COMPUTATIONAL STUDIES OF THE ENERGETICS AND DYNAMICS OF PROTEIN-PROTEIN BINDING

by

Reza Salari

B.S., University of Tehran, 2005

M.D., Tehran University of Medical Sciences, 2008

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Reza Salari

It was defended on

September 13, 2012

and approved by

Rob D. Coalson, Professor, Chemistry

W. Seth Horne, Assistant Professor, Chemistry

Daniel M. Zuckerman, Associate Professor, Computational and Systems Biology

Thesis Advisor: Lillian T. Chong, Associate Professor, Chemistry

COMPUTATIONAL STUDIES OF THE ENERGETICS AND DYNAMICS OF PROTEIN-PROTEIN BINDING

Reza Salari, PhD

University of Pittsburgh, 2012

Protein-protein binding is crucial to various processes in living organisms including signal transduction and cell regulation and also plays a central role in various diseases. Therefore, detailed understanding of protein binding is of great importance and is an active area of research in many fields including chemistry, molecular biology and biophysics. In this dissertation, a series of five computational studies were completed to provide molecular details of the energetics and dynamics of model protein-protein complexes.

The first three studies focus on the role of solvent in protein-protein binding. The presence of solvent is very important to the formation of protein-protein complexes through both favorable and unfavorable contributions. For example, the extent to which that salt bridges contribute to the binding stability is predominantly determined by their desolvation penalties, which is difficult to examine experimentally but has been previously studied using implicit solvent models. Here, extensive implicit and explicit solvent simulations were carried out to directly compare the two solvent models in estimating the desolvation penalties of salt bridges upon protein binding. In addition, the effects of high temperature and salt concentration on the desolvation penalties were also explored.

In the fourth study, molecular simulations were employed to model rearrangements of an intermolecular beta sheet in a protein-peptide complex, providing insight into how nature might correct for mistakes in binding orientation for protein-protein interactions involving the

formation of beta sheets. The rearrangement mechanism includes a hydrophobic residue of the peptide anchoring itself to a transient hydrophobic pocket on the protein and helping the peptide to “crawl” back to its native state.

Finally, in the fifth study, the relative stabilities of the dimeric and newly discovered trimeric states for a model coiled-coil protein, the GCN4 leucine zipper were compared in isolation. Parallel tempering molecular dynamic simulations in implicit solvent, performed on the microsecond timescale, revealed that while the dimer fold is more stable at room temperature, both oligomers have similar stabilities at temperatures well below the melting temperatures and therefore the same sequence can populate both folds depending on the environment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This thesis would not have been possible without the support of many wonderful and talented people. First and foremost, I would like to express my gratitude to my advisor Professor Lillian Chong for her continuous support throughout my graduate studies and for her patience, guidance and encouragement. I would also like to thank the members of my thesis committee, Professor Rob Coalson, Professor Seth Horne and Professor Daniel Zuckerman for their participation and insightful discussions.

I am also thankful to all members of Chong group for helpful discussions and creating such an enjoyable working environment. I extend thanks to XSEDE and University of Pittsburgh CMMS for computational resources. Additionally, I would like to thank the Chemistry department staff which has been so ready to lend a hand when needed; in particular Fran Nagy and Elaine Springel.

Finally, I would like to thank my family for their wonderful support. To Maria, words cannot express how much I love you and how much I am looking forward to our time together. Thank you for all your support.

1.0 INTRODUCTION

The focus of this dissertation is to study the energetics and dynamics of protein-protein binding using computational simulations. The formation of specific protein complexes is crucial to life and is involved in many biological processes such as cell regulation, signal transduction and the immune system. In addition, pathological protein binding is the cause of many diseases including toxin-mediated infections and Alzheimer's. Therefore it is of great importance to characterize protein-protein interactions in molecular detail in order to better understand protein binding.

In biological systems protein binding occurs in the solution environment and is the result of a subtle balance between both favorable interactions (e.g. hydrophobic effect, electrostatic interactions between oppositely charged residues and hydrogen bond formation) and unfavorable interactions (e.g. desolvation cost of charged residues and entropy loss of proteins upon complex formation) whose net effect determines the formation of the complex and its stability.[1-3] In contrast to vacuum, many of these interactions are either due to the presence of water or are affected by it, which makes the binding process in solution substantially different and much more complex. Consequently, proper treatment of water is crucial when simulating protein binding.

Computational simulations, in particular atomistic molecular dynamics (MD) simulations, can offer very detailed views of protein-protein interactions. They allow direct observation of the binding process and provide useful information of the binding energetics and the molecular basis for binding specificity. In addition, they can assist in designing improved

binding interfaces[4, 5] and help in exploring certain aspects of binding that may be difficult to characterize using experimental techniques, such as transient binding pockets.[6]

However, while MD simulations of proteins in explicit solvent can provide the most detailed views of the binding process, for many proteins it is not always feasible to run them long enough to cover the relevant timescales of motions involved in protein binding (Figure 1-1). One way to circumvent this limitation is to employ more computational power such as supercomputers and GPU clusters. Additionally, it might also be required to apply enhanced sampling techniques[7] or use approximate models for the solvent (e.g. implicit solvent models) and/or protein (e.g. "coarse graining"). In this dissertation, a number of these approaches have been used to study the energetics and dynamics of several protein-protein complexes.

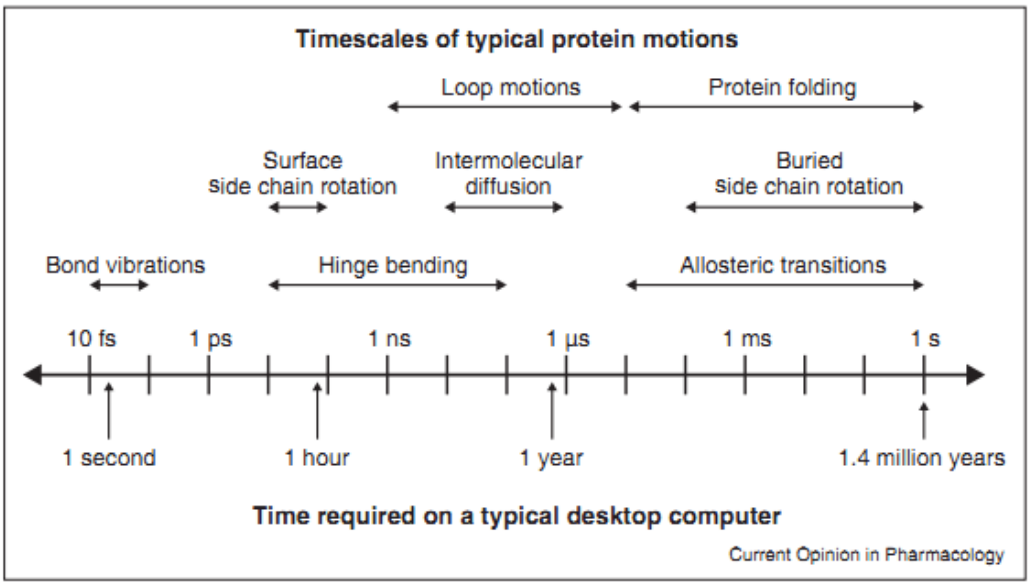

Figure 1-1. Timescales of typical protein motions. Figure was adapted from reference 7.

Salt bridges are among the electrostatic interactions involved in protein folding and binding and are formed when the side chains of two oppositely charged residues are within

hydrogen-bonding distance. Therefore, they are typically composed of positive charges from Lys, Arg, His and N-terminal amino group, and negative charges from Asp, Glu and C-terminal carboxyl group. Salt bridges can form within proteins[8] or at the binding interfaces,[9] be single or networked[10] and buried[11] or solvent-exposed[12]. While the interactions between the opposite charges are attractive, perhaps surprisingly, salt bridges are believed to make little (or even no) favorable contribution to the protein folding or binding due to their significant desolvation penalty.[13] This is mainly based on theoretical studies employing implicit solvent calculations using the Poisson-Boltzmann (PB) equation,[14-17] as it is very difficult to directly measure the stability gain of salt bridges using experimental techniques.

The Poisson-Boltzmann (PB) equation. The PB equation is a second order, nonlinear partial differential equation that describes the interaction between molecules in solutions. It is an implicit solvation model and involves two high dielectric and low dielectric media where fixed point charges are embedded in the low dielectric medium. The Poisson's equation can be derived from Gauss's law which states that the electric flux through any closed surface is proportional to the enclosed electric charge (in the cgs units where $\frac{1}{4\pi\epsilon_0} = 1$):

$$\oiint \mathbf{E}.\mathbf{n}\, dS = 4\pi \sum_i q_i \qquad (1)$$

where **E** is the electric field, **n** is the unit vector normal to the surface S and q represents a discrete set of enclosed charges. For a continuous charge density inside a surface, $\rho(\mathbf{r})$, we can write:

$$\oiint \mathbf{E}.\mathbf{n}\, dA = 4\pi \iiint \rho(\mathbf{r})\, d^3r \qquad (2)$$

The divergence theorem, which is the Stokes' theorem in 3D space, states that the outward flux of a vector field through a closed surface is equal to the volume integral of the divergence of the region inside the surface. So we can write the left side of Equation 2 as:

$$\oiint \mathbf{E}.\mathbf{n}\, dA = \iiint (\nabla.\mathbf{E})\, d^3r \qquad (3)$$

And then:

$$\iiint (\nabla.\mathbf{E})\, d^3r = 4\pi \iiint \rho(\mathbf{r})\, d^3r \qquad (4)$$

$$\iiint (\nabla.\mathbf{E} - 4\pi\rho(\mathbf{r}))\, d^3r = 0 \qquad (5)$$

$$\nabla.\mathbf{E} = 4\pi\rho(\mathbf{r}) \qquad (6)$$

Equation 6 is the differential form of Gauss's law of electrostatics. This equation relates the charge density to the electric field (**E**) in vacuum. When the charge density is located in a dielectric medium with the dielectric constant of ϵ, Equation 6 is written in terms of the electric displacement field **D** defined as $\mathbf{D} = \epsilon\mathbf{E}$:

$$\nabla.\mathbf{D} = 4\pi\rho(\mathbf{r}) \qquad (7)$$

Therefore:

$$\nabla.\mathbf{E} = \frac{4\pi\rho(\mathbf{r})}{\epsilon} \qquad (8)$$

Since it is often simpler to deal with the scalar functions rather than vector fields, we can use the following relation between electric field and electric potential (Φ):

$$\mathbf{E} = -\nabla\Phi(\mathbf{r}) \qquad (9)$$

to combine Equations 8 and 9 and write:

$$\nabla.(-\nabla\Phi(\mathbf{r})) = \frac{4\pi\rho(\mathbf{r})}{\epsilon} \qquad (10)$$

$$\nabla^2\Phi(\mathbf{r}) = -\frac{4\pi\rho(\mathbf{r})}{\epsilon} \qquad (11)$$

Equation 11 is a form of Poisson's equation in mathematics (i.e. $\nabla^2\phi = f$ in Euclidian space) and hence is called the Poisson's equation of electrostatics. This equation is a second order linear differential equation which can be solved to calculate the electric potential around a charged density in homogeneous dielectric medium. For inhomogeneous dielectric media, it is easy to rewrite Equation 11 using a dielectric constant that is a function of position:

$$-\nabla.\epsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) = 4\pi\rho(\mathbf{r}) \qquad (12)$$

To include the effect of mobile salt ions, we can use the Boltzmann distribution law to find the concentration of mobile ions around a charged object as a function of distance, n(r):

$$n(r) = n_0 e^{-\frac{ze\Phi}{kT}} \qquad (13)$$

Where n_0 is the concentration of ions in the bulk solution, z is the valence of the ion, e is the elementary charge and k is the Boltzmann constant. The corresponding charge density of this distribution for all of the mobile ions then can be written as:

$$\rho(r) = \sum_j e\, z_j\, n_{0,j} e^{-\frac{ze\Phi}{kT}} \qquad (14)$$

Now we can combine Equations 11 and 13 and write:

$$-\nabla.\epsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) = -4\pi(\rho_{molecule} + \rho_{ions}) \quad (15)$$

$$-\nabla.\epsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) = -4\pi\left(\sum_i q_i\delta(x-x_i) + \sum_j e\ z_j\ n_{0,j}e^{-\frac{ze\Phi - V_j(r)}{kT}}\right) \quad (16)$$

Where V_j is the steric interaction between the biomolecule and ions of species *j* which prevents overlap between the biomolecular and mobile counterion charge distributions. Equation 16 is called the Poisson-Boltzmann equation. The PB equation is a second order, nonlinear partial differential equation and is typically solved using numerical methods. In the absence of salt, the PB equation reduces to the Poisson's equation.

For many systems where the mean-field linear dielectric approximations implicit in the PB equation are valid, the nonlinear PB equation can be reduced to a linear equation. In this case the exponential can be truncated at first order in the Taylor series. By also assuming all steric factors are the same V_j=V, this linearization yields:

$$-\nabla.\epsilon(x)\nabla\varphi(x) = 4\pi\sum_{i=1}^{M} Q_i\,\delta(x-x_i) + \frac{8\pi e^2 I}{kT}e^{\frac{-V(x)}{kT}}\varphi(x) \quad (17)$$

where *I* is the ionic strength. Equation 17 is called the linearized PB equation or the Debye-Huckel equation. The linearized approximation is appropriate only when the electrostatic potential is small. This condition can be fulfilled when the overall charge density is low and/or when the salt concentration is high (due to the strong screening effects of salt).

Studies performed in this dissertation. As previously mentioned, according to the implicit solvent calculations using the PB equation salt bridges are believed to minimally contribute to the stability of protein binding due to their significant desolvation penalty. While the PB-based fast solvation model is very popular in estimating the electrostatic energies of

proteins, it lacks important molecular details such as hydrogen bonding and bridging water molecules. In **Chapters 2** to **4** of this dissertation, the PB-implicit solvent model is directly compared to explicit solvent models in calculating the desolvation penalties of 14 salt bridges across protein-protein interfaces. **Chapter 2** focuses on how the two models perform in general and investigates the overall agreement between the two models and potential explanations for the observed discrepancies.

Interestingly, while salt bridges are thought to provide little stability gain for proteins at room temperature, they appear to be critical in the adaptation of proteins for stability at extremely high temperatures. Based on the PB model it has been proposed that this stability gain is due to the reduced desolvation penalty of salt bridges at high temperatures. **Chapter 3** studies how the desolvation penalties of salt bridges change upon increasing temperature from 25^{o}C to 100^{o}C and how the two implicit and explicit solvent models compare in capturing these changes. **Chapter 4** extends the work in the previous chapters and examines how the two implicit and explicit solvent models estimate the change in the desolvation penalties of salt bridges when various concentrations of a monovalent salt are used.

Chapter 5 studies the rearrangements in the hydrogen bond register of an intermolecular protein-peptide β-sheet. By employing a combination of distributed computing and supercomputer resources, it provides the first direct observation of such rearrangements using atomically detailed simulations. The simulations show that the rearrangements involve a hydrophobic residue of the peptide anchoring to a transient pocket on the protein and helping the peptide to "crawl" to the native state.

Chapter 6 explores the relative stability of two GCN4 leucine zipper oligomers, the known dimer fold vs. the newly-discovered trimer fold, in isolation by employing an enhanced

sampling technique (parallel tempering molecular dynamics simulations) on the microsecond timescale. The simulations allow estimating the melting temperatures of the two oligomers and provide insights to their relative stabilities at different temperatures.

Chapter 7 summarizes the overall findings of the previous chapters and discusses future directions.

Finally, **Appendix A** provides a brief introduction to the Multistate Bennett Acceptance Ratio (MBAR) method that is used for extracting thermodynamic properties from simulations performed at multiple equilibrium states.

1.1 REFERENCES

(1) Jones, S., and Thornton, J. M. Principles of protein-protein interactions, *Proc Natl Acad Sci.* **1996**, 93, 13-20.

(2) Veselovsky, A. V., Ivanov, Y. D., Ivanov, A. S., Archakov, A. I., Lewi, P., and Janssen, P. Protein-protein interactions: mechanisms and modification by drugs, *Journal of molecular recognition.* **2002**, 15, 405-422.

(3) Prabhu, N., and Sharp, K. Protein-solvent interactions, *Chemical reviews*. 2006, 106, 1616-1623.

(4) Kuroda, D., Shirai, H., Jacobson, M. P., and Nakamura, H. Computer-aided antibody design, *Protein engineering, design & selection*. **2012**, 25, 507-522.

(5) Levin, A. M., Bates, D. L., Ring, A. M., Krieg, C., Lin, J. T., Su, L., Moraga, I., Raeber, M. E., Bowman, G. R., Novick, P., Pande, V. S., Fathman, C. G., Boyman, O., and Garcia, K. C. Exploiting a natural conformational switch to engineer an interleukin-2 'superkine', *Nature*, **2012**, 484, 529-533.

(6) Wells, J. A., and McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces, *Nature* **2007**, 450, 1001-1009.

(7) Zwier, M. C., and Chong, L. T. Reaching biological timescales with all-atom molecular dynamics simulations, *Curr. Opin. Pharmacol.* **2012**, 10, 745-752.

(8) Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. PH-Induced Denaturation Of Proteins - A Single Salt Bridge Contributes 3-5 Kcal Mol To The Free-Energy Of Folding Of T4-Lysozyme, *Biochemistry* **1990**, 29, 2403-2408.

(9) Buckle, A. M., Schreiber, G., and Fersht, A. R. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution, *Biochemistry*, **1994**, 33, 8878-8889.

(10) Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. Strength and co-operativity of contributions of surface salt bridges to protein stability, *Journal of Molecular Biology* **1990**, 216, 1031-1044.

(11) Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. Are Buried Salt Bridges Important For Protein Stability And Conformational Specificity, *Nature Structural Biology* **1995**, 2, 122-128.

(12) Sun, D. P., Sauer, U., Nicholson, H., and Matthews, B. W. Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis, *Biochemistry* **1991**, 30, 7142-7153.

(13) Hendsch, Z. S., and Tidor, B. Do Salt Bridges Stabilize Proteins - A Continuum Electrostatic Analysis, *Protein Science* **1994**, 3, 211-226.

(14) Jackson, J. K. Classical Electrodynamics. John Wiley & Sons. **1962**.

(15) Dill K. A. and Bromberg S. Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology. Garland Science. **2002**.

(16) Sharp, K.A. and B. Honig, Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem*. **1990**. 94(19): p. 7684-7692.

(17) Baker, N.B.; Bashford, D.; Case, D. A., Implicit Solvent Electrostatics in Biomoleular Simulation. *New Algorithms for Macromolecular Simulation*, **2005**: p. 265-295.

2.0 DESOLVATION COSTS OF SALT BRIDGES ACROSS PROTEIN BINDING INTERFACES: SIMILARITIES AND DIFFERENCES BETWEEN IMPLICIT AND EXPLICIT SOLVENT MODELS

This work was published as: R. Salari and L. T. Chong (2010). "Desolvation Costs of Salt Bridges across Protein Binding Interfaces: Similarities and Differences between Implicit and Explicit Solvent Models." *J. Phys. Chem. Lett.* 1(19): 2844-2848.

2.1 INTRODUCTION

Protein binding interactions often involve salt bridges, that is, pairs of oppositely charged residues that are within hydrogen-bonding distance. On the basis of theoretical studies, salt bridges are thought to make surprisingly little (or even no) favorable contribution to protein folding or binding due to the significant cost of desolvating the two charged salt-bridge partners.[1-5] For efficient computations, these previous studies all used a dielectric continuum solvent model based on the Poisson-Boltzmann (PB) equation. This model, which is the "gold standard" of implicit solvent models, has been successfully parameterized to reproduce solvation free energies of small molecules determined by either experiment[6] or simulations[7,8] with explicit

water molecules. However, the PB model lacks important molecular details of the first solvation shell and a description of nonpolar contributions to solvation.[9] Valuable insights about modeling solvation can therefore be obtained by comparing explicit and implicit solvent calculations.[9-13]

Here, for the first time, we directly compare the PB implicit solvent model with several explicit water models in computing the desolvation penalties of salt bridges across protein-protein binding interfaces. We performed both implicit and explicit solvent calculations on all 14 salt bridges across the binding interfaces of four protein-protein complexes (Figure 2-5, Supporting Information), that were identified by a previous study as having a wide range of desolvation penalties.[5] We computed the desolvation penalty for each salt bridge upon binding relative to its hydrophobic 'isostere', that is, a hypothetical mutant version that has all partial charges on the salt bridge side chains set to zero; this desolvation penalty is reported as $\Delta\Delta G_{solv}$. In the explicit solvent calculations, this desolvation penalty was computed using thermodynamic integration techniques (see Methods). As done in previous theoretical studies, we focused on "rigid" binding, with the unbound conformations of the proteins being identical to the corresponding bound conformations. To circumvent convergence problems associated with net-charged systems in explicit solvent calculations,[14] we represented the unbound state – in both the implicit and explicit solvent calculations – with proteins separated by 30 Å (between their centers of mass) and simultaneously turned off the charges of the oppositely charged side chains of the salt bridge; this was done in both the unbound and bound states of the proteins.

In order to directly compare the solvation thermodynamics of the implicit and explicit water models, it was essential to keep the proteins *completely* rigid, even in the explicit solvent molecular dynamics (MD) simulations. A direct comparison also required that we fix all other parameters common to the two approaches to ensure that they remained absolutely identical i.e.

protein coordinates, atomic charges and radii (OPLS-AA/L force field),[15] box volume, and temperature. MD simulations were performed with periodic boundary conditions and a PME treatment of long-range electrostatics.[16] Periodic boundary conditions were also employed in the PB calculations, implicitly including long-range electrostatic interactions with all periodic images. Implicit and explicit solvent calculations were performed using the DelPhi[17] and GROMACS[18] software packages, respectively. Three different water models were explored in the explicit solvent calculations: TIP3P,[19] TIP4P,[19] and SPC/E.[20] To represent the boundary between the low-dielectric protein region and high-dielectric solvent region in the implicit solvent calculations, we focused primarily on the molecular surface of the protein,[21] which is the standard representation; calculations were also performed using the van der Waals surface, which has been proposed as an alternative[22,23] but led to comparable results (see below).

2.2 METHODS

To directly compare the solvation thermodynamics of the implicit and explicit solvent approaches, we kept the proteins rigid and fixed all parameters common to the approaches to be identical i.e. protein coordinates, atomic charges and radii (OPLS/AA-L force field),[15] box volume, and temperature. To enable a consistent treatment of long-range electrostatics, periodic boundary conditions were employed in both approaches, enabling the use of the PME method[16] for the explicit solvent calculations; all systems were electrically neutral. Details of the protein models are provided in Supporting Information. Desolvation penalties of salt bridges upon protein binding ($\Delta\Delta G_{solv}$) were computed according to the thermodynamic cycle shown in Figure 2-6 in Supporting Information.

