Mixtures, Assemblies, Flexible Systems

Maxim Petoukhov
EMBL, Hamburg Outstation
SAXS on polydisperse sample

- Oligomeric equilibrium
- Complex formation/dissociation
- Stoichiometric studies (excess of a component)
- Flexibility (conformational polydispersity)
- (With caution!): 3D modelling
Complex vs Mixture

• Equimolar ratio A and B

Correlated

\[ A_{a+b}(s) = A_a(s) + A_b(s) \]

\[ l_0 \sim M_a + M_b \]

Independent

\[ l_{a+b}(s) = c_a l_a(s) + c_b l_b(s) \]

\[ l_0 \sim (M_a^2 + M_b^2) / (M_a + M_b) \leq \text{MAX}(M_a, M_b) \]
Mixture

\[ \text{log}_{10} I(q) + q, \text{ nm}^{-1} \]
Scattering from mixtures

\[ I(s) = \sum_{k} v_k I_k(s) \]

For mixtures, solution scattering permits to determine the number of components and, given their scattering intensities \( I_k(s) \), also the volume fractions.

Oligomeric composition of mixtures

Monomer/dimer equilibrium of *Drosophila* kinesin

Ionic strength, concentration, pH, ligands, temperature

*J. Biol. Chem.* **276**, 1267
Complex Stoichiometry

SVD: Number of independent components

Mixture of monomers and dimers
SVDplot in “old” Primus

For Q=0.02 (P=0.98), Ncomp = 2

Non-parametric test for the number of components.

In this test shapes of left-side singular vectors are tested for randomness of sequencing positive and negative elements. The number of components equals the number of singular vectors which shape is non-random (non-noisy).

In this table:

- "No.": ordinal number of the singular vector. All the vectors are arranged correspondingly to singular values of the data matrix arranged in decreasing order.
- "criterion": testing statistics.
- "p": confidence level. In the corresponding columns two numbers (columnwise pairs) are left and right critical values.

If the criterion lies outside critical values then the shape of the vector is NONrandom.

<table>
<thead>
<tr>
<th>No.</th>
<th>criterion</th>
<th>P=0.95</th>
<th>P=0.99</th>
<th>P=0.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-18.92</td>
<td>-2.017</td>
<td>-2.386</td>
<td>-2.633</td>
</tr>
<tr>
<td>2</td>
<td>-12.11</td>
<td>-2.013</td>
<td>-2.379</td>
<td>-2.626</td>
</tr>
<tr>
<td>3</td>
<td>-1.185</td>
<td>-2.012</td>
<td>-2.379</td>
<td>-2.628</td>
</tr>
<tr>
<td>4</td>
<td>0.9633</td>
<td>-2.012</td>
<td>-2.379</td>
<td>-2.628</td>
</tr>
</tbody>
</table>

For Q=0.02 (P=0.98), Ncomp = 2

Press to continue

NCOMP = 2
Molecular Assembly of Lumazine Synthase

LS catalyzes the formation of 6,7-dimethyl-8-ribityllumazine in the penultimate step of riboflavin biosynthesis. Depending on the buffer it forms pentamers, dimers of pentamers and icosahedral capsids.

Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing

- Timely inhibition of Cdt1 by Geminin is essential to this DNA replication licensing.
- The mechanism of DNA licensing inhibition by Geminin, is analyzed by combining MX, SAXS and functional studies.
- The Cdt1:Geminin complex can exist in two distinct forms, a “permissive” heterotrimer and an “inhibitory” heterohexamer.

Endophilin-A1 BAR domain interaction with arachidonyl CoA

Endophilin-A1 belongs to the family of BAR domain containing proteins that catalyze membrane remodeling processes via sensing, inducing and stabilizing membrane curvature.

Extensive titration with the lipid micelles

Word of Caution: Glutamate Synthase Story

GltS is a complex iron-sulfur flavoprotein that catalyse the reductive transfer of L-glutamine (L-Gln) amide group to the C$_2$ carbon of 2-oxoglutarate (2-OG), yielding two molecules of L-glutamate (L-Glu)

αβ protomer: 160+50=210 kDa

β homology model

MW~900 kDa => tetramer

α Xtal dimer

Collaboration: M.A. Vanoni (Milano Univ.)
Glutamate synthase: EM model
Comparison of the “Old” and “New” data

\[ \text{MM}_{\text{Old}} = 900 \text{ kDa} \]
\[ \text{MM}_{\text{New}} = 1200 \text{ kDa} \]
Glutamate synthase: salt / substrate dependence

**OLIGOMER with EM model**

GlтS $\Rightarrow$ 90%$(\alpha \beta)_6$ +10% $\alpha \beta$

GlтS+NaCl $\Rightarrow$ 100% $\alpha \beta$

GlтS+2OG $\Rightarrow$ 90%$(\alpha \beta)_6$ +10% $\alpha \beta$

GlтS+L-MetS $\Rightarrow$ 90%$(\alpha \beta)_6$ +10% $\alpha \beta$

GlтS+L-MetS+2OG $\Rightarrow$ 90%$(\alpha \beta)_6$ +10% $\alpha \beta$

GlтS+NADP $\Rightarrow$ 70%$(\alpha \beta)_6$ +20%$(\alpha \beta)_2$ +10% $\alpha \beta$

GlтS+NADP+L-MetS+2OG $\Rightarrow$ 70%$(\alpha \beta)_6$ +20%$(\alpha \beta)_2$ +10% $\alpha \beta$
In case of a polydisperse sample SAS can be applied to characterize (in terms of volume fractions)

- Oligomeric equilibrium
- Conformation abundancies
- Complex stoichiometry
Flexible systems

Patterns of globular and flexible proteins

Kratky plot permits to detect disorder

Unfolded

Folded

Multi-domains with flexible linkers
Detection of Flexibility

SAXS curves

Analysis of the overall size descriptors ($R_g$, $p(r)$, Kratky)

Go for flexibility!

