



Interpreting SAXS/WAXS Data with Explicit-Solvent Simulations: A Practical Guide

Markus R. Hermann and Jochen S. Hub

Abstract

Small- and wide-angle X-ray scattering (SAXS/WAXS/SWAXS) have evolved to be accurate tools used to gain structural information of biomolecules in solution. However, the interpretation of SWAXS data remains challenging owing to the low information content of the data and scattering contributions from the solvent. In recent years, methods for the interpretation of SWAXS data based on explicit-solvent molecular dynamics (MD) simulations have become increasingly popular. The physicochemical information in the MD force fields complements the low-information SWAXS data, thereby greatly reducing the risk of overfitting, and the explicit-solvent models may accurately account for scattering contributions from the solvent. In this chapter, we provide a practical introduction to MD-based methods for the interpretation of SWAXS data. First, we present the back-calculation of a SWAXS curve from an MD trajectory as required to validate an MD simulation against experimental SWAXS data. Second, we present the structure refinement of an atomic model against SWAXS data using SAXS-driven MD simulations. Common technical problems together with appropriate solutions are discussed.

Key words SAXS, WAXS, Molecular dynamics simulation, Structure refinement, Proteins, Explicit solvent

1 Introduction

Small- and wide-angle X-ray scattering (SAXS/WAXS/SWAXS) are well-established experimental methods used to obtain structural information on biomolecules in solution. Owing to better light sources and detectors, improved standards for sample preparation, and progress in analysis software, SWAXS has become increasingly popular in the structural biology community. For a typical SWAXS experiment, two samples are prepared. The first sample contains the biomolecule in buffer, typically at a concentration range of 0.5–10 mg/ml; the second sample contains the same buffer but no biomolecule. Subtracting the scattering profile of the buffer from the profile of the biomolecule solution yields the final SWAXS curve $I(q)$, typically reported as a function of momentum

transfer q . The small-angle (SAXS) regime of this one-dimensional curve contains information on the overall shape and on the radius of gyration whereas the wide-angle (WAXS) regime contains additional information on the internal structure of the biomolecule.

SWAXS is often applied complementarily to other structural biology techniques since the method holds a number of key advantages. Unlike crystallography, SWAXS provides information on biomolecules in solution at near-native conditions. The solution environment allows the study of large-scale conformational transitions that would be inhibited by a crystal lattice. Compared to NMR spectroscopy, SWAXS may be applied to larger biomolecules. Unlike many spectroscopic methods that yield highly reduced structural information, such as the electrostatic environment of a dye or the distance between two dyes, SWAXS yields information on the overall size and shape of the biomolecules. Further, SWAXS does not require labeling of the molecule; it allows for time-resolved measurements, and it may be used in high-throughput setups. However, SAXS experiments make relatively high demands on the sample: large amounts of the biomolecule are required, the sample must be pure and monodisperse, and the buffer of the biomolecule sample must accurately match the pure-buffer sample. For experimental reviews of SWAXS, we refer to the literature [1–4].

The interpretation of SWAXS data in terms of structures, ensembles, and/or conformational transitions of biomolecules is an “inverse problem.” Hence, starting from an initial structural model, the model is updated until the SWAXS curve back-calculated from the model agrees with the experimental data. Evidently, such a protocol requires a forward model, that is, a method for accurately predicting the SWAXS curve from a proposed structural model. Accurate forward models have to take the following contributions to the SWAXS curve into account:

- The density of the hydration layer of many biomolecules differs from the density of bulk solvent. This difference in density thereby contributes to the overall electron density contrast reflected by the SWAXS curve. Hence, the forward model must include a model for the hydration layer. To account for the density of the hydration layer, simplified implicit-solvent methods require the fitting of the hydration layer against the experimental data, increasing the risk of overfitting.
- The volume of solvent displaced by biomolecules is required to compute the buffer-subtracted SWAXS curve; hence the forward model requires knowledge about the atomic volumes. Many SWAXS prediction methods use tabulated atomic volumes to account for the displaced solvent, leading to *reduced form factors* for the atoms. However, the atomic volumes are often unclear

and, indeed, tabulated atomic volumes greatly differ, leading to some uncertainty when relying on reduced form factors.

- Since the experimental SWAXS curve reports an ensemble average over thermal fluctuations, forward models should, if possible, take fluctuations into account.

In recent years, SWAXS prediction methods based on explicit-solvent molecular dynamics (MD) simulations have developed to an accurate forward model used for SWAXS interpretation efforts. For a list of MD-related methods for SWAXS prediction we refer to ref. 5. Critically, explicit-solvent MD provides an accurate model for the hydration layer and for the displaced solvent, thereby avoiding the use of solvent-related fitting parameters or reduced form factors. Further, MD simulations naturally model thermal fluctuations.

In this chapter we describe the protocols developed in our group for the interpretation of SWAXS data using MD simulations. For details on theory, implementation, and examples we refer to the literature [6–10].

This tutorial shall enable the reader to:

- Calculate a SWAXS curve from a given MD trajectory.
- Carry out SAXS-driven MD simulations with the aim of refining a protein against SWAXS data, that is, to drive the protein into conformations that match the experimental SWAXS curve.

For this tutorial, basic knowledge of Linux/Unix shells (such as bash) is required. Basic experience with the MD software Gromacs is recommended. Notably, for the calculation of only a few SWAXS curves from a given structure, the web server WAXSiS provides simple access to explicit-solvent SWAXS predictions for nonexperts [11].

1.1 Hands On: Getting Ready

As a test case we use the tungstate-binding protein TupA, which is a typical two-domain periplasmic binding protein involved in tungsten uptake via the ABC-type transporter system. The crystal structure and the SAXS curve of the apo form of TupA are available from the protein data bank (PDB ID 5MY5) and from the Small-angle Scattering Biological Data Bank (www.sasbdb.org, ID SASDBD9) [12]. Importantly, the protein was attached to a 23-residue-long expression tag that was not resolved in the crystal structure.

The calculations presented in this tutorial were conducted with the SWAXS extension implemented into Gromacs 2018, which is freely available for download at <https://biophys.uni-saarland.de/swaxs.html> and via GitLab. To compile the modified Gromacs, follow the installation instructions on the Gromacs website for version 2018 (<http://manual.gromacs.org/documentation/2018/install-guide/index.html>). The code implements both:

SWAXS curve predictions from a given MD trajectory and experiment-supported simulations for structure refinement against SWAXS data. To run this program, a Linux or MacOS environment and a powerful workstation are recommended. The code also benefits from outsourcing calculations on a CUDA-supporting graphics processing unit (GPU).

When predicting a SWAXS curve from an MD simulation, the electron density of the solvated protein is compared with the density of water. To this end, an MD simulation of the protein and an MD simulation of a pure-solvent box are required. To compute the X-ray scattering intensity, atomic form factors as tabulated by the Cromer-Mann parameters of all atom types are needed. A good starting point is to create a new folder for running a pure-solvent simulation:

```
$ mkdir waterbox; cd waterbox
```

In the folder `waterbox`, an MD simulation of the pure-solvent box should be carried out using a simulation box at least of the size of the protein simulation box. We recommend using the same water force field (in this tutorial TIP3P) and, ideally, the same MD parameters (`mdp` file) as used for the protein simulation. This ensures that the bulk solvent in the protein system closely matches the solvent in the pure-solvent system, thereby avoiding systematic errors in the buffer subtraction. If the protein system contains salt (in addition to counter ions), salt with the same concentration may be added to the water simulations. We recommend at least 20 ns of simulation for the pure-solvent system. For detailed tutorials on MD simulations, we refer to many excellent Gromacs tutorials. When the pure-solvent simulation is set up and running, it is time to prepare the MD simulation of the solvated protein in a separate directory:

```
$ cd ..
$ mkdir md; cd md
```

In the folder `md`, the MD simulation of the protein will be carried out. Starting from the PDB file, the MD system is set up using standard protocols, involving the generation of a topology (Gromacs module `pdb2gmx`), solvating the protein in a box (modules `editconf`, `solvate`), neutralizing the system with counter ions (`genion`), followed by energy minimization, equilibration with position restraints on the backbone atoms, and possibly further equilibration (modules `grompp`, `mdrun`). For the SWAXS calculation shown below, it is critical to use a relatively large water layer around the protein, for instance, by applying `gmx editconf -d 1.8` when generating the box. As soon as the MD simulations of

the pure-solvent box and the protein have finished, we turn to calculating the SWAXS curves.