2.2.1 Implicit Solvent Calculations

Implicit solvent calculations were performed using finite difference methods, as implemented in the DelPhi 4.0 software package,[17] to solve the linearized form of the PB equation (which reduces to the Poisson equation in the absence of salt, as in our calculations). In particular, electrostatic contributions to solvation free energies were computed for each wild-type salt bridge and its hydrophobic isostere in the unbound and bound protein states; these contributions were determined by first directly calculating the induced polarization charges and then calculating the interaction between the protein charges and the reaction field due to the polarization charges.[21] The electrostatic contribution to the solvation free energy of the salt bridge relative to its hydrophobic isostere in the unbound or bound state yields the solvation free energies $\Delta G_{solv}^{unbound}$ or ΔG_{solv}^{bound}, respectively. The desolvation penalty of each salt bridge upon protein binding relative to its hydrophobic isostere was computed using $\Delta\Delta G_{solv} = \Delta G_{solv}^{unbound} - \Delta G_{solv}^{bound}$ (Figure 2-6, Supporting Information).

Calculations of each state of the system were performed 14 separate times with systematic molecular translations on the grid at 25 ^{o}C. Results reported are averages of 14 calculations, with uncertainties represented by the standard deviation. Each calculation was carried out for 10,000 steps to satisfy a convergence criterion of 0.001 kT/e in the potential. To avoid errors in the dielectric boundary, the OPLS/AA-L radii of polar hydrogen atoms were converted from 0 to the default value of 1.0 Å. To represent the dielectric boundary, we tested both the molecular (default)[21] and van der Waals surfaces of the protein. Consistent with keeping the proteins rigid, a dielectric constant of 1 was used for the protein region; to represent the dielectric properties of water at 25 ^{o}C, a dielectric constant of 78.4 was used for the solvent

region. A grid resolution of 0.33 Å/(grid units) was used for all protein systems except for the neuraminidase-antibody complex, which was limited to a slightly lower resolution of 0.37 Å/(grid units) due to its large size. Grid dimensions for the barnase-barstar, growth hormone-receptor, neuraminidase-antibody and RafRBD-Rap1A complexes were 343 x 343 x 343, 403 x 403 x 403, 479 x 479 x 479, and 361 x 361 x 361, respectively. Each calculation required 1.5 to 4 CPU hours on a single core of a dual-core 2.6 GHz Opteron node.

2.2.2 Explicit Solvent Calculations

Explicit solvent calculations were performed using the thermodynamic integration approach[27] with explicit solvent MD simulations, as implemented in the GROMACS 4.0.4 software package.[18] In particular, we computed first computed differences in the overall free energy of each salt bridge relative to its hydrophobic isostere in its unbound and bound states $\Delta G^{(un)bound}$, which is the sum of contributions from both nonbonded protein-protein and protein-solvent interactions, $\Delta G^{(un)bound}_{protein}$ and $\Delta G^{(un)bound}_{solv}$, respectively. Next, to obtain differences in solely the solvation free energies, all nonbonded protein-protein interactions were subtracted from differences in the overall free energies. Finally, the desolvation penalty of each salt bridge upon protein binding relative to its hydrophobic isostere was computed using $\Delta\Delta G_{solv} = \Delta G^{unbound}_{solv} - \Delta G^{bound}_{solv}$.

Differences in the overall free energies of each salt bridge relative to its hydrophobic isostere in its unbound and bound states were computed using the following:

$$\Delta G^{(un)bound} = \int_0^1 d\lambda \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda$$

where $H(\lambda)$ is the system Hamiltonian as a function of the coupling parameter λ and the brackets represent ensemble averaging at a given λ value; the λ values of 0 and 1 represent the wild-type and hydrophobic-isostere versions of the salt bridge, respectively. Separate MD simulations of the proteins (unbound and bound states) were performed at each of the following eight λ values, linearly discharging the side chains of the salt bridge: 0, 0.15, 0.3, 0.45, 0.6, 0.75, 0.9, and 1. The trapezoidal method was then used to numerically solve the thermodynamic integral to obtain $\Delta G^{(un)bound}$. Uncertainties in the free energies are derived from sampling errors in $\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda}$; errors at each λ value were estimated using block averaging,[28] as implemented in the g_analyze utility of GROMACS.[18]

MD simulations were performed with explicit solvent (TIP3P,[19] TIP4P,[19] or SPC/E)[20] in the NVT ensemble, with the number of atoms in the unbound and bound states of each system enforced to be exactly the same (see Supporting Information). Proteins were kept rigid throughout the simulations using the GROMACS "frozen" option, setting velocities of all protein atoms to zero. Real-space electrostatic interactions were truncated at 10 Å while the long-range components of these interactions were calculated using the PME method[16] with periodic boundary conditions, a spline order of 6, Fourier spacing of 1.0 Å, and relative tolerance of 10^{-6} between long- and short-range energies. Van der Waals interactions were switched off smoothly between 8 and 9 Å. Each λ simulation was performed for 1 ns at constant temperature (25 ^{o}C) and volume. Prior to each λ simulation, the solvent was equilibrated in two stages: 1) 10 ps at constant temperature (25 ^{o}C) and volume, and 2) 100 ps at constant temperature (25 ^{o}C) and pressure (1 atm). The Langevin thermostat (frictional constant of 1 ps^{-1}) and a weak Berendsen barostat[29] (coupling time constant of 5 ps) were used to maintain constant temperature and

pressure, respectively. A 2-fs time step was used for all simulations. Each λ simulation required 1 to 6 CPU days on a dual quad-core 2.66 GHz Xeon node.

2.3 RESULTS

As shown in Figure 2-1A, the desolvation penalties estimated by implicit solvent calculations are strongly correlated with those from explicit solvent calculations with the TIP3P water model (R^2 = 0.996). An equally strong correlation results when the TIP4P and SPC/E water models are used (R^2 of 0.993 and 0.992, respectively; Figure 2-7, Supporting Information). The overall agreement between the results from implicit and explicit water models is surprisingly good, given the dramatic differences in their representations of solvent and given the large range of the desolvation penalties (~10 to ~210 kcal/mol). These results provide important reinforcement, therefore, of the widely appreciated utility of Poisson-based calculations for modeling solvation effects in charged, biomolecular systems.

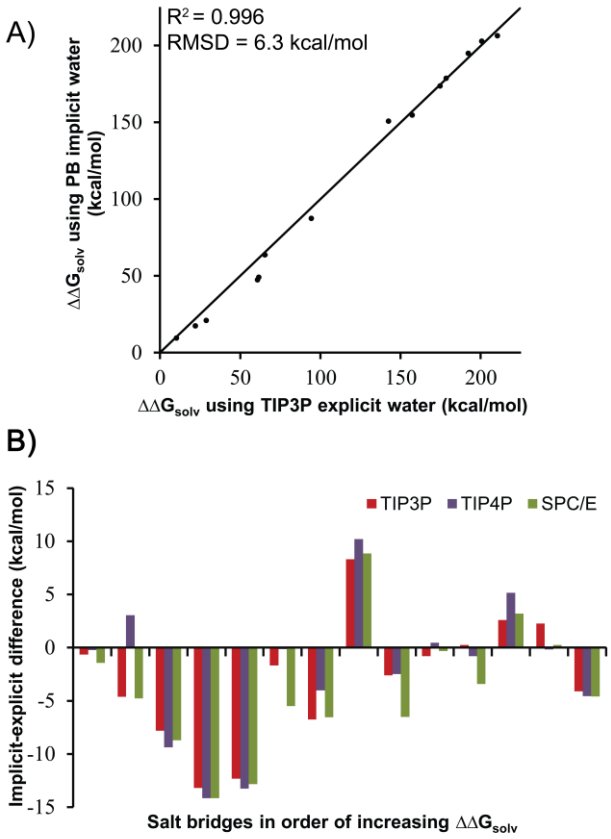

Figure 2-1. (A) Comparison of implicit and explicit solvent models for computing desolvation penalties of salt bridges upon protein binding ($\Delta\Delta G_{solv}$). The dielectric boundary in the implicit solvent calculations was represented by the molecular surface. The diagonal line represents perfect agreement. Error bars are included, but difficult to see since they are small (< 2 kcal/mol). (B) Implicit-explicit differences for each salt bridge.

That said, a closer examination of the results reveals significant discrepancies between the implicit and explicit solvent predictions: the absolute rms deviations between the predictions for all salt bridges are 6.3, 6.8, and 7.1 kcal/mol for the TIP3P, TIP4P, and SPC/E water models, respectively, which correspond to relative rms deviations of 5.5, 6.0, and 6.2 %, respectively (absolute rms deviation divided by the average $\Delta\Delta G_{solv}$ of the explicit water model).

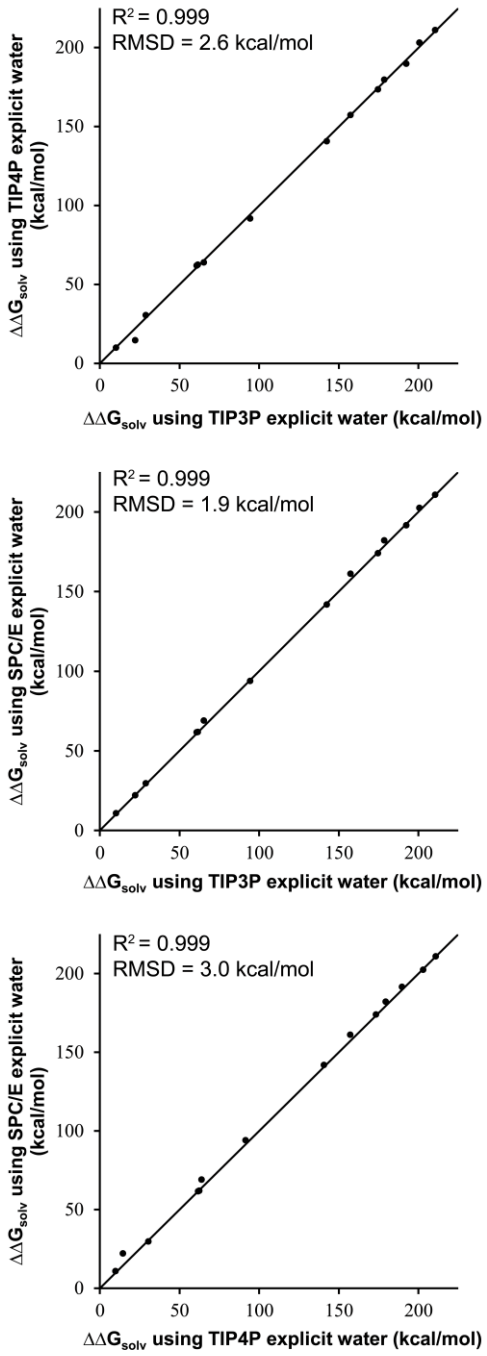

Figure 2-2. Comparison of explicit solvent models for computing desolvation penalties of salt bridges upon protein binding ($\Delta\Delta G_{solv}$). Diagonal lines represent perfect agreement. Error bars are included, but difficult to see since they are small.

Notably, the implicit-explicit discrepancies for individual salt bridges are largely independent of the explicit water model (Figure 2-1B). Results among the three explicit solvent

models are comparable, with rms deviations of 2.6, 1.9, and 3.0 kcal/mol for TIP4P vs. TIP3P, SPC/E vs. TIP3P, and SPC/E vs. TIP4P, respectively (Figure 2-2). These findings not only provide further confidence in the explicit solvent calculations, they also strongly suggest that the discrepancies reflect key differences between implicit and explicit models of solvation.

To investigate the source of the discrepancies between the implicit and explicit solvent results, we also performed calculations on the same 14 salt bridges in the absence of the protein environment, that is, with the same geometries, but in solution and with the residues capped with acetyl and N-methyl groups at the N- and C-termini, respectively. In this second set of explicit solvent calculations, only the TIP3P water model was used. The rms deviation between the implicit and explicit solvent results is significantly reduced from 6.3 kcal/mol to 1.8 kcal/mol when the protein environment is replaced by solvent (Figure 2-3). It appears, therefore, that the protein environment – and the solvent's response to it – is the primary source of the deviations observed between implicit and explicit solvent calculations.

Representation of the protein environment in the implicit solvent calculations is influenced not only by the protein dielectric constant, but also the dielectric boundary between the protein and solvent regions. In addition to using the molecular surface of the protein to represent the dielectric boundary, which is traced out by a spherical "water" probe with a radius of 1.4 Å, we also tested the van der Waals surface. However, the resulting (implicit) desolvation penalties were found to be comparable to those associated with the molecular surface, with rms deviations of 5.9 kcal/mol from the TIP3P explicit solvent calculations, for example (Figure 2-7, Supporting Information). Interestingly, although the molecular surface with the current set of atomic radii underestimates the solvation free energies of the salt bridge relative to their hydrophobic isosteres ($\Delta G_{solv}^{(un)bound}$) in the unbound and bound states (Figure 2-8, Supporting

Information), the difference between the two states ($\Delta\Delta G_{solv}$) are underestimated for some of the salt bridges and overestimated for others when compared to explicit solvent calculations (Figure 2-1A).

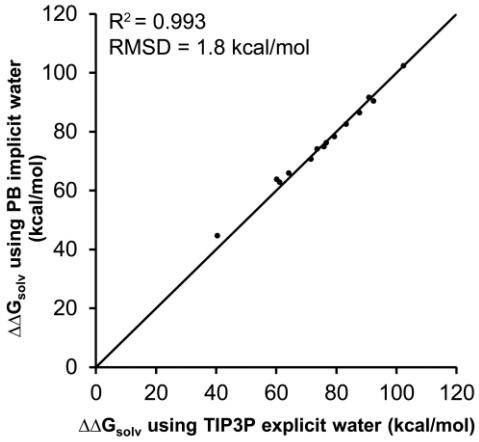

Figure 2-3. Comparison of implicit and explicit solvent models for computing desolvation penalties of salt bridges upon association in the absence of the protein environment ($\Delta\Delta G_{solv}$). The dielectric boundary in the implicit solvent calculations was represented by the molecular surface. The diagonal lines represent perfect agreement. Error bars are included, but difficult to see since they are small.

To determine why certain salt bridges have larger implicit-explicit differences than others, we plotted these differences vs. (a) $\Delta\Delta G_{solv}$ and (b) the percent burial upon binding; both van-der-Waals and molecular-surface implicit solvent results were considered. Only the plot involving the van der Waals surface and percent burial resulted in any correlation (R^2 of 0.320; Figure 2-9, Supporting Information). We also looked for a correlation between implicit-explicit differences and involvement of the salt bridge in a "network" where at least one of the charged partners forms another salt bridge[2,24] however, no correlation was found (Table 2-2, Supporting Information). We did find a significant correlation between the magnitude of implicit-explicit

differences in the solvation free energy of the salt bridge in its bound state relative to its hydrophobic isostere, ΔG_{solv}^{bound}, and the probability of observing "bridging" water molecules in the explicit solvent simulations when the molecular surface was used (Figure 2-4).; no correlation was found when the van der Waals surface was used (Figure 2-10, Supporting Information).

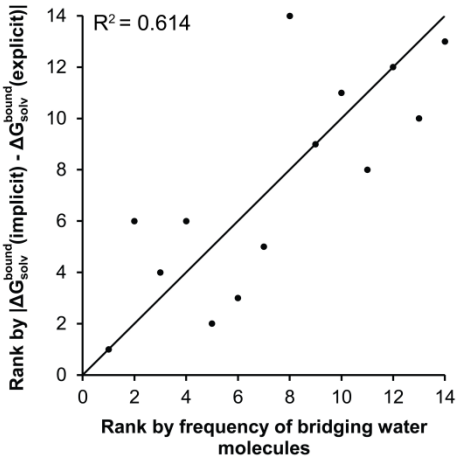

Figure 2-4. Correlation of the magnitude of implicit-explicit differences in ΔG_{solv}^{bound} vs. probability of observing bridging water molecules during simulations in TIP3P explicit water when the dielectric boundary in the implicit solvent calculations is represented by the molecular surface. Probabilities were computed from conformations sampled every ps during the 1-ns simulations. Water molecules were defined as "bridging" if they form hydrogen bonds with both salt bridge partners in their bound state. A hydrogen bond was defined as having a hydrogen-acceptor length of 2.5 Å and a donor-hydrogen-acceptor angle of more than 90°. Diagonal lines represent perfect agreement.

We tried reducing implicit-explicit differences by using a lower solvent dielectric constant for the implicit solvent calculations that is more representative of the explicit solvent models (i.e., 52, the dielectric constant of TIP4P, since this value is the lowest among the explicit solvent models tested),[25] but this lower value had no effect (Table 2-3, Supporting Information). Finally, we considered reducing implicit-explicit differences by either increasing the protein dielectric

constant or scaling the atomic radii in the implicit solvent calculations. However, these approaches would underestimate the desolvation penalties for some salt bridges and overestimate those for others, even in the absence of their protein environments (as is evident in Figure 2-3).

2.4 CONCLUSION

In closing, we have performed the first direct comparison of implicit and explicit solvent models for use in evaluating free energy contributions of salt bridges to protein-protein binding. We have demonstrated that the desolvation penalties of salt bridges upon protein binding are of similar magnitudes when estimated using implicit and explicit solvent models. Nonetheless, significant discrepancies exist for particular salt bridges. Given that bridging water molecules have been shown to be a source of discrepancies in other studies,[10,11] hybrid implicit/explicit solvent models might be an attractive alternative approach.[26] Since the set of salt bridges[5] studied here highlights challenges to be faced in the application of implicit solvent methods, it might also provide valuable test cases for the development of improved fast solvation models.

2.5 ACKNOWLEDGEMENT

This work was supported by NSF CAREER Award MCB-0845216 to LTC, and the University of Pittsburgh's A&S fellowship to RS. We thank the University of Pittsburgh's CMMS for use of its Linux cluster.

2.6 SUPPORTING INFORMATION

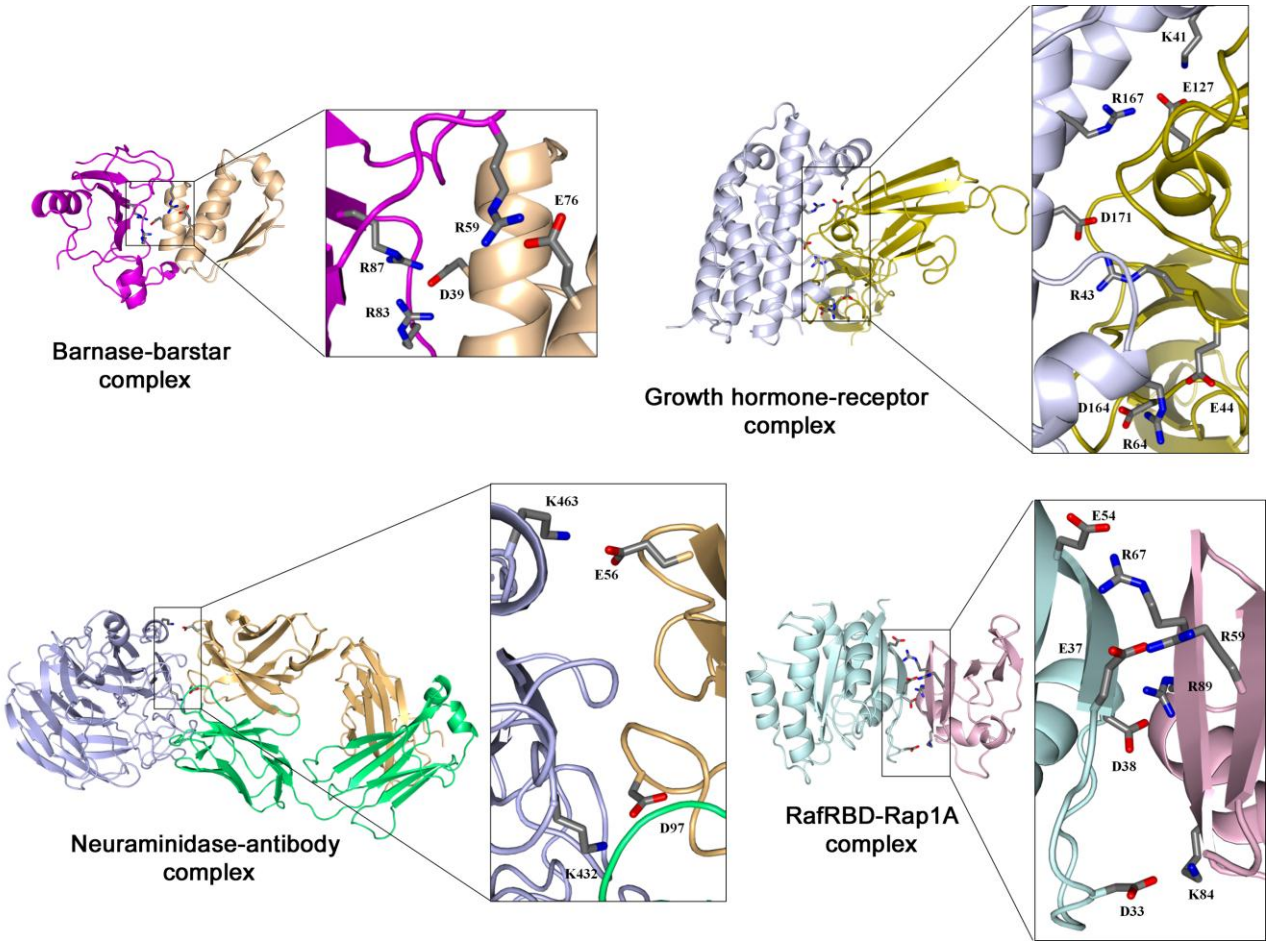

Figure 2-5. Locations of salt bridges across the binding interfaces of each protein-protein complex in this study.

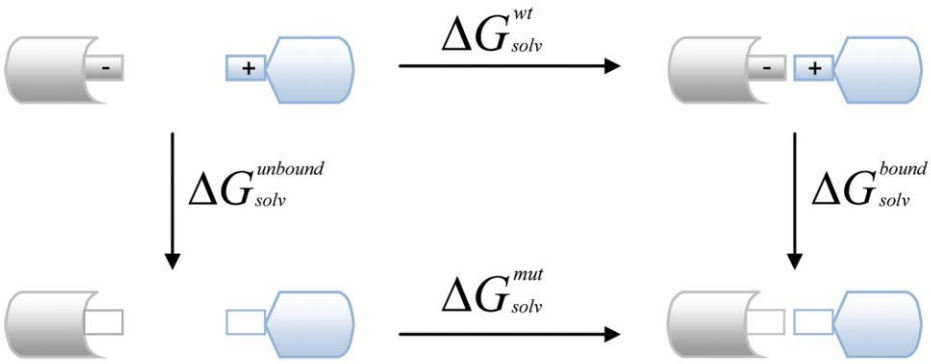

Figure 2-6. Thermodynamic cycle used for computing the desolvation penalty of a salt bridge upon protein binding relative to its hydrophobic isostere ($\Delta\Delta G_{solv}$). The wild-type salt bridge and its "mutant" hydrophobic isostere are represented by the filled and empty rectangles, respectively. The $\Delta\Delta G_{solv}$ of each salt bridge was computed using the vertical arrows of the thermodynamic cycle in Equation (2), which is much more straighforward to evaluate than Equation (1) as it circumvents the need to simulate the diffusional association of the proteins:

$$\Delta\Delta G_{solv} = \Delta G_{solv}^{wt} - \Delta G_{solv}^{mut} \quad (1)$$

$$\Delta\Delta G_{solv} = \Delta G_{solv}^{unbound} - \Delta G_{solv}^{bound} \quad (2)$$

where and are the desolvation penalties of the salt bridge upon binding to form the wild-type and mutant complexes, respectively; and are the solvation free energies of the wild-type unbound and bound states, respectively, relative to the corresponding mutant states.

