Accelerating protein unfolding simulations: Effect of ionic liquids on villin headpiece unfolding time

Hsin-Ju Tung

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2016

Reading Committee:

Jim Pfaendtner
David Beck

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

Abstract

Accelerating protein unfolding simulations: Effect of ionic liquids on villin headpiece unfolding time

Hsin-Ju Tung

Chair of the Supervisory Committee:
Professor Jim Pfaendtner, Ph.D.
Chemical Engineering

We demonstrate an approach to accelerate protein unfolding simulation using the metadynamics (MetaD) method, and the new rate calculation method “infrequent metadynamics”. The accelerated simulation time can be 200 ~ 700 times faster than normal molecular dynamics (MD) simulations. Using protein villin headpiece (HP35) as an example, we are the first to this approach to systematically investigate the effect of ionic liquids (ILs) on protein unfolding time. We reveal HP35 unfolding time in four different 20 % (w/w) ILs / water mixtures. Comparing to experimental results of ribonuclease A (RNase A), we find the anion’s ability to decrease protein unfolding time shares the same order as our simulation results. A detailed examination of protein and solvent behavior at molecular-level supports the current view in the literature about protein stability in ILs; namely that the ion-protein interaction play a more significant role than the ion-induced change in the bulk water structure. This study provides the understanding of effects of

ILs on protein stability and makes the achievement of calculating protein unfolding times in a greatly accelerated timescale.

TABLE OF CONTENTS

List of Figures vi

List of Tables vii

1. Introduction 10

1.1 Molecular Dynamics 10

1.2 Ionic liquids solvents effects on protein unfolding 11

2. Method 13

2.1 Infrequent metadynamics 13

2.2 Model system 14

2.2.1 Selection of collective variables 15

2.2.2 Optimizing simulation parameters 16

2.2.3 ILs system setup 18

3. Results and discussion 19

3.1 Validation of using infrequent metadynamics on protein unfolding 19

3.2 Comparison of HP 35 unfolding times in ionic liquid 20

4. Conclusion 24

Bibliography 25

Supporting Material 27

LIST OF FIGURES

Figure 2.1 Simulation workflow 14
Figure 2.2 Collective variables of HP35 16
Figure 2.3 Gaussian deposition stride vs p-value 18
Figure 3.1 Water oxygen-oxygen RDF in 20 % ILs (solvent-only) 21
Figure 3.2 Protein Cα-anion RDF in different ILs solution 23
Figure S1 Sampling example 27
Figure S2 Ions structure 28
Figure S3 Two dimensions free energy plot of HP35 28
Figure S4 Transition temperature Tm for the thermal denaturation of RNase [10] 29

LIST OF TABLES

Table 1.1 Abbreviations for complex ions of ILs 12
Table 2.1 The results of different CVs test 16
Table 2.2 Various deposition stride τ results of protein chignolin 18
Table 3.1 Mean unfolding time of chignolin and HP35 (2F4K)..... 19
Table 3.2 HP35 unfolding time results in ILs 20
Table 3.3 Time-average anion numbers within (Å) around protein 22
Table 3.4 MSD analysis of anions in the spherical shells of HP35. 22
Table S1 MetaD parameters setting of CVs test 29
Table S2 Summarize the MeteD parameter for HP35 unfolding simulation in ILs 29

ACKNOWLEDGEMENTS

I thoroughly enjoy working in PRG group, I feel proud to be one of the team. I would first like to thank my thesis advisor Professor Jim Pfaendtner. The door to Jim was always open whenever I ran into a trouble or had a question about my research or writing. He is always supportive and patient when talking to me. He consistently gave me right direction whenever he thought I needed it. I have learned a lot under his instructions. I would also like to thank all the PRG group members: Vance, Wesley, Kayla, Kelly, Josh, Arushi, Blake and Yeneneh. They are awesome lab mates. They always provide good solutions to my problems. I really enjoy working with them. I would also like to thank Professor David Beck as second reader of this thesis. He helped me build my interest and skills in data science. It became a useful tool for all the data analysis in this work. Lastly, simulations were able to complete by using the HYAK super computer at University of Washington.

DEDICATION

To all my family who supported me to have higher education opportunity and who gave me the chance to broaden my horizons of this world

1. INTRODUCTION

1.1 MOLECULAR DYNAMICS

Molecular dynamics (MD) simulation has become a standard tool used in many branches of science because its fully atomistic description can provide detailed mechanisms at the molecular level that are hard to observe in experiments. The principle of MD is the numerical, step-by-step, integration of the classical equation of motion, which for a simple atomic system may be written:

$$fi = m\ddot{r} \quad \text{and} \quad fi = \frac{\partial \mu}{\partial ri} \tag{1.1}$$

The purpose is to be able to calculate the forces fi acting on the atoms and then once we know the force, we can predict the atoms next move by solving the classical equations. The forces are derived from the potential energy u. The potential energy includes Non-bonded interaction (Lennard-Jones potential and Coulomb potential) and Bonding potential (bonds, angles and torsion). It is not hard to notice, the integration process is really computationally expensive. Thanks to the continuous and tremendous progress of computing power, MD simulation nowadays become more and more affordable to solve complex problems, interpret experimental results and make novel predictions.

However, in our study, the main challenge of using MD simulation to reveal protein unfolding time is that such dynamical features take place as the system moves from one free energy basin to another through infrequent rare events, which can occur after waiting times often well exceeding the millisecond time scale [1]. For example, in simple water system could take 100−2900 μs simulation time for normal MD, depending on the proteins [2], which need 2 ~ 58 years using 4 nodes 16 cores computer power. Not to mention, calculating protein unfolding time in ILs is expected to take longer due to viscosity effects. Metadynamics (MetaD) is a commonly used enhanced sampling method to accelerate simulations. In Metadynamics, sampling is accelerated by addition of a history-dependent bias potential, which is adaptively constructed in the space of carefully chosen collective variables (CVs) that encourage the system to escape from one local minimum energy state to another [3]. Furthermore, a recent extension of the MetaD approach by Tiwary and Parrinello, referred as “Infrequent Metadynamics”, provides a simple yet powerful method for calculating the rates of transition between different metastable

states [1]. This exciting new approach gives us a chance to explore protein transition times between folded and unfolded states in an affordable simulation time scale. In fact, this approach has successfully recovered alanine dipeptide transition time between two well-known conformations [1, 4] and the symmetric S_N2 chemical reaction rate [5]. The details of the approach will be discussed later. Most excitingly, we have shown this approach can accelerate protein villin headpiece and protein chignolin unfolding times in explicit water by a factor of 96 and 110 times, respectively. This breakthrough development allows one to systematically explore protein dynamics in a large accelerated timescale. On the basis of these developments in aqueous simulations, we immediately set out to study how is the protein unfolding times change in the presence of ionic liquid solvent.