If the user aims to compute the SAXS curve of a given fixed structure, the production simulations should be carried out with position restraints on the C_α atoms, thereby allowing fluctuations of solvent and side chains, but keeping the backbone fixed. Note that while the SWAXS calculations are implemented into a modified variant of Gromacs 2018, the MD simulations may be carried out with the most recent Gromacs version since the format of the compressed trajectory (`xtc`) later used for computing the SWAXS curve has not changed.

2 Hands On, Part 1: Calculating SWAXS Curves from MD Trajectories

2.1 Preparation of the Pure-Solvent System

In the folder with the pure-solvent simulation (`waterbox`), the first 100 ps of the trajectory are removed for equilibration. Further, a time step of a few picoseconds (*e.g.*, 10 ps) is used to ensure that the frames are statistically independent:

```
$ cd waterbox
$ gmx trjconv -f waterbox.xtc -o water-b100-dt10.xtc -b 100 -dt 10
```

Having a total of 1000 frames in the final trajectory `water-b100-dt10.xtc` is usually sufficient for computing a fully converged SWAXS curve.

Next, we generate a run-input file `waterbox.tpr` of the pure-solvent system that contains information on the atomic form factors (or Cromer-Mann parameters) of water atoms and, if present, salt ions. To this end, we use the `mdp` file `rerun.mdp` (*see* Appendix) with the following two parameters changed:

```
> waxes-solute =
> waxes-solvent = Water ; or: Water_and_ions in case ions are present
```

With

```
$ source /path/to/swaxs-gromacs/bin/GMXRC
$ gmx grompp -c conf.gro -f rerun.mdp -o waterbox.tpr
```

the run-input file `waterbox.tpr` is created. Here, the first line ensures that `grompp` from the SAXS-modified Gromacs is used.

2.2 Calculating the SAXS Curve from an `.xtc` File

To prevent the output files of MD simulation from being overwritten, the SWAXS calculations will be carried out in the subfolder `swaxs` within the `md` folder:

```
$ cd md; mkdir swaxs; cd swaxs
```

A run-input (`tpr`) file of the protein system is required that contains information on atomic form factors (*i.e.*, Cromer-Mann parameters) of all protein and solvent atoms. Since the chemical elements are by default not stored in a `tpr` file, a tool `gmx genscatt` has been devised that detects the chemical element from the atomic masses in combination with the atom names. `gmx genscatt` takes any `tpr` file of the protein system as input. Such a `tpr` file may be created with `grompp` using an empty `mdp` file (`empty.mdp`) as follows.

```
$ source /path/to/waxs-gromacs/bin/GMXRC
$ rm -f empty.mdp; touch empty.mdp
$ grompp -f empty.mdp -c ../confout.gro -p ../topol.top -o tmp./tpr
```

Next, `gmx genscatt` reads the atom names and masses from the `tmp.tpr` file and writes “include topology” files (`itp`) for each protein chain, named, *e.g.*, `scatter.itp`:

```
$ gmx genscatt -s tmp.tpr [-n index.ndx] [-vsites]
```

When prompted, the physical atoms of the protein should be selected by the user, *e.g.*, `Protein` or `Prot-Masses`. An index file may be required in case the protein contains nonstandard groups such as ligands. If the protein contains virtual sites, the option `-vsites` instructs `gmx genscatt` to expect uncommon atomic masses.

The generated `scatter.itp` has to be added to the topology file:

```
#ifndef SCATTER
#include "../scatter-prot.itp"
#endif
```

Critically, `scatter.itp` defines the Cromer-Mann parameters for one molecule type, similar to the common position restraint definitions. Hence, a good location for including `scatter.itp` is right behind `#include "posre.itp"`. If the protein contains multiple chains, multiple `itp` files with scattering information must be created and added to the individual `.itp` files of the protein chains. Now, the topology contains all required information for computing X-ray scattering intensities. The new topology file is henceforth named `swaxs.top`.

A new `.tpr` file for the SWAXS calculation is generated from `swaxs.top` by `grompp` using the `rerun.mdp` input file shown in the Appendix:

```
$ gmx grompp -f rerun.mdp -p swaxs.top -c ../confout.gro [-n index.ndx]
-o waxs.tpr.
```

2.3 Generating the Envelope

The thickness of the hydration layer included in the SWAXS calculation is defined by a spatial envelope (Fig. 2), which is generated prior to the SWAXS calculation with the `genenv` module:

```
$ echo "C-alpha Prot-Masses" | gmx genenv -s tmp.tpr -f ../
protein.xtc -d 0.7 -o envelope.dat -or envelope-ref.gro
```

The first input group “C-alpha” indicates the group of atoms used to fit the protein onto a reference structure. The option `-d` indicates the distance of the envelope from the protein atoms. The recommended value is 0.7 nm for computing the SWAXS curve from a given trajectory. If the protein is expected to expand during the simulation, for instance, when a closed structure is refined against SWAXS data of an open structure, a larger envelope is required. The resulting envelope is then saved in the file `envelope.dat`, and the reference coordinates for the least-square fit in `envelope-ref.gro`. The envelope can be visualized in PyMol with the file `envelope.py` (using `run envelope.py` in the PyMol terminal, also *see* Fig. 2). VMD users may use `gmx genenv --vmdout` to generate a tcl version of the envelope.

The envelope files used during the SWAXS calculations are specified with environment variables. With bash, use:

```
$ export GMX_ENVELOPE_FILE=/PATH/TO/CALCULATION/envelope.dat
$ export GMX_WAXS_FIT_REFFILE=PATH/TO/CALCULATION/envelope-ref.gro
```

2.4 The SWAXS Calculation

The SWAXS calculation is executed with the `mdrun` module using the `-rerun` option. Hence, `mdrun` does not carry out a new simulation, but instead reads and reanalyzes frames from a previous simulation:

```
$ gmx mdrun -sw ../../waterbox/waterbox.tpr \
-fw ../../waterbox/water-b100-dt10.xtc \
-rerun ../protein.xtc -deffnm waxes
```

This command reads the trajectories of the MD simulations of the protein in solvent (`protein.xtc`) and the pure-solvent system (`water-b100-dt10.xtc`), along with the run-input files containing the atomic form factors (`waxes.tpr` and `waterbox.tpr`, respectively), and calculates the average buffer-subtracted SWAXS curve.

2.5 Analysis of the Results

The SWAXS calculation writes several new files into the `swaxs` directory:

- `waxes_final.xvg`—the computed SWAXS curve.
- `waxes_contrib.xvg`—contributions to the SWAXS curve (*see* Eq. 10 of ref. 8).

- `waxs_spectra.xvg`—SWAXS curves over time, illustrating the convergence of the SWAXS curve with the number of frames.
- `waxs.log`—several statistics, most importantly the radii of gyration R_g from Guinier analysis and from the bare protein (excluding the hydration layer).
- `Protein+solvationlayer.pdb`—example frame of the protein and the hydration layer included into the SWAXS calculations. May be visualized in PyMol together with `envelope.py`.
- `excluded_solvent.pdb`—example frame of the pure-solvent droplet within the envelope, used to compute the buffer subtraction.

The calculated SWAXS curve can now be compared with the experimental curve that was downloaded from SASBDB. Since experimental SAXS data are typically not normalized to the intensity per solute, the experimental SAXS curve has to be scaled to the calculated curve. (Since the experimental data for TupA are limited to small angles, we refer to it here as SAXS.)

As evident from Fig. 1, the experimental curve and the curve computed from the closed structure greatly differ. Furthermore, the radius of gyration computed by the Guinier analysis differs between the calculated and the experimental curves (2.00 nm *versus* 2.40 nm). This demonstrates that the crystal structure is too compact to match the SAXS data, and that the SAXS curve instead reflects an extended, more open state. Hence, in the next step, we carry out a SAXS-driven (or SAXS refinement) simulation to find a structure that is compatible with the SAXS curve.