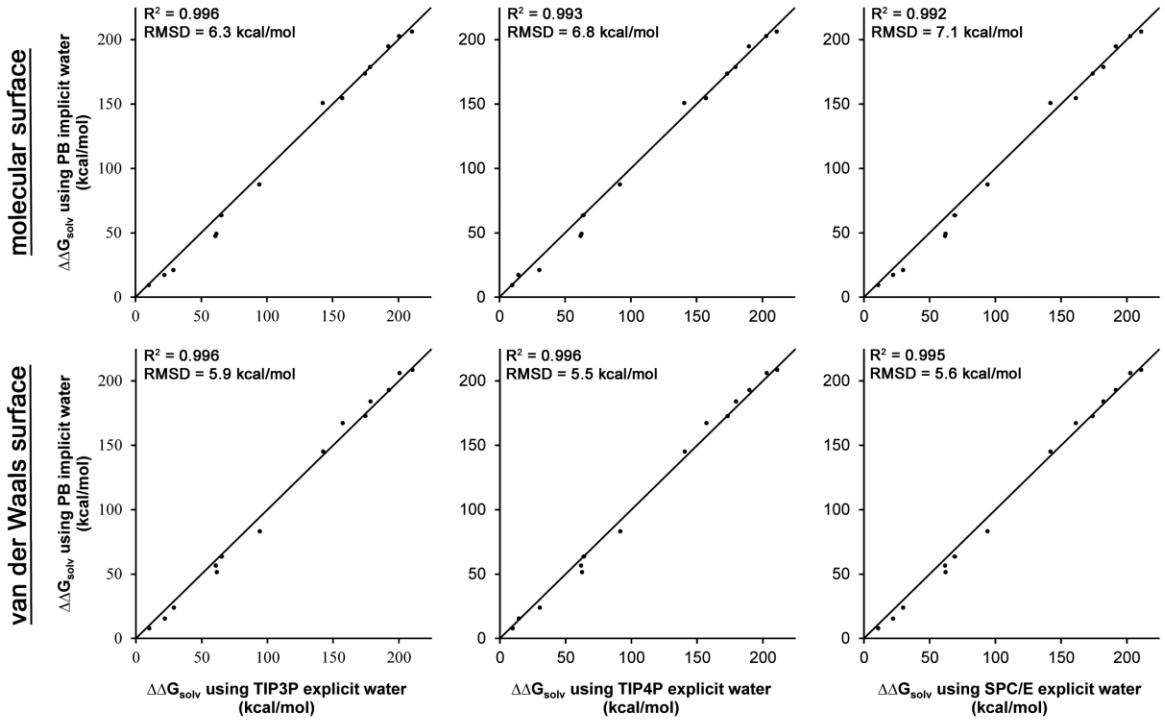

Figure 2-7. Comparison of implicit vs. explicit solvent models for computing desolvation penalties of salt bridges upon protein binding ($\Delta\Delta G_{solv}$). In the implicit solvent calculations, two different representations of the dielectric boundary were tested: the molecular surface of the protein (top row) and the van der Waals surface of the protein (bottom row). Diagonal lines represent perfect agreement. Error bars are included, but difficult to see since they are small.

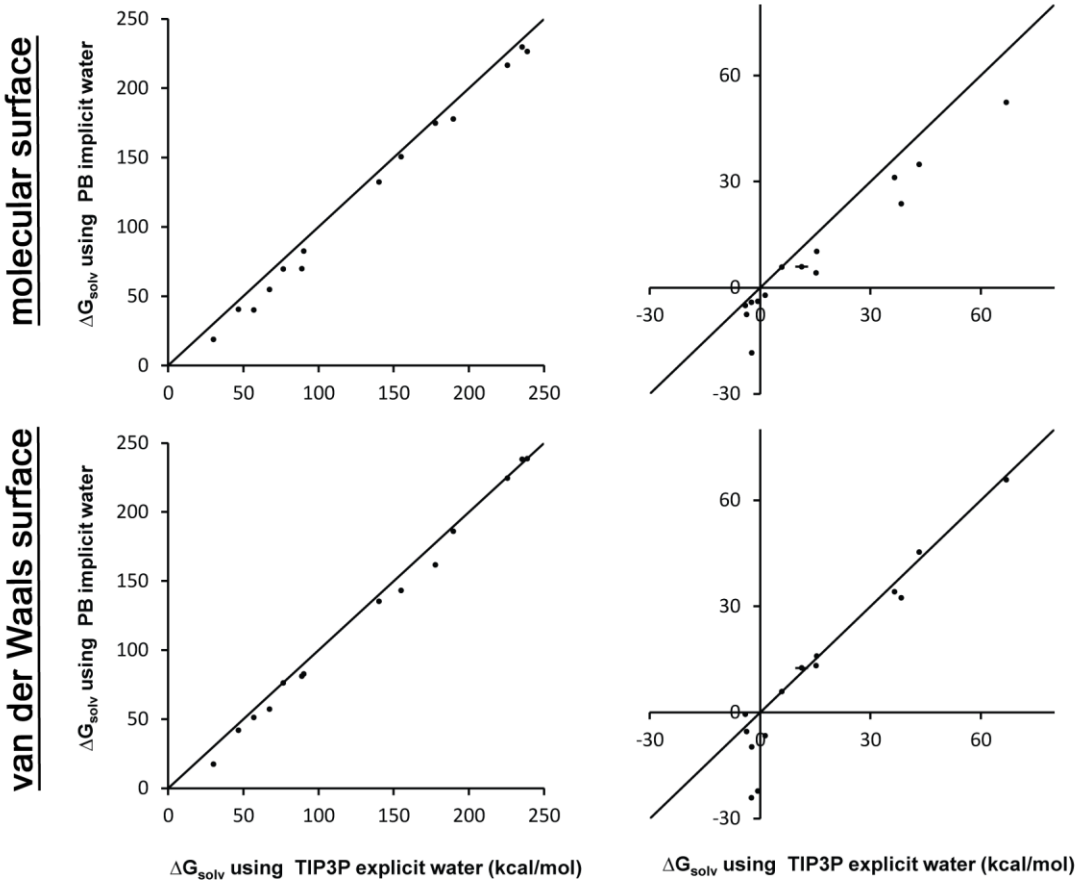

Figure 2-8. Comparison of implicit vs. explicit solvent models for computing the solvation free energies of salt bridges (ΔG_{solv}) in their unbound (left) and bound states (right). In the implicit solvent calculations, both the molecular (top row) and van der Waals surfaces (bottom row) were tested as representations of the dielectric boundary. The diagonal lines represent perfect agreement. Error bars are included, but difficult to see since they are very small.

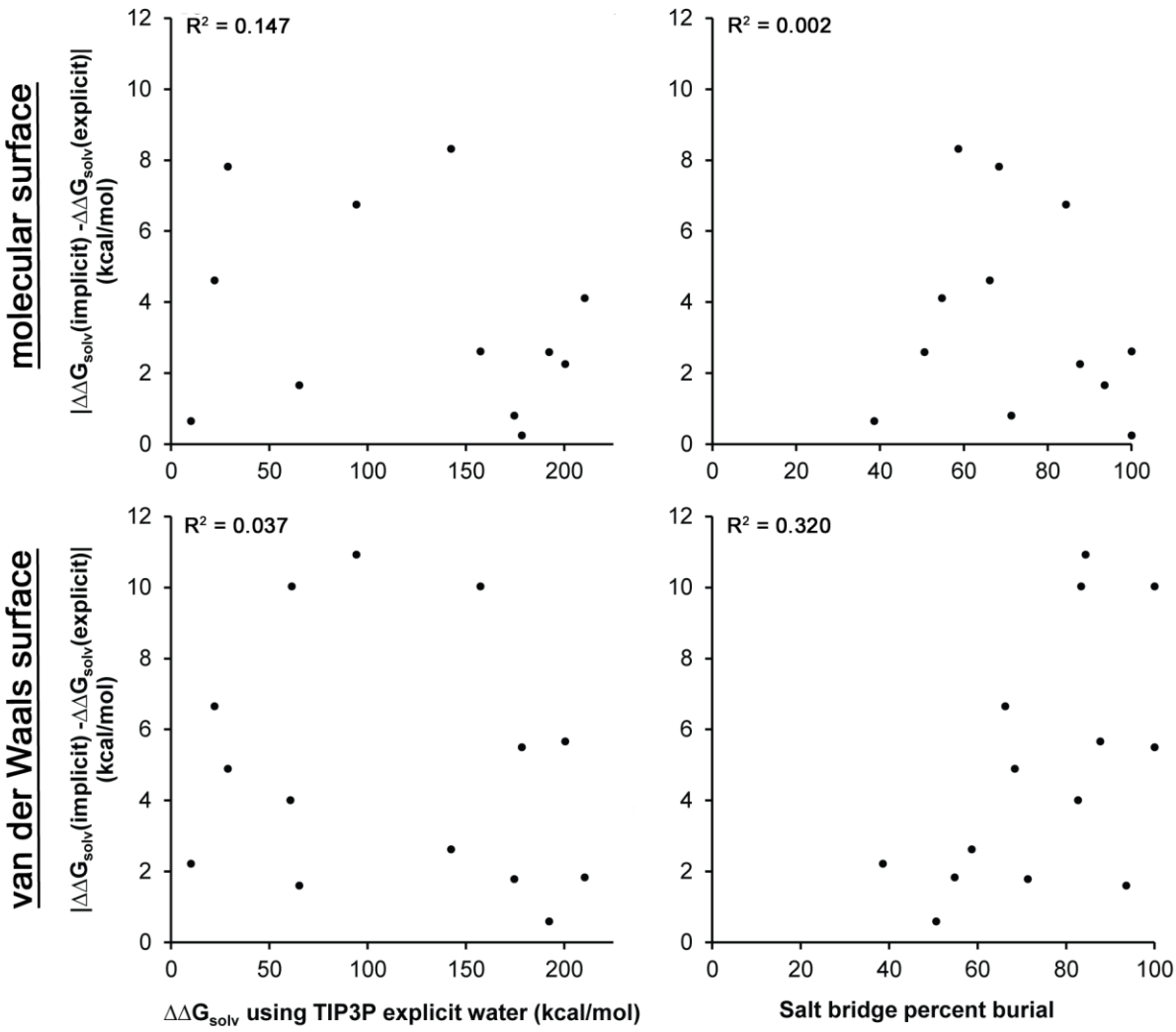

Figure 2-9. Correlation of the magnitudes of implicit-explicit differences in $\Delta\Delta G_{solv}$ vs. $\Delta\Delta G_{solv}$ (left) and percent burial of salt bridges (right). Implicit solvent calculations were performed separately using the molecular (top row) and van der Waals surfaces (bottom row). The TIP3P water model was used for the explicit solvent calculations.

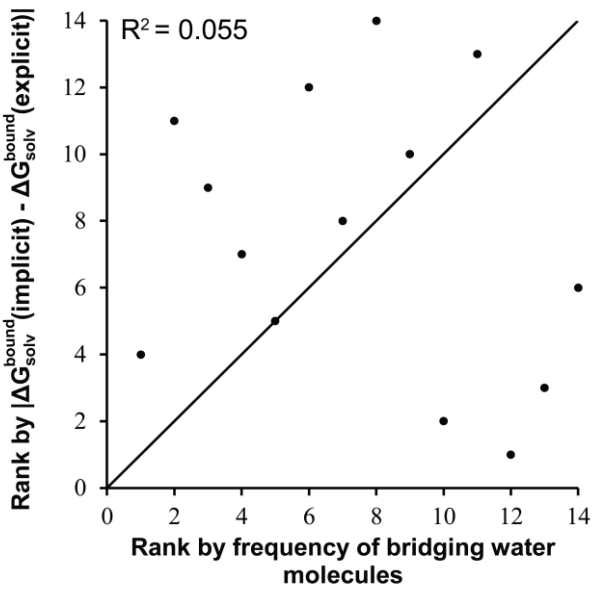

Figure 2-10. Correlation of the magnitude of implicit-explicit differences in ΔG_{solv}^{bound} vs. probability of observing bridging water molecules during simulations in TIP3P explicit water when the dielectric boundary in the implicit solvent calculations is represented by the van der Waals surface.

Table 2-1. Computed desolvation penalties of salt bridges upon protein binding ($\Delta\Delta G_{solv}$) using implicit and explicit solvent models. In the implicit solvent calculations, two different representations of the dielectric boundary were tested: the molecular surface and van der Waals surface of the protein.

		implicit solvent		*explicit solvent*		
complex	*salt Bridge*	*molecular surface*	*van der Waals surface*	*TIP3P*	*TIP4P*	*SPC/E*
Barnase-barstar	R59-E76	173.7 (0.2)	172.8 (0.2)	174.6 (0.2)	173.3 (0.9)	174.1 (0.3)
	R83-D39	178.7 (0.2)	184.0 (0.3)	178.5 (0.3)	179.5 (0.2)	182.1 (0.5)
	R87-D39	154.7 (0.1)	167.3 (0.3)	157.3 (0.3)	157.2 (0.2)	161.2 (0.3)
Growth hormone-receptor	K41-E127	49.2 (0.2)	51.5 (0.2)	61.5 (0.2)	62.4 (0.2)	62.0 (0.4)
	R64-E44	17.4 (0.2)	15.4 (0.3)	22.1 (0.2)	14.4 (0.3)	22.2 (0.4)
	R64-D164	21.0 (0.2)	24.0 (0.3)	28.9 (0.2)	30.4 (0.2)	29.7 (0.3)
	R167-E127	47.5 (0.2)	56.7 (0.3)	60.7 (0.3)	61.7 (0.2)	61.7 (0.3)
	D171-R43	63.6 (0.1)	63.7 (0.2)	65.3 (1.7)	63.8 (1.8)	69.1 (1.9)
Neuraminidase-antibody	K432-D97	87.5 (0.2)	83.3 (0.3)	94.2 (0.3)	91.5 (0.2)	94.0 (0.5)
	K463-E56	9.5 (0.2)	7.9 (0.2)	10.1 (0.2)	9.7 (0.2)	10.9 (0.3)
Raf1-Rap1A	D33-K84	194.9 (0.2)	192.9 (0.2)	192.3 (0.3)	189.7 (0.5)	191.7 (0.4)
	E37-R59	206.3 (0.3)	208.6 (0.2)	210.4 (0.2)	210.9 (0.2)	210.9 (0.3)
	D38-R89	202.8 (0.1)	206.2 (0.2)	200.5 (0.3)	203.0 (0.7)	202.5 (0.3)
	E54-R67	150.8 (0.2)	145.1 (0.2)	142.5 (0.2)	140.6 (0.3)	141.9 (0.4)

Table 2-2. Implicit-explicit differences in $\Delta\Delta G_{solv}$ for each salt bridge and whether or not the salt bridge is networked. Salt bridges are considered "networked" if at least one of the charged partners forms another salt bridge. In the implicit solvent calculations, both the molecular and van der Waals surfaces were tested as representations of the dielectric boundary. Implicit-explicit differences are presented in order of increasing magnitude according to implicit solvent results with the molecular surface. Explicit solvent calculations are associated with the TIP3P water model.

salt bridge	$\lvert\Delta\Delta G_{solv}$*(implicit)-* $\Delta\Delta G_{solv}$*(explicit)*$\rvert$ *(kcal/mol)*		*networked?*
	molecular surface	*van der Waals surface*	
R83-D39	0.2 (0.3)	5.5 (0.4)	X
K463-E56	0.7 (0.3)	2.2 (0.3)	
R59-E76	0.8 (0.2)	1.8 (0.2)	
D171-R43	1.7 (1.7)	1.6 (1.7)	
D38-R89	2.3 (0.3)	5.7 (0.3)	
D33-K84	2.6 (0.3)	0.6 (0.3)	
R87-D39	2.6 (0.3)	10.0 (0.4)	X
E37-R59	4.1 (0.3)	1.8 (0.2)	
R64-E44	4.6 (0.2)	6.7 (0.4)	X
K432-D97	6.7 (0.3)	10.9 (0.4)	
R64-D164	7.8 (0.2)	4.9 (0.3)	X
E54-R67	8.3 (0.3)	2.6 (0.3)	
K41-E127	12.3 (0.2)	10.0 (0.2)	X
R167-E127	13.2 (0.3)	4.0 (0.4)	X

Table 2-3. Effect of the solvent-region dielectric constant on RMS deviations between implicit and explicit predictions of $\Delta\Delta G_{solv}$ for the TIP3P, TIP4P, and SPC/E explicit water models.

implicit solvent		*RMS deviations from explicit solvent results (kcal/mol)*		
dielectric boundary	*dielectric constant*	*TIP3P*	*TIP4P*	*SPC/E*
molecular surface	*78.4*	6.3	6.8	7.1
	52.0	6.4	6.8	7.3
van der Waals surface	*78.4*	5.9	5.5	5.6
	52.0	5.5	4.7	5.4

2.6.1 Protein Models

All model building was performed using the GROMACS 4.0.4 software package.[18] Coordinates of heavy atoms for each protein-protein complex were taken from the highest-resolution crystal structure of the complex from the Protein Data Bank: 1BRS (chains A and D) for the barnase-barstar complex,[30] 1C1Y for the complex between the Ras binding domain of c-Raf1 kinase (RafRBD) and Ras homologue, Rap1A,[31] 3HHR (chains A and B) for the complex between human growth hormone and its receptor,[32] and 1NCA for the complex between influenza virus N9 neuraminidase and the NC41 antibody.[33] Hydrogen atoms were added using ionization states present at neutral pH. The N- and C- termini were modeled as charged except for the N-terminus of the human growth hormone receptor, which was capped with an acetyl group. All crystallographic water molecules were removed. In addition, the following ions and molecules that are distant from the binding sites were removed: Mg^{2+} ions, Ca^{2+} ions, and the GTP-analog from the RafRBD-Rap1A complex; the Ca^{2+} ion and sugar molecules from the neuraminidase-antibody complex. The following missing residues, which are all distant from the binding site, were reconstructed using the MODELLER 9v6 software package:[34] residues 1 and 2 of barnase (chain A), residues 64 and 65 of barstar (chain D), residues 149-153 of the human growth hormone (chain A), and residues 57-61, 74, 235 and 236 of the human growth hormone receptor (chain B). As necessary for the PME treatment of long-range electrostatics[15] in the explicit solvent calculations, the net charge of each protein-protein complex was neutralized by introducing the following mutations at the most distant locations from the binding site: E8K and E57K of barstar; E129K of Rap1A; E1K, E60K, D121K, E152K, D174K, and E178K of the human growth hormone receptor; E185K of the NC41 antibody (chain L).

To relieve unfavorable interactions, each protein-protein complex was energy minimized in two stages with the OPLS-AA/L force field[16] in a cubic box of TIP3P water molecules,[19] with each stage involving 1000 steps of steepest descent minimization. During the first stage, the positions of all hydrogens, reconstructed missing residues, and water molecules were minimized while applying harmonic position restraints to the heavy atoms of residues that are present in the crystal structure. During the second stage, minimization was conducted with no position restraints. The resulting, minimized protein structures (after removing all water molecules) were used as the bound states for both implicit and explicit solvent calculations. To generate the unbound states for these calculations, the two proteins in each protein-protein complex were separated by 30 Å between their centers of mass, resulting in a minimum distance of 15 Å between the proteins. For the explicit solvent calculations, the unbound and bound states of each protein-protein complex were solvated in cubic boxes of explicit water (TIP3P,[19] TIP4P,[19] or SPC/E[20]) with identical volumes that allowed for a minimum solute-wall distance of 12 Å. The number of atoms in the unbound and bound states of each system was enforced to be exactly the same by removing extra water molecules in the bound state, which contained more water molecules than the unbound state in all cases. When the TIP3P or SPC/E water model was used, the total number of water molecules in each protein system was the following: 47,753 for the barnase-barstar complex; 56,116 for the RafRBD-Rap1A complex; 77,780 for the growth hormone-receptor complex; and 178,663 for the neuraminidase-antibody complex; when the TIP4P water model was used, the numbers are the following (in the same order): 45,644; 55,685; 78,203; and 176,538.

2.7 REFERENCES

(1) Novotny, J.; Sharp, K.A. Electrostatic Fields in Antibodies and Antibody/Antigen Complex. *Prog. Biophys. Mol. Biol.* **1992,** *58*, 203-224.

(2) Hendsch, Z.S.; Tidor, B. Do Salt Bridges Stabilize Proteins? A Continuum Electrostatic Analysis. *Protein Sci*. **1994**, *3*, 211-226.

(3) Elcock, A.H. The Stability of Salt Bridges at High Temperatures: Implications for Hyperthermophilic Proteins. *J. Mol. Biol.* **1998,** *284*, 489-502.

(4) Hendsch, Z.S.; Tidor, B. Electrostatic Interactions in the GCN4 Leucine Zipper: Substantial Contributions Arise from Intramolecular Interactions Enhanced on Binding. *Protein Sci*. **1999**, *8*, 1381-1392.

(5) Sheinerman, F. B.; Honig, B. On the Role of Electrostatic Interactions in the Design of Protein-Protein Interfaces. *J. Mol. Biol*. **2002**, *318*, 161-177.

(6) Sitkoff, D.; Sharp, K.A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994,** *98*, 1978-1988.

(7) Nina, M.; Beglov, D.; Roux, B. Atomic Radii for Continuum Electrostatics Calculations Based on Molecular Dynamics Free Energy Simulations. *J. Phys. Chem.* **1997,** *101*, 5239-5248.

(8) Swanson, J.M.J.; Adcock, S.A.; McCammon, J.A. Optimized Radii for Poisson− Boltzmann Calculations with the AMBER Force Field. *J. Chem. Theory Comput.* **2005,** *1*, 484-493.

(9) Wagoner, J.A.; Baker, N.A. Assessing Implicit Models for Nonpolar Mean Solvation Forces: The Importance of Dispersion and Volume Terms. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 8331-8336.

(10) Zhang, L.Y.; Gallicchio, E.; Friesner, R.A.; Levy, R.M. Solvent Models for Protein-Ligand Binding: Comparison of Implicit Solvent Poisson and Surface Generalized Born Models with Explicit Solvent Simulations. *J. Comput. Chem.* **2001,** *22*, 591-607.

(11) Yu, Z.; Jacobson, M.P.; Josovitz, J.; Rapp, C.S.; Friesner, R.A. First-Shell Solvation of Ion Pairs: Correction of Systematic Errors in Implicit Solvent Models. *J. Phys. Chem. B* **2004,** *108*, 6643-6654.

(12) Tan, C.; Yang, L.; Luo, R. How Well Does Poisson− Boltzmann Implicit Solvent Agree with Explicit Solvent? A Quantitative Analysis. *J. Phys. Chem. B* **2006,** *110*, 18680-18687.