1.2 IONIC LIQUIDS SOLVENTS EFFECTS ON PROTEIN UNFOLDING

Protein folding and unfolding in ionic liquids (ILs) has generated a lot of interest in both simulation and experimental communities due to the highly-tunable properties of ILs. It has been shown that an estimated ~10^6 combinations of known cations and anions can form ILs [6]. Thus, the resulting possibility to enhance or decrease thermal and functional stability of protein by manipulating ILs solvents environment can revolutionize chemical and biochemical methodologies. Since 1990s, several studies of rationally designed IL formulations affect protein stability have been reported [6–13]. However, up to now, no reliable method exists to predict the stabilizing properties of any novel ILs for biomolecules, although some guidelines have been reported [7]. A common guideline to predict a given IL's effect is the Hofmeister series, which classifies ions either kosmotropes (structure makers) or chaotropes (structure breakers) according to their relative abilities to induce the structuring of water [7, 13]. Ions of high charge and small size are believed to be kosmotropic, which globally enhance the H-bonded network and stabilize the protein. Large ions of low charge should act as chaotropes, which destroy this network and destabilize the protein. It has been known that this series has large effect on a wide range of biological phenomena, but they still remain unexplained on protein folding and unfolding. A widely quoted Hofmeister series ranking the anion according to their protein-stabilizing efficiency is [14, 15]:

$$[SO_4]^{2-} > [dhp]^- > [ac]^- > F^- > Cl^- \parallel > Br^- > I^- > [SCN]^-$$

The double bar (||) indicates the crossover from stabilizing to destabilizing behavior. Abbreviations for complex ILs are defined in Table 1.1. However, many studies have shown that protein stability in ILs does not always follow Hofmeister series. There are cases where the Hofmeister series is found reversed [16–18]. There is also evidence where strong denaturants were found to be acting as stabilizer to some extent. One study used differential scanning calorimetry (DSC) to systematically characterize the effect of ILs on the thermal denaturation of ribonuclease A (RNase A) and provide a slightly different anion series, in terms of decreasing T_m values [10]:

$[SO_4]^{2-} > [HPO_4]^{2-} > Cl^- > [EtOSO_3]^- > [BF_4]^- \approx Br^- > [MeOSO_3]^- > [TfO]^- > [SCN]^- \approx [dca]^- > [Tf_2N]^-$

Different than Hofmeister series, this series simply compare the destabilizing efficiency among each anion, instead of general effect on stabilizing/destabilizing. In general, anion variations have lager consequences and more efficient on protein stability than cation variations [6, 10]. We selected the common cation 1-butyl-3methylimidazolium $[BMIM]^+$ along with four different anions: Cl^-, $[dca]^-$, Br^-, and $[MeOSO_3]^-$, based on both Hofmeister series and experimental results. Some ions structures that used in this study are shown in Figure S2.

Table 1.1 Abbreviations for complex ions of ILs

$[BMIM]^+$	1-butyl-3methylimidazolium
$[dhp]^-$	Dihydrogen phosphate
$[ace]^-$	Acetate
$[EtOSO3]^-$	Ethylsulfate
$[dca]^-$	Dicyanamide
$[TfO]^-$	Trifluoromethanesulfonate
$[Tf_2N]^-$	Bis(trifluoromethanesulfonyl)imide
$[SCN]^-$	Thiocyanate
$[BF4]^-$	Tetrafluoroborate
$[MeOSO_3]^-$	Methylsulfate

HP35 is a 35-residue subdomain of the headpiece of actin-binding protein villin. It has been the target of wide variety of experimental and computational efforts to characterize its folding and

unfolding mechanism, due to its fast-folding property [19–24]. A wild type villin headpiece is known to fold in 4-5 microseconds; a fast-folding mutant exists which folds in under a microsecond. This allows researcher explore complete atomic scale of the folding mechanism by using MD simulation. However, there has been no exploration of dynamics of the HP35 in an IL containing system. A major reason for this is that to obtain the folding and unfolding times, one has difficulties and requires other supporting techniques as we discussed earlier. In the first part of this thesis, we demonstrate how to apply the infrequent metadynamics approach to accelerate protein unfolding simulations. The second part we will focus on the discussion of ionic effects on HP35 unfolding time.

2. METHOD

2.1 INFREQUENT METADYNAMICS

Infrequent metadynamics is an algorithm that is applied to the results of a MetaD simulation that has biased one rare event to occur. In this approach, an acceleration factor α was introduced [1]. An individual transition time (t^{eff}) can be calculated as summation of the MetaD time step (Δt) multiplied by the acceleration factor, which is derived from the instantaneous value of the growing MetaD potential $V_{bias}(s,t)$ at i^{th} point in the transition trajectory:

$$\alpha \approx e^{\beta V_{bias}(s,t)} \tag{2.1.1}$$

$$t^{eff} = \sum_{i=0}^{N} \Delta t_i^{MetaD} e^{\beta V_{bias}(s,t)} \tag{2.1.2}$$

where $\beta = (1/k_B T)$ is the inverse of temperature multiplied by the Boltzmann constant. The mean transition time is ensemble average of a set of individual transition times. To be more clear, every time when the rare events happened, we stopped the simulation and calculated the individual transition time (t^{eff}). After collecting 20 ~ 30 individual events, we can calculate the mean transition time. The underlying principle is that, in order to calculate correct transition time, two conditions must to meet when performing infrequent metadynamics algorithm 1) The biased

CVs must be able to distinguish between all relevant states in the system 2) No bias added in the transition state (TS) [4]. The first condition can be met by carefully choose representative CVs. The second condition can be met by increasing the time lag between successive Gaussian depositions (increasing the stride τ values). A statistical p-value analysis was performed to check whether these two conditions are satisfied or not. According to Tiwary and Parrinello's previous study, if the transition between two basins is rare event type, the transitions are independent from one another, the statistics of transition times must follow a Poisson distribution, the so-called law of rare events [4]. A Kolmogorov-Smirnov analysis with p-value reflects that the probability of the data extract from MetaD follow a Poisson distribution. The conventional assumption is to set p-value equals 0.05. The null hypothesis is accepted if p-value > 0.05. Other statistic characteristics include its mean value μ, standard deviation σ and median t_m. According to their paper, μ/σ and $\mu\ ln2/t_m$ value should be 1 in an exponential distribution Figure 2.1 showed the MetaD simulation along with infrequent metadynamics algorithm workflow. A bootstrapping sampling method is performed to obtain some statistical uncertainty of our results. The sampling method is shown in supporting material.