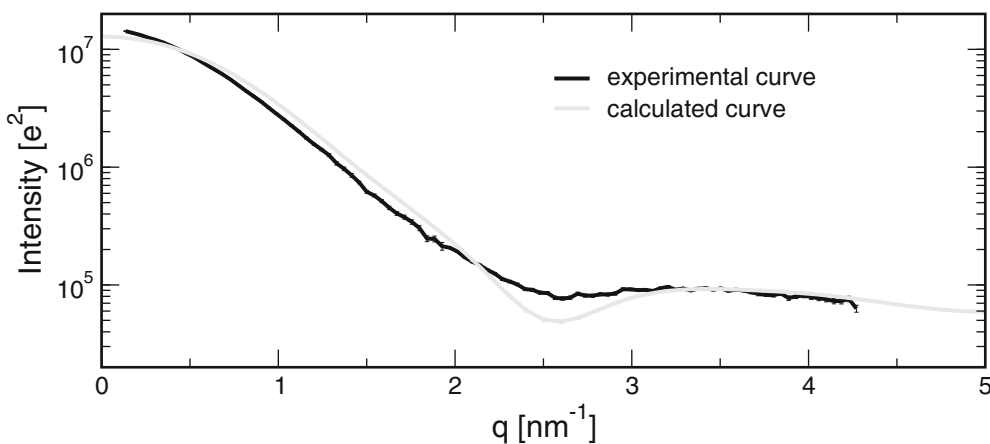


Fig. 1 SAXS curve calculated from the closed TupA crystal structure (gray) and obtained experimentally (black). The discrepancy between the two curves reveals that the SAXS data reflect a different, more open structure

3 Hands On, Part 2: SAXS-Driven MD Simulation, Refining a Structure Against Experimental Data

For the SAXS-driven MD simulations, a new folder is needed:

```
$ mkdir saxsmd
```

Again, the simulation of the water box and the envelope files are required. These can be taken from the previous calculation given that the envelope is large enough to enclose an extended protein structure. If the envelope is too small, you need to generate a larger envelope with, *e.g.*, `gmx genenv -d 1.5`.

During the SAXS-driven simulations, the SAXS curve will be calculated on-the-fly and the protein conformation will be optimized such that the calculated curve fits the experimental curve. Since some `mdp` options require modifications, we create a local copy of the `mdp` file along with a local copy of the topology:

```
$ cp ../md/md.mdp saxsmd.mdp
$ cp ../md/swaxs/waxs.top saxs.top
```

All `mdp` parameters relevant for SAXS-driven MD simulation are provided in the Appendix. The most important parameters are:

- `waxs-tau = 250`

This parameter τ controls the memory time used for on-the-fly averaging of the SWAXS curve during the simulation. Here, $\tau = 250$ ps. Hence, the SWAXS curve at time t of the simulation depends not only on the present simulation frame but also on frames of approximately the previous $1-2\tau$ (with exponentially decaying weights). In addition, within $0 < t < \tau$, all forces are set to zero, and the forces are gradually switched on during $\tau < t < 2\tau$. This protocol ensures that the computed SWAXS curve is converged before any SWAXS-derived forces are applied.

- `waxs-fc = 1`

This sets the force constant for coupling the calculated to the experimental curve. Values around between 1 and 5 are often suitable. We recommend to try a few force constants and to detect, for your system, a force constant that leads to a reliable conformational transition within reasonable simulation time while not perturbing the integrity of the protein. A force constant of 1 allows the interpretation of the refined ensemble as a *posterior* distribution in Bayesian statistics [6], providing rigorous confidence intervals for SAXS-refined structures.

- `waxs-nstcalc = 250`

The number of integration steps (and time) after which the SWAXS curve is updated. A common update time step is 0.5 ps (250 steps with a 2 fs integration time step).

- `waxs-nq = 20`

The number of q -points used for the coupling. A reasonable choice is at least 1.5 times the number of Shannon channels of the SWAXS curve.

- `waxs-startq = 0.2`
`waxs-endq = 4.2`

The first and last q -point to be calculated should be present in the experimental curve.

With these parameters, the run-input `tpr` file for the SAXS-driven simulation is generated.

```
$ gmx grompp -f saxsmd.mdp -p saxs.top \  
-c ../md/md.gro [-n index.ndx] -o saxsmd.tpr
```

The q -values for the experimental SAXS curve must be provided as $q = 4\pi\lambda^{-1} \sin \theta$, where λ is the wavelength of the X-ray beam, and 2θ is the scattering angle. Further, Gromacs standard units for q must be used, that is, nm^{-1} . If the experimental curve is provided in inverse Å, the command line tool `awk` may be used to do the conversion:

```
$ awk '{print $1 * 10, $2, $3}' experimental_data.dat > Iq_target.svg
```

Note that experimental data are at times reported as a function of $s = 2\lambda^{-1} \sin \theta$, hence requiring another factor 2π in the conversion.

Now that the run-input files, the envelope, and the experimental curve with the correct q units are set up, the SAXS-driven MD simulation can be started.

```
$ gmx mdrun -sw ../../waterbox/waterbox.tpr \  
-fw ../../waterbox/water-b1000-dt10.xtc \  
-is Iq_target.svg -deffnm saxsmd
```

On a reasonably modern compute node with one Nvidia Pascal GPU and six CPU cores, a 10-ns SAXS-driven MD simulation of TupA takes 5–6 h. Approximately 90% of the computational effort is spent for the normal MD simulation whereas ~10% is spent for the calculation of the SAXS curve and the SAXS-derived forces. Here, also the SAXS calculations have been ported to CUDA, hence they greatly benefit from a GPU.

As shown in Fig. 2, TupA carried out an opening transition during the SAXS-driven MD simulation as evident from the structure (Fig. 2, top) and from the increased center-of-mass distance between both domains (Fig. 2, bottom). Hence, the TupA apo

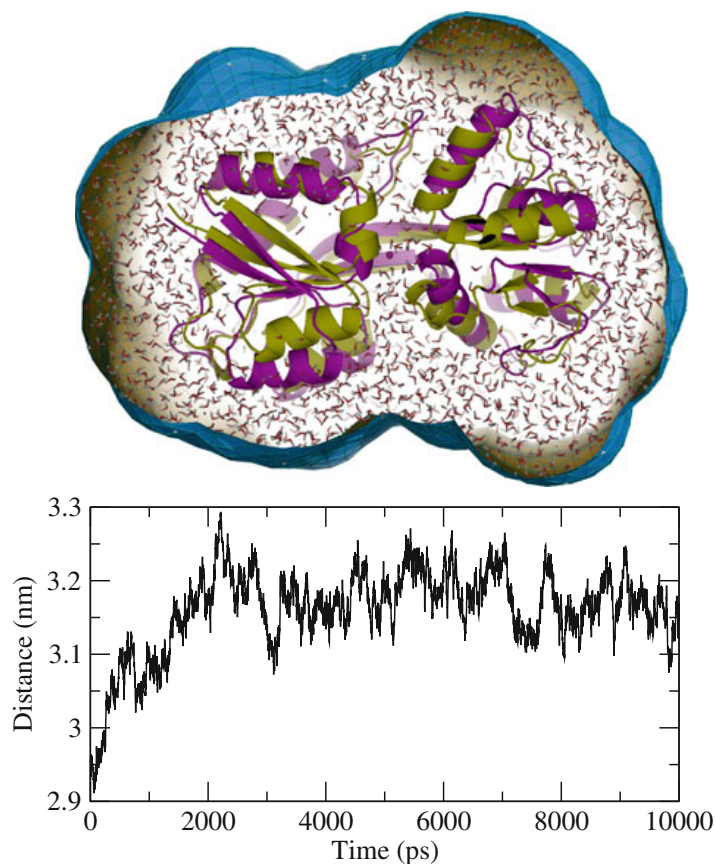


Fig. 2 Top: overlay of apo TupA crystal structure (olive) and the optimized TupA structure after SAXS-driven MD (magenta). Water molecules inside the envelope contributed to the SAXS calculations. Water molecules outside the box are not shown for clarity. Bottom: increasing center-of-mass distance between the two domains during SAXS-driven MD, demonstrating an opening transition within the first 2 ns

crystal structure is too compact to match the SAXS curve, in line with the findings reported previously [12].

