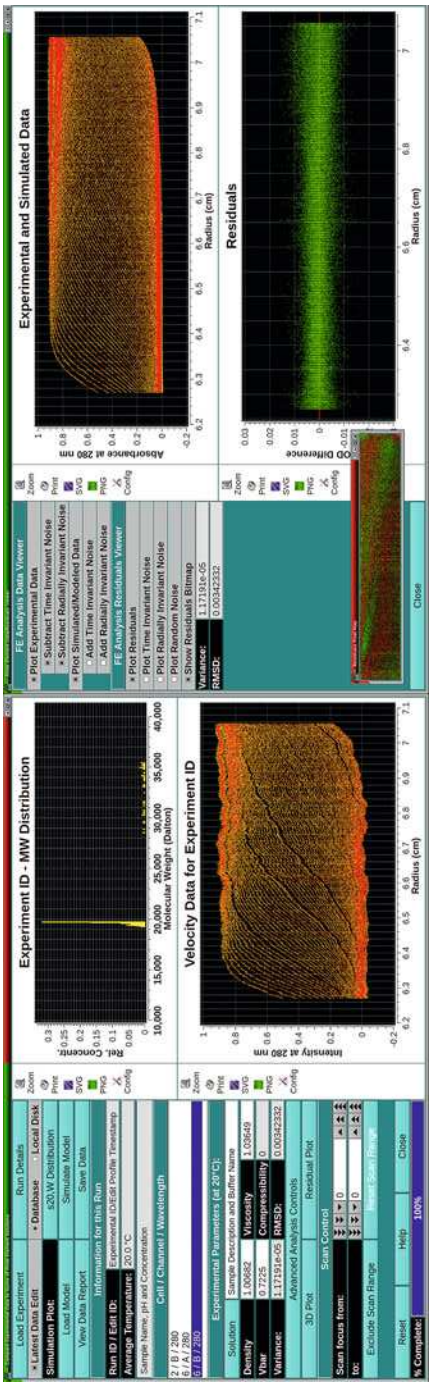

Fig. 8.8 Visualization of experimental intensity SV data and modeled simulations in US3. Experimental data (*yellow*) overlaid by simulations (in *red*) including (*lower left*) and after subtraction of systematic noise components (*upper right*), stochastic noise (lower right), residual bitmap (*low center*) and molar mass distribution information, which can be displayed for all hydrodynamic parameters and transformations (*upper left*). Systematic noise components can be displayed in the lower right panel (*not shown*), and various controls provide flexibility in the display and selection of data

