USING MOLECULAR DYNAMICS SIMULATIONS TO INTERPRET SAXS EXPERIMENTS

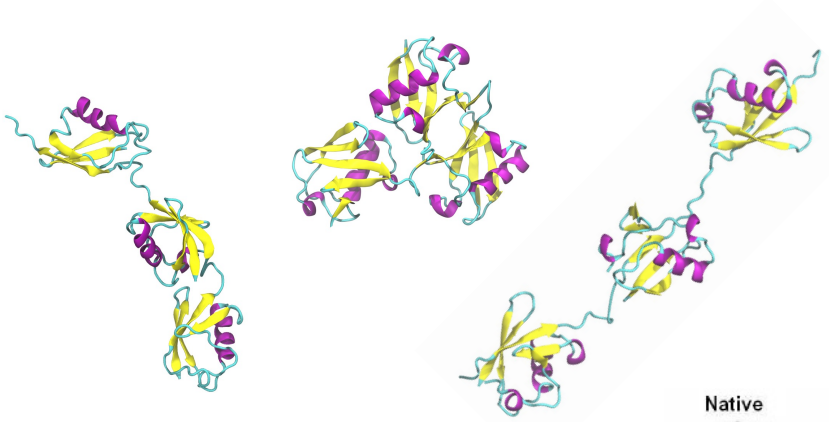
Jeff Wereszczynski
Department of Physics, Illinois Institute of Technology

Outline

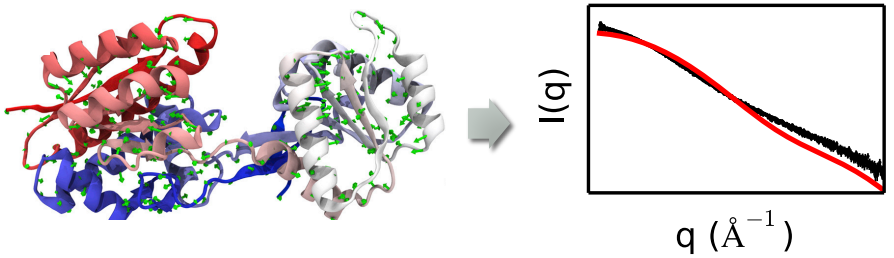
- Brief overview of Molecular Dynamics (MD) simulations
- Using MD simulations to compute SAXS and WAXS curves
- Biasing MD simulations to model SAXS experiments
- Post-hoc analysis of unbiased simulations for flexible systems:
  - Plausible Structure Generation
  - Minimal Ensemble Approaches
  - Maximum Entropy Approaches

Key Idea: How much information do you have from your simulations and experimental data, and how do you balance those?

Why Should I Care about MD?

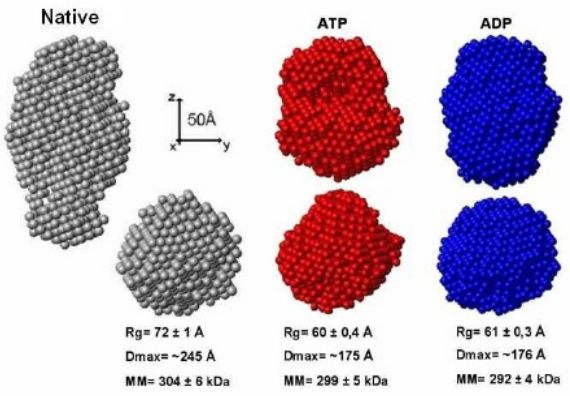


Flexible Systems



X^2>>1

Poor Fits to Experiments



Lacking High Resolution Data

MD and SAXS are Natural Compliments

SAXS:

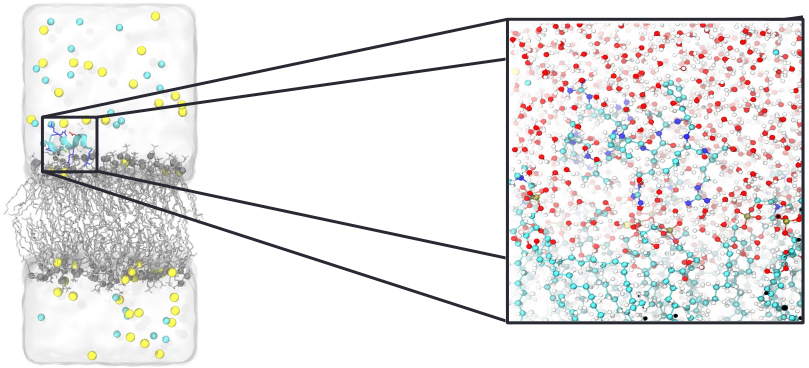
- Relatively easy to perform on diverse systems
- Provides information on large-scale conformational changes
- Relatively low-information content

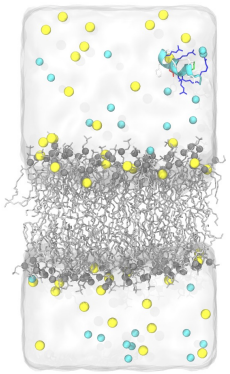
MD:

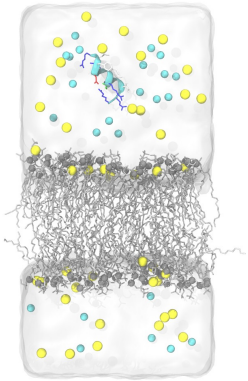
- High-resolution data
- Hard to sample large-scale conformational changes
- Limited by models in use (force fields, fixed charges, etc)

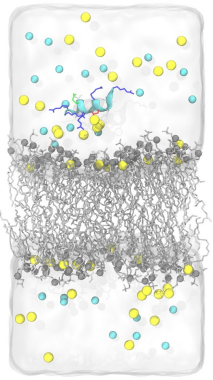
Molecular Dynamics (MD) Simulations Model Biomolecular Motions

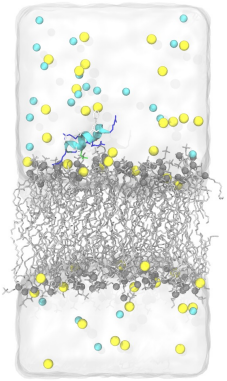
$$U(r) = \sum_{bonds} K_r (b - b_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{R_{ij}} \right]$$









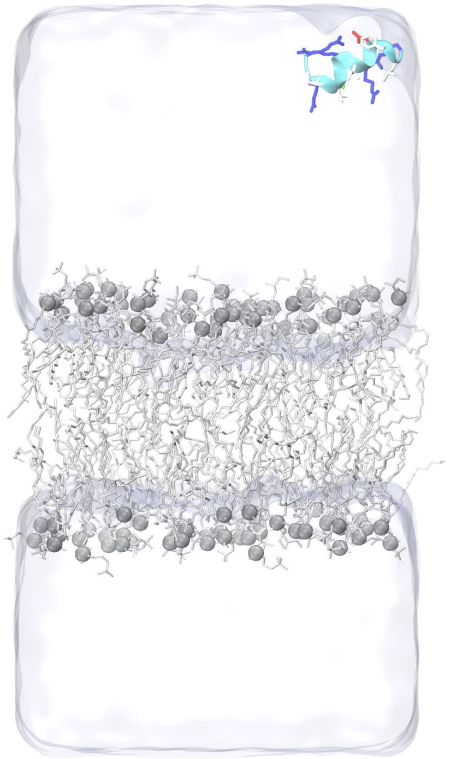


calculate F(r) → Δt → calculate F(r) → Δt → calculate F(r) → Δt → calculate F(r) → Δt

Molecular Dynamics (MD) Simulations Model Biomolecular Motions

Advantages:

- All-atom representation
- Can be applied to model diverse systems
- Can be used to compute kinetic and thermodynamic data



Disadvantages:

- Computationally expensive
- Can be slow to converge
- Limited system sizes
- Fixed-charged force fields limits the physics that can be modeled

CHARMM-GUI: A User-Friendly Tool to Setup MD Simulations

CHARMM-GUI
Effective Simulation Input Generator and More

CHARMM is a versatile program for atomic-level simulation of many-particle systems, particularly macromolecules of biological interest. - M. Karplus

about us :: input generator :: Q&A :: archive :: charmm docs :: lectures :: movie gallery :: video demo :: citations :: update log :: jobs & events :: giving

Some lectures, job postings, and FAQ are now available. See update log for update history and giving for donation. Contact info is given below. Logout

CHARMM-GUI
- About Us
- Input Generator
- Questions & Answers
- Archive
- CHARMM Docs
- Lectures
- Movie Gallery
- Video Demo
- Citations
- Update Log
- Jobs & Events
- Giving
- ST-analyzer

Geographical Visitors

Front Page

User Profile

Since its original development in 2006, CHARMM-GUI has proven to be an ideal web-based platform to interactively build complex systems and prepare their inputs with well-established and reproducible simulation protocols for state-of-the-art molecular simulations using widely used simulation packages such as CHARMM, NAMD, GROMACS, AMBER, GENESIS, Tinker, LAMMPS, Desmond, and OpenMM. The CHARMM-GUI development project has been widely adopted for various purposes and now contains a number of different modules designed to set up a broad range of molecular simulation systems in Input Generator. Many original modules were developed as an in-house effort, but we have established close collaborations with the developers of CHARMM and other MD simulation packages for addition of newer modules.

