Probing Protein Ensembles with X-ray Solution Scattering

Lee Makowski
Northeastern University
Crystallography provides a ‘snap shot’ of one structure of the ensemble…

*It might be highly representative or not – it might be highly abundant – or not*

It represents the basis for structural and functional hypotheses

*We are developing WAXS/MD as an experimental approach to testing those hypotheses*
• **So we study them in solution:**

• **WAXS (wide angle x-ray solution scattering)**

  • What can you gain by collecting to the highest possible resolution?
  • *Just like SAXS (small-angle x-ray solution scattering)*

  • Cannot use waxes to directly calculate structure (ambiguous once away from the SAXS regime)

• **BUT can be an excellent tool for testing molecular models (because we can quantitatively calculate WAXS data from molecular models)**
SAXS/WAXS Experiment

Flow cell (100 ms x-ray exposure)

Temperature controlled

1.5 mm path length (typical)

10 microliter sample volumes possible

>10 mg/ml concentration preferred
WAXS Data Set from Hb – 150 mg/ml

I_{prot} = I_{obs} - I_{cap} - (1-\text{vol\%})I_{ solvent}

Wide angle scatter largely due to buffer;

Each data set is composed of circularly averaged scattering from:

(i) Empty capillary
(ii) Buffer-filled capillary
(iii) Protein solution-filled capillary
KEY to WAXS – can predict quantitatively the data expected from a given molecular model…

Model of water of hydration as a uniform layer is inadequate

- WAXS patterns were computed from proteins using explicit atomic representations for water.
- Proteins were placed in droplets generated by MD simulations and scattering was calculated using an average over 100 snapshots.
- Water contribution was accounted for by subtraction of scattering from droplets containing water without proteins.

Simulated x-ray scattering of protein solutions using explicit-solvent models

Sanghyun Park,1,8† Jaydeep P. Bardhan,2 Benoît Roux,2,3 and Lee Makowski2

1Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA
2Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, USA
(Received 29 December 2008; accepted 25 February 2009; published online 7 April 2009)
Success of this approach also provides strong evidence that MD approaches are getting water of hydration correct.
… but there’s an ensemble of proteins in the scattering volume…
WAXS studies of protein ensemble – how does polymorphism effect the scattering pattern?

scattering from spheres of 14; 15 and 16 Å radius
minima at ~ 1/(radius)

so…the broader the ensemble; the greater the effect

WAXS is highly sensitive to this effect…

scattering from a solution of all three spheres looks like the average sphere but with minima filled in and maxima muted
Hemoglobin - Ensemble changes as function of protein concentration (molecular crowding…)

At concentrations below about 50 mg/ml, the scattering from met-Hb indicates a progressive increase in polydispersity - broadening of the structural ensemble.

Seen most easily at the positions of extrema
How sensitive is the ensemble to alterations in the protein?

di-α Hb is a variant in which the two α-chains are covalent linked at their termini

di-α-Hb appears far more rigid than CO-Hb

An engineered variant is far more rigid than the native protein
Doesn’t behave as it should…

Wide angle intensities are maintained more or less constant
Want to quantitatively model motions

molecular dynamics

simple models defined by global parameters

WAXS pattern always gives you a pair-distribution function \( p(r) \);
A simple model for increased fluctuations is to replace every

interatomic distance in the protein by a distribution of distances
\[
\sigma_{13} = (\sigma_{12}^2 + \sigma_{23}^2)^{1/2}
\]
\[
\sigma \sim r^{1/2}
\]

Alternatively, for rigid body motion, may find \( \sigma \sim r \)

The effects of fluctuations on scattering can be modeled by convoluting the pair-distribution function by a length-dependent function (e.g. Gaussian)
In the most general sense, the ensemble can be ‘broadened’ in many ways...
HIV Protease

Active site protected by two 'flaps'
When inhibitors (or substrate) bind to active site flaps fold down over them
Their flexibility is required for access to active site
Extensive information available on mutants including MDR
All studies use Q7K to prevent self-digestion

Consider T80N
(position T80 is invariant in both treated and untreated populations)

T80N abolishes both protease activity and viral infectivity
But thr80 is not in the active site (contacts active site residues)

And its structure is virtually indistinguishable from WT
In spite of the similarity in crystal structures – the solution scattering from WT and T80N is quite different. Structure? Or dynamics?

Differences between WAXS patterns +/- inhibitor

Indicate

Wt binds inhibitor ($\Delta I$ large)

MDR mutant binds inhibitor ($\Delta I$ large)

t80n does not bind inhibitor ($\Delta I \sim 0$)
MD simulations suggest T80N inhibits motion of the flap regions
Form of T80N scattering suggests WT looks similar but exhibits structural fluctuations.

