Combined Use of NMR and SAS for Flexible Proteins

Pau Bernadó

Centre de Biochimie Structurale - Montpellier

pau.bernado@cbs.cnrs.fr
Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS

Nathalie Sibille and Pau Bernadó†
Centre de Biochimie Structurale, INSERM U1054, CNRS UMR 5048, Université Montpellier 1 and 2, 29 rue de Navacelles, 34090 Montpellier, France


Molecular BioSystems
Cite this: Mol. BioSyst., 2012, 8, 151–167
www.rsc.org/molecularbiosystems

Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering†

Pau Bernadó* and Dmitri I. Svergun*†

Received 4th July 2011, Accepted 2nd September 2011
DOI: 10.1039/c1mb05275f
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

The Importance of Being Unfolded

- Increased interaction surface area
- Accessible to post-translational modifications
- Conformational flexibility to interact with several targets
- Fine tuning of the thermodynamic properties of the complex

ID Linkers restricts the space sampled by folded domains in multidomain proteins
Small Angle X-ray Scattering Experiments

Sample

\[ I(s) \text{, a.u.} \]

\[ s (\text{A}^{-1}) \]

**Guinier**

**Kratky**

\[ R_g \text{, Molecular Weight} \]

\[ \text{Mass Density} \]
Small Angle X-Ray Scattering Resolution

Resolution (Å) = \( \frac{2\pi}{s_{\text{max}}} \)

Size and Shape
Fold
Atomic structure
Traditional Use of SAXS in Structural Biology

Nothing known about the structure:  
*ab initio* low resolution structure

Complete high resolution structure  
Compatibility of 3D models with SAXS data

Structure of domains or complexes:  
Rigid body refinement
In flexible proteins the measured SAXS curve is the weighted average of all conformations present in solution.
Nuclear Magnetic Resonance

Atomic Resolution Technique
Each active nucleus can be identified
Isotopic labelling is necessary
NMR a versatile Source of Structural Information

Local Conformations
Chemical Shifts, J-Couplings

Interatomic Distances
$^1$H-$^1$H NOEs, PREs, Pseudocontact Shifts

Orientational Restrictions
RDCs, Relaxation ($R_2/R_1$), Pseudocontact Shifts

Overall Size
Diffusion Coefficients, Relaxation (Rotational Diffusion Tensor), RDCs (Alignment Tensor)

Time-Scale of Motions
Relaxation ($R_1, R_2, \text{Het-NOE}$), Relaxation Dispersion, $R_1\rho$, PREs
NMR Residual Dipolar Couplings in Partially Aligned Systems

\[ D_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{8 \pi^3} \left\langle \frac{P_2(\cos \theta(t))}{r_{ij}^3} \right\rangle \]

In solution, where molecules tumble isotropically dipolar interaction averages to zero.
Residual Dipolar Couplings in Partially Aligned Systems

Residual Dipolar Couplings by NMR

\[
D_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{8 \pi^3} \left\langle P_2(\cos \theta(t)) \right\rangle \frac{r_{ij}^3}{r_{ij}^3}
\]

\[
D_{ij} = -S \frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij}^3} \left( A_a (3\cos^2 \theta - 1) + \frac{3}{2} A_r \sin^2 \theta \cos 2\phi \right)
\]

Residual dipolar couplings depend on the orientation of internuclear vectors relative to the alignment frame.

RDCs are a valuable source of information to study structure and dynamics of biomolecules.
Paramagnetic Relaxation Enhancement (PRE)

Free electron Radical attached to a Cys mutant

Residues in the proximity experience an enhancement of their $R_2$ relaxation

Effect up to $r < 20 \, \text{Å}$

$r^6$ dependence

Although it is a relaxation phenomena, in ensembles properties are normally averaged
Highly Flexible Proteins do not have a permanent structure. Shape changes. Different interdomain motions are probably active at the same time-scale.

