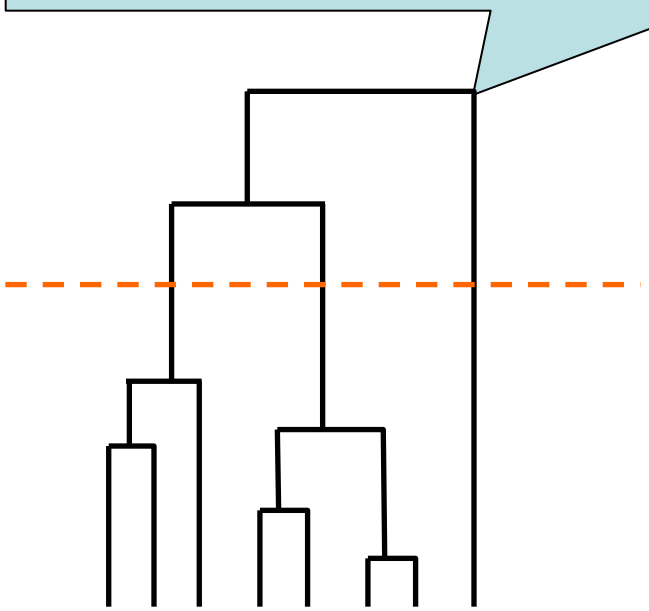
DamClust: Assessment of multimodality (*has damaver & friends inside*)

Creates the complete graph by iteratively joining the clusters (singles)

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

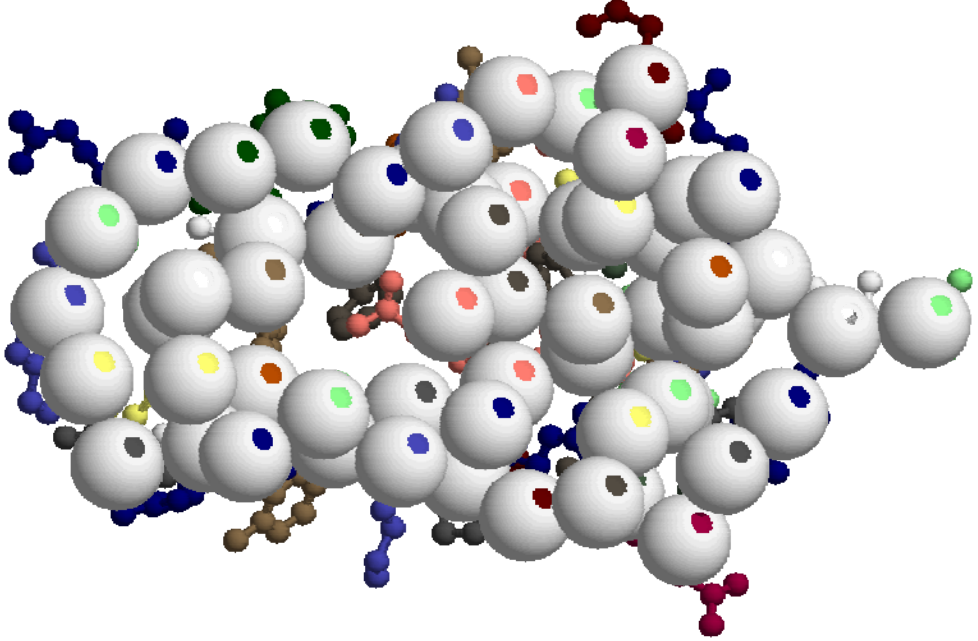


Clustering of multiple SAS models

- 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)

Dummy Residues Model

- 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* DRs those are
  - Identical
  - Have no ordinal number
  - For simplicity are centered at the Cα positions