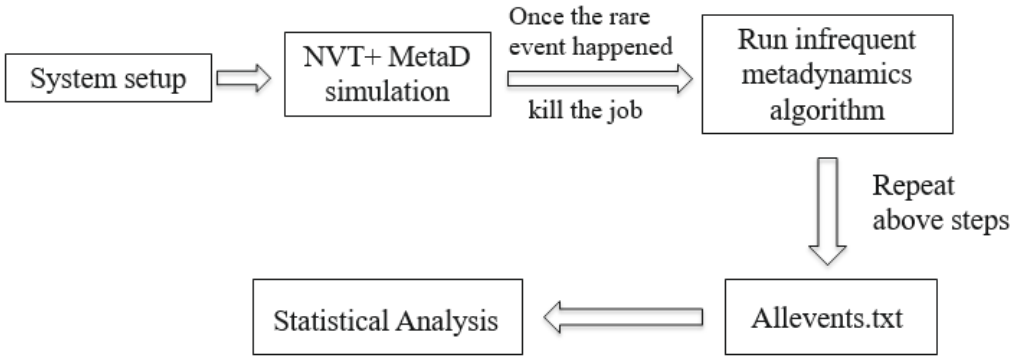

Figure 2.1 Simulation workflow

2.2 MODEL SYSTEM

From our understanding, this study is the first time using infrequent metadynamics approach to accelerate protein unfolding simulation. Before we are able to correctly use it, we did some trial and error to understand it more. The main challenges are finding the right CVs, knowing when

the system is in the transition state, and avoiding the biased potential added on the transition state. This section will go through how we got there and provide our final simulation protocol.

2.2.1 *Selection of collective variables*

A previous study by DEshaw research group has reported several proteins unfolding time and their CVs cutoff, which performed over 100 μs MD simulation by their supercomputer Anton [2]. It has been used to verify our unfolding simulation. Selection of correct collective variables that can distinguish between folded and unfolded states is our first problem. We have tested the method on three different CVs that have been used a lot to describe protein folding/unfolding in MetaD simulations: root-mean-square deviation (RMSD) of the Cα atoms, alpha helical contact (COOR) and hydrophobic contact (SS) of protein HP35 (PDB: 1YRF), see Figure 2.2. We first carried out four trials of 50 ns MetaD simulation biasing on COOR and SS to have general idea when unfolding happened in the CV space. Figure S3 shows the two dimensional free energy plots of those four trials. We can see that all of them have different folded states (indicated in the square box) and each of them may be not converged yet. However, based on the plots, we roughly defined the transition state cutoff where CVs escape from the folded region. Two cutoffs has been used, the square cutoff: COOR < 12 and SS < 32 and the circle cutoff: within radius = 1.0, center (COOR, SS) = (14, 35). From Table 2.1, we see that square cutoff failed the p-value test, circle cutoff although has p-value > 0.05, its other statistical test are way off, and the unfolding 0.05 μs is not close to what we saw in the literature. But, as we know, low p-value and failed statistical test might be caused from the fast Gaussian deposition rate (τ) so that the potential has been added on the transition state. Therefore, this result does not give us too much information. We later used an RMSD cutoff 0.4 nm for HP35 from DEshaw paper and the results are shown in Table 2.1.

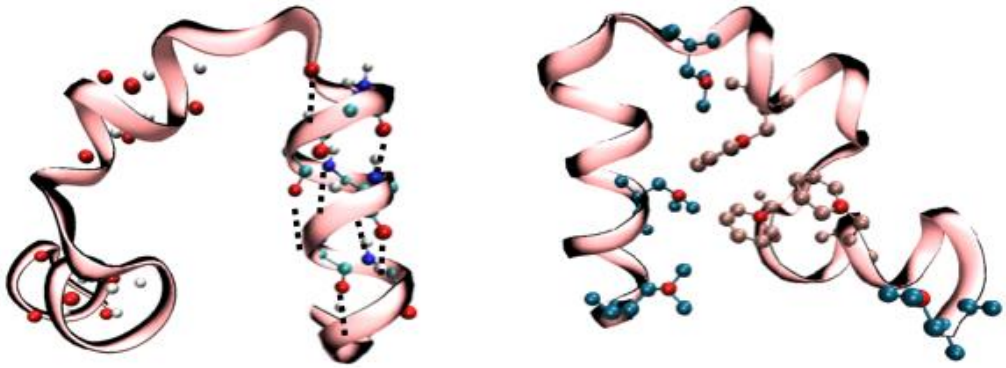

Figure 2.2 Collective variables of HP35 (a) Alpha helical contact, oxygen atom of carboxyl groups and hydrogen atom of amino groups (b) Hydrophobic contact, second carbon atom (red) of hydrophobic residue LEU (green) and PHE (gray)

Table 2.1 The results of different CVs test

CVs	(COOR, SS)	(COOR, SS)	RMSD
Cutoff	Square	Circle	0.4 nm
T_uf [μs]	191.22	0.05	3.26
p-value	6.58×10^{-7}	0.24	0.123
μ/σ	0.272	0.132	0.49
$uln2/t_m$	29.5	93	2.74

It can be seen that using RMSD as cutoff has overall better results and the unfolding time makes more sense to us. In addition, RMSD cutoff is easier to define than COOR and SS cutoff. Therefore, we decided to use RMSD as our CV for the other simulations, and referenced the protein unfold cutoff from DEshaw Research group. Table S1 provides further simulation details.

2.2.2 *Optimizing simulation parameters*

After deciding using RMSD as our CV, next step is to make sure no bias added on the transition state. In the other words, the MetaD Gaussian hills deposition rate between two steps should be long enough that a transition time can occur without additional hills being deposited in the TS. We tuned the wait periods between depositions of consecutive Gaussian hills deposition stride (τ) of protein chignolin (PDB: 1UAO) simulation, which is also one of the proteins reported in

DEshaw group's paper. In order to compare our results to the reference, these simulations were performed using Charm22* force field developed by DEshaw research group [25], implemented in GROMACS 4.6 and using the TIP3P water model. Systems were minimized using a steepest descent method for 5000 steps. For our MD simulations, a time step of 2 fs was performed. Periodic boundary conditions were applied to all dimensions. Lennard-Jones interaction cutoff is 1.2 nm and shifted to avoid discontinuities in energy. Particle mesh Ewald summation is used to account for the long-range electrostatic interaction. Temperature was maintained using the velocity rescaling algorithm [26] and pressure was held at 1 bar using a Berendsen barostat [27]. We equilibrated the system in the NPT ensemble for 1 ns and running simulation in NVT ensemble. The simulation temperature is 360 K and the Cα - RMSD cutoff 0.6 nm was defined as transition state based on the results from DEShaw.