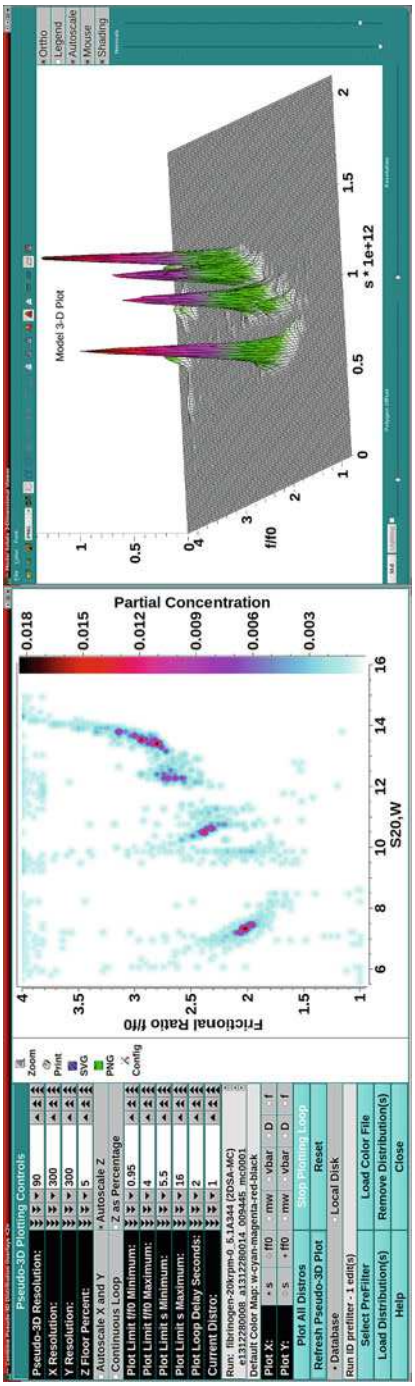

Fig. 8.9 Model visualization options (the same multicomponent 2DSA Monte Carlo model is shown in both viewers). *Left:* The pseudo-three-dimensional viewer. Low concentration or noise points are shown in light cyan. *Right:* A flexible three-dimensional viewer module based on the Qt framework is integrated in US3. Both viewers provide a flexible choice of visualizing all hydrodynamic parameter combinations for x- and y-axes to accommodate the Custom Grid method

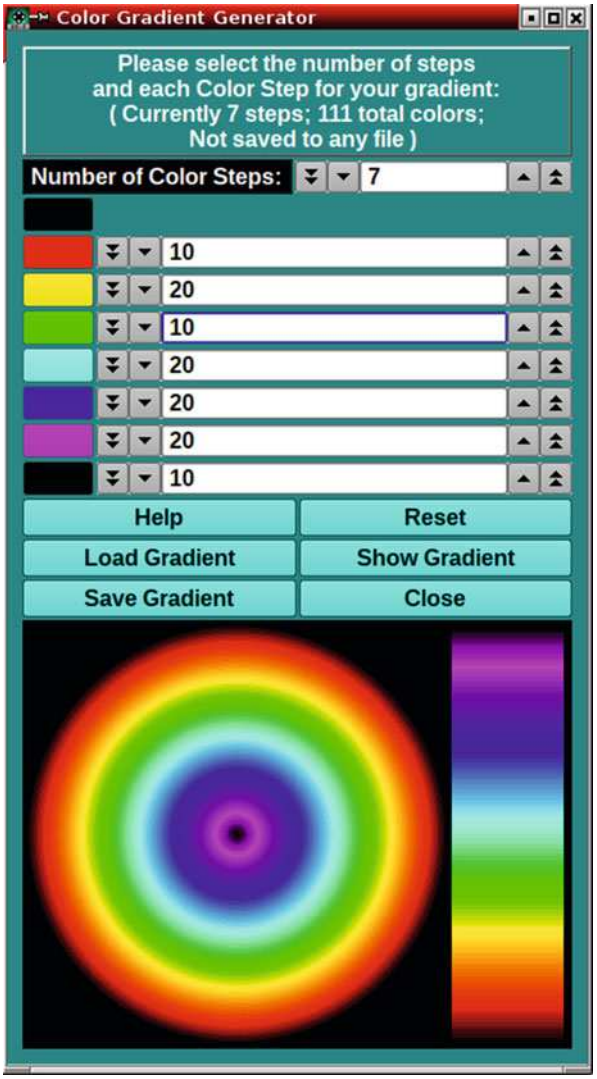

Fig. 8.10 Gradient editor

to select any two of the variable hydrodynamic parameters for the X- and Y-axis. A gradient editor (see Fig. 8.10) provides a convenient interface to define custom color gradients, which can be used to differentially color the three-dimensional plots. In the PCSA analysis, an arbitrary equation is used to specify a constraint within the two-dimensional hydrodynamic parameter space to be probed. The analysis will provide a heatmap of the root mean square deviation (RMSD) values for each variation of the parametrization, providing a visual feedback of the solution's error surface (Fig. 8.11). Two-dimensional combination graphs can be created both from enhanced van Holde–Weischet analyses (see Fig. 8.6) or any of the whole boundary