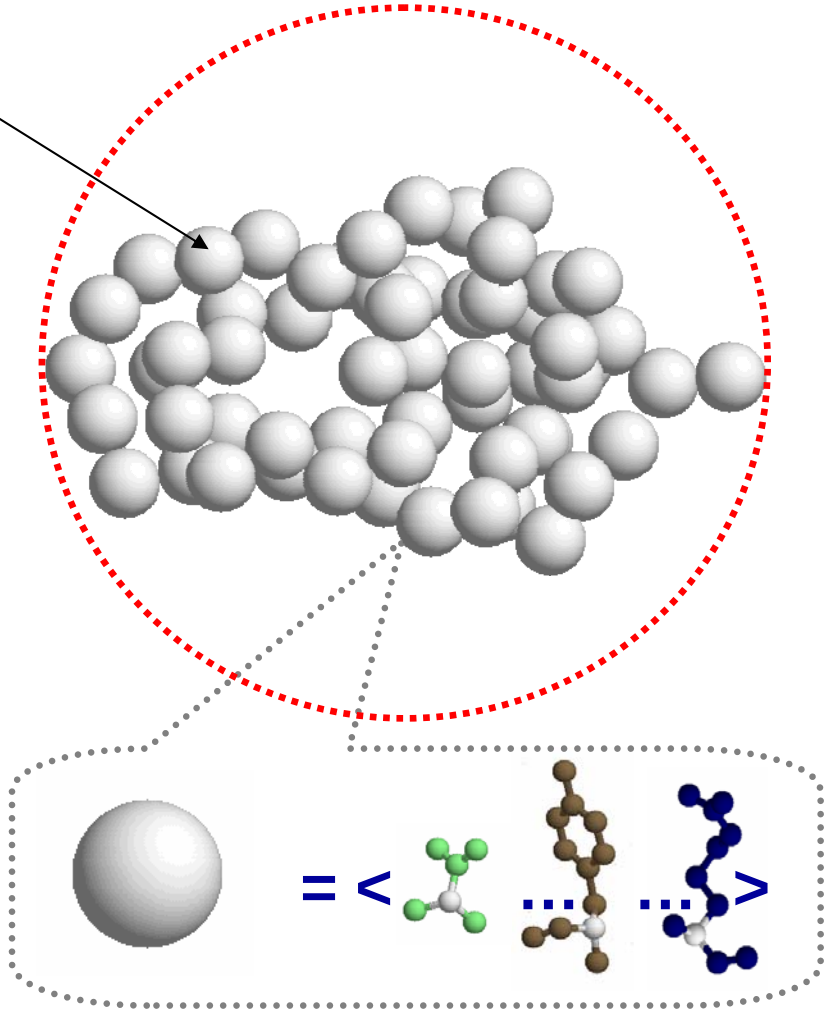


DR Modelling: GASBOR

- Finds coordinates $\{r_i\}$ of *K* DRs within the spherical search volume
- Scattering is computed using the Debye (1915) formula

$$I_{DR}(s) = \sum_{i=1}^{K}\sum_{j=1}^{K} g_i(s) g_j(s) \frac{\sin sr_{ij}}{sr_{ij}}$$

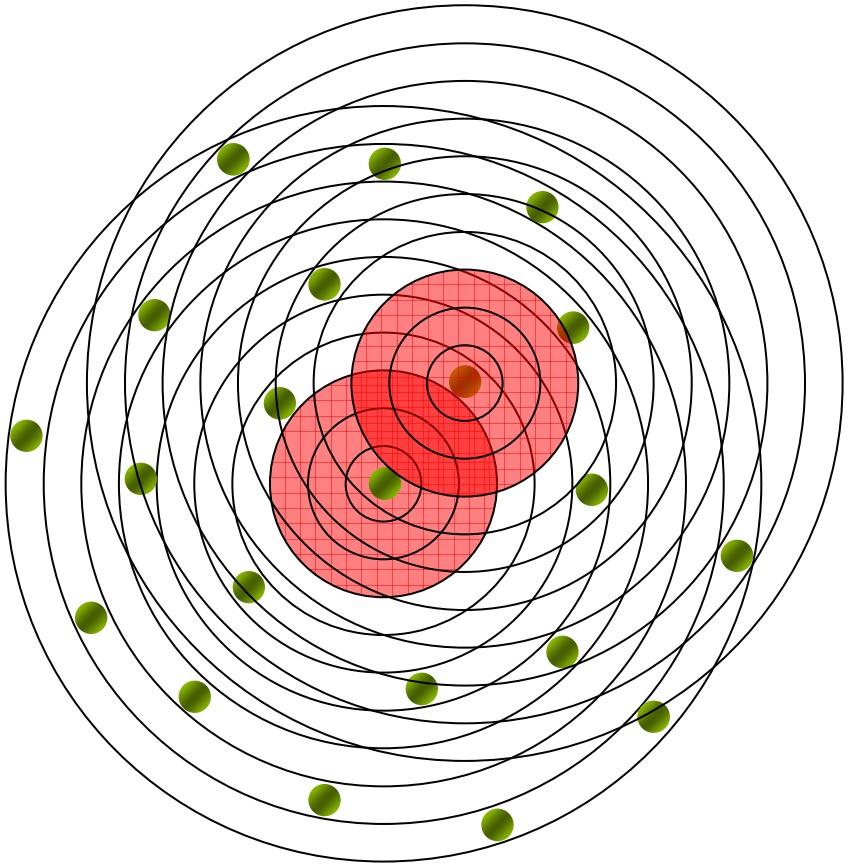
- Requires polypeptide chain-compatible arrangement of DRs

