UltraScan: High-resolution Modeling of Analytical Ultracentrifugation Experiments on TeraGrid

Emre Brookes
Borries Demeler
Department of Biochemistry
5 August 2010
Overview

• Background
  – What is Analytical Ultracentrifugation (AUC)?
  – What does an AUC experiment provide?
  – How do we obtain results?
  – Example results

• TG Science Gateway

• User community
  – Publications

• Future
Background: What is AUC?

- AUC is an important technique for the **solution** study of macromolecules
- Molecules are not fixed to a microscope grid
- Molecules are not distorted by crystal packing forces (vs X-Ray crystallography)
- Very large size range (complements cryo-EM and NMR)
- Dynamic processes can be studied
- Conformational changes
Background: What is AUC?

- Sample placed in cell
- Run Ultracentrifuge
  - Usually 20-60k RPM
- Collect data
  - 4 to 24 hours or more
- Analyze the data
Background: What does it provide?

- Time series of radial concentration profiles
Background: What does it provide?

- Time series of radial concentration profiles

From this, we can determine:
Background: What does it provide?

- Time series of radial concentration profiles

From this, we can determine:

- How many different types of molecules are present in the solution
Background: What does it provide?

- Time series of radial concentration profiles

From this, we can determine:

- How many different types of molecules are present in the solution
- What are there molecular weights and shapes (one dimension of shape – spherical to ellipsoidal)
Background: What does it provide?

- Time series of radial concentration profiles

From this, we can determine:

- How many different types of molecules are present in the solution
- What are there molecular weights and shapes (one dimension of shape – spherical to ellipsoidal)
- Are interactions present
Background: How do we obtain results?

- The Lamm equation is a PDE which describes an ideal solute (a molecule in solution).

\[
\left( \frac{\partial C}{\partial t} \right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - Dr \frac{\partial C}{\partial r} \right]_t
\]

- Two solute-specific parameters (s, D)
  - Other parameters are global to the experiment
- Build finite element models (FEM) for each (s,D) in range
- Fit collections of (s,D) FEMs to experimental data.
  - Multistep procedure
Background: How do we obtain results?

- Key breakthrough algorithms:
  - 2DSA
    - Solves arbitrarily large non-negative least squares problems
    - Breakthrough algorithm which allowed our advanced analysis techniques
      - Brookes, Boppana, Demeler, *Computing Large Spare Multivariate Optimization Problems with an Application in Biophysics*. SC06 ACM 0-7695-2700-0/06
  - GA
    - Parsimonious regularization “sharpens” for sparse solution space
      (vs. “smoothing” of max-entropy or Tikhonov)
Example Results: Fit of DNA digest

Digest of pPOL-1 208-12. Individual fragments can be distinguished.
Example Results: Protein – 1 day old

CuZn Superoxide Dismutase Mutant - Metalated form, freshly purified
(Data provided by P.J. Hart & Sai Venkatesh Seetharaman)
Example Results: Protein – 7 days old

Apo-form showing clear signs of degradation
Example: Clathrin baskets assembling from clathrin triskelia (A). The sample also displays several nonglobular species which represent the building block subunits required for assembly of intact baskets (B, D). Sample shows a large heterogeneity of different sized baskets that assume a mostly globular form with a unity frictional ratio (B,C). (Data kindly provided by E. Lafer, UTHSCSA, Dept. of Biochemistry)
TeraGrid Science Gateway

- Center for Analytic Ultracentrifugation of Macromolecular Assemblies

URL: http://cauma.uthscsa.edu
Welcome to the UltraScan LIMS Portal User Area

Active Account: Emre Brookes

Help:

- Help and Instructions

1. Project Area:

- Start a new Project

2. Enter Associated Data:

- Enter Peptide Sequence
- Enter Nucleic Acid Sequence
- Enter Buffer Information
- Upload an Image (Gel Picture, Absorbance Spectrum, etc.)

3. Sample Information:

- Enter Sample Information

4. Experiment Request:

- Enter Experiment Request

5. Analysis Request:

- Enter Analysis Request
- View Analysis Queue
- View Current Supercomputer Load
- HPC Hours Calculation
Queue Viewer

Datasets currently in the queue:
5-30.veloc.11  Remove

<table>
<thead>
<tr>
<th>Setup 2DSA Control</th>
<th>Setup 2DSA Control with MW Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup GA Control</td>
<td>Setup GA Control with MW Constraint</td>
</tr>
<tr>
<td>Select Experiment</td>
<td>Select Additional Cell</td>
</tr>
<tr>
<td>Clear Queue</td>
<td></td>
</tr>
</tbody>
</table>
## HPC Data Analysis Queue Viewer:

Queue status snapshot as of Thu Apr 17 10:49:45 CDT 2008:

<table>
<thead>
<tr>
<th>Job</th>
<th>Name</th>
<th>Owner</th>
<th>Status</th>
<th>Analysis Type</th>
<th>Submitted on</th>
<th>Running on</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5-30.veloc.11</td>
<td><a href="mailto:ebrookes@cs.utsa.edu">ebrookes@cs.utsa.edu</a></td>
<td>Active</td>
<td>GA</td>
<td>04/17/08. at 10:48:25</td>
<td>alamo.uthscsa.edu</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5-30.veloc.11</td>
<td><a href="mailto:ebrookes@cs.utsa.edu">ebrookes@cs.utsa.edu</a></td>
<td>Active</td>
<td>2DSA-MC-40</td>
<td>04/17/08. at 10:48:48</td>
<td>laredo.uthscsa.edu</td>
<td></td>
</tr>
</tbody>
</table>

Delete all jobs?

Delete All

Back to user home
2DSA control parameters were defined as follows:

Monte Carlo iterations: 25
s minimum: 1
s maximum: 10
s resolution: 12
f/f0 minimum: 1
f/f0 maximum: 4
f/f0 resolution: 12
Grid repetitions: 16
Fit TI noise: on

The results of your 2DSA analysis involved datasets:

Experiment 5-60, cell 1, wavelength 1, solutes 633, rmsd 0.00578876, iterations 0 submitted at 10:17:13 on 03/08/2008 has completed.

The results are attached.

Results processed by TIGRE on bcf.uthscsa.edu
### Information for this Run:

- **Run Id:** 5-60
- **Temperature:** 20.000 °C
- **Available Cells:** 1

### Simulated Velocity Data

<table>
<thead>
<tr>
<th>Cell 1</th>
<th>Wavelength 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell 2</td>
<td>Wavelength 2</td>
</tr>
<tr>
<td>Cell 3</td>
<td>Wavelength 3</td>
</tr>
<tr>
<td>Cell 4</td>
<td></td>
</tr>
<tr>
<td>Cell 5</td>
<td></td>
</tr>
</tbody>
</table>

### Experimental Parameters (at 20°C):

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>0.998200</td>
</tr>
<tr>
<td>Viscosity</td>
<td>1.000400</td>
</tr>
<tr>
<td>(v_{\text{bar}})</td>
<td>0.7144</td>
</tr>
<tr>
<td>RMSD</td>
<td>7.3903e-03</td>
</tr>
</tbody>
</table>

### Velocity Data for 5-60

The image shows a graph of velocity data for the experiment with a range of radii from 5.6 to 7.2 in cm, and a range of absorbance at 999 nm from 0 to 0.8.
<table>
<thead>
<tr>
<th>Genetic Algorithm Control Window</th>
<th>Distribution Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Initial Solutes:</strong> 0</td>
<td></td>
</tr>
<tr>
<td><strong>f/f0 Minimum:</strong> 1</td>
<td></td>
</tr>
<tr>
<td><strong>f/f0 Maximum:</strong> 4</td>
<td></td>
</tr>
<tr>
<td><strong>Width of s Bucket:</strong> 0.1498</td>
<td></td>
</tr>
<tr>
<td><strong>Width of f/f0 Bucket:</strong> 0.0170</td>
<td></td>
</tr>
</tbody>
</table>

**Pseudo-3D Controls**

- **Pseudo-3D Resolution:** 90
- **X Resolution:** 800
- **Y Resolution:** 800
- **X-pixel width:** 2
- **Y-pixel width:** 2
- **Automatic Plot Limits** (unselect to override)
  - **Plot Limit f/f0 min.:** 1.12496
  - **Plot Limit f/f0 max.:** 2.14511
  - **Plot Limit s min.:** 1.901
  - **Plot Limit s max.:** 10.889

**Load Color File**

<table>
<thead>
<tr>
<th>Help</th>
<th>1-Dimensional Plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print</td>
<td>2-Dimensional Plot</td>
</tr>
<tr>
<td>Load Distribution</td>
<td>Pseudo 3-D Plot</td>
</tr>
<tr>
<td>Manually Draw Bins</td>
<td>Shrink Bins</td>
</tr>
<tr>
<td>Autobassign Solute Bins</td>
<td>Reset Solute Bins</td>
</tr>
<tr>
<td>Close</td>
<td>Save</td>
</tr>
</tbody>
</table>

Now either auto-assign the solute bins, or manually select bins by clicking first on the lower and then on the upper limit of the bin range. If you auto-assign the bins you should first select the number of solute bins you want to use. UltraScan will space the bins proportional to the integral value of each peak, such that each bin contains the same integral value. You can select each solute bin from the listbox on the left and modify the frictional ratio limits by selecting them first with the respective counters. To change the frictional ratios for the solutes, change to the desired f/f0 values in the counter, then double-click on the...
TG SG
Usage 2007-10

- Job statistics for UltraScan project for approximately the last 4 years.