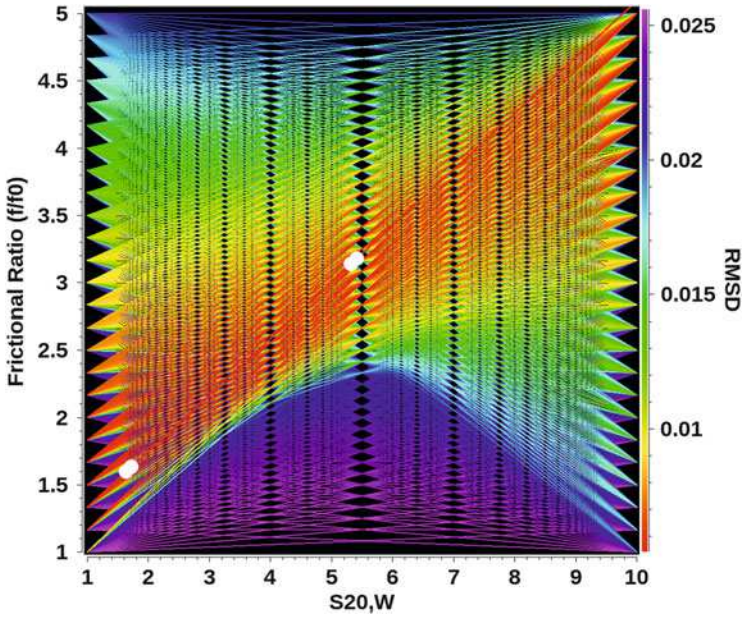

Fig. 8.11 PCSA error surface heat map for SV data from a mixture of lysozyme and a 208 bp DNA fragment using a straight line parametrization. The white dots show the solutes found for the line that best fits the SV data. The color gradient indicates the RMSD value for each line

modeling programs (Fig. 8.12). The former program is well suited to compare the results from multiple experiments where a parameter like the solute concentration, the ionic strength or the pH of the buffer is modulated, or mutants and wildtype are compared, or a titration of a binding partner is followed. The latter program provides detailed views of all hydrodynamic parameters and is well suited to compare results from multiple analysis methods for the same sample. For multi-wavelength data, either three-dimensional plots or movies can be used to visualize the data. Three-dimensional plots are appropriate for simultaneously viewing spectral and hydrodynamic sample properties in a single plot (Fig. 8.13). Movies can be created by showing individual pseudo-three-dimensional plots (Fig. 8.9, left panel), where each frame represents the next wavelength in a sequence of wavelengths. To enable customization of visualization output in US3, each graph has a configuration button that allows customization of every plot element, such as axes, labels, legends, lines, symbols, colors, grid lines, plot canvas, and graphing items.

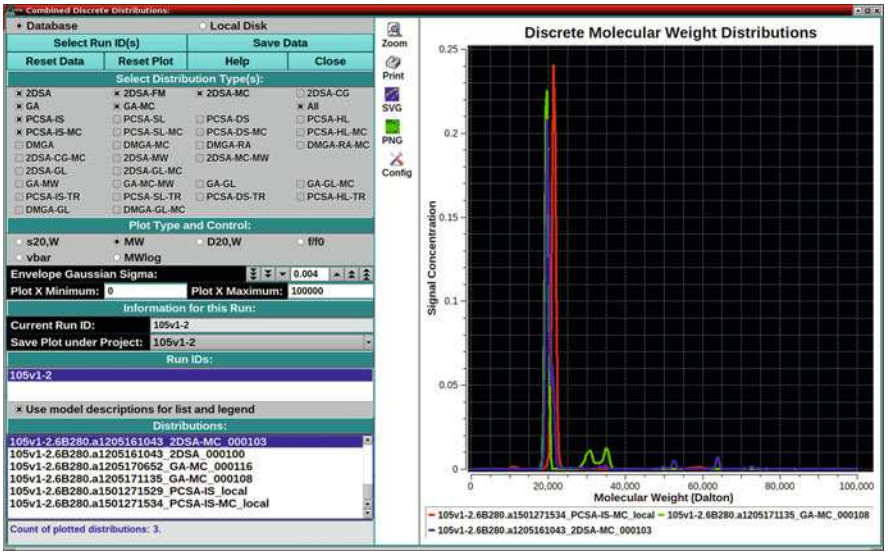

Fig. 8.12 Discrete molecular weight distributions for a 19.7 kDa protein analyzed with Monte Carlo for 2DSA, PCSA and GA

