Computer simulation study of folding thermodynamics and kinetics of proteins in osmolytes and denaturants

Apichart Linhananta, Gianluca Amadei, and Timothy Miao
Department of Physics, Lakehead University, Thunder Bay, Ontario, Canada

E-mail: apichart.linhananta@lakeheadu.ca

Abstract. In living cells, the presence of macromolecular crowders, such as osmolytes or denaturants, strongly affects the stability of folded proteins. The overall effects depend on the size, concentration, and chemical properties of the crowding agents. This work uses an all-atom Gō model of the Trp-cage in spherical solvents to probe the physical origin of protein stabilization/destabilization by osmolytes/denaturants. The solvent quality is controlled by the solvent-protein contact $\varepsilon_{ps}$ that can represent repulsive osmolytes ($\varepsilon_{ps} > 0$) or attractive denaturants ($\varepsilon_{ps} < 0$). The model is used to show that protein stabilization by osmolytes proceeds by an excluded volume, entropy-driven mechanism. Protein destabilization by denaturant is shown to be driven by changes in enthalpy. It is found that small osmolytes are the most effective stabilizer of proteins. Folding simulations of the Trp-cage in osmolytes observe a two-fold increase in folding rates, for small osmolytes. This is due to an osmolyte-induced shift to more compact unfolded protein conformations.

1. Introduction
Osmolytes are intracellular organic molecules that stabilize proteins against unfolding under environmental stresses such as high temperatures, desiccation, or the presence of chemical denaturants such as urea.[1] The stabilizing property of osmolytes has been correlated with the exclusion of osmolytes from protein domains, resulting in the accumulation of water near protein domains.[2, 3] The exclusion of osmolytes from proteins arise from repulsive interactions between osmolytes and the backbone of proteins.[4, 5] A possible stabilization mechanism suggests an osmolyte-induced loss of protein conformational entropy, with the greater entropic loss by the unfolded (U) state, leading to an overall shift in equilibrium towards the native (N) state. The entropy loss mechanism is consistent with experimental works that observed increased compactness in unfolded states of cutinase[6], protein S6,[7] and Rnase S[8] due to osmolytes.

In contrast to protecting osmolytes, denaturants (such as urea) are nonprotecting osmolytes that denature proteins from folded to unfolded state. The attractive protein-denaturant interactions lead to the accumulation of denaturants near proteins.[9] This lowers the free energy of both the N and U states, but due to its larger solvent exposed surface area the free energy of the U state is lowered by a greater amount.[5] Consequently, the addition of denaturants to protein solutions results in an enthalpy-driven mechanism that shifts the equilibrium to the unfolded state.

There has been relatively few theoretical or computer simulation investigations of proteins in aqueous solution of osmolytes and denaturants. Daggett et al.[10, 11] performed all-atom molecular
dynamics (MD) simulations of proteins in aqueous solution of osmolytes and urea to show that the stabilization property of the osmolyte trimethylamine N-oxide (TMAO) arises from enhance water structures due to their interactions with TMAO. All-atom MD simulations have been performed to show that urea destabilizes proteins by direct and water-assisted urea-protein interactions.[9, 12, 13] There have been simulations of polymer and hydrocarbons,[14, 15] and of RNA hairpin,[16] in solutions of osmolytes or urea. However, the high computational cost of all-atom MD simulations of proteins in explicit solvents has thus far prevented a complete thermodynamics study required to fully understand the mechanisms of protein stabilization/destabilization by osmolyte/denaturant. The exception is recent landmark study by Canchi et al.[17] MD simulation of Trp-cage in solution of urea, which found an enthalpy-driven protein denaturation mechanism - a result that is supported by experiments.[18]

A universal description of protecting and denaturing (urea) osmolytes has been proposed by Bolen.[5] Based on a solvent quality paradigm, it classifies solvents as good or poor. In poor solvents (solvophobic) protein intramolecular interactions dominate, which favors compact folded native states that minimize solvent exposed protein surface area. In good solvents (solvophilic), protein-solvent interactions dominate, which favors unfolded states that maximize protein-solvent contacts. Water is a poor solvent since the effective water-protein interactions lead to the hydrophobic effects, one of the major forces that fold proteins. Aqueous osmolytes and aqueous denaturant solution are poor and good solvent, respectively. The solvent quality paradigm led to several molecular free energy transfer models, which calculate free energy change of proteins transferred from pure water to aqueous osmolyte/urea solutions.[19, 20] A recent study combines protein conformations from simulation data of G6 model simulations with transfer free energy models to infer the thermodynamics property of proteins in solutions of osmolytes and urea.[21] The study predicted that solution of osmolyte and urea raises and lowers, respectively, the folding temperatures of the G6 model of proteins. However, free energy transfer models do not take into account the size of solvents, and may lead to an incomplete description of the change in protein conformational entropy due to solvents. Excluded volume due to solvents can reduce the amount of accessible protein conformations, changing the unfolded protein state to a more compact native-like unfolded state, as observed in experiments.[6-8] In addition, the binding of solvent to protein depends on the size of water, osmolytes and urea, and solvent size must be included in theoretical or MD models to fully assess the enthalpy change due to solvents.[22]

