SAXS Studies of Flexible Systems

Pau Bernadó

Centre de Biochimie Structurale – Montpellier

pau.bernado@cbs.cnrs.fr
Proteins are Dynamic Biomolecules

Concerted Motions
On/off mechanisms
Allostery

IDPs
Signalling and control
PTM
Recognition

ORDER

Protein vibrations
Catalysis
Partner Recognition

Flexible Multidomain
Proteins
Springs
Entropic Linkers
Fishing

DISORDER
Protein Motions: Conformational disorder and Time-Dependence

Conformational Disorder ↔ DYNAMICS ↔ Time-dependence

- vibrational motions
- overall tumbling
- enzyme catalysis
- fast loop motions
- slow loop motions
- domain motions
  - side chain rotation/reorientation
  - aromatic ring flips
- protein folding
In a flexible Protein a SAXS curve is the average of all conformations coexisting in solution.
Kratky and Porod: Useful Tools to Detect Disorder

Using Gaussian Chain Models, Kratky and Porod established approximate relationships between $I(s)$ vs $s$ for folded and unfolded proteins:

- $I(s) \sim 1/s$
- $I(s) \sim 1/s^4$

Scattering of a Rod

Porod representation ($s^4 I(s)$ vs $s^4$) displays a plateau for rigid particles.

Rambo & Tainer Biopolymers 2011
Two Important Concepts

Reliable and meaningful dynamic information of biomolecules by SAXS can only be derived if can model it structurally.

Assumption of a dynamic system implies the use of the concept of ‘ensemble of conformations’ to properly describe the data (SAXS and others).
Outline of the Presentation

- SAXS as a tool to validate structural models with flexibility
- Detection of flexibility in SAXS data
- Ensemble Methods: Theory and Examples
Outline of the Presentation

► SAXS as a tool to validate structural models with flexibility

► Detection of flexibility in SAXS data

► Ensemble Methods: Theory and Examples
Validation of Structural Models with Flexibility

► SAXS is used in combination with molecular dynamics simulations or other theoretical approaches

► Therefore is a *validation method* for a dynamic model that either works or does not work

► Snapshots are collected, their individual SAXS curve are computed with CRYSOL (or others), averaged and compared with the experimental one

► In relatively rigid systems, the effects of moderate dynamics can be ‘compensated’ with small structural perturbation or changes in the hydration
Flexible Regions in Proteins

Tiago Cordeiro

Col. William Bourguet (CBS)
MD Simulations

PROBLEMS:
Limited conformational sampling in MD
Solvent treatment can compensate flexibility

Structural Models for Intrinsically Disordered Proteins

► IDPs lack stable tertiary and/or secondary structure
► IDPs are more common in eukaryotes than in bacteria and archaea: Probably linked with their major biological complexity… Up to ~30-50% of genome
► Very specific amino acid sequences rich in P G R K and E, and depleted on hydrophobic residues

Both Ordered and Disordered regions are associated with distinct functions
Disordered Proteins complement the functions of ordered protein regions


<table>
<thead>
<tr>
<th>ORDER</th>
<th>DISORDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMP biosynthesis</td>
<td>Differentiation</td>
</tr>
<tr>
<td>Amino-acid biosynthesis</td>
<td>Transcription</td>
</tr>
<tr>
<td>Transport</td>
<td>Transcription regulation</td>
</tr>
<tr>
<td>Electron transport</td>
<td>Spermatogenesis</td>
</tr>
<tr>
<td>Lipid A biosynthesis</td>
<td>DNA condensation</td>
</tr>
<tr>
<td>Aromatic hydrocarbons catabolism</td>
<td>Cell cycle</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>mRNA processing</td>
</tr>
<tr>
<td>Purine biosynthesis</td>
<td>mRNA splicing</td>
</tr>
<tr>
<td>Pyrimidine biosynthesis</td>
<td>Mitosis</td>
</tr>
<tr>
<td>Carbohydrate metabolism</td>
<td>Apoptosis</td>
</tr>
<tr>
<td>Branched-chain amino acid biosynthesis</td>
<td>Protein transport</td>
</tr>
<tr>
<td>Lipopolysaccharide biosynthesis</td>
<td>Mitosis</td>
</tr>
<tr>
<td>Sugar transport</td>
<td>Cell division</td>
</tr>
<tr>
<td>Antibiotic resistance</td>
<td>Ub conjugation pathway</td>
</tr>
<tr>
<td>Lipid synthesis</td>
<td>Wnt signaling pathway</td>
</tr>
<tr>
<td>Tricarboxylic acid cycle</td>
<td>Neurogenesis</td>
</tr>
<tr>
<td>Arginine biosynthesis</td>
<td>Chromosome partition</td>
</tr>
<tr>
<td>Ion transport</td>
<td>Ribosome biogenesis</td>
</tr>
<tr>
<td>Rhamnose metabolism</td>
<td>Chondrogenesis</td>
</tr>
<tr>
<td>Peptidoglycan synthesis</td>
<td>Growth regulation</td>
</tr>
</tbody>
</table>
The **Flexible-Meccano Approach**

Randomly chosen residue-specific $\phi/\psi$ and a very simple steric potential are used to define a single conformation.

