Advanced Data Analysis:  
SEC, SEC-MALS, SEC-SAXS, Reconstructions

Everything BioSAXS 5  
Getting Started in Biological Small-Angle X-ray Solution Scattering  
Tuesday 11/5/19

Kushol Gupta  
Research Asst. Professor  
Department of Biochemistry and Biophysics,  
Perelman School of Medicine (Univ. of Penn.)  
kgupta@pennmedicine.upenn.edu
Advanced Data Analysis

- **SEC and SEC-MALS**
- **SEC-SAXS**
  - SVD-EFA
- **Ab Initio Reconstructions**
  - DAMMIN/F
  - GASBOR
  - DENSS
  - MONSA
Principles of SEC

"Size-Exclusion Chromatography: Principles and Methods,” GE Healthcare Life Sciences
Principles of SEC

“Size-Exclusion Chromatography: Principles and Methods,” GE Healthcare Life Sciences
Principles of SEC

“Size-Exclusion Chromatography: Principles and Methods,” GE Healthcare Life Sciences
Principles of SEC

“Size-Exclusion Chromatography: Principles and Methods,” GE Healthcare Life Sciences

V ~ B - A * log(MW)
Principles of SEC


The use of gel chromatography for the determination of sizes and relative molecular masses of proteins

Interpretation of calibration curves in terms of gel-pore-size distribution

Marc Le MAIRE,*‡ Alexandre GHAZI,† Jesper V. MØLLER‡ and Lawrence P. AGGERBECK*‡

*Centre de Génétique Moléculaire, Laboratoire propre du Centre National de la Recherche Scientifique, Associé à l’Université Pierre et Marie Curie (Paris VI), 91190 Gif-sur-Yvette, France, †Laboratoire des Biomembranes (UA 1116), Bat. 433, Université Paris-Sud, 91405 Orsay, France, and ‡Institute of Medical Biochemistry, Aarhus University, DK 8000 Aarhus, Denmark

<table>
<thead>
<tr>
<th>Protein</th>
<th>Source</th>
<th>$R_s$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cobalamin</td>
<td></td>
<td>0.85</td>
</tr>
<tr>
<td>Cytochrome $c$</td>
<td>Horse heart</td>
<td>1.7</td>
</tr>
<tr>
<td>Ribonuclease II-A</td>
<td>Bovine pancreas</td>
<td>1.75</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>Horse skeletal muscle</td>
<td>1.9</td>
</tr>
<tr>
<td>Haemoglobin</td>
<td>Human</td>
<td>2.4</td>
</tr>
<tr>
<td>Ovalbumin</td>
<td>Chicken egg</td>
<td>2.8</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td><em>Escherichia coli</em></td>
<td>3.3</td>
</tr>
<tr>
<td>Albumin</td>
<td>Bovine serum</td>
<td>3.5</td>
</tr>
<tr>
<td>Transferrin (iron-free)</td>
<td>Human</td>
<td>3.6</td>
</tr>
<tr>
<td>Aldolase</td>
<td>Rabbit muscle</td>
<td>4.6</td>
</tr>
<tr>
<td>Catalase</td>
<td>Ox liver</td>
<td>5.2</td>
</tr>
<tr>
<td>Aspartate transcarbamylase</td>
<td><em>Escherichia coli</em></td>
<td>6.0</td>
</tr>
<tr>
<td>Ferritin</td>
<td>Horse spleen</td>
<td>6.3</td>
</tr>
<tr>
<td>$\beta$-Galactosidase</td>
<td><em>Escherichia coli</em></td>
<td>6.9</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>Ox thyroid</td>
<td>8.6</td>
</tr>
</tbody>
</table>
Principles of SEC

SEC-MALS

18+1+1=20 channels of data being collected every second!
\[
\frac{Kc}{R(\theta, c)} = \frac{1}{M_w P(\theta)} + 2A_2c
\]

- \(K\): constants
- \(c\): concentration from UV or RI
- \(R\): from measured intensity
- \(P\): from scattering angle
- \(A_2\): 2\textsuperscript{nd} virial coefficient (fit)

\[
M_w = \frac{\sum c_i m_i}{\sum c_i}
\]

(Weight-average molecular weight; mass concentrations)

- Global fit of \(M_w\) values with 18 \(R(\theta), P(\theta)\) values
- Calculation is done “on-the-fly”, every few seconds of SEC column elution
- Larger particles have stronger scattering signal; can use less material
- Need \(~20-100\) uL of concentrated protein; gets diluted on column
SEC-MALS

results graph

control graph
SEC-MALS

Molar Mass vs. Time

Hemoglobin 52-62kD

BSA 132kD dimer

BSA 66kD monomer
Example: Preparation of Group II Intron from Lactococcus Lacti.
SEC-MALS

What is M for this protein?

If M=26 kDa, what is the oligomeric state?  

