Can a continuum solvent model reproduce the free energy landscape of a $\beta$-hairpin folding in water?

Ruhong Zhou$^1$ and Bruce J. Berne$^2$

$^1$ IBM Thomas J. Watson Research Center; and $^2$ Department of Chemistry, Columbia University, New York, NY 10027

Proceedings of the National Academy of Science Vol. 99, No. 20

BBSI Journal Club
June 24, 2004

Presented By:
Jason Funt, Anes Aref, Jerome Nilmeier
Protein Folding/Unfolding

- **Structure/Function**, how do proteins achieve their structure?
- Experiments probe proteins at different stages of folding, elucidate kinetic mechanisms
- Computer modeling can help clarify what is unknown
Why study a $\beta$-hairpin?

- C-terminus hairpin of G protein is believed to be one of the smallest naturally occurring systems that exhibits many features of a full-size protein.
- Fast folder (folds in $\approx 6 \mu$s).
- Understanding key secondary structure elements, such as the $\beta$-hairpin or $\alpha$-helix helps in understanding much more complex proteins.

National Energy Research Scientific Computing Center
http://old-www.nersc.gov/research/annrep98/pande.html
Explicit v. Implicit Solvent Modeling

Molecular mechanics takes into account the interactions between a molecule and the solvent in which it is immersed. There are two different methods; explicit and implicit modeling.

- **Explicit solvent models:**
  - hundreds or thousands of discrete solvent molecules.
  - Slow convergence of calculations because of the number of particles involved.
  - Orders of magnitude more CPU time than corresponding gas phase calculations.
Explicit v. Implicit Solvent Modeling

Because explicit models are computationally expensive, there is a significant interest in developing the more rapid implicit solvent models.

- **Implicit Solvent Models:**
  - treat the solvent as a continuous medium, with the average properties of the real solvent
  - surround the solute beginning at the van der Waals surface.
Research Summary on this Protein

Implicit Models

- Generalized-Born (GB) continuum solvent model
- CHARMM19 with continuum solvent model EEF1

Explicit Models

- AMBER94 with a 9Å cutoff
- OPSLAA with no cutoffs (with P3ME)

References:
A surface-generalized Born (SGB) implicit model is used in conjunction with a highly parallel replica-exchange method (REM). The model uses the OPLSAA force field, which is optimized for ≈200 small organic molecules.

The are 18 replicas simulated from 270-690K *

The free energy landscape of this model is then compared with the authors’ previous work in an explicit model.

* Accurate results were not expected at high temperatures, but were helpful for transitioning energy barriers rapidly, allowing for more efficient sampling.
Methodology

REM was implemented by a molecular-modeling package, IMPACT.

- Run in parallel (on parallel processors) at sequence of temperatures
- Neighboring replicates would periodically exchange, acceptance determined by a Metropolis MC criterion
- Replicas can be generated by MC, HMC, or MD; HMC was used for this study

Hybrid Monte Carlo

- HMC uses MD to generate possible conformations.
- “bad MD, but good MC”
- HMC generally scales poorly with system size, but was comparable to MD for this study
Methodology

The algorithm used for REM is:

\[ T(x_i | x_j) = \begin{cases} 
1, & \text{for } \Delta \leq 0 \\
e^{-\Delta}, & \text{for } \Delta > 0 
\end{cases} \]

Where \( \Delta = (\beta_i - \beta_j) [V(x_j) - V(x_i)] \)

\( \beta_i \) and \( \beta_j \) are the two reciprocal temperatures

\( x_i \) is the configuration at \( \beta_i \) (also for \( x_j \) and \( \beta_j \))

\( V(x_n) \) is the potential energy function
Methodology of SGB

From Dr. Madura’s Lecture slides:

\[
G_{elec} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\varepsilon r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \frac{q_i^2}{a_i}
\]
**Methodology**

Solvation free energy of a protein is expressed as the sum of the “reaction field energy” and the “cavity energy”

\[
U_{SGB} = U_{RXN} + U_{CAV}
\]

Total reaction field energy is expressed as:

\[
U_{RXN} = U_{SE}(q_i, \mathbf{r}_i) + \sum_{i<j} U_{pr}(q_i, q_j, \mathbf{r}_i, \mathbf{r}_j)
\]

Where \( U_{SE} \) (single energy) is

\[
U_{SE} = -1/8\pi \left(1/\varepsilon_i - 1/\varepsilon_o\right) \int_{s} q_i^2 / |\mathbf{R} - \mathbf{r}_k|^4 (\mathbf{R} - \mathbf{r}_k) \cdot \mathbf{n}(\mathbf{r}) d^2\mathbf{R}
\]

Where \( \varepsilon_i\) is the dielectric constant for the interior of the solute. For proteins, typically \(\approx 1.0-4.0\). \(\varepsilon_o\) is the dielectric constant for outside water (78.5)

Pairwise screened Coulomb Energy

\[
U_{pr} = -\left(1/\varepsilon_i - 1/\varepsilon_o\right) q_i q_j / \sqrt{(\mathbf{r}_{ij}^2 + \alpha_{ij}^2 e^D)}
\]

\[
\alpha_{ij} = \sqrt{(\alpha_i \alpha_j)}, \text{ where } \alpha \text{ is the Born radius, and parameter } D = \mathbf{r}_{ij}^2 / (2 \alpha_{ij}^2)
\]
Experiment

- Residues 41-56 from the C terminus of Protein G, capped with Ace and Nme groups, totals 256 atoms.
- The SGB continuum solvation model is used with a dielectric constant of 2.0, but 1.0 and 4.0 were also tried.
- The MD simulations are carried out with IMPACT
- 18 replicas are simulated with temperatures ranging from 270 to 690 K
- Before the production run, a conjugate gradient minimization is performed
A 100-ps MD equilibration is followed with temperature ramping from 0K to the specified temp.