Protein Dynamics from Relaxation Rates

Hydrodynamic Properties
Rotational Diffusion Tensor (ns)

S², τₑ (ps)
Complementarity between SAXS and NMR in Structural Biology

**NMR**

- Close Contacts: $^1$H-$^1$H nOes
- LR relationship: RDCs, $R_2/R_1$

**SAXS**

- Low Resolution Overall Shape

**Globular Proteins**

- Interfaces: $^1$H-$^1$H nOes, CS, PRE
- Orientation: RDCs, $R_2/R_1$

**Complexes and Rigid Multi-Domain Proteins**

**Unstructured Proteins**

- Conform. Sampling: CS, RDCs
- LR Contacts: PREs

**Flexible Multi-Domain Proteins**

- Dimensions of the Ensemble

- Time-Scale Information through Spin Relaxation

- Volume Sampled by the Domains
Needs for Structural characterization of Disordered States

Structural Models for Disordered Proteins

Calculation of NMR and SAXS Properties from Individual Conformations
Capacity to address complex biological systems

Structural models with better resolution

More complete models embedding structure and dynamics
Flexible-Meccano Model for Unfolded Proteins

500 X-ray single chain structures
Resolution < 1.8 Å and B < 30
Presenting low structural/sequence homology among them


Only residues non placed in secondary structure elements were used

X-Pro is treated as the 21st Residue type
Symmetrization of Glycines

Bernadó et al. PNAS 2005, 102, 17002
Ozenne et al. Bioinformatics 2012, 28, 463
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, 463
Conformational Sampling of in Disordered States

Properties Conformation 1

Bernadó et al. PNAS 2005, 102, 17002
Ozenne et al. Bioinformatics 2012, 28, 463
Conformational Sampling of in Disordered States

Bernadó et al. *PNAS* 2005, 102, 17002
Ozenne et al. *Bioinformatics* 2012, 28, 463
Conformational Sampling of in Disordered States

Bernadó et al. *PNAS* 2005, 102, 17002
Ozenne et al. *Bioinformatics* 2012, 28, 463
Calculation of Properties from Conformations

**SAXS:**
- CRYSOL, AXES, FOXS, AquaSAXS

**NMR:**
- CS: Sparta, Sparta-2, ShiftX, CamShift
- RDCs: PALES, Flexible-Meccano
- PREs: Flexible-Meccano
- Hydrodynamics: HydroPro
Calculation of Properties from Conformations

SAXS:
CRYSOL, AXES, FOXS, AquaSAXS

NMR:
CS: Sparta, Sparta-2, ShiftX, CamShift
RDCs: PALES, Flexible-Meccano
PREs: Flexible-Meccano
Hydrodynamics: HydroPro

Be careful !!!
Not all properties can be averaged
Not all properties require the same number of conformations
Two Philosophies to Characterize Residual Structure

► Definition of the structural rules for building ensembles

Definition of the nature of Random Coil
Partially structured regions (class, location and population)
Identification of transient long-range contacts (location and population)

Trial and error strategy difficult to automatize

► Data-driven ensembles

Search for a subensemble that properly describes all data measured
Analysis of the subensemble searching for structural features

Limited to the structural content of the data
Two Philosophies to Characterize Residual Structure

► Definition of the structural rules for building ensembles

- Definition of the nature of Random Coil
- Partially structured regions (class, location and population)
- Identification of transient long-range contacts (location and population)

Trial and error strategy difficult to automatize

► Data-driven ensembles

- Search for a subensemble that properly describes all data measured
- Analysis of the subensemble searching for structural features

Limited to the structural content of the data
Conformational Sampling of in Disordered States

Averaging of RDCs and SAXS curves for the ensemble of conformations

RDCs by NMR

2,000

100,000

\[ \langle I(q) \rangle \]

\[ \langle \text{RDC} \rangle \]
Modelling Alignment Properties of PX

The Model accounts for the main conformational properties probed by RDCs and the overall features as demonstrated by SAXS.

Interpreting RDCs with *FM*: Local Structural Information

Experimental FM Simulation

Random Coil Model

Tiago Cordeiro - Col. Francisco Blanco (Biogune, Bilbao)
RDCs in Disordered States

Disordered and Extended Regions

\[ P_2(\cos \theta) \]

Negative RDCs
Interpreting RDCs with FM: Local Structural Information

Experimental FM Simulation

Random Coil Model

Extended

α-Helical

Extended

Refined Model

Tiago Cordeiro - Col. Francisco Blanco (Biogune, Bilbao)
Interpreting RDCs with \textit{FM}: Local Structural Information

\begin{itemize}
  \item SV PX
  \item Unique SRC
  \item \(\alpha\)-Synuclein
  \item SV Nucleoprotein
  \item p53 (TAD)
  \item Tau (K18)
\end{itemize}