In this work we construct an all-atom G6 model protein immersed in solvent molecules to investigate how osmolyte/urea stabilizes/destabilizes proteins. The previously studied Trp-cage Go model is selected for its small size (i.e. low computational cost) and cooperative behavior.[23] The Trp-cage model is immersed in spherical solvents, where of the solvent-protein interaction is controlled by the solvent-protein contact energy $\epsilon_{ps} \cdot \epsilon_{ps}$ is the control parameter of the solvent quality paradigm discussed earlier. Positive/negative value of $\epsilon_{ps}$ corresponds to bad/good solvents, which mimics the effects of aqueous osmolyte/denaturant solutions. In the next section, the detail of the discontinuous molecular dynamics (DMD) model and the methods used to calculate the free energies and entropy are discussed. The model is used to demonstrate thermal stabilization/destabilization by osmolyte/denaturant. It will be shown that the protein stabilization by osmolytes ($\epsilon_{ps} > 0$) is driven by entropy change, and that protein destabilization (denaturation) by denaturant is driven by enthalpy change. Solvent size effects and the effects of osmolyte/denaturant on folding kinetics are also presented.

2. Methods

2.1. Models of Protein in Solvents

Discontinuous Molecular Dynamics (DMD) is an efficient method that has been used to study protein folding, protein aggregation,[24-26] and ab initio protein structure prediction[27]. This work is based on a previous DMD all-atom model of the Trp-cage protein.[23] A brief description of DMD is presented here. The initial heavy-atom positions are obtained from the NMR structure (structure 1 of PDB ID 1L2Y) and the missing polar hydrogen molecules are constructed as before. The resulting structure is...
comprised of 189 heavy atoms and polar hydrogen atoms, which in this “discontinuous” model are represented as hard spheres. Two bonded atoms \(i\) and \(j\), as well any 1,3 angle-constrained pair and 1,4 aromatic pair, are constrained by an infinite square-well potential

\[
u_{ij}^{\text{bond}} = \begin{cases} 
\infty, & r < 0.9\sigma_{ij} \\
0, & 0.9\sigma_{ij} < r < 1.1\sigma_{ij} \\
\infty, & r > 1.1\sigma_{ij},
\end{cases}
\]

where \(\sigma_{ij}\) is the separation distance of the bonded \(i,j\) pair in the NMR structure. The model also includes discontinuous improper dihedral potential, to maintain chirality about tetrahedral heavy atoms (such as \(\alpha\) carbon without explicit hydrogen), and certain planar atoms. The potential has the form

\[
u_{ij}^{\text{improp}} = \begin{cases} 
\infty, & \omega < \omega_0 - 20' \\
0, & \omega_0 - 20' < \omega < \omega_0 + 20' \\
\infty, & \omega > \omega_0 + 20',
\end{cases}
\]

where \(\omega\) represents the dihedral angles of the constrained atoms, which are restricted to values near \(\omega_0 = 35.26439\)° for tetrahedral heavy atoms, and \(\omega_0 = 0°\) for planar atoms. Two non-bonded atoms \(i,j\) pair interacts by a hard-core and square-well potential

\[
u_{ij}^{\text{non-bond}} = \begin{cases} 
\infty, & r < 0.8\sigma_{ij}^{vdW} \\
B_{ij} \epsilon_{Go}, & 0.8\sigma_{ij}^{vdW} < r < 1.2\sigma_{ij}^{vdW} \\
0, & r > 1.2\sigma_{ij}^{vdW},
\end{cases}
\]

where \(\sigma_{ij}^{vdW}\) are the van der Waals (vdW) diameters of the CHARMM potential set 19 and \(B_{ij}\) and \(\epsilon_{Go}\) are the \(\epsilon_{Go}\) interaction strength parameters. Using these parameters we performed a short discontinuous molecular dynamics (DMD) to remove bad contact from the initial experimental structure to produce the ground-state structure of the model Trp-cage, as shown in Figure 1A. The \(\epsilon_{Go}\) model potential is implemented by setting the non-bonded square-well depth to \(-\epsilon_{Go}\) (\(B_{ij} = -1\)) for \(ij\) pairs with van der Waals overlap in the ground-state structure, otherwise it is set to 0 (\(B_{ij} = 0\)).