(13) Thomas, A.S.; Elcock, A.H. Direct Observation of Salt Effects on Molecular Interactions through Explicit-Solvent Molecular Dynamics Simulations: Differential Effects on Electrostatic and Hydrophobic Interactions and Comparisons to Poisson−Boltzmann Theory. *J. Am. Chem. Soc.* **2006,** *128*, 7796-7806.

(14) Bogusz, S.; Cheatham, T.E., 3rd; Brooks, B.R. Removal of Pressure and Free Energy Artifacts in Charged Periodic Systems via Net Charge Corrections to the Ewald Potential. *J. Chem. Phys.* **1998,** *108*, 3017-3020.

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

(16) Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995,** *103*, 8577-8593.

(17) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995,** *268*, 1144-1149.

(18) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Gromacs 4: Algorithms for Highly Efficient, Load-Balanced and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008,** *4*, 435-447.

(19) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.

(20) Berendsen, H.J.C.; Grigera, J.R.; Straatsma, T.P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987,** *91*, 6269-6271.

(21) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid Grid-Based Construction of the Molecular Surface and the Use of Induced Surface Charge to Calculate Reaction Field Energies: Applications to the Molecular Systems and Geometric Objects. *J. Comput. Chem.* **2002,** *23*, 128-137.

(22) Dong, F., Vijayakumar, M.; Zhou, H.-X. Comparison of Calculation and Experiment Implicates Significant Electrostatic Contributions to the Binding Stability of Barnase and Barstar. *Biophys. J.* **2003,** *85*, 49-60.

(23) Dong, F.; Zhou, H.-X. Electrostatic Contribution to the Binding Stability of Protein-Protein Complexes. *Proteins* **2006,** *65*, 87-102.

(24) Kumar, S.; Nussinov, R. Salt Bridge Stability in Monomeric Proteins. *J. Mol. Biol*. **1999**, 293, 1241–1255.

(25) Hess, B.; van der Vegt, N. F. A. Hydration Thermodynamic Properties of Amino Acid Analogues: A Systematic Comparison of Biomolecular Force Fields and Water Models. *J. Phys. Chem. B* **2006**, 110 (35), 17616–17626.

(26) Woo, H.-J.; Dinner, A.R.; Roux, B. Grand Canonical Monte Carlo Simulations of Water in Protein Environments. *J. Chem. Phys.* **2004,** 6392-6400.

(27) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935,** *3*, 300-313.

(28) Bishop, M.; Frinks, S. Error Analysis in Computer Simulations. *J. Chem. Phys.* **1987,** *87*, 3675-3676.

(29) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.

(30) Buckle, A. M.; Schreiber, G.; Fersht, A. R. Protein-Protein Recognition: Crystal Structural Analysis of a Barnase-Barstar Complex at 2.0 Ang. Resolution. *Biochemistry* **1994**, 33, 8878-8889.

(31) Nassar, N.; Horn, G.; Herrmann, C.; Scherer, A.; McCormick, F.; Wittinghofer, A. The 2.2 Å Crystal Structure of the Ras-Binding Domain of the Serine/Threonine Kinase C-Raf1 in Complex with Rapla And a GTP Analogue. *Nature* **1995**, 375, 554-560.

(32) de Vos, A. M.; Ultsch, M.; Kossiakoff, A. A. Human Growth Hormone and Extracellular Domain of its Receptor: Crystal Structure of the Complex. *Science* **1992**, 255, 306-312.

(33) Tulip, W. R.; Varghese, J. N.; Laver, W. G.; Webster, R. G.; Colman, P. M. Refined Crystal Structure of the Influenza Virus N9 Neuraminidase-NC41 Fab Complex. *J. Mol. Biol*. **1992**, 227, 122-148.

(34) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Protein Restraints. *J. Mol. Biol*. **1993**, 234, 779-815.

3.0 EFFECTS OF HIGH TEMPERATURE ON DESOLVATION COSTS OF SALT BRIDGES ACROSS PROTEIN BINDING INTERFACES: SIMILARITIES AND DIFFERENCES BETWEEN IMPLICIT AND EXPLICIT SOLVENT MODELS

This work was published as: Reza Salari and Lillian T. Chong (2012). "Effects of High Temperature on Desolvation Costs of Salt Bridges across Protein Binding Interfaces: Similarities and Differences between Implicit and Explicit Solvent Models." *J. Phys. Chem. B.* 116(8): 2561-7.

3.1 INTRODUCTION

Salt bridges are thought to make little contribution to the stability of protein-protein complexes at room temperature;[1-3] however, they are particularly abundant in hyperthermophilic proteins[4-7] and therefore appear to play critical roles in the adaptation of proteins for stability at extremely high temperatures (e.g. 100 ^{o}C). The latter point has been rationalized on the basis of theoretical studies, which determined the thermodynamic costs of desolvating the oppositely charged members of the salt bridge upon binding to be significant at room temperature,[1-3] but markedly reduced at high temperatures.[8] For efficient computations, these studies all employed a dielectric continuum solvent model based on the Poisson-Boltzmann (PB) equation, which is the gold standard of implicit solvent models. Despite the simplicity of the PB model, it has been possible to parameterize the model to reproduce solvation free energies of small organic

molecules determined by either experiment[9] or more costly simulations with explicit water molecules.[10,11] In addition, PB calculations have been found to be comparable to explicit solvent simulations in capturing temperature-dependent effects for the association of salt bridge analogues (i.e. acetate and methyl ammonium) from 0 to 100 ^{o}C,[12] provided that certain physical parameters are adjusted according to temperature.[13] Nevertheless, the PB model lacks important features such as molecular details of the first solvation shell, including "bridging" water molecules.[14,15] Valuable insights about modeling solvation effects can therefore be obtained by comparing implicit solvent calculations with more detailed explicit solvent calculations.[12,14-21]

Recently, we conducted a direct comparison of the PB implicit solvent model with several explicit solvent models in computing the desolvation penalties of salt bridges across a number of protein-protein interfaces at 25 ^{o}C and found overall agreement between the implicit and explicit solvent results.[14] Here, for the first time, we directly compare implicit and explicit solvent models in computing the desolvation penalties of salt bridges across protein-protein interfaces at high temperature (i.e. 100 ^{o}C). Both our comparisons at 25 and 100 ^{o}C involve the same set of salt bridges – namely, all 14 salt bridges across the binding interfaces of four protein-protein complexes (Figure 3-1) that had been identified by others as having a wide range of desolvation penalties.[3] As done in previous theoretical studies on the desolvation penalties of salt bridges,[2,3,14,22-25] we focused on a) rigid binding, with the unbound conformations of the proteins being identical to the corresponding bound conformations, and b) evaluating the desolvation penalties relative to those obtained when the charged side-chains are replaced by hydrophobic side-chains of identical size and shape (isosteres) i.e. $\Delta\Delta G_{solv}$; these hydrophobic isosteres are hypothetical mutant versions in which all partial charges on the salt bridge side chains are set to 0. In the implicit solvent calculations, desolvation penalties were computed using the PB model;

in the explicit solvent calculations, they were computed using thermodynamic integration techniques (see Methods). We explored the same three explicit water models as our previous study at 25 ^{o}C, namely TIP3P,[26] TIP4P,[26] and SPC/E.[27] Although the dielectric constants of these explicit water models may not be accurate for a given temperature, their relative changes from one temperature to another are likely to be in good agreement with experiment.[28] We therefore focused our comparison of implicit and explicit solvent models on their computed desolvation penalties of salt bridges at 100 ^{o}C *relative to 25 ^{o}C*.

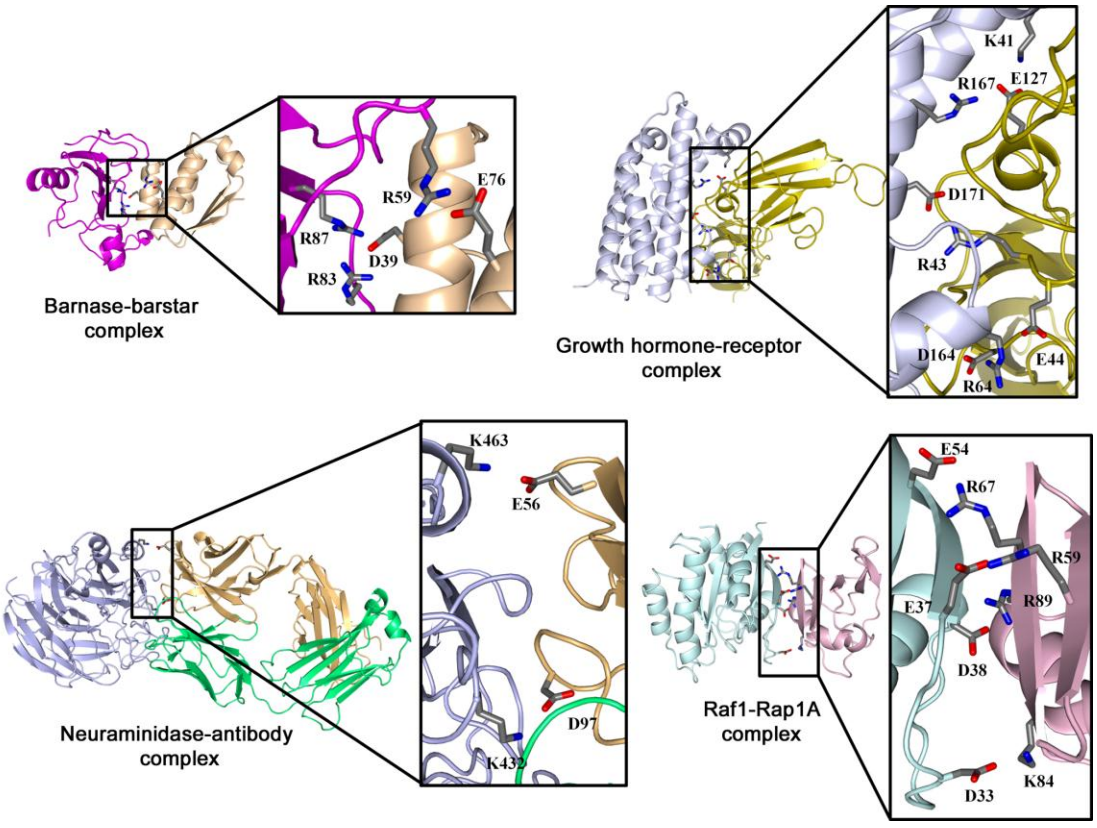

Figure 3-1. Locations of salt bridges across the binding interface of each protein-protein complex in this study.

3.2 METHODS

In order to directly compare the solvation thermodynamics of the implicit and explicit solvent models, it was essential to keep the proteins completely rigid, even in the explicit solvent molecular dynamics (MD) simulations. It was also necessary to fix all parameters common to the two approaches to ensure that they remained absolutely identical – that is, protein coordinates, atomic charges and radii (OPLS-AA/L force field)[29], box volume, and temperature. Periodic boundary conditions were employed in both approaches, enabling the use of the PME treatment of long-range electrostatics[30] for the explicit solvent calculations. As required by the PME method,[31] all systems were constructed to be electrically neutral by implementing the following in both implicit and explicit solvent calculations: 1) neutralizing the net charge of each protein-protein complex by introducing mutations at the most distant locations from the binding interface,[14] 2) representing the unbound state with the proteins separated by a distance (30 Å between their centers-of-mass) at which electrostatic interactions between the proteins were found to be negligible, and 3) simultaneously mutating the oppositely charged side chains of the salt bridge to their hydrophobic isosteres (i.e. turn off their partial charges) in both the unbound and bound states of the proteins.

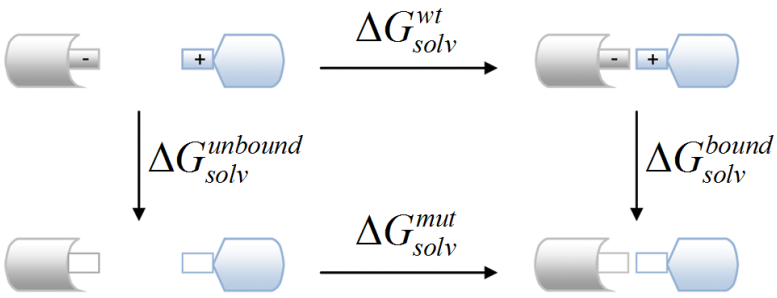

Figure 3-2. Thermodynamic cycle used for computing the desolvation penalty of a salt bridge upon protein binding relative to its hydrophobic isostere ($\Delta\Delta G_{solv}$). The wild-type salt bridge and its "mutant" hydrophobic isostere are represented by the filled and empty rectangles, respectively.

Desolvation penalties of salt bridges upon protein binding relative to their hydrophobic isosteres ($\Delta\Delta G_{solv}$) were computed according to the thermodynamic cycle shown in Figure 3-2. In particular, the $\Delta\Delta G_{solv}$ of each salt bridge was computed using the following equation, which circumvents the need to simulate the diffusional association of the proteins:

$$\Delta\Delta G_{solv} = \Delta G_{solv}^{unbound} - \Delta G_{solv}^{bound}$$

where $\Delta G_{solv}^{unbound}$ and ΔG_{solv}^{bound} are the solvation free energies of the wild-type unbound and bound states, respectively, relative to the corresponding "mutant" hydrophobic isostere states. Full details of the protein models as well as the implicit and explicit solvent calculations are provided in our previous publication involving the same set of salt bridges at 25 ^{o}C.[14] We summarize the key details of the calculations below.

3.2.1 Implicit Solvent Calculations

Implicit solvent calculations were performed using finite difference methods, as implemented in the DelPhi 4.0 software package,[32] to solve the linearized form of the PB

equation; this equation reduces to the Poisson equation in the absence of salt, as in our calculations. To represent the boundary between the low-dielectric protein region and high-dielectric solvent region, the standard molecular surface was used.[33] Consistent with keeping the proteins rigid, a dielectric constant of 1 was used for the protein region; to model solvation at 100 ^{o}C, the solvent dielectric constant was set to the experimental value for water at 100 ^{o}C (55.55, compared to 78.4 at 25 ^{o}C).[34] In addition, we tested the effects of scaling the atomic radii using a temperature-dependent radius scaling factor (RSF) that has been empirically derived for 100 ^{o}C: 1.012 for the NH_3^+ and guanidinium groups, 1.036 for carboxyl groups, and 1.046 for all other groups.[13] Consistent with the explicit solvent calculations, periodic boundary conditions were employed, implicitly including long-range electrostatic interactions with all periodic images. Electrostatic contributions to solvation free energies were determined by first directly calculating the induced polarization charges and then calculating the interaction between the protein charges and the reaction field due to the polarization charges.[35] The resulting electrostatic contributions are averages of 14 calculations involving systematic molecular translations on the grid, with uncertainties represented by the standard deviation. The electrostatic contribution to the solvation free energy of each salt bridge relative to its hydrophobic isostere in the unbound or bound state yields the solvation free energies $\Delta G_{solv}^{unbound}$ or ΔG_{solv}^{bound}, respectively. The desolvation penalty of each salt bridge relative to its hydrophobic isostere was then calculated using $\Delta\Delta G_{solv} = \Delta G_{solv}^{unbound} - \Delta G_{solv}^{bound}$. Nonpolar contributions to the solvation free energies were not calculated since these contributions are identical for the wild-type salt bridge and its hydrophobic isostere, cancelling out in the evaluation of $\Delta G_{solv}^{unbound}$ and ΔG_{solv}^{bound}.

3.2.2 Explicit Solvent Calculations

Explicit solvent calculations were performed using the thermodynamic integration approach with MD simulations in explicit solvent, as implemented in the GROMACS 4.0.4 software package.[36] In particular, we first calculated differences in the overall free energies of each salt bridge relative to its hydrophobic isostere in its unbound and bound states. To obtain differences in solely the solvation free energies, all nonbonded protein-protein interactions were subtracted from differences in the overall free energies. The desolvation penalty of each salt bridge upon protein binding relative to its hydrophobic isostere was then calculated as the difference of the unbound and bound solvation energies. Separate MD simulations of the proteins (unbound and bound states) were performed in the NVT ensemble at each of eight λ values, linearly reducing the partial charges of the side chains of the salt bridge from 0 (wild-type) to 1 (hydrophobic isostere). Results were considered converged if the uncertainty of each λ simulation was small (< 5%) and if the plot of $\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda}$ vs. λ was linear ($R^2 > 0.997$). Uncertainties in the free energies are derived from sampling errors in $\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda}$; errors at each λ value were estimated using block averaging,[37] as implemented in the g_analyze utility of GROMACS.[36]

Each λ simulation was performed for 1 ns at 100 ^{o}C in the NVT ensemble (constant number of atoms, volume, and temperature) using the Langevin thermostat (frictional coefficient of 1 ps^{-1}). Constant volume was enforced by solvating the unbound and bound states of each protein-protein complex in cubic boxes of explicit water (TIP3P,[26] TIP4P,[26] or SPC/E[27]) with identical volumes that allowed for a minimum solute-wall distance of 12 Å. To ensure a constant number

of atoms in the unbound and bound states, extra water molecules were removed from the bound state, which contained more water molecules than the unbound state in all cases. This removal was done before energy minimization of the entire system and subsequent equilibration of the solvent; both minimization and equilibration were performed prior to the production phase of each λ simulation. Equilibration of the solvent was performed in two stages: 1) 10 ps at constant temperature (100 ^{o}C) and volume, and 2) 100 ps at constant temperature (100 ^{o}C) and pressure (1 atm). Throughout all stages of the simulations, the proteins were kept rigid using the GROMACS "frozen" option, which sets the velocities of all protein atoms to 0. Real space electrostatic interactions were truncated at 10 Å, while the long-range components of these interactions were calculated using the PME method[30] with periodic boundary conditions. Van der Waals interactions were switched off smoothly between 8 and 9 Å. A 2-fs time step was used for all simulations.

3.3 RESULTS

As discussed above, we focused our comparison of implicit and explicit solvent models on computing desolvation penalties of salt bridges at 100 ^{o}C relative to 25 ^{o}C. We first performed calculations on the 14 salt bridges in the absence of the protein environment – that is, with the same geometries, but in solution and with the residues capped with acetyl and N-methyl groups at the N- and C-termini, respectively. We then performed calculations on the same set of salt bridges in the context of the proteins. In both the absence and presence of the protein environment, we examined the effect of including a temperature-dependent radius scaling factor (RSF) on the implicit solvent results.

3.3.1 Salt Bridges in the Absence of the Protein Environment

In our calculations involving the salt bridges in solution, only the TIP3P water model was used for the explicit solvent calculations. Figure 3-3C shows the correlation between the desolvation penalties of the salt bridges at 100 ^{o}C relative to 25 ^{o}C (in reference to their hydrophobic isosteres) for the PB implicit calculations vs. TIP3P explicit solvent calculations. As shown by previous studies,[8,28] the solvation free energies of the salt bridges become less favorable at high temperature, with the unbound state more adversely affected than the bound state, thereby reducing the magnitude of desolvation penalties incurred upon salt bridge formation. The resulting changes in the solvation free energies (in the unbound and bound states) as well as the desolvation penalties of the salt bridges are therefore expected to be negative in sign, as is the case in our results (see Table 3-1, Supporting Information). This reduction in desolvation penalties is underestimated by the implicit solvent calculations when the atomic radii are not scaled with temperature. Upon scaling the atomic radii, the rms deviation is reduced from 0.8 to 0.4 kcal/mol and the slope of the linear regression line increases from 0.16 to 0.86. The agreement of the implicit and explicit solvent results improves even more dramatically for the solvation free energies of the salt bridges in their unbound and bound states at 100 ^{o}C relative to 25 ^{o}C (Figures 3-3A and 3-3B, respectively; see also Tables 3-1 and 3-2 in Supporting Information). These improvements, which are consistent with those reported for acetate and methyl ammonium associations, are encouraging given that the RSF was derived to reproduce the solvation free energies of amino acids from experiments rather than explicit solvent simulations.[13]

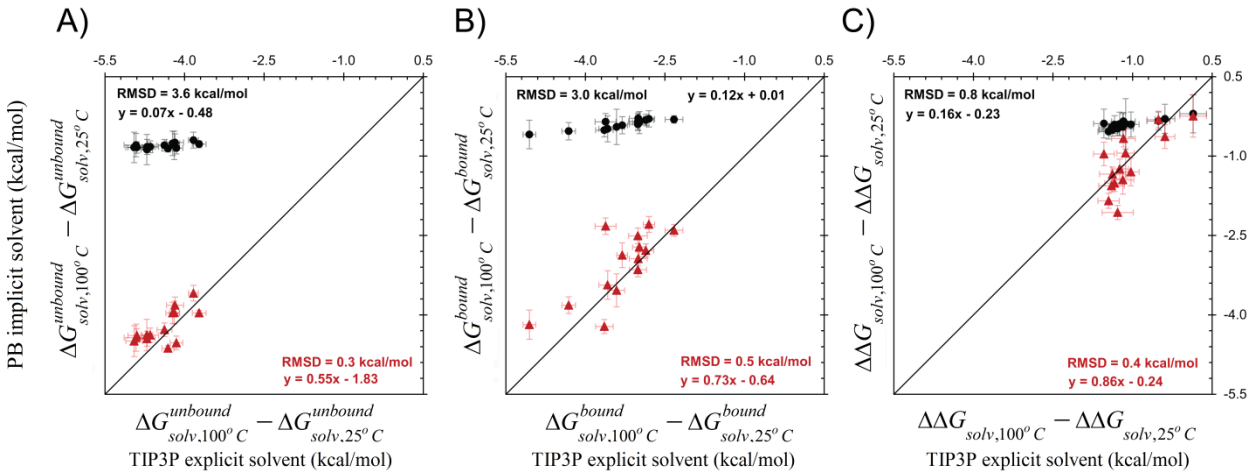

Figure 3-3. Comparison of the PB implicit solvent model and the TIP3P explicit solvent model for computing the solvation thermodynamics of salt bridges at 100 ^{o}C relative to 25 ^{o}C (in reference to their hydrophobic isosteres) in the absence of the protein environment: A) solvation free energies in the unbound state $\Delta\Delta G^{unbound}_{solv}$, B) solvation free energies in the bound state $\Delta\Delta G^{bound}_{solv}$, and C) desolvation penalties upon association $\Delta\Delta G_{solv} = \Delta\Delta G^{unbound}_{solv} - \Delta\Delta G^{bound}_{solv}$. Implicit solvent simulations were performed with and without a radius scaling factor (RSF) (red triangles and black circles, respectively). The diagonal lines represent perfect agreement; the rms deviations (RMSD) and equations for the linear regression in the bottom right and upper left corners of the plots correspond to implicit solvent calculations with and without the inclusion of an RSF, respectively. Error bars were calculated as described in Methods.