- Bernadó \textit{PNAS} 2005, 102, 17002.
- Wells \textit{PNAS} 2008, 105, 5762.
- Pérez \textit{JMB} 2009, 391, 136.
- Jensen \textit{JACS} 2008, 130, 8055.
- Pérez \textit{JACS} 2007, 129, 5235.
Overall Properties of IDPs with SAXS

SAXS can not validate NMR-refined ensembles as transient structuration is not always reflected in the curve… They can not be distinguished from Random coil

Pérez et al. Unpublished

Mukrash et al. 2007, 129, 5235
Mylonas et al. 2008, 47, 10345
The Conformational Nature of the Denatured State

The size of denatured proteins has been traditionally studied with SAXS and Hydrodynamic Measurements. Flory’s relationship correlates the radii with the length of the chain

\[ \frac{R_g}{r_h} = R_0 \cdot N^\nu \]

\[ \log \frac{R_g}{r_h} = k + \nu \cdot \log N \]

- \( R_0 \) Persistence Length
- \( \nu \) Solvent ‘quality’

Several experimental (\( R_g \) and \( r_h \)) and theoretical studies establish \( \nu \approx 0.6 \) as an indication of the ‘random coil’ in chemically denatured (Urea or GuHCl) proteins.


Wilkins et al. *Biochemistry*, 1999, 38, 16424
Reparametrizing Flory’s Equation for IDPs

$$R_g = R_0 \cdot N^\nu$$

<table>
<thead>
<tr>
<th>Proteins</th>
<th>$\nu$</th>
<th>$R_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denatured Proteins</td>
<td>0.59</td>
<td>1.93</td>
</tr>
<tr>
<td>IDPs</td>
<td>0.55</td>
<td>2.54</td>
</tr>
</tbody>
</table>

Generality of $FM$ allows for the derivation of universal properties for IDPs

Bernadó & Blackledge BJ 2009
Bernadó & Svergun MolBioSyst 2011
Reparametrizing Flory’s Equation for IDPs

N-tail Measles virus

EXTENDED

COMPACT

K18 Tau

IB5

α-Syn

RESUIN

II-1

MeCP276-305

N-Tail MV
Two Philosophies to Characterize Residual Structure

Definition of the structural rules for building ensembles

- Definition of the nature of Random Coil
- Partially structured regions (class, location and population)
- Identification of transient long-range contacts (location and population)

Trial and error strategy difficult to automatize

Data-driven ensembles

- Search for a subensemble that properly describes all data measured
- Analysis of the subensemble searching for structural features

Limited to the structural content of the data
Ensemble Methods in SAXS

- 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)
Elitism

Crossing

Mutations

Experimental Curve

\begin{align*}
\text{Experimental Curve} &= 1 - 1.5 s (\text{A}^{-1}) \\
\text{Experimental Curve} &= \frac{1}{1 + 1.5 s (\text{A}^{-1})}
\end{align*}
Elitism

Experimental Curve

Mutations

Crossing

G Generations

R1 R2 R3

Rg Distribution

Bernadó et al. JACS 2007, 129, 5656
N-Tail of VS Virus Phosphoprotein, an IDP

A Bimodal distribution is obtained when using EOM
Bimodality is perturbed (as expected) by modifying the buffer

NMR detects interaction between the tail and the two HMG-Boxes. No information on the relative population of both conformations.

EOM shows that the compact conformation is predominant. Deletion of the tail breaks the interaction and the protein behaves as a Random Coil.

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
Relaxation Study of the Ribosomal L12 Protein

The two domains have different correlation times. Are they completely independent?

The linker region is highly flexible, but behaves asymmetrically.

No Structural model can be derived from relaxation data.

