

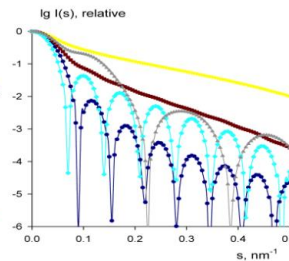
Ab Initio methods: What are they and how do they work?

Clement Blanchet

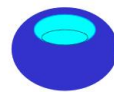
SAXS team – EMBL Hamburg

SAXS: Size, shape

Solid sphere



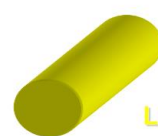
Hollow sphere



Dumbbell



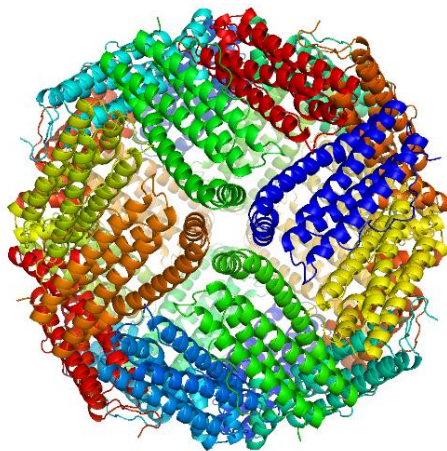
Flat disc



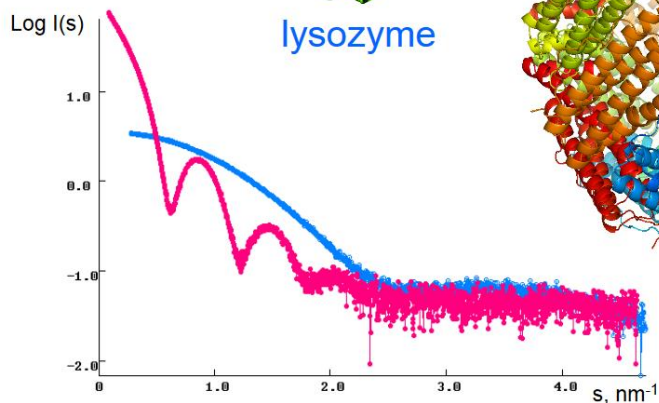
Long rod



lysozyme

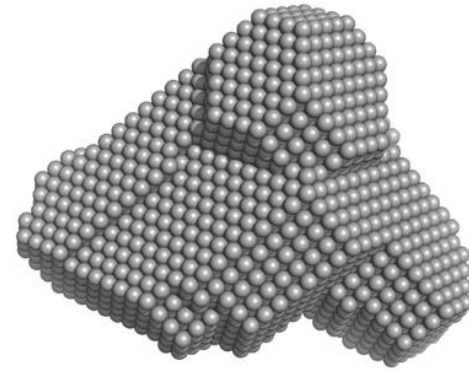
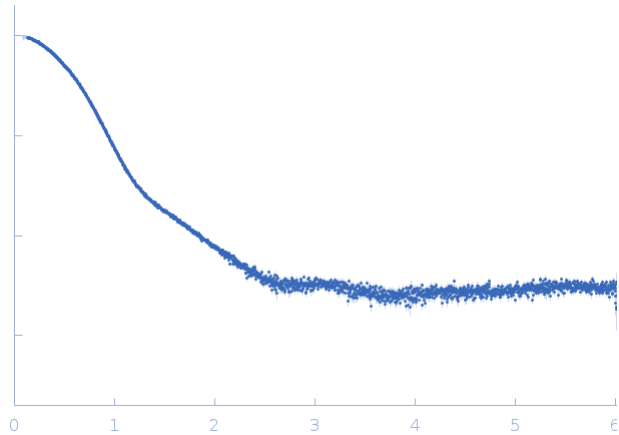


apoferritin



Ab initio modeling

From a 1d curve to a 3d model from the SAXS data only



How is it done?

Outline

- Form factor from geometrical shape
- Envelope function using spherical harmonics
- **Bead/dummy residue modeling:**
 - Principles
 - Different flavors of ab initio modeling
 - Words of caution
- Reconstructing electron density, DENSS

Computing form factor from simple geometrical shape

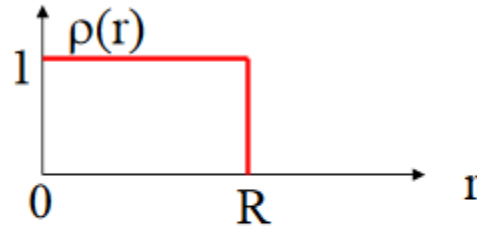
$$I(\mathbf{s}) = A(\mathbf{s}) \cdot A^*(\mathbf{s})$$

$$A(\mathbf{s}) = \int_{V_r} \rho(\mathbf{r}) \cdot e^{-i\mathbf{r} \cdot \mathbf{s}} dV_r$$

For simple geometrical shapes, the form factor can be computed from the electron density.

Example: form factor of a solid sphere

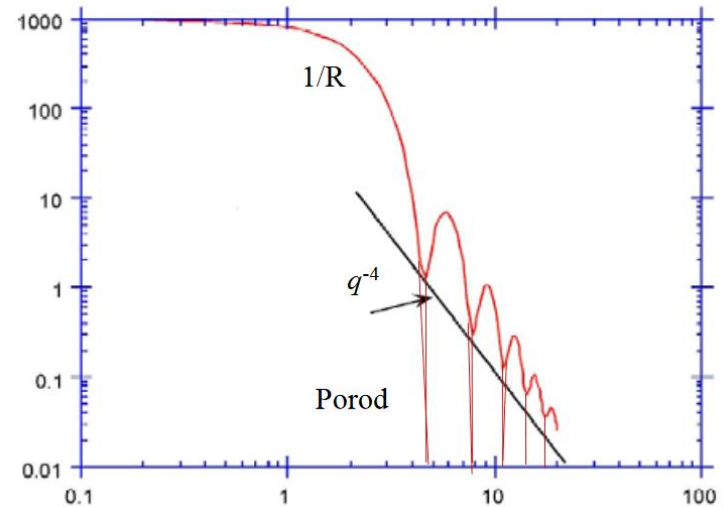
Form factor of a solid sphere



$$A(q) = 4\pi \int_0^{\infty} \rho(r) \frac{\sin(qr)}{qr} r^2 dr = 4\pi \int_0^R \frac{\sin(qr)}{qr} r^2 dr$$
$$= \frac{4\pi}{q} \int_0^R \sin(qr) r dr =$$

Form factor of sphere

$$P(q) = A(q)^2/V^2$$



The sphere case is trivial,
It quickly become complicated

Table 3.4. Equations for Scattering Intensities of Simple Bodies

Uniform sphere of radius R	$9 \left(\frac{\sin t - t \cos t}{t^3} \right)^2 = \phi^2(t), \quad t = sR$
Spherical layer with radii $R_1 > R_2$	$(R_1^3 - R_2^3)^{-2} [R_1^3 \phi(sR_1) - R_2^3 \phi(sR_2)]^2$
Triaxial ellipsoid (semiaxes a, b, c)	$\int_0^1 \int_0^1 \phi^2 \{ s [a^2 \cos^2(\frac{1}{2}\pi x) + b^2 \sin^2(\frac{1}{2}\pi x)(1 - y^2) + c^2 y^2]^{1/2} \} dx dy$
Ellipsoid of rotation $a: a: va$	$\int_0^1 \phi^2 [sa(1 + x^2(v^2 - 1))^{1/2}] dx$
Parallelepiped (edges A, B, C)	$\int_0^1 \Psi_p[s, B(1 - x^2)^{1/2}, A] S^2(sBCx/2) dx; \quad S(t) = \sin(t)/t$ $\Psi_p(s, B, A) = \frac{2}{\pi} \int_0^{\pi/2} S^2[sA \sin(y/2)] S^2[sB \cos(y/2)] dy$
Right elliptical cylinder with height H , semiaxes of ellipse a, va	$\int_0^1 \Psi_{ec}[s, a(1 - x^2)^{1/2}] S^2(sHx/2) dx$ $\Psi_{ec}(s, a) = \frac{1}{\pi} \int_0^\pi A_1^2 \left[sa \left(\frac{1 + v^2}{2} + \frac{1 - v^2}{2} \cos y \right)^{1/2} \right] dy$ $A_1(t) = 2J_1(t)/t$
Right hollow cylinder with height H , outer radius R_1 , inner radius R_2	$\int_0^1 \Psi_{hc}[s, R_1(1 - x^2)^{1/2}, R_2(1 - x^2)^{1/2}] S^2(sHx/2) dx$ $\Psi_{hc}(s, R_1, R_2) = \frac{1}{1 - \gamma^2} [A_1(sR_1) - \gamma^2 A_1(sR_2)]$ $\gamma = R_2/R_1$
Right circular cylinder of radius R , height H	$4 \int_0^1 \frac{J_1^2[sR(1 - x^2)^{1/2}]}{[sR(1 - x^2)^{1/2}]^2} S^2(sHx/2) dx$
(a) $R = 0$ (infinitely thin rod, height H)	$2 \text{Si}(sH)/sH - S^2(sH/2), \quad \text{Si}(t) = \int_0^1 S(x) dx$
(b) $H = 0$ (infinitely thin disk, radius R)	$[2 - A_1(2sR)]/s^2 R^2$

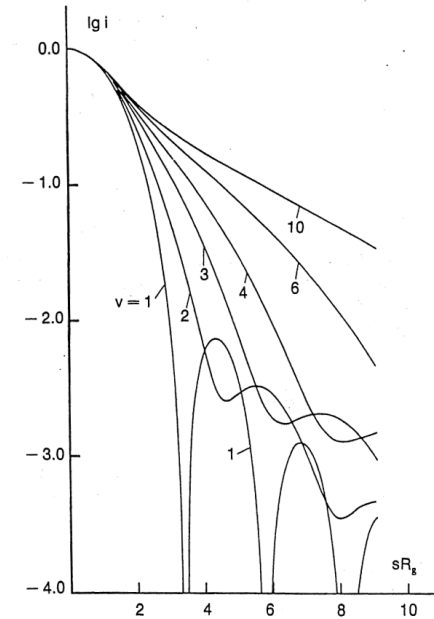
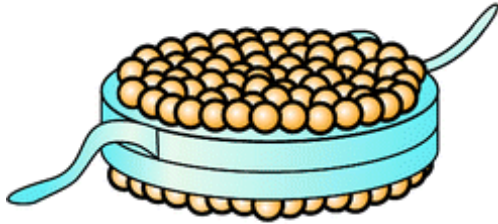
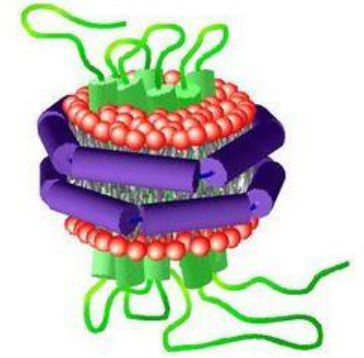


Figure 3.12. Scattering curves for prolate ellipsoids of rotation with $v = c/a$ (after Kratky and Pilz, 1972).

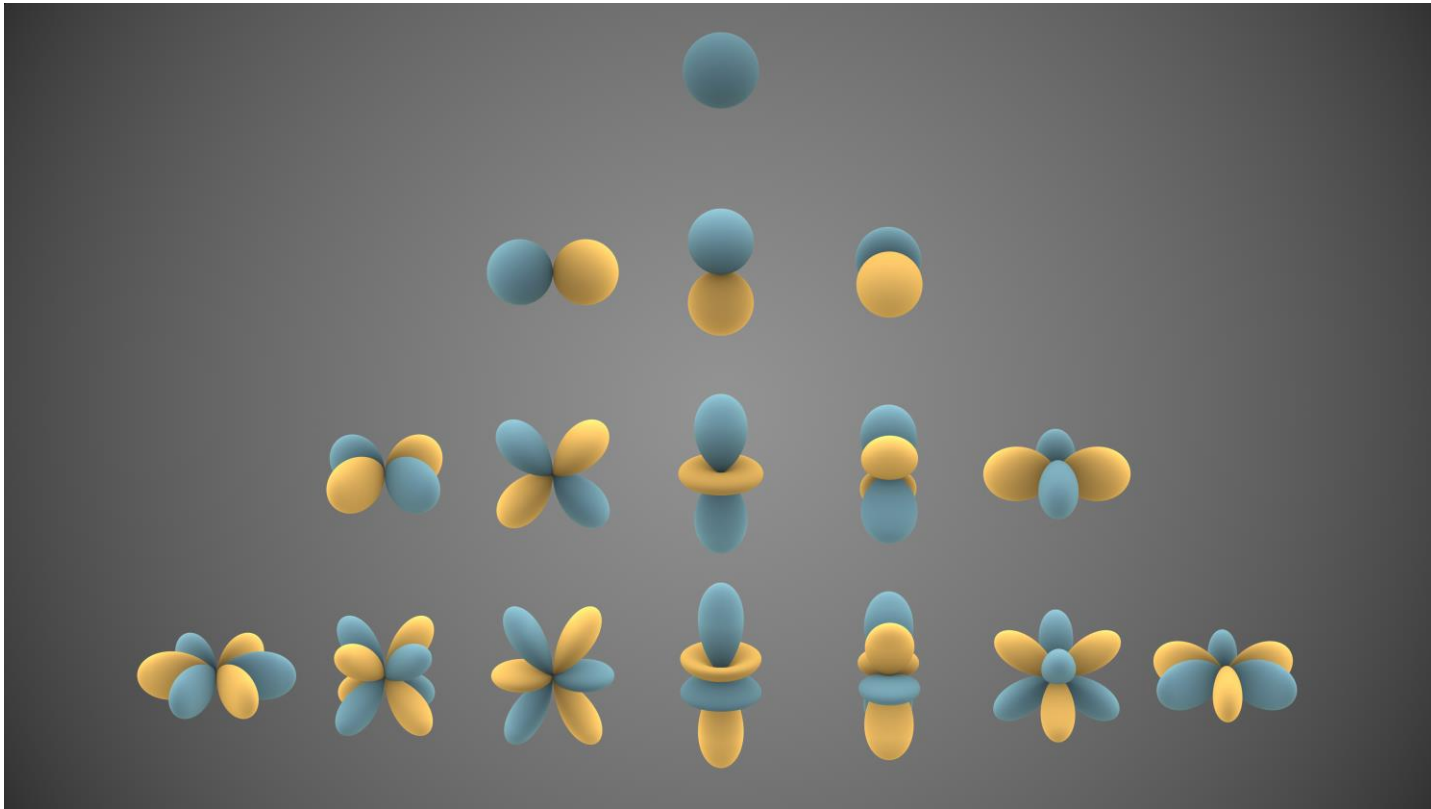
$I(s) = I(s, \mathbf{X})$ [e.g., for an ellipsoid $\mathbf{X} = (a, b, c)$], one can apply algorithm described in Section 3.2. From the approximate classes of bodies we can choose that providing the best fit with experiment (namely, with the scattering curve and the invariants). It should be noted, however, that the scatterer sufficiently large s , even in a region of homogeneity, cannot be represented by the scattering curve from a simple body;