Simulations were biased with MetaD method. Our biased simulations were performed with PLUMED [28] and used the well-tempered metadynamics [3], which the decreases the deposition rate over the time. The so-called biasfactor (γ) was set to 10. The starting hill height was 1.0 kJ/mol. The Gaussian widths of σ = 0.025 nm. A variety of deposition stride τ values have been tested. The results are shown in Figure 2.3 and Table 2.2. T_uf indicates the mean unfolding time for a given series or ensemble of simulations. A red dashed line divide the p-value threshold 0.05. We can see that except τ = 2 and 5 ρs, the p-values of all deposition strides are above recommended threshold of 0.05. As we expected, the p-value started to converge when the deposition rate became slower. τ = 100 ρs and 120 ρs have similar p-value and unfolding time results and τ = 100 ρs almost gives us same unfolding time to the literature (2.2 μs). By tuning the Gaussain hills deposition stride (τ), we can find the optimized MetaD parameters that both meet the infrequent metadynamics requirements and able to reproduce the unfolding time obtained from 106 μs long normal MD simulation. Noting that, in order to make sure τ values are not case by case and to safely apply the τ value to other system, we actually chose τ = 120 ρs for the latter simulations.

Table 2.2 Various deposition stride τ results of protein chignolin

$\tau(\rho s)$	T_uf (μs)	p-value	Events record
2	12732	0.019	45
5	1328	0.044	45
10	43.16	0.118	45
14	48.12	0.236	45
20	8.17	0.42	45
50	4.16	0.235	45
60	5.50	0.386	45
100	2.11	0.489	45
120	1.12	0.498	45

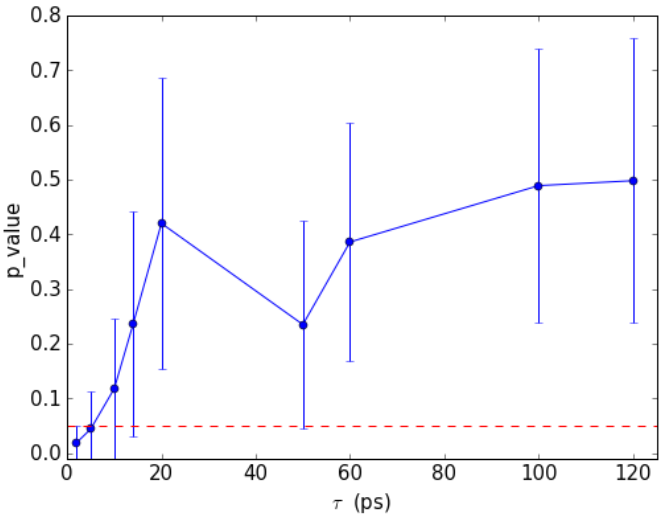

Figure 2.3 Gaussian deposition stride vs p-value

2.2.3 *ILs system setup*

ILs systems set up are as same as describe above, except using AMBER99SB-ILDN force field. The Gaussian deposition stride (τ) are using the optimized one τ = 120 ρs. The ionic liquid force field was obtained from the code developed by Vance Jaeger [29] which calculates ion's partial charges from quantum calculations using Gaussian 09 [30]. We scaled the charges by 0.8 to more accurately represent dynamics properties of solvent, as has been demonstrated in the literature [31]. Solvent boxes with periodic boundary condition are added using the Packmol software package. Wild type HP35 (PDB: 1YRF) was used and simulation were performed at

moderate temperature 330K. Five sets of simulations, each representing HP35 unfolding events, have been carried out in different solvents including, pure water and 20% (w/w) [BMIM][Cl], [BMIM][dca], [BMIM][Br], [BMIM][$MeOSO_3$] ILs/water co-solvent. The Cα - RMSD cutoff 0.4 nm was defined as transition state as reported from the reference [2].

3. RESULTS AND DISCUSSION

3.1 VALIDATION OF USING INFREQUENT METADYNAMICS ON PROTEIN UNFOLDING

In the methods section, we actually have successfully used infrequent metadynamics approach to reproduce protein chignolin unfolding time by tuning MetaD Gaussian hills stride (τ) value. However, in order to have more evidences to prove this approach works, specifically for the protein HP35, we adopt the same simulation settings that used to reproduce protein chignolin to reproduce HP35 unfolding from DEshaw research group's paper. Since they used mutant HP35 (PDB: 2F4K), which substitute two buried lysine residues by norleucine residues, we here used same PDB file to have complete comparison to their results. Table 3.1 summarized the results of using our approach and normal MD simulation. The acceleration factor showed our simulations are 110 and 96 times faster than normal MD simulation. If we compare the total simulation time of the two studies, we reduced the total computational cost by a factor of about 262 and 431 times for chignolin and HP35, respectively. With two examples shown as evidence here, we have proved that we can obtain protein unfolding time in a massively accelerated timescale by applying the infrequent metadynamics algorithm. Most importantly, this success allows us to explore protein unfolding times in the IL solvents, which has never been done before.

Table 3.1 Mean unfolding time of chignolin and HP35 (2F4K)

Protein	T (K)	T_uf (μs)	*T_uf (μs)	α	MD time	*MD time
Chignolin	340	2.11 (0.5)	2.2 (4)	110	10.1 ns/events	106 μs
HP35(PDB:2F4K)	360	0.83 (0.17)	0.9 (2)	96	6.9 ns/events	125 μs

T_uf is the mean unfolding time. Standard errors are reported in parentheses.
MD time is average simulation time per event (Chignolin: 40 events; HP35: 42 events)
* represent reference data [2]
α is acceleration factor

3.2 COMPARISON OF HP 35 UNFOLDING TIMES IN IONIC LIQUID

Table 3.2 showed the results of protein HP35 unfolding times in 20% ILs with $[BMIM]^+$ as cation and in pure water. Rejecting rate was calculated from the rejected events out of 2000 iterations of bootstrapping procedure. Table S2 summarized the MetaD parameters we used. In terms of unfolding times, the resulting anion series reads: $Cl^- > Br^- > [MeOSO_3]^- > [dca]^-$. Interestingly, this is in an agreement with the experimental data from protein RNase A (Figure S4), the anion's ability of decrease protein unfolding time share the same order. The increase of hydrophobicity of the anion is seen to have a destabilizing effect on the protein. When compared to the water the other ILs all been observed to be destabilizing to the proteins with the exception of Cl^-. As we expected, ILs have dramatic effects on protein stability. Due to the fact that cations have been shown to have less effect of protein stability than anions, we focus here on discussing the anions effect. The similar anions ranking to the experimental result suggests the anions have specific interaction within the system, based on their nature, size, and properties.