Monomer
You can often establish that your protein is monomeric from a SEC experiment alone. You cannot conclude anything about your protein’s oligomeric state from SEC alone!

Gupta et al 2015, JBC
Advanced Data Analysis

- SEC and SEC-MALS
- **SEC-SAXS**
  - SVD-EFA
- **Ab Initio** Reconstructions
  - Calculating Profiles
  - DAMMIN/F
  - GASBOR
  - DENSIS
  - MONSA
SEC-SAXS: Suggested Reading


Figure S9: $R_g$ and $I(0)$ across the elution peak of wt-PheH with and without L-Phe. Although only a single elution peak is observed in each case, the $R_g$ values determined by Guinier analysis vary, suggesting the presence of multiple species. While the low signal-to-noise of the profiles at the beginning of the run complicates Guinier analysis, it is clear that the $R_g$ is significantly higher than 40 Å (dotted line).

SEC-SAXS: SVD

Applications for biological studies:

- Oligomerization
- Multiple assembly forms
- Temperature dependent transitions
- Ligand-dependent transitions

Polydisperse & interactive systems:

- Equilibrium oligomeric mixtures (OLIGOMER, COSMIC)
- Assembly/disassembly processes (SVDPLOT, MIXTURE)
- Natively unfolded proteins and multidomains proteins with flexible linkers (EOM, SASSIE, BILBOMD)
- SVD-EFA (RAW), US-SOMO – SEC-SAXS
The number of significant singular vectors in SVD (i.e. non-random curves with significant singular values) yields the minimum number of independent curves required to represent the entire data set by their linear combinations (e.g. for mixtures). (Konarev)
Figure S10: Conventional SVD of SEC-SAXS data from wt-PheH in 0 mM L-Phe. Three significant singular values are observed (top panel). Although the corresponding right singular vectors (columns of $V$) have shapes that are reminiscent of elution peaks, we find that there are sign changes within the curves. Moreover, when the scattering basis set is recovered by multiplying the columns of $U$ with the experimental error, we find that there are non-physical sign changes within the curves. Because SVD produces orthogonal singular vectors, they cannot represent physical states, such as elution peaks and scattering intensities, which must be positive numbers.
Evolving Factor Analysis
(Maeder M. Anal Chem. 1987;59:527–530)

Variant of SVD that allows for the identification of ranges within each elution peak where separate species elute from the abrupt changes in the number of significant singular values as scattering profiles are added or removed from the matrix.

SEC-SAXS: Troubleshooting

1. Radiation Damage: sticking to windows
   1. Flow Rate increase
   2. Add glycerol
   3. Attenuation

2. Column Resolution
   1. Overloading
   2. Injection volume too large
   3. Media interactions

---

SEC-SAXS: Troubleshooting

3. $R_g$ changes within Peak (also a problem with SEC-MALS)

- **Smiling:**
  - repulsive inter-particle interference.
  - Consider lowering sample concentration, adding salt to the buffer.

- **Frowning:**
  - aggregating post-column
  - attractive structure factor (stickiness)
  - weakly-associated oligomer

- **Sloping:** (especially if peak is also broad)
  - exchanging oligomers
  - overlapping peaks from multiple species

Steven Meisburger
The C-terminal tail of the NEIL1 DNA glycosylase interacts with the human mitochondrial single-stranded DNA binding protein

Nidhi Sharma\textsuperscript{a}, Srinivas Chakravarthy\textsuperscript{b}, Matthew J. Longley\textsuperscript{c}, William C. Copeland\textsuperscript{c}, Aishwarya Prakash\textsuperscript{a, *}

\textsuperscript{a} University of South Alabama, Mitchell Cancer Institute, 1660 Springhill Avenue, Mobile, AL 36604, United States
\textsuperscript{b} Illinois Institute of Technology, Advanced Photon Source, Argonne National Laboratory, 9700 S Cass Avenue, Argonne, IL 60439-4800, United States
\textsuperscript{c} Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States

Table 2
SAXS data collection parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Homo sapiens</th>
<th>Homo sapiens</th>
<th>Homo sapiens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>E. coli</td>
<td>E. coli</td>
<td>E. coli</td>
</tr>
<tr>
<td>UniProt sequence ID (residues in constructs)</td>
<td>Q04527 (17–148)</td>
<td>Q06414 (1–300)</td>
<td></td>
</tr>
<tr>
<td>SEC-SAXS column, l/300 mm Superdex S200</td>
<td>1.313</td>
<td>0.725</td>
<td></td>
</tr>
<tr>
<td>Loading concentration (mg ml(^{-1}))</td>
<td>7</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Injection volume (μl)</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Flow rate (ml/min)</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>Solvent (blanks taken from SEC flow through prior to elution of protein)</td>
<td>25 mM HEPES pH 7.4, 5% glycerol, 300 mM NaCl, and 1 mM DTT</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