The SAXS-driven simulation writes several output files. The output SAXS curve in the file `waxsm�_final.xvg` reports the final on-the-fly calculated SAXS curve, hence representing purely the last part of the simulations (last $1-2\tau$). To obtain the average SAXS curve of the refined structural ensemble, a SAXS prediction calculation following the first part of this tutorial can be used. To this end, a new folder `swaxs_rerun` is created within `waxsm�`:

```
$ mkdir swaxs_rerun; cd swaxs_rerun
```

Because the conformational transition occurred within the first 2 ns of a 10 ns trajectory (*see* Fig. 2, bottom), we instruct `mdrun` to ignore the first 2000 ps of the trajectory with:

```
$ export GMX_WAXS_BEGIN=2000
```

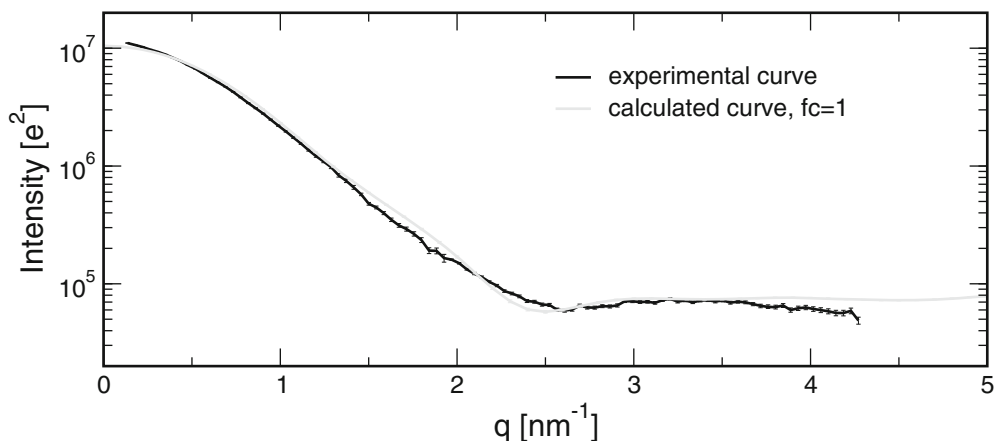


Fig. 3 SAXS curves of refined TupA obtained by a SAXS-driven MD simulation (gray) and experimental SAXS curve (black). The agreement between the experimental curve and the curve from SAXS-MD shows only small deviations

Next, `mdrun -rerun`, using the `rerun.mdp` file described in Appendix 2, computes the SAXS curve as an average over the final 8 ns of the 10 ns trajectory. The computed SAXS output curve (`waxs_final.xvg`) is compared with the experimental curve in Fig. 3. Evidently, the SAXS curve of the refined TupA reasonably agrees with the experimental curve, demonstrating that the simulation was capable of generating an ensemble compatible both with the shape of the SAXS curve and with the physical information contained in the force field. It is interesting to note, however, that R_g obtained from Guinier analysis of the refined ensemble (2.25 nm) is smaller than the experimental R_g of TupA (2.40 nm), which was increased owing to the presence of the 23 residue-long expression tag. Hence, the force field prohibited further expansion of the protein as would have been required to match the experimental R_g , and thereby precluded a misinterpretation of the SAXS curve in terms of an over-expanded TupA structure. This is an, albeit simple, example demonstrating that the simulation balances (1) the physicochemical information contained in the force field with (2) the experimental data in the SAXS curve to obtain the most plausible apo TupA model.

The SAXS-derived potential energy can be extracted from the energy output file (`edr` file) with the `gmx energy` module.

```
$ g_energy -f waxsmc.edr
```

and choosing the term “Xray-coupl” when prompted (Fig. 4). At the beginning of the simulation, the SAXS-derived potential is gradually turned on, finds a maximum at ~ 1500 ps, and relaxes as the protein opens (compare with Fig. 2, bottom). After the opening transition, the potentials fluctuate around a relatively high value of ~ 55 kJ/mol, revealing that the simulation was not capable of

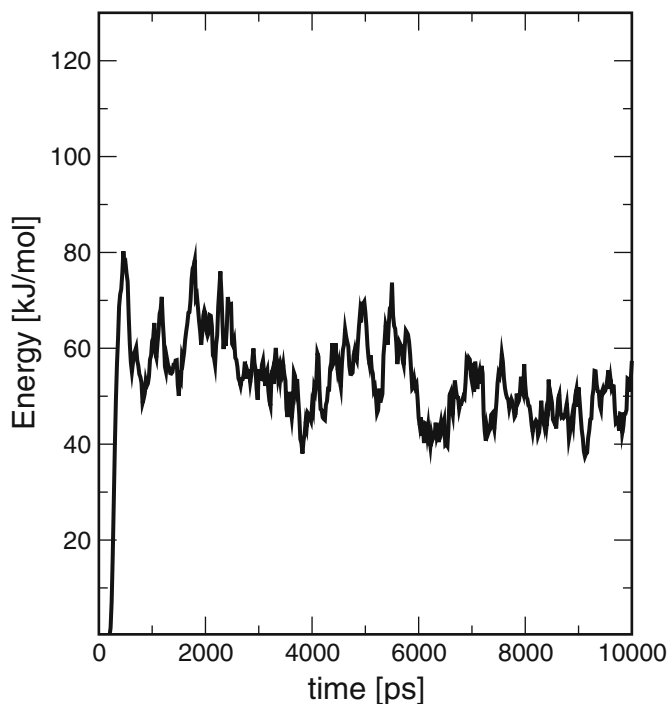


Fig. 4 SAXS-derived biasing potential that drives the simulation into conformations compatible with the data

finding a conformation that accurately matches the experimental data. This finding is rationalized by the increased R_g owing to the presence of the expression tag.

4 Conclusions and Outlook

This tutorial demonstrates the application of MD simulations for deriving an atomic interpretation of SWAXS data. We presented the back-calculation of a SAXS curve from a given MD trajectory, followed by structure refinement using SAXS-driven MD simulations. The explicit-solvent SWAXS calculations employed accurate models for the hydration layer and for the excluded solvent into account, thereby avoiding any solvent-related fitting parameters. Further, we emphasized how the physicochemical information in the MD force field complements the experimental data, thereby reducing the risk of overfitting the data. More recent developments, not discussed in this chapter, include the interpretation of small-angle neutron scattering (SANS) data by means of explicit-solvent MD [13] as well as the development of a Bayesian framework for the interpretation of SWAXS data [14]. In addition, the SAXS-driven MD simulations have been extended to carry out maximum-entropy ensemble refinement by coupling parallel-replica simulations to a SAXS curve [15].

5 Appendix 1: Common Error Messages and Problems

In this section, we summarize the most common problems and error messages, and we outline possible solutions.

Error: Atom nr 1 (name MN1) is in WAXS scattering group Solute, but no Cromer-Mann parameters are defined for this atom.

Solution: Your system contains virtual atoms, which do not contribute to the X-ray scattering. Set `waxs-solute = Prot-Masses` in the `mdp` file to compute the SAXS intensity purely from the physical atoms.

Error: `grompp` reports: “No default Xray coupl. types.”

Solution: The `grompp` module found atom in the `waxs-solute` or `waxs-solvent` groups, for which no Cromer-Mann parameters were found. Maybe definitions are missing in the `scatter.itp` file, or Cromer-Mann definitions are missing in the force field folder for water or ions.

Error: The envelope does not fit into the compact unit cell.

Solution: Restart your simulation with a larger simulation box with more water.

Warning: Atom `xx` is outside of the envelope.

Solution: Your protein expanded during the simulation, for instance, owing to a close-to-open conformational transition; hence the protein no longer fits into the envelope. Build a larger envelope, *e.g.*, with `gmx genenv -d 1.5`.

Error: Some atoms are outside of the compact box...

Solution: This indicates that making the protein whole and shifting it to the box center (as required to compute the SWAXS curve) failed. Possible reasons are:

- (a) You used an unsuitable PBC atom. Specify a `waxs-pbc atom` atom closer to the center of the protein.
- (b) Your protein expands or unfolds in an unphysical manner; see next point.

Error: The simulation is unstable, or the protein is unfolding in an unexpected manner.

Solution: SAXS-derived forces might be too large. Check if the protein unfolds or if the SAXS-derived potential is too large. Possible solutions include:

- (a) Ensure that the experimental curve uses the correct q -units (nm *vs.* Å, 2π convention).
- (b) Increase `waxs-solvdens-uncert` to absorb systematic errors in the experimental curve.
- (c) Remove experimental small-angle data points that are obviously affected by aggregation or solute repulsion.
- (d) Reduce the force constant `waxs-fc`.

- (e) Use a larger `waxst-target`, for instance, 10000ps, instructing the simulation to turn on the SAXS-derived forces more slowly at the beginning of the simulation, thereby allowing the protein to gradually adapt the SAXS bias.

Error: Incorrect number of atoms in the reference coordinates used to fit to the WAXS envelope! Please provide the whole system, the solute, or the fit group only.

Solution: Make sure that `waxs-rotfit` indicates the same group as specified as the fit group in `gmx genenv`.