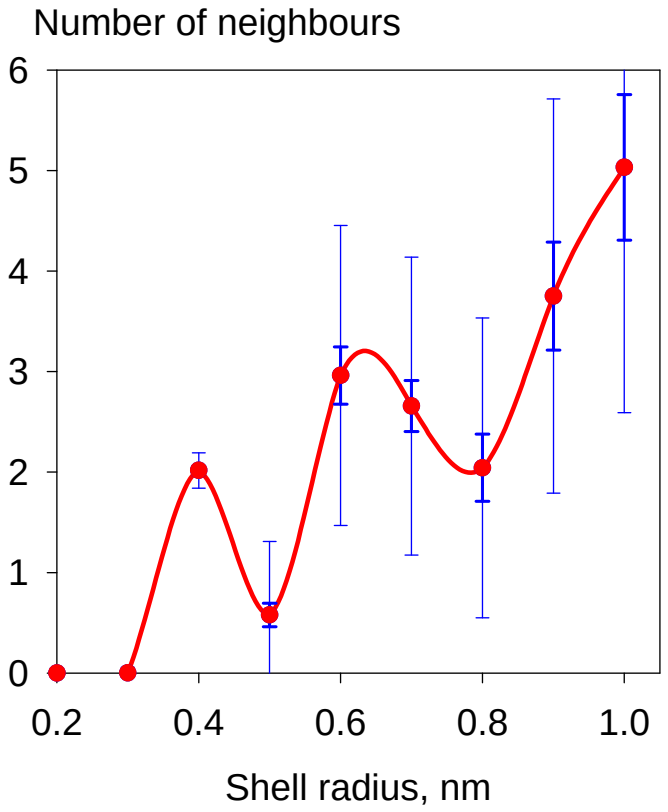


D.I. Svergun, M.V. Petoukhov, & M.H.J. Koch (2001) *Biophys. J.* 80, 2946-53

GASBOR Restraints

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

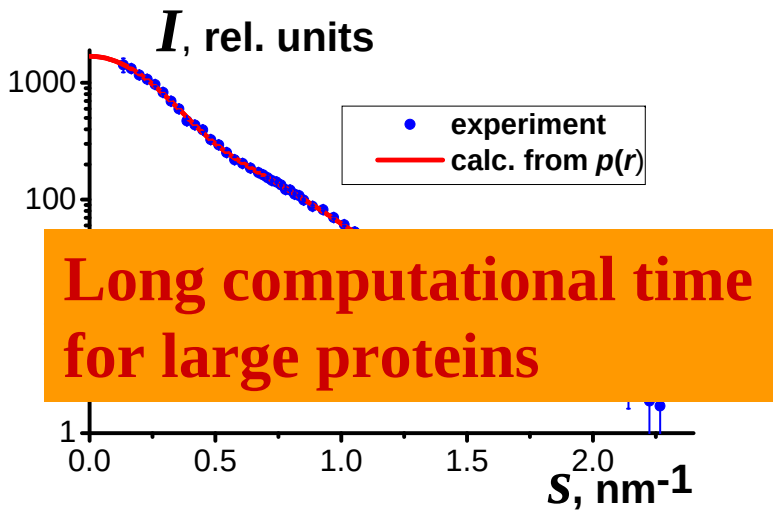




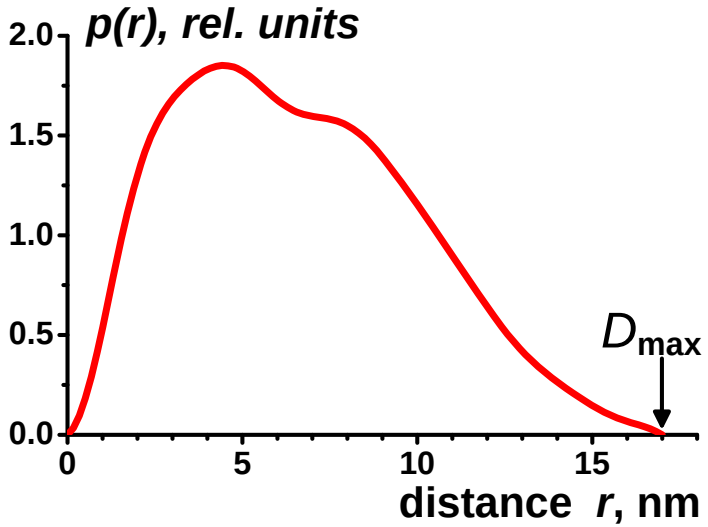
DR modelling program GASBOR

searches for a chain-like arrangement of dummy residues fitting the scattering data by minimizing

$$f(X) = \chi^2 + \Sigma \alpha_i P_i(X)$$



Long computational time for large proteins



- Symmetry can be taken into account; groups supported: Pn, Pn2 (n=2..19), P23, P432 and icosahedral symmetry

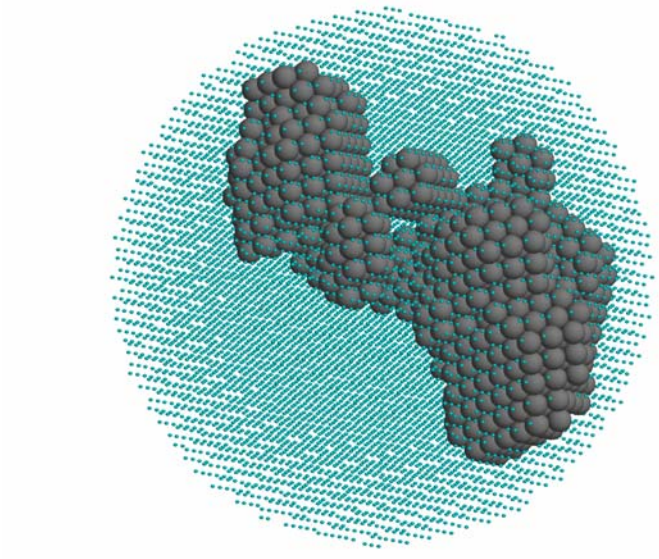
GASBOR Summary

- **Task:** Searches for a chain-like arrangement of dummy residues fitting the scattering data.