(B) Data collection parameters

Beamline | APS | APS | APS |
Wavelength (Å) | 1.03 | 1.03 | 1.03 |
Q Range (Å\(^{-1}\)) | 0.0059–0.3892 | 0.0059–0.3892 | 0.0059–0.3892 |
Temperature (°C) | 25 | 25 | 25 |

(C) Software used for data reduction, analysis and interpretation

SAXS data reduction

PRIMUS (ATSAS 2.3.1)

Extinction coefficient estimate

PROtparam [63], Scater [43], BioXTAS RAW [47]

Shape/bead modeling

DAMMIF and DAMMIF (ATSAS 2.8.1)

(D) Structural parameters

R(0) from P(r) | 44.88 | 51.49 | 22.74 |
Rg (Å) from P(r) | 27.52 | 37.15 | 28.18 |
R(0) from Guinier | 44.83 ± 0.10 | 49.20 ± 0.24 | 22.77 ± 0.03 |
Rg from Guinier | 27.44 ± 1.21 | 33.04 ± 2.46 | 28.40 ± 0.86 |
Dmax (Å) | 103.21 | 149.96 | 98.74 |

(E) Molecular weight determination (kDa)

Expected Theoretical (Exapty) | 60.78 | 44.72 | 78 |
MW(Vo) | 61 | 46 | 79 |
MW(Vo) | 59 | 36 | 71 |
MW (MALS) | 59.8 ± 1.67% | 48.5 ± 2.23% | 69 ± 1.46% |

(E) Modeling parameters

Symmetry | P4 | P4 | P4 |
Particle anisometry | Oblate | Unknown | Oblate |
# of modeling iterations | 10 | 10 | 10 |
X\(^{2}\) of the model | 1.072 | 1.041 | 0.945 |
Advanced Data Analysis

- SEC and SEC-MALS
- SEC-SAXS
  - SVD-EFA
- *Ab Initio* Reconstructions
  - Calculating Profiles
  - DAMMIN/F
  - GASBOR
  - DENSS
  - MONSA
Ab Initio Reconstructions

Simple shapes

Collections of simple shapes

Spherical harmonics envelopes

Bead modeling

Hybrid modeling

Direct Electron Density Calculation

Thomas Grant
Calculating SAXS profiles from Models

• Debye equation:

\[ I(q) = \sum_i \sum_j f_i(q) f_j(q) \frac{\sin(q \cdot r_{ij})}{q \cdot r_{ij}} \]

• Calculation of reciprocal space scattering from model

• Accomplished vis Debye Equation or Spherical Harmonics Approximation

• Modeling solvent shell and excluded volume is the technical challenge
## Calculating SAXS profiles from Models

<table>
<thead>
<tr>
<th>Method</th>
<th>Spherical Averaging</th>
<th>Hydration Layer</th>
<th>Representation</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRY SOL [34]</td>
<td>Multipole expansion</td>
<td>Implicit water layer based on envelope function</td>
<td>Atomic</td>
<td>Server, download <a href="http://www.embl-hamburg.de/biosaxs/crysol.html">http://www.embl-hamburg.de/biosaxs/crysol.html</a></td>
</tr>
<tr>
<td>Fast-SAXS [38]</td>
<td>Debye formula</td>
<td>Explicit placement of water molecules</td>
<td>Coarse-grained, residue level</td>
<td><a href="http://yanglab.case.edu/software.html">http://yanglab.case.edu/software.html</a></td>
</tr>
<tr>
<td>Stovgaard et al. [40]</td>
<td>Debye formula</td>
<td>-</td>
<td>Coarse-grained, 1 or 2 points per-residue</td>
<td><a href="http://salilab.org/foxs/">http://salilab.org/foxs/</a></td>
</tr>
<tr>
<td>Virtanen et al. [44]</td>
<td>Debye formula or Cube model</td>
<td>HyPred based on MD simulations</td>
<td>Atomic, MD simulation</td>
<td>Source code, server, download <a href="http://sastb.als.lbl.gov/cgi-bin/intensity.html">http://sastb.als.lbl.gov/cgi-bin/intensity.html</a></td>
</tr>
<tr>
<td>Zernike Polynomials [45]</td>
<td>Zernike polynomial expansions</td>
<td>Hydration layer from voxelized representation</td>
<td>Atomic</td>
<td>Source code, server, download <a href="http://sastb.als.lbl.gov/cgi-bin/intensity.html">http://sastb.als.lbl.gov/cgi-bin/intensity.html</a></td>
</tr>
</tbody>
</table>

Schneidman-Duhovny, et.al. (2012) *BMC Structural Biology*
Ab Initio Reconstructions

- Synchrotron/Reactor Sources
- Home Source Rotating Anodes
- Detectors
- Software and Computing Power
- Algorithms for Shape Reconstruction