Our philosophy in CHARMM-GUI development is less about providing the nuts and bolts of molecular modeling, but instead focused on helping users to achieve a task, such as building a membrane system or solvating a protein, by providing a streamlined interface. This design principle helps us to think of the workflow critically when designing the interface, which leads CHARMM-GUI to be accessible to users with little experience in modeling tools and remains useful to experts, especially for batch generation of systems. CHARMM-GUI has been used by many researchers, and it is a well-recognized tool in the molecular modeling and simulation communities (see Google Scholar Citations).

The CHARMM-GUI development project is still ongoing. These functionalities are not only based on requests from general users and developers, but also on an emerging need for a unified platform to prepare and execute various advanced simulation approaches that have been developed and will be developed by many developers in diverse simulation communities and packages. CHARMM-GUI will continue to help expert and non-expert researchers from a broader range of the modeling and simulation community to build the complex molecular systems of their interest and prepare the input files for any general and advanced modeling and simulation through the large and unique scope of CHARMM-GUI functionality. It will also provide an effective one-stop online resource for the biomedical research community to carry out innovative and novel molecular modeling and simulation research.

Visit our COVID-19 Archive for collection of SARS-CoV-2 protein systems.
Follow CHARMM-GUI on Twitter: https://twitter.com/CharmmGui.

LEHIGH UNIVERSITY

Lehigh University / Department of Biological Sciences / Department of Chemistry / Department of Bioengineering / Im Lab
Problems, Questions, & Comments? E-Mail / Copyright(c) 2006-2022 by the Im Lab

https://www.charmm-gui.org

Using MD to Treat SWAXS Hydration Effects

Instead of implicit model for hydration effects, treat hydration as in experiments:

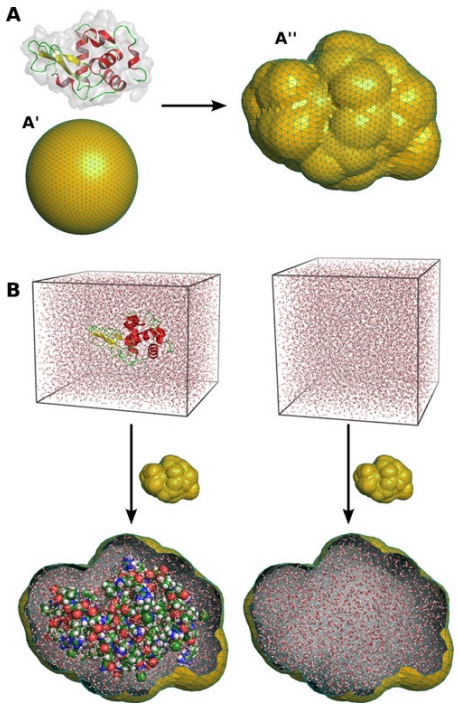
$$I(q) = I_{sam}(q) - I_{buf}(q)$$

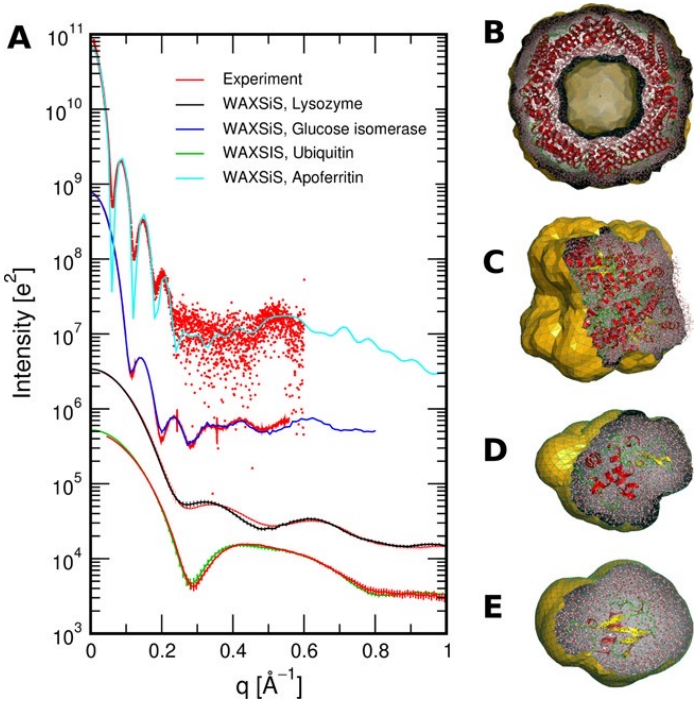
Generate $I_{sam}(q)$ and $I_{buf}(q)$ from solute restrained atomistic simulations.

$$I(q) = \left\langle \tilde{A}_i(q) - \tilde{B}_i(q) \right\rangle_{\Omega}$$

Uses fourier transform of atomic densities:

$$\tilde{A}_i(q) = \sum_{j=i}^{N_A} f_i(q)\, e^{-iq\cdot r}$$

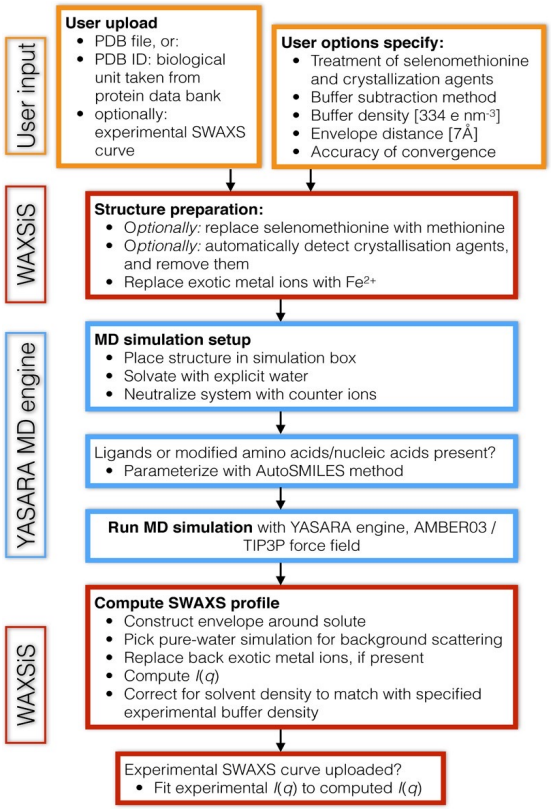


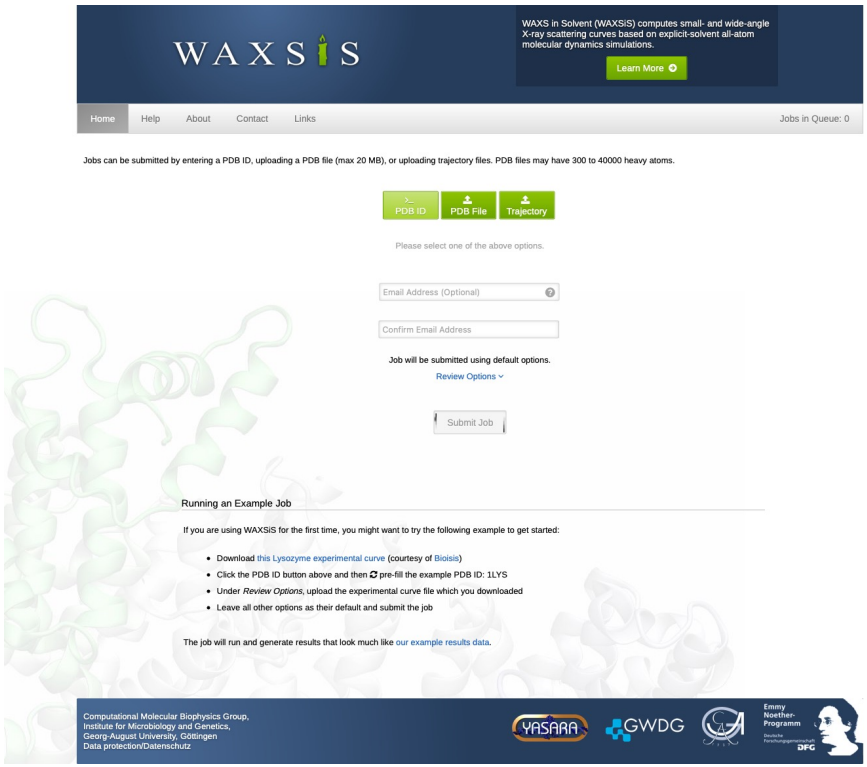


Nucleic Acids Res, Volume 43, Issue W1, 1 July 2015, Pages W225–W230, https://doi.org/10.1093/nar/gkv309

WAXSIS: An online tool for small- and wide-angle X-ray scattering curves

http://waxsis.uni-goettingen.de





Nucleic Acids Res, Volume 43, Issue W1, 1 July 2015, Pages W225–W230, https://doi.org/10.1093/nar/gkv309

Advantages and Disadvantages of WAXSIS

- Advantages:
  - No free solvation parameters
  - Reproduces SWAXS curves beyond ~q=0.3
  - Available as a user-friendly web server or standalone Gromacs code for power users