Example: modeling nano-disc

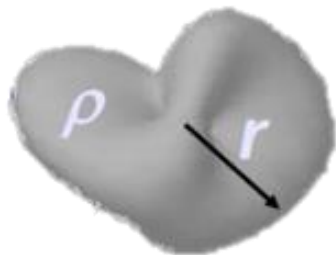


Skar-Gislinge *et al.* *J. Am. Chem. Soc.*

Envelope reconstruction using spherical harmonics



Envelope representation



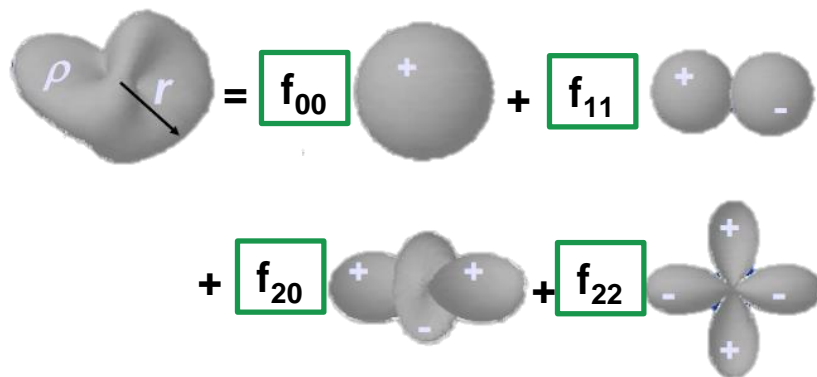
Scattering Density in Spherical Coordinates:

$$\rho(r) = \begin{cases} 1, & \text{if } 0 \leq r \leq F(\omega) \\ 0, & \text{if } r > F(\omega) \end{cases}$$

where r is the radial distance, and $F(\omega)$ is an envelope function describing the shape of the particle.

Shape Parameterization by Spherical Harmonics:

$$F(\omega) \approx F_L(\omega) = \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega)$$



where:

- $Y_{lm}(\omega)$ are the orthogonal spherical harmonics,
- f_{lm} are the parameterization coefficients,
- L is the maximum order of harmonics used for the shape description.

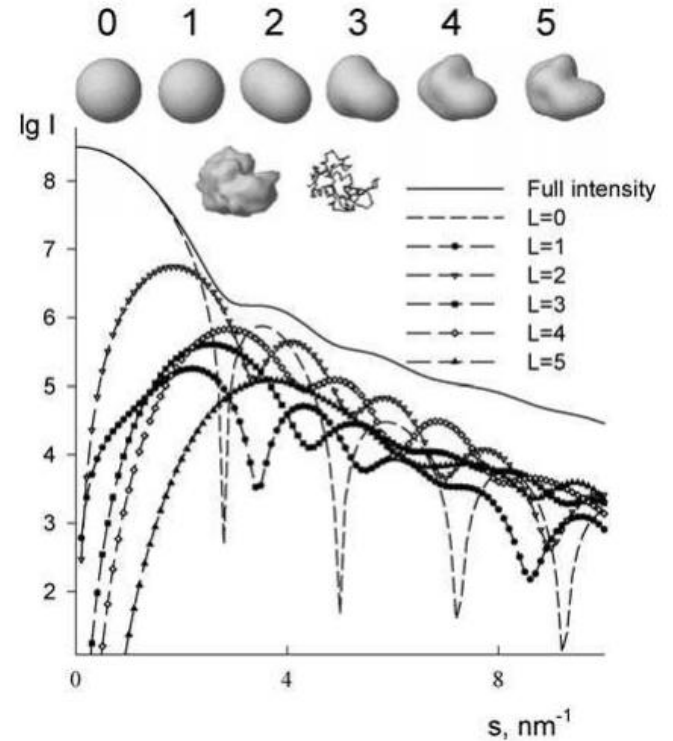
Scattering curves computed using spherical harmonics

Theoretical Scattering Intensity:

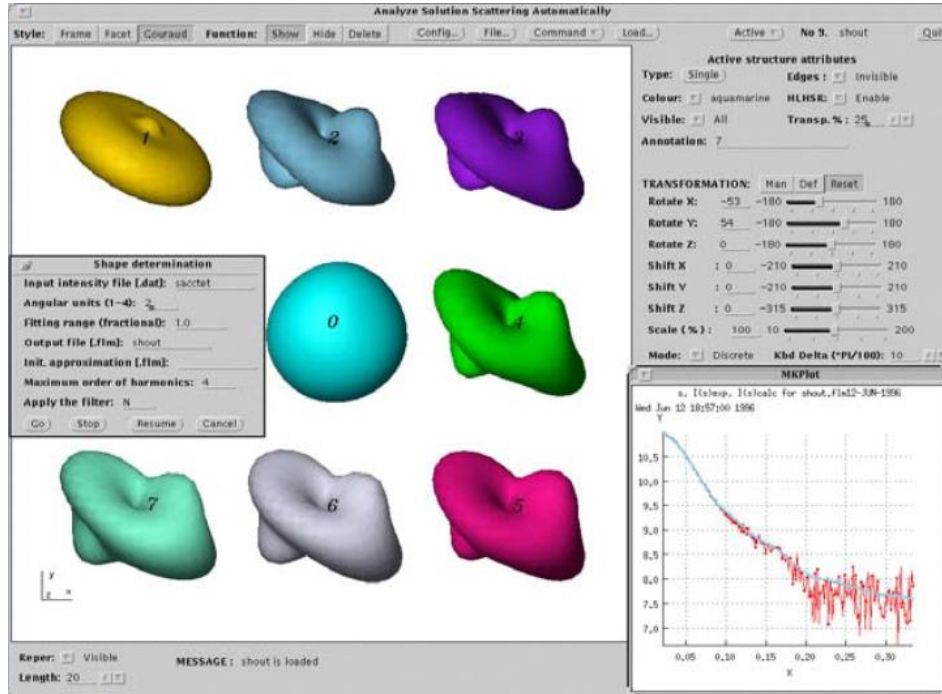
$$I_{\text{theor}}(s) = \sum_{l=0}^L \left(\sum_{m=-l}^l 2\pi^2 |A_{lm}(s)|^2 \right)$$

where:

- $I_{\text{theor}}(s)$ is the theoretical scattering intensity,
- $A_{lm}(s)$ represents the amplitude contributions from partial harmonics
- s is the magnitude of the scattering vector.



Sasha: building envelope from SAXS data using spherical harmonics



But..

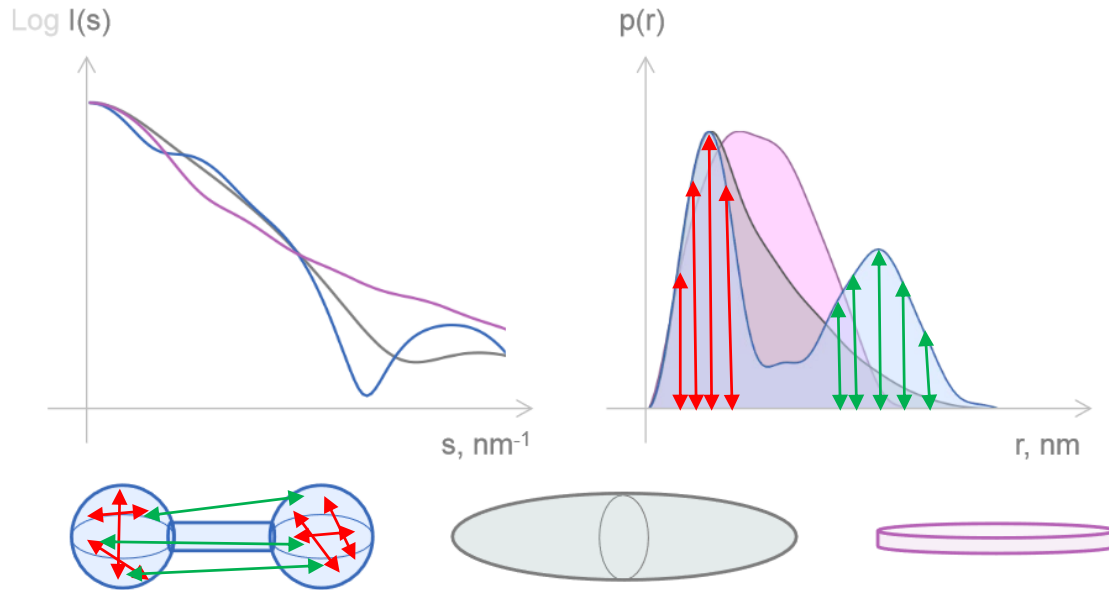
- limited in resolution,
- cannot reconstruct hollow shape.

→ Superseded by bead modeling methods

Bead modeling, dummy residue modeling

Basic idea

SAXS provides information about distances

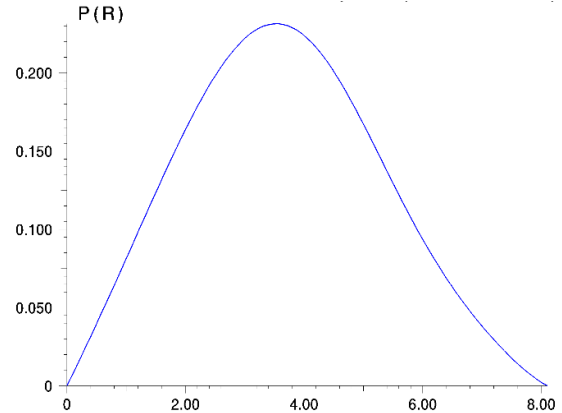
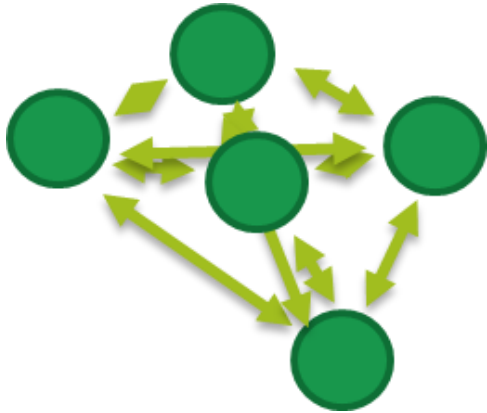


$$p(r) = \rho^2 \gamma_0(r) V r^2$$

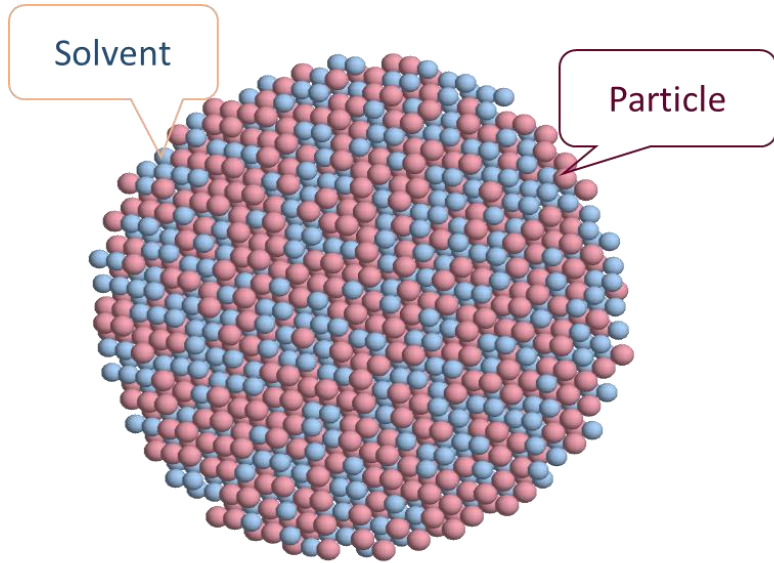
Where $\gamma_0(r)$ is the probability of finding a point within the particle at a distance r from a given points.