Table 3.2 HP35 unfolding time results in ILs

Solution	T_uf (μs)	p-value	Rejecting rate	α	MD time (ns/event)	Events record
Water	10. 0 (2.17)	0.44 (0.24)	0.047	688	12.7	26
[BMIM][Cl]	10.49 (3.72)	0.35 (0.22)	0.05	701	11.6	27
[BMIM][Br]	8.70 (2.18)	0.41 (0.24)	0.07	571	9.72	23
[BMIM][MTS]	8.56 (2.5)	0.315 (0.22)	0.15	488	12.65	28
[BMIM][DCA]	2.588 (0.65)	0.4 (0.24)	0.01	220.1	12.39	29

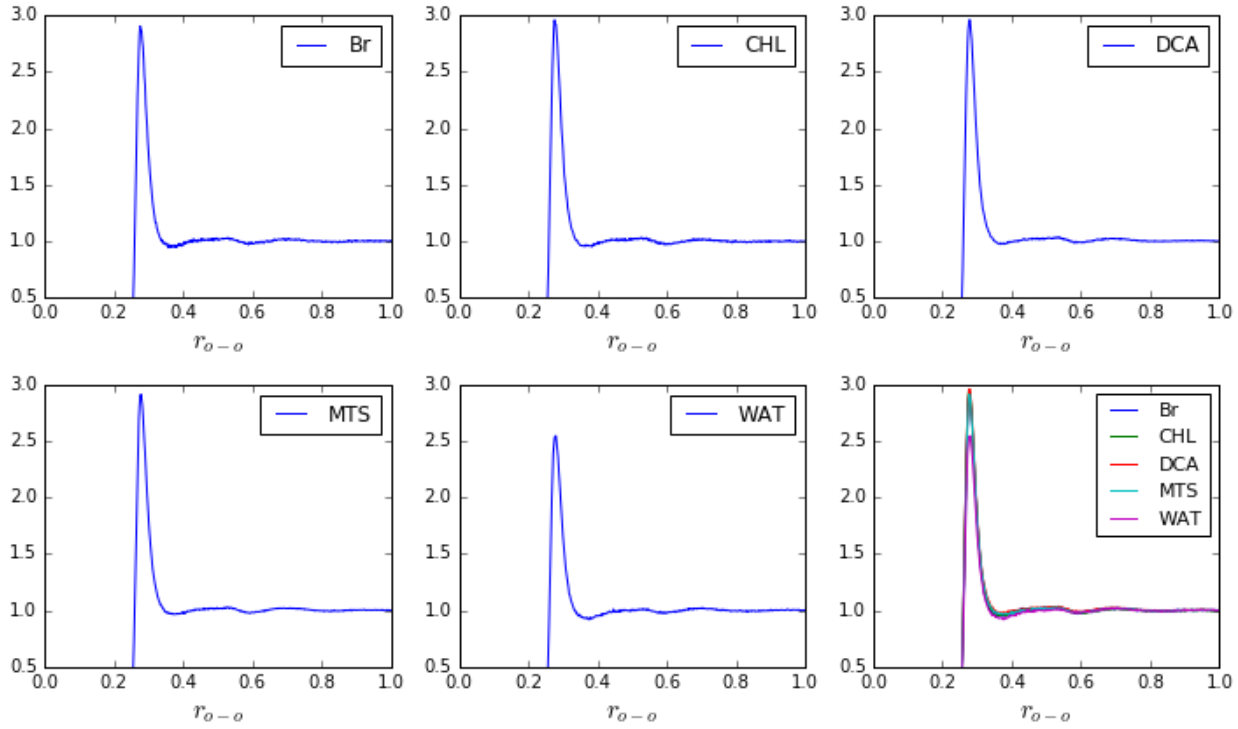

Figure 3.1 Water oxygen-oxygen RDF in 20 % ILs (solvent-only)

Well-known in the literature, the Hofmeister effect on proteins proposes that the ion-induced changes in the water structure increase or decrease hydrogen bond networking of water and consequentially affect protein fold and unfold rates. We performed 10 1ns solvent-only simulations and averaged their water radial distribution function (RDF). Figure 3.1 shows that the bulk water structure in different solvents have no difference. Except that the pure water has a slightly lower peak, the other ILs solvent have all same RDF 1st shell peak at ~ 0.3 nm. This indicates that the ion-induced water structure changes do not have significant influence in our system. Another possible hypothesis comes from a paper published by Joseph L. Baker and Jeffrey Furbish [8]. They examined the effect of the ionic liquid [C_4mpy][Tf_2N] on the protein Trp-cage using MD simulation. Starting from an unfolded structure, they found Trp-cage in the IL showed the tendency to fold back to native state at high temperature, while it stayed in unfolded state at low temperature. They suggested protein folding and unfolding requires the ILs molecules move out of the way of the protein, allowing the intra-bonding interactions of the protein to form. Simply speaking, ILs molecules will restrict the protein's motion. Three trials of

5 ns normal MD simulations of protein-solvent systems have been carried out. The time-average ions number around HP35 and the mean square displacement (MSD) have been calculated using python package MDAnalysis. Table 3.3 showed the anion $[dca]^-$ has highest numbers within 4 Å while it has fastest unfolding time. Table 3.4 provided the MSD value in concentric spherical shells about HP35 geometric center and bin sizes 5 Å. In the sense of ILs restricting the protein motion hypothesis, the higher MSD value should give faster unfolding times, because high mobility of ions decrease the chances to block the protein intra-interaction. However, our MSD results do not have significant correlation between the unfolding time.

Table 3.3 Time-average anion numbers within (Å) around protein

Solution	2	4	6	8	10
[BMIM][Cl]	0.46	5.48	11.13	17.75	25.44
[BMIM][DCA]	2.211	10.579	15.308	21.374	28.057
[BMIM][MTS]	3.41	9.6	13.85	18.87	24.78
[BMIM][Br]	0.04	5.41	10	15.4	21.77

Table 3.4 MSD analysis of anions in the spherical shells of HP35. R=12 Å, bin sizes = 5 Å

Solution	20-25	25-30	30-35	35-40
[BMIM][Cl]	160	351.3	527.7	802.3
[BMIM][DCA]	175.9	286	429.9	667.14
[BMIM][MTS]	285.5	349.4	424	523.2
[BMIM][Br]	558.7	394.4	595.63	600.6

Although several studies have shown that high concentration ILs increase the stability of protein due to their high viscosity, in other word, ILs indeed have effect on hindering the protein's motion, we think 20 % (w/w) ILs solvent might not high enough to observe this phenomenon in this study. We noticed that the average number of anions within 4 Å around protein have sort of reverse trend of the unfolding time. We further calculated the radial distribution function (RDF) to gain more concrete picture of how the anions around protein behave. The values of the RDFs presented here are the average anions density varies as functions of distance from the protein Cα atoms. Figure 3.2 shows the RDF profiles for each ILs solutions. Clear trend can be seen near 0.5 nm from Cα, $[dca]^-$ anion has highest peak, follow by $[MeOSO_3]^-$, Br^- and Cl^-. Interestingly, it has a totally reverse order of our unfolding results, which suggests that the high density of anions around protein does not restrict the protein's motion, instead it helps protein unfold faster.

It becomes more and more clear that anion and protein interaction has significant effect on protein stability. Another interesting feature of RDF but not relevant to our study is that, owing to the anion structure, $[dca]^-$ and $[MeOSO_3]^-$ have similar RDF profile which both have another raising peak near 1.0 nm.

