

DAMMIF Update

Get the latest version of DAMMIF together with the
ATSAS-2.4 release package!

Ab-Initio Modelling

DAMMIN and DAMMIF

Daniel Franke

European Molecular Biology Laboratory

2010/10/27

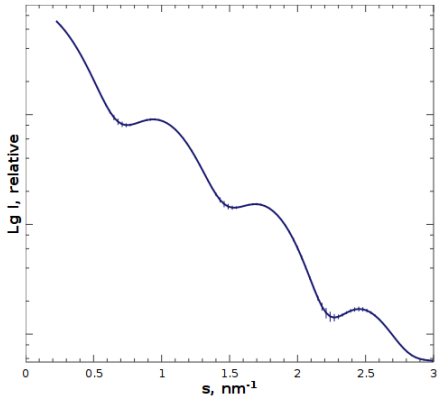
The following slides describe the **how**,
not the **why**!

Outline

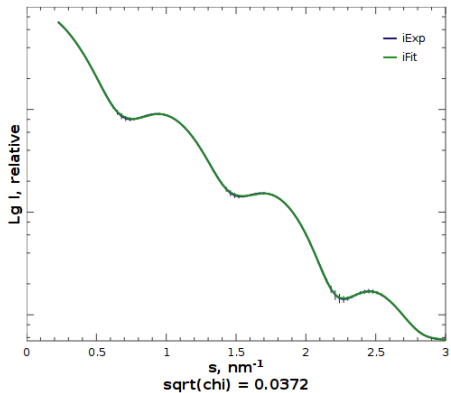
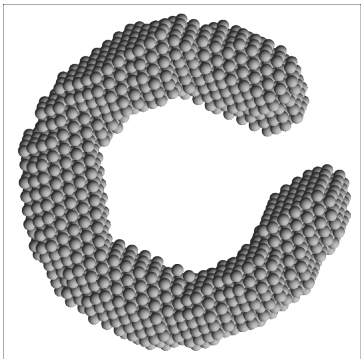
- 1 Introduction
- 2 Ab-Initio Modelling
- 3 Exercises
- 4 Postprocessing Models

Basic Idea

Find a three dimensional object whose theoretical scattering curve would resemble the experimental data best.

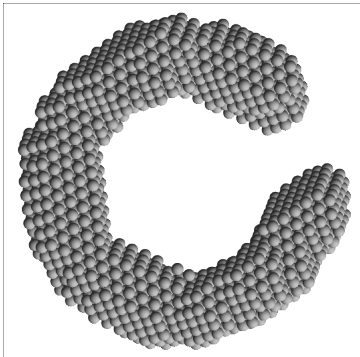


Results



The Dummy Atom Model

Many little scatterers ...



A **Dummy Atom Model (DAM)** is build by a tightly packed group of dummy atoms. The volume occupied by dummy atoms in any state (particle, solvent) is also known as **search volume**.

The Dummy Atom

One little scatterer ...

- Acts as a placeholder for, but does not resemble, a real atom
- Occupies a known position in space
- Has a known scattering pattern
- May either contribute to the solvent or the particle

Dummy atoms are also referred to as **beads**.

Basic Idea

Revisited.

Find a three dimensional object whose theoretical scattering curve would resemble the experimental data best.

Find the set of dummy atoms within a search volume whose accumulated scattering resembles the experimental data best.

Basic Idea

Revisited.

Find a three dimensional object whose theoretical scattering curve would resemble the experimental data best.

Find the set of dummy atoms within a search volume whose accumulated scattering resembles the experimental data best.

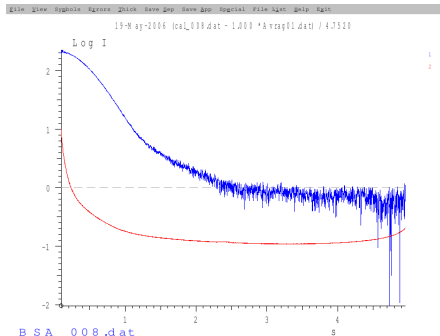
Validity of Input

Garbage In – Garbage Out

Validate input data; check for

- aggregation at the beginning
- noise at higher angles

Remember: noise can be modelled nicely



Outline

- 1 Introduction
- 2 Ab-Initio Modelling**
- 3 Exercises
- 4 Postprocessing Models

An estimate on the problem's size.

The Universe is not enough

A search volume of 2000 dummy atoms has

$$2^{2000} \approx 10^{600}$$

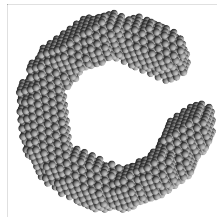
possible **conformations**, i.e. scattering curves.

On 40.000.000 conformations per hour per CPU, 1000 CPUs, 24 hours a day, 365 days a year one would spend the next couple of universes' time on enumerating all scattering curves!

Imposing restrictions in solution space.

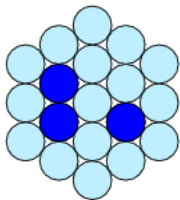
A **valid conformation** is ...