Basic idea

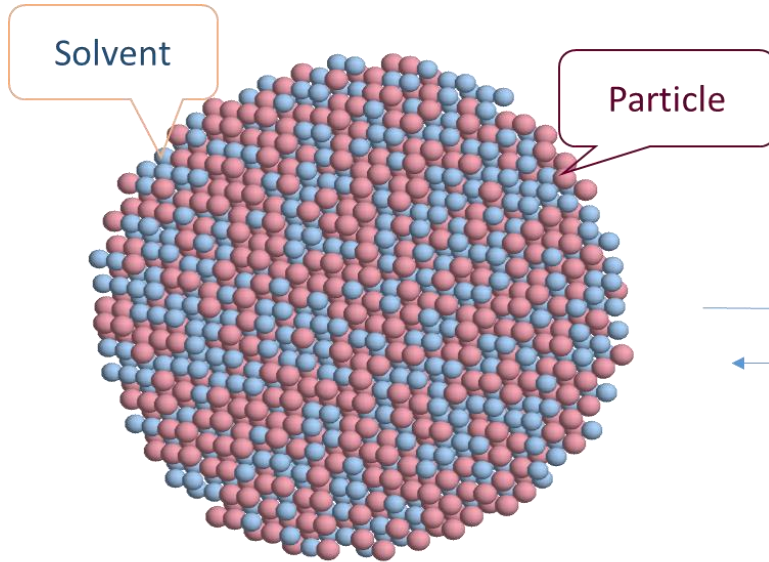
- Find an ensemble of beads with the inter-bead distances that are consistent with the $p(r)$



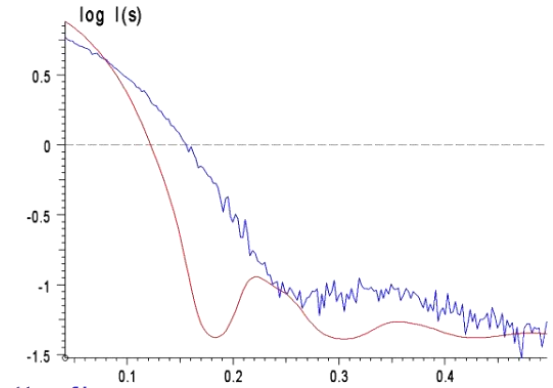
Beads on a grid



Beads on a grid



Trial and error

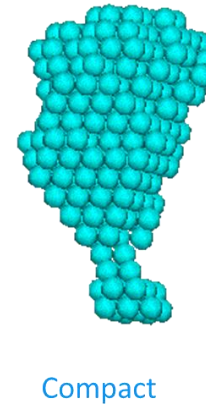
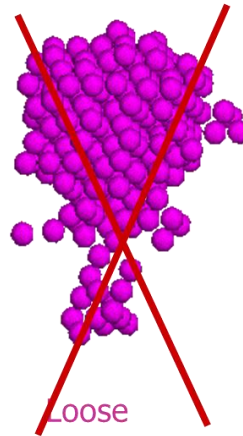
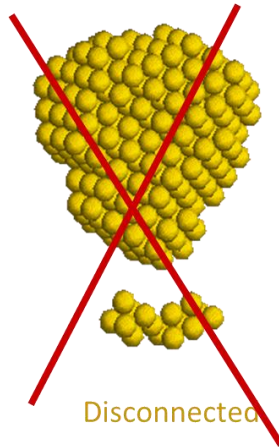


SAXS curve computed from the bead ensemble and compared with the experimental data.



Penalty terms

- Bead configuration should not only fit the data but also be compact and connected:
- The “looseness” and “disconnectivity” are quantified from the bead configuration
- Added to the target function to be minimized



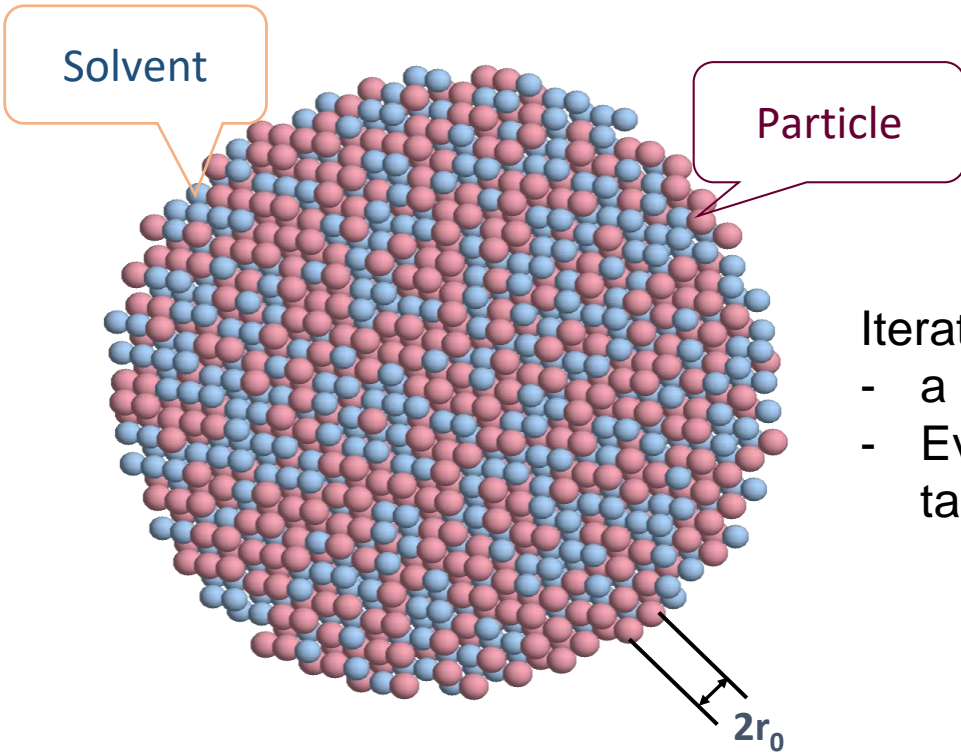
Minimization of the target function

Parameterization:
a binary vector,
0 if solvent, 1 if particle

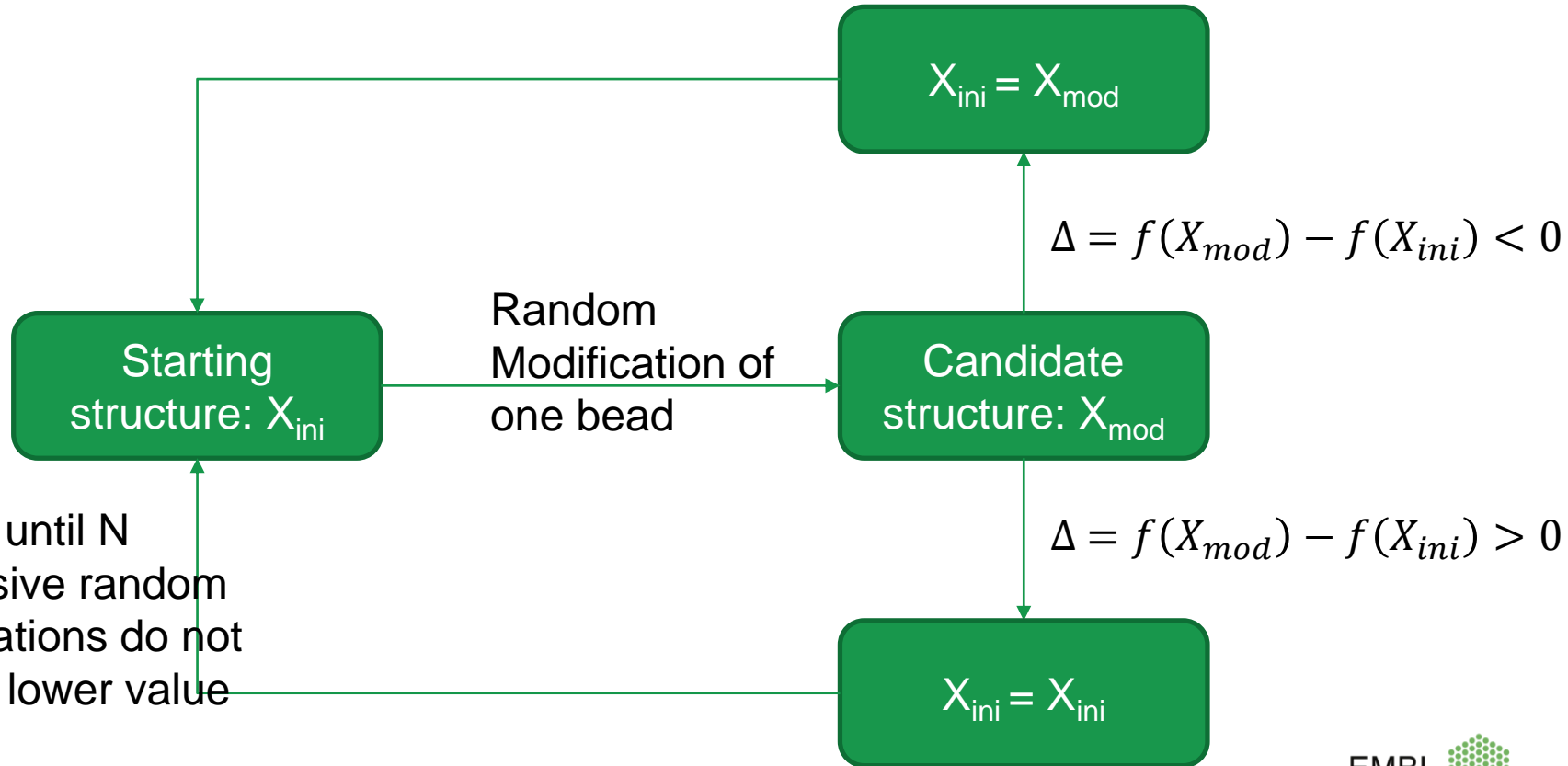
$$f(X) = \chi^2 + \alpha \cdot P(X)$$

Iterative approach:

- a bead is changed
- Evaluate the effect of the change on the target function



Pure Monte Carlo



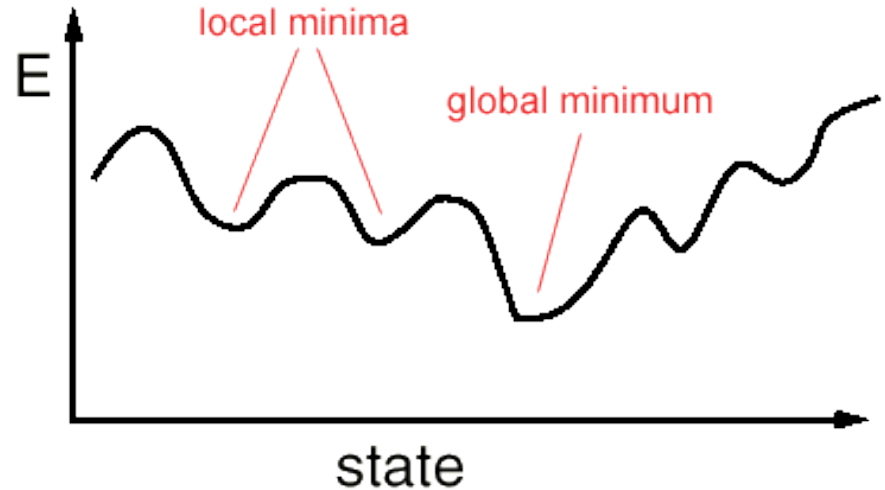
Repeat until N successive random modifications do not provide lower value

Local Minima vs global minimum

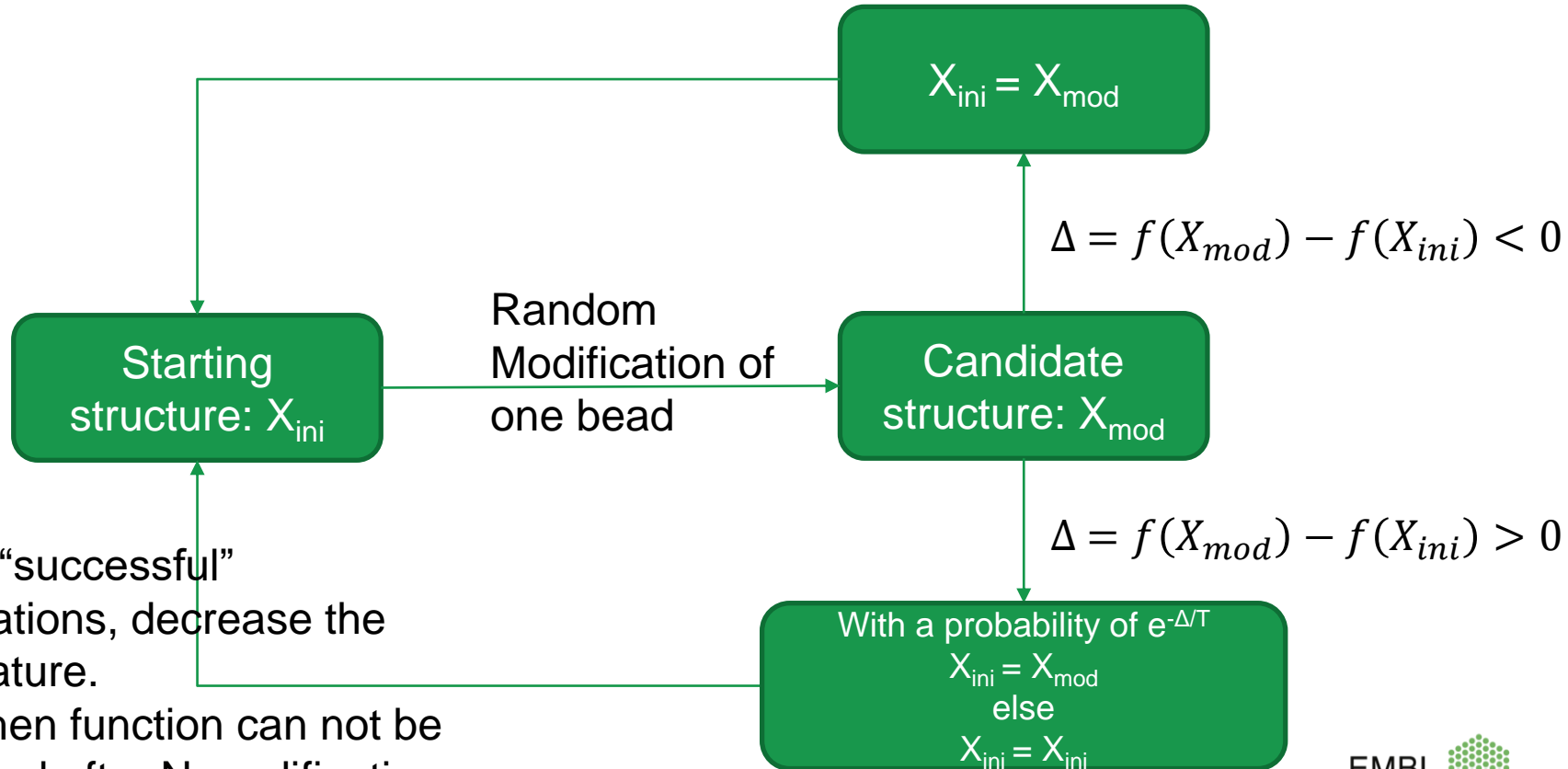
Pure Monte-Carlo search always goes to the closest local minimum (nature: rapid quenching and vitreous ice formation)

To get out of local minima, global search must be able to (sometimes) go to a worse point.

