Small Angle Scattering
Ab initio modelling
Requirements for Data Collection

- sample must be pure and monodisperse
- sufficient volumes for multiple concentrations
- know your sample concentrations
- do not make up buffers from scratch, use dialysis buffer

Know what you do not know and what you want to learn from SAXS!
SAXS Data Processing: Bird’s Eye View

Primary Processing

Regularized Data

Structural Parameters

Experimental Data

Ab Initio Modelling

Model Refinement

Dummy Atom Models

Dummy Residue Models

Electron Microscopy Maps

Variability, Resolution

Superposition

Hybrid Methods

Validation

Rigid Body Models

Missing Fragments

Flexibility

Mixture Analysis

Contacts

Atomic Models

Sequence
Small Angle X-Ray Scattering

a. Radial averaging

b. Diagram showing log(I(s), relative) vs. s, nm\(^{-1}\) with regions labeled for size, shape, and fold.
Fingerprinting of SAS Curves

Solid sphere

Hollow sphere

Flat disc

Dumbbell

Long rod
**Principle of SAS Modeling**

1D scattering data
(or multiple data sets)

**3D search model**

\[ \mathbf{x} = \{X\} = \{X_1 \ldots X_M\} \]

M parameters

**Non-linear search**

**Trial-and-error**

**discrepancy:**

\[ \chi^2 = \frac{1}{N-1} \sum_j \left[ \frac{I_{\text{exp}}(s_j) - cI(s_j)}{\sigma(s_j)} \right]^2 \]

Additional information is ALWAYS required to resolve or reduce ambiguity of interpretation at given resolution.
Ab Initio Modeling

- Dummy Residue Models (GASBOR, GASBORMX)
- Single Phase Dummy Atom Models (DAMMIN, DAMMIF)
- Multi Phase Dummy Atom Models (MONSA)
- Obtaining Models
- Model Validity, Uniqueness and Stability
- Model Post Processing (DAMAVER, DAMCLUST)
- MONSA and contrast variation
Dummy Residue Models

- Proteins typically consist of folded polypeptide chains composed of amino acid residues.
- At a resolution of 0.5 nm each amino acid can be represented as one entity (dummy residue).
- In GASBOR a protein is represented by an ensemble of $K$ dummy residues that are
  - Identical
  - Have no ordinal number
  - For simplicity are centered at the C$_\alpha$ positions.
Dummy Residue Models

- GASBOR finds coordinates of $K$ dummy residues within its search volume (red)
- Scattering is computed using the Debye (1915) formula
- Requires polypeptide chain-compatible arrangement of dummy residues
Dummy Residue Models for Mixtures

- GASBORMX extension to equilibrium mixtures
- Reconstructs the monomer and a symmetric multimer together
- Interconnectivity is required for the monomer and the multimer
Single Phase Dummy Atom Models

Dummy atoms:
• Act as a placeholder for, but does not resemble, a real atom
• Occupy a known position in space
• Have a known scattering pattern
• May either contribute to solvent or particle
• Are also known as beads
Single Phase Dummy Atom Models

A volume is filled by densely packed beads of radius $r_0 << D_{\text{max}}$

Parametrization:
a binary vector,
0 if solvent, 1 if particle
Single Phase Dummy Atom Models

A volume is filled by densely packed beads of radius $r_0 \ll D_{\text{max}}$

Parametrization:
- a binary vector,
- 0 if solvent, 1 if particle
Single Phase Dummy Atom Models

At the current iteration:
• dark blue particle, might become solvent
• light blue solvent, might become particle
• white solvent, won’t change
Single Phase Dummy Atom Models

- Scattering intensity is computed using spherical harmonics
- Penalty terms ensure compactness and connectivity

compact

loose

disconnected
Multi Phase Dummy Atom Models

Single phase shape determination

- One can differentiate between distinct parts of the particle
- Several curves are required
- Assuming the same arrangement of the parts in different samples
Multi Phase Dummy Atom Modeling