Modelling: *ab initio* (DAMMIN/DAMMIF) and Rigid body (BUNCH/CORAL)

Analysis of the differences
Detection of Flexibility

PolyUbiquitin Molecules

2, 3, 4 and 5 Ubiquitin (72 AA) domains connected by 20 AA linker (RanCH)

Flexible Multidomain Proteins present less features than rigid counterparts

Flexibility as a mix of different conformations

\[ I(s) = 4\pi \int_0^D p(r) \frac{\sin sr}{sr} dr \]

For monodisperse systems the scattering is proportional to that of a single particle averaged over all orientations

\[ I(s) = \sum_k \nu_k I_k(s) \]

\( \nu_k \) = volume fraction

\( I_k(s) \) = scattering intensity from the \( k \)-th component
Instead: find an ensemble with the same characteristics as the sample
Ensemble Optimization Method
Ensemble methods in SAXS

\[ I(s) = \frac{1}{N} \sum_{n=1}^{N} I_n(s) \]
Genetic Algorithm (optimized ensemble size)

Chromosome → Mutation → Crossing → Elitism

Generation 1

Generation 2

Elitism → Crossing → Mutation → Chromosome

\[ I(s) = \frac{1}{N} \sum_{n=1}^{N} I_n(s) \]
Ensemble Optimization Method

actual structures

curve

Optimized ensemble
EOM 2.0 can handle multimeric assemblies
(full length protein measured in two buffers with low and high ionic strength respectively)
IDP and unfolded
Assessing conformation variability: lysozyme unfolding

8M Urea, 10 mM DTT

$R_g = 26.3 \text{ Å}$

$R_g$ (native) $= 15.1 \text{ Å}$

8M Urea, 10 mM DTT

8M Urea, 100 mM DTT

$R_g = 30.0 \text{ Å}$
Simultaneous fitting of multiple curves
Multiple Curve Fitting with EOM: Reaching «Higher Resolution»

$M$ CONFORMATIONS (POOL)

Experimental data

$\text{C}^{\alpha}-\text{C}^{\alpha}$ Average Distance Matrix

Log $y$, Relative

Log $y$, Relative

Log $y$, Relative
Application to Tau Protein

Adult Tau

- N
- I1
- I2
- P1
- P2
- R1
- R2
- R3
- R4
- C

- P2
- R1
- R2
- R3
- R4

- P1
- P2
- R1
- R2
- R3
- R4

- R1
- R2
- R3
- R4

- K32
- K16
- K18

- ht40

Fetal Tau

- N
- P1
- P2
- R1
- R3
- R4
- C

- P2
- R1
- R3
- R4

- P2
- R1
- R3
- R4

- R1
- R3
- R4

- K27
- K17
- K19
- K44
- K10
- K25

- ht23

Mylonas et al. Biochemistry 2008, 47, 10345
Multidomain protein
Flexibility of Chitinase from Psychrophilic Bacterium *M. marina*

The flexibility of the hinge regions between the domains makes chitinases difficult to crystallize. In addition to the elongated state found in the crystal, the protein can adapt other conformations in solution ranging from fully extended to compact.

Multidomain & multimeric
**E. coli Flavorubredoxin**

Modular enzyme endowed with nitric oxide and/or oxygen reductase activity

**Collaboration:**
J. Vicente and P. Crowley (ITQB, Lisboa)
*E. coli* Flavorubredoxin: *ab initio* modelling
E. coli Flavorubredoxin: Xtal vs SAXS
E. coli Flavorubredoxin: various constraints

Tai Chi!
*E. coli* Flavorubredoxin: EOM
Multidomain nucleoprotein
Heterogeneous Assemblies with Flexibility

Tumor suppressor p53 and its complex with DNA

Cross-shaped extended
p53 from SAXS and NMR

Compact P53/DNA from SAXS also confirmed by EM

P53 is a transcription factor that regulates genes involved in cell cycle and apoptosis. Misfunction of p53 is related with cancer

The homotetrameric p53 consists of folded core and tetramerization domains, linked and flanked by intrinsically disordered segments

Flexible Systems Summary

- SAS can address protein flexibility, conformational changes and assembly/dissociation processes
- Ensemble Methods are appropriate tools to study (potentially) flexible molecules
- Unique structural information can be obtained based on distributions of descriptors whereas structures collected are simply a TOOL to describe the shape distributions
3D Reconstruction from Polydisperse Data?

**GASBOR_MX:** quaternary structure of weak (symmetric) oligomers

**SASREF_MX:** structural analysis of transient complexes
Ab initio Approach for Oligomeric Mixtures

Dummy residues model

GASBORMX: 
ab initio modelling from oligomeric equilibrium

Fit by linear combination
Rigid Body Modelling from Mixtures

Intact model

Dissociation products

SASREFMX:
rigid body modelling
of transient
assemblies

Fit
by linear
combination
3D Modelling from Equilibrium Mixtures

- Transient complexes and weak oligomers might be modelled (with caution!) against SAXS data
- Volume fractions of the entire assembly and of dissociation products are additional minimization parameters
- *Ab initio* analysis of weak symmetric homo-oligomers is done by dummy residues approach
- Quaternary structure of dissociating multisubunit assemblies (including nucleoproteins) is reconstructed by rigid body modelling
- Multiple scattering profiles (e.g. from concentration series) can be fitted simultaneously