Slower annealing allows to search for a global minimum (nature: normal, e.g. slow freezing of water and ice formation)

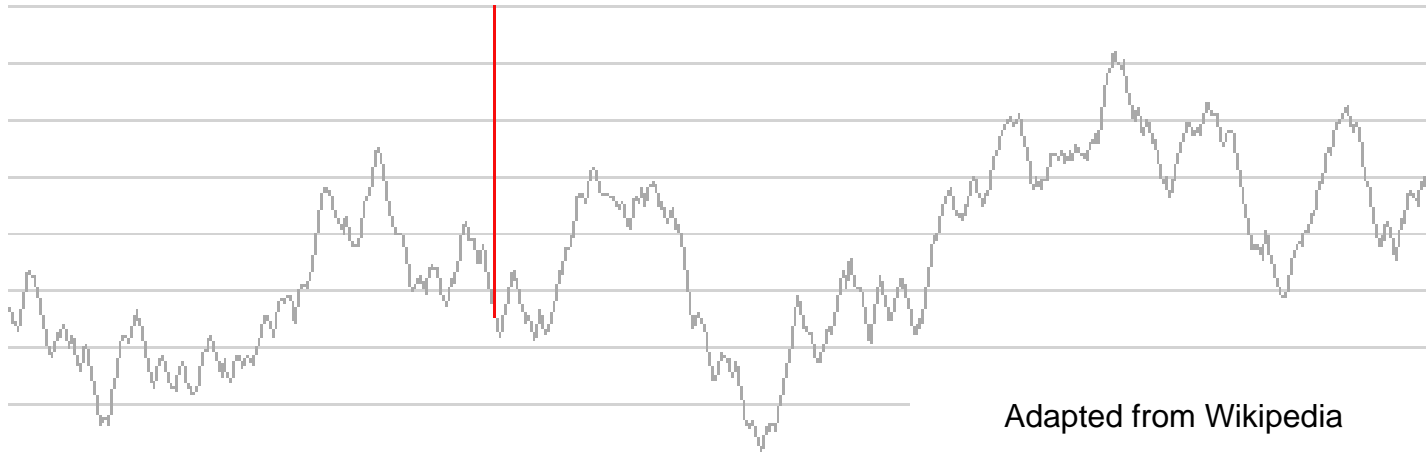


Simulated annealing



Repeat.
after M “successful”
modifications, decrease the
temperature.
Stop when function can not be
minimized after N modification.

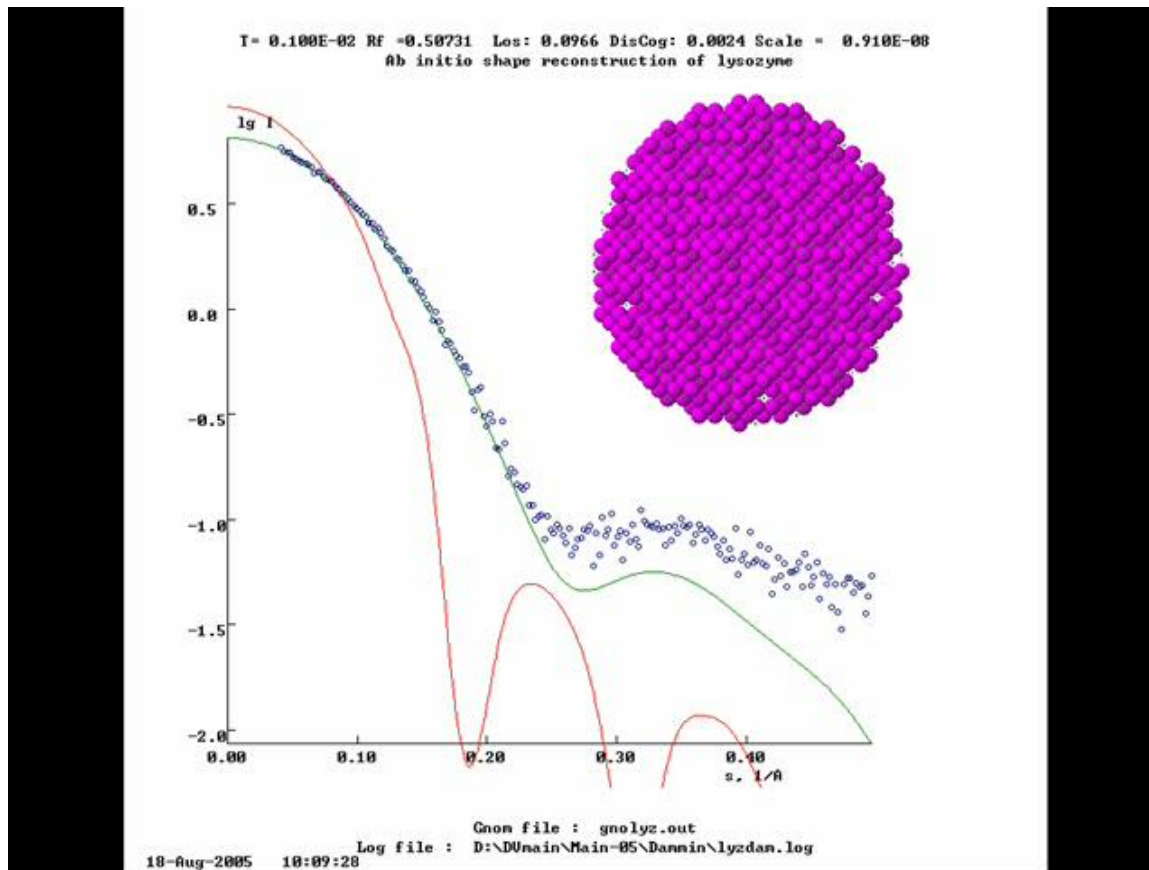
Simulated annealing



Ab initio programs



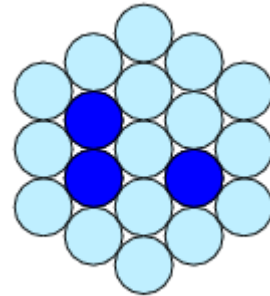
Bead modeling



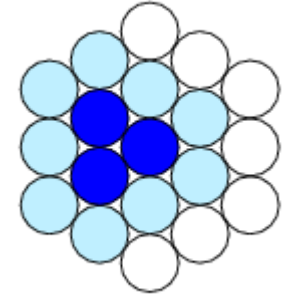
Bead modeling



- Reimplementation of DAMMIN written in object oriented code
- About 25 to 40 times faster (about a min for fast run on a PC)
- Make use of multiple CPU
- Use adaptive search volume



DAMMIN

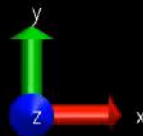
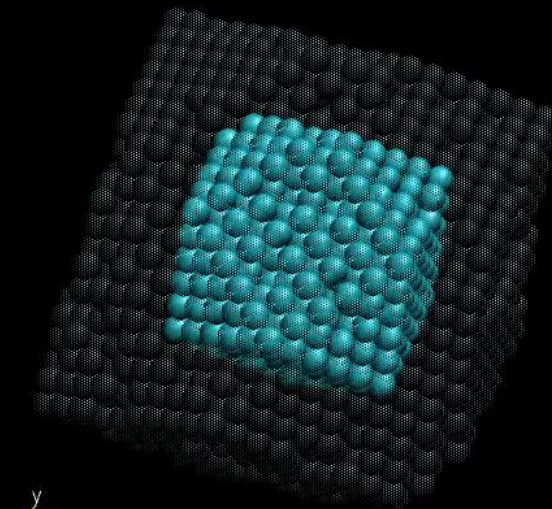


DAMMIF

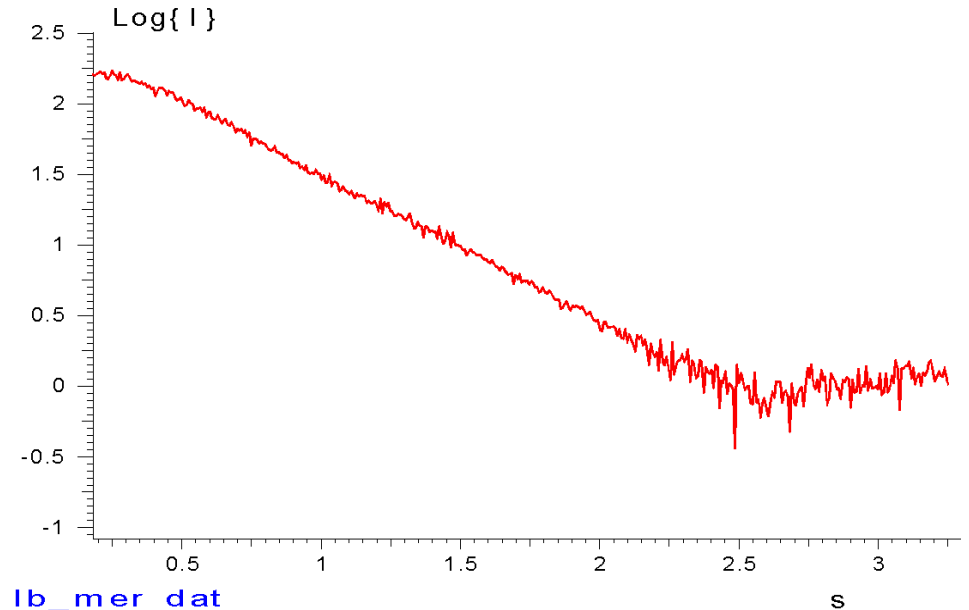
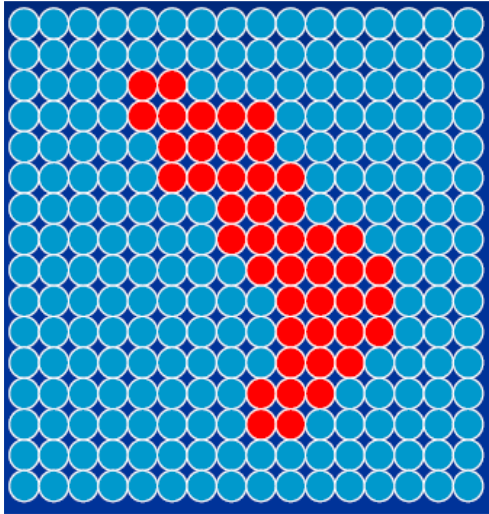
At the current iteration:

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

DAMMIF in action

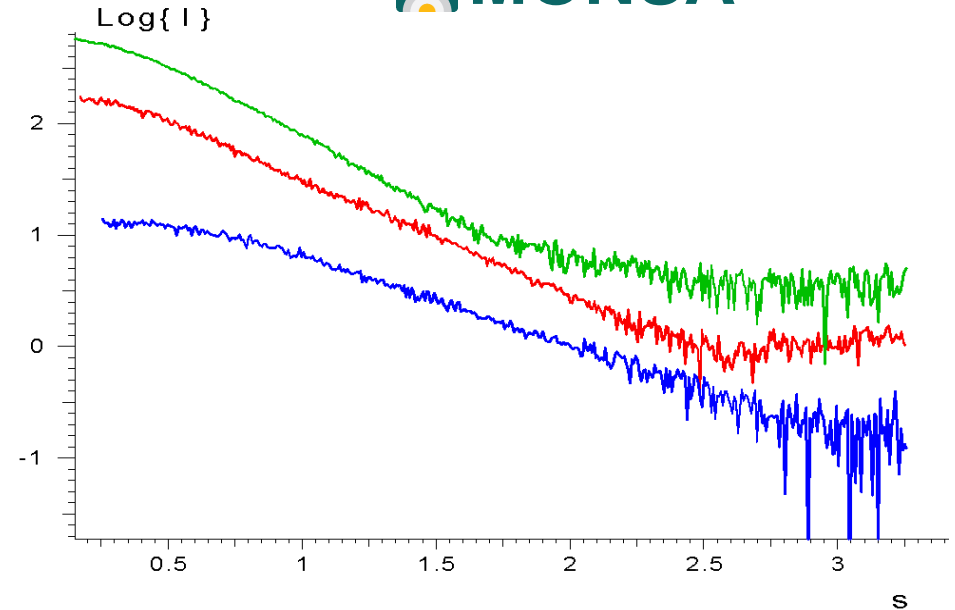
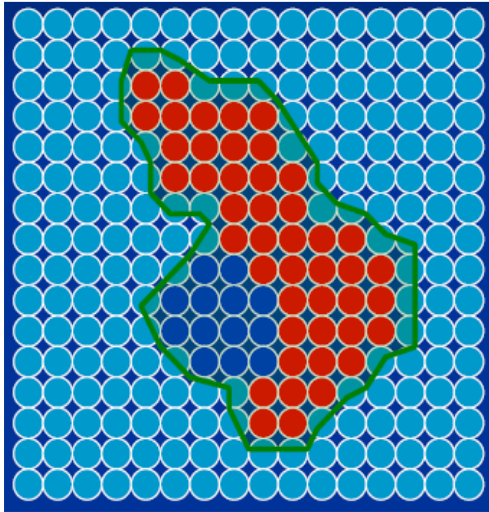