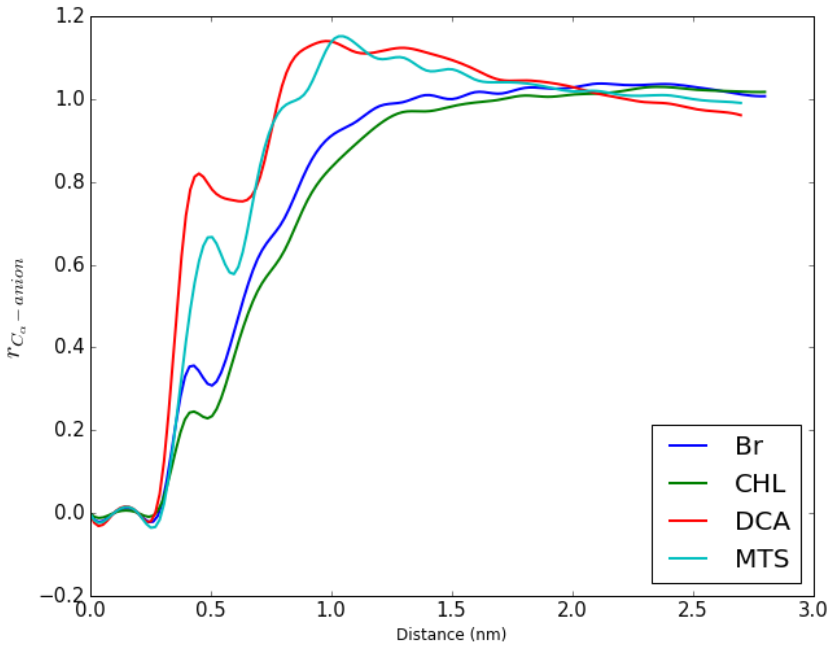

Figure 3.2 Protein Cα-anion RDF in different ILs solution

According to average numbers around protein and RDF profiles, we hypothesize it is the interactions between the anion and protein that affect the protein's unfolding time. From the HP35 structural point of view, the three main conserved phenylalanine residues (PHE 6, PHE 10 and PHE 17) establish the stability of the three alpha helices [22]. We hypothesize that the strong hydrophobic nature of the anion, upon its interaction with the hydrophobic cores of HP35, causes the unfolding time of the protein to decrease. The hydrophobic cores of protein and hydrophobic anions attract each other, leading to an increased presence of the strongly hydrophobic anions around the protein and thus destabilization of the protein Conversely, the RDF profiles show a low density of weakly hydrophobic anions such as Cl^- around the protein, indicating they have little-no-no effect on the unfolding time of the protein. Additionally, the surrounding cation $[BMIM]^+$ might hinder the protein's motion, causing it to have a slightly higher unfolding time

compared to in water (i.e. a stabilizing effect). To summarize, (1) our results indicate protein stability in ILs is largely influenced by ion-protein interactions rather than by the water bulk structure effect, (2) the presence of anions with strong hydrophobicity around the protein leads to faster protein unfolding times, and (3) predicting a priori whether a given IL will have a stabilizing or destabilizing effect on a protein compared to water depends on complex ion-protein interactions.

4. CONCLUSION

The infrequent metadynamics approach is a powerful post-processing method to extract relevant dynamic information about a process from a biased simulation. Herein, we have comprehensively shown this approach can be used to predict protein unfolding in greatly accelerated timescales. Specifically, we used this approach to systematically analyze the unfolding of the protein HP35 in IL solvents. In this study, we provide unique atomic level insight into the interactions between proteins and IL solvents, and help to resolve several hypotheses previously made in the literature. Although Hofmeister effect has provided a guideline for protein stability for years and seems to be sufficiently general to explain some biological events, there are many exceptions to this rule. For example, while our results indicate the anions Cl^- and Br^- follow the Hofmeister series in terms of their relative influence on protein stability, our analysis of the structure of water in ILs does not support the Hofmeister theory. A new explanation of the effects of ILs on protein stability arises from our study. The RDF profiles between the protein and surrounding anions combined with analysis of the average number of surrounding anions indicates the interaction between hydrophobic anions and the hydrophobic core of the protein plays a decisive role in determining whether the IL will stabilize or destabilize the protein. Strongly hydrophobic anions interact with the protein more than weakly hydrophobic anions and thus lead to faster protein unfolding times. From this molecular insight, we conclude that our results support the current view in the literature that ion-protein interactions play a more significant role in determining protein stability than ion-induced bulk water changes.

Our study provides a new strategy to understand the effects of IL behavior on protein stability, which can help to progress many biochemical and bioprocess applications. In addition, the

computational approach here is convenient to apply and affordable on standard computational power. Although it is difficult to generalize the IL's effect on protein stability in all types of situations, we expect that the understanding of it will evolve as the ion-protein interactions become clearer with additional testing.

BIBLIOGRAPHY

[1] P. Tiwary and M. Parrinello, "From metadynamics to dynamics," *Phys. Rev. Lett.*, vol. 111, no. 23, pp. 1–5, 2013.

[2] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold.," *Science (80-.).*, vol. 334, no. 6055, pp. 517–520, 2011.

[3] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 1, no. 5, pp. 826–843, 2011.

[4] M. Salvalaglio, P. Tiwary, and M. Parrinello, "Assessing the reliability of the dynamics reconstructed from metadynamics," *J. Chem. Theory Comput.*, vol. 10, no. 4, pp. 1420–1425, 2014.

[5] K. L. Fleming, P. Tiwary, and J. Pfaendtner, "A New Approach for Investigating Reaction Dynamics and Rates with Ab Initio Calculations," *J. Phys. Chem. A*, p. acs.jpca.5b10667, 2015.

[6] H. Weingärtner, C. Cabrele, and C. Herrmann, "How ionic liquids can help to stabilize native proteins.," *Phys. Chem. Chem. Phys.*, vol. 14, no. 2, pp. 415–26, 2012.

[7] A. Kumar and P. Venkatesu, "Does the stability of proteins in ionic liquids obey the Hofmeister series?," *Int. J. Biol. Macromol.*, vol. 63, pp. 244–253, 2014.

[8] J. L. Baker, J. Furbish, and G. E. Lindberg, "Journal of Molecular Graphics and Modelling Influence of the ionic liquid [C 4 mpy][Tf 2 N] on the structure of the miniprotein Trp-cage," vol. 62, pp. 202–212, 2015.

[9] K. Fujita, D. R. Macfarlane, M. Forsyth, M. Yoshizawa-fujita, K. Murata, N. Nakamura, and H. Ohno, "Solubility and Stability of Cytochrome c in Hydrated Ionic Liquids : Effect of Oxo Acid Residues and Kosmotropicity," pp. 2080–2086, 2007.

[10] D. Constantinescu, H. Weingärtner, and C. Herrmann, "Protein Denaturation by Ionic Liquids and the Hofmeister Series : A Case Study of Aqueous Solutions of Ribonuclease A **."

[11] F. Van Rantwijk and R. A. Sheldon, "Biocatalysis in Ionic Liquids," 2007.