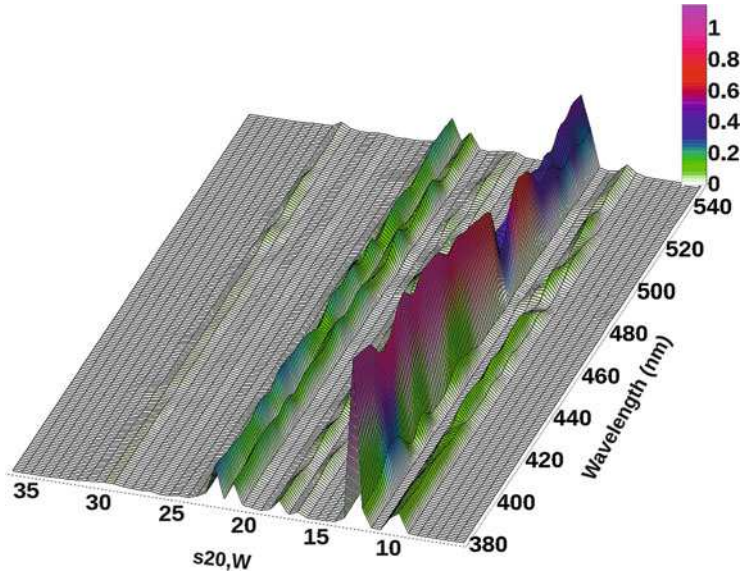

Fig. 8.13 Three-dimensional view of a global multi-wavelength analysis showing the relative absorbance for different hydrodynamic species for each fitted wavelength (Credits: Robert Whetten, German Plascencia, UTSA)

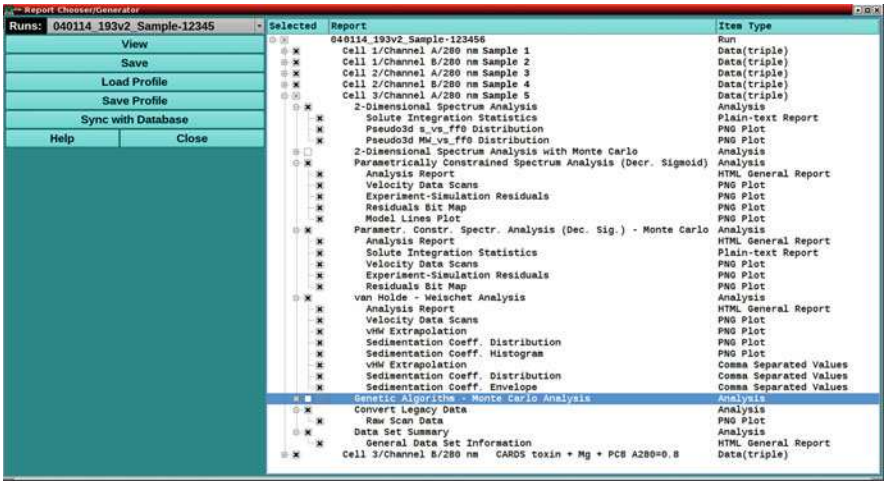

Fig. 8.14 Report Generator. Individual report items from each triple can be included into a PDF-formatted report, and the selection profile can be stored