Ab initio modeling for multi-components systems



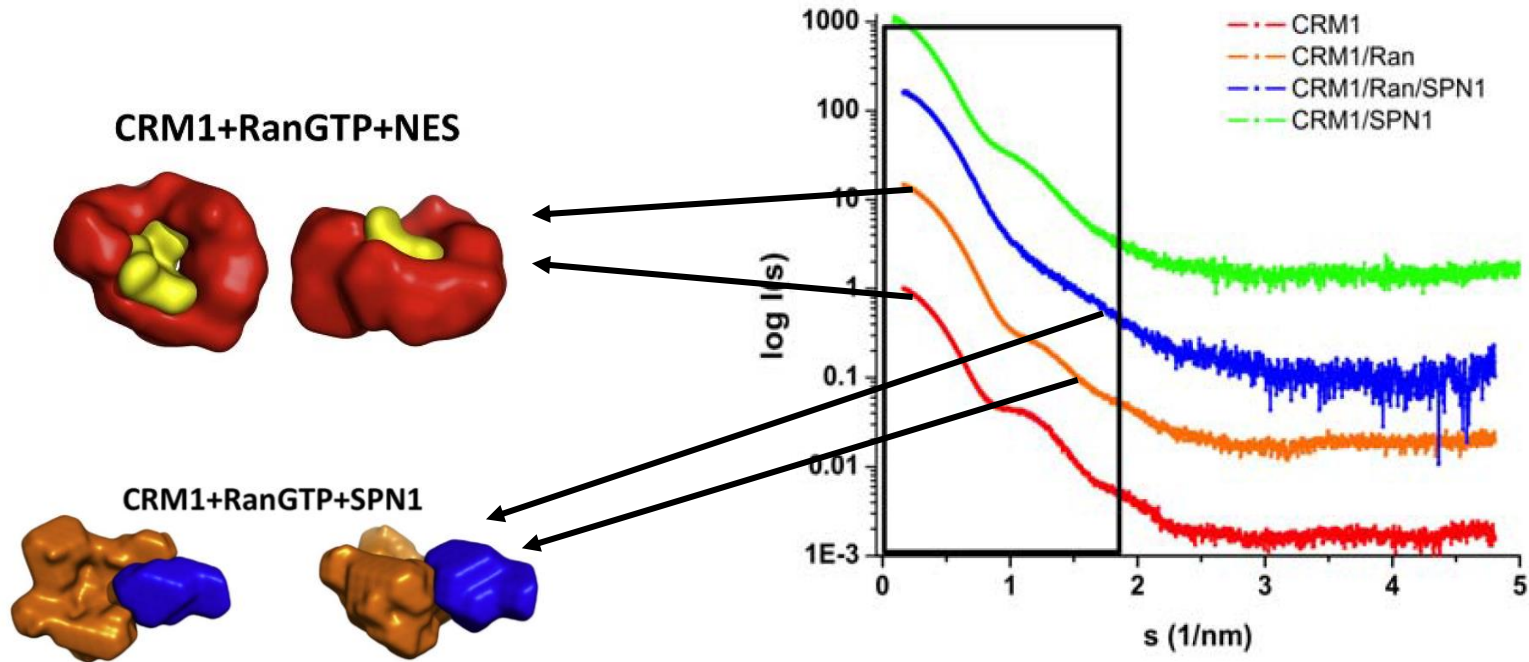
One component, one scattering curve:
Single phase ab initio modeling

Ab initio modeling for multi-components systems



Several component and several scattering curves:
Multiphase ab initio modeling

Example multi-component system



Extremely useful for contrast matched data



Ab initio modeling on mixtures

- Dummy atom modeling on mixture with known volume fraction

$$I_k(s) = v_{mk}I_m(s) + v_{ak}I_a(s) + v_{ik}I_i(s)$$

where v_{mk} , v_{ak} , and v_{ik} are the volume fractions of the components, $v_{mk} + v_{ak} + v_{ik} = 1$.

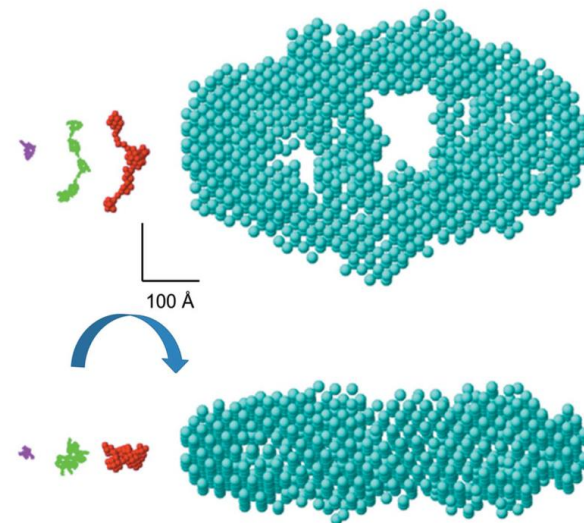
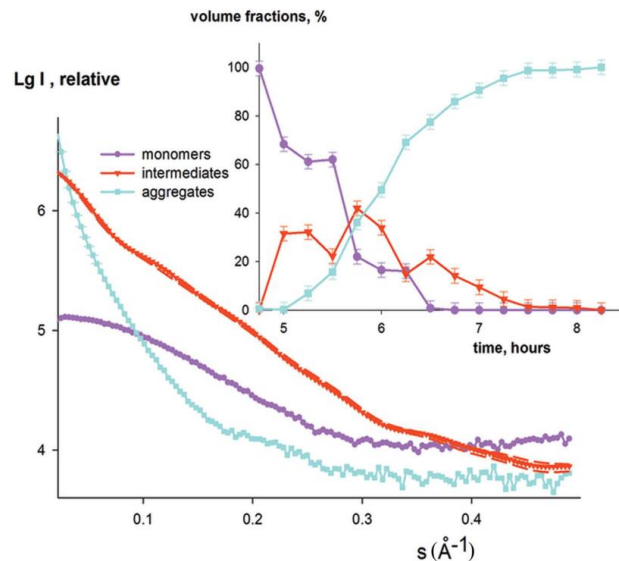
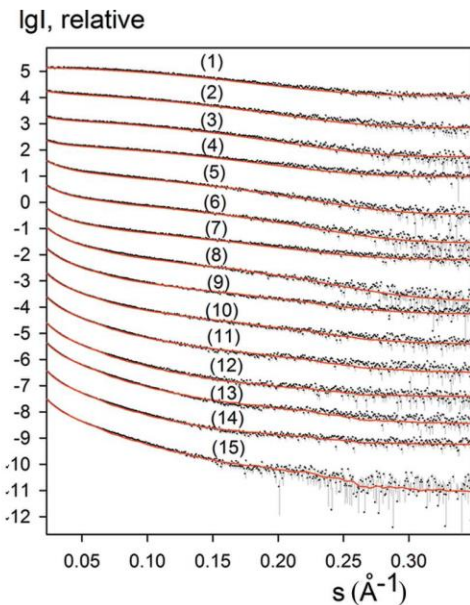
$$F(X) = \chi^2(X) + P(X)$$

$$F(X) = \sum_{k=1}^K \chi_k^2(X) + \sum_j W_j \times P_j(X)$$

Example amyloid fibrils

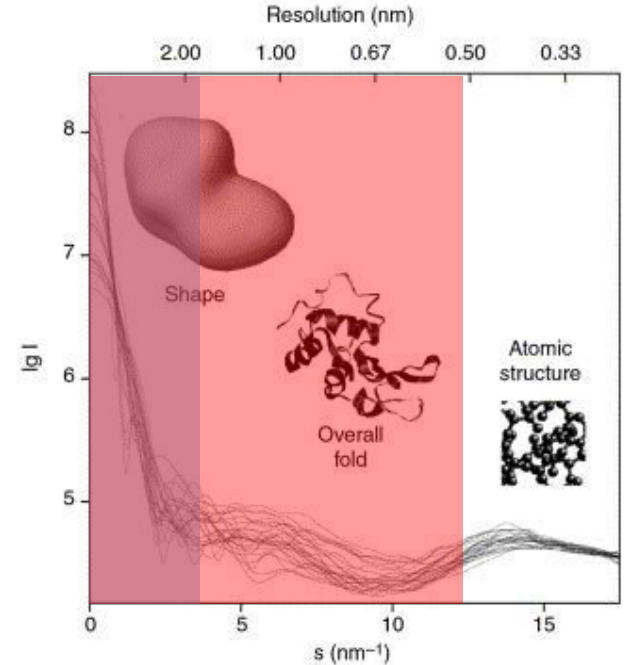
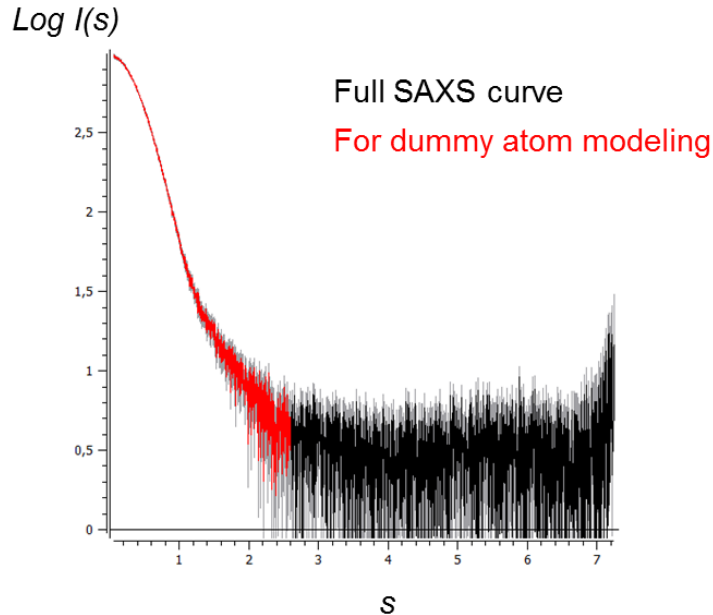
Vestergaard B, Groenning M, Roessle M, Kastrup JS, de Weert Mv, et al. (2007) A Helical Structural Nucleus Is the Primary Elongating Unit of Insulin Amyloid Fibrils . PLOS Biology 5(5): e134. <https://doi.org/10.1371/journal.pbio.0050134>

Konarev, P. V. & Svergun, D. I. (2018). IUCrJ 5, 402-409.



Resolution limit of dummy atom model

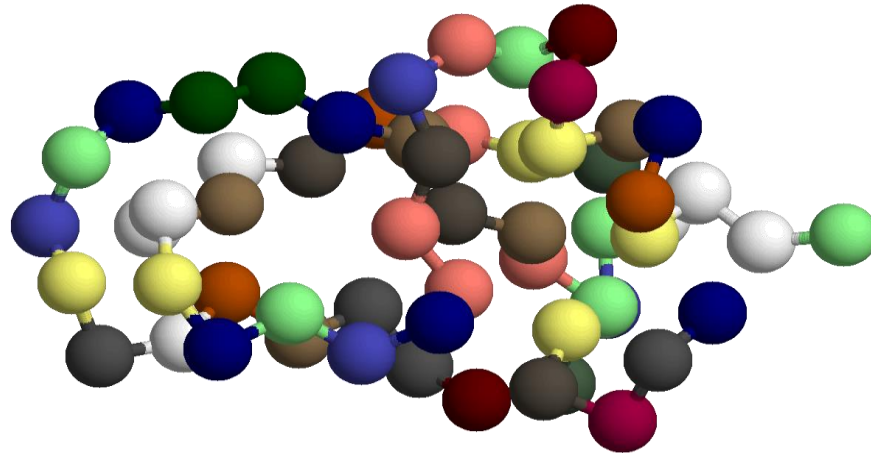
For dummy atom models, the electron density is considered as homogeneous \rightarrow only the shape is obtained, limited in resolution



\rightarrow Use of dummy residue modeling to increase the resolution.