Using t80n as a model for a rigid protease - adding fluctuations results in a predicted WAXS pattern indistinguishable from WT.

A single site mutation (t80n) causes a functionally significant suppression of structural fluctuations.
Ubiquitin

Is very rigid in aqueous solution

Calculated ~ observed

Little change in intensity as function of concentration

Change L50 to E…

What is effect?

Figure 4b

With Tobin Sosnick…
L50E

Intensity changes dramatically from WT ...

Intensity is a strong function of concentration

Modeling demonstrates that all of the intensity changes can be explained by structural fluctuations $\sigma \sim 0.7-0.8r^{0.5}$

For $r = 10 \ldots \sigma \sim 2.1 \text{ Å} - 20\%$ - so denaturing...

4.7 Å peak decreases more rapidly in model than in data… as might be expected for a global model
Alcohol dehydrogenase

+/- NADH

With ATP – differences are small; correspond to what you calculate from crystal structures

In absence of ATP –

Binding of NADH to ADH dramatically alters the global fluctuations – leading to intensity changes much greater than predicted from the crystal

$$\sigma \approx 0.1 - 0.8 r^{0.5}$$

Or – motions of ~ 2 Å or more in the lowest concentration apo form – suggestive of partial unfolding
... change in fluctuations may lead to change in function...

But what about detection of functional conformational changes in solution?

If we can generate representative structures of all abundant conformations, WAXS might be able to calculate their relative abundances
Multidomain assembled states of Hck tyrosine kinase in solution

Representatives of families of conformations abundant under different conditions

Catalytic domain – blue
SH2 domain green
SH3 domain yellow

Hck-YEEI – high-affinity mutant
Adenylate kinase +/- inhibitor  
(with George Phillips)

Catalyzes the interconversion of AMP; ADP and ATP

$$2 \text{ADP} \leftrightarrow \text{AMP} + \text{ATP}$$

Flaps cover the reagents once in the active site

Form of scattering suggests flaps very flexible in unliganded form
Adenylate kinase +/- inhibitor … and during catalysis

During catalysis closely resembles inhibited form

but cannot be constructed as linear combination of the two endpoints

Reduced chi-squares
apo – inhib 42.75
apo-adp 6.58
Inhib-adp 5.23
Cycling Adenylate Kinase

Can we see what some of these states look like?

Aden and Wolf-Watz, 2007
Collect WAXS data from AdK under many different conditions

- AMP, binds both sites - apparent $K_d$ of 3.2 mM
- ATP, binds LID domain - $K_d$ of 43 $\mu$M
- AP5A, binds both sites - tight binding (nM)
- ADP, enzyme is catalytic - apparent $K_d$ of 267 $\mu$M

From thermodynamic measurements can predict relative abundances of different species at any given ligand concentrations
Singular Value Decomposition

\[ D = USV' \]

- \( A = \text{number of angles} \)
- \( N = \text{number of experiments} \)

Data Matrix, \( D \) (\( A \times N \))
Orthonormal Basis, \( U \) (\( A \times N \))
Singular Values, \( S \) (\( N \times N \))
Coordinates, \( V \) (\( N \times N \))

4 significant vectors
the future...

• Decomposition of scattering from ensembles into terms associated with each individual species has the potential to provide detailed structural information on multiple intermediates

• MD modeling to predict structural ensembles under different conditions

• Comparison of WAXS and MD to evaluate the MD models and refine structures of functionally important intermediates that cannot be captured crystallographically

• E.g…. use of intermediate structures to inform drug design

• Coordination with thermodynamic (etc..) studies to characterize the components present under each condition studied
• There will always be multiple conformations in the scattering volume… where more than one is abundant we will need to tease their contributions apart.

• Continuous flow
• Stopped flow
• Steady state reacting systems
• Pump probe (TR-WAXS)
Thanks

Bob Fischetti
Dave Gore
Diane Rodi
Suneeta Mandava
Amina Aziz
Lynda Dieckman
Sanghyun Park (ANL)
Jaydeep Bardhan (Rush)
David Minh (ANL)
Jyotsana Lal (ANL)
Sichun Yang (Case)
Benoit Roux (UCh)
Tobin Sosnick (UCh)
Karl Freed (UCh)
Juoko Virtanen (UCh)

Chien Ho (CMU)
George Phillips (UW)
Steve Kent (U Ch)
Vladimir Torbeev (U Ch)
Celia Schiffer (UMass)

DOE
ANL
NIH