The \(\epsilon_{Go}\) model in solvents system is implemented by placing the Trp-cage protein in a 40Å x 40Å x 40Å box, with randomly generated (without overlaps) spherical solvent molecules. Standard periodic boundary conditions are implemented. Solvent molecules interact with each other by a hard-core and square-well potential \(U_{ij}^{\text{solvent-solvent}}\)

\[
u_{ij}^{\text{solvent-solvent}} = \begin{cases} 
\infty, & r < 0.8\sigma_{ij}^{SS} \\
\frac{\epsilon_{SS}}{\epsilon_{Go}} \epsilon_{Go}, & 0.8\sigma_{ij}^{SS} < r < 1.2\sigma_{ij}^{SS} \\
0, & r > 1.2\sigma_{ij}^{SS},
\end{cases}
\]

where \(\sigma_{ij}^{SS}\) is the vdW diameter of the solvent. \(\epsilon_{SS}^*\) is the solvent-solvent square-well depth in units of the \(\epsilon_{Go}\) contact energy \(\epsilon_{Go}\). Solvent-protein molecules interact with each other by a hard-core and square-well potential \(U_{ij}^{\text{solvent-protein}}\)
\[
\frac{\epsilon_{ij}^{\text{protein-solvent}}}{\epsilon_{\text{Go}}} = \begin{cases} 
\infty, & r < 0.8 \sigma_{ij}^{\text{PS}} \\
\epsilon_{\text{PS}}^*, & 0.8 \sigma_{ij}^{\text{PS}} < r < 1.2 \sigma_{ij}^{\text{PS}} \\
0, & r > 1.2 \sigma_{ij}^{\text{PS}},
\end{cases}
\]

where \( \sigma_{ij}^{\text{PS}} = \left( \sigma_{ij}^{\text{vdW}} + \sigma_j^S \right) / 2 \) is the average vDW diameter of the protein-solvent \((i,j)\) pair, where \( \sigma_{ij}^{\text{vdW}} \) is the CHARMM potential set 19 vDW diameter of the \(i\)th atom of the protein and \( \sigma_j^S \) is the vDW diameter of the \( j \)th solvent molecule. \( \epsilon_{\text{PS}}^* \) is the protein-solvent square-well depth in unit of \( \epsilon_{\text{Go}} \).

**Figure 1.** A) Ground-state native state structure of the all-atom Trp-cage protein (note the short helix and the hydrophobic core.); B) Unfolded Trp-cage in denaturants; C) Folded Trp-cage in osmolytes.

\( \epsilon_{\text{PS}}^* \) is the solvent quality scale and is varied from a minimum value of \( \epsilon_{\text{PS}}^* = -0.8 \), to represent a strongly denaturing denaturant solution (Figure 1B), up to a maximum value of \( \epsilon_{\text{PS}}^* = +1.0 \), to represent a strongly protein-stabilizing osmolyte solution (Figure 1C). The solvent-solvent square-well depth is usually set \( \epsilon_{\text{SS}}^* = -1.0 \) to reflect the solvent particles overall preference to interact with itself rather than with the protein \( \left( \epsilon_{\text{SS}}^* < \epsilon_{\text{PS}}^* \right) \). In some simulations it is set to zero.

As in previous DMD studies the energy scale is set by the Go contact energy, with native Go contact energy \( -\epsilon_{\text{Go}} = -1 \).[23, 28-31] Free energies (such as \( E' = E / \epsilon_{\text{Go}} \)) and temperature \( (T' = k_B T / \epsilon_{\text{Go}}) \) are scaled in units of \( \epsilon_{\text{Go}} \). A reduced time unit \( t' = \sqrt{\epsilon_{\text{Go}} / M \sigma^2_L} \), where \( M \) is the average atomic mass of the protein and \( \sigma_L = 1 \, \text{Å} \), is used.

### 2.2. Implementation of Discontinuous Molecular Dynamics

Standard continuous molecular dynamics (MD) are usually performed by numerical integration of Newton’s equation in femtosecond time step. In most algorithms, the potential energy and force acting on the system of particles are calculated at every time step. DMD avoid these time consuming steps, since the discontinuous particles interact only at the instant of collision. The algorithm involves calculating all collision times of the model protein.[32, 33] The collision time, \( t_{\text{collision}} \), of bonded (Equation 2.1) and
non-bonded (Equation 2.3 and 2.4) pairs are found by solving a quadratic equation in time, with solution,

$$t_{\text{collision}} = \frac{2\vec{v}_{ij} \cdot \vec{r}_{ij} \pm \sqrt{4\left(\vec{v}_{ij} \cdot \vec{r}_{ij}\right)^2 - 4v_{ij}^2r_{ij}^2}}{2v_{ij}^2}$$ (2.6)

where $\vec{v}_{ij} = \vec{v}_i - \vec{v}_j$ is the relative velocity, and $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$ is the current relative position of the $ij$ pair. The calculation of the time of a dihedral collision (Equation 2.2) involves solving a cubic equation by a Newton-Raphson scheme, to find the smallest positive root, as documented in a previous publication.[34] The collision times are compare to find the next collision time, $t_{\text{next}}$ and to identify the colliding pair, or the colliding quartet in the case where the next collision is a dihedral collision. All atoms are propagated for the duration of $t_{\text{next}}$, at which point the colliding pair or quartet collide, and the new velocities are determined by conservation of momentum. The velocities of non-colliding particles are unchanged. New collision times for all pairs or quartets that include the colliding particles are recalculated as above. The collision times of pairs or dihedral quartets that do not include the colliding particles do not require recalculations, but are adjusted by $t_{\text{new collision}} = t_{\text{old collision}} - t_{\text{next}}$. A neighbour link list is used to further reduce the computational cost[33].