Dummy residue models

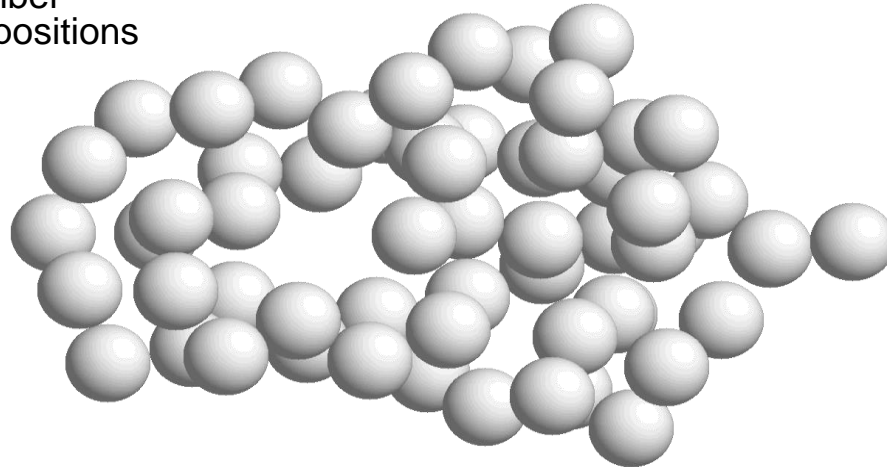
- Proteins consist of folded polypeptide chains composed of amino acid residues
- At a resolution of 0.5 nm each amino acid can be represented as one entity (dummy residue)



Dummy residue models

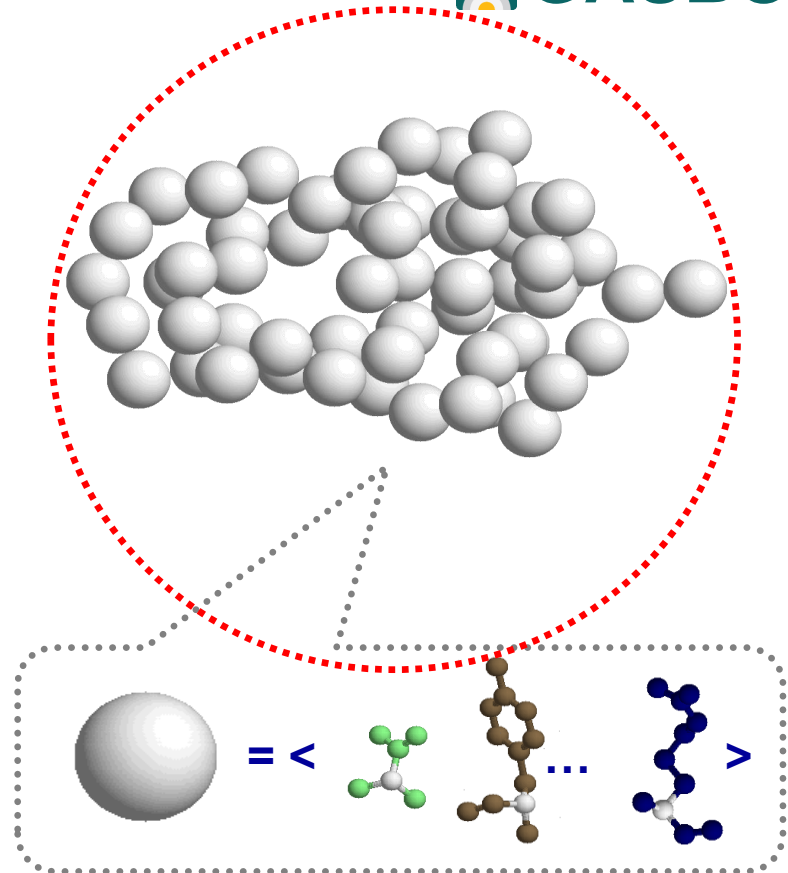
- Proteins consist of folded polypeptide chains composed of amino acid residues
- At a resolution of 0.5 nm each amino acid can be represented as one entity (dummy residue)
- In GASBOR a protein is represented by an ensemble of K dummy residues that are

Identical
Have no ordinal number
Centered at the $C\alpha$ positions



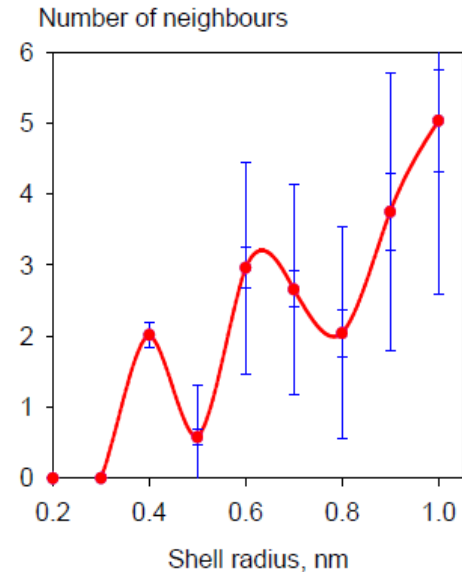
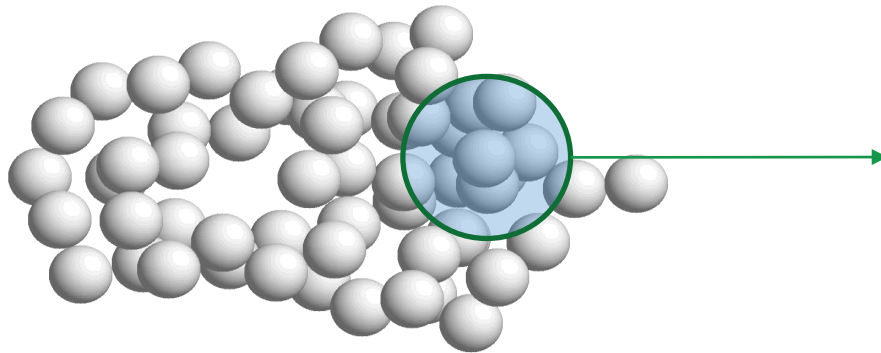
Dummy residue models

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

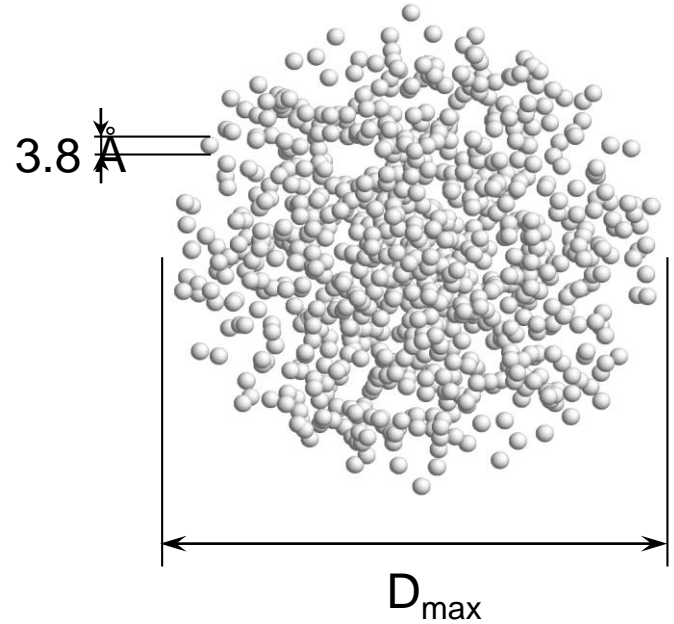


Distribution of neighbours

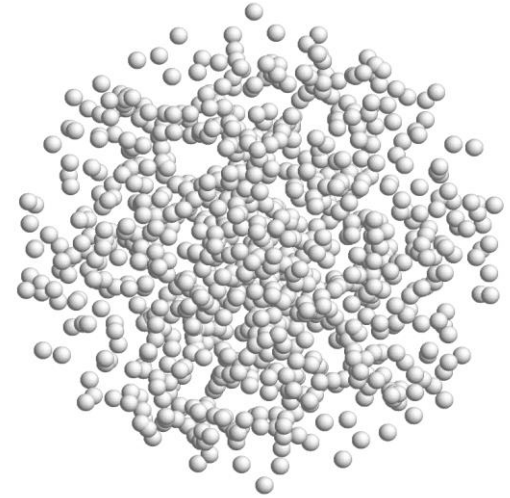
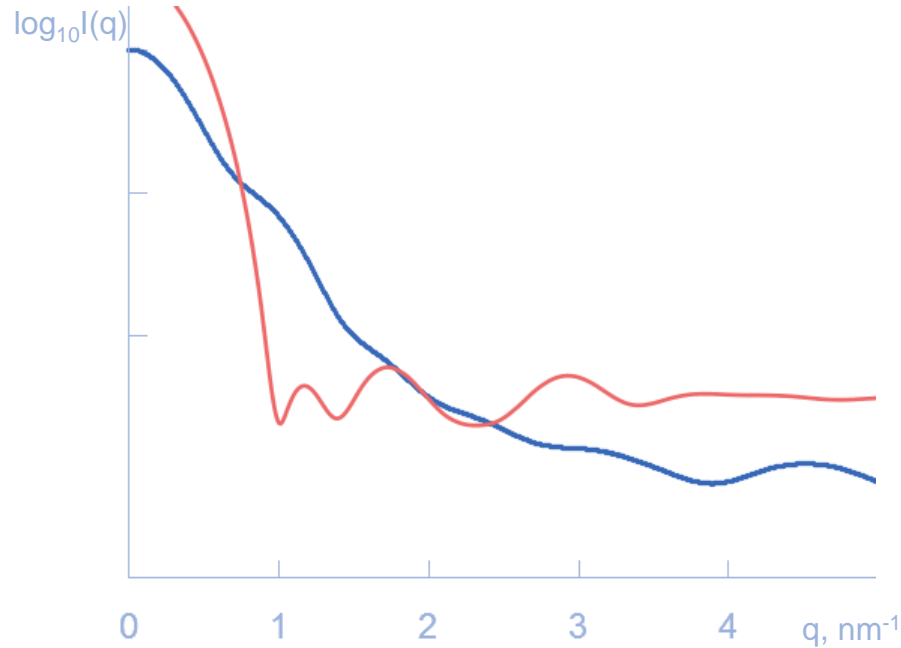
- Excluded volume effects and local interactions lead to a characteristic distribution of nearest neighbors around a given residue in a polypeptide chain



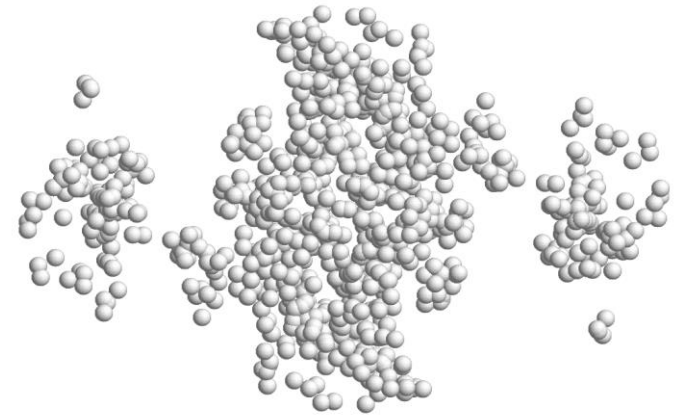
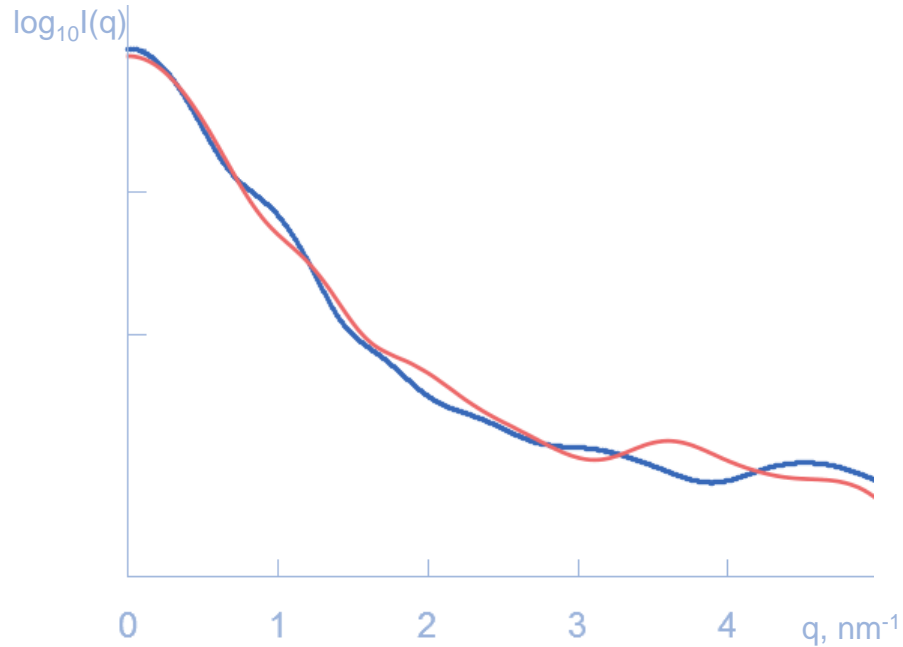
Ab initio reconstruction: dummy residue modeling



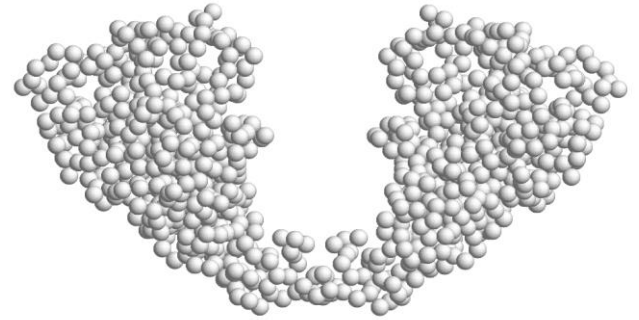
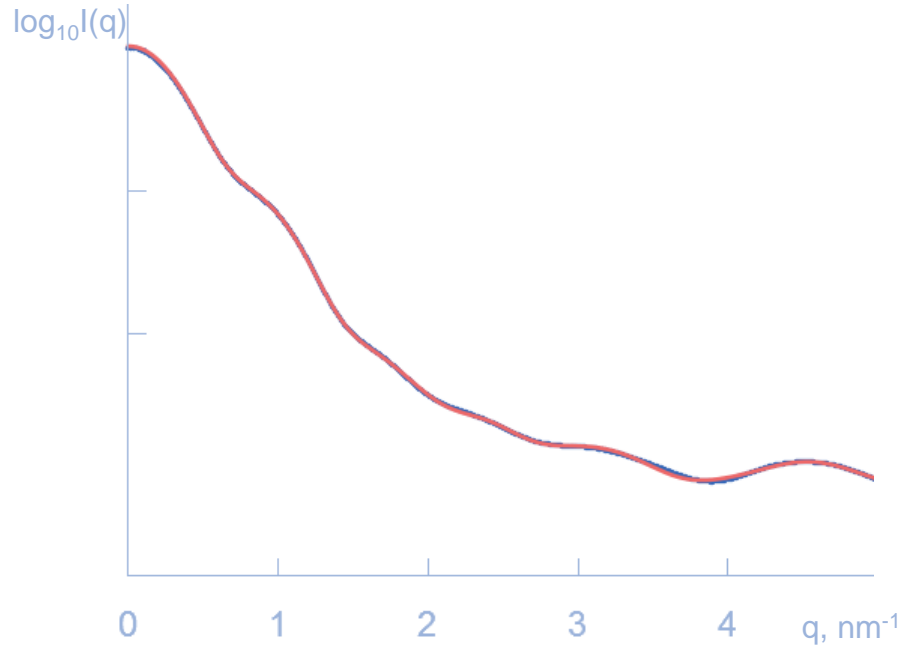
Ab initio reconstruction: dummy residue modeling