8.4.10 Reports

Completed analysis results are captured and saved to disk and database. Each analysis generates a selection of graphs, analysis report records, and portable spreadsheets of the resulting data. The user can retrieve all result documents either from the USLIMS or through the desktop version. In the desktop version, a hierarchy of completed analysis methods and record items is presented for each experiment. The user can create a custom report by selecting desired items from the tree and then either generate a pre-formatted PDF document of all selected items or print it directly. If a particular selection profile is used repeatedly, the user can save this profile and re-apply it to other experiments to regenerate the desired report selection. The user can also migrate report records to a new computer by synchronizing the local storage with the database (Fig. 8.14). In the USLIMS, a report can be created dynamically from any record stored in the database. After selecting the experiment, any triple in the experiment provides a link to all report records belonging to this triple, sorted by analysis method, which the user can view or download. Graphics are provided both in PNG compressed bitmaps and also in scalable vector graphics format, suitable for post-editing at arbitrary resolution.

8.5 Simulation Programs

US3 offers a range of simulation modules to assist the user in designing experiments and interpreting results. These include a graphical finite element modeling program for the simulation of SV data from any model stored locally or in the database.

All parameters and boundary conditions for the run can be modified using an instrument control panel (rotor speed, length of run, number of scans, scan delay, rotor acceleration rate, temperature), as well as data parameters such as noise levels for TI, RI and stochastic noise components, selection of boundary or band-forming centerpieces, meniscus and cell bottom positions. Pre-defined buffers can be chosen to apply viscosity and density corrections. Multi-speed experiments can be simulated by defining speed profiles. Finite element solutions can be generated using multiple discretization schemes (radius: adaptive space, fixed Claverie mesh, moving hat, specified mesh file, or finite volume meshes needed for advanced simulations, time: fixed or adaptive). All profiles can be saved and applied to future simulations. The ASTFEM/ASTFVM solutions programmed in US3 can produce simulations for arbitrary advanced models, including reacting systems with kinetics, concentration dependent non-ideality, co-sedimenting solutes and solvent compressibility. Models for such experiments can be created with the US3 model editor. Simulations can be shown in accelerated time as movies. US3 has an equilibrium simulation program that predicts the time it takes to reach equilibrium based on molecular weight, rotor speed, centerpiece position, and column height. Multiple speed steps can be simulated. A self-association simulator can be used to predict relative concentrations of individual oligomers in a reversible self-associating monomer n-mer m-mer system as a function of concentration, where the equilibrium constants for each association reaction can be supplied by the user. Finally, US3 offers two hydrodynamic calculators. The first takes as input the molar mass, partial specific volume and the solution conditions, as well as an axial ratio to predict s, D, f, f/f_0, and the dimensions of the two axes of a prolate and oblate ellipsoid, and for a long rod model and a sphere. The second module takes two of the three parameters molar mass, s and D to predict axial ratios and dimensions for the same two ellipsoids and the long rod model, and predicts the Stokes radius, f/f_0, and the remaining hydrodynamic parameters.

8.6 Utilities

The US3 software offers a number of utilities that support sedimentation analysis. First, an export module can be used to convert OpenAUC formatted data to traditional Beckman ASCII file format. For intensity data, the user has the choice to export either intensity or pseudo-absorbance data. This makes US3 fully backward compatible with other analysis packages and offers a conversion path between traditional Optima XLA/I acquired data and newer multi-wavelength instruments collecting data in binary OpenAUC format. A second utility is used to calibrate rotor stretching. This calibration allows US3 to predict the precise displacement of the cell bottom due to stretching of the rotor as a function of speed and will calculate an exact position of the cell bottom at any speed, since an exact centerpiece geometry is also measured and stored for each centerpiece in US3. The exact bottom position is critical for a correct solution of the Lamm equation in all whole boundary fitting

