Combining Solution Scattering Data with Protein Folding Simulations

Tobin Sosnick, Lee Makowski, and Karl Freed

Goals

The goals of this project are to improve protein structure determination and to characterize the structural fluctuations of the ground state of the protein. The Freed/Sosnick group has developed an algorithm for predicting the structure of a protein from the amino acid sequence. The algorithm operates by minimizing the energy obtained from a statistical potential. One goal is to improve the folding algorithm by incorporating data from WAXS (Wide Angle X-ray Scattering) experiments at Argonne. The protein structure in the folding simulations will be optimized by varying the protein structure so as to minimize the sum of the energy and the fit of the simulations to the WAXS data.

Another goal of this project is to use WAXS data to study structural fluctuations of proteins in solution by averaging over an ensemble of structures from simulations of the dynamics rather than the common practice of calculating the scattering from the static crystal structure.

Progress

The publicly available program CRYSOL for calculating WAXS patterns is only available in executable form, which would make it impossible to test and especially to improve upon the myriad of assumptions and approximations inherent in this program, especially in the description of the contribution from the solvent and from the hydration layer surrounding the protein. Thus, as a first step, a program has been written to calculate the X-ray scattering of proteins in solution and to calculate the time-averaged WAXS pattern over the course of a Langevin dynamics simulation. (Details of the program are provided in the appendix.)

Other computational efforts have focused on the refinement of structure within the condensed state (using a move set involving large torsional motions that keep the structure intact) and the prediction of secondary structure in order to provide a starting model for the refinement.

As part of this collaboration, Makowski and coworkers have run experiments for systems amenable to Langevin dynamics simulations and have thereby generated an extensive data set for a variety of proteins. Of particular interest are the data for native and partially unfolded ubiquitin. In the native state, ubiquitin is a well folded protein, and its WAXS pattern exhibits no protein concentration dependence, consistent with the native state undergoing no significant ground state fluctuations (Fig. 1, left). However, a variant has a partially unfolded loop. This difference in flexibility is apparent in both the experimental data, which show a concentration dependence (Fig. 1, right), and theoretical data (Fig. 2). The calculated time-averaged WAXS pattern for the native state obtained using Langevin dynamic simulations with a variety of force fields agrees better with the experimental data than does the pattern calculated from the (single) crystal structure (Fig. 2). Moreover, the new program yields predictions in far better agreement with experiment than the standard program CRYSOL.

Appendix

The program requires both xyz and pdb files. The xyz files contain the Cartesian coordinates of the hydrogen atoms, whereas most pdb files do not. The pdb files are needed because xyz files do not specify the residue to which the atoms belong. The interatom distances are histogramed into 0.1 Å bins since smaller bins slow the calculation and do not improve the fit with experiment. The experiments determine the WAXS profile of the empty capillary, the pure solvent, and the protein in solution, the difference is taken to provide the "protein contribution". Computationally, this difference must be obtained by subtracting the average electron density of the solvent that is displaced by the protein[4]. This procedure is equivalent to subtracting a dummy atom scattering factor from the scattering factor of the atom, so the scattering factors become F'(S)=F(S)-G(S), where F'(S) is the solvent corrected scattering factor, F(S) is the scattering factor of the atom, and G(S) is the scattering factor of the dummy atom. Various forms for the dummy atoms have been tried in comparison with data from Makowski, e.g., for the lysozyme and for the ubiquitin example provided below.

The density of water near a protein is about 10% higher relative to the bulk in aqueous solution. Thus, when the solvent scattering is subtracted using the bulk solvent density, the contributions from the hydration shell are incorrectly represented and are treated by use of dummy hydration shell atoms. The parameters for the dummy atoms to represent the displaced solvent and the hydration layer have been used for lysozyme and will be used for all proteins.