- Disadvantages:
  - Relies on atomistic models of solvation (corrections applied to try to account for this)
  - Only determines a SWAXS curve for one structure
  - Does not allow changes in solute structure

SAXS-Biased MD Simulations

- MD force-fields can be biased by experimental data such as SAXS:

$$E_{Total} = E_{FF} + E_{SAXS}$$

- Where a biasing force is defined by:

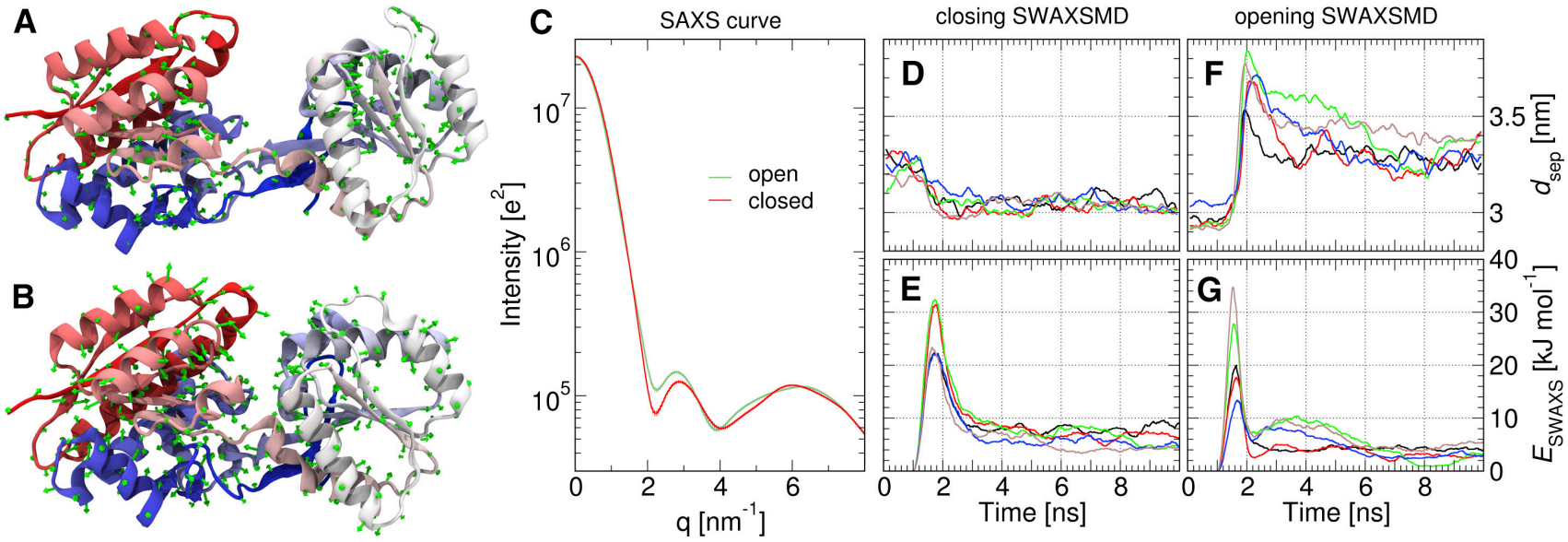
$$E_{SAXS} = \frac{k_r k_B T}{n_q} \sum_{i=1}^{n_q} \frac{\left(I_c\left(q_i, R\right) - I_{exp}\left(q_i\right)\right)^2}{\sigma_i^2}$$

- Requires calculating the intensity at each snapshot, I_c, with a bias applied by a force constant k_T

Biophysical Journal 2015 1082573-2584DOI: (10.1016/j.bpj.2015.03.062)

SAXS-Biased MD Simulations: Comparison to Theoretical Data

Theoretical data shows Leucine Binding Protein can be biased quickly between closed and open states



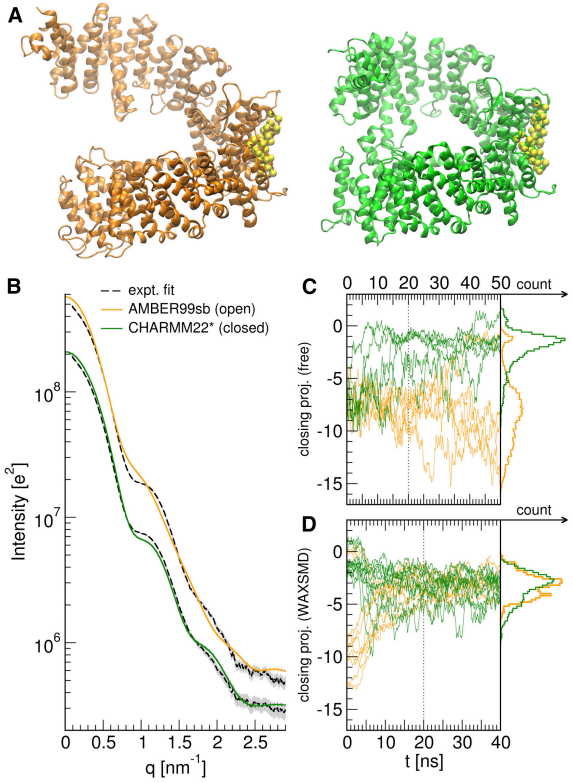
Biophysical Journal 2015 1082573-2584DOI: (10.1016/j.bpj.2015.03.062)

SAXS-Biased MD Simulations: Comparison to Experimental Data

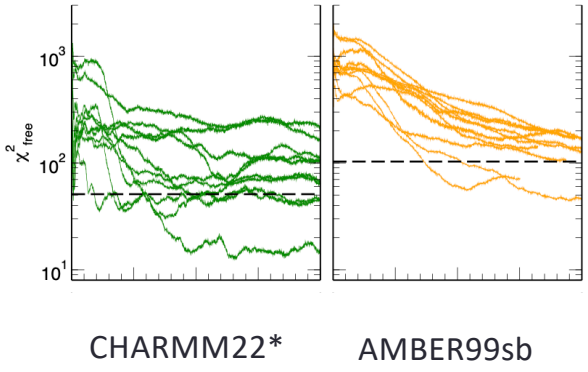
CRM1: 21 repeat nuclear exportin observed in open and closed states.

Unbiased simulations with different force fields showed different propensities for open vs closed states (c).

SAXS-biased simulations with both force fields showed a quick convergence to an intermediate state (d).



Simulation replicates showed quick, but varied, convergence



But…cryo-EM experiments suggest a mix of states is present in solution (2:1 open to closed)

Biophysical Journal 2015 1082573-2584DOI: (10.1016/j.bpj.2015.03.062)

Advantages and Disadvantages of SAXS-Biased MD

Advantages:

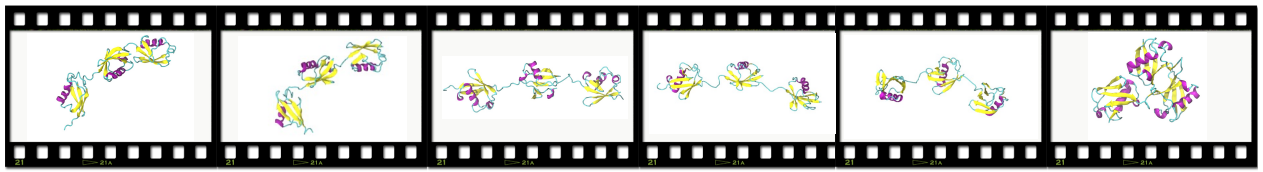
- Can quickly refine structures to match experimental data
- Allows for changes in large-scale and small-scale solute structures

Disadvantages:

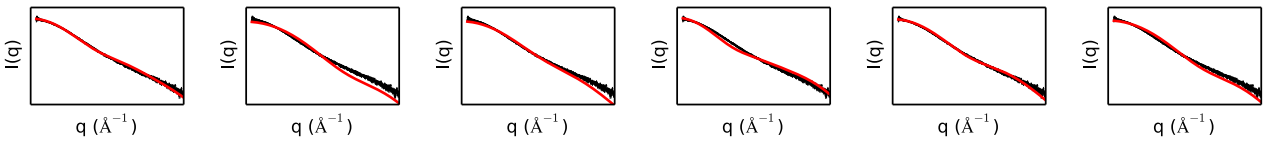
- May be best for local-refinement, may not be appropriate for drastic structural changes
- May bias to non-physiological intermediate states when solution data is from an ensemble
  - Solutions to this exist, such as running multiple interacting MD replicas

Determining Potential Structures with Conventional Simulations

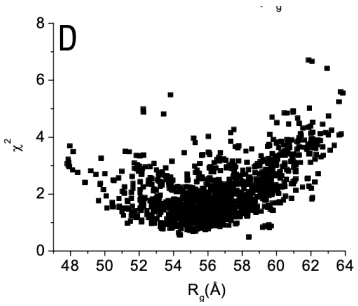
Idea: Use simulations to determine potential structures