Bernadó et al. *PNAS* 2005, 102, 17002
Ozenne et al. *Bioinformatics* 2012, 28, 1463
The ensemble is in agreement with the SAXS curve.

The size and shape properties of PX are well reproduced.

The Flexible-Meccano approach is a good model to describe IDPs as it reproduces the overall (SAXS) and the local (NMR) structural properties.

Bernadó et al. PNAS 2005, 102, 17002
Model Validation of IDPs

Wells et al. PNAS 2008, 105, 5762.

Di Biasio et al. BJ 2014, 106, 865

Pérez et al. Unpublished

Bernadó et al. PNAS 2005, 102, 17002.

Mukrash et al. JACS 2007, 129, 5235

Mylonas et al. Biochemistry 2008, 47, 10345
Reparametrizing Flory’s Equation for IDPs

\[ R_g = R_0 \cdot N^\nu \]

Denatured Proteins
\[ \nu = 0.59 \]
\[ R_0 = 1.93 \]

IDPs
\[ \nu = 0.55 \]
\[ R_0 = 2.54 \]

Departures from the predicted values (random coil model) indicate the presence of structural features

Bernadó & Blackledge BJ 2009
Bernadó & Svergun MolBioSyst 2011
Outline of the Presentation

► SAXS as a tool to validate structural models with flexibility

► Detection of flexibility in SAXS data

► Ensemble Methods: Some Examples
Detection of Flexibility: A Crucial Issue

Artefacts exerted by Flexibility into Structural Models obtained from SAXS data

Calculation of experimental-like SAXS curves

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

Modelling: ab initio (DAMMIN) and Rigid body (BUNCH)

Analysis of the differences
Detection of Flexibility: A Crucial Issue

PolyUbiquitin Molecules

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

Flexible Multidomain Proteins present less features in the SAXS curve than their rigid counterparts

Detection of Flexibility: A Crucial Issue

Flexible Proteins have large $D_{\text{max}}$ values and smooth ending to $p(D_{\text{max}})=0$

Only two peaks in the $p(r)$, indicating distal correlation between folded domains, appear in the flexible scenario

Additional peaks are only present in the rigid scenario

Modelling the Flexible Scenario with Single Conformation Strategies

Good fits are obtained in both ab initio and rigid body modelling.

No structural variation is observed between solutions.

Homogeneous densities are observed in DAMMIN solutions. There is a systematic decrease of resolution.

Domains appear isolated, no interdomain contacts are observed in BUNCH solutions.
Indications (not Proofs!!!) of Flexibility

► Smooth Scattering profiles and featureless Kratky Plots

► Large $R_g$ and $D_{max}$

► Absence of correlation peaks in the $p(r)$ function

► Low correlation densities in ab initio reconstructions

► Isolated domains in rigid body modelling

► No Plateau in Porod’s representation

► Lower excluded volume ($V_r$) than a globular protein (1.35 -1.37 g/cm$^3$)

► Excellent fitting of the data using rigid-body or ab initio methods

► Equivalent solutions in different runs
Outline of the Presentation

► SAXS as a tool to validate structural models with flexibility

► Detection of flexibility in SAXS data

► Ensemble Methods: Some Examples
- The Ensemble Optimization Method (EOM)


- Minimal Ensemble Search (MES)

Pelikan, Hura, and Hammel. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys* 2009, **28**:174–189.

- Basis-Set Supported SAXS (BSS-SAXS)


- Ensemble Refinement of SAXS (EROS)

Calculation of a conformational ensemble

Computation of the theoretical SAXS profiles

Optimization of a subensemble

Structural interpretation of the subensemble
- The Ensemble Optimization Method (EOM)


- Minimal Ensemble Search (MES)


- Basis-Set Supported SAXS (BSS-SAXS)


- Ensemble Refinement of SAXS (EROS)

Bernadó & Svergun *MolBioSyst* 2011
Unfolded States of Lysozyme

8M Urea, 10 mM DTT

$R_g = 26.3 \text{ Å}$

$R_g$ (native) = 15.1 Å

8M Urea, 10 mM DTT

8M Urea, 100 mM DTT

$R_g = 30.0 \text{ Å}$

Bernadó et al. JACS 2007, 129, 5656
FAQ about EOM for Flexible Proteins

Structural Meaning of the Selected Conformations

None Unstructured system can be described with only 50 conformations… But SAXS data of a disordered state can be. SAXS is a low resolution technique!!!!

You can get more or less the same results with smaller subensembles… These ensembles are just representations of reality

It is tempting to look at the structures at atomic/residue level… Don’t do that because (Remember that) SAXS is a low resolution technique and the information content is limited

Structures collected are simply a TOOL to describe the shape distributions...