[12] S. Studies, "Protein Structure and Dynamics in Ionic Liquids . Insights from Molecular Dynamics," pp. 2566–2572, 2008.

[13] R. D. Hutton, J. Wilkinson, M. Faccin, E. M. Sivertsson, A. Pelizzola, A. R. Lowe, P. Bruscolini, and L. S. Itzhaki, "Mapping the Topography of a Protein Energy Landscape,"

2015.
[14] K. D. Collins and M. W. Washabaugh, “The Hofmeister effect and the behaviour of water at interfaces.,” *Q. Rev. Biophys.*, vol. 18, no. 4, pp. 323–422, Nov. 1985.
[15] W. Kunz, P. Lo Nostro, and B. W. Ninham, “The present state of affairs with Hofmeister effects,” vol. 9, pp. 1–18, 2004.
[16] M. M. Ries-kautts and A. F. Ducruix, “Relative Effectiveness of Various Ions on the Solubility and Crystal Growth of Lysozyme *,” vol. 264, no. 2, pp. 745–748, 1989.
[17] A. Tardieu, “The Hofmeister effect as seen by SAXS in protein solutions,” vol. 9, pp. 112–116, 2004.
[18] Y. Zhang and P. S. Cremer, “The inverse and direct Hofmeister series for lysozyme,” vol. 106, no. 36, 2009.
[19] L. E. Packer, B. Song, D. P. Raleigh, and C. J. Mcknight, “Competition between Intradomain and Interdomain Interactions: A Buried Salt Bridge Is Essential for Villin Headpiece Folding and Actin Binding,” pp. 3706–3712, 2011.
[20] M. Wang, Y. Tang, S. Sato, L. Vugmeyster, C. J. Mcknight, and D. P. Raleigh, “Dynamic NMR Line-Shape Analysis Demonstrates that the Villin Headpiece Subdomain Folds on the Microsecond Time Scale,” pp. 6032–6033, 2003.
[21] J. K. Chung, M. C. Thielges, and M. D. Fayer, “Dynamics of the folded and unfolded villin headpiece (HP35) measured with ultrafast 2D IR vibrational echo spectroscopy,” vol. 108, no. 9, pp. 1–6, 2011.
[22] T. K. Chiu, J. Kubelka, R. Herbst-irmer, W. A. Eaton, J. Hofrichter, and D. R. Davies, “High-resolution x-ray crystal structures of the villin headpiece subdomain , an ultrafast folding protein,” vol. 87, pp. 1–6, 2005.
[23] F. B. Activity, J. Meng, D. Vardar, Y. Wang, H. Guo, J. F. Head, and C. J. Mcknight, “High-Resolution Crystal Structures of Villin Headpiece and Mutants with Reduced,” pp. 11963–11973, 2005.
[24] A. Ferna, M. Shen, T. R. Sosnick, R. S. Berry, and K. F. Freed, “Large-Scale Context in Protein Folding : Villin Headpiece †,” pp. 664–671, 2003.
[25] S. Piana, K. Lindorff-larsen, and D. E. Shaw, “How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization ?,” *Biophysj*, vol. 100, no. 9, pp. L47–L49, 2011.
[26] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity-rescaling.”
[27] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, “Molecular dynamics with coupling to an external bath Molecular dynamics with coupling to an external bath,” vol. 3684, no. 1984, 2012.
[28] L. Kleinman and D. M. Bylander, “Efficacious Form for Model Pseudopotentials,” *Phys. Rev. Lett.*, vol. 48, no. 20, pp. 1425–1428, May 1982.
[29] V. W. Jaeger and J. Pfaendtner, “Structure, Dynamics, and Activity of Xylanase Solvated in Binary Mixtures of Ionic Liquid and Water,” 2013.
[30] G. A. P. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, *Gaussian 09, Revision B.01*. .
[31] Y. Zhang and E. J. Maginn, “A Simple AIMD Approach to Derive Atomic Charges for Condensed Phase Simulation of Ionic Liquids,” 2012.

SUPPORTING MATERIAL

Sampling Method

For every set of simulations, we performed a sampling procedure by randomly selecting (with replacement) subsets of data and calculated p-value. If p-value < 0.05 which suggested this randomly chosen subset didn't follow Poisson process, we then rejected this subset. Rejecting rate was calculated from the rejected events out of 2000 iterations of sampling procedure. The same sampling method also used for sampling mean unfolding time, p-value and other statistic value. To rule out the outliers from the original data sets, we suppose the data (X) are Poisson distributed data, and transform the data to $Y = 2\sqrt{X}$, and exclude an observation as an outlier if $Y > mad(Y) + 3$, where mad is mean absolute deviation. The Figure S1 shows an example:

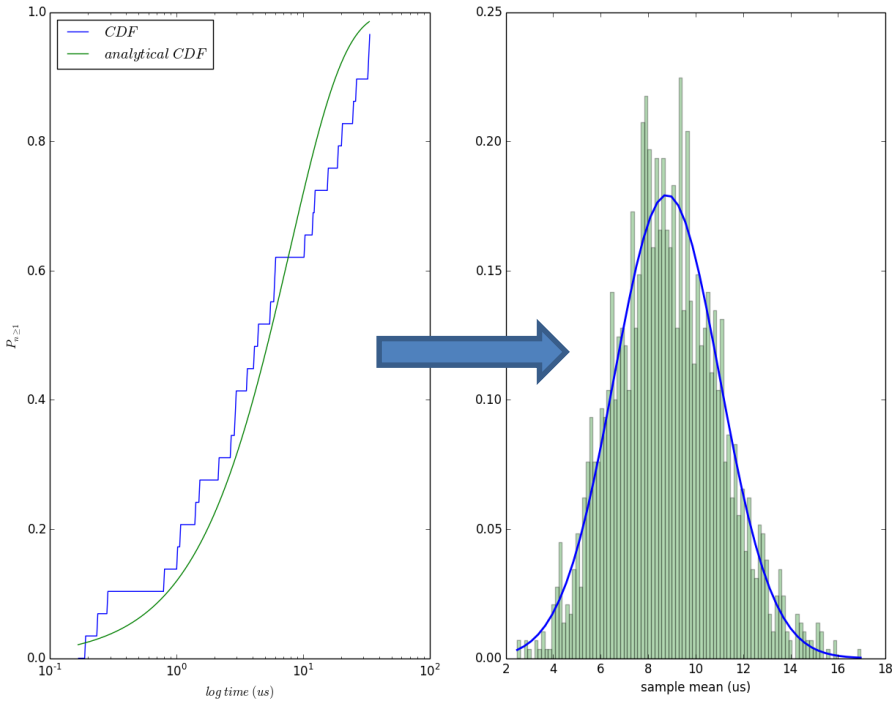

Figure S1 Sampling example Left plot is the cumulative distribution functions and analytical CDFs of original datasets (23 events) from [BMIM][Br] IL's unfolding simulation.