- **Parameters:** 3D coordinates of DRs describing Cα positions.
- **Objects:** Applicable for polypeptide chains (*i.e.* proteins and their assemblies) with $K \lesssim 5000$.
- **Capabilities:** Fits the scattering curves at higher angles. Takes into account symmetry and anisometry. Reciprocal and real space fitting options.
- **Limitations:** Fits only SAXS data and applicable for proteins only. CPU time is quadratically proportional to *K*.
- **Theoretical intensity:** computed using the Debye formula.

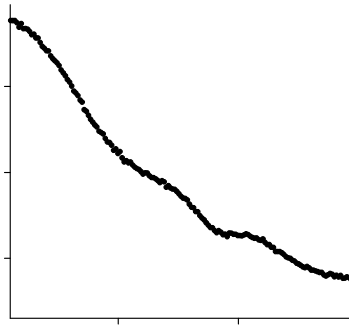
GASBOR examples

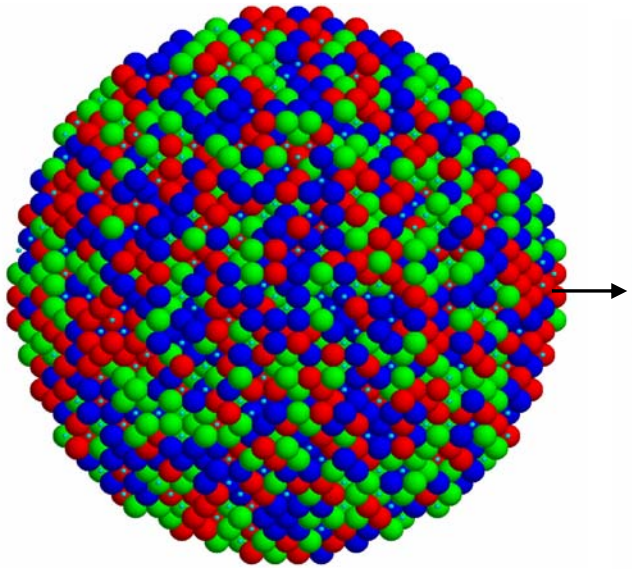
- NEOCARZINOSTATIN, 113 AA: Patrice
- GST homodimer with P2 symmetry, 240 AA per monomer: Kate