Ab initio reconstruction: dummy residue modeling

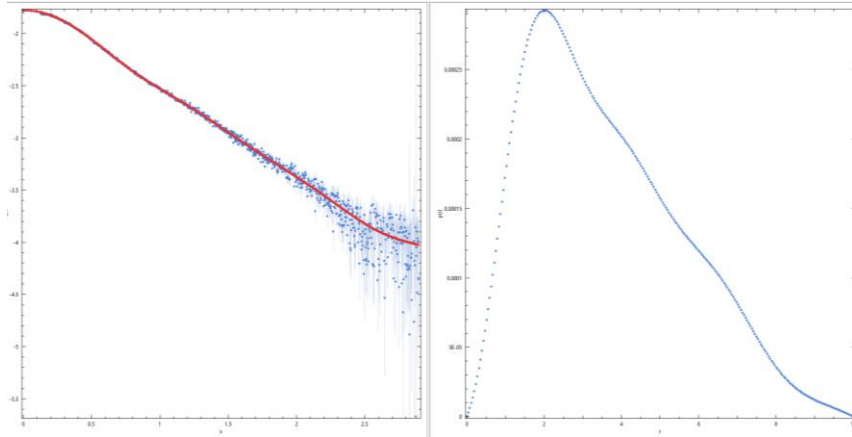


Ab initio reconstruction: dummy residue modeling



Flavors of Gasbor

GarborI, GarborP, GasborMX



Use GASBORI:

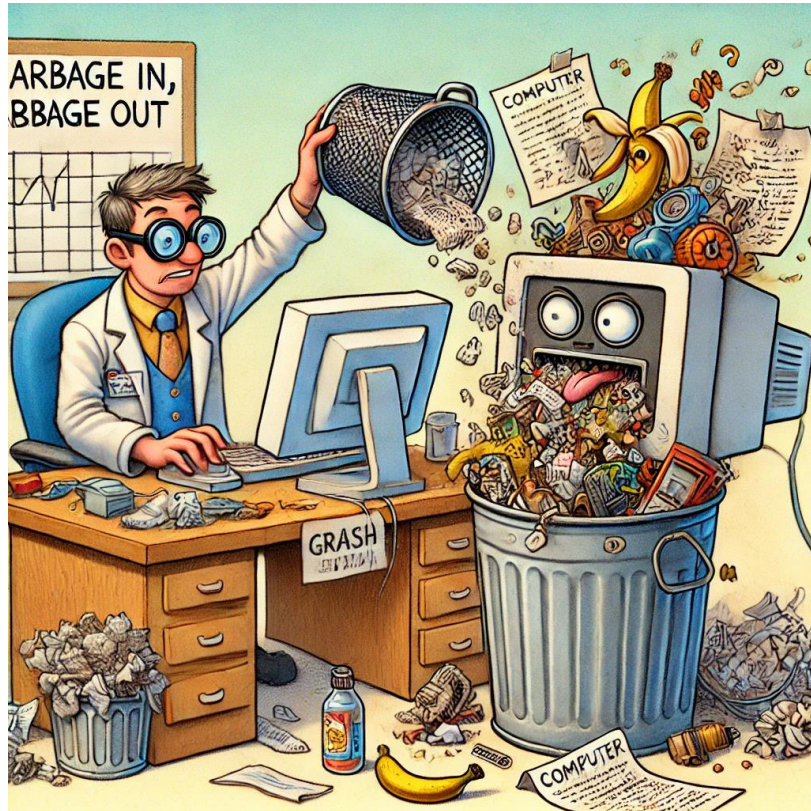
- When you have high-quality, noise-free SAXS data.
- When you want to capture detailed structural features.
- When a direct fit to the experimental $I(s)$ data is needed.

Use GASBORP:

- When you want to model the particle in real space based on the $p(r)$ function.
- When you have noisy $I(s)$ data and prefer to work with a smoother, less noisy representation.
- When the overall shape and size of the macromolecule are the primary focus.

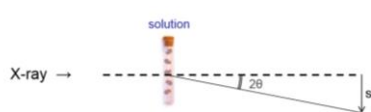
Words of caution

Garbage in, garbage out



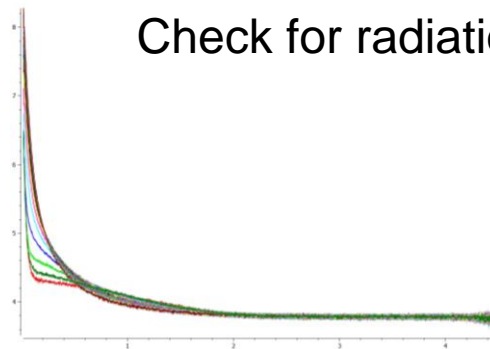
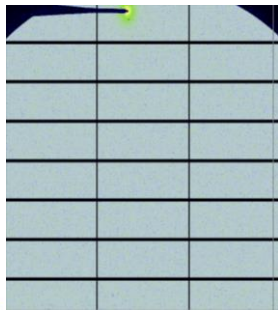
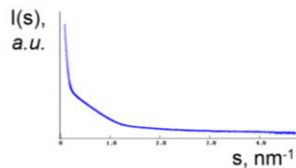
- Ab initio modeling will always give a **nice looking model that fit the data, even if the data are completely wrong.**
- Make sure that the SAXS curves used for ab initio correspond to the form factor of the solutes you are trying to measure.

Careful measurement and data reduction



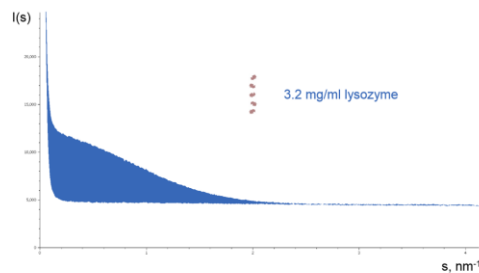
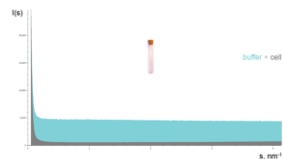
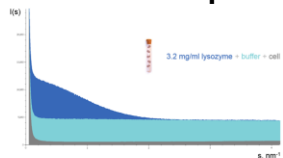
$$|s| = 4\pi \sin\theta/\lambda$$

2θ – scattering angle
 λ – wavelength
 s – scattering vector
 $I(s)$ – intensity



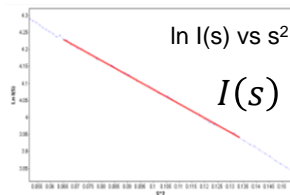
Check for radiation damage

Proper buffer subtraction



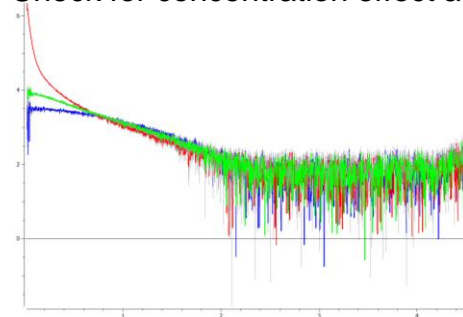
Check overall parameters

Radius of gyration (Guinier)



$$I(s) = I_0 \exp\left(-\frac{1}{3} s^2 R_g^2\right)$$

Check for concentration effect and aggregation



MW using forward scattering

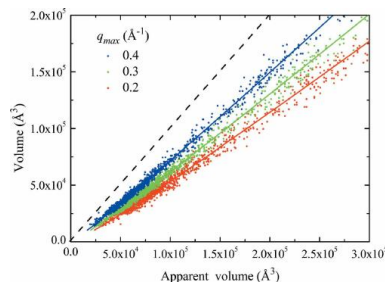
$$MW = \frac{I(0)}{c} \cdot \frac{(c_{st} \cdot MW_{st})}{I(0)_{st}}$$

Porod Volume

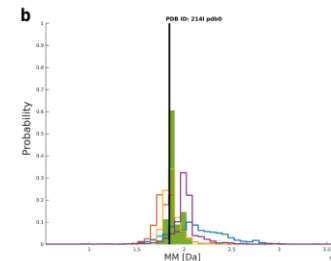
Volume of correlation

$$V_C = \frac{I(0)}{\int q \cdot I(q) dq}$$

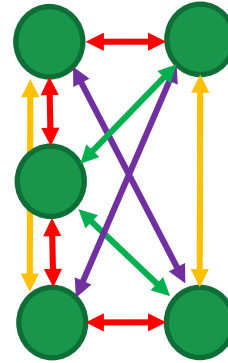
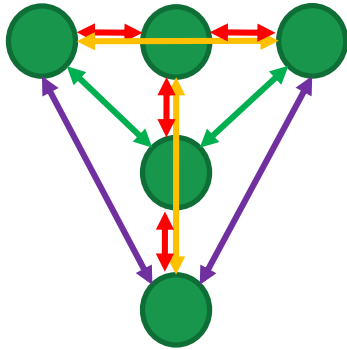
SAXS mow



Bayesian MW



Ambiguity in SAXS

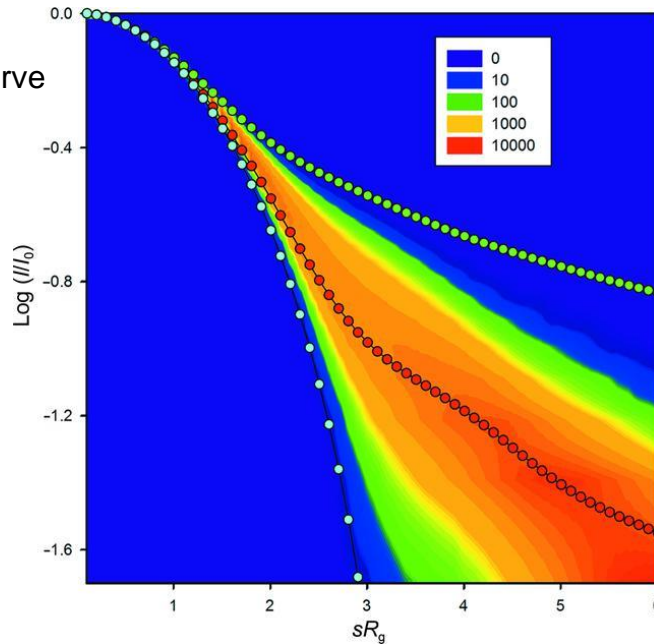
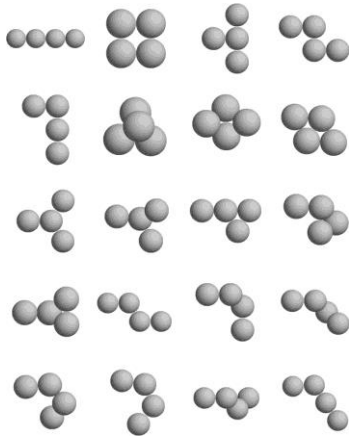


- Same distances in the two conformations
- Leads to the exact same SAXS pattern

Estimate the ambiguity within your curve



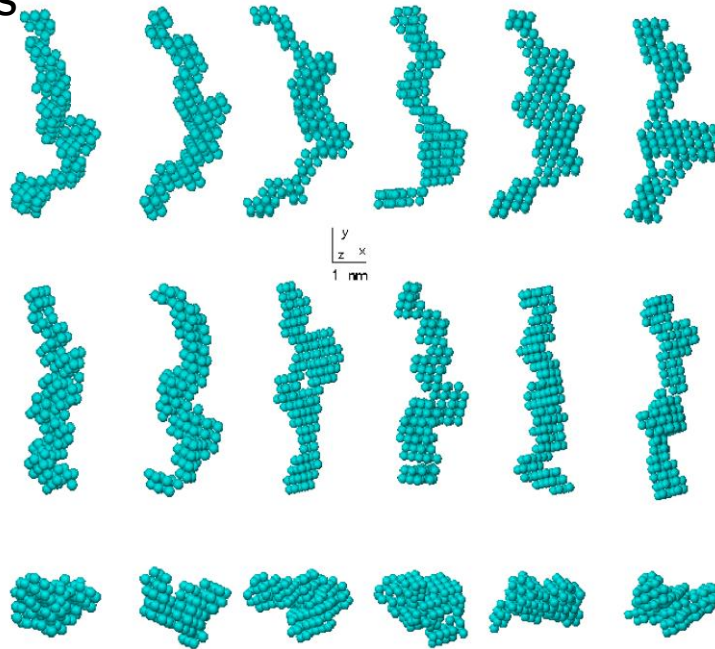
Quantitative measure of the ambiguity from the SAXS curve



- 14000 shape topologies generated (up to seven beads closely packed on hexagonal grid).
- Scattering curves computed and rescaled to keep only shape topology information.
- Scattering map computed from these curves.
- By plotting the experimental SAXS curves on the map, ambiguity intrinsic to the curve can be estimated .