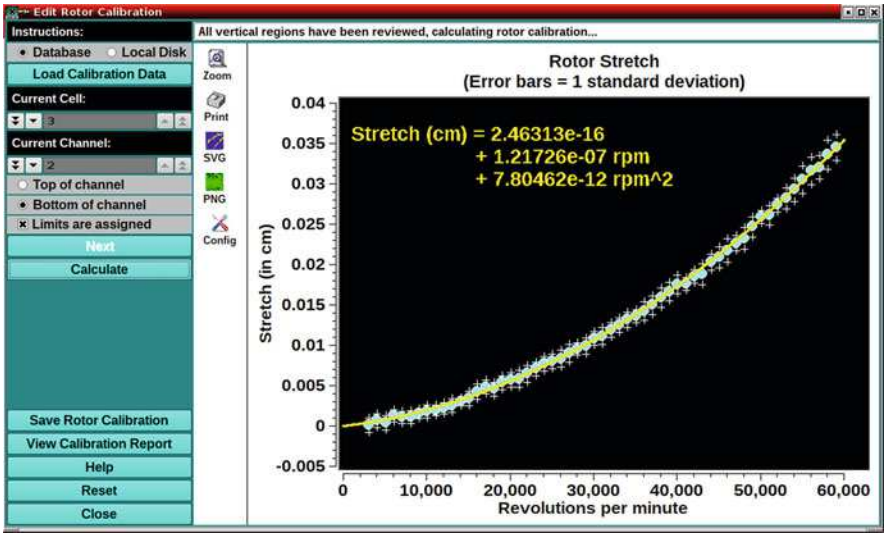

Fig. 8.15 Rotor Calibration program showing a stretch calibration for a Beckman An60Ti 4-hole rotor

programs, especially for multi-speed experiments. First, a calibration experiment is performed where an empty two- or six-channel centerpiece is placed into each rotor hole. All channels are radially scanned in intensity mode. This records positions for the edges of each channel in the centerpiece. The first scan of all cells, including the counterbalance, is collected at 3000 rpm. Subsequently, the rotor is accelerated in 1000 rpm increments to maximum speed, pausing every 1000 rpm and scanning all cells and channels, generating a total of 57 scans for each cell and channel (47 for the 8-hole rotor). The data are cropped to each edge of a centerpiece channel. For each speed, the center of the edge is determined by the calibration algorithm, and the average displacement for each speed increment is calculated and plotted against rotor speed. These data are then fitted to a second-order polynomial, and the baseline is adjusted such that the displacement at zero rpm is zero. A fit for a typical rotor calibration is shown in Fig. 8.15. The stored calibrations for each rotor are read automatically based on the associations made for each cell of each experiment every time finite element calculations are made. This mechanism provides a more accurate alternative to the introduction of another fitting parameter for the bottom of the cell position. A configuration utility offers flexible configurations for personal preferences such as fonts, color schemes, advanced interface options, debug levels, default file locations, and for database connectivity options, including passwords. Individual database dialogs provide access to tables which may be stored locally or in a remote database. These dialogs allow the user to retrieve and edit existing records, and create new records for investigator information, buffer files, analytes, solutions, experimental data, edit profiles, models, noise files, projects, rotors, and rotor calibrations. Database users can use the data management tool to synchronize

any computer with the contents of a remote database, which makes US3 data inherently portable and independent of location. The data managed with these dialogs will always be synchronized with the USLIMS information. US3 also features a complete online help menu that covers each routine with a context-specific help module.

References

Bhattacharyya SK et al (2006) Development of fast fiber based UV-vis multiwavelength detector for an ultracentrifuge. Progr Colloid Polym Sci 131:9–22

Brookes EH, Demeler B (2007) Parsimonious regularization using genetic algorithms applied to the analysis of analytical ultracentrifugation experiments. In: GECCO '07 proceedings of the 9th annual conference on genetic and evolutionary computation. ACM, New York, pp 361–368. ISBN: 978-1-59593-697-4. doi:10.1145/1276958.1277035

