Ab initio approaches: resolution and uniqueness

D.I. Svergun, V.V. Volkov, M.B. Kozin, M.V. Petoukhov
Plan of the talk

- Information content
- *Ab initio* shape determination
- Analysis of domain structure
- Use of contrast variation
Information content in solution scattering

\[ I(s) = \sum_{k=1}^{\infty} s_k I(s_k) \left( \frac{\sin D(s - s_k)}{D(s - s_k)} - \frac{\sin D(s + s_k)}{D(s + s_k)} \right) \]

A solution scattering curve from a particle with maximum size \( D \) can be represented by its values taken at discrete points (Shannon channels)

\[ s_k = k\pi/D \]

Information content in solution scattering is usually low. Typically $N_s = 5-15$

Particle structure must be represented by a model with a small number of independent parameters.
Crystal *versus* Solution

- Thousands of reflections
- 3D, high resolution

- A few Shannon channels
- 1D, low resolution

\[ \Delta s = \frac{2\pi}{D} \]

\[ \Delta s \ll \frac{\pi}{D} \]
Crystal \textit{versus} Solution
What may solution scattering yield

Resolution, nm

Shape

Fold

Atomic structure
**Ab initio** shape determination

- **3D search model**
  - M parameters

- **Trial-and-error**

- **Non-linear search**

- **1D scattering data**

  - Poor fit ➔ Model incorrect? **Yes**
  - Good fit ➔ Model correct? **??**
Two possible strategies

♦ Use models described by a few parameters only
  ♦ In the past: simple geometrical bodies (triaxial ellipsoids, cylinders etc)

♦ Use models described by many parameters, but try to constrain the solution
  ♦ In the past: multibody models (e.g. consisting of many densely packed spheres) constrained using information from other methods
Low resolution envelope

\[ F(\omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\omega) \]

- \( M=(L+1)^2 - 6 \) parameters, \( M \approx 10^1 \)
- Unique shape restoration if \( M \approx 1.5N_s \) (typically, \( L=4 \))
- Limited ability to describe complex shapes


Ab initio program SASHA
Bead models

\[ \text{Position}(j) = X(j) = 1 \text{ or } 0 \]

- \( M \approx (D_{\text{max}}/r_0)^3 \approx 10^3 >> N_s \) parameters, too many for conventional minimization
- No unique shape restoration unless constrained
- Able to describe complex shapes


Local and global search

- Local search always goes to a better point and can thus be trapped in a local minimum.
- To avoid local minima, global search must be able go to a worse point.
Simulated annealing

Aim: find a configuration of $M$ variables $\{x\}$ minimizing a function $f(x)$ of these variables.

1. Start from a random configuration $x$ at a “high” temperature $T$.
2. Make a small step (random modification of the configuration) $x \rightarrow x'$ and compute the difference $\Delta = f(x') - f(x)$.
3. If $\Delta < 0$, accept the step; if $\Delta > 0$, accept it with a probability $e^{-\Delta/T}$.
4. Make another step from the old (if the previous step has been rejected) or from the new (if the step has been accepted) configuration.
5. Anneal the system at this temperature, i.e. repeat steps 2-4 “many” (say, $100M$ tries or $10M$ successful tries, whichever comes first) times, then decrease the temperature ($T' = cT$, $c<1$).
6. Continue cooling the system until no improvement in $f(x)$ is observed.
Ab initio program DAMMIN

Using simulated annealing, finds a compact dummy atoms configuration $X$ that fits the scattering data by minimizing

$$f(X) = \chi^2 [I_{\text{exp}}(s), I(s,X)] + \alpha P(X)$$

where $\chi$ is the discrepancy between the experimental and calculated curves, $P(X)$ is the penalty to ensure compactness and connectivity, $\alpha > 0$ its weight.
A screenshot of DAMMIN
Test: shape of the myosin head S1
Test: globular particle (cube)
Test: elongated particle (cylinder 1:5)
Test: flat particle (disk 10:1)
The use of additional information

• Both SASHA and DAMMIN permit to include the following *a priori* information:
  - Particle symmetry (point groups P1 to P6, P22 to P62 supported)
  - Particle anisometry (prolate/oblate/unknown)
Shape determination of symmetric proteins

zym

pox

ttc

1pvd-t
Clinton and Gore

or

Additional information must be used with caution
Shape determination of $V_1$ ATPase

$N_s = 14.3$
Other *ab initio* shape determination methods

- ‘Interconnected ellipsoids’ procedure of Trewhella et al. (2001)

- Both these methods always go to a better point
Limitations of shape determination

- Very low resolution
- Ambiguity of the models
- Accounts for a restricted portion of the data

How to construct *ab initio* models accounting for higher resolution data?
Dummy residues model

- Proteins typically consist of folded polypeptide chains composed of amino acid residues

- At a resolution of 0.5 nm a protein can be represented by an ensemble of $K$ dummy residues centered at the Cα positions with coordinates $\{r_j\}$

- Scattering from such a model is computed using the Debye (1915) formula

$$I_{DR}(s) = \sum_{i=1}^{K} \sum_{j=1}^{K} g_i(s) g_j(s) \frac{\sin sr_{ij}}{sr_{ij}}$$
Excluded volume effects and local interactions lead to a characteristic distribution of nearest neighbors around a given residue in a polypeptide chain.