Each configuration is run for 2.0 ns for data collection, replica exchange attempted every 200 fs, with protein configuration saved every 80 fs, giving a total of 0.45 million configurations.

The temperatures were from 270, 282, 295, 310, …, 649 to 690K. Sufficient evidence for adequate sampling was demonstrated by the all the replicas visiting all the temperatures during a simulation, and at a given temperature, all the replicas visited many times during the same MD run.
The free energy landscape is determined by calculating the normalizing probability distribution function from a histogram analysis.

\[ P(x) = e^{-\beta W(x)} / Z \]

Where \( X \) is the specified choice set of reaction coordinates and \( Z \) (should look familiar from the statistical mechanics lecture) is the partition function.

The next slide shows the contour map, where \( A \) is the result of the explicit model, and \( B \) is the result of the implicit model. The \( R_g \) (core) is the radius of gyration of the side-chain atoms on the four hydrophobic residues.
Fig. 1.  Comparison of the free energy contour maps versus the number of β-sheet hydrogen bonds $N_{\text{H}}$ and the hydrophobic core radius gyration $R_g^{\text{core}}$ for explicit (a) and implicit (b) solvent simulations at 310 K. A hydrogen bond is counted if the distance between two heavy atoms (N and O in this case) is less than 3.5 Å, and the angle N–H...O is larger than 150.0°. The free energy is in units of RT, and contours are spaced at intervals of 0.5 RT.
Results

The models prove to be surprisingly different!

1. The native state is no longer the lowest free energy state in the continuum solvent model
2. The most heavily populated state has no meaningful hydrogen bonds
3. Higher radius of gyration for the hydrophobic core

However, the overall shape of the contour map is still an “L”
Results

Fig. 2. Comparison of the representative structures with the lowest free energy from the explicit (a) and implicit (b) solvent simulations. The hydrophobic residues (W43, Y45, F52, and V54) are represented by space-fill, charged residues (E42, D46, D47, K50, and E56) are represented by sticks with positively charged residues colored blue and negatively charged residues colored red, and the rest are represented by ribbons. The implicit solvent structure show very different features compared with the explicit solvent structure (see text for details).
Results

1. The hydrophobic residue F52 is expelled from the hydrophobic core in the implicit solvent model, but remains in the core in the explicit solvent model.

2. The explicit solvent model’s side chains extend into the solvent and are fully solvated. The implicit solvent model’s residues form salt bridges between opposite charges.
Results

The authors further calculated the $\beta$-hairpin population at various temperatures to the populations determined experimentally from measurements of fluorescence quantum yields.

Experimentally, it is found that at 282K, the $\beta$-hairpin population is $\approx 80\%$. The explicit model calculates this population at 72%, while the implicit model seriously underestimates this population at 39%.

This trend remains true throughout the range of dielectric constants tried.
Results

The authors applied a similar analytical technique, calculating the hydrogen-bond populations and comparing the NMR results.

Experimental NMR data show $\approx 42\%$ of the $\beta$-sheet hydrogen-bond population at 310K, compared to explicit model calculation of 45% and implicit model calculation of only 10%.

Again, the implicit model significantly underestimates the population. It should be pointed out that at higher temperatures, however, the explicit model will overestimate hydrogen-bonds.
Results

STRIDE\textsuperscript{1} was used to calculate the $\beta$-sheet and $\alpha$-helix content suggested by the implicit solvent model.

Results are consistent with experimental evidence; only 1-2\% of the conformations exhibit helical content at all temperatures, which is also consistent with the explicit solvent model.

Results

Further analysis was done on OPLSAA/SGB implicit solvent model, to see if it could still be used as a scoring function in protein structure prediction.

The next slide shows a histogram for the structures and their corresponding energy from the lowest energies, derived from the implicit solvent model.
Fig. 5. The OPLSAA/SGB energy histogram for structures in state H from implicit solvent simulation. The native structure is found to have the lowest OPLSAA/SGB energy in this case (marked in the figure). The fact that the native structure is found to have the lowest energy validates its use in protein structure prediction as a scoring function, since the OPLSAA/SGB energy scoring function still picks the native structure as the best structure.
Discussion

Contour Map
- Electrostatic and Hydrophobic Interactions not preserved
- Same “L” shape suggests internal consistency with hydrophobic collapse
- Most favored state had no H-bonds, we know this to be incorrect
- Protein structure expelled part of the hydrophobic core, in favor of an over-stabilized salt bridge
Discussion

Additional Calculations

- β-hairpin population at various temperatures was significantly closer to experimentally determined values with the explicit model.

- Hydrogen bonds were much closer with the experimentally determined values, but as temperature increased this was overestimated.
The authors concluded that a better implicit solvent model should be devised, because the overestimation of the salt bridge stability is the principle aspect of the erroneous results.

Also, the OPLSAA/SGB can still be used a scoring function, as it calculated the lowest free energy at the native state.
How does this relate to my research?

- MMC
  - Boltzmann probability
  - Simulation
- Free Energy
- Protein Folding
Acknowledgement

I would like to thank:

- Anes Aref and Jerome Nilmeier
- Dr. Hagai Meirovitch
- NSF-NIH
- Duquesne University, University of Pittsburgh, Carnegie Mellon University, and the Pittsburgh Supercomputing Center
- Everyone in the BBSI