Compute scattering profiles for each structure



Consider which structures fit experimental data

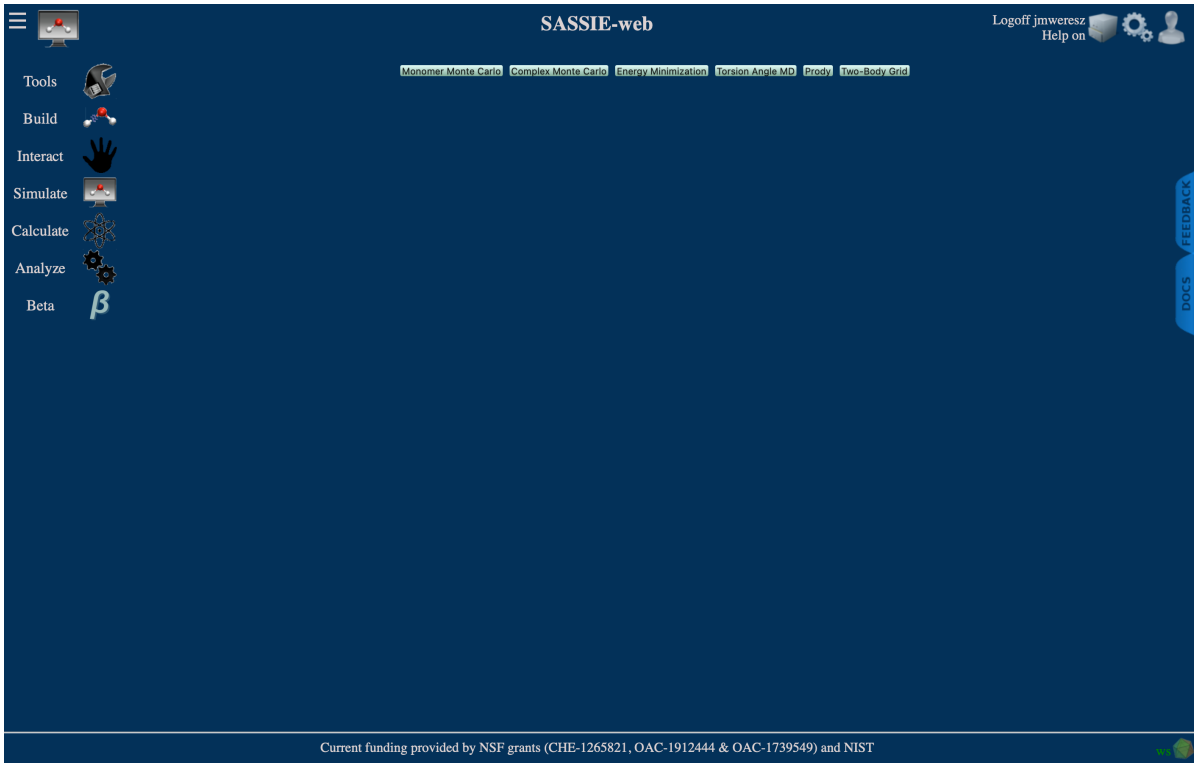


Using SASSIE-Web to Interpret SAXS Experiments

Web-based tool for connecting atomic structures to scattering data:

- Builds structures
- Perform basic MD and MC calculations
- Calculate scattering curves for structures
- Perform chi-squared and other analysis

https://sassie-web.chem.utk.edu/sassie2/

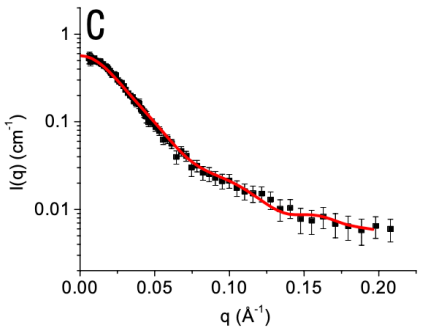


Example: Using SAXS + MD to understand α-catenin structures

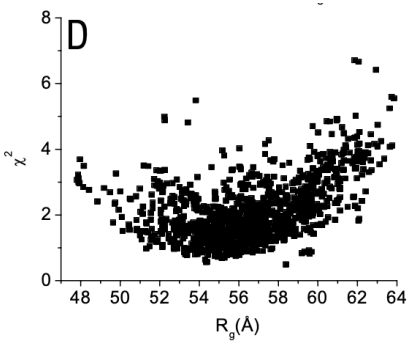
α-catenin: primary link between cadherins and actin cytoskeleton. Contains three domains (N, M, ABD) with flexible linkers

Monte Carlo simulations used to determine the structural pool

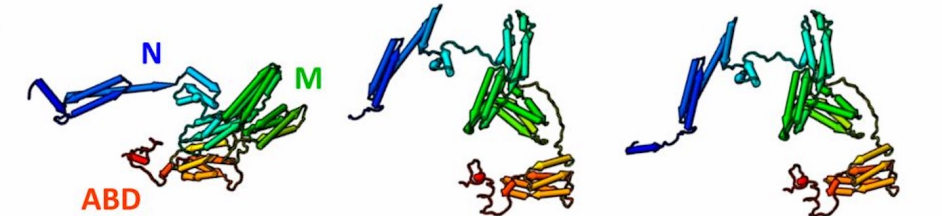
Results show multiple conformations of M and ABD domains



Best-fit curve to experiments



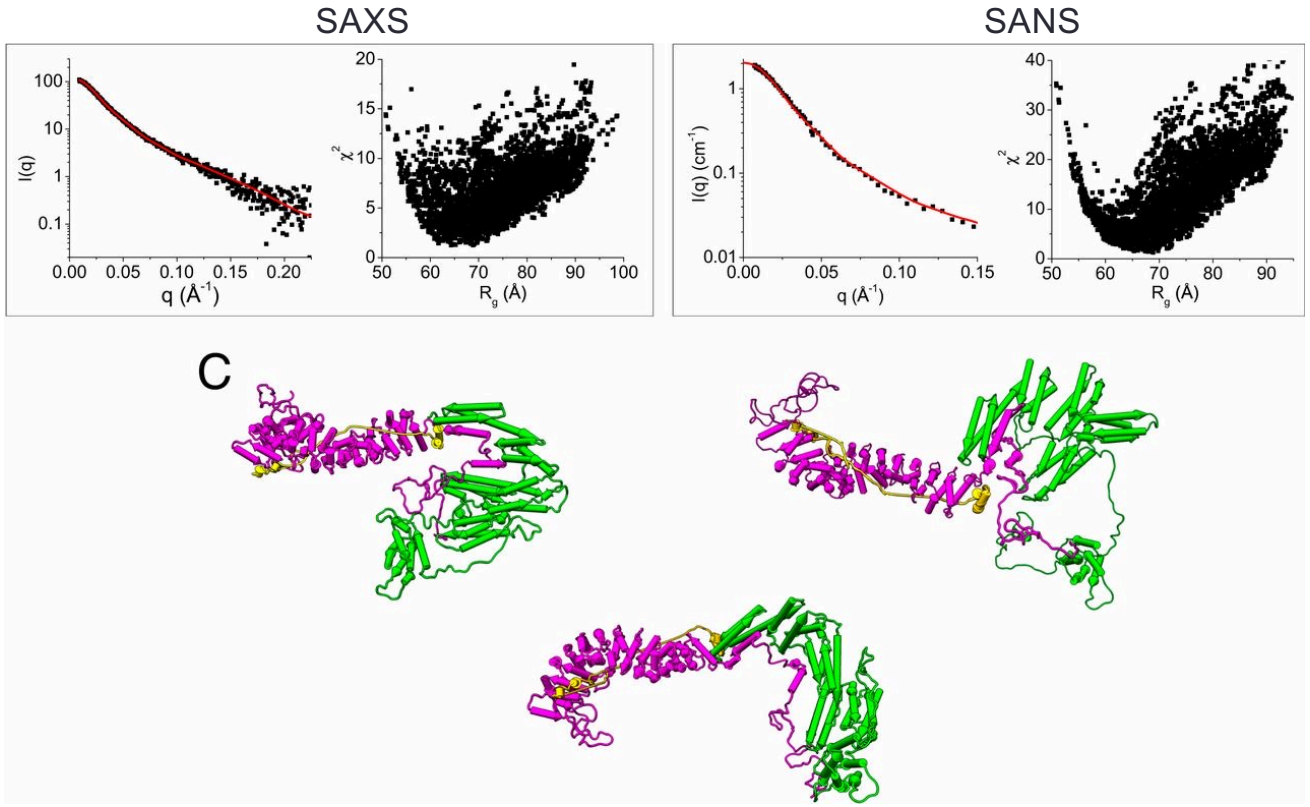
Fit vs R_g for all structures



Bush *et al.* PNAS. (2019)