Distribution of neighbors

Number of neighbors

Shell radius, nm

0.2 0.4 0.6 0.8 1.0
0 1 2 3 4 5 6
Ab initio program GASBORB

Using simulated annealing, finds a spatial distribution of $K$ dummy residues within a sphere with diameter $D_{\text{max}}$ to fit the scattering data by minimizing

$$f (\{r_i\}) = \chi^2 \left[ I_{\text{exp}} (s), I_{DR} (s, \{r_i\}) \right] + \alpha P (\{r_i\})$$

where $\chi$ is the discrepancy between the experimental and calculated curves, $P (\{r_i\})$ is the penalty to ensure a chain-like distribution of neighbors, $\alpha > 0$ its weight.

**Ab initio models of lysozyme**

- Experimental
- DR model
- Bead model

**Graph:**

- **x-axis:** \(s, \text{nm}^{-1}\)
- **y-axis:** \(\lg I, \text{relative}\)

**Diagram:**

- Typical bead model
- Most divergent DR models
Ab initio modelling using dummy residues

- Yields much more stable and detailed models than those provided by shape determination
- Reveals domain structure of proteins
- Can be used in protein folding prediction
- Has potential for future development (neutron scattering, phase problem in low resolution crystallography)
Contrast variation yields additional information about shape and internal structure:

- by changing the solvent density and/or
- by selective labeling of specific structure fragments

Use of contrast variation: study of ribosome

- Molecular mass 2.3 Mda, diameter about 27 nm
- Two unequal subunits, small (30S) and large (50S)
- 30S: 21 individual proteins (TP30) + 16S RNA (RNA30)
- 50S: 34 individual proteins (TP50) + 5S RNA + 23S RNA (RNA50)
X-rays versus Neutrons

Addition of sucrose or salts

RNA, 550 e/nm³

60% sucrose, 430 e/nm³

Protein, 410 e/nm³

H₂O, 344 e/nm³

Isotopic H/D substitution

D-Protien, 130% D₂O

D₂O, 6.38×10¹⁰ cm⁻²

H-RNA, 70% D₂O

H-Protein, 40% D₂O

H₂O, -0.59×10¹⁰ cm⁻²
Contrast variation on hybrid ribosomes

0% D$_2$O

Protonated 70S ribosome, HH30+HH50

40% D$_2$O

Hybrid 70S with 23S RNA deuterated, HH30+HD50

70% D$_2$O
Scattering data from hybrid ribosomes

Contrast variation by solvent exchange

- HH30+HH50 DD30+HH50 DH30+HH50 in 0, 35, 50, 75, 100% D₂O  
  15 curves
- HH30+DD50 in 0, 35, 50, 75% D₂O  
  4 curves
- DH30+DD50 and HH30+DH50 in 0, 40, 60, 100% D₂O  
  4 curves
- HH30 and HH50 in 0, 100% D₂O  
  4 curves
- DD30 and DD50 in 0% D₂O  
  2 curves

Spin-dependent contrast variation data

- HH30+DD50, DD30+HH50, DH30+DH50  
  Polarization = 0 and 1  
  2 curves
- X-ray scattering curves from 70S, 30S and 50S  
  3 curves

TOTAL 42 curves
Search volume for the 70S ribosome

- Yellow pixels: cryo-EM model of Frank et al. (1995)
- Red and blue circles: dummy atoms belonging to the 30S and 50S subunits, respectively

Number of atoms $M = 7860$

Packing radius $r_0 = 0.5 \text{ nm}$
Ribosomal data fitted

Neutrons

X-rays


*J.Biol. Chem.* **275**, 14432-14439
A protein-RNA map in the 70S ribosome *E.coli*
Solution versus crystal

3 nm resolution model of the 30S subunit in the 70S ribosome *E.coli*
(Svergun & Nierhaus, May 2000)

0.33 nm resolution model of the 30S subunit *Th. Thermophilus*
(Yonath group, September 2000)
Solution versus crystal

3 nm resolution neutron scattering model of the 50S subunit in the 70S ribosome *E.coli* (Svergun & Nierhaus, May 2000)

0.24 nm resolution crystallographic model of the 50S subunit *H.marismortui* (Steitz group, August 2000)
Solution versus crystal

3 nm resolution neutron scattering model of the 50S subunit in the 70S ribosome *E.coli* (Svergun & Nierhaus, May 2000)

0.24 nm resolution crystallographic model of the 50S subunit *H.marismortui* (Steitz group, August 2000)
Conclusions
Conclusions