The DMD simulation is conducted in the canonical ensemble with the temperature fixed by the Andersen ghost particle method, where the atoms of the proteins and solvent particles experience a random collision with heat-bath ghost particles. The rate of collision follows a distribution $\exp(-\nu t)$ with the collision frequency given by standard kinetic theory $\nu = n\sigma^2\sqrt{4\pi k_BT/m}$, where $\sigma$ and $m$ are the average size and mass, respectively, of the system, and $n$ is the ghost particle density. For the current protein-solvent model a ghost density of 0.1 per cubic angstrom gives satisfactory temperature equilibration, without altering folding behaviour. The ghost collision time of all particles are calculated by the algorithm $t_{\text{ghost}} = -(1/\nu)\ln R$, where $R$ is a random number. The ghost collision times are compared with other collision times to determine the next collision time. When the next collision is a ghost collision the velocity of the colliding atom is recalculated according to the Maxwell-Boltzman distribution

$$\exp\left(-\frac{1}{2}\frac{mv^2}{k_BT}\right) \approx \frac{1}{\sqrt{2\pi}} \left(\frac{1}{k_BT/\epsilon}\right)^2 \frac{1}{\frac{mv^2}{\epsilon}} \frac{1}{\sqrt{2\pi}} \left(\frac{1}{k_BT/\epsilon}\right)^2,$$

using the algorithm presented in reference[32].

The efficiency of the DMD model enables the folding of an all-atom Trp-cage model in solution in about 12 hr to 72 CPU hours by a serial job submission on a 2.2 GHz Opteron processor. In all, roughly 10000 Trp-cage all-atom models were folded to the native state under a wide variety of thermodynamic conditions and solutions. Taking into account the $4\mu s$ folding time of the Trp-cage[35], this corresponds to about 40 ms in real time. In contrast, the recent landmark study by Garcia et al.[17] on an all-atom model of the Trp-cage in solution of urea, involves a total sampling of about $35\mu s$. They used the TIP3P model for water, and the Kirkwood-Buff model for urea. The model presented here takes an alternative approach by using a coarse-grain solvent quality model that reflects the solvent quality paradigm of Bolen.[5] In the supplementary section of a previous work,[36] it is shown how the effect of water on proteins in an osmolyte/urea solution can be included implicitly into an effective interaction between a united spherical solvent particle and an atom of the protein, controlled by the effective solvent-protein contact energy $\epsilon_{PS}^\ast$. Setting $\epsilon_{PS}^\ast > 0$ captures some of the effective repulsion between a protein and osmolytes due to the interaction between osmolytes and the water molecules that hydrates the surface of the protein. Setting $\epsilon_{PS}^\ast < 0$ captures the favourable protein-denaturant interactions that some researchers believe drives protein denaturation. Though such united-solvent models cannot examine specific aspects of protein denaturation/stabilization by urea/osmolyte, its computational efficiency and simple structure facilitate the study of universal properties, that, as discussed earlier, are still computationally unfeasible.
by more sophisticated models. Of particular interest are the effects of solvent size, concentration, and interaction strength, on protein stability and cooperativity, and protein folding kinetic.

2.3. Free Energy and Entropy Calculations

The free energy calculations are based on standard multiple weight histogram method (WHAM), [37-39] which calculates thermodynamics quantities by approximating the density of states \( g(\epsilon) \) (i.e. the number of states with energy \( \epsilon \)) from simulation data. This defines the partition function \( Z = \sum g(\epsilon) \exp(-\beta \epsilon) \), the free energy \( F = -k_B T \ln Z \), the internal energy \( U = \langle E \rangle = \sum \epsilon g(\epsilon) \exp(-\beta \epsilon)/Z \), and the entropy \( S = (U - F)/T \). This is generalized by defining the free energy functional as a function of an order parameter \( x \):

\[
f(x,\Delta x) = -k_B T \ln \sum \epsilon g(\epsilon,x,\Delta x) \exp(-\beta \epsilon),
\]

where \( \epsilon \) indicates a summation over all energy states within the order parameter range from \( x \) to \( x + \Delta x \), and \( g(\epsilon,x,\Delta x) \) is the density of state of energy \( \epsilon \) in the order parameter range \( x \) to \( x + \Delta x \). The probability that the protein systems occupy states with order parameter range from \( x \) to \( x + \Delta x \), \( P(x,\Delta x) \) can be determined directly from the protein configurations recorded during simulations, and it obeys the relation

\[
P(x,\Delta x) = \frac{\sum \epsilon g(\epsilon,x,\Delta x) \exp(-\beta \epsilon)}{Z},
\]

where \( Z \) is the partition function. Some manipulations lead to the relation for the free energy density

\[
f(x,\Delta x) = -k_B T \ln P(x,\Delta x) - k_B T \ln Z = -k_B T \ln P(x,\Delta x) + F,
\]

where \( F \) is the free energy calculated by standard WHAM. We also defined the internal energy functional \( u(x,\Delta x) \), the average energy for order parameter range \( x \) to \( x + \Delta x \), which can be obtained directly from the simulation data. This leads to the entropy functional:

\[
s(x,\Delta x) = \frac{u(x,\Delta x) - f(x,\Delta x)}{T}.
\]