Example: Using SAXS + MD to understand α-catenin structures

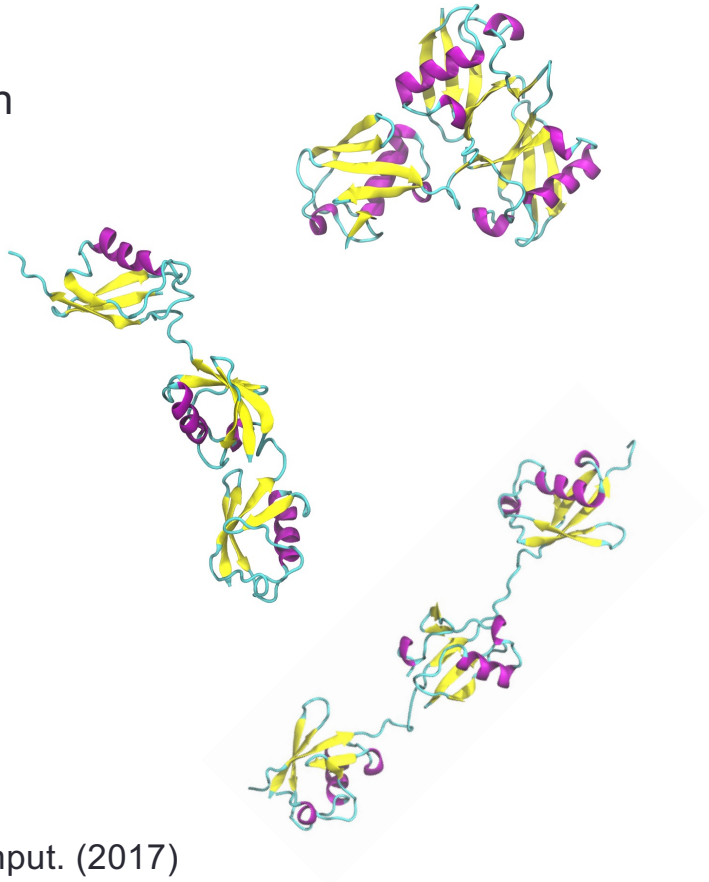
α-catenin•β-catenin•epithelial (ABE) complex also shows similar heterogeneity with SAXS and SANS



Bush *et al.* PNAS. (2019)

Example of a Flexible System: Tri-Ubiquitin Chains

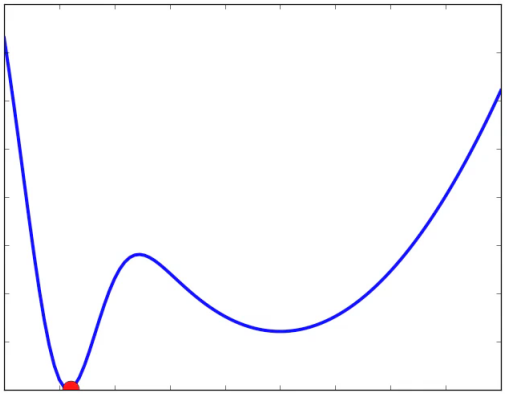
- SAXS experiments performed on diverse tri-ubiquitin systems (Eric Strieter, UMass Amherst)
- Conventional + accelerated molecular dynamics simulations of similar systems (us)
- Bayesian refinement of simulation ensembles to determine the minimal basis set to match experiments



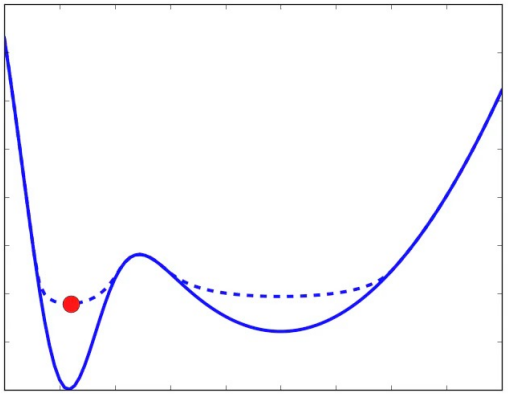
Bowerman *et al.* J. Chem. Theory Comput. (2017)

Accelerated Molecular Dynamics (aMD) Speeds Sampling (In Theory)

Conventional MD



Accelerated MD



$$V^*(r) = V(r) + \Delta V(r)$$

$$\Delta V(r) = \begin{cases} 0 & V(r) \geq E \\ \dfrac{(E - V(r))^2}{\alpha + E - V(r)} & V(r) < E \end{cases}$$

Hamelberg, Morgan, & McCammon *J. Chem. Phys.* (2004)

Determining Minimal Ensembles of Structures to Fit SAXS Data

Generate candidate structures
- aMD, cMD, Monte Carlo, TAMD, etc…

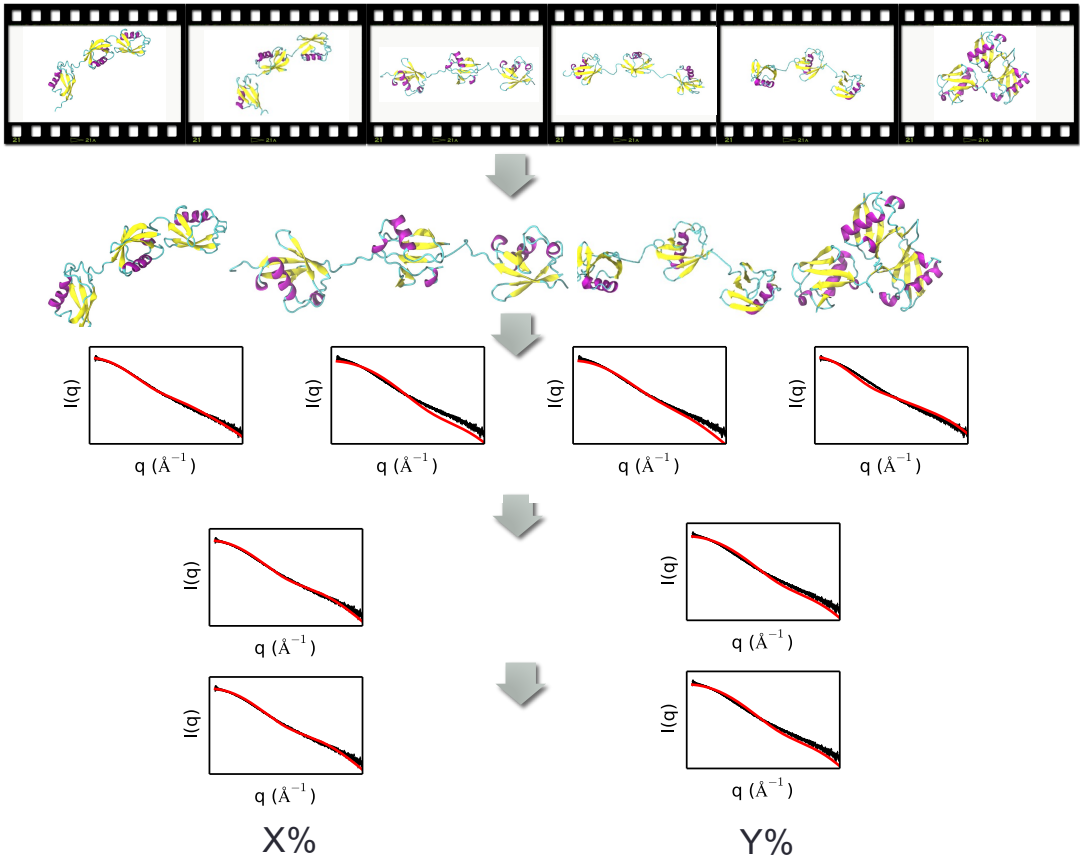
Pare down the structures into a manageable number
- RMSD based clustering

Compute theoretical scattering profiles for each structure
- Crysol (here), SasCalc, FoXS, etc.

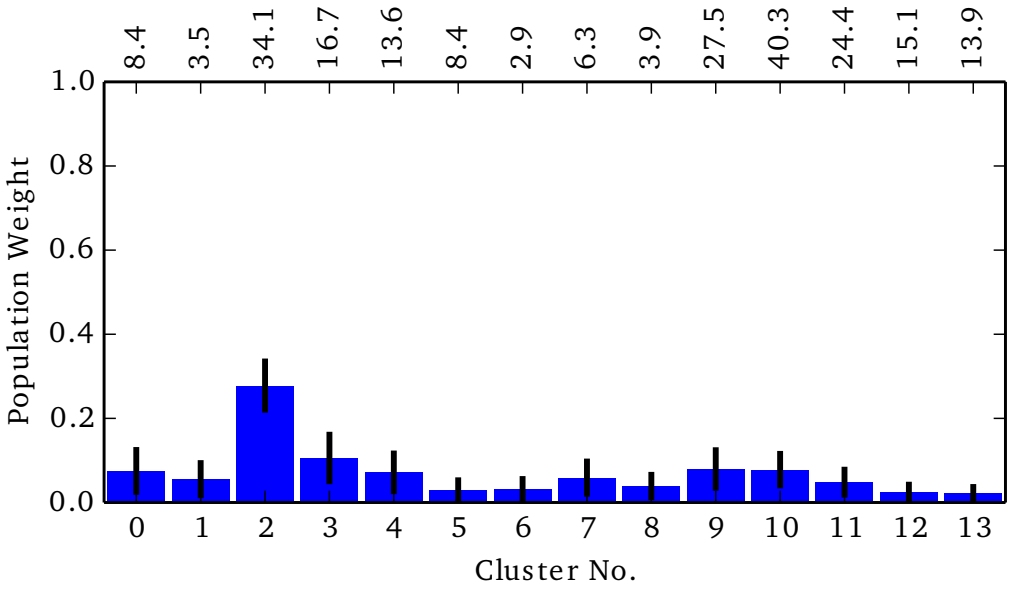
Cluster scattering profiles
- $\chi 2_{free}$ based hierarchical clustering