6 Appendix 2: MD Parameters for SAXS Curve Prediction Calculations

The SWAXS part of an `mdp` file to be used with `rerun -mdrun` for predicting a SAXS curve from a given trajectory:

```
; Required to read Cromer-Mann parameters from the topology tree
define = -DSCATTER
; Scattering type (xray/neutron) - here, we focused on X-ray
scattering
scatt-coupl = xray
; Group of the scattering solute (e.g. Protein, or Prot-Masses)
waxs-solute = Protein
; Group of the scattering solvent (e.g. Water, or Water_and_ions)
waxs-solvent = Water_and_ions
; Group used for rotational fit, as specified for gmx genenv
(e.g. C-alpha or Backbone)
waxs-rotfit = Backbone
; Atom index of solute group near the solute COM, used to make the
protein whole. See output of gmx genenv for a suitable atom.
waxs-pbcatom = 102
; Coupling time constant (ps). -1 means uniform average, as
common for computing a SWAXS curve from a given trajectory with
mdrun -rerun
waxs-tau = -1
; Frequency of I(q) calculation (steps). Use 1 for mdrun -rerun
waxs-nstcalc = 1
; Number of pure solvent frames used for solvent scattering
waxs-nfrsolvent = 1000
; Frequency of writing updated I(q) to wax_spectra.xvg
waxs-nstlog = 1
; Nr of q values, starting and ending q in inverse nanometer
waxs-nq = 101
waxs-startq = 0
waxs-endq = 10
; Nr of points used for spherical quadrature. Use at least ~0.2*
```

```

(D*waxs-endq)^2, where D is the protein diameter. Better use 3-4
times more.
waxs-nsphere = 1000
; Electron density of solvent in e/nm3 (0 = no correction, 334 =
experiment), used to correct small deviations of the solvent
density.
waxs-solvdens = 334
; Correct I(q=0) to a given value, by adding a very small density
to the pure-solvent system. This may speed up the convergence of
I(q) in systems with a small contrast between solute and solvent,
such as a small disordered peptide.
waxs-scale-i0 = no

```

7 Appendix 3: SWAXS Part of a SAXS-Driven MD Simulation

Here, we list additional `mdp` options that are relevant or specific for SAXS-driven simulations. Parameters not shown here may be chosen following Appendix 2.

```

; Coupling time constant (ps), indicating the memory time used
for on-the-fly updates of I(q). Reasonable values are between
50 and 500ps.
waxs-tau = 250
; Force constant for SAXS-derived forces. 1 is often fine, but
larger values may be required in certain cases to enforce a good
match with the experimental data.
waxs-fc = 3
; Turn on SAXS-derived forces gradually within this time (ps).
This allows the biomolecule to gradually adapt to the SAXS
restraints, avoiding strong, sudden forces.
waxs-t-target = 2000
; Frequency of I(q) calculation (steps). A value corresponding to
0.5ps is recommended.
waxs-nstcalc = 250
; Nr of q values. A reasonable value is 1.5 times the number of
Shannon channels.
waxs-nq = 30
waxs-startq = 0.2
waxs-endq = 4.2
; Specify how to fit the experimental curve I[exp] to the
calculated curve via I[fit]=f*I[exp]+c: none (f=1,c=0), scale
(c=0), or scale-and-offset. The parameter f adjusts the overall
scale of I(q), and c may absorb an uncertainty in the buffer
subtraction. In doubt, first try with "scale".
waxs-Iexp-fit = scale

```


; Uncertainty of the solvent density. This translates into a systematic uncertainty of the calculated SWAXS curve, denoted $\sigma[\text{buf}]$ in previous work. A reasonable value is between 0.1% and 0.5%. In practice, this parameter reduces the weights of the smallest-angle data points.

waxs-solvdens-uncert = 0.001

References

1. Rambo RP, Tainer JA (2013) Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu Rev Biophys* 42:415–441
2. Tuukkanen AT, Spilotros A, Svergun DI (2017) Progress in small-angle scattering from biological solutions at high-brilliance synchrotrons. *IUCrJ* 4:518–528
3. Jeffries C, Trehwella J (2013) Small-angle scattering. In: Wall ME (ed) *Quantitative biology: from molecular to cellular systems*. CRC Press, Boca Raton, pp 113–151
4. Bizien T, Durand D, Roblina P *et al* (2016) A brief survey of state-of-the-art BioSAXS. *Protein Pept Lett* 23:217–231
5. Hub JS (2018) Interpreting solution X-ray scattering data using molecular simulations. *Curr Opin Struct Biol* 49:18–26
6. Shevchuk R, Hub JS (2017) Bayesian refinement of protein structures and ensembles against SAXS data using molecular dynamics. *PLoS Comput Biol* 13:e1005800
7. Ivanović MT, Bruetzel LK, Lipfert J, Hub JS (2018) Temperature-dependent atomic models of detergent micelles refined against small-angle X-ray scattering data. *Angew Chem Int Ed* 57:5635–5639
8. Chen P, Hub JS (2014) Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophys J* 107:435–447
9. Chen P, Hub JS (2015) Structural properties of protein–detergent complexes from SAXS and MD simulations. *J Phys Chem Lett* 6:5116–5121
10. Chen P, Hub JS (2015) Interpretation of solution X-ray scattering by explicit-solvent molecular dynamics. *Biophys J* 108:2573–2584
11. Knight CJ, Hub JS (2015) WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res* 43:W225–W230
12. Otrelo-Cardoso AR, Nair RR, Correia MAS *et al* (2017) Highly selective tungstate transporter protein TupA from *Desulfovibrio alaskensis* G20. *Sci Rep* 7:5798–5798
13. Chen P, Shevchuk R, Strnad F, Lorenz C, Karge L, Gilles R, Stadler A, Hennig J, Hub JS (2019) Combined small-angle X-ray and neutron scattering restraints in molecular dynamics simulations. *J Chem Theory Comput* 15:84687–84698
14. Shevchuk R, Hub JS (2017) Bayesian refinement of protein structures and ensembles against SAXS data using molecular dynamics. *PLoS Comp Biol* 13:e1005800
15. Hermann MR, Hub JS (2019) SAXS-restrained ensemble simulations of intrinsically disordered proteins with commitment to the principle of maximum entropy. *J Chem Theory Comput* 15(9):5103–5115



Determining the Free Energies of Outer Membrane Proteins in Lipid Bilayers

Gerard H. M. Huysmans, Dagan C. Marx, Sheena E. Radford, and Karen G. Fleming

Abstract

The thermodynamic stabilities of membrane proteins are of fundamental interest to provide a biophysical description of their structure-function relationships because energy determines conformational populations. In addition, structure-energy relationships can be exploited in membrane protein design and in synthetic biology. To determine the thermodynamic stability of a membrane protein, it is not sufficient to be able to unfold and refold the molecule: establishing path independence of this reaction is essential. Here we describe the procedures required to measure and verify path independence for the folding of outer membrane proteins in large unilamellar vesicles.

Key words Outer membrane protein, PagP, Large unilamellar vesicles, Folding, Stability, Equilibrium, Thermodynamics, Path independence

1 Introduction

Understanding the balance of the physical forces that determine the thermodynamic stability and dynamic properties of a protein holds the key to understanding function, predicting structure, and enabling rational design. Protein stabilities are determined by measuring the change in free energy between the unfolded and folded conformations [1]. For outer membrane proteins (OMPs), a hydrophobic cosolvent is required to enable the proteins to fold, and the closest, most informative mimics for biological membranes are large unilamellar vesicles (LUVs) of defined composition.

Free energy change (ΔG^0) is a state function: its value is only determined by the thermodynamic equilibrium of the system under study, not by the path by which the system reached that equilibrium. Finding conditions under which OMP folding and unfolding are path independent remains a challenge because both the folded and unfolded conformations must be simultaneously populated at

levels high enough to measure their concentrations. This condition is compromised by the propensity of membrane proteins to aggregate in aqueous solution. Denaturant titrations are the standard strategy employed to populate the folded and unfolded conformations simultaneously, and the equilibrium free energy change is extracted from the response of the population distribution of the protein at every point in the denaturant gradients in the unfolding and refolding directions [1]. These unfolding and refolding curves must overlay exactly upon each other to meet the condition of thermodynamic reversibility. If this is not the case, either one or the other or both of the curves represent a delayed response by the protein to the change in denaturant concentration. This is a phenomenon called hysteresis and has been interpreted in terms of a rugged, free energy landscape on which proteins get trapped, sometimes for very long times, in local minima [2].

We provide a protocol to determine the equilibrium stabilities and use PagP as an example OMP as we have achieved reversible path-independent conditions for this protein using both protocols discussed here [3–5]. PagP is a membrane-embedded palmitoyl transferase that acts in the stress response mechanism to reinforce the bacterial outer membrane [6]. PagP folds into an eight-stranded β -barrel, preceded by an N-terminal α -helix [7]. The PagP β -barrel is tilted approximately 25° with respect to the membrane normal and has been proposed to have a lateral gate that aligns with the outer leaflet of the membrane to provide lipid access to the catalytic site [8].