- Only partial data is available for 2007 (2nd half) and 2010 (thru June), and only successful runs are included.

**A.** Totals of CPU hours consumed from TeraGrid, UTHSCSA and international resources

**B.** Number of investigators whose data were analyzed (left Y-axis), and number of submitted jobs (right Y-axis).

- Both panels indicate increasing usage and need for TeraGrid resources and an increasing number of investigators requiring access to these resources.
User Community: Publications

- Since the development of our advanced methods, virtually every publication from our lab has used these methods.
- We currently count 35 peer reviewed journal publications and poster abstracts.
- Many additional presented talks where these methods have provided important new detail to the investigations of biological as well as synthetic polymer systems.
- We are aware of at least another 25 publications that were facilitated by our methods from other laboratories using our TeraGrid applications.
User Community: Publications

• Major problems successfully addressed with our methods include:


• Studies of band-gaps of CdTe nanoparticles using the innovative multi-wavelength analysis
User Community: Publications

• A few additional publications:


User Community: Laboratories

1. Academia Sinica, Institute of Biological Chemistry - Taiwan
2. Amylin Pharmaceuticals, Inc. - San Diego, California
3. A. H. Bakh Institute of Biochemistry - Russian Academy of Sciences, Moscow, Russia
4. Burnet Institute - Melbourne, Australia
5. Charles University - Prague, Czech Republic
6. Center for Analytical Ultracentrifugation, EMBL - Heidelberg, Germany
7. Colorado State University - Fort Collins, Colorado
8. Ecole Polytechnique Fédérale de Lausanne - Lausanne, Switzerland
9. Florida State University - Tallahassee, Florida
10. Forschungsinstitut für Molekulare Pharmakologie - Berlin, Germany
11. Heinrich Heine University - Düsseldorf, Germany
12. IGBMC - Strasbourg, France
13. Indiana University - Bloomington, Indiana
15. Marshall University - Huntington, West Virginia
16. Max Planck Institute for Colloids and Interfaces - Golm, Germany
17. McGill University - Montreal, Canada
18. MD Anderson Cancer Center - Houston, Texas
19. NIH - NHLBI - Bethesda, Maryland
20. Oregon State University - Oregon State University, Corvallis, Oregon
21. Osaka University - Osaka Japan
22. Pasteur Institute - Paris, France
23. Public LIMS Portal - Open Access
24. Rice University - Houston, Texas
25. Shriners Hospital for Crippled Children - Portland, Oregon
27. State University New York (SUNY) - Albany, New York
28. Stony Brook University - Stony Brook, New York
29. Technische Universität München - Garching, Germany
30. Texas A&M University - College Station, Texas
31. University of Bristol - Bristol, UK
33. University of Mainz, Molecular Biophysics - Mainz, Germany
34. University of Massachusetts, Medical School - Worcester, Massachusetts
35. University of Melbourne, Australia - Melbourne, Victoria
36. University of Michigan - Ann Arbor, Michigan
37. University of Missouri, Columbia - Columbia, Missouri
38. University of Montana - Missoula, Montana
39. National Center for Macromolecular Hydrodynamics, University of Nottingham - Nottingham, UK
40. University of Texas at Austin - Austin, Texas
41. University of Texas Health Science Center - San Antonio, Texas
42. University of Toronto - Toronto, Canada
43. University of Utrecht - Utrecht, Netherlands
44. University of Victoria - British Columbia, Canada
45. University of Washington - Seattle, Washington
46. University of Washington, Klevit Lab - Seattle, Washington
Future

● Improved gateway
  – ASTA, GFAC

● Multiwavelength data
  – Datasets 3 orders of magnitude larger

● MD simulation
  – DMD
  – BD

● Global solution studies
  – Multiple AUC experiments
  – DLS
  – SAXS/SANS
Acknowledgments

Thanks to:

- Our team at UTHSCSA:
  - Jeremy Mann, Bruce Dubbs, Dan Zollars, Gary Corbet, Virgil Schiff.
- TACC:
  - Chris Hempel, Margaret Murray, Jay Boisseau
- IU
  - Raminder Singh, Suresh Marru, Marlon Pierce
- SDSC
  - Nancy Wilkins-Diehr
- LONI
- All the people behind the scenes that help make it all happen.
- & If we have forgotten someone, you are doubly thanked!
- Supported by
  - To Demeler:
    - NIH/NCRR 1R01 RR022200
    - NIH/NCRR 3R01RR022200-03S1
    - NSF TG-MCB070039,40
    - NSF/ASTA TG-MCB07038
  - To Brookes
    - NIH/NIGMS 1K25GM090154-01A1