Determine populations of states
- Bayesian Monte Carlo algorithm

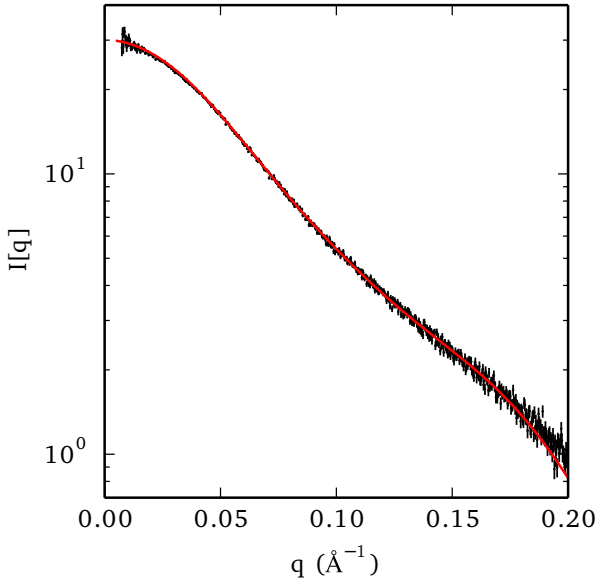
χ^2_{free}: Rambo & Tainer *Nature* (2013)



(Over)Fitting SAXS Data to a Population of States



$\chi^2_{free} = 1.12$

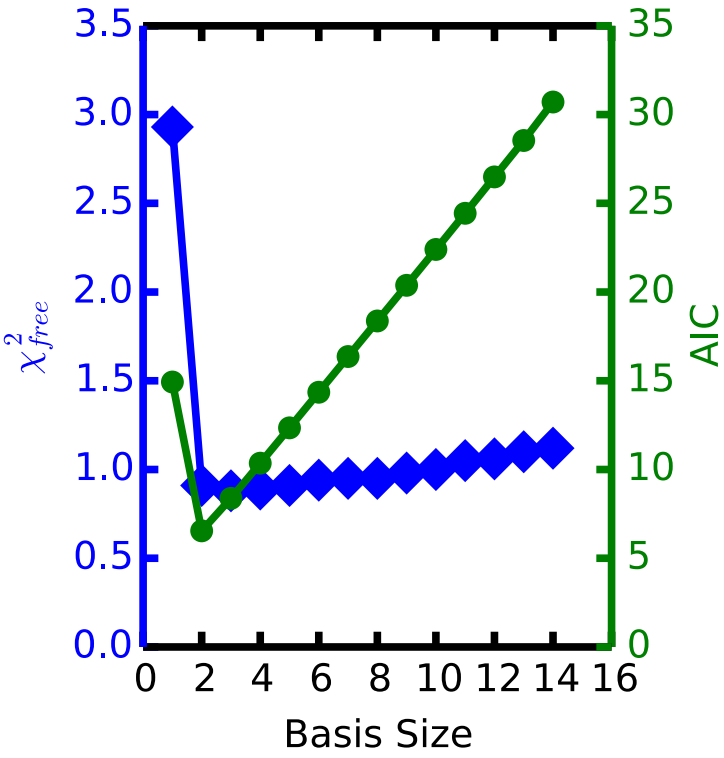


Resisting Overfitting with Iterative Refinement to Find Minimal Basis Set

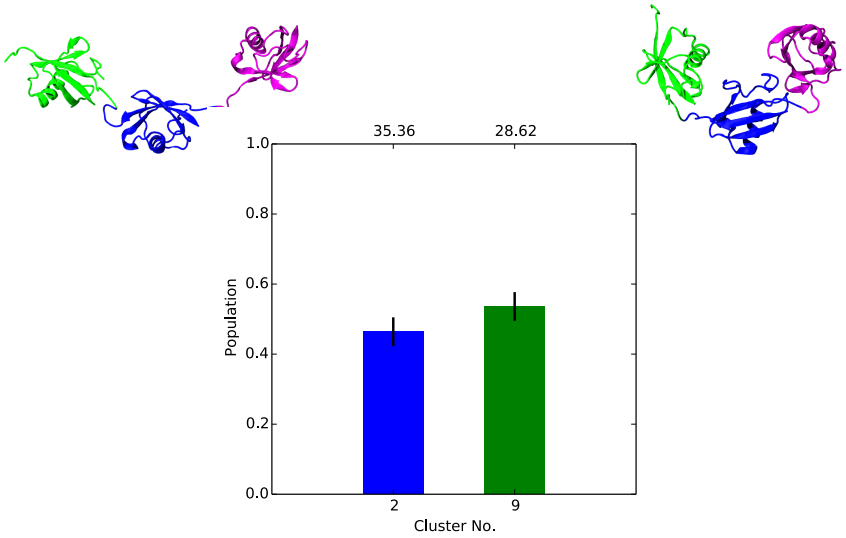
1. Compute populations with single scatterer
2. Compute each permutation of two scatterer basis sets, take the value with minimal χ^2
3. Repeat N times until all scatterers in basis set
4. Choose ensemble size that minimizes and the Akaike information criterion (AIC)

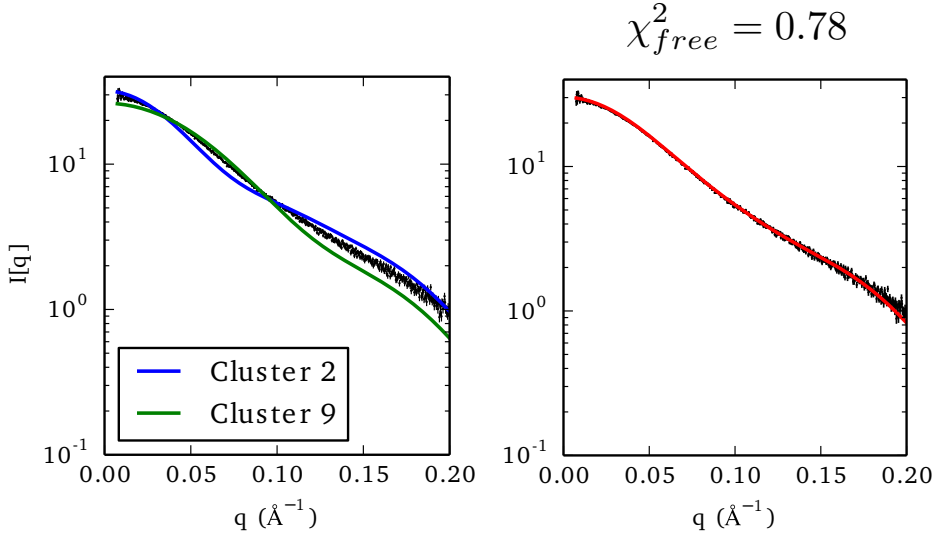
$$AIC = 2k - 2\ln\widehat{L} = 2k + \widehat{\chi^2}$$

$$BIC = \ln(n)\,k - 2\ln\widehat{L} = \ln(n)\,k + \widehat{\chi^2}$$



Iterative Refinement to Find Minimal Ensemble Set

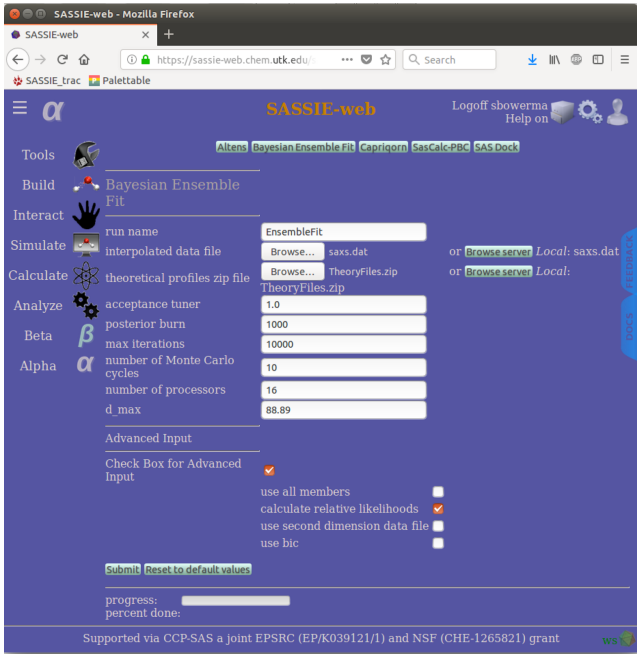


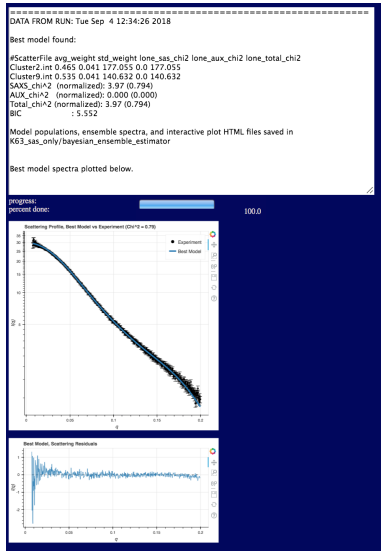


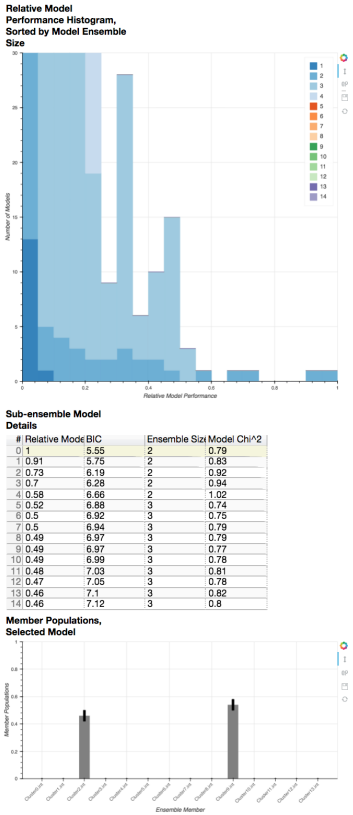
	R_g (Å)	Distal Group Distance (Å)	Interdomain Angle
2 (46±4%)	32.6 ± 0.2	68.9 ± 0.4	120°±0.1°
9 (54±4%)	23.3 ± 0.3	41.0 ± 1.5	97°±8.3°
Combined	27.5 ± 0.4	~	~
Experiment	28.0 ± ~1.0	~	~