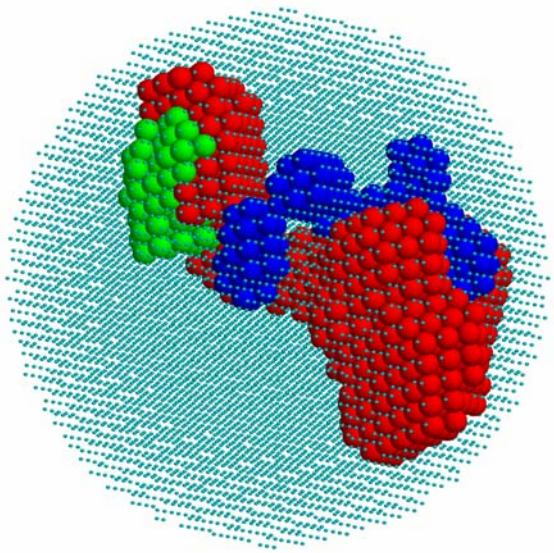
MONSA (multiphase modelling)



DAMMIN/F shape determination



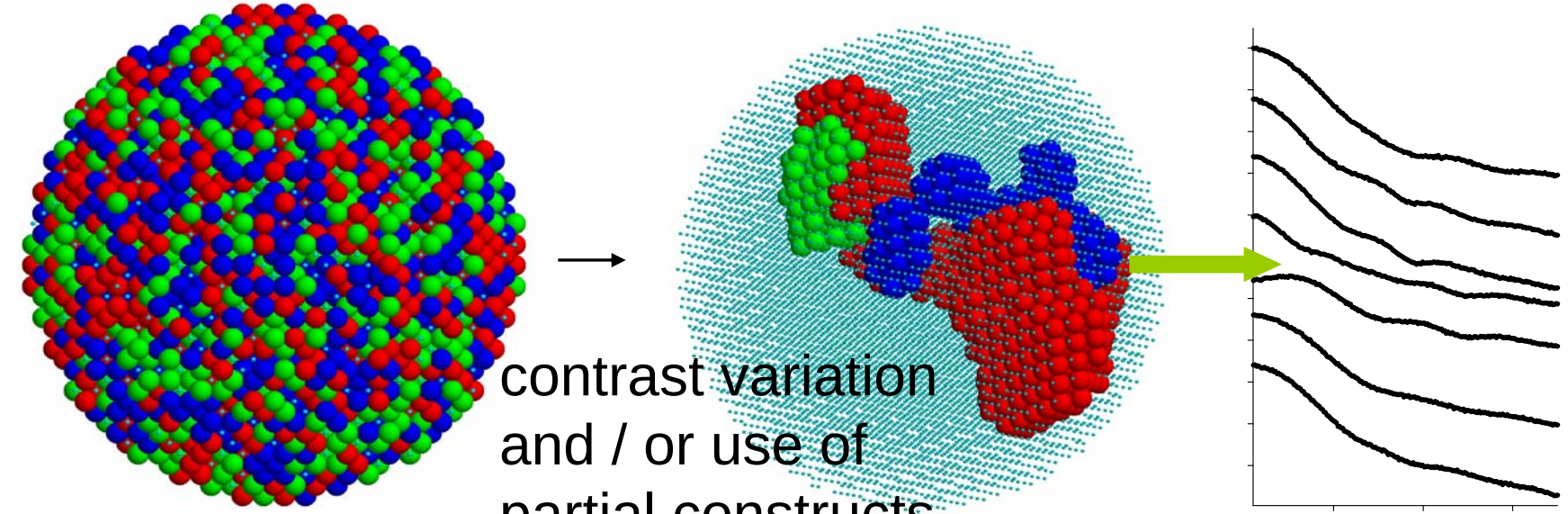




- One can differentiate between distinct parts of the particle
- Several curves are required
- Assuming the same arrangement of the parts in different samples

MONSA (multiphase modelling)

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



Case1: protein-RNA complex

252 AA protein **(two domains)**

67 nucleotides RNA

Three curves in total:

- Free RNA
- Complex
- Free protein