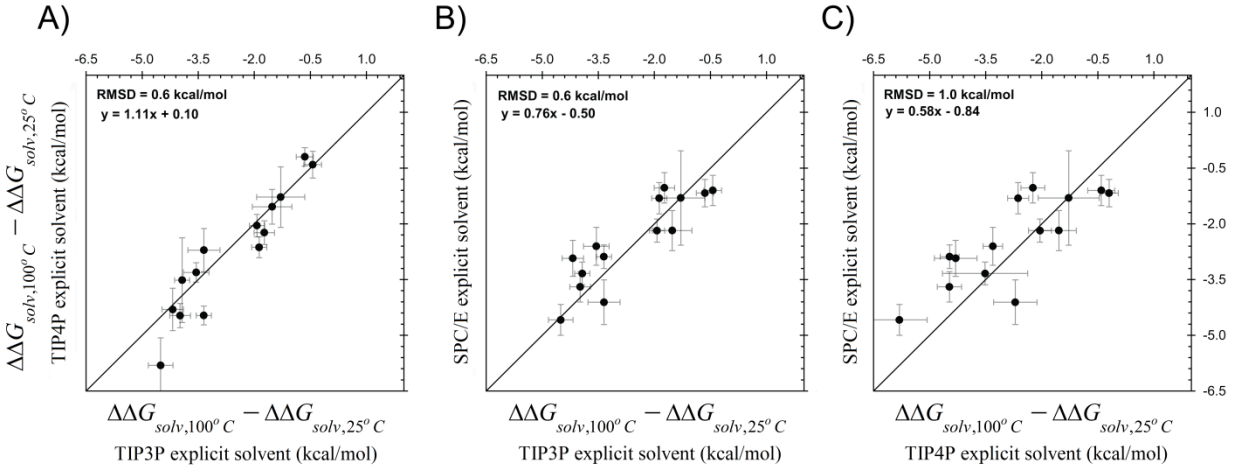

Figure 3-4. Comparison of explicit solvent models for computing the desolvation penalty of salt bridges upon protein-protein binding at 100 ^{o}C relative to 25 ^{o}C (in reference to their hydrophobic isosteres): A) TIP4P vs. TIP3P, B) SPC/E vs. TIP3P and C) SPC/E vs. TIP4P.

Diagonal lines represent perfect agreement; the rms deviations (RMSD) and equations for the linear regression are displayed in the upper left corners of the plots.

3.3.2 Salt Bridges in the Context of the Proteins

In our calculations involving the salt bridges in their protein environments, all three water models – TIP3P, TIP4P, and SPC/E – were tested for the explicit solvent calculations. The desolvation penalties of the salt bridges at 100 ^{o}C relative to 25 ^{o}C are similar for all of these explicit solvent models, with rms deviations of 0.6, 0.6, and 1.0 kcal/mol for TIP4P versus TIP3P (Figure 3-4A), SPC/E versus TIP3P (Figure 3-4B), and SPC/E versus TIP4P (Figure 3-4C), respectively. Figure 3-5C shows the correlation between the desolvation penalties for the PB implicit calculations vs. TIP3P explicit solvent calculations. As determined earlier for the corresponding salt bridges in the absence of their protein environments, these results reveal that the desolvation costs are reduced at high temperature (i.e. 100 ^{o}C) relative to room temperature. In contrast, however, inclusion of an RSF in the implicit solvent calculations only slightly improves the agreement between results for implicit and explicit solvent, regardless of the explicit water model (TIP3P, TIP4P, or SPC/E). For example, the rms deviation is lowered from 1.8 to 1.4 kcal/mol and the slope of the trend line is increased from 0.37 to 0.62 for the TIP3P explicit water model (Figure 3-5C).

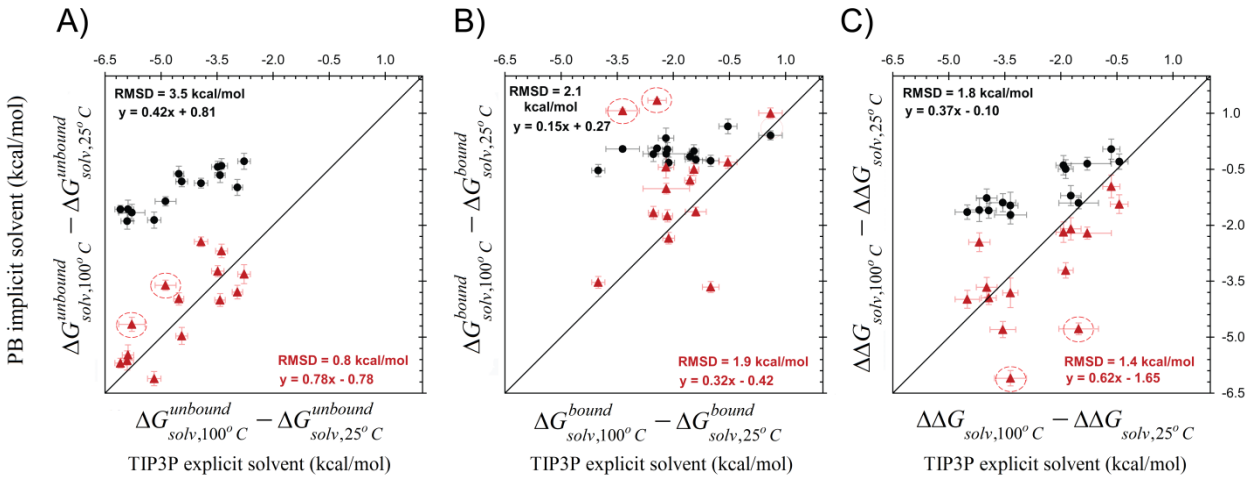

Figure 3-5. Comparison of implicit and explicit solvent models for computing the solvation thermodynamics of salt bridges at 100 °C relative to 25 °C (in reference to their hydrophobic isosteres) in the context of the proteins: A) solvation free energies in the unbound state $\Delta\Delta G^{unbound}_{solv}$, B) solvation free energies in the bound state $\Delta\Delta G^{bound}_{solv}$, and C) desolvation penalties upon association $\Delta\Delta G_{solv} = \Delta\Delta G^{unbound}_{solv} - \Delta\Delta G^{bound}_{solv}$. Implicit solvent simulations were performed with and without a radius scaling factor (RSF) (red triangles and black circles, respectively). The diagonal lines represent perfect agreement; the rms deviations (RMSD) and equations for the linear regression trend line in the bottom right and upper left corners of the plots correspond to implicit solvent calculations with and without the inclusion of an RSF, respectively. Error bars were calculated as described in Methods. Outliers discussed in the text are highlighted with dashed red circles.

To determine the source of this improvement (and why it is small), we examined the correlations between implicit and explicit solvent calculations in terms of the solvation free energies of the salt bridges at 100 °C relative to 25 °C in their unbound and bound states (Tables 3-3 to 3-6, Supporting Information). Inclusion of an RSF significantly reduces the implicit-explicit differences in the solvation free energies in the unbound states (Figure 3-5A), lowering the rms deviation from 3.5 to 0.8 kcal/mol and increasing the slope from 0.42 to 0.78. At first glance, inclusion of the RSF appears to have little effect on the solvation free energies in the bound states (Figure 3-5B), with the rms deviation remaining essentially the same (2.1 reduced

to 1.9 kcal/mol) while the slope increases from 0.15 to 0.32. However, the apparent absence of improvement is largely due to two outliers with negative values from the explicit solvent calculations, but unexpected positive values from the implicit solvent calculations. Once these outliers are removed, the rms deviation improves from 1.9 to 1.1 kcal/mol and the slope increases from 0.32 to 0.74. For the differences in desolvation penalties, the rms deviation improves from 1.4 to 0.9 kcal/mol (the slope of ~0.6 remains essentially unchanged). These differences are comparable to those between different explicit solvent models (i.e. rms deviation of 1.0 kcal/mol and slope of 0.58 for SPC/E vs. TIP4P), which is remarkable given the dramatic differences between implicit and explicit solvent models.

3.3.3 Further Examination of Outliers

The two outliers in the bound state of the proteins (Figure 3-5B) correspond to the R83-D39 and R87-D39 salt bridges across the binding interface of the barnase-barstar complex (Figure 3-1). These salt bridges are the only ones among our set of 14 that are (a) completely buried in the bound state according to the implicit solvent model and (b) involved in a "network" where at least one of the charged partners (i.e. D39) forms another salt bridge[3,38] (Table 3-7, Supporting Information). We therefore wondered if this unique combination of features might be responsible for the large implicit-explicit discrepancies associated with these salt bridges, causing the implicit solvent results to be too unfavorable and/or the explicit solvent results to be too favorable.

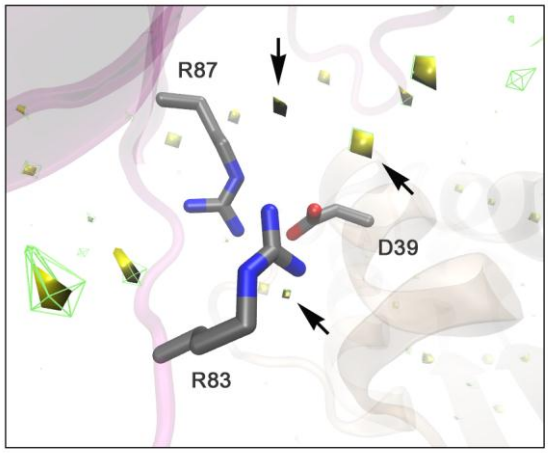

Figure 3-6. Most frequently visited positions of water molecules in the vicinity of the R83-D39 and R87-D39 salt bridges during explicit solvent simulations of the wild-type barnase-barstar complex (no hydrophobic isosteres) at 100 and 25 ^{o}C (yellow solid and green mesh regions, respectively). The arrows indicate regions in the protein cavity surrounding the R83-D39 and R87-D39 salt bridges that are occupied by water molecules that are within 5 Å of the salt bridges. These regions correspond to the locations of three crystallographic water molecules, two of which form hydrogen bonds with the barstar Asp39 residue. To map out these positions, the simulation box was first divided into ~1 Å^3 cubes and then the number of oxygen atoms of the water molecules were counted in each cube using snapshot configurations that were collected every ps from a 1-ns simulation of the barnase-barstar complex (total of 1000 configurations).

We first compared the degree to which these salt bridges are buried in both the implicit and explicit solvent calculations. While these salt bridges are completely buried within the context of the implicit solvent model, they are partially solvated by three crystallographic water molecules[39] that remain relatively fixed in the surrounding protein cavity at the barnase-barstar interface throughout the explicit solvent simulations (Figure 3-6); two of these water molecules form hydrogen bonds with the barstar Asp39 residue. To determine the effect of this partial solvation, we first removed these confined water molecules and then repeated the explicit solvent calculations at both 100 and 25 ^{o}C. The resulting solvation free energies of the salt bridges in their bound states at 100 ^{o}C relative to 25 ^{o}C ($\Delta G^{bound}_{solv,100^oC} - \Delta G^{bound}_{solv,25^oC}$) became less favorable and thereby closer in agreement with the implicit solvent results, increasing from -2.4 ± 0.2 to -

1.7 ± 0.4 kcal/mol for the R83-D39 salt bridge and from -3.4 ± 0.4 to -1.8 ± 0.3 kcal/mol for the R87-D39 salt bridge (uncertainties were computed as described in Methods). It appears therefore that the difference in solvent exposure of these salt bridges in the implicit vs. explicit solvent calculations is a source of the large implicit-explicit discrepancies in the effects of increasing temperature on the solvation free energy of the bound state. This difference arises because the implicit solvent model is not sufficiently detailed to capture the full complexity of the molecular surface at the protein-protein interface.

Next, we examined the contribution of the R83-D39-R87 salt bridge network to the implicit-explicit solvent discrepancies associated with the R83-D39 and R87-D39 salt bridges. In particular, for each of these two salt bridges, we first disrupted the network by mutating the other member of the network to its hydrophobic isostere (i.e. for the R83-D39 salt bridge, R87 is mutated; for the R87-D39 salt bridge, R83 is mutated) and then performed another set of both implicit and explicit solvent calculations for that salt bridge at 100 and 25 ^{o}C to evaluate the solvation free energies of the salt bridges in their bound states at 100 ^{o}C relative to 25 ^{o}C. While the explicit solvent results became even more negative (decreasing from -2.4 ± 0.2 to -6.2 ± 0.8 kcal/mol for the R83-D39 salt bridge and from -3.4 ± 0.4 to -4.4 ± 0.3 kcal/mol for the R87-D39 salt bridge), the implicit solvent results decreased from positive values to just zero. These results indicate that the implicit treatment of solvent somehow falls short of explicit solvent models in capturing the effects of high temperature on the solvation free energies of buried, networked salt bridges in their bound states. This limitation, in combination with the difference in solvent exposure between the implicit and explicit solvent calculations, appears to be at least partially responsible for the large implicit-explicit discrepancies associated with these salt bridges.

3.4 CONCLUSION

In conclusion, we performed a direct comparison of implicit and explicit solvent models in computing the desolvation penalties of salt bridges across a number of protein-protein interfaces at 100 ^{o}C relative to 25 ^{o}C. With the exception of two outliers, the implicit and explicit solvent results are of similar magnitudes and significantly reduced at 100 ^{o}C relative to 25 ^{o}C. As proposed previously based on solely implicit solvent calculations, the reduction in desolvation penalties at high temperature is a potential explanation for salt bridges playing crucial roles in promoting hyperhermostability in proteins despite making little favorable contribution to protein stability at room temperature.[8] Our study demonstrates that this proposal is also supported by more detailed explicit solvent calculations, based on the general agreement between our implicit and explicit solvent results. This agreement also demonstrates that implicit solvent models can be comparable to explicit solvent models in their ability to quantitatively account for the effects of increasing the temperature from 25 to 100 ^{o}C on the solvation thermodynamics of proteins. Nonetheless, significant discrepancies exist for particular salt bridges i.e. the two pairs in which the salt bridges are part of a salt bridge network that is completely buried in the implicit solvent model, but partially exposed to solvent in the explicit solvent simulations. For these salt bridges, the implicit solvent model does not appear to be sufficiently detailed to capture the effects of increasing temperature on the solvation thermodynamics, even after appropriate adjustment of its temperature-dependent parameters. Given the potential importance of salt bridge networks in proteins[38] and protein-protein complexes,[40] these challenging cases should be considered in the development of fast solvation approaches.

3.5 ACKNOWLEDGEMENT

We thank Matt Zwier for constructive comments on this manuscript. This work was supported by the NSF (CAREER MCB-0845216 and XSEDE MCB070090) and an Arts & Sciences Fellowship from the University of Pittsburgh to RS. Computational resources were also provided by the Center for Simulation and Modeling at the University of Pittsburgh.

3.6 SUPPORTING INFORMATION

Table 3-1. Computed desolvation penalties (kcal/mol) of salt bridges upon association in the absence of the protein environment at 100 ^{o}C. Implicit solvent calculations were performed with and without an empirically derived radius scaling factor (RSF). Uncertainties in parentheses are calculated as described in Methods.

$\Delta\Delta G_{solv,100^{o}C}$

Complex	*Salt Bridge*	*implicit (with RSF)*	*implicit (without RSF)*	*TIP3P*
Barnase-barstar	R59-E76	76.8 (0.1)	77.9 (0.1)	78.0 (0.1)
	R83-D39	81.1 (0.2)	82.1 (0.2)	82.1 (0.1)
	R87-D39	73.6 (0.2)	74.5 (0.2)	74.8 (0.1)
Growth hormone-receptor	K41-E127	100.5 (0.1)	101.8 (0.1)	100.9 (0.1)
	R64-E44	62.2 (0.1)	62.5 (0.2)	60.7 (0.1)
	R64-D164	65.3 (0.1)	65.6 (0.1)	63.0 (0.1)
	R167-E127	73.2 (0.1)	73.8 (0.2)	72.0 (0.1)
	D171-R43	75.0 (0.1)	75.9 (0.2)	75.4 (0.1)
Neuraminidase-antibody	K432-D97	63.5 (0.1)	63.5 (0.1)	59.6 (0.1)
	K463-E56	90.3 (0.1)	91.1 (0.1)	89.5 (0.1)
Raf1-Rap1A	D33-K84	84.9 (0.1)	86.0 (0.1)	86.3 (0.1)
	E37-R59	88.4 (0.1)	90.0 (0.1)	91.0 (0.1)
	D38-R89	69.8 (0.3)	70.4 (0.3)	70.4 (0.1)
	E54-R67	44.4 (0.3)	44.4 (0.2)	40.5 (0.1)

Table 3-2. Computed changes in the desolvation energy ($\Delta\Delta G_{solv,100^oC} - \Delta\Delta G_{solv,25^oC}$) of salt bridges upon association in the absence of the protein environment due to increasing the temperature from 25 to 100 ^oC. Also shown are the effects on the solvation free energies of the salt bridges in the unbound state ($\Delta G^{unbound}_{solv,100^oC} - \Delta G^{unbound}_{solv,25^oC}$) and bound state ($\Delta G^{bound}_{solv,100^oC} - \Delta G^{bound}_{solv,25^oC}$). Implicit solvent calculations were performed with and without an empirically derived radius scaling factor (RSF). Uncertainties in parentheses are calculated as described in Methods.

		$\Delta G^{unbound}_{solv,100^oC} - \Delta G^{unbound}_{solv,25^oC}$			$\Delta G^{bound}_{solv,100^oC} - \Delta G^{bound}_{solv,25^oC}$			$\Delta\Delta G_{solv,100^oC} - \Delta\Delta G_{solv,25^oC}$		
Complex	*Salt Bridge*	*implicit (with RSF)*	*implicit (without RSF)*	*TIP3P*	*implicit (with RSF)*	*implicit (without RSF)*	*TIP3P*	*implicit (with RSF)*	*implicit (without RSF)*	*TIP3P*
Barnase-barstar	R59-E76	-4.4 (0.0)	-0.8 (0.0)	-4.6 (0.0)	-2.9 (0.0)	-0.4 (0.0)	-3.3 (0.0)	-1.5 (0.0)	-0.4 (0.0)	-1.3 (0.0)
	R83-D39	-4.0 (0.1)	-0.8 (0.1)	-4.2 (0.2)	-2.5 (0.2)	-0.4 (0.2)	-3.0 (0.1)	-1.5 (0.2)	-0.4 (0.2)	-1.2 (0.2)
	R87-D39	-3.6 (0.2)	-0.7 (0.3)	-3.8 (0.1)	-2.3 (0.1)	-0.3 (0.1)	-2.8 (0.2)	-1.3 (0.3)	-0.4 (0.2)	-1.0 (0.2)
Growth hormone-receptor	K41-E127	-4.6 (0.1)	-0.9 (0.2)	-4.3 (0.1)	-2.8 (0.1)	-0.3 (0.2)	-2.9 (0.1)	-1.8 (0.3)	-0.5 (0.2)	-1.5 (0.2)
	R64-E44	-4.4 (0.1)	-0.8 (0.1)	-4.7 (0.1)	-3.8 (0.1)	-0.5 (0.1)	-4.3 (0.1)	-0.6 (0.1)	-0.3 (0.1)	-0.4 (0.2)
	R64-D164	-3.8 (0.1)	-0.7 (0.1)	-4.2 (0.1)	-3.1 (0.2)	-0.4 (0.2)	-3.0 (0.1)	-0.7 (0.2)	-0.3 (0.3)	-1.2 (0.2)
	R167-E127	-4.5 (0.2)	-0.8 (0.2)	-5.0 (0.2)	-3.5 (0.1)	-0.4 (0.1)	-3.4 (0.2)	-1.0 (0.2)	-0.4 (0.2)	-1.5 (0.2)
	D171-R43	-4.0 (0.3)	-0.8 (0.3)	-4.2 (0.2)	-2.7 (0.3)	-0.4 (0.3)	-3.0 (0.2)	-1.2 (0.2)	-0.4 (0.3)	-1.2 (0.2)
Neuraminidase-antibody	K432-D97	-4.5 (0.2)	-0.8 (0.2)	-4.2 (0.1)	-4.2 (0.2)	-0.5 (0.2)	-3.6 (0.1)	-0.3 (0.2)	-0.3 (0.2)	-0.5 (0.2)
	K463-E56	-4.3 (0.1)	-0.8 (0.1)	-4.4 (0.1)	-2.9 (0.1)	-0.3 (0.1)	-3.0 (0.2)	-1.3 (0.2)	-0.5 (0.2)	-1.4 (0.2)
Raf1-Rap1A	D33-K84	-4.0 (0.1)	-0.8 (0.1)	-3.7 (0.1)	-2.4 (0.1)	-0.3 (0.1)	-2.3 (0.2)	-1.6 (0.1)	-0.5 (0.1)	-1.4 (0.2)
	E37-R59	-4.4 (0.1)	-0.8 (0.1)	-4.9 (0.1)	-2.3 (0.1)	-0.4 (0.1)	-3.6 (0.2)	-2.1 (0.1)	-0.5 (0.1)	-1.3 (0.2)
	D38-R89	-4.4 (0.1)	-0.9 (0.1)	-4.7 (0.2)	-3.4 (0.2)	-0.5 (0.2)	-3.6 (0.2)	-0.9 (0.1)	-0.4 (0.1)	-1.1 (0.3)
	E54-R67	-4.4 (0.3)	-0.8 (0.3)	-4.9 (0.1)	-4.2 (0.3)	-0.6 (0.2)	-5.1 (0.2)	-0.2 (0.5)	-0.2 (0.5)	0.1 (0.2)

Table 3-3. Computed desolvation penalties (kcal/mol) of salt bridges upon protein-protein binding at 100°C. Implicit solvent calculations were performed without and with an empirically derived radius scaling factor (RSF). Uncertainties in parentheses are calculated as described in Methods.

$$\Delta\Delta G_{solv,100^{o}C}$$

		implicit solvent		*explicit solvent*		
Complex	*Salt Bridge*	*with RSF*	*without RSF*	*TIP3P*	*TIP4P*	*SPC/E*
Barnase-barstar	R59-E76	169.8 (0.1)	172.1 (0.1)	170.7 (0.1)	169.8 (0.7)	170.8 (0.2)
	R83-D39	172.6 (0.1)	177.0 (0.2)	175.2 (0.3)	176.9 (0.6)	178.1 (0.3)
	R87-D39	149.9 (0.1)	153.3 (0.1)	155.9 (0.4)	155.7 (0.4)	159.1 (0.4)
Growth hormone-receptor	K41-E127	45.9 (0.1)	48.6 (0.2)	59.6 (0.1)	59.8 (0.1)	60.7 (0.2)
	R64-E44	16.0 (0.2)	17.1 (0.1)	21.6 (0.1)	14.0 (0.1)	21.1 (0.2)
	R64-D164	18.8 (0.2)	20.6 (0.2)	26.9 (0.1)	28.4 (0.2)	27.6 (0.2)
	R167-E127	44.0 (0.2)	46.4 (0.2)	56.8 (0.1)	57.2 (0.2)	58.0 (0.2)
	D171-R43	61.4 (0.1)	63.3 (0.1)	61.9 (0.6)	61.8 (0.3)	64.8 (1.0)
Neuraminidase-antibody	K432-D97	82.7 (0.1)	86.1 (0.2)	90.9 (0.2)	88.0 (0.1)	91.4 (0.2)
	K463-E56	8.5 (0.2)	9.5 (0.1)	9.5 (0.2)	9.5 (0.1)	9.7 (0.2)
Raf1-Rap1A	D33-K84	192.4 (0.1)	193.3 (0.2)	188.1 (0.1)	185.4 (0.2)	188.7 (0.2)
	E37-R59	202.5 (0.3)	204.8 (0.2)	207.1 (0.1)	206.4 (0.1)	208.0 (0.2)
	D38-R89	198.8 (0.2)	201.1 (0.1)	196.0 (0.2)	197.2 (0.2)	198.0 (0.3)
	E54-R67	148.7 (0.2)	149.6 (0.2)	140.8 (0.2)	138.4 (0.2)	140.9 (0.2)