Mulder et al. Biochemistry, 2004, 43, 5930
Bocharov et al. J.Biol.Chem. 2004, 279, 17697
Highly Flexible Proteins do not have a permanent structure. Shape changes
Different interdomain motions are probably active at the same time-scale

S², τe (ps)

Internal Dynamics

Protein Dynamics from Relaxation Rates

Hydrodynamic Properties
Rotational Diffusion Tensor (ns)
The iRED Approach

The isotropic Reorientational Eigenmode Dynamics offer an alternative data analysis

A Spherical Harmonic Normal Mode Analysis is performed from a MD trajectory or Ensemble

\[ M_{kl} = \langle P_2(\cos(\Omega_k - \Omega_l)) \rangle \] Average over all members of the ensemble

Eigenvectors represent the essential motional modes, \( |m\rangle_j \)

N-H vectors (residues) contribute differently to these eigenmodes

\[ \delta S^2 = \lambda_m \cdot | |m\rangle_j |^2 \]

The NMR relaxation data are fitted to the motional modes and their time-scale are obtained

\[ J_j(\omega) = \sum_m \delta S^2_{j,m} \frac{\tau_m}{1 + \omega^2 \tau_m^2} \]

Prompers & Brüschweiler JACS 2001, 123, 7305
Prompers & Brüschweiler JACS 2002, 124, 4522
L12 Eigenmodes

Pool Eigenmodes

EOM Eigenmodes

Three Rigid Bodies
15 Eigenmodes

POOL:
10 non-degenerate Modes
Domain PURE

EOM:
12 non-degenerate Modes
Modes with mixed components

Bernadó et al. Biophys J 2010 98, 2374
Time-scale of the Interdomain Motions

We have characterized the conformational plasticity and the time-scale of the motions for L12

This is a complete picture of a highly flexible protein that includes structure and dynamics

Bernadó et al. *Biophys J* 2010 98, 2374
Conformational Space of Flexible Biological Macromolecules from Average Data

Ivano Bertini,*†‡ Andrea Giachetti,† Claudio Luchinat,†‡ Giacomo Parigi,†‡
Maxim V. Petoukhov,§ Roberta Pierattelli,†‡ Enrico Ravera,†‡ and
Dmitri I. Svergun§

CERM, University of Florence, Via L. Sacconi 6, and Department of Chemistry, University of
Florence, Via della Lastraecia 3, 50019 Sesto Fiorentino, Italy. EMBL, Hamburg Outstation,
Notkestrasse 85, D-22603 Hamburg, Germany, and Institute of Crystallography, Russian
Academy of Sciences, Lentinsky pr. 59, 117333 Moscow, Russia

Received July 19, 2010; E-mail: ivanobertini@cerm.unifi.it

DOI 10.1007/s10858-012-9638-1

MaxOcc: a web portal for maximum occurrence analysis

Ivano Bertini · Lucio Ferella · Claudio Luchinat ·
Giacomo Parigi · Maxim V. Petoukhov ·
Enrico Ravera · Antonio Rosato · Dmitri I. Svergun

Synergistic application of SAXS with NMR data (RDCs and pseudocontact shifts) to characterize the conformational space sampled by a flexible multidomain protein
The Concept of Maximum Occurrence

Maximum percent of time that a flexible protein can spend in a given conformation.

Creation of a large ensemble of conformations sampling the complete available space

Capacity to describe experimental data of an ensemble when increasing the population of a given conformation
Conformational Sampling in Calmodulin

SAXS + RDCs with Tb$^{3+}$, Tm$^{3+}$, and Dy$^{3+}$ + PCSs with Tb$^{3+}$, Tm$^{3+}$, and Dy$^{3+}$

Blue < 5%
Red > 40%
Differential Sensitivity of Data

Despite the overall picture is independent of the data used, SAXS is the less sensitive data to restrain the conformational sampling in Calmodulin.
Conclusions

- In highly flexible proteins, NMR provides the conformational sampling at residue level. SAXS provides the overall size and shape.

- Synergistic application of NMR and SAXS (embedded in computational tools) provides more accurate structural/dynamic models for flexible proteins.

- SAXS provides information about large-amplitude motions in biomolecules and reaches novel and biologically relevant information.

- Progress in the structural interpretation of SAXS data (in terms of dynamics) will come from the development of theoretical methods with the capacity to perturb 3D structures… and also a deeper comprehension of the information content of SAXS data.
Acknowledgements

Flexible Meccano
Martin Blackledge (IBS-Grenoble)

Ensemble Optimization Metod (EOM)
Dmitri Svergun (EMBL-Hamburg)
Efstratios Mylonas (EMBL-Hamburg)
Maxim Petoukhov (EMBL-Hamburg)

L12
Mikael Akke (Lund)
Kristofer Modig (Lund)
Miquel Pons (IRB-Barcelona)
Marek Tchorzewski (Lublin)
Przemek Grela (Lublin)
Dmitri Svergun (EMBL-HH)