- **connected:** particle beads must be interconnected
- **tightly packed:** particle beads shall be tightly packed, avoid loose strands
- **centered:** assemble the particle within the search volume, avoid boundary contact
- **in right shape:** oblate or prolate shapes can be enforced

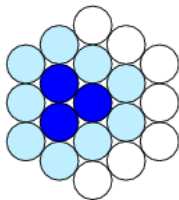


Advances And Differences In Programs

Selection Scheme



DAMMIN



DAMMIF

At the current iteration:

dark blue	particle, might become solvent
light blue	solvent, might become particle
white	solvent, won't change

DAMMIF Walkthrough

```
$> dammif shape.out
```

DAMMIF Output

Reading the output of DAMMIF

```

Step:      1, T: 0.130E-03, 42/1941,
          Succ: 1229, Eval: 20001, CPU: 00:00:03
Rf: 0.0875, Los: 0.17, Dis: 0.00, Rg: 0.15,
          Cen:22.57, Ani: 0.00, Fit: 0.0989

```

Step	Step number
T	Temperature, artificial
p/a	Number of particle beads of all beads
Succ	Number of successfull iterations at current T
Eval	Accumulated number of iterations
CPU	Accumulated runtime

DAMMIF Output

Reading the output of DAMMIF (cont.)

```
Step:      1, T: 0.130E-03, 42/1941,  
          Succ: 1229, Eval: 20001, CPU: 00:00:03  
Rf: 0.0875, Los: 0.17, Dis: 0.00, Rg: 0.15,  
          Cen:22.57, Ani: 0.00, Fit: 0.0989
```

Rf Goodness of Fit, data only
Los Contribution of Looseness Penalty
Dis Contribution of Disconnectivity Penalty
Per Contribution of Peripheral Penalty
Ani Contribution of Anisometry Penalty
Fit Goodness of Fit, data and penalties

Outline

- 1 Introduction
- 2 Ab-Initio Modelling
- 3 Exercises**
- 4 Postprocessing Models

Exercises

Run DAMMIF on shape.out in ...

- fast mode (bigger beads, less iterations)
- slow mode (smaller beads, more iterations)
- fast mode settings, without penalties
- fast mode settings, one penalty set to 1.0 in turn
- ...

Run multiple times, compare ...

Outline

- 1 Introduction
- 2 Ab-Initio Modelling
- 3 Exercises
- 4 Postprocessing Models**

Postprocessing Models

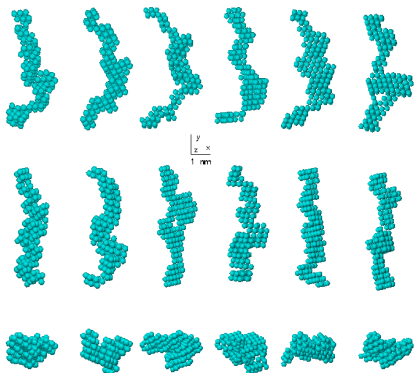
How to proceed ...

With multiple models:

- find those that are most similar
(uniqueness of reconstruction is not guaranteed)
- superimpose and average them
- restart fitting process using the averaged model

Multiple models

Funari et al. (2000) J. Biol. Chem. 275, 31283–31288.



5S RNA, multiple solutions with equally good fit.

Selecting Models

DAMSEL

Computes the similarities between all pairs of input files.

Measure of similarity of models:

Normalized Spatial Discrepancy (*NSD*)

$NSD < 1$ implies similar models

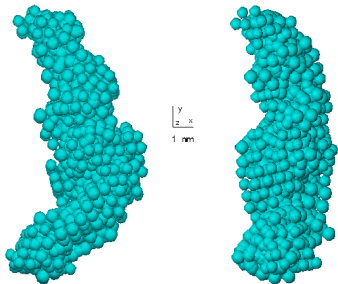
Superimposing Models.

SUPCOMB, DAMSUP

- SUPCOMB: superimpose any two models (principle axis alignment, gradient minimization, local grid search)
- DAMSUP: superimpose multiple models on a reference using SUPCOMB.

Superimposing models

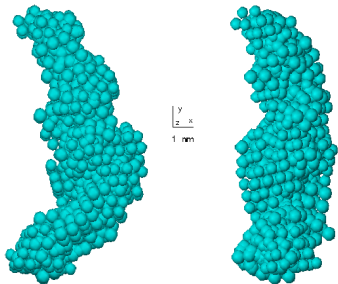
5S RNA continued ...



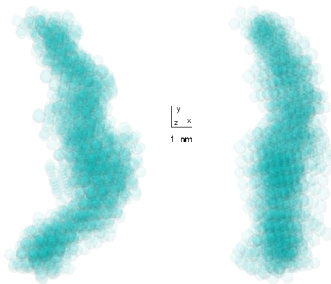
Solution spread region.

Superimposing models

5S RNA continued ...



Solution spread region.



Most populated volume.

Averaging Models

DAMAVER, DAMFILT

- DAMAVER: Creates a bead probability density map within the search volume.
- DAMFILT: Generates the averaged model, using a user-defined probability threshold. Will give a valid model, violating the threshold if necessary.

Ab-Initio Modelling

Options at this point.

- take the averaged model – but this will not fit the data
- take the model that has the least NSD to all others – this fits the data
- use averaged model and restart DAMMIN to fit the experimental data (DAMSTART)

Ab-Initio Modelling

Options at this point.

- take the averaged model – but this will not fit the data
- take the model that has the least NSD to all others – this fits the data
- use averaged model and restart DAMMIN to fit the experimental data (DAMSTART)

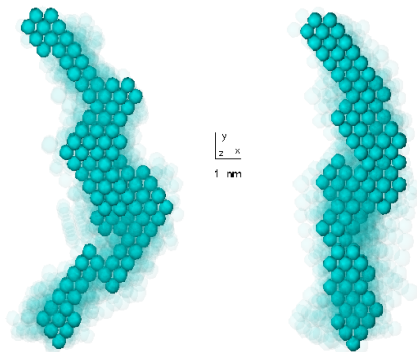
Ab-Initio Modelling

Options at this point.

- take the averaged model – but this will not fit the data
- take the model that has the least NSD to all others – this fits the data
- use averaged model and restart DAMMIN to fit the experimental data (DAMSTART)

Ab-Initio Modelling

5S RNA continued ...



Finalized model, refined by DAMSTART.

That's all folks.

Questions? Visit

<http://www.saxier.org/forum>