Table 3-4. Computed changes in the solvation free energy (kcal/mol) of salt bridges in the unbound states of the protein-protein complexes due to increasing the temperature from 25 to 100°C. Implicit solvent calculations were performed without and with radius scaling factor (RSF). Uncertainties in parentheses are calculated as described in Methods.

$$\Delta G^{unbound}_{solv,100^o C} - \Delta G^{unbound}_{solv,25^o C}$$

		implicit solvent		*explicit solvent*		
Complex	*Salt Bridge*	*with RSF*	*without RSF*	*TIP3P*	*TIP4P*	*SPC/E*
Barnase-barstar	R59-E76	-5.7 (0.1)	-1.6 (0.1)	-6.1 (0.1)	-6.2 (0.2)	-5.5 (0.2)
	R83-D39	-4.7 (0.2)	-1.7 (0.2)	-5.8 (0.4)	-5.6 (0.2)	-3.6 (0.3)
	R87-D39	-3.6 (0.1)	-1.4 (0.1)	-4.9 (0.3)	-3.3 (0.2)	-3.6 (0.4)
Growth hormone-receptor	K41-E127	-4.0 (0.2)	-0.7 (0.2)	-3.4 (0.1)	-3.6 (0.2)	-3.2 (0.2)
	R64-E44	-5.0 (0.2)	-0.8 (0.1)	-4.4 (0.2)	-4.4 (0.3)	-4.2 (0.3)
	R64-D164	-2.7 (0.2)	-0.4 (0.2)	-3.4 (0.2)	-2.8 (0.2)	-3.0 (0.2)
	R167-E127	-4.0 (0.2)	-0.6 (0.2)	-4.5 (0.1)	-4.9 (0.2)	-4.1 (0.2)
	D171-R43	-3.2 (0.1)	-0.4 (0.2)	-3.5 (0.2)	-3.5 (0.2)	-2.6 (0.3)
Neuraminidase-antibody	K432-D97	-3.8 (0.2)	-1.0 (0.2)	-3.0 (0.1)	-3.0 (0.2)	-2.3 (0.2)
	K463-E56	-3.3 (0.3)	-0.3 (0.2)	-2.8 (0.2)	-2.2 (0.2)	-2.5 (0.2)
Raf1-Rap1A	D33-K84	-6.1 (0.2)	-1.9 (0.2)	-5.2 (0.2)	-5.3 (0.3)	-4.3 (0.3)
	E37-R59	-5.5 (0.3)	-1.6 (0.2)	-5.9 (0.2)	-6.3 (0.2)	-4.9 (0.3)
	D38-R89	-5.6 (0.3)	-1.9 (0.2)	-5.9 (0.2)	-6.8 (0.2)	-5.8 (0.2)
	E54-R67	-2.4 (0.1)	-0.9 (0.1)	-3.9 (0.2)	-3.9 (0.2)	-3.4 (0.3)

Table 3-5. Computed changes in the solvation free energies (kcal/mol) of salt bridges in the bound states of the protein-protein complexes due to increasing the temperature from 25 to 100°C. Implicit solvent calculations were performed with and without an empirically derived radius scaling factor (RSF). Uncertainties in parentheses are calculated as described in Methods.

$$\Delta G^{bound}_{solv,100^oC} - \Delta G^{bound}_{solv,25^oC}$$

		implicit solvent		*explicit solvent*		
Complex	*Salt Bridge*	*with RSF*	*without RSF*	*TIP3P*	*TIP4P*	*SPC/E*
Barnase-barstar	R59-E76	-1.7 (0.2)	0.0 (0.2)	-2.2 (0.1)	-2.7 (1.1)	-2.2 (0.2)
	R83-D39	1.3 (0.0)	0.1 (0.1)	-2.4 (0.2)	-2.9 (0.5)	0.5 (0.5)
	R87-D39	1.1 (0.1)	0.0 (0.1)	-3.4 (0.4)	-1.8 (0.4)	-1.4 (0.3)
Growth hormone-receptor	K41-E127	-0.8 (0.1)	-0.2 (0.1)	-1.6 (0.1)	-1.0 (0.2)	-1.9 (0.4)
	R64-E44	-3.5 (0.2)	-0.5 (0.2)	-4.0 (0.2)	-4.0 (0.2)	-3.1 (0.3)
	R64-D164	-0.5 (0.2)	0.0 (0.2)	-1.5 (0.1)	-0.8 (0.2)	-0.8 (0.2)
	R167-E127	-0.3 (0.2)	0.6 (0.2)	-0.5 (0.2)	-0.4 (0.3)	-0.4 (0.4)
	D171-R43	-1.0 (0.1)	-0.1 (0.1)	-2.2 (0.6)	-2.2 (0.8)	-1.3 (1.2)
Neuraminidase-antibody	K432-D97	1.0 (0.1)	0.4 (0.1)	0.6 (0.3)	0.3 (0.2)	0.4 (0.5)
	K463-E56	-2.3 (0.2)	-0.3 (0.2)	-2.1 (0.2)	-2.0 (0.2)	-1.4 (0.3)
Raf1-Rap1A	D33-K84	-3.7 (0.1)	-0.3 (0.1)	-1.0 (0.2)	-1.0 (0.5)	-1.3 (0.4)
	E37-R59	-1.7 (0.2)	-0.1 (0.2)	-2.5 (0.1)	-1.8 (0.2)	-2.1 (0.2)
	D38-R89	-1.6 (0.1)	-0.2 (0.1)	-1.4 (0.3)	-0.9 (0.7)	-1.2 (0.3)
	E54-R67	-0.4 (0.3)	0.3 (0.3)	-2.2 (0.2)	-1.6 (0.2)	-2.3 (0.3)

Table 3-6. Computed changes in the desolvation penalties (kcal/mol) of salt bridges upon protein-protein binding due to increasing the temperature from 25 to 100 ^{o}C. Implicit solvent calculations were performed with and without an empirically derived radius scaling factor (RSF). Uncertainties in parentheses are calculated as described in Methods.

$$\Delta\Delta G_{solv,100^{o}C} - \Delta\Delta G_{solv,25^{o}C}$$

		implicit solvent		*explicit solvent*		
Complex	*Salt Bridge*	*with RSF*	*without RSF*	*TIP3P*	*TIP4P*	*SPC/E*
Barnase-barstar	R59-E76	-3.9 (0.2)	-1.6 (0.2)	-3.9 (0.2)	-3.5 (1.1)	-3.3 (0.3)
	R83-D39	-6.1 (0.2)	-1.7 (0.2)	-3.3 (0.4)	-2.7 (0.6)	-4.1 (0.6)
	R87-D39	-4.8 (0.2)	-1.4 (0.2)	-1.5 (0.5)	-1.5 (0.5)	-2.2 (0.5)
Growth hormone-receptor	K41-E127	-3.2 (0.2)	-0.5 (0.3)	-1.9 (0.2)	-2.6 (0.3)	-1.3 (0.4)
	R64-E44	-1.4 (0.3)	-0.3 (0.2)	-0.4 (0.2)	-0.4 (0.4)	-1.1 (0.4)
	R64-D164	-2.2 (0.3)	-0.4 (0.3)	-1.9 (0.2)	-2.0 (0.3)	-2.2 (0.3)
	R167-E127	-3.7 (0.3)	-1.3 (0.2)	-4.0 (0.3)	-4.5 (0.3)	-3.7 (0.4)
	D171-R43	-2.2 (0.2)	-0.4 (0.2)	-1.3 (0.6)	-1.3 (0.8)	-1.3 (1.3)
Neuraminidase-antibody	K432-D97	-4.8 (0.2)	-1.4 (0.2)	-3.6 (0.3)	-3.3 (0.3)	-2.6 (0.5)
	K463-E56	-1.0 (0.3)	0.0 (0.3)	-0.6 (0.2)	-0.2 (0.2)	-1.2 (0.4)
Raf1-Rap1A	D33-K84	-2.5 (0.3)	-1.6 (0.3)	-4.2 (0.3)	-4.3 (0.6)	-2.9 (0.5)
	E37-R59	-3.8 (0.4)	-1.5 (0.4)	-3.4 (0.2)	-4.5 (0.3)	-2.9 (0.3)
	D38-R89	-4.0 (0.2)	-1.6 (0.2)	-4.5 (0.3)	-5.8 (0.7)	-4.6 (0.4)
	E54-R67	-2.1 (0.3)	-1.2 (0.3)	-1.7 (0.3)	-2.2 (0.3)	-1.0 (0.4)

Table 3-7. The percent burial of the salt bridges upon protein-protein binding and whether or not the salt bridge is networked. Salt bridges are considered "networked" if at least one of the charged partners forms another salt bridge.

Complex	*Salt Bridge*	*% burial*	*Networked?*
Barnase-barstar	R59-E76	71	
	R83-D39	100	x
	R87-D39	100	x
Growth hormone-receptor	K41-E127	83	x
	R64-E44	66	x
	R64-D164	68	x
	R167-E127	83	x
	D171-R43	94	
Neuraminidase-antibody	K432-D97	84	
	K463-E56	39	
Raf1-Rap1A	D33-K84	51	
	E37-R59	55	
	D38-R89	88	
	E54-R67	59	

3.7 REFERENCES

(1) Novotny, J.; Sharp, K. A. Electrostatic Fields in Antibodies and Antibody/Antigen Complex. *Prog. Biophys. Mol. Biol.* **1992**, *58*, 203.

(2) Hendsch, Z. S.; Tidor, B. Electrostatic Interactions in the GCN4 Leucine Zipper: Substantial Contributions Arise from Intramolecular Interactions Enhanced on Binding. *Protein Sci.* **1999**, *8*, 1381.

(3) Sheinerman, F. B.; Honig, B. On the Role of Electrostatic Interactions in the Design of Protein-Protein Interfaces. *J. Mol. Biol.* **2002**, *318*, 161.

(4) Sterner, R.; Liebl, W. Crit. Rev. Biochem. Thermophilic Adaptation of Proteins. *Mol. Biol.* **2001**, *36*, 39.

(5) Vieille, C.; Zeikus, G. J. Microbiol. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Mol. Biol. Rev.* **2001**, *65*, 1.

(6) Kumar, S.; Nussinov, R. How Do Thermophilic Proteins Deal with Heat? *Cell. Mol. Life Sci.* **2001**, *58*, 1216.

(7) Karshikoff, A.; Ladenstein, R. Ion Pairs and the Thermotolerance of Proteins from Hyperthermophiles: a "Traffic Rule" for Hot Roads. *Trends Biochem. Sci.* **2001**, *26*, 550.

(8) Elcock, A. H. The Stability of Salt Bridges at High Temperatures: Implications for Hyperthermophilic Proteins. *J. Mol. Biol.* **1998**, *284*, 489.

(9) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978.

(10) Nina, M.; Beglov, D.; Roux, B. Atomic Radii for Continuum Electrostatics Calculations Based on Molecular Dynamics Free Energy Simulations. *J. Phys. Chem.* **1997**, *101*, 5239.

(11) Swanson, J. M. J.; Adcock, S. A.; McCammon, J. A. Optimized Radii for Poisson-Boltzmann Calculations with the AMBER Force Field. *J. Chem. Theory Comput.* **2005**, *1*, 484.

(12) Zhu, S.; Elcock, A. H. A Complete Thermodynamic Characterization of Electrostatic and Hydrophobic Associations in the Temperature Range 0 to 100 °C from Explicit-Solvent Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2010**, *6*, 1293.

(13) Elcock, A. H.; McCammon, Continuum Solvation Model for Studying Protein Hydration Thermodynamics at High Temperatures. J. A. *J. Phys. Chem.* B **1997**, *101*, 9624.

(14) Salari, R.; Chong, L. T. Desolvation Costs of Salt Bridges across Protein Binding Interfaces: Similarities and Differences between Implicit and Explicit Solvent Models. *J. Phys. Chem. Lett.* **2010**, *1*, 2844.

(15) Yu, Z. Y.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. First-Shell Solvation of Ion Pairs: Correction of Systematic Errors in Implicit Solvent Models. *J. Phys. Chem. B* **2004**, *108*, 6643.

(16) Wagoner, J. A.; Baker, N. A. A. Assessing Implicit Models for Nonpolar Mean Solvation Forces: The Importance of Dispersion and Volume Terms. *Proc. Natl. Acad. Sci.* USA **2006**, *103*, 8331.

(17) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. Solvent Models for Protein-Ligand Binding: Comparison of Implicit Solvent Poisson and Surface Generalized Born Models with Explicit Solvent Simulations. *J. Comput. Chem.* **2001**, *22*, 591.

(18) Tan, C.; Yang, L.; Luo, R. HowWell Does Poisson-Boltzmann Implicit Solvent Agree with Explicit Solvent? A Quantitative Analysis. *J. Phys. Chem. B* **2006**, *110*, 18680.

(19) Thomas, A. S.; Elcock, A. H. Direct Observation of Salt Effects on Molecular Interactions through Explicit-Solvent Molecular Dynamics Simulations: Differential Effects on Electrostatic and Hydrophobic Interactions and Comparisons to Poisson-Boltzmann Theory. *J. Am. Chem. Soc.* **2006**, *128*, 7796.

(20) Shang, Y.; Nguyen, H.; Wickstrom, L.; Okur, A.; Simmerling, C. Improving the Description of Salt Bridge Strength and Geometry in a Generalized Born Model. *J. Mol. Graph. Model.* **2011**, *29*, 676.

(21) Wang, J.; Tan, C.; Chanco, E.; Luo, R. Quantitative Analysis of Poisson-Boltzmann Implicit Solvent in Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1194.

(22) Green, D. F.; Tidor, B. Design of Improved Protein Inhibitors of HIV-1 Cell Entry: Optimization of Electrostatic Interactions at the Binding Interface. *Proteins* **2005**, *60*, 644.

(23) Altman, M. D.; Nalivaika, E. A.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Tidor, B. Computational Design and Experimental Study of Tighter Binding Peptides to an Inactivated Mutant of HIV-1 Protease. *Proteins* **2008**, *70*, 678.

(24) Kajander, T.; Sachs, J. N.; Goldman, A.; Regan, L. Electrostatic Interactions of Hsp-Organizing Protein Tetratricopeptide Domains with Hsp70 and Hsp90: Computational Analysis and Protein Engineering. *J Biol Chem* **2009**, *284*, 25364.

(25) Carrascal, N.; Green, D. F. Energetic Decomposition with the Generalized-Born and Poisson-Boltzmann Solvent Models: Lessons from Association of G-protein Components. *J Phys Chem B* **2010**, *114*, 5096.

(26) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.

(27) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269.

(28) Thomas, A. S.; Elcock, A. H. Molecular Simulations Suggest Protein Salt Bridges are Uniquely Suited to Life at High Temperatures. *J. Am. Chem. Soc.* **2004**, *126*, 2208.

(29) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474.

(30) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.

(31) Bogusz, S.; Cheatham, T. E., 3rd; Brooks, B. R. Removal of Pressure and Free Energy Artifacts in Charged Periodic Systems via Net Charge Corrections to the Ewald Potential. *J. Chem. Phys.* **1998**, *108*, 3017.

(32) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144.

(33) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid Grid-Based Construction of the Molecular Surface and the Use of Induced Surface Charge to Calculate Reaction Field Energies: Applications to the Molecular Systems and Geometric Objects. *J. Comput. Chem.* **2002**, *23*, 128.

(34) Archer, D. G.; Wang, P. The Dielectric Constant of Water and Debye-Hückel Limiting Law Slopes. *J. Phys. Chem. Ref. Data* **1990**, *19*, 371.

(35) Dong, F.; Zhou, H.-X. Electrostatic Contribution to the Binding Stability of Protein-Protein Complexes. *Proteins* **2006**, *65*, 87.

(36) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Gromacs 4: Algorithms for Highly Efficient, Load-Balanced and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(37) Bishop, M.; Frinks, S. Error Analysis in Computer Simulations. *J. Chem. Phys.* **1987**, *87*, 3675.

(38) Kumar, S.; Nussinov, R. Salt Bridge Stability in Monomeric Proteins. *J. Mol. Biol.* **1999**, *293*, 1241.

(39) Buckle, A. M.; Schreiber, G.; Fersht, A. R. Protein-protein Recognition: Crystal Structural Analysis of a Barnase-Barstar Complex at 2.0 Å Resolution. *Biochemistry* **1994**, *33*, 8878.

(40) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic Aspects of Protein-protein Interactions. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.

4.0 EFFECTS OF SALT ON DESOLVATION COSTS OF SALT BRIDGES ACROSS PROTEIN BINDING INTERFACES: SIMILARITIES AND DIFFERENCES BETWEEN IMPLICIT AND EXPLICIT SOLVENT MODELS

4.1 INTRODUCTION

The contribution of salt bridges to the stability of protein-protein complexes is mainly determined by the cost of desolvating the salt bridge residues upon binding. Based on the theoretical studies, it is thought that such desolvation penalties are not fully compensated upon complex formation and therefore salt bridges are believed to make minor (or no) favorable contribution to protein binding at room temperature.[1-3] These desolvation costs are also subject to the environmental conditions such as temperature and salt. Salt ions can have significant effects on proteins, influencing their stability and solubility (e.g. the Hofmeister effect).[4] The effect of salt on the desolvation penalty of salt bridges, which is difficult to examine experimentally, has been investigated in several theoretical studies in the past.[5,6] For efficient computations, all these previous studies have been limited to a dielectric continuum solvent model based on the Poisson-Boltzmann (PB) equation. While the PB model is considered the gold standard of implicit solvation models, it treats both solvent and salt implicitly and therefore lacks important features

such as molecular details of the first solvation shell, bridging water molecules and ion-ion correlations.[7-9] Thus, important insight about modeling solvation can be obtained by comparing implicit and explicit solvent simulations.[7,9-13]

Previously we performed a direct comparison of the PB implicit solvent model with several explicit solvent models in computing the desolvation penalties of salt bridges across a number of protein-protein interfaces at 25 ^{o}C and found an overall agreement between the implicit and explicit solvent results.[8] We also explored how the two solvent models compare in capturing the effect of high temperature on the solvation free energies and found that, with the exception of two salt bridges, both models performed comparably and predicted a significant reduction of the desolvation penalties at 100 ^{o}C relative to 25 ^{o}C.[14] Here, we extend this direct comparison on the same set of 14 salt bridges to study how the implicit and explicit solvent models perform in capturing the effects of various concentrations of NaCl on the desolvation penalties at 25 ^{o}C.

As done in previous theoretical studies on the desolvation penalties of salt bridges,[2,3,8,14-16] we focused on rigid binding, with the unbound conformations of the proteins being identical to the corresponding bound conformations. The desolvation penalty for each salt bridge upon binding was calculated relative to its hydrophobic isostere, a hypothetical mutant where all partial charges on the salt bridge side chains are set to 0. In the implicit solvent calculations, desolvation penalties were computed using the PB model. For the explicit solvent calculations, the TIP3P[17] water model was used and the thermodynamic integration (TI) technique was employed to compute desolvation penalties.

4.2 METHODS

As in our previous studies,[8,14] to enable a direct comparison between the solvation thermodynamics of the implicit and explicit solvent models, proteins were kept completely rigid in both models and all parameters common to the two approaches were kept identical; namely protein coordinates, atomic charges and radii (OPLS-AA/L forcefield),[18] box volume, temperature and the application of periodic boundary conditions. In order to investigate the effect of NaCl on solvation free energies, five different concentrations of salt were tested: 0.1, 0.3, 0.5, 1.0 and 2.0 M. At each of these concentrations, implicit and explicit solvent simulations were performed and the desolvation penalties of salt bridges upon protein binding ($\Delta\Delta G_{solv}$) were calculated using the thermodynamic cycle shown in Figure 4-1. Full details of the protein models as well as implicit and explicit solvent calculations have been previously described.[8] Here we summarize the procedure and outline the key differences related to this work.

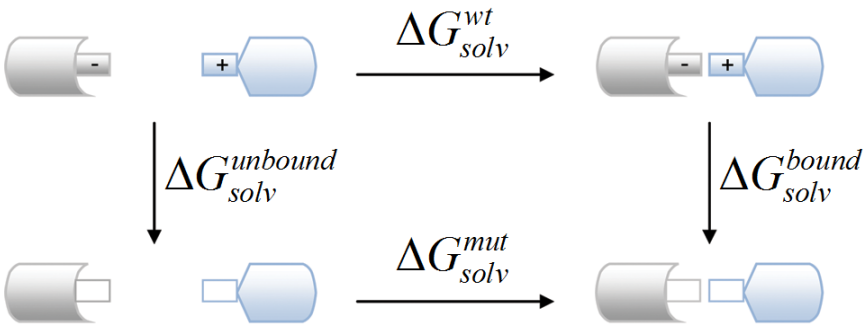

Figure 4-1. Thermodynamic cycle used for computing the desolvation penalty of a salt bridge upon protein binding relative to its hydrophobic isostere ($\Delta\Delta G_{solv}$). Filled and empty rectangles were used to represent the wild-type salt bridge and its "mutant" hydrophobic isostere, respectively.

4.2.1 Implicit solvent calculations

Implicit solvent calculations were performed using finite difference methods, as implemented in the DelPhi 4.0 software package,[19] to solve the linearized form of the PB equation. NaCl was represented as monovalent salt with an ion exclusion radius of 2.0 Å. Consistent with keeping the proteins rigid, a dielectric constant of 1 was used for the protein region; to represent the dielectric properties of water at 25 ^{o}C, a dielectric constant of 78.4 was used for the solvent region. Molecular surface was used to represent the boundary between the proteins and solvent.[20] Electrostatic contributions to solvation free energies were determined by the sum of two terms:[6,21] the reaction field energy, calculated as the interaction between the protein charges and the reaction field due to the polarization charges, and the external ion contribution, calculated as the difference between grid energy at particular salt concentration minus grid energy at zero salt. The reported solvation energies are averages of 14 calculations involving systematic molecular translations on the grid (in order to estimate the discretization error resulting from finite difference method in DelPhi)[19] and uncertainties represented by the standard deviation. The electrostatic contributions to the solvation free energy of each salt bridge (relative to its hydrophobic isostere) in the unbound or bound state, as well as the desolvation penalties, were computed according to Figure 4-1. Nonpolar contributions to the solvation free energies were not calculated since these contributions are identical for the wild-type salt bridge and its hydrophobic isostere, cancelling out in the evaluation of $\Delta G_{solv}^{unbound}$ and ΔG_{solv}^{bound}.