In this work the non-local native fraction \( Q \), defined previously in reference [23], is employed as the order parameter, and varies from 0 (completely unfolded) to 1 (completely folded). For the free energy functional calculation, a bin size \( \Delta Q = 0.02 \) is used. As mentioned earlier, the computed functional of energy \( u^*(Q, \Delta Q) \), free energy \( f^*(Q, \Delta Q) \), and entropy \( s^*(Q, \Delta Q) \) are in units \( \epsilon_{go} \).

3. Results and Discussions

To study the thermodynamic, we performed equilibrium simulations of Trp-cage in 1000 spherical solvents, of vdW diameter \( \sigma^* = 3.0 \) Å (roughly the size of a water molecule) in a 40Å X 40Å X 40Å box. The solvent-solvent interaction is set to \( \epsilon_{ss}^* = -1.0 \), and the protein-solvent contact energy varies from \( \epsilon_{ps}^* = -0.8 \) (strong denaturants) to \( \epsilon_{ps}^* = 1.0 \) (repulsive osmolytes). All simulations start from the ground-state native structure of Figure 1a, and is equilibrated for a scaled time of \( t^* = 20000 \), follow by \( t^* = 60000 \) of simulations. A serial run takes up to two days on a SHARCNET cluster. To ensure good statistic for WHAM, five independent runs at a fixed temperature are performed, for \( T^* = 3.0 \) to 6.0, in steps of \( \Delta T^* = 0.2 \). Figure 2a plots the heat capacity vs. temperature of the simulations. As expected the folding temperature, \( T_f^* \), defined as the peak of the heat capacity, increases with \( \epsilon_{ps}^* \). Figure 2b is the Temperature vs. solvent quality, \( \epsilon_{ps} \), phase diagram constructed by assuming that \( T_f^* \) defines the phase
boundary. As $\varepsilon_{ps}$ increases, and the solvent quality transforms from denaturant-like ($\varepsilon_{ps} < 0$) to osmolyte-like ($\varepsilon_{ps} > 0$), the protein remains in the native (N) state up to higher temperatures. Hence, the model reproduces thermal stabilization by osmolytes. In the following subsections, the physical origins of protein stabilization by osmolytes are discussed.

![Figure 2](image_url)

**Figure 2.** a) Heat Capacity vs. Temperature for Trp-cage and 1000 spherical solvents in a 40Å X 40Å X 40Å box with $\varepsilon_{ps} = -0.6$ to 0.8; b) Temperature vs. solvent quality ($\varepsilon_{ps}$) phase diagram.

### 3.1. Stabilization Mechanism

![Figure 3](image_url)

**Figure 3.** a) Free energy functional ($f^*$) vs. native fraction (Q) at temperature $T^* = 4.8$, with solvent quality $\varepsilon_{ps} = 0$ (solid), 0.4 (dashed), -0.2 (dotted); b) Internal energy functions ($u^*$) vs. Q; c) Entropy $S$ vs. Q; d) Entropy change $\Delta S$. 
functional ($s^*$) vs. Q; d) Relative entropy $\Delta S$ vs. Q, $\Delta S = s^*(\varepsilon_{ps}) - s^*(0)$, where $s^*(0)$ is the entropy of the model with $\varepsilon_{ps} = 0$, and $s^*(\varepsilon_{ps})$ is for $\varepsilon_{ps} = 0.4$ (dashed) and $\varepsilon_{ps} = -0.2$ (dotted).