Reversibility of OMP refolding into lipid vesicles from a completely unfolded OMP-conformation without hysteresis has only been reported for β -barrel proteins from the bacterial outer membrane [3–5, 9–12]. In addition, partial unfolding and refolding without hysteresis within lipid vesicles has been achieved for the α -helical transmembrane transporter LeuT [13].

2 Materials

2.1 Reagents for Refolding and Unfolding Experiments

1. Buffer stock solutions (sodium citrate for buffers below pH 6, glycine or Tris for buffers above pH 9, and sodium phosphate or glycyl-glycine in the range of pH 7 and pH 8).
2. Guanidinium-HCl (ultrapure), urea (ultrapure), ethylenediaminetetraacetic acid (EDTA), 3-(*N,N*-dimethyl-myristylammonio) propanesulfonate (SB3-14).
3. 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (*diC*_{12:0}PC, [850335C]), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (*diC*_{14:0}PC, [850345C]), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (*diC*_{18:1}PC, [850375C]), 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (*diC*_{18:1}PE, [850725C]) [Avanti Polar Lipid catalogue numbers are given between brackets].

2.2 Equipment

1. *Essential equipment*: a water bath; an N₂ or Ar₂ cylinder with pressure regulator and adaptor tubing to hold a glass Pasteur pipette; a desiccator connected to a vacuum line/vacuum pump; an Avanti miniextruder with 1 ml syringes, polycarbonate membranes with a pore size of 100 nm and filter supports; a cell disruptor working at 10,000–15,000 psi or a sonicator; a heated stir plate; a fluorimeter; UV-spectrometer; a rotating incubator.
2. *Recommended equipment*: an electrophoresis chamber with power supply and house-made or commercially purchased polyacrylamide gels; a circular dichroism spectrometer.

3 Methods

The thermodynamics of OMP folding is determined by measuring the free energy change between the folded and unfolded states of OMPs [1]. Hence, key prior knowledge of the experimental signatures of the refolded and unfolded conformations of an OMP is required for setting the endpoint observables for these experiments. These are typically obtained by determining the spectral properties of chromogenic amino acids, like tryptophan (Trp), or the secondary structure content by circular dichroism spectroscopy. Both methods can be used to document the structural change between refolded and unfolded OMPs. However, conformational transitions are most commonly detected through changes in solvent accessibility of tryptophan (Trp) as detected using fluorescence intensity because of its high sensitivity, the possibility to access a wide range of protein concentrations and buffer conditions, and its ability to focus on specific sites [14].

3.1 Fluorescence Parameters and Their Usage

The technical considerations when collecting Trp-fluorescence emission spectra are discussed in detail by Fleming and coworkers [15]. Here, we highlight some key features of the spectral properties of OMPs that are important to enable taking and interpreting accurate fluorescence spectra.

Figure 1 shows characteristic Trp-fluorescence emission spectra for unfolded and refolded PagP. Trp-fluorescence emission occurs between 300 and 400 nm upon excitation at 295 nm. Excitation at 280 nm will also excite tyrosine side chains, resulting in higher total intensities, but this practice could complicate the spectral response of the refolding and unfolding titrations. For this reason, direct excitation of Trp at 295 nm is the most applied method of choice. The excitation light scattering peak (gray dots) partially overlaps with the fluorescence emission band. Introducing polarizers can reduce the contribution of light scattering to the fluorescence spectrum at the expense of total signal [16]. Huysmans *et al.* [3]

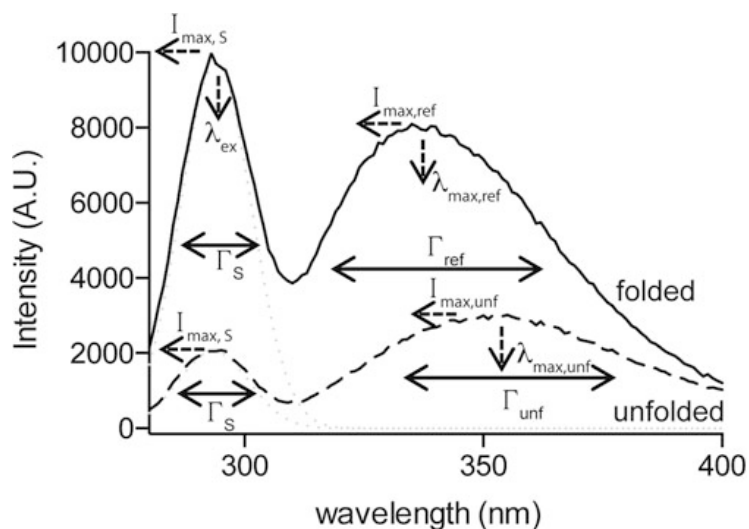


Fig. 1 Example tryptophan emission spectra for OMPs in folding experiments. Depicted are two emission spectra upon excitation 295 nm (λ_{ex}): “refolded” PagP (full line) and “unfolded” PagP (dashed line). The data were collected with 0.4 μ M PagP, *diC*_{12:0}PC LUVs at an LPR of 2000:1, pH 3.8, and in the presence of 2 and 6 M Gdn-HCl, respectively. Indicated are the maximum intensity for both the scattering ($I_{max,S}$) and tryptophan fluorescence ($I_{max,ref/unf}$), wavelength of maximum emission Trp fluorescence ($\lambda_{max,ref/unf}$), and peak width at half maximum intensity for scattering (Γ_S) and Trp fluorescence (Γ). These parameters can be accurately determined by fitting to a sum of a normal distribution (scattering peak, gray dotted line) and a log-normal distribution (Trp fluorescence)

subtracted buffered liposome spectra to correct for scattering contributions. Moon *et al.* [4] and Marx and Fleming [5] collected the light scattering peak around 295 nm and subtracted its contribution to the Trp-fluorescence spectrum by curve fitting. Unfolded PagP has a broad (large Γ_{unf}) Trp-fluorescence emission spectrum with low intensity ($I_{max,unf}$) and a maximum around 350 nm ($\lambda_{max,unf}$) (Fig. 1, dashed line), indicative of Trp residues that are exposed to the aqueous solvent. The spectrum increases in intensity ($I_{max,ref}$), sharpens (smaller Γ_{ref}), and becomes blue shifted to a maximum around 335 nm ($\lambda_{max,ref}$) when the Trp residues are buried in the hydrophobic environment of the *diC*_{12:0}PC-bilayer upon refolding (Fig. 1, full line).

The goal of acquiring Trp fluorescence spectra is to extract the fractional contribution of the folded and unfolded OMP to the fluorescence signal. It is important to note that not all spectral parameters vary linearly with the fraction of refolded or unfolded OMP. Ladokhin *et al.* [16] showed that neither the λ_{max} nor Γ varies linearly with the folded fraction. The only spectral properties that vary linearly with the fraction of folded OMP are the intensity at a chosen wavelength ($I(\lambda)$), usually the $I(\lambda_{max})$ of the refolded OMP, and the average wavelength, if it is corrected for the

difference in total intensity between the folded and unfolded OMP over the spectrum (*see* Subheading 3.4.3). The average wavelength ($\langle \lambda \rangle$) is calculated as $\langle \lambda \rangle = \frac{\sum I(\lambda) \cdot \lambda}{\sum I(\lambda)}$ in which $I(\lambda)$ is the fluorescence intensity at wavelength λ .

3.2 Preparation of LUVs

The proper preparation of LUVs is critical for the success of OMP-folding studies. We recommend purchasing the lipids in solvent from Avanti Polar Lipids (most lipids will be delivered in chloroform). The most widely used lipid compositions are *di*C_{12:0}PC, *di*C_{14:0}PC, and *di*C_{18:1}PC/*di*C_{18:1}PE (90/10) [3–5, 9–12]. Glass test tubes should be used, and they should be cleaned extensively with a saturating KOH solution or detergent/10% HCl solution before use. Lipid quantities are exclusively measured using glass pipettes or Hamilton syringes; plasticware should not be used at any stage where organic solvents are present. Lipids normally arrive in sealed containers under inert gas (*e.g.*, N₂). Upon breaking the vials, unused lipid should be aliquoted, dried into a thin film in glass tubes under inert gas, covered with inert gas, and stored at –80 °C to minimize lipid oxidation.