If certain structure is collected at each run… It does not necessarily mean that it is prevalent in solution
Multidomain Proteins: Detection of Interdomain Dynamics

The $R_g$ distribution in the optimized ensemble is diagnostic of interdomain motion.
FA Critizism about EOM

“With EOM you can fit an elephant if you want…”

Maria Garcia-Parajo

THIS IS NOT TRUE
EOM analysis of L12 Protein SAXS data

Three samples at 7, 15 and 20 mg/ml were measured and merged.

Data measured at X33 Beamline in DESY-Hamburg.

Although highly dynamic, L12 samples a reduced conformational space, and mainly adopts highly elongated (anisotropic) structures.

This induces large distances between both CTD domains that leave the NTD dimer in the middle.

Bernadó et al. Biophys. J. 2010, 98, 2374
Structural Analysis of the Optimized Ensemble

The linker is partially structured
Conformational Analysis of IgG Antibodies


IgG1  EPKSCDKTHTCPPCPAPELLGG
IgG2  ERK---CCVECPPCPAPPVA-G
IgG4  ESK---YGPPCPPCPAPEFLGG
Conformational Analysis of IgG Antibodies

IgG1: EPKSCDKTHTCPPCPAPELLGG
IgG2: ERK---CCVECPPCPAPPVA-G
IgG4: ESK---YGPPCPPCPAPEFLGG

Each monoclonal antibody display a distinct conformational behavior being the degree of flexibility
IgG1 > IgG4 > IgG2

The D_{max} distributions derived by the EOM analysis display a shift towards more extended conformations

Indirect information about the conformational behavior of the linker regions is obtained
Shape Clustering of the EOM results

IgG1

Cluster 9: 35.7%
Cluster 8: 16.5%
Cluster 10: 10.0%
Cluster 1: 8.3%
Cluster 4: 8.3%
Cluster 2: 7.0%
Cluster 5: 4.8%
Cluster 6: 4.3%
Cluster 3: 3.5%
Cluster 7: 1.7%

T-Shape
Y-Shape
Y/T-Shape

IgG2

Cluster 4: 53.4%
Cluster 1: 26.9%
Cluster 3: 13.4%
Cluster 5: 6.0%

IgG4

Cluster 7: 52.9%
Cluster 6: 32.2%
Cluster 5: 9.4%
Cluster 8: 3.3%
Conformational Preferences in IgG2

Isoform A (53%)

Isoform A/B₁ (6%)

Isoform B (13%)

Isoform A/B₂ (27%)

Presence of two additional Cysteines can induce multiple disulphide bridges that will modify the shapes

Resulting clusters are coherent with present knowledge of disulphide bridge pairing

Dillon et al., JBC 283: 16206-16215 (2008)
Structure of Transilin: A Nucleic Acid Binding Protein


Tian et al. NSMB 2011

Perez-Cano et al. Nucl Acid Res. 2013
Flexible N- and C- terminal flexible tails do not explain the disagreement with experimental data
4237 MODELS calculated with Normal Modes
Use of Minimal Ensemble Search (MES)

Pelikan et al. Gen Physiol Biophys 2009
Conformational Equilibrium in Translin

SAXS data of Translin can only be explained if a large population (≈ 80%) of open conformations are open.
The open conformation explains the interaction of Transilin with Nucleic Acids

RNA can enter inside Translin cavity

SAXS provides in a unique manner insights into the biologically relevant form
Translin: ssRNA Complex

- Similar Topology
- Same $D_{\text{max}}$
- Slightly Larger $R_g$ for Free translin
- Translin: ssRNA is more compact

Simulated @ scattering.tripod.com
Modelling Translin:RNA Complex

Flexibility of the Target !!!

Flexibility of the ssRNA !!!!
Translin:ssRNA Complex seen by SAXS

The same equilibrium open-close is observed in the complex.

Multiple conformations of ssRNA are needed for the fit.

RNA does not reduce the flexibility in translin.

Translin:ssRNA complex is a highly plastic assembly where the two partners display flexibility.

This could explain the putative role of translin in NA transport.
Conclusions

- Protein Flexibility is not a problem for SAS…

- In fact it is the most powerful technique to study large amplitude motions and one can reach novel and biologically relevant information.

- Discerning between flexible and rigid scenarios is fundamental.

- Ensemble Methods (EOM, MES, BSS-SAXS) are appropriate tools to study (potentially) flexible molecules.

- Information is based on distributions of structural parameters… a different (and less intuitive) way to describe structures.

- SAXS becomes more powerful in combination with other structural biology methods and/or advanced modeling tools SAXS.
Acknowledgements

Intrinsically Disordered Proteins
Martin Blackledge (IBS-Grenoble)

EOM
Maxim Pethoukhov (EMBL-HH)
Efstratios Mylonas (EMBL-HH)
Dmitri Svergun (EMBL-HH)

Transilin
Laura Pérez (BSC-Barcelona)
Juan Fernández-Recio (BSC-Barcelona)
Fabián Glaser (IIT-Haifa)
Haim Manor (IIT-Haifa)

GalNac-T2
Ramon Hurtado (BIFI-Zaragoza)
Erandi Lira (BIFI-Zaragoza)

Xinsheng Tian & Bente Vestergaard for sharing slides