Spherical Harmonics (Envelope Function)
Stuhrmann, 1970; Svergun & Stuhrmann, 1991; Svergun et al., 1996

Bead Models (DAMMIN, MONSA)
Chacon et al., 1998; Svergun, 1999; Walther, Cohen & Doniach, 2000
Svergun, Petoukhov & Koch, 2001

Dummy Residue Models (GASBOR)
Svergun, Petoukhov & Koch, 2001

http://www.embl-hamburg.de/workshops/2008/embo/d-sv_abinitio.html
Ab Initio Reconstructions: DAMMIN/F/N

- DAMMIN/F uses a dummy atom/bead modeling approach that applies spherical harmonics

- Relies on low resolution data where s*Rg < 7-8 or q < 0.3, as contributions from solvent at high q can lead to errors

- 3D volume fits the data with physical constraints applied

- Penalties for envelopes that are loose, compact, or disconnected

- Simulated Annealing Method – each calculation is slightly different.
Ab Initio Reconstructions: DAMMIF/N

S1 shape reconstruction

Step 0 Temperature = 0.100E-02  Chi= 36.38

http://www.embl-hamburg.de/workshops/2008/embo/d-sv_abinitio.html
Ab Initio Reconstructions: DAMMIF/N

Process:

- DAMMIN/F: Calculate 5-20 ab initio reconstructions
- DAMSUP/DAMAVER/DAMFILT:
  - Models are aligned
- Normalized Spacial Discrepancy (NSD) calculated:
  - ~Avg of 0.5 implies good stability
  - ~0.7-0.9 implies fair stability, but more common with anisotropic particles
  - ~>1.0 implies poor stability
- Ensemble then averaged and filtered to yield final result
Ab Initio Reconstructions: DAMMIF/N

Gallery of ten DAMMIN shape reconstructions for a yeast Holliday junction, rendered as bead models with a sphere radius of 3.25 Å. The overall normalized spatial discrepancy (NSD) for these ten reconstructions range from 0.70 to 0.76, indicating a stable solution.

Final averaged and filtered reconstruction is rendered in grey spheres with a radius of 3.25 Å, rotated around the y axis. For reference, the averaged envelope without filtering is shown as small spheres. The shape has dimensions of 104.0 Å x 68.3 Å x 64.4 Å. According to HYDROPRO analysis, the shape has a predicted Stokes radius of 38 Å and a predicted S value of 5.3, largely consistent with DLS (R_s=35 Å) and Sedimentation velocity measurements (S_{20,w} of 4.4).
Ab Initio Reconstructions: DAMMIF/N and GASBOR Examples
Ab Initio Reconstructions: Uniqueness

Ab Initio Reconstructions: GASBOR example

- GASBOR uses dummy residue approach with explicit solvent models, which then allows for use of higher resolution data.

Ab Initio Reconstructions: DENSs

Bead modeling (uniform density)

1 2 3
4 5 6
7 8 9
10 11 12
13

Direct Electron Density Calculation

1 (16Å) 2 (25Å) 3 (23Å)
4 (24Å) 5 (24Å) 6 (27Å)
7 (22Å) 8 (24Å) 9 (26Å)
10 (24Å) 11 (28Å)


Ab Initio Reconstructions: MONSA


A.

![NSD Comparison](image)

B.

- Multiple Profiles → Increased Data-to-Parameters
- Experimental Contrast Information
- Simultaneous Solution

Gupta et al 2012 Structure
Ab Initio Reconstructions: MONSA

MONSA Shape Reconstruction

Gupta et al 2012 Structure
Ab Initio Reconstructions: MONSA

Supplemental Table S3 — Comparison of the hydrodynamic parameters from shape reconstructions and experimentally determined values

<table>
<thead>
<tr>
<th></th>
<th>ERRγLBD</th>
<th>Binary Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sedimentation Coefficient of SAXS reconstructed shape*</td>
<td>3.5</td>
<td>3.4</td>
</tr>
<tr>
<td>Sedimentation Velocity Analysis</td>
<td>3.8</td>
<td>3.95</td>
</tr>
<tr>
<td>Rₜ of SAXS reconstructed shape*</td>
<td>35.3 Å</td>
<td>50.6 Å</td>
</tr>
<tr>
<td>Rₜ from SEC Analysis</td>
<td>33.2 Å</td>
<td>51.7 Å</td>
</tr>
</tbody>
</table>

* As calculated by HYDROPRO

Disorder-to-Order Structural Transition in the Assembly of the PGC-1α/ERRg Metabolic Hub, Srikripa Devarakonda, Kushol Gupta, Michael J. Chalmers, John F. Hunt, Patrick R. Griffin, Gregory D. Van Duyne, Bruce M. Spiegelman, PNAS 2011

Biophysical properties determined by orthogonal approaches should agree.