BEES: Bayesian Ensemble Estimation from SAS

- SASSIE: Online portal for modelling SAXS data
  - Result of joint US/UK funded CCPSAS project, collaborative with NIST
- Python version available on GitHub for “power users”



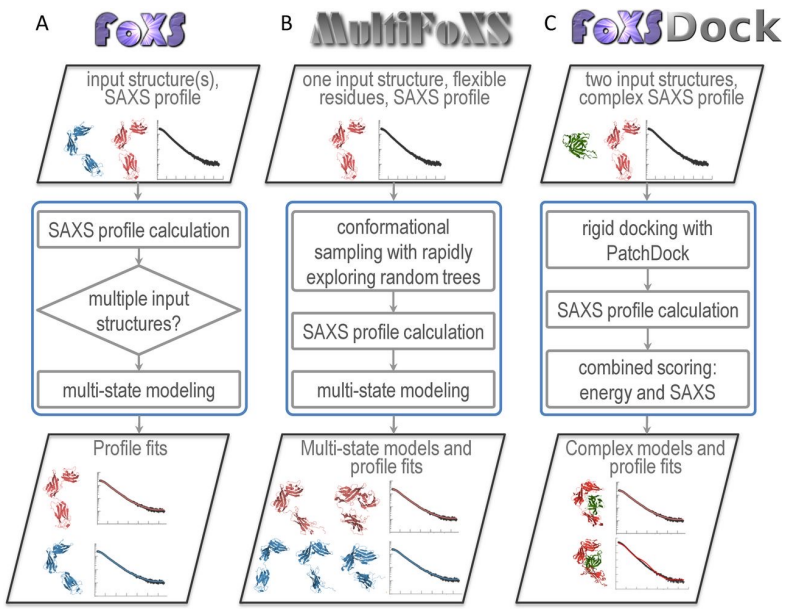




#	Relative Model	BIC	Ensemble Size	Model Chi^2
0	1	5.55	2	0.79
1	0.91	5.75	2	0.83
2	0.73	6.19	2	0.92
3	0.7	6.28	2	0.94
4	0.58	6.66	2	1.02
5	0.52	6.88	3	0.74
6	0.5	6.92	3	0.75
7	0.5	6.94	3	0.79
8	0.49	6.97	3	0.79
9	0.49	6.97	3	0.77
10	0.49	6.99	3	0.78
11	0.48	7.03	3	0.81
12	0.47	7.05	3	0.78
13	0.46	7.1	3	0.82
14	0.46	7.12	3	0.8

Bowerman *et al.* (Biophys. J. 2019)

Other Popular Minimal Ensemble Tools: MultFoxs and EOM



Ensemble Optimization Methods (EOM)

- Uses genetic algorithm to determine ensemble that best fits experimental results
- Part of the ATSAS package

BILBOMD: Webserver that uses MD + MultiFOXS

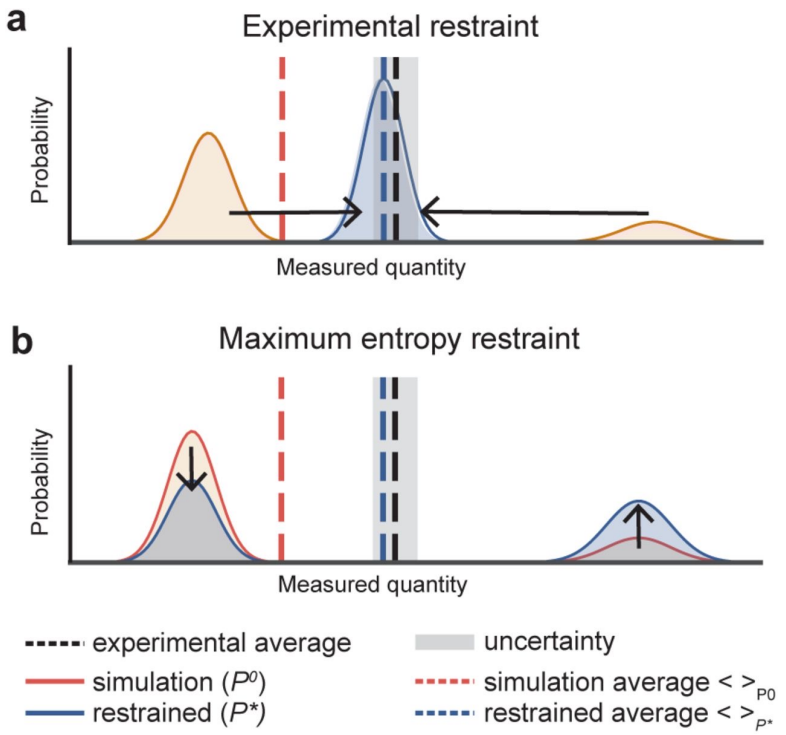
https://bl1231.als.lbl.gov/bilbomd

Nucleic Acids Res, Volume 44, Issue W1, 8 July 2016, Pages W424–W429, https://doi.org/10.1093/nar/gkw389

Maximum Entropy Approaches: Use as Many Structures as Possible

- Principle of maximum entropy: modify the simulation ensemble as little as possible to match the experimental data

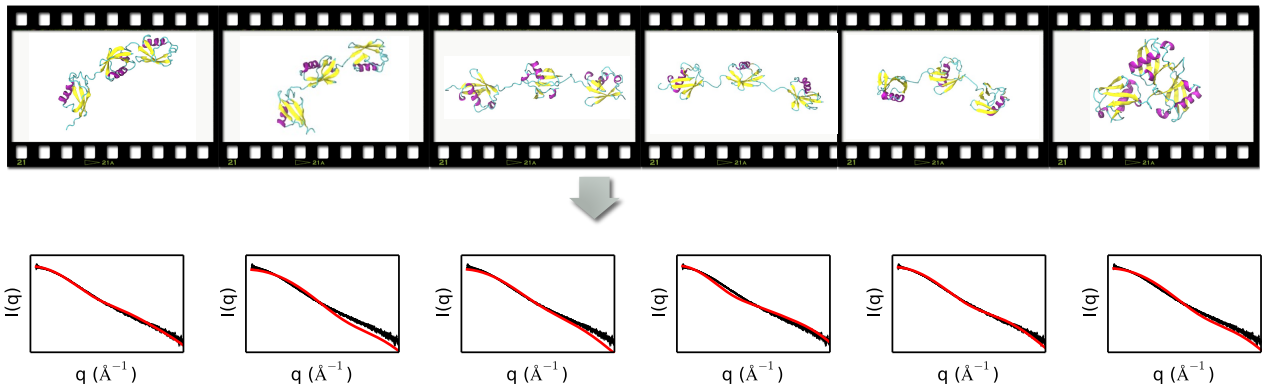
- Requires extensive simulation and the use of Lagrange multipliers



Bottaro *et al.* (Struct. Bio. 2020)

Theory of Maximum Entropy

Standard (extensive) simulations are run. Each simulation frame is given an equal weight of $w_j^0=1$



$$w_1^0 = 1 \quad w_2^0 = 1 \quad w_3^0 = 1 \quad w_4^0 = 1 \quad w_5^0 = 1 \quad w_6^0 = 1$$

Define two terms:

relative entropy

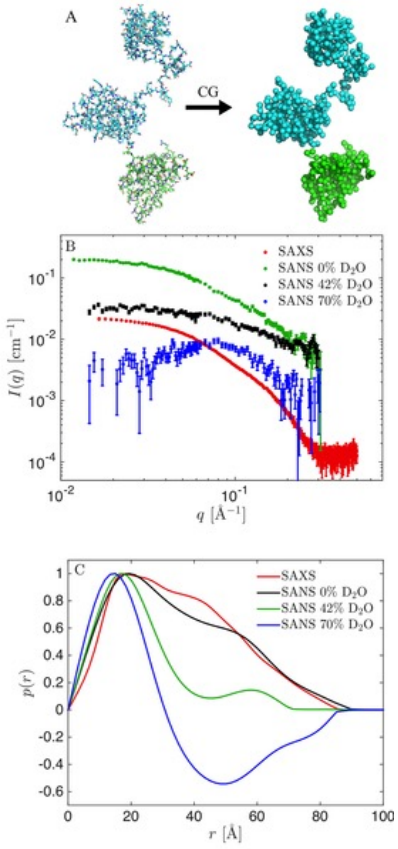
$$S(w) = -\sum_{j=1}^{N} w_j \cdot \log\left(\frac{w_j}{w_j^0}\right)$$

experimental fit

$$\chi^2(w) = \sum_{i=1}^{M} \left(\frac{\sum w_j \cdot I_{sim,j,i} - I_{exp,i}}{\sigma_i}\right)^2$$