4.2.2 Explicit solvent simulations

Explicit solvent calculations were performed using the thermodynamic integration (TI) approach with the TIP3P[17] explicit solvent, as implemented in the GROMACS 4.0.4 software package.[22] The desolvation penalty of each salt bridge upon protein binding relative to its hydrophobic isostere was calculated as the difference of the unbound and bound solvation energies ($\Delta\Delta G_{solv} = \Delta G_{solv}^{unbound} - \Delta G_{solv}^{bound}$, Figure 4-1). In order to calculate the solvation free energies for the unbound and bound states, first the overall free energies of each salt bridge relative to its hydrophobic isostere were computed in its unbound and bound states. Then the solvation free energies were obtained by subtracting all nonbonded protein-protein interactions from the overall free energies.

At each of the five salt concentrations, separate MD simulations of the proteins (unbound and bound states) were performed for each salt bridge using eight λ values, leading to 1120 λ simulations. Each λ simulation was performed in the NVT ensemble (at 25 ^{o}C) for 5 ns and took on average one week to finish on an 8-core 2.6 GHz Intel Xeon node. The partial charges of the side chains of the salt bridge were linearly turned off from λ = 0 (wild-type) to λ = 1 (hydrophobic isostere). To evaluate the convergence of the simulations, three criteria were applied: (1) a small uncertainty of each λ simulation (<5%), (2) linearity of $\langle \partial H(\lambda)/\partial\lambda \rangle_{\lambda}$ vs. λ (R^2 > 0.9), and (3) plot of the solvation free energies vs. 1-ns blocks of the simulations should reach a plateau (Figure 4-2). For simulations that met all three criteria, the 4 ns to 5 ns interval of each λ simulation was used in the final free energy analysis. Uncertainties in the free energies are derived from sampling errors in $\langle \partial H(\lambda)/\partial\lambda \rangle_{\lambda}$. Errors at each λ value were estimated using block averaging.[23]

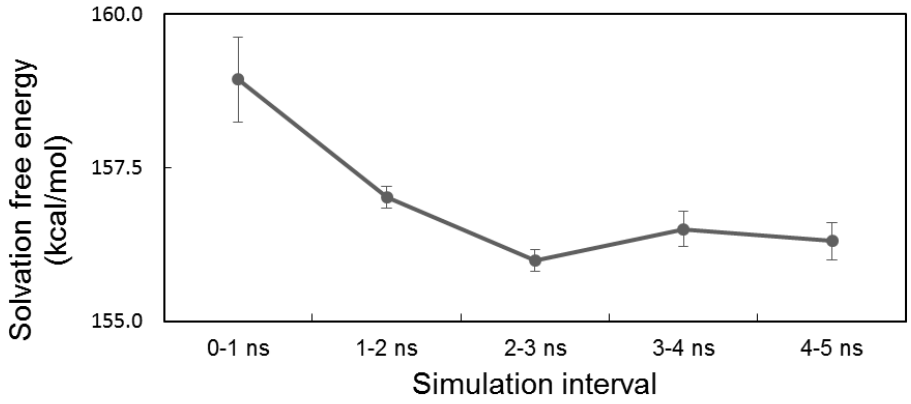

Figure 4-2. Example of assessing convergence by plotting the solvation energy vs. 1 ns blocks of the simulations. The data was taken from the unbound state of the D33-K84 salt bridge in Raf1-Rap1A complex in 2.0 M NaCl solution. Error bars were calculated as described in Methods.

Prior to each λ simulation, the solvent was equilibrated in two stages: 10 ps at constant temperature (25 ^{o}C) and volume, followed by 2 ns at constant temperature (25 ^{o}C) and constant pressure (1 atm). All simulations were performed using the Langevin thermostat (frictional coefficient of 1 ps^{-1}) and a 2 fs time step. Proteins were kept rigid throughout the simulations using the GROMACS "frozen" option, setting velocities of all protein atoms to 0. Real space electrostatic interactions were truncated at 10 Å and the PME[24] method was used to calculate the long-range components. Van der Waals interactions were switched off smoothly between 8 and 9 Å. For the Na^{+} and Cl^{-} nonbonding parameters, the default OPLS-AA/L forcefield parameters were used.[18,25]

4.3 RESULTS

As might be expected, the presence of ions in the simulations necessitates longer simulations for reaching convergence, especially at higher salt concentrations for the large protein-protein systems. Compared to our previous studies in pure water,[8,14] in order to obtain converged results for the systems with ions, we 1) equilibrated each system in the NPT ensemble for 2 ns (instead of 100 ps), and 2) performed each λ simulation five times longer, for 5 ns (instead of 1 ns) (see Methods). According to our convergence criteria it appears that all of the explicit solvent simulations of the salt bridges, both in solution and in the protein context are converged within 5 ns of the TI simulations.

We focused our comparison on computing the desolvation penalties at each of five salt concentrations of 0.1, 0.3, 0.5, 1.0 and 2.0 M *relative* to zero salt. We first describe the results for the salt bridges in the absence of the protein environment, i.e. salt bridges with the same geometries but in solution with the residues capped (with acetyl and N-methyl groups). Then, the results for the salt bridges in the proteins will be discussed.

4.3.1 Salt bridges in the absence of the protein environment

Figure 4-3C shows the correlation between the desolvation penalties estimated by the PB implicit model vs. TIP3P explicit solvent model at five different salt concentrations relative to zero salt. The results of the two models are of similar magnitude (all rmsd's ≤ 0.4 kcal/mol) but there is no significant linear correlation between the explicit and implicit results. The latter appears to be due to the fact that the implicit solvent, at each salt concentration, predicts similar

desolvation penalties independent of salt bridge type or configuration (i.e. horizontal correlation line).

A closer examination of the plots in Figure 4-3C reveals that the rmsd between the two solvent models, while low overall, increases as the salt concentration increases. This is especially more obvious for the solvation free energies in the unbound and bound states (Figure 4-3A, B, Tables 4-1 and 4-2, Supporting Information). As previously mentioned, implicit solvent models ignore ion-ion correlations due to the implicit representation of the salt. Such correlations are naturally expected to occur more at the higher salt concentrations and can be one potential explanation for increasing discrepancies between the two models as the salt concentration increases.

In a previous study of salt bridge analogues (acetate and methyl ammonium) in salt solutions,[9] it had been found that in the implicit solvent calculations setting all the regions of space that are available to the solvent also accessible to the salt (i.e., no Stern (ion-exclusion) la-

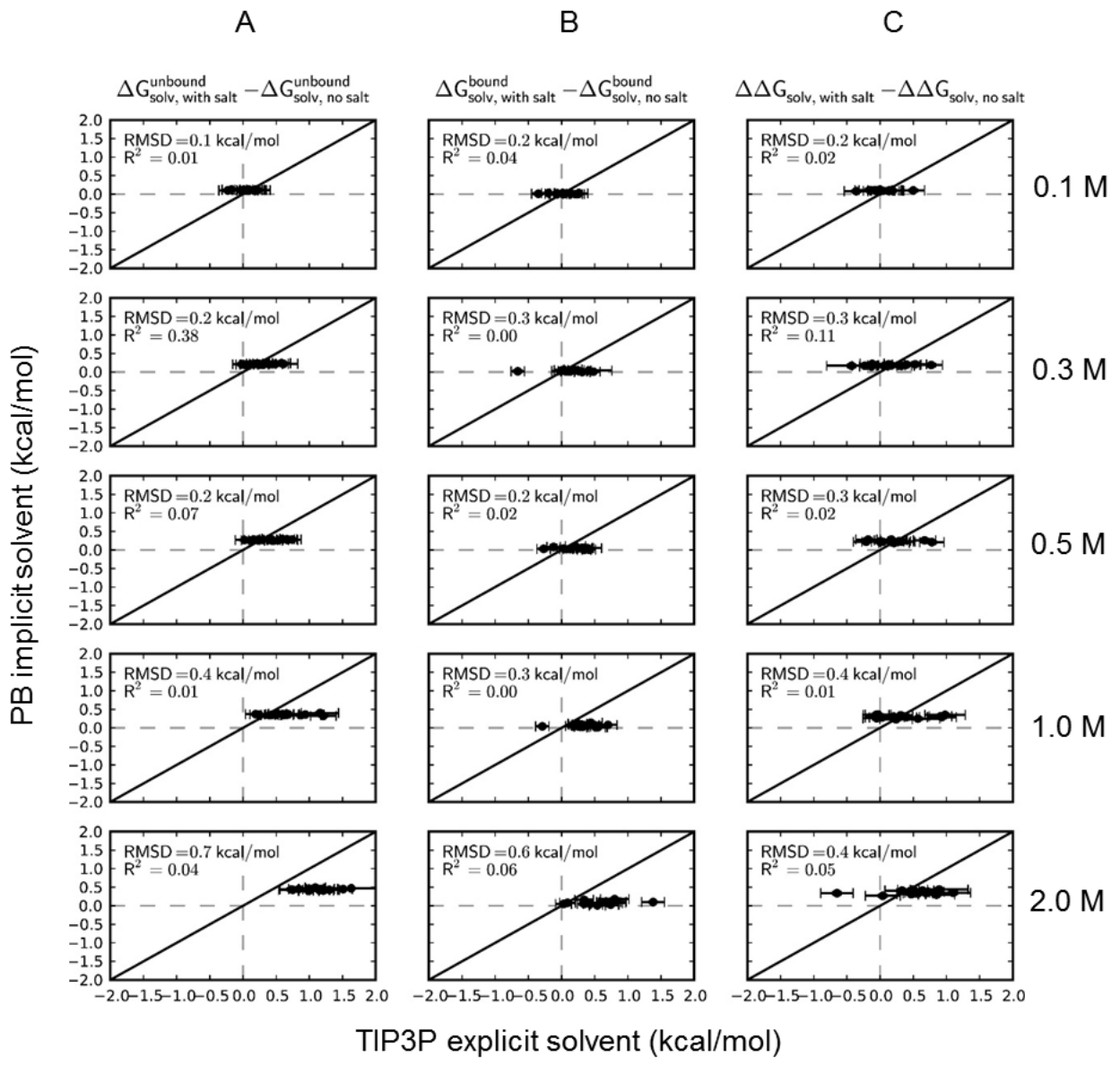

Figure 4-3. Comparison of the PB implicit and the TIP3P explicit solvent models for computing the solvation thermodynamics of salt bridges (in reference to their hydrophobic isosteres) at five different concentrations of salt relative to zero salt in the absence of the protein environment: change in the solvation free energies in the unbound state (A), bound state (B) and change in the desolvation penalties (C). Each row represents one salt concentration. The rmsd and r^2 values are provided for each plot. The diagonal line represents perfect agreement and the dashed lines represent *x=0* and *y=0* axes.

-yer) improves the agreement between implicit and explicit solvent results. To test this, we performed another set of the PB implicit solvent calculations where the ion-exclusion radius was set to zero. As shown in Figure 4-7 in Supporting Information, this only slightly improves the implicit-explicit solvent agreement, especially for the solvation free energies in the unbound and bound states at higher salt concentrations (Figure 4-7A, B, Supporting Information).

4.3.2 Salt bridges in the context of the proteins

Figure 4-4C shows the correlation between the desolvation penalties of the salt bridges (in the context of proteins) estimated by the PB implicit model vs. TIP3P explicit solvent model at five different salt concentrations relative to zero salt. Compared to the salt bridges in solution (Figure 4-3C), the range of the desolvation penalties is much higher but a similar trend is observed: while there is no significant linear correlation, the results of implicit and explicit solvents are of similar magnitude at lower salt concentration but the discrepancies between the two models increase as the salt concentration increases. In addition, setting the ion-exclusion radius to zero in the implicit solvent calculations has a minor effect in improving the results (Figure 4-8, Supporting Information). In contrast to the salt bridges in solution, however, there are several outliers in the results particularly at higher salt concentrations

At the higher salt concentrations (i.e., 1.0 and 2.0 M) the agreement between the results of the two solvent models is poor as reflected in the large rmsd values. As shown in Figure 4-4C, this is mainly due to the presence of multiple outliers among the desolvation penalties. Interestingly, it appears that the source of these outliers can be either from the unbound or bound states of the proteins (Figure 4-4A, B). In particular, the highest implicit-explicit discrepancies in either the unbound or bound states at 1.0 and 2.0 M salt are due to three salt bridges: D38-R89 in

the unbound state of the Raf1-Rap1A, R83-D39 in both the unbound and bound stats of barnase-barstar and, R87-D39 in the unbound state of barnase-barstar (Figure 4-4A, B and Tables 4-3 and 4-4, Supporting Information). These salt bridges have been labeled in Figure 4-4 using numbers 1, 2 and 3 respectively. If these three salt bridges are not included in the correlation plots, the implicit-explicit rmsd values for the desolvation penalties at higher salt concentrations reduce significantly (Figure 4-9, Supporting Information). One source of such discrepancies can be specific ion effects, i.e. ion binding. Such effects are only present in the explicit solvent simulations and cannot be captured by the treatment of ions in the implicit solvent calculations.[5] Therefore we examined the presence of bound ions in the vicinity of the residues of these three salt bridges.

At 1.0 M salt, the salt bridges D38-R89 and R83-D39 are the two outliers in the correlation plots of the unbound and bound states (labeled as "1" and "2" respectively in Figure 4-4 A, B). For the D38-R89 salt bridge in the unbound state of Raf1-Rap1A (Figure 4-4, labeled as "1"), there is a very large negative change in the unbound solvation free energy (relative to zero salt). Examining the corresponding explicit solvent trajectories reveals that there is a bound Na^+ ion ~ 5.5 Å from the residue R89 throughout the simulations (Figure 4-5). Since both Na^+ and R89 are positively charged, it can explain the large negative change in the unbound solvation free energy (relative to zero salt) when the residues are mutated to their hydrophobic isosteres. For the other outlier from the 1.0 M salt simulations, the R83-D39 salt bridge in the bound state of barnase-barstar (Figure 4-4, labeled as "2"), there is a large positive change in the bound solvation free energy (relative to zero salt). As shown is Figure 4-6, there is a buried Cl^- ion

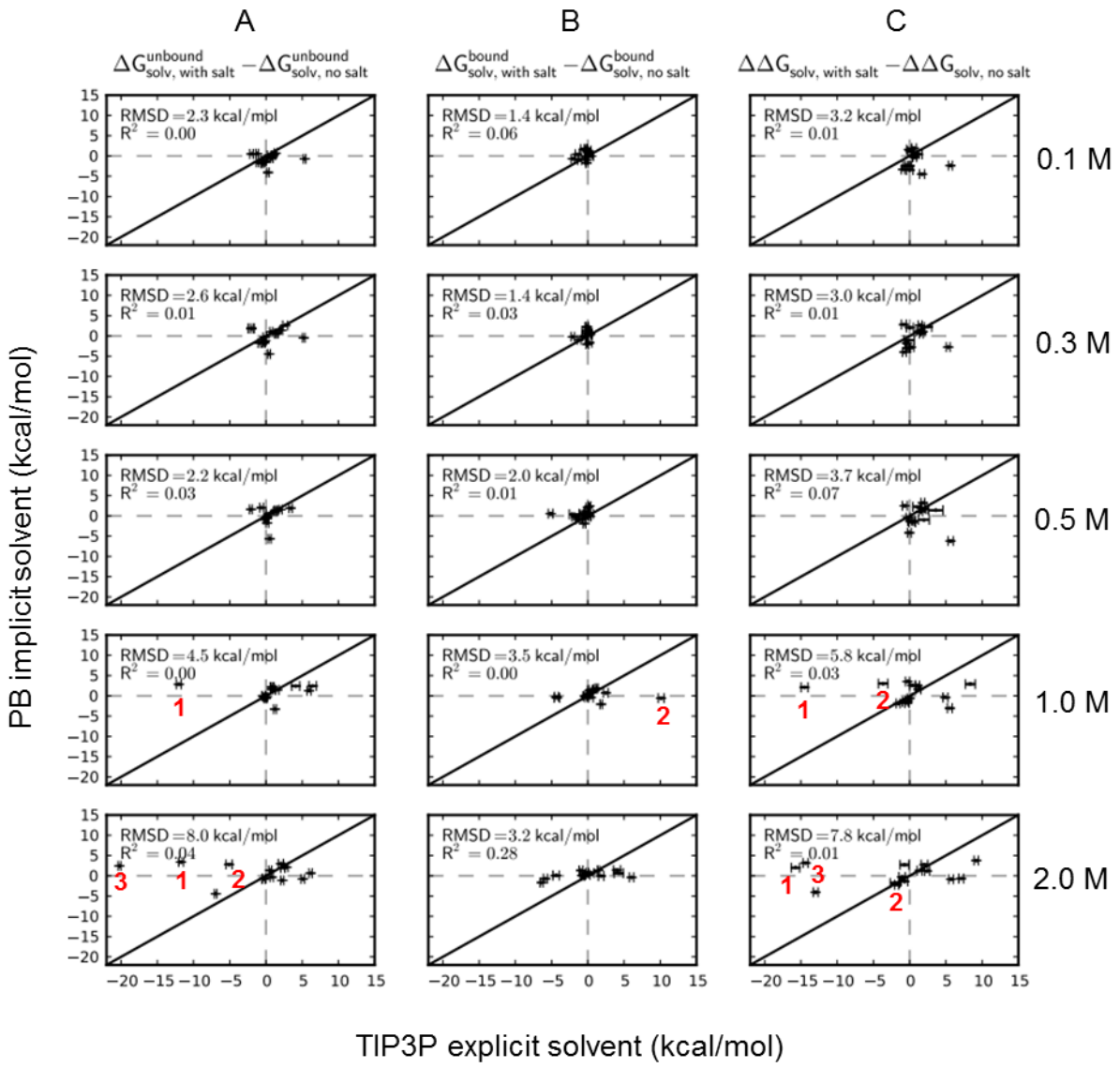

Figure 4-4. Comparison of the PB implicit and the TIP3P explicit solvent models for computing the solvation thermodynamics of salt bridges (in reference to their hydrophobic isosteres) at five different concentrations of salt relative to zero salt in the context of the proteins: change in the solvation free energies in the unbound state (A), bound state (B) and change in the desolvation penalties (C). Each row represents one salt concentration. The rmsd and r^2 values are provided for each plot. The outlier salt bridges discussed in the text have been labeled: 1) D38-R89 in the Raf1-Rap1A complex, 2) R83-D39 in the barnase-barstar complex, and, 3) R87-D39 in the barnase-barstar complex. The diagonal line represents perfect agreement and the dashed lines represent *x=0* and *y=0* axes.

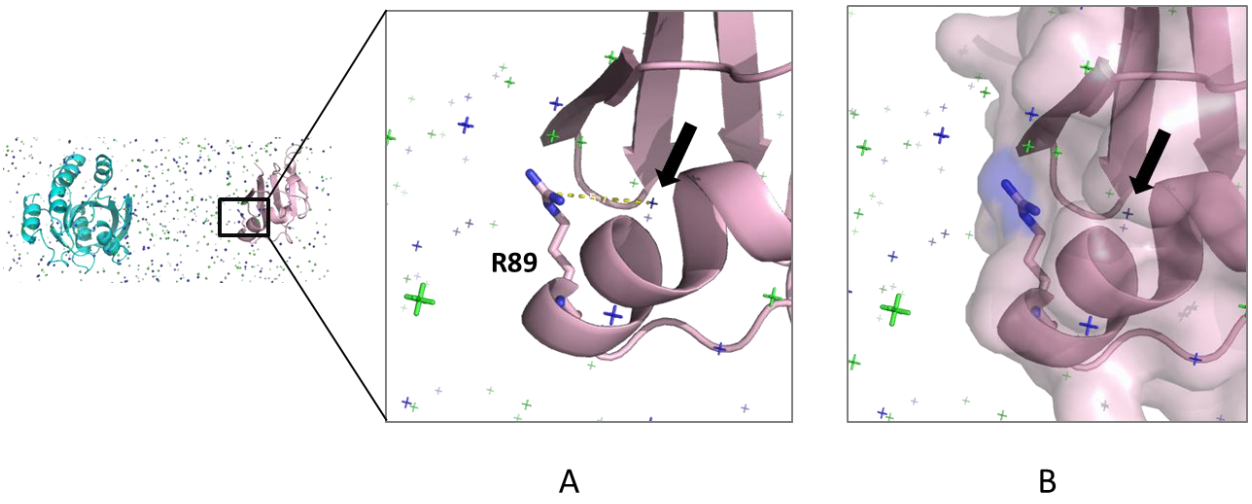

Figure 4-5. Ions in the vicinity of the residue R89 in the unbound state of the Raf1-Rap1A complex at 1.0 M salt concentration. A) The arrow points to a buried Na^+ ion that is ~5.5 Å from the R89 residue throughout the simulation. B) Same as A, with the addition of the surface representation to the proteins, which shows that the ion is buried. Proteins and salt bridges are represented in cartoon and stick respectively and blue and green dots correspond to Na^+ and Cl^- ions. Explicit solvent molecules were not included in the pictures.

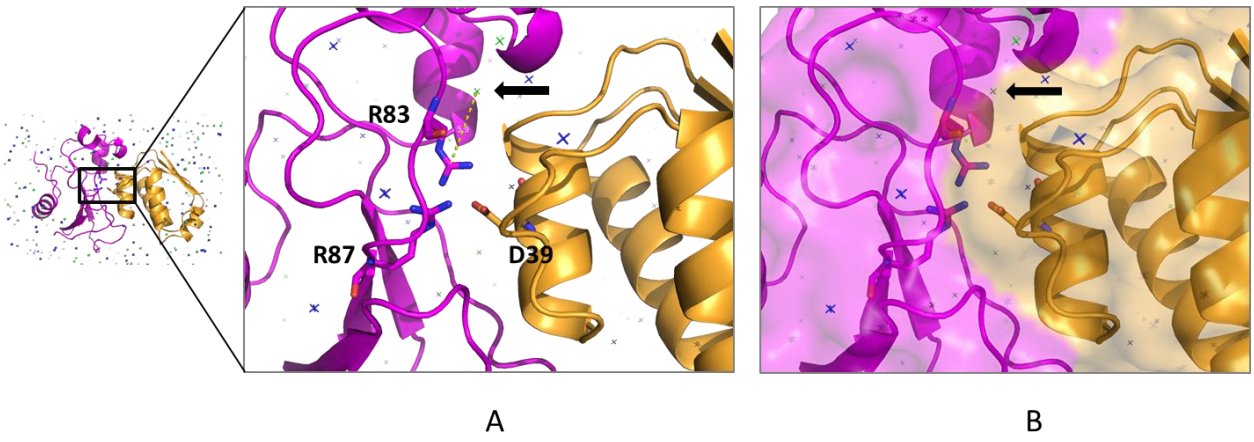

Figure 4-6. Ions in the vicinity of the salt bridge R83-D39 in the bound state of the barnase-barstar complex at 1.0 M salt concentration. A) The arrow points to a buried Cl^- ion that is ~5.8 Å from the R83 residue throughout the simulation. B) Same as A, with the addition of the surface representation to the proteins, which shows that the ion is buried. Proteins and salt bridges are represented in cartoon and stick respectively and blue and green dots correspond to Na^+ and Cl^- ions. Explicit solvent molecules were not included in the pictures.

located ~ 5.8 Å from the R83 residue. The loss of favorable interaction between the R83 residue and Cl^- ion upon mutation of the salt bridge to its hydrophobic isostere is consistent with the large positive change in the bound solvation free energy. Similarly, for all the outliers from the 2.0 M salt simulations, there is a bound ion within 7 Å of one of the salt bridge residues that can explain the large explicit-implicit deviations. In particular, for the R87-D39 salt bridge in the unbound state of barnase-barstar, which has the largest negative change in the unbound solvation free energy (Figure 4-4, labeled as "3"), the bound Na^+ ion is very close to the R87 residue (~ 4 Å).