Ab initio model validity

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

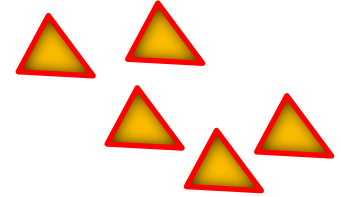
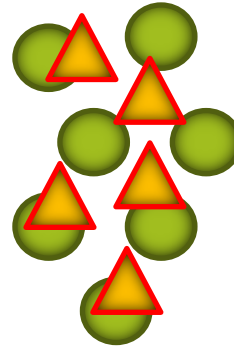


Ab initio model validity

- To reduce ambiguity, several models are built, compared and averaged

- Align models (CIFSUP):

- Different methods for alignment
 - ICP (interactive closest point, default)
 - NSD (normal spatial discrepancy)
 - NCC (normalized cross correlation)



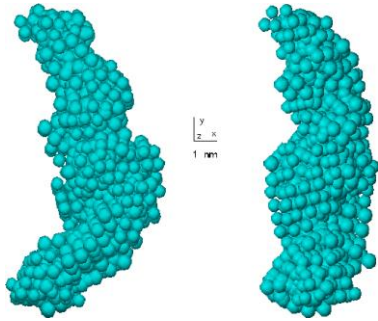
- Compare models:

- Compute “Distances” between the model (how different are they?)
- Possibly create cluster of domain

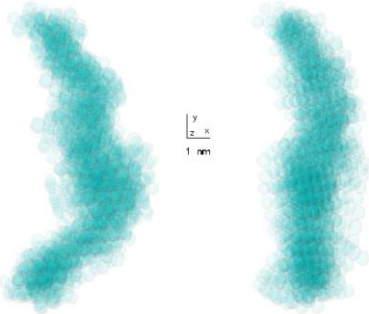
- Indicate the most typical model (+ filtered model, and solution spread region)

- Refine!

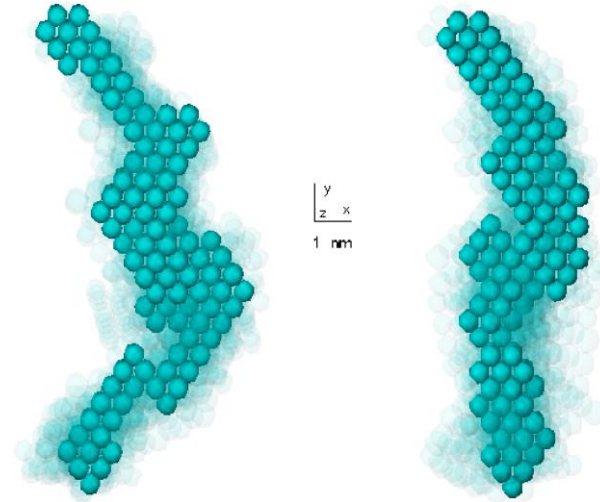
Model Validity



5S RNA – Solution spread region

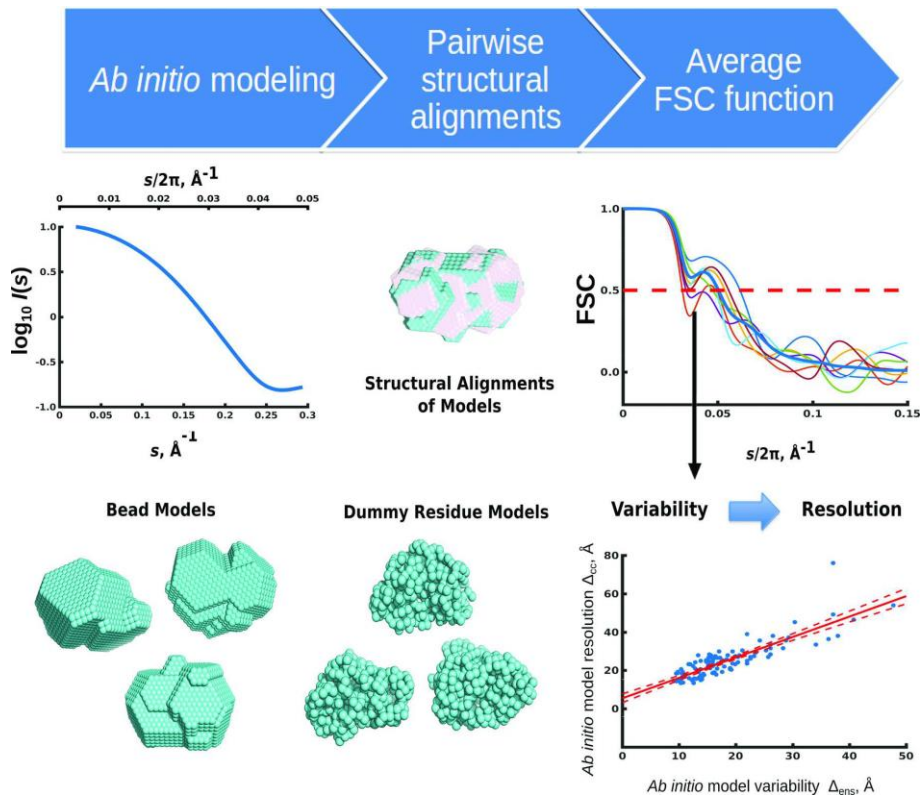


5S RNA – Most Populated Volume



5S RNA – Final Solution
within the Spread Region

Estimate the resolution of your ab initio model



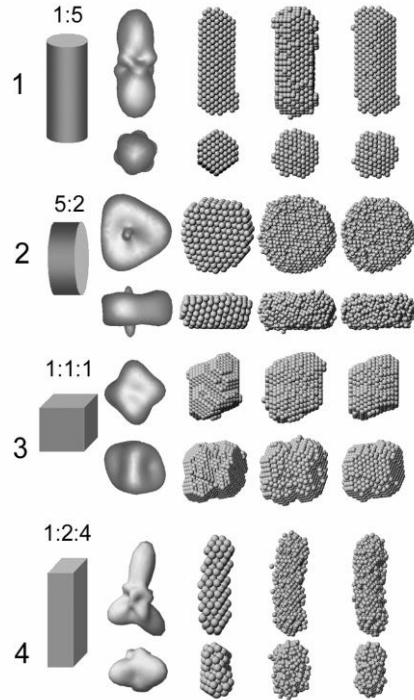
“Measure ambiguity to estimate resolution”

Resolution estimated from a set of (10-20) bead model.

Model compared and aligned.

Measure of the variability gives an estimation of the resolution

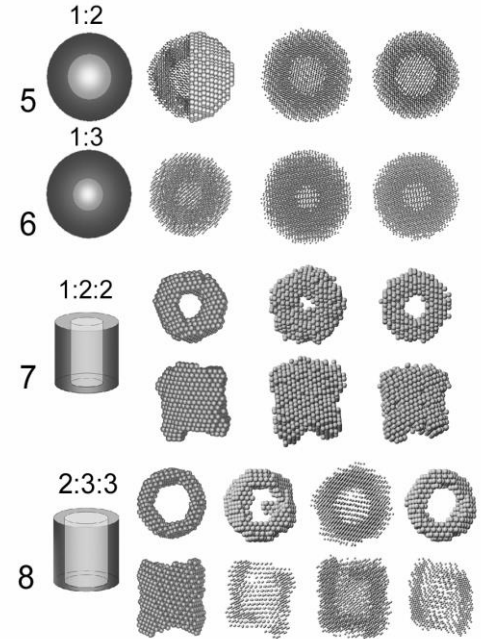
Can all shapes be reconstructed by ab initio modeling?



Globular solid

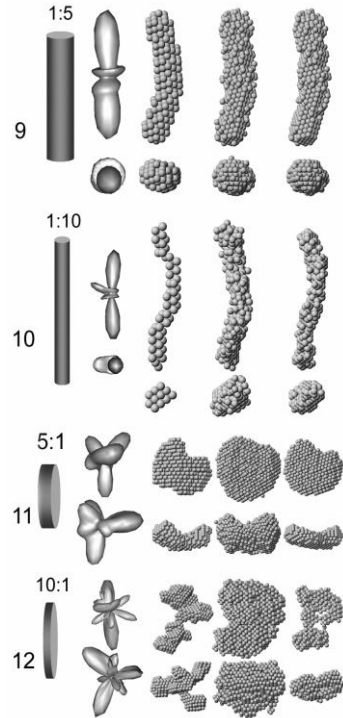
Solid bodies with moderate anisotropy (elongated particles up to 1:5 and flattened up to 5:2) can be reliably reconstructed from the SAXS data. Mean value NSD : 0.4-0.7

Hollow globular models can also be well reconstructed



Hollow globular

Can all shapes be reconstructed by ab initio modeling?



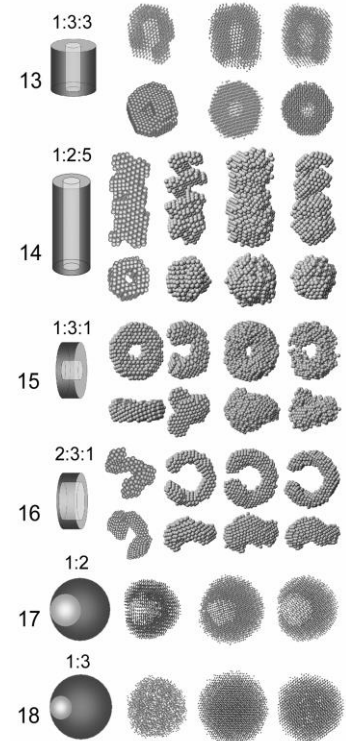
anisotropic solid

Shape reconstructions of anisometric particles are less stable and reliable.

Elongated hollow body: the channels may appear closed from one or both sides

For hollow flattened the resulting shapes may show a helical turn instead of a hollow disk, even after the averaging.

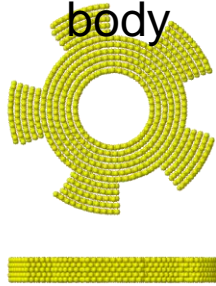
Acentric voids in hollow spheres are only reconstructed if r/R is about 0.5



Hollow anisotropic and acentric

Use of symmetry

Original
body



Reconstruction of electron density from the SAXS data

Reconstructing electron density

$$A(\mathbf{s}) = \int_{V_r} \rho(\mathbf{r}) \cdot e^{-i\mathbf{r} \cdot \mathbf{s}} dV_r = |A(\mathbf{s})| e^{i\Phi(\mathbf{s})}$$

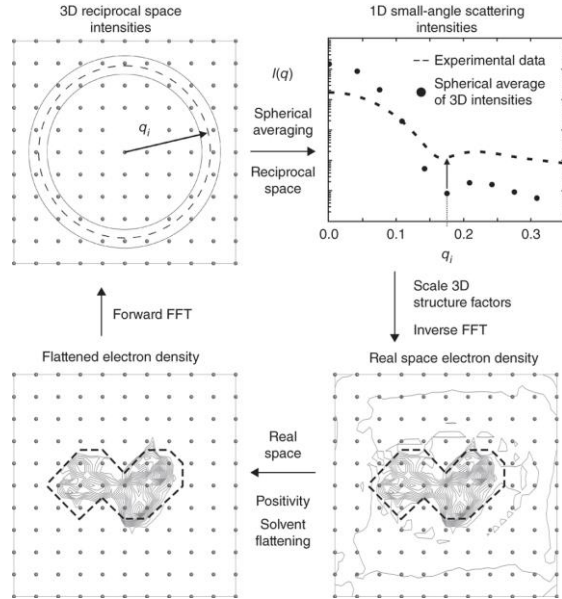
$$I(\mathbf{s}) = A(\mathbf{s}) \cdot A^*(\mathbf{s}) = |A(\mathbf{s})|^2$$

$I(\mathbf{s})$ is measured

→ the phase information ($\Phi(\mathbf{s})$) is lost

→ the electron density $\rho(\mathbf{r})$ cannot be directly reconstructed

DENSS

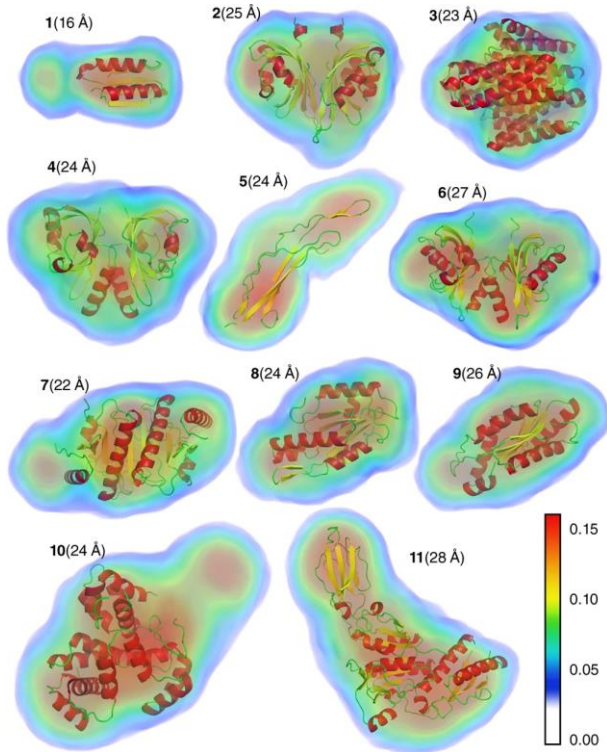


DENSS is a program allowing to reconstruct the electron density from the SAXS data through iterative phase retrieval:

- One starts from an initial (random) electron density
- SAXS data are computed from this electron density
- The computed curves are compared to the experimental ones
- Electron density is modified to improve the fit.
- Iterative process until a stable solution is found

Grant, T. *Ab initio* electron density determination directly from solution scattering data. *Nat Methods* **15**, 191–193 (2018). <https://doi.org/10.1038/nmeth.4581>

DENSS



Compared to bead model:

- Continuous electron density (not envelop)
- Nice model
- Much more computer resourced intensive

Stay careful about ambiguity, resolution

Summary

- Ab initio methods allows one to obtain a three dimensional shape of the solutes from the SAXS curves
- Different program exist for ab initio methods for differeents tasks:
 - single phase dummy atom modeling (DAMMIN, DAMMIF)
 - multiphase dummy atom modeling (MONSA)
 - Dummy atom modeling of mixture (DAMMIX)
 - Dummy residue modeling (GASBORI, GASBORP)
 - Dummy residue modeling of mixture (GASBORMX)
 - Electron density reconstruction (DENSS)
- Ab initio methods generally give “nice” models, that will fit your data → Make sure you have nice data (GIGO)
- Ab initio model have limited resolution and can be ambiguous.
 - Repeat modeling, average, filter, refine...