Right plot is the normal distribution results from randomly select 20 events and do the same p-value and ensemble average of individual transition time analysis 2000 times.

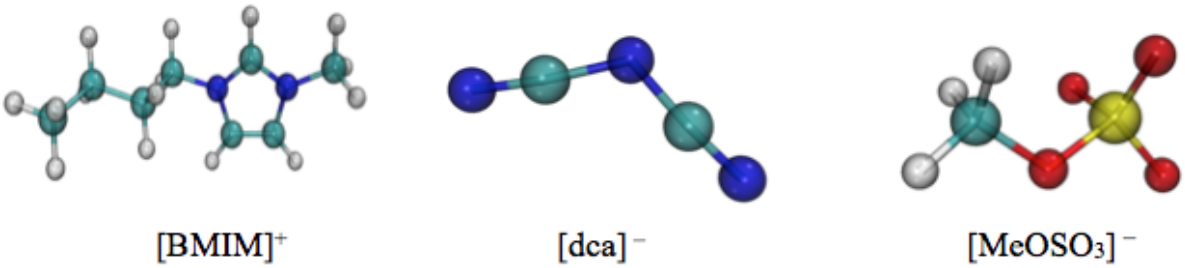

Figure S2 Ions structure

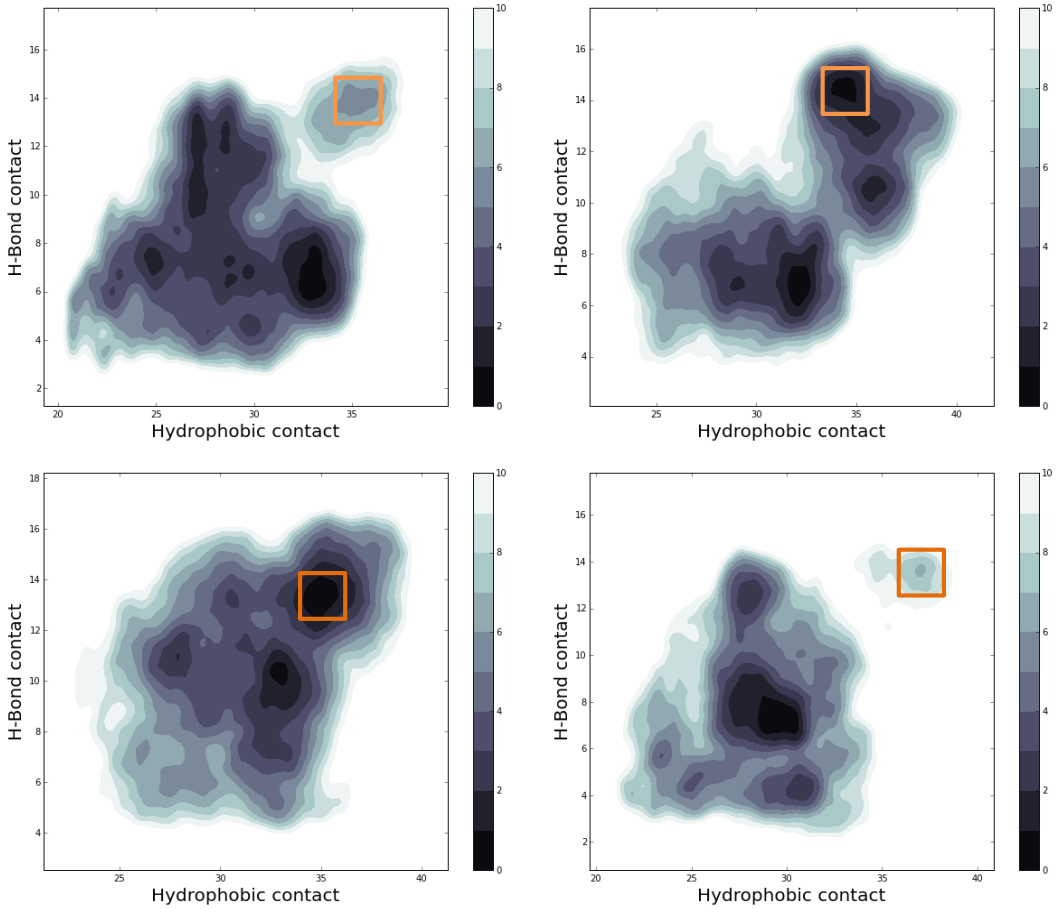

Figure S3 Two dimensions free energy plot of HP35 biasing on alpha helical contact and hydrophobic contact. Four trials. 290K, 50ns MetaD simulation. Orange box indicate folded region.

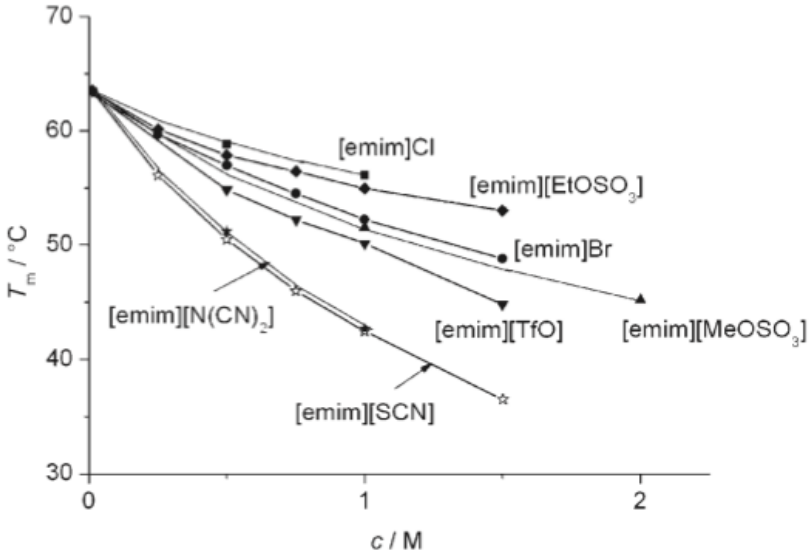

Figure S4 Transition temperature T_m for the thermal denaturation of RNase A as a function of the concentration c of added ILs with [emim] as a common cation. [10] 20 % (w/w) ILs ~ 1M

Table S1 MetaD parameters setting of CVs test

MetaD setting	(COOR,SS)	RMSD
Deposition stride (τ)	1000	2500
Sigma	0.3	0.025
Gaussian Height (kJ/mol)	1.0	0.5
T (K)	290	360
Biasfactor (γ)	10	10

Table S2 Summarize the MeteD parameter for HP35 unfolding simulation in ILs

MetaD setting	
Deposition stride (τ)	60000
Sigma	0.025
Gaussian Height (kJ/mol)	1
T (K)	330
Biasfactor (γ)	10
RMSD cutoff	0.4 nm