The amounts to prepare and the buffers to use vary with respect to the procedure of choice as detailed in Subheadings 3.4.1 and 3.4.2. The formation of a thin lipid film without precipitating the lipid during solvent evaporation is key to preparing LUVs of good quality. A thin film is obtained by rotating the glass vial while evaporating the solvent under a gentle stream of N₂ gas. Turbid film formation indicates that the lipid has precipitated during chloroform evaporation. This is prevented by warming the lipids during solvent evaporation in a water bath above the melting temperature of the lipid. Keeping the water bath at 37 °C is appropriate for typical mixtures. Residual solvent is evaporated overnight in a desiccator. The lipid film is hydrated at room temperature for 30 min upon the addition of the appropriate buffer. Applying a number (*e.g.*, three) of freeze-thaw cycles by alternating immersion of the test tube in liquid nitrogen and room temperature water may aid lipid hydration. Hydration will result in the formation of multilamellar vesicles of varying size. LUVs are formed by extrusion through polycarbonate membranes with a pore size of 100 nm (or other size, if required) using an Avanti miniextruder. Prewet two filter supports and one polycarbonate membrane in buffer and assemble the chamber of the miniextruder following the manufacturer's instructions. LUVs are obtained by extruding the liposome solution a minimum of 11 times. It is always useful to confirm the size distribution of the LUVs formed, for example, using dynamic light scattering [13, 17].

3.3 Preparation of OMP Stock

Folding experiments require milligram amounts of protein. We have found that a successful way to produce large quantities of OMP is to overexpress the OMP into inclusion bodies in *E. coli*.

To achieve this, the OMP open reading frame must be cloned without its signal sequence into a T7-inducible plasmid and the codons for an initiator methionine appended at the beginning of the open reading frame. Inclusion bodies are rapidly isolated from cell lysates by centrifugation (*e.g.*, 30 min at $25,000 \times g$). Typically, protocols include washing the inclusion bodies with detergent (we used 2% (v/v) Triton X-100 [3] or 0.1% (v/v) Brij-L23 [5]) to solubilize contaminating membranes. This procedure provides sufficiently pure protein used by Fleming and coworkers [4, 5]. Huysmans *et al.* [3] used a C-terminal His₆ affinity tag to provide higher purity in a trade-off with lower yields. In this case, PagP was purified from inclusion bodies by affinity chromatography under denaturing conditions in buffered 6 M guanidinium-HCl (Gdn-HCl) [3]. Purified OMP inclusion bodies or purified OMP, precipitated by dialysis after affinity chromatography, can be stored frozen at $-80\text{ }^{\circ}\text{C}$ for up to 6 months.

3.4 Titration Protocols

The many thermodynamic protein folding experiments conducted over the past 50 years make extensive use of two different chemical denaturants that enable simultaneous observation of the folded and denatured populations [18]. These are urea and guanidinium-HCl. Gdn-HCl is the “stronger” of the two denaturants, and it is empirically observed that the concentration required to denature soluble proteins is two- to threefold lower compared to urea [1]. The origins of this effect are not well understood, but it is worth recognizing that urea is a neutral molecule whereas Gdn-HCl is a salt and carries a charge. Another difference is urea’s reactivity with proteins. Incubation and storage of proteins in urea solutions can lead to carbamylation of primary amine moieties on lysine side chains. Both denaturants have been used to determine the thermodynamic stability of the OMP PagP [3–5], and we compare both methods here.

Two key experiments must be carried out to meet the requirements for a thermodynamic equilibrium. These are referred to as (1) a folding titration and (2) an unfolding titration. The protocols in Subheadings 3.4.1 and 3.4.2 were optimized to be free from hysteresis. We provide ways to troubleshoot the procedures in case hysteresis is observed in Subheading 3.5.

3.4.1 Urea Method

The workflow of the urea method is illustrated in Fig. 2a. All steps are executed at $25\text{ }^{\circ}\text{C}$. The required reagents for this method are 0.5 M sodium phosphate (pH 8.0), 6 M Gdn-HCl (in 50 mM sodium phosphate (pH 8.0)), 10 M urea (in 50 mM sodium phosphate (pH 8.0)), 11.5 M urea (in 50 mM sodium phosphate (pH 8.0)); solubilization requires gentle heating), 20 and 40 mM *di*C_{12:0}PC LUVs, and precipitated PagP. Ideally, urea solutions are prepared in volumetric flasks, and urea concentrations should be verified by refractometry for accuracy. Typical volumes are given throughout.

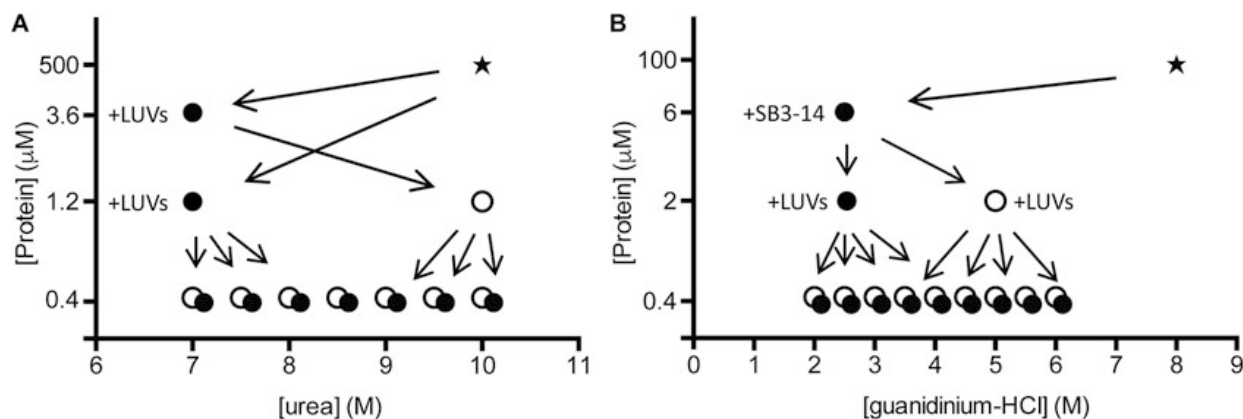


Fig. 2 Dilution scheme for reversible PagP folding into $d/C_{12:0}PC$ LUVs. **(a)** PagP is dissolved at 500 μM in 6 M guanidinium HCl (star). PagP is then diluted to 3.6 μM or 1.2 μM in the presence of 7 M urea and LUVs to refold PagP. Next, 1.2 μM PagP is diluted threefold to 0.4 μM into a range of urea concentrations to obtain an unfolding titration curve. Refolded PagP at 3.6 μM is first diluted threefold to unfold PagP in urea and further diluted to 0.4 μM to obtain a refolding titration curve. **(b)** PagP, dissolved to a concentration under 100 μM in 8 M guanidinium-HCl (star), is diluted to 6 μM in the presence of SB3-14 detergent slightly above its critical micelle concentration and 2.5 M guanidinium-HCl. PagP is then diluted threefold to 2 μM in the presence of LUVs. For refolding experiments PagP is diluted into 5 M guanidinium-HCl; for unfolding experiments PagP is diluted into 1 M guanidinium-HCl. Finally, PagP is diluted fivefold to a final concentration of 0.4 μM at a range of guanidinium-HCl concentrations to obtain a titration curve. Unfolding is represented by open circles, refolding by closed circles

Preparation of the Unfolded PagP Stock

1. Solubilize precipitated PagP in 6 M Gdn-HCl to a concentration of 0.5 mM. Remove PagP aggregates by centrifugation at $25,000 \times g$ for 20 min. Determine the exact concentration by measuring the absorbance at 280 nm [19].

Preparation of the Unfolding Titration

1. Make 3 ml refolded PagP by diluting Gdn-HCl-solubilized PagP to 1.2 μM PagP into 50 mM sodium phosphate buffer (pH 8.0) supplemented with liposomes at a final LPR of 3200:1 (3.85 mM final LUV concentration, diluted from the 20 mM LUV stock) and 7 M urea (diluted from the 10 M urea stock) to prevent PagP aggregation.
2. Incubate this refolding reaction for at least 8 h.
3. The unfolding titration is obtained by diluting the refolded PagP to 0.4 μM into 500 μl of the same buffer with final urea concentrations between 7 and 10 M with 0.2 M urea intervals.
4. Incubate the unfolding reaction for at least 8 h.

Preparation of the Refolding Titration

1. Prepare 1 ml of refolded PagP per titration by diluting 0.5 mM Gdn-HCl-solubilized PagP to 3.6 μM into 50 mM sodium phosphate buffer (pH 8.0) supplemented with liposomes at an LPR of 3200:1 mol/mol (11.55 mM final LUV concentration, diluted from a 40 mM LUV stock) and 7 M urea (diluted from the 10 M urea stock) to prevent PagP aggregation.
2. Incubate the mixtures for at least 8 h.