Goal: balance these two by minimizing expression: $L(w) = \chi^2(w)/2 - \theta S(w)$

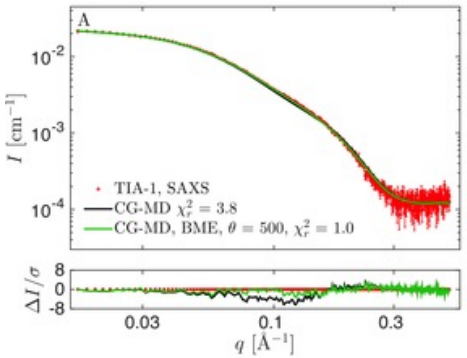
Example of Maximum Entropy

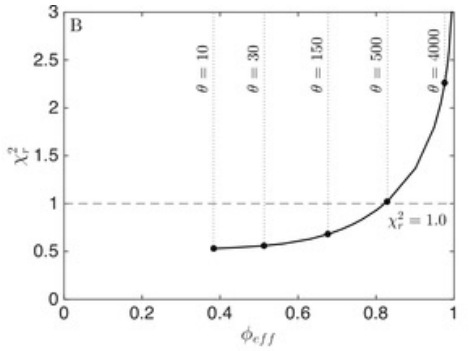


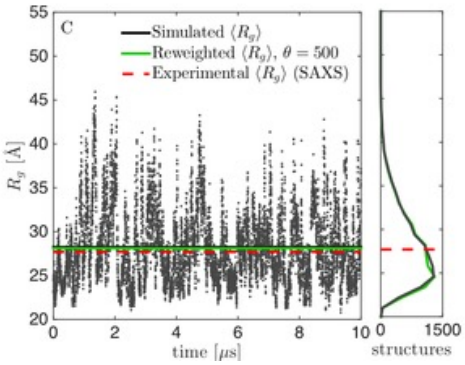
Combined SAXS, SANS, and coarse-grain (Martini) simulations of the three-domain TIA-1 protein

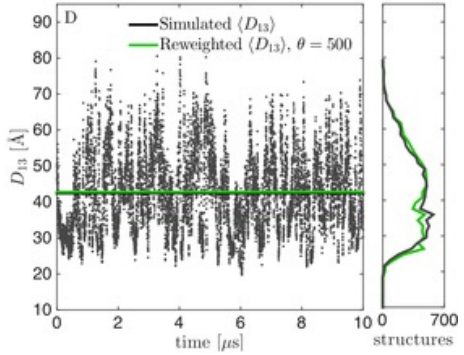
Using a “good” force field, the initial fit to experiments was decent (chi-squared=3.8) and improved to 1.0 with BME

This reweighting used ~80% of the simulation data and required minimal changes



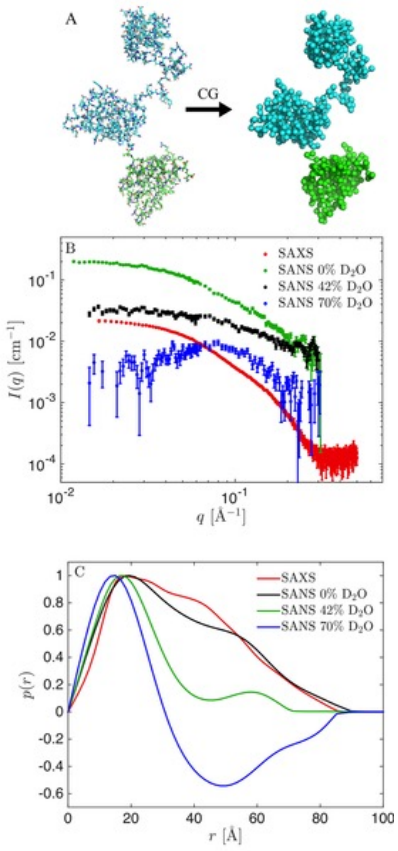






Larsen AH, Wang Y, Bottaro S, Grudinin S, Arleth L, et al. (2020)

Example of Maximum Entropy

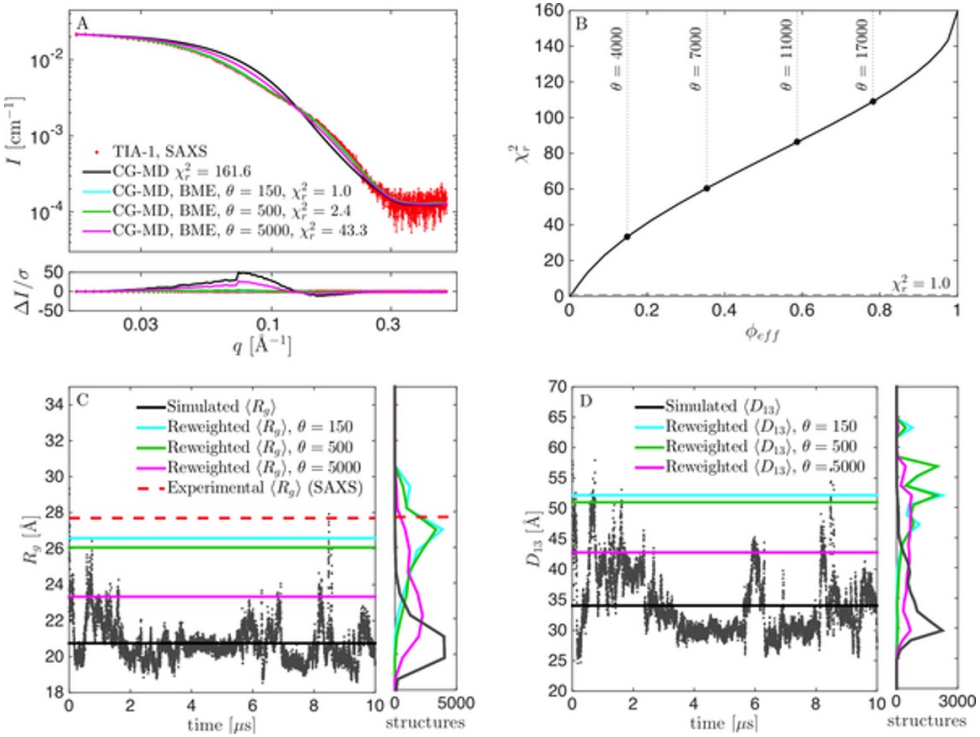


Combined SAXS, SANS, and coarse-grain (Martini) simulations of the three-domain TIA-1 protein

Using a “bad” force field, the initial fit to experiments was bad (chi-squared=161.6) and improved to 1.0 with BME

This reweighting used 0.4% of the simulation data

Reweighted ensembles are **qualitatively** and **quantitatively** different between the two force fields!



Larsen AH, Wang Y, Bottaro S, Grudinin S, Arleth L, et al. (2020)

Maximum Entropy Approaches: Advantages and Disadvantages

Advantages:

- Uses the most data from simulations
- Balances experimental and theoretical models
- Can include data from different experimental sources

Disadvantages:

- Requires extensive simulations
- Simulations must be fairly accurate (simulations can not be extensively perturbed)
- Requires a free parameter to balance theoretical and experimental data

Conclusions

A hierarchy of methods exists for using MD to interpret SAXS

Some methods are more complicated (maximum entropy)

Some methods are more straightforward to interpret (structure generation)

In all methods its important to be aware of limitations of the simulations, as well as uses and limitations of the method

References

- MD with CHARMM-GUI: https://www.charmm-gui.org
- WAXSIS: http://waxsis.uni-goettingen.de
  - Nucleic Acids Res, Volume 43, Issue W1, 1 July 2015, Pages W225–W230, https://doi.org/10.1093/nar/gkv309
- SAXS-Biased MD: Biophysical Journal 2015 1082573-2584DOI: (10.1016/j.bpj.2015.03.062)
- SASSIE: https://sassie-web.chem.utk.edu/sassie2/
- Minimal Ensemble Search:
  - Bowerman *et al.* J. Chem. Theory Comput. (2017)
  - https://modbase.compbio.ucsf.edu/multifoxs/
  - https://www.embl-hamburg.de/biosaxs/manuals/eom.html
  - https://bl1231.als.lbl.gov/bilbomd
- Maximum Entropy Approaches:
  - Larsen AH, Wang Y, Bottaro S, Grudinin S, Arleth L, et al. (2020)
  - Metainference: A Bayesian inference method for heterogeneous systems, Volume: 2, Issue: 1, DOI: (10.1126/sciadv.1501177)