Brookes EH, Demeler B (2008) Parallel computational techniques for the analysis of sedimentation velocity experiments in UltraScan. Colloid Polym Sci 286:138–148

Brookes EH, Boppana RV, Demeler B (2006) Computing large sparse multivariate optimization problems with an application in biophysics. In: Proceedings of the SC 2006 conference. ACM/IEEE. doi:10.1109/SC.2006.18 E-ISBN: 0-7695-2700-0 Print ISBN: 0-7695-2700-0 INSPEC Accession Number: 9343095

Brookes EH, Cao W, Demeler B (2010a) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J 39:405–414

Brookes E, Demeler B, Rosano C, Rocco M (2010b) The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. Eur Biophys J 39(3):423–435

Brookes E, Demeler B, Rocco M (2010c) Developments in the US-SOMO bead modeling suite: new features in the direct residue-to-bead method, improved grid routines, and influence of accessible surface area screening. Macromol Biosci 10(7):746–753. doi:10.1002/mabi.200900474

Brookes E, et al (2013) US-SOMO cluster methods: year one perspective. XSEDE '13. In: Proceedings of the conference on extreme science and engineering discovery environment: gateway to discovery, at San Diego, California 01/2013. doi:10.1145/2484762. 2484815. ISBN: 9781450321709

Cao W, Demeler B (2005) Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation. Biophys J 89(3):1589–1602

Cao W, Demeler B (2008) Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution for multi-component reacting systems. Biophys J 95(1):54–65

Cölfen H et al (2010) The open AUC project. Eur Biophys J 39(3):347–359

Demeler B (2009) High-resolution modeling of hydrodynamic experiments with UltraScan. http://www.XSEDE.org/gateways/projects.php?id=77

Demeler B (2010) Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Curr Protoc Protein Sci 07: Unit 7.13. doi:10.1002/0471140864.ps0713s60 PMCID: PMC4547541

Demeler B, Brookes EH (2008) Monte Carlo analysis of sedimentation experiments. Colloid Polym Sci 286:129–137

Demeler B, van Holde KE (2004) Sedimentation velocity analysis of highly heterogeneous systems. Anal Biochem 335:279–288

Demeler B et al (2010) Characterization of reversible associations by sedimentation velocity with UltraScan. Macromol Biosci 10(7):775–782. PMID: 20486142

Demeler B et al (2014) Characterization of size, anisotropy, and density heterogeneity of nanoparticles by sedimentation velocity. Anal Chem 86(15):7688–7695

Gorbet G et al (2014) A parametrically constrained optimization method for fitting sedimentation velocity experiments. Biophys J 106:1741–1750

Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton JC (eds) Analytical ultracentrifugation in biochemistry and polymer science. Royal Society of Chemistry, Cambridge, pp 90–125

Marru S, et al (2011) Apache Airavata: a framework for distributed applications and computational workflows. In: GCE '11 Proceedings of the 2011 ACM workshop on Gateway computing environments, pp 21–28, ACM New York, NY, USA. Apache Airavata: http://airavata.apache.org/

Memon S, et al (2013) Improvements of the UltraScan scientific gateway to enable computational jobs on large-scale and open-standards based cyberinfrastructures. In: XSEDE '13 Proceedings of the conference on extreme science and engineering discovery environment: gateway to discovery No. 39, ACM New York, ISBN: 978-1-4503-2170-9, doi:10.1145/2484762.2484800

Memon S, et al (2014) Advancements of the UltraScan scientific gateway for open standards-based cyberinfrastructures. Concurr Computat Pract Exper, Wiley. doi: 10.1002/cpe.3251

Schilling K (2014) A multiwavelength capable detector for the analytical ultracentrifuge (Nanolytics, Potsdam, Germany). Personal Communication

Schuck P, Demeler B (1999) Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. Biophys J 76(4):2288–2296

Stafford WF (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(2):295–301

UltraScan-III wiki: http://wiki.bcf2.uthscsa.edu/ultrascan3/