3. Make 3 ml unfolded PagP stock containing 1.2 μM PagP by unfolding 1 ml refolded PagP in the presence of 10 M urea in 50 mM sodium phosphate, pH 8.0.
4. Incubate the unfolding reactions for at least 8 h.
5. The refolding titration is obtained by diluting the unfolded PagP stock into 500 μl 50 mM sodium phosphate, pH 8.0, with final urea concentrations between 7 and 10 M with 0.2 M urea intervals.
6. Incubate the refolding reactions for at least 8 h.

Preparation of Buffer Blanks

1. Prepare 500 μl containing 50 mM sodium phosphate buffer (pH 8.0), 1.28 mM LUV, and 7–10 M urea with 0.2 M intervals. These will provide background spectra to be subtracted from the OMP spectra.
2. Incubate the reactions for at least 8 h as for refolding and unfolding reactions.

Acquisition and Analysis of Trp-Fluorescence Spectra

1. Warm the lamp for 15 min and heat the sample chamber to 25 $^{\circ}\text{C}$.
2. Set the excitation and emission slit widths to 2 mm and the integration time to 1 s. The bandwidth is 1 nm and the excitation wavelength 280 nm. Spectra are recorded between 300 and 400 nm.
3. Clean a 500 μl quartz cuvette (10 mm pathlength) with an appropriate detergent like Hellmanex (2%), rinse with deionized water and ethanol, and blow dry with air or an inert gas, like N_2 .
4. Accurate Trp-fluorescence emission spectra are obtained by averaging three spectra for each refolding, unfolding, and buffer reaction.
5. Export the spectra in ASCII format for data processing.
6. Subtract buffer spectra from OMP spectra in a data spreadsheet (*e.g.*, Microsoft Excel).
7. At each denaturant concentration, path-independent data will result in spectra from the unfolding curves that overlay with those obtained from the refolding curves.
8. Calculate the average wavelength (*see* Subheading 3.1) to construct the equilibrium refolding and unfolding curve (Fig. 3a).

3.4.2 Guanidinium-HCl Method

The workflow of the Gdn-HCl method is illustrated in Fig. 2b. The required reagents for this method are 100 mM citrate supplemented with 1 mM EDTA (pH 3.8), 10 mM SB3-14 detergent (3-(*N,N*-dimethyl-myristylammonio) propanesulfonate), 8 M Gdn-HCl (in 100 mM citrate supplemented with 1 mM EDTA (pH 3.8)), 40.2 mM *diC*_{12:0}PC LUVs, and PagP inclusion bodies.

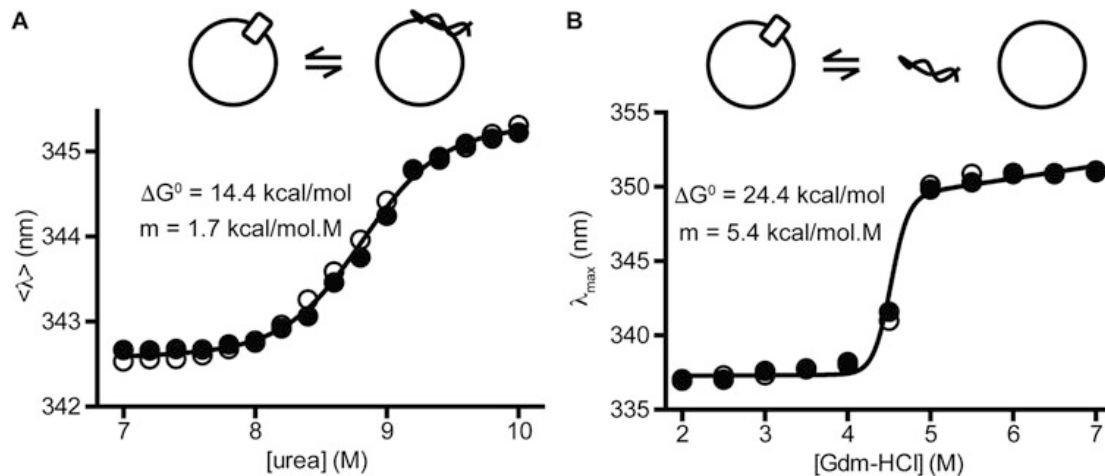


Fig. 3 Overlay of equilibrium unfolding and refolding of PagP into *d/C*_{12:0}PC LUVs following the urea method (left) and the guanidinium-HCl method (right). The cartoon on top of each graph represents the end states of the titrations: the urea method measures the free energy between unfolded LUV-adsorbed PagP and folded PagP while the guanidinium-HCl method measures the free energy between unfolded PagP free in solution and folded PagP. Curves represent fits to a two-state transition

Preparation of the Unfolded PagP Stock

1. Resuspend inclusion bodies in 100 mM citrate containing 1 mM EDTA (pH 3.8) and aliquot this into ten tubes for a 500 mL growth.
2. Collect inclusion bodies by centrifugation in a tabletop centrifuge for 30 s at $14,000 \times g$. Store at -80°C .
3. Resuspend one tube of inclusion bodies in 1 ml 8 M Gdn-HCl (in 100 mM citrate, 1 mM EDTA (pH 3.8)).
4. Remove contaminating nucleic acids by centrifugation for 20 min at $14,000 \times g$ in a tabletop centrifuge when the inclusion bodies are completely dissolved.
5. Pass the solubilized inclusion bodies through 0.22 μm filters.
6. Determine the exact concentration by measuring the absorbance at 280 nm [19] and bring the concentration to 100 μM with 8 M Gdn-HCl (in 100 mM citrate, 1 mM EDTA (pH 3.8)).

Preparation of the Unfolding Titration

1. Heat a stir plate (400 rpm) to 42°C . Add a stir bar to two glass vials.
2. Make 2.5 ml unfolded PagP stock by diluting 100 μM PagP dropwise to a concentration of 6 μM into 2.5 M Gdn-HCl containing 1.4 mM SB3-14, 100 mM citrate, and 1 mM EDTA (pH 3.8). SB3-14 acts as a holdase to prevent PagP aggregation.
3. This step should follow immediately after **step 2** to avoid aggregation. Prepare 3.3 ml refolded PagP by slowly adding (1 drop every 5–10 s) unfolded PagP from **step 2** to a concentration of 2 μM into 4 mM LUV, 2.5 M Gdn-HCl, 0.47 mM SB3-14, 100 mM citrate, 1 mM EDTA, pH 3.8.

4. *Note:* in **steps 1** and **2** the stir speed and the slow dropwise addition are critical to avoid aggregation.
5. Equilibrate the reaction overnight in a rotating incubator (37 °C, 6 rpm).
6. Dilute refolded PagP 1:5 into final Gdn-HCl concentrations between 2 and 6 M with 0.5 M intervals. Typically, 1.1 ml per reaction is needed; this will depend on the volume needed for the fluorimeter and cuvette being used.
7. Return all titration samples to the rotating incubator for at least 40 h to reach equilibrium.

Preparation
of the Refolding Titration

1. Heat a stir plate (400 rpm) to 42 °C. Add stir bars to two glass vials.
2. Make 2.5 ml unfolded PagP stock by diluting 100 μM PagP dropwise to a concentration of 6 μM into 2.5 M Gdn-HCl containing 1.4 mM SB3-14, 100 mM citrate, and 1 mM EDTA (pH 3.8).
3. This step should follow immediately after **step 2** to avoid aggregation. Prepare 3.3 ml unfolded PagP by slowly adding (1 drop every 5–10 s) PagP to a concentration of 2 μM into 4 mM LUV, 5 M Gdn-HCl, 0.47 mM SB3-14, 100 mM citrate, 1 mM EDTA, pH 3.8.
4. *Note:* in **steps 1** and **2** the stir speed and the slow dropwise addition are critical to avoid aggregation.
5. Equilibrate the reaction overnight in a rotating incubator (37 °C, 6 rpm).
6. Dilute refolded PagP 1:5 into final Gdn-HCl concentrations between 2 and 6 M with 0.5 M intervals. Typically, 1.1 ml per reaction is needed (see unfolding titration).
7. Return all titration samples to the rotating incubator for 40 h to reach equilibrium.

Acquisition and Analysis
of Trp-Fluorescence Data

1. Warm the lamp for 15 min and heat the sample chamber to 37 °C.
2. To decrease light scattering from LUVs, cross-polarization is used by setting the excitation polarizer to 90° and the emission polarizer to 0°.
3. Clean a 10 mm path length quartz cuvette and use a flea stir bar to ensure good mixing of the sample.
4. Each sample is removed from the rotating incubator one at a time and added to the cuvette. The sample is then placed into the sample housing in the fluorimeter for 2 min to reach thermal equilibrium.