To determine the stabilization mechanism, analysis that compare solvent models near their folding temperatures are done. Here comparison of three models at $T^* = 4.8$ are presented: $\varepsilon_{ps} = 0$ (neutral, $T^*_i = 4.5$); $\varepsilon_{ps} = -0.2$ (denaturant-like, $T^*_i = 4.3$); $\varepsilon_{ps} = 0.4$ (osmolyte-like, $T^*_i = 5.0$). Figure 3a plots the free energy functional obtained by combining WHAM with Eqs.(2.8) and (2.9). The neutral solvent and denaturant-like models have free-energy minima at low Q (<0.2) indicating an unfolded (U) state, while the osmolyte-like model has a minimum at high Q (>0.5) consistent with a stable folded native (N) state. The free energy is divided into the internal energy and entropic (Eq.(2.10)) components. Figure 3b shows that the difference between the internal energy of the three models do not vary significantly with Q, and hence the stabilization mechanism is not driven by enthalpy (internal energy). Figure 3c,d shows that the entropy differences between the models are substantial only for unfolded states (low Q), and that, compare to the neutral solvent ($\varepsilon_{ps} = 0$), the Trp-cage in osmolyte-like solvent ($\varepsilon_{ps} = 0.4$) shows a substantial loss of entropy for U state (Q < 0.3, $\Delta S < 0$), but not for the N state (Q > 0.5, $\Delta S = 0$). This agrees with our hypothesis that the repulsive osmolytes ($\varepsilon_{ps} > 0$) reduce the entropy of unfolded states more than folded states, resulting in a free-energy bias that produces a more stable native state. This entropy-driven stabilization mechanism is consistent with several experiments.[6-8] For the weak denaturant-like system ($\varepsilon_{ps} = -0.2$), Figure 3c,d shows significant entropy gain ($\Delta S > 0$) only for the unfolded state, also consistent with an entropy-driven denaturing mechanism. Other analysis (not shown) also found entropy-driven mechanism for osmolyte-like or weak denaturant solvents ($\varepsilon_{ps} \leq -0.2$).

Figure 4. $\Delta H$ vs. Q. $\Delta H = u^*(\varepsilon_2) - u^*(\varepsilon_1)$, where $u^*(\varepsilon_i)$ is the internal energy functional of the solvent model $\varepsilon_i$. The osmolyte models (solid line) compares the internal energy of two osmolyte models $\varepsilon_1 = 0$ and $\varepsilon_2 = 0.4$. The denaturant models (dotted) compares the internal energy of two strongly denaturing solvent models $\varepsilon_1 = -0.4$ and $\varepsilon_2 = -0.6$. As defined above, $\Delta H$ is the internal energy difference. The use of $\Delta H$ facilitates comparisons with other works.

For strong denaturant ($\varepsilon_{ps} < -0.4$) the attractive solvent-protein interactions changes the stabilization/destabilization mechanism to enthalpy driven. In Figure 4, the osmolyte models’ curve, $\varepsilon_1 = 0$ and $\varepsilon_2 = 0.4$ (see Figure 3c), shows no variation of $\Delta H = u^*(\varepsilon_2) - u^*(\varepsilon_1)$ with Q. In contrast, the denaturant models’ curve, $\varepsilon_1 = -0.4$ and $\varepsilon_2 = -0.6$, shows $\Delta H$ decreasing with Q, which indicates an energetic preference for unfolded protein conformations in the presence of strong denaturants. The entropy difference $\Delta S$ (not shown) between the strong denaturant models does not show the clear variation with Q found in the osmolyte models. This finding is consistent with the work of the Garcia group that performed all-atom AMBER simulation of Trp-cage in urea [17], and also observed an
enthalpy-driven protein denaturation mechanism. In that work, the thermodynamics are obtained by fitting simulation data to a phenomenological Hawley-type free energy surface. In contrast, in the work presented here, the free energy, enthalpy (in our model $\Delta U = \Delta H$), are directly calculated by the method outlined in subsection 2.3.

3.2. Size effects and cooperativity

![Figure 5](image_url)

**Figure 5.** a) Heat capacity ($C^*$) vs. Temperature ($T^*$) for Trp-cage, with no solvents (vacuum), and in hard-sphere solvent of radius 1.5, 2.0, 2.5 Å, in a periodic box of 40Å X 40Å X 40Å; b) Probability distribution vs. Energy ($E^*$), where the Trp-cage in vacuum is the thick solid line, and the peaks corresponding to the N and U states of the Trp-cage in vacuum are indicated. The distribution for solvent radius 1.5Å is also indicated. Both graphs were obtained by WHAM.

In view of the depletion-type (entropy driven) stabilization mechanism of osmolyte-like solvent, $\varepsilon_{ps} > 0$, it is expected that smaller solvents are better protein stabilizer. To examine this, simulations of Trp-cage in solvents are performed at solvent volume fraction of 30% for solvent vDW radii $r_{vdW} = 1.5, 2.0, 2.5, \text{ and } 3.0$ Å, which, for a 40Å X 40Å X 40Å box, correspond to 1358, 573, 293, and 170 solvent molecules, respectively. The solvent-solvent and solvent-protein contact energy are set to zero ($\varepsilon_{cd} = 0, \varepsilon_{ps} = 0$). Hence the solvents are pure hard spheres. **Figure 5a** shows that the folding temperature increases from $T_f^* = 4.0$, for the Trp-cage in vacuum, to $T_f^* = 4.4, 4.7$ and 5.0 for solvent radius $r_{vdW} = 2.5, 2.0, \text{ and } 1.5$ Å, respectively. This shows that smaller osmolytes are more effective at stabilizing the native state of proteins. Simulations of Trp-cage in denaturants, $\varepsilon_{ps} < 0$, also found that smaller denaturants are more effective at destabilizing proteins (data not shown).

