Introduction to SAS Model Determination
Ab Initio Modelling
Shape and Size

Log I(s)

lysozyme

apoferitin
Fingerprinting of SAS Curves

Solid sphere

Hollow sphere

Flat disc

Dumbbell

Long rod
Principle of SAS Modeling

1D scattering data (or multiple data sets)

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

3D search model
\[ x = \{x\} = \{X_1 \ldots X_M\} \]
M parameters

Trial-and-error

Non-linear search

Additional information is ALWAYS required to resolve or reduce ambiguity of interpretation at given resolution.
Simulated Annealing

http://www.youtube.com/watch?v=iaq_Fpr4KZc
Outline

• Ab Initio Modeling
  • Classification (DATCLASS)
  • Single Phase Dummy Atom Models (DAMMIN, DAMMIF)
  • Obtaining Models
  • Model Validity, Uniqueness and Stability (SUPCOMB, SUPALM)
  • Model Post Processing (DAMAVER, DAMCLUST)

• DAMMIF Walkthrough

• Maxim
  • Multi Phase Dummy Atom Models (MONSA)
  • Dummy Residue Models (GASBOR, GASBORMX)
  • Ambivalence Score (AMBIMETER)
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 $r_0 \ll D_{\text{max}}$

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 $r_0 \ll D_{\text{max}}$

Solvent

Particle

Parametrization:

a binary vector,

0 if solvent, 1 if particle

$D_{\text{max}}$

$2r_0$
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

DAMMIN  DAMMIF
Single Phase Dummy Atom Models

- Scattering intensity is computed using spherical harmonics

- Penalty terms ensure compactness and connectivity
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 – ATSAS Online

DAMMIN online

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

Submit
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`
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 cannot be restored without use of additional information.
Model Validity – Stability

Disk 5:1

Disk 10:1

This structure can not be restored without use of additional information

Spread region, most probable volume

Spread region, most probable volume
Model Validity – Stability

Typical solution with P5 symmetry

Typical solution with no symmetry

Original body
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

- $\text{NSD}_i = \langle \text{NSD}_{ij} \rangle_j$
- $\text{MIN}( \text{NSD}_i ) \Rightarrow \text{typical (most probable) model}$
- $\langle \text{NSD} \rangle + 2 \sigma (\text{NSD}) \Rightarrow \text{threshold for outliers}$
Model Post Processing – SUPALM (NEW!)

- Superimpose models by minimizing the Normalized integrated Cross-Term (NCT)
- Also accepts EM maps

\[ NCT(A, B) = \frac{\int_0^{Sm} A_{lm}(s) B_{lm}^*(s) s^2 ds}{2 \left( \int_0^{Sm} I_A(s) s^2 ds \right)^{1/2} \left( \int_0^{Sm} I_B(s) s^2 ds \right)^{1/2}} \]
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 – Final Solution within the Spread Region

5S RNA – Most Populated Volume
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

$\textit{damaver} -a \ast-1.pdb$
Model Post Processing – DAMCLUST

Clustering of multiple models
- Discrepancies (distances) between multiple models as criteria for grouping
- Normalized spatial deviation serves as a distance between heterogeneous models (e.g. bead models)
- R.m.s.d. is employed for those with atom-to-atom correspondence (e.g. rigid body models)

Creates the complete graph by iteratively joining the clusters.

Selects the optimal threshold as a compromise between the number of clusters and averaged spread within the cluster.
Ab Initio Walkthrough

$ dammif

$ damaver -a

$ dammin
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

$\text{damaver \ --a \ *-1.pdb}$
DAMMIF fits
DAMAVER fit
DAMFILT fit
DAMMIN fit