- 1 phase = 1 component of a complex particle
- For each phase, \( R_g \), \( V \) and its scattering curve can be given
- For each curve, contrast of each phase are specified

contrast variation and / or use of partial constructs
## Dummy Atom Models

<table>
<thead>
<tr>
<th>Objects</th>
<th>DAMMIN</th>
<th>DAMMIF</th>
<th>MONSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max # of phases</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Angular range</td>
<td>lower part</td>
<td>lower part</td>
<td>lower part</td>
</tr>
<tr>
<td>Resolution</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Search volume</td>
<td>fixed</td>
<td>growing</td>
<td>fixed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Constrains</th>
<th>DAMMIN</th>
<th>DAMMIF</th>
<th>MONSA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Symmetry, Interconnectivity, Compactness</td>
<td>Symmetry, Interconnectivity, Compactness</td>
<td>Symmetry, Interconnectivity, Compactness</td>
</tr>
</tbody>
</table>

| Performance      | slow       | fast      | very slow |

| Limitations      | DAMMIN has better symmetry support |

**Warning:** results are not atomic models, just a filled volume!
Obtaining Models – primus/qt
Obtaining Models – Windows

dammif lyz.out --mode=slow --prefix FMRP1

dammif lyz.out --mode=slow --prefix FMRP2

dammif lyz.out --mode=slow --prefix FMRP3

dammif lyz.out --mode=slow --prefix FMRP4

dammif lyz.out --mode=slow --prefix FMRP5

dammif lyz.out --mode=slow --prefix FMRP6

dammif lyz.out --mode=slow --prefix FMRP7

dammif lyz.out --mode=slow --prefix FMRP8

dammif lyz.out --mode=slow --prefix FMRP9

dammif lyz.out --mode=slow --prefix FMRP10
for i in 'seq 1 10' ; do
dammif --prefix=lyz-$i --mode=slow lyz.out;
done
Obtaining Models – local cluster

Please contact your system administrator for details of your cluster and how to submit jobs.

Important: as processes are being run in parallel, multiple may be started at the same time – with the same random seed – resulting in exactly the same model.

Make sure to redefine the random seed for each run!
Obtaining Models – fine tuning

- Run dammif in slow mode once
- Find the $prefix.in file
- Modify as needed
- Run dammif as
  `dammif -prefix=. --mode=i < modified.in`
Obtaining Models – ATSAS Online

DAMMIN online

- Project ID
- GNOM file (*.out)
- Symmetry: P1 (no symmetry)
- Anisometry: Unknown
- Mode: Slow (smaller beads)

SUBMIT
Model Validity

• Validate your input data
• Check for
  • Aggregation
  • Noise at higher angles
• Keep in mind: it is easy to model noise

→ Garbage in, garbage out
Model Validity – Stability

This structure can not be restored without use of additional information
Model Validity – Stability

Disk 5:1

Spread region, most probable volume

Disk 10:1

Spread region, most probable volume

This structure can not be restored without use of additional information
Model Validity – Stability

Typical solution with P5 symmetry

Original body

Typical solution with no symmetry
Model Post Processing – SUPCOMB

• Superimpose models by minimizing the Normalized Spatial Discrepancy (NSD)

• Steps
  • Principle axes alignment
  • Gradient minimization
  • Local grid search
Model Post Processing – DAMAVER

- \( \text{NSD}_i = \langle \text{NSD}_{ij} \rangle_j \)
- \( \text{MIN} (\text{NSD}_i) \Rightarrow \text{typical (most probable) model} \)
- \( \langle \text{NSD} \rangle + 2 \sigma (\text{NSD}) \Rightarrow \text{threshold for outliers} \)

<table>
<thead>
<tr>
<th>NSD</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>...</th>
<th>i</th>
<th>...</th>
<th>j</th>
<th>...</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NSDij</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NSDji</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Model Post Processing – Example

Shape determination of 5S RNA: a variety of DAMMIN models yielding identical fits

Model Post Processing – Example

5S RNA – Solution spread region

5S RNA – Most Populated Volume

5S RNA – Final Solution within the Spread Region
Model Post Processing – Options

- Take the model with the least NSD to all others (fits the data)
- Take the averaged and filtered model (will not fit the data)
- Restart DAMMIN with the averaged model to obtain a model that fits the data

$\texttt{damaver \ –a \ *-1.pdb}$
Model Post Processing – Options

Notes:

• GASBOR models can generally not be post processed – dummy residues would be reduced to dummy atoms
• MONSA models may not be post processed with the distributed DAMAVER program, specialized tools may be available on request
### Method Applicability to SAXS/SANS data

<table>
<thead>
<tr>
<th>Program</th>
<th>SAXS</th>
<th>SANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GASBOR/GASBORMX</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td>DAMMIN/DAMMIF</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>MONSA</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

* May be used if contrast is high and the particle is homogeneous
** Dummy residue form factors are available for X-rays only

“Do not measure SANS where the answers can be given by SAXS” – D. Svergun
DAMMIF fits
DAMAVER fit
DAMFILT fit
DAMMIN fit