It is well established that proteins are cooperative, in that they exist only in two states: native or unfolded.[40] In a previous study[23] the Trp-cage $G0$ in vacuum was shown to be cooperative. This is evident in the **Figure 5b**, where at its folding temperature $T_f^* = 4.0$, the energy probability distribution is bimodal, with the low and high energy peaks corresponding to the N and U state, respectively. In contrast, the Trp-cage in solvent systems show a progressive loss of two-state behavior, with a near complete loss of cooperativity for solvent size $r_{vdW} = 1.5$ Å, at its folding temperature $T_f^* = 5.0$, where the molten globular state (a state intermediate to the N and U state) is stable. The cooperative behavior of proteins in osmolytes and denaturants is a complex topic, and has been partially dealt with in a previous publication.[36]
3.3. Folding Kinetics

Folding simulations is performed for Trp-cage in pure HS solvent systems, by performing 100 unfolding simulations at $T^* = 6.0$ for a period of $t^* = 10000$ to produce 100 distinct unfolded conformations. The conformations are the initial structures of 100 folding simulations at $T^* = 3.0$, each lasting $t^* = 100000$. The trajectories are analyzed for the first passage time, defined as the time it takes for the protein to fold to its native state. A protein is defined to be folded if it has a main chain root-mean-square deviation (rmsd) of less than 1.5 Å and an all-heavy atom rmsd of less than 2.0 Å. For the Trp-cage in vacuum the 100 folding simulation gives an average folding time, defined as the average first passage time, of $\tau_{\text{fold}} = 17350$. Figure 6 plots the $Q_{\text{core}}$ vs. $Q_{\text{helix}}$ contour plot, which shows that the folding follows two basic paths: path 1, the fast diffusion-collision path; path 2, is the hydrophobic collapse path. In a previous publication[23], it was shown that path 2 is much slower because the core contacts hinders helical formation. That over 90% of trajectories fold by path 1 is the reason the Trp-cage is one of the fastest-folding proteins.[35] Folding simulations were performed on Trp-cage in a 40Å X 40Å X 40Å box with 20% solvent volume fraction of solvent radius $r_{\text{sdw}} = 1.5, 2.0, 3.0, 4.0$ and $6.0$ Å, which correspond to 906, 382, 113, 48 and 14 spherical solvent molecules, respectively. For larger solvents $r_{\text{sdw}} \geq 3.0$ Å, the folding times are $\tau_{\text{fold}} = 16000$, similar to that of the Trp-cage in vacuum. For smaller solvents $r_{\text{sdw}} = 1.5$ and 2.0 Å, the folding times are 8890 and 13150, respectively. Similarly for system of 30% solvent fraction of size $r_{\text{sdw}} = 2.0$ and 2.5 Å, which correspond to 573 and 293 solvents, the folding times are 7760 and 9920. It is clear that smaller repulsive osmolytes, as well as higher osmolyte concentration, leads to higher protein folding rates (decrease of $\tau_{\text{fold}}$) compare to the vacuum system. Analysis found (not shown here) that the Trp-cage in osmolytes still fold mainly by path 1 (about 90%), just like for Trp-cage in vacuum. The increase in folding rates is due mainly to osmolyte-induced entropy reduction of unfolded state, to a more compact native-like unfolded state, represented by $U^*$ in Figure 6. This conclusion is supported by experiments.[41]
4. Conclusions
This work presents a DMD model of Trp-cage in spherical solvents, constructed by the spherical solvent paradigm[5], where the solvent quality is controlled by the protein-solvent contact energy $\varepsilon_{PS}$ to produce good (denaturant, $\varepsilon_{PS} < 0$) and bad solvents (osmolyte, $\varepsilon_{PS} > 0$). Equilibrium simulations observe thermal stabilization, where proteins remain in the native state up to a higher temperature, as $\varepsilon_{PS}$ increases, and the solvents become more osmolyte-like. Using a modified WHAM method the protein stabilization mechanism is determined to be entropy driven, where the repulsive osmolytes reduce the entropy of unfolded conformations more than that of folded conformations. This mechanism is supported by experiments,[6-8] For strong denaturants ($\varepsilon_{PS} < -0.4$), it is demonstrated that the destabilization of the native state proceeds by enthalpy-driven mechanism. The result is largely consistent with the much more sophisticated model of the Garcia group[17], which suggests that generic aspects are important in protein stabilization/destabilization by osmolyte/denaturant. It is found that smaller osmolytes stabilize proteins more effectively, which is expected, in view of the depletion (entropy) driven mechanism. The model produces a loss in cooperative behavior due to osmolytes, which is surprising since osmolytes stabilize the native state proteins. The cooperative behavior of proteins in solvent is a complex topic, which has been partially dealt with in a different publication,[36] and is a topic of future study. DMD folding simulations found a substantial increase in folding rates, in the presence of small osmolytes, at osmolyte volume fraction of 20% and 30%. In the experiment of Silow et al.[41], folding rates do increase at low or moderate osmolyte concentration, but decreases at very high concentration. Reproducing this result with DMD simulation at high osmolyte concentration is a challenging future project.