Figures

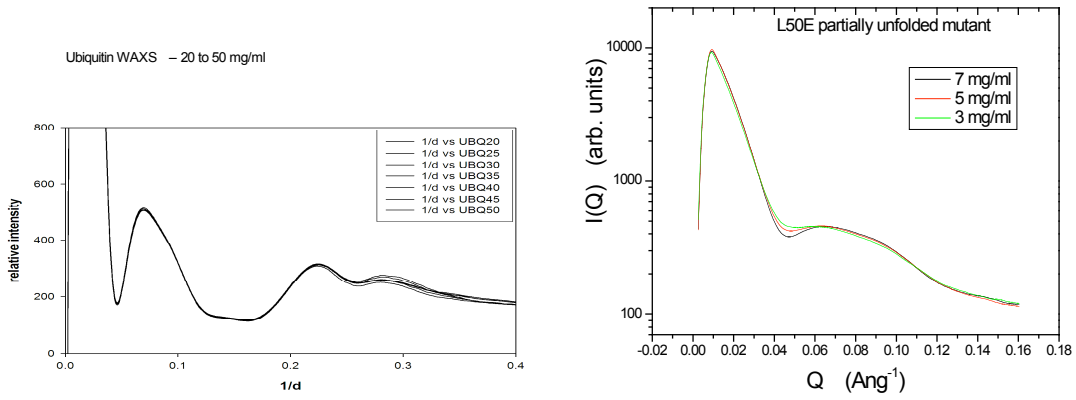

Fig. 1 Concentration dependence of the WAXS pattern for native ubiquitin (left) and a partially unfolded mutant L50E (right). The native protein has minimal concentration dependence while the mutant has motions which can be suppressed at higher concentration.

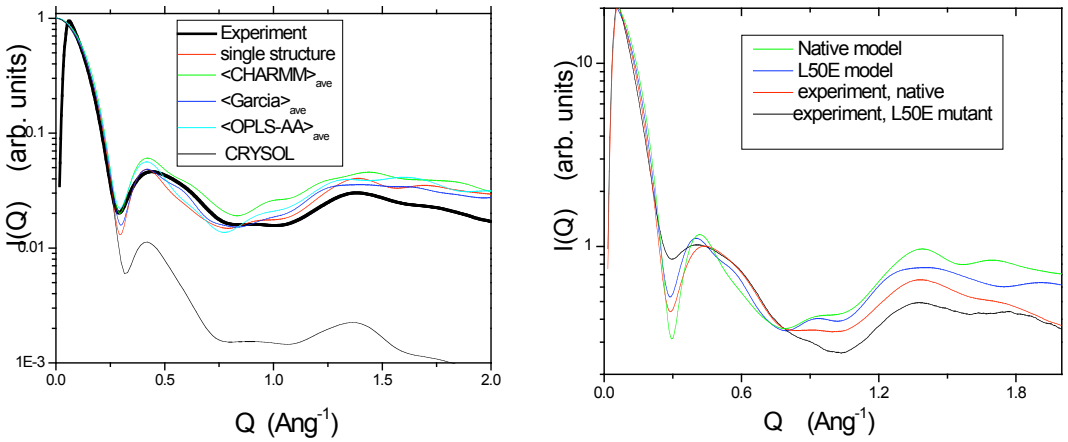

Fig. 2 Comparison between experiment data for the protein ubiquitin and calculations from models. Left: For the native state, the time-averaged WAXS pattern from Langevin dynamic simulations using a variety of force fields agrees better with the experimental data than the pattern from the static crystal structure. Our calculation agrees better with the data than the standard program CRYSOL (bottom trace). Right: Comparison between native ubiquitin and a partially unfolded L50E mutant. In both cases, the predicted pattern has more features (deeper minima at 0.25 $Å^{-1}$) than the experimental data which reflects the time-averaged structure. In addition, the pattern for the native structure shows more structure than the mutant protein, as expected.

1 A. D. MacKerrell, Jr., et al., "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins", J. Phys. Chem. B, 102, 3586-3616 (1998).

2 G. A. Kaminsky, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculationson Peptides", J. Phys. Chem. B, 105, 6474-6487 (2001)

3 Cantor R., Schimmel R., Biophysical Chemistry Part II: Techniques for the study Biological structure and function, W. H. Freeman and Company, 1980.

4 Fraser, R. D. B., MacRae, T. P. & Suzuki, E. (1978). *J. Appl. Cryst.* **11**, 693-694

5 Creighton E., Proteins Structures and Molecular Properties, Second Edition, W. H. Freeman and Company, 1993.

6 Lattman, E. E. (1989). *Proteins*, **5**, 149-155.