Exploring multiple structural states and protein interactions in solution with Small Angle X-ray Scattering (SAXS) profiles

Dina Schneidman
Silberman Institute of Life Sciences

THE HEBREW UNIVERSITY OF JERUSALEM
Small Angle X-Ray Scattering

Protein in solution

Monochromatic beam

X-ray: $\lambda=0.1-0.2$ nm

Large distance

$2\theta$

Detector

$\log(\text{intensity})$

SAXS profile

- Rapid data collection (**minutes**)!
- No crystallization required
- Wide range of targets (sizes) is suitable for SAXS
- Measures can be made for a variety of conditions
Reciprocal vs. real space

\( I(q) \) - intensity as a function of distance from the center of the detector

\( P(r) \) - distribution of distances between all the electrons

Fourier transform
Docking with SAXS profile of the complex

**Debye formula**

\[ I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(qd_{ij})}{qd_{ij}} \]

**Schneidman-Duhovny D, Hammel M, Sali A. NAR 2010**

**Schneidman-Duhovny D, Hammel M, Sali A. J Struct Biol. 2011**

**Schneidman-Duhovny D, Hammel M, Tainer J, Sali A. Biophys J 2013**
Uncertainty in Integrative Structure Modeling

Information → Scoring → Sampling → Analysis

Data satisfied?

- yes → Single structure
- no → Multiple structures

more data is needed

Schneidman-Duhovny, Pellarin, Sali. COSB 2014
DNA Ligase III with and without DNA

Essential functions in nuclear and mitochondrial DNA replication and repair

Cotner-Gohara et al. Biochemistry 2010
Dynamics Comes in Flavors and it is Common

“rigid”  
“flexible”  
“disordered”

short disordered fragments (≥10 and <30 residues)
long disordered fragments (> 40 residues)

PDB  
~ 40% of structures  
~ 10% of structures

SwissProt  
> 25% of sequences

(Romero et al. 2001; Dunker et al 2000; Le Gall 2007)
SAXS data can be easily collected for proteins that include disordered regions.

Data interpretation is challenging.
Heterogeneous Sample Requires Multi-State Model

**Heterogeneous sample**
compositional or conformational heterogeneity in the sample used to generate the data

**Multi-state model**
a model that specifies two or more co-existing structural states and values for any other parameter

**Ensemble of models**
an ensemble of (good scoring) single or multi-state models
Algorithm for multi-state modeling with SAXS

### Information
- Standards for data collection and validation
- Sparse data
- Heterogeneous sample

### Scoring
- Debye formula
- Heterogeneity model
- \[ \chi = \frac{1}{S} \sum_{i=1}^{S} \left( \frac{I_{\exp}(q_i) - c \sum_{n} w_n I_n(q_i, c_1, c_2)}{\sigma(q_i)} \right)^2 \]

### Sampling
- Conformational sampling with Rapidly exploring Random Trees (RRTs)
- Enumeration of multi-state models that fit the data within noise

### Analysis
- Quality of fit to data, variance among top scoring models

---

Berlin, Castaneda, Schneidman-Duhovny, Sali, Nava-Tudela, Fushman *JACS* 2013
1. Conformational sampling

Proteins and robots have similar degrees of freedom

We rely on methods for Motion Planning developed in Robotics (La Valle, Latombe, Kavraki, Cortes)
Mapping collision free space with Rapidly exploring Random Tree (RRT)

Collision free space for robot

Collision free space for protein chain
2. Enumeration of multi-state models that fit the data within noise

Enumeration

**branch & bound** deterministic algorithm

Multi-state models of size $i+1$ are generated by extending the best $K$ (=10000) multi-state models of size $i$.

- **best K multi-state models of size 1:**
- **best K multi-state models of size 2:**
- **best K multi-state models of size 3:**

...
weights optimization is needed for each set of structural states

Non-negative least square fitting (NNLS, Lawson & Hanson 1974)
DNA Ligase III with and without DNA

Essential functions in nuclear and mitochondrial DNA replication and repair

Cotner-Gohara et al. Biochemistry 2010
Conformational and compositional heterogeneity

DNA ligase III with and without DNA

<table>
<thead>
<tr>
<th></th>
<th>DNA</th>
<th>No DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>45±1%</td>
<td>26±5%</td>
<td>26±2%</td>
</tr>
<tr>
<td>50±2%</td>
<td>40±8%</td>
<td>10±4%</td>
</tr>
</tbody>
</table>
The power of integrative structure modeling

Use all the available information to optimize the accuracy, precision, and resolution of the structural models.

Construct single-state and multi-state models of large and dynamic macromolecular complexes.

Infer functional mechanism from the models.

Information → Scoring → Sampling → Analysis

- Protein interactions
- Protein dynamics

- X-ray crystallography
- NMR spectroscopy
- 2D electron microscopy
- 3D electron microscopy
- Small Angle X-ray Scattering
- Shape
- Statistical potential
- Crosslinking
- Residue Type Content from NMR spectroscopy

Use all the available information to optimize the accuracy, precision, and resolution of the structural models.

Construct single-state and multi-state models of large and dynamic macromolecular complexes.

Infer functional mechanism from the models.
Acknowledgements

Haim Wolfson
Ruth Nussinov
Yuval Inbar
Oranit Dror
Max Shatsky
Alex Shulman

Andrej Sali
Seung Joong Kim
Riccardo Pellarin
Pat Weinkam
Barak Raveh
Yannick Spill
Guangqiang Dong
Natalia Khuri
Daniel Russel
Elina Tjioe
Ben Web
Ursulla Pieper

Michal Hammel
Greg Hura
Rob Rambo
Susan Tsutakawa
John Tainer

Lester Carter
Tsutomu Matsui
Thomas Weiss

CAPRI organizers

Looking for students!