Acknowledgement
AL acknowledges financial support of the National Science and Engineering Research Council of Canada and the Shared Hierarchical Academic Research Computing Network (SHARCNET), as well as the computing resources provided by SHARCNET and Compute/Calcul Canada.

References
[1] P. H. Yancey et al., Science 217, 1214 (1982).
[2] G. F. Xie, and S. N. Timasheff, Biophys. Chem. 64, 25 (1997).
[3] G. F. Xie, and S. N. Timasheff, Protein Sci. 6, 211 (1997).
[4] D. W. Bolen, and L. V. Baskakov, J. Mol. Biol. 310, 955 (2001).
[5] D. W. Bolen, and G. D. Rose, Annu. Rev. Biochem. 77, 339 (2008).
[6] R. P. Baptista et al., Biopolymers 89, 538 (2008).
[7] L. Chen et al., J. Mol. Biol. 351, 402 (2005).
[8] G. S. Ratnaparkhi, and R. Varadarajan, J. Biol. Chem. 276, 28789 (2001).
[9] B. J. Bennion, and V. Daggett, Proc. Natl. Acad. Sci. USA 100, 5142 (2003).
[10] Q. Zou et al., J. Am. Chem. Soc. 124, 1192 (2009).
[11] B. J. Bennion, and V. Daggett, Proc. Natl. Acad. Sci. USA 101, 6433 (2004).
[12] D. Klimov, J. E. Straub, and D. Thirulamai, Proc. Natl. Acad. Sci. USA 101, 14760 (2004).
[13] L. Hua et al., Proc. Natl. Acad. Sci. USA 105, 16928 (2008).
[14] M. V. Athawale, J. S. Dordick, and S. Garde, Biophys. J. 89, 858 (2005).
[15] R. Zangi, R. Zhou, and B. J. Berne, J. Am. Chem. Soc. 131, 1535 (2009).
[16] D. L. Pincus, C. Hyeon, and D. Thirulamai, J. Am. Chem. Soc. 133, 7364 (2009).
[17] D. R. Canchi, D. Paschek, and A. E. Garcia, J. Am. Chem. Soc. 132, 2338 (2010).
[18] L. N. R. Wafer, W. W. Streicher, and G. Makhatadze, Proteins-Structure Function and Bioinformatics 78, 1376 (2010).
[19] T. O. Street, D. W. Bolen, and G. D. Rose, Proc. Natl. Acad. Sci. USA 103, 13997 (2006).
[20] J. Rosgen, B. M. Pettitt, and D. W. Bolen, Biophys. J. 89, 2988 (2005).
[21] E. O'Brien et al., Proc. Natl. Acad. Sci. USA 105, 13403 (2008).
[22] R. Mountain, and D. Thirulamai, J. Phys. Chem. B 108, 6826 (2004).
[23] A. Linhananta, J. Boer, and I. MacKay, Journal of Chemical Physics 122 (2005).
[24] H. Jang, C. K. Hall, and Y. Q. Zhou, Biophys. J. 86, 31 (2004).
[25] F. Ding, and J. J. LaRocque, J. Biol. Chem. 280, 40235 (2005).
[26] F. Ding et al., Structure 14, 5 (2006).
[27] F. Ding et al., Structure 16, 1010 (2008).
[28] Y. Q. Zhou, and A. Linhananta, Proteins-Structure Function and Genetics 47, 154 (2002).
[29] Y. Q. Zhou, and A. Linhananta, Journal of Physical Chemistry B 106, 1481 (2002).
[30] A. Linhananta, and Y. Q. Zhou, Journal of Chemical Physics 117, 8983 (2002).
[31] A. Linhananta, H. Y. Zhou, and Y. Q. Zhou, Protein Science 11, 1695 (2002).
[32] M. P. Allen, and D. J. Tildesley, Computer Simulation of Liquids (Oxford University Press, Oxford, 1987).
[33] D. Frenkel, and B. Smit, Understanding Molecular Simulations (Academic Press, 2002).
[34] Y. Q. Zhou, and M. Karplus, Journal of Molecular Biology 293, 917 (1999).
[35] J. Kubelka, J. Hofrichter, and W. A. Eaton, Curr. Opin. Struct. Biol. 14, 76 (2004).
[36] A. Linhananta, S. Hadizadeh, and S. S. Plotkin, Biophys. J. 100, 459 (2011).
[37] A. M. Ferrenberg, and R. M. Swendsen, Phys. Rev. Lett. 61, 2635 (1988).
[38] J. Lee, and J. M. Kosterlitz, Phys. Rev. Lett. 65, 137 (1990).
[39] S. Kumar et al., J. Comput. Chem. 13, 1011 (1992).
[40] H. S. Chan, Proteins-Structure Function and Genetics 40, 543 (2000).
[41] M. Silow, and M. Oliveberg, J. Mol. Biol. 326, 263 (2003).