Therefore our findings indicate that the presence of the bound ions in the vicinity of salt bridges can explain the large implicit-explicit discrepancies observed at higher salt concentrations upon mutation of the salt bridge residues to their corresponding hydrophobic isosteres. This can be further verified by rerunning the outlier simulations in explicit solvent after removing the bound ions and placing them in the bulk solution. Implicit solvent models inherently cannot capture specific ion effects (i.e. ion binding) as the ions are not treated explicitly. Our results show that such specific ion effects can be significant and can cause significant deviation of the implicit solvent calculations from explicit solvent results.

4.4 CONCLUSION

In closing, we performed a direct comparison between the implicit and explicit solvent models in computing the effects of five different NaCl concentrations on the desolvation penalties of salt bridges across several protein-protein interfaces. The results of the two solvent models are of similar magnitude, particularly at lower salt concentrations, but no significant

linear correlation was observed. The results also show that substantial discrepancies can arise from specific ion effects, such as bound ions in the vicinity of salt bridges. Therefore for the systems with known bound ions (e.g. coordinated or buried ions in the crystal structure), the implicit solvent calculations might not capture the energetics of the system properly. In performing fast implicit solvent calculations for such cases, a combination of the implicit treatment of ions in the bulk solution in conjunction with explicit presentation of the bound ions might be more accurate.[26,27]

4.5 SUPPORTING INFORMATION

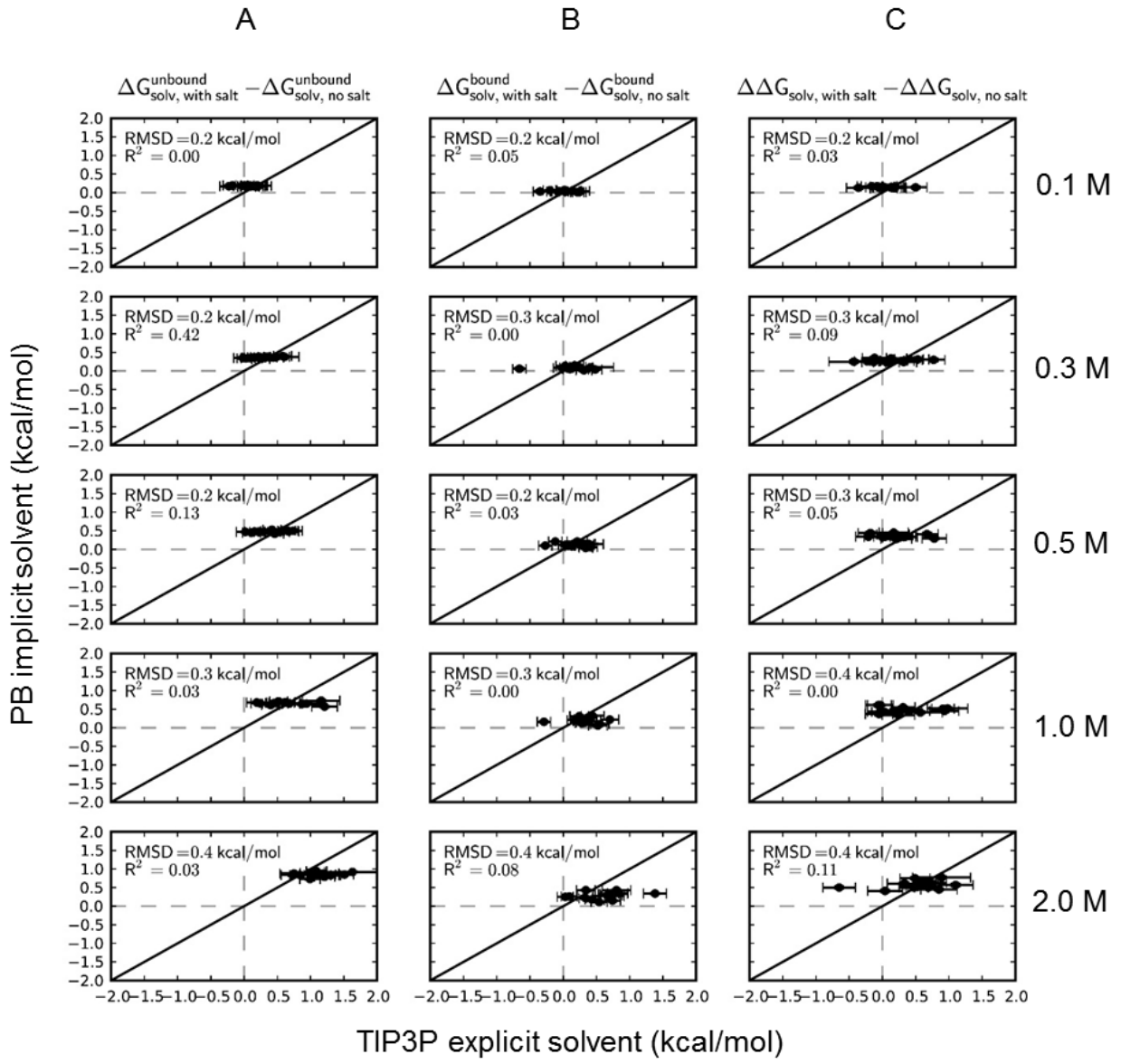

Figure 4-7. Effect of setting the thickness of the Stern layer to zero on the solvation free energies of salt bridges (in reference to their hydrophobic isosteres) in the absence of the protein environment. This plot shows the comparison of solvation thermodynamics computed by the PB implicit and the TIP3P explicit solvent models at five different concentrations of salt relative to zero salt in the absence of the protein environment: change in the solvation free energies in the unbound state (A), bound state (B) and change in the desolvation penalties (C). The rmsd and r^2 values are provided for each plot. The diagonal line represents perfect agreement.

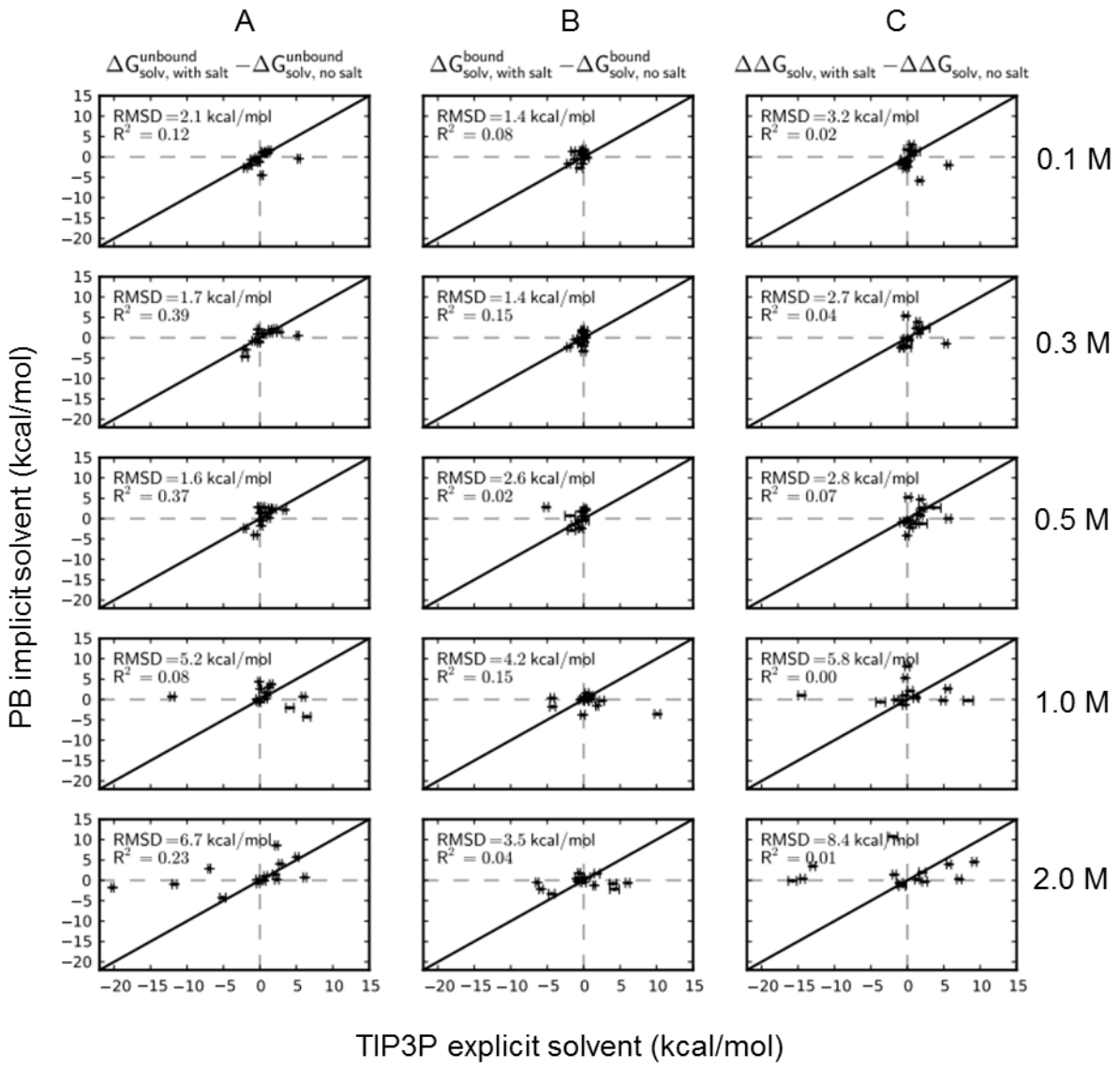

Figure 4-8. Effect of setting the thickness of the Stern layer to zero on the solvation free energies of salt bridges (in reference to their hydrophobic isosteres) in the context of the proteins. This plot shows the comparison of the solvation thermodynamics computed by the PB implicit and the TIP3P explicit solvent models at five different concentrations of salt relative to zero salt in the context of the proteins: change in the solvation free energies in the unbound state (A), bound state (B) and change in the desolvation penalties (C). The rmsd and r^2 values are provided for each plot. The diagonal line represents perfect agreement.

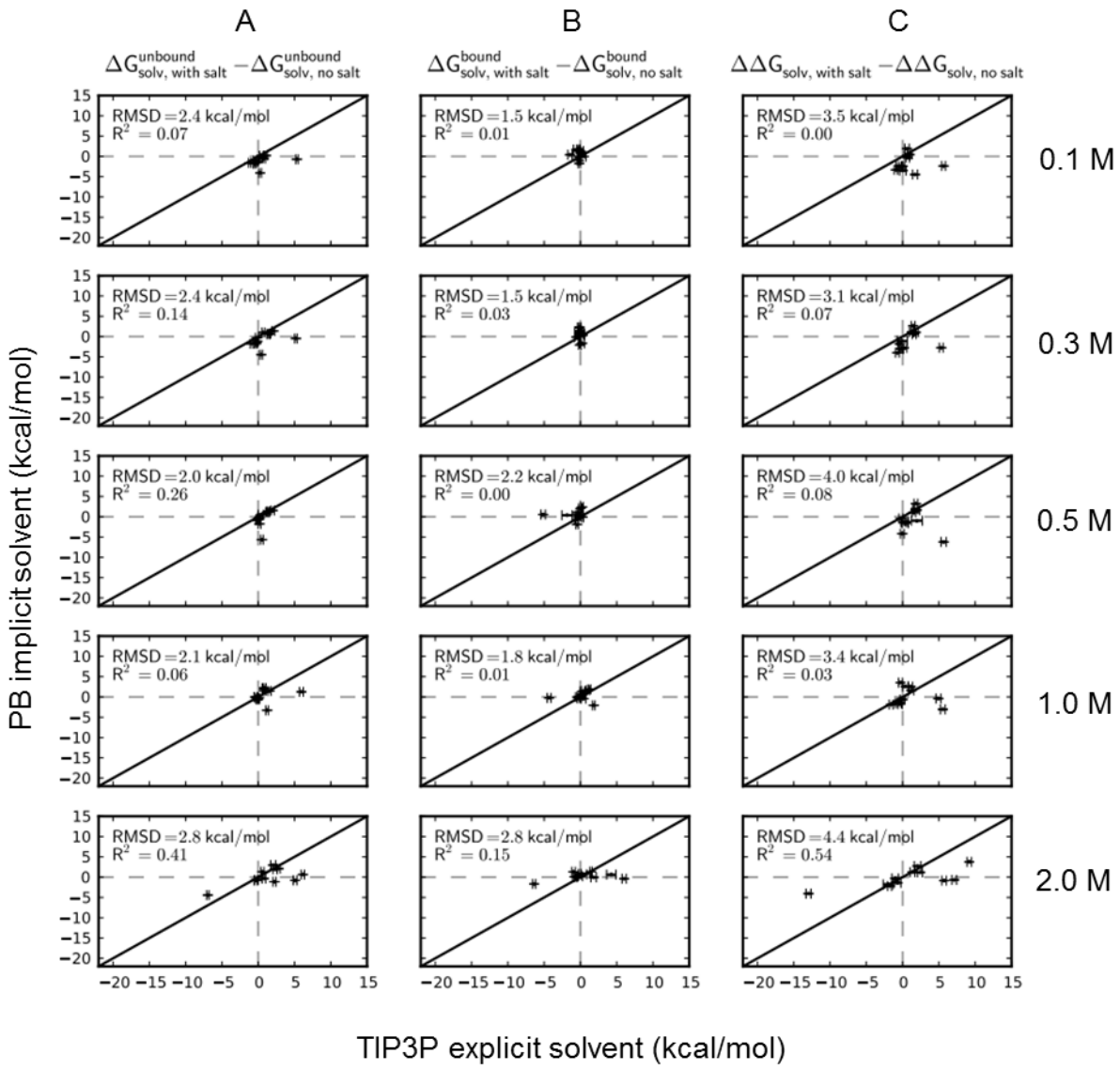

Figure 4-9. Comparison of the solvation thermodynamics of salt bridge (in reference to their hydrophobic isosteres) computed by the PB implicit and the TIP3P explicit solvent models at five different concentrations of salt relative to zero salt in the context of the proteins, when the three outliers discussed in the text are not included: change in the solvation free energies in the unbound state (A), bound state (B) and change in the desolvation penalties (C). The rmsd and r^2 values are provided for each plot. The diagonal line represents perfect agreement.

Table 4-1. Computed desolvation penalties ($\Delta\Delta G_{solv,\ with\ salt} - \Delta\Delta G_{solv,no\ salt}$) of salt bridges upon association in the absence of the protein environment at different salt concentrations relative to no salt using the explicit solvent simulations. Also the effect on the solvation energies of the salt bridges in the unbound ($\Delta G^{unbound}_{solv,\ with\ salt} - \Delta\Delta G^{unbound}_{solv,no\ salt}$) and bound ($\Delta G^{bound}_{solv,\ with\ salt} - \Delta\Delta G^{bound}_{solv,no\ salt}$) states are shown.

		Barnase-barstar			***Raf1-Rap1A***				***Growth hormone - receptor***					***Neuraminidase-Ab***	
		R59-E76	R83-D39	R87-D39	D33-K84	E37-R59	D38-R89	E54-R67	K41-E127	R64-E44	R64-D164	R167-E127	D171-R43	K432-D97	K463-E56
Conc. (M)															
$\Delta G^{unbound}_{solv,with\ salt} - \Delta G^{unbound}_{solv,no\ salt}$	*0.1*	0.2	0.1	0.0	0.2	0.1	-0.2	0.2	0.1	0.2	0.1	-0.2	0.0	0.0	0.1
	0.3	0.1	0.2	0.3	0.5	0.1	0.4	0.2	0.6	0.3	0.0	0.0	0.4	0.6	0.2
	0.5	0.5	0.2	0.5	0.7	0.0	0.7	0.4	0.6	0.7	0.1	0.4	0.3	0.4	0.2
	1.0	0.9	0.4	1.2	0.6	0.6	0.2	0.7	0.5	0.2	0.4	0.9	0.7	1.2	0.5
	2.0	1.2	1.2	1.0	1.0	0.7	1.2	0.8	1.6	1.5	1.2	0.7	1.3	1.1	1.0
$\Delta G^{unbound}_{solv,with\ salt} - \Delta G^{unbound}_{solv,no\ salt}$	*0.1*	0.1	-0.1	-0.1	0.0	-0.3	-0.2	0.0	0.2	0.3	-0.1	0.1	-0.1	0.0	0.2
	0.3	0.2	0.0	0.1	0.1	-0.7	0.0	0.2	0.5	0.2	0.3	0.4	0.2	0.0	0.3
	0.5	0.3	0.0	0.1	0.1	-0.3	-0.1	0.2	0.4	0.4	0.3	0.4	0.3	0.3	0.3
	1.0	0.3	0.2	0.3	0.3	-0.3	0.2	0.4	0.6	0.3	0.4	0.7	0.3	0.2	0.5
	2.0	0.7	0.1	0.4	0.3	0.0	0.3	0.8	0.7	0.8	0.7	1.4	0.7	0.8	0.5
$\Delta\Delta G_{solv,with\ salt} - \Delta\Delta G_{solv,no\ salt}$	*0.1*	0.0	0.2	0.1	0.2	0.5	0.0	0.2	-0.2	-0.1	0.1	-0.4	0.0	0.0	-0.1
	0.3	-0.1	0.2	0.2	0.4	0.8	0.3	0.1	0.1	0.1	-0.2	-0.4	0.3	0.5	-0.1
	0.5	0.2	0.2	0.3	0.7	0.3	0.8	0.2	0.2	0.3	-0.2	0.0	0.0	0.2	-0.2
	1.0	0.6	0.2	0.9	0.3	0.9	0.0	0.2	0.0	0.0	0.0	0.2	0.4	1.0	-0.1
	2.0	0.5	1.1	0.6	0.6	0.7	0.8	0.0	0.9	0.7	0.5	-0.7	0.6	0.3	0.5

Table 4-2. Computed desolvation penalties ($\Delta\Delta G_{solv,\ with\ salt} - \Delta\Delta G_{solv,no\ salt}$) of salt bridges upon association in the absence of the protein environment at different salt concentrations relative to no salt using the implicit solvent calculations. Also the effect on the solvation energies of the salt bridges in the unbound ($\Delta G^{unbound}_{solv,\ with\ salt} - \Delta\Delta G^{unbound}_{solv,no\ salt}$) and bound ($\Delta G^{bound}_{solv,\ with\ salt} - \Delta\Delta G^{bound}_{solv,no\ salt}$) states are shown.

		Barnase-barstar			***Raf1-Rap1A***				***Growth hormone - receptor***					***Neuraminidase-Ab***	
		R59-E76	R83-D39	R87-D39	D33-K84	E37-R59	D38-R89	E54-R67	K41-E127	R64-E44	R64-D164	R167-E127	D171-R43	K432-D97	K463-E56
Conc. (M)															
$\Delta G^{unbound}_{solv,with\ salt} - \Delta G^{unbound}_{solv,no\ salt}$	*0.1*	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	0.5	0.3	0.3	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	1.0	0.3	0.4	0.3	0.4	0.3	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.4	0.4
	2.0	0.4	0.4	0.4	0.5	0.4	0.5	0.5	0.5	0.4	0.4	0.4	0.4	0.5	0.5
$\Delta G^{unbound}_{solv,with\ salt} - \Delta G^{unbound}_{solv,no\ salt}$	*0.1*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.3	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.5	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0
	1.0	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.0
	2.0	0.1	0.1	0.0	0.0	0.1	0.2	0.2	0.0	0.1	0.1	0.1	0.1	0.1	0.0
$\Delta\Delta G_{solv,with\ salt} - \Delta\Delta G_{solv,no\ salt}$	*0.1*	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	0.5	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.3	0.3
	1.0	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.4	0.3	0.3	0.3	0.3	0.3	0.4
	2.0	0.3	0.3	0.4	0.4	0.4	0.3	0.3	0.4	0.3	0.3	0.3	0.3	0.4	0.4

Table 4-3. Computed desolvation penalties ($\Delta\Delta G_{solv,\ with\ salt} - \Delta\Delta G_{solv,no\ salt}$) of salt bridges upon association in the context of the protein at different salt concentrations relative to no salt using the explicit solvent simulations. Also the effect on the solvation energies of the salt bridges in the unbound ($\Delta G^{unbound}_{solv,\ with\ salt} - \Delta\Delta G^{unbound}_{solv,no\ salt}$) and bound ($\Delta G^{bound}_{solv,\ with\ salt} - \Delta\Delta G^{bound}_{solv,no\ salt}$) states are shown.

		Barnase-barstar			***Raf1-Rap1A***				***Growth hormone - receptor***					***Neuraminidase-Ab***	
		R59-E76	R83-D39	R87-D39	D33-K84	E37-R59	D38-R89	E54-R67	K41-E127	R64-E44	R64-D164	R167-E127	D171-R43	K432-D97	K463-E56
	Conc. (M)														
$\Delta G^{unbound}_{solv,with\ salt} - \Delta G^{unbound}_{solv,no\ salt}$	*0.1*	0.4	-2.0	-1.2	1.1	0.5	1.3	0.8	-0.3	5.3	-1.2	-0.5	-0.4	0.3	-0.1
	0.3	1.4	-2.0	-2.0	2.1	1.5	2.7	0.7	-0.3	5.2	-0.9	-0.3	-0.2	0.4	-0.1
	0.5	1.4	-0.6	-2.1	2.0	1.4	3.4	1.2	0.0	0.4	0.2	-0.2	0.1	0.5	0.2
	1.0	1.6	6.4	4.1	0.8	0.9	-12.0	0.8	0.1	5.9	-0.3	-0.1	-0.3	1.2	-0.1
	2.0	2.8	-5.1	-20.2	2.1	2.3	-11.7	0.6	5.1	6.2	-0.3	2.2	0.8	-6.9	-0.3
$\Delta G^{unbound}_{solv,with\ salt} - \Delta G^{unbound}_{solv,no\ salt}$	*0.1*	-0.2	-2.1	-1.1	0.5	-0.2	0.1	-0.2	0.1	-0.3	-0.2	-0.5	0.0	-1.4	-0.1
	0.3	-0.1	-2.1	-1.2	0.3	-0.1	0.1	-0.6	-0.1	-0.1	-0.2	0.0	-0.1	0.2	0.3
	0.5	-0.5	-1.7	-1.4	0.1	-0.1	-0.2	-0.5	0.3	0.0	-0.3	-0.3	-1.8	-5.1	0.2
	1.0	1.8	10.1	-4.3	0.4	-0.2	2.5	-0.4	0.4	1.0	1.2	-0.1	0.3	-4.3	0.3
	2.0	-6.4	-4.4	-5.9	0.0	-0.1	4.0	-0.8	-0.6	-1.0	1.5	4.2	1.8	6.0	0.5
$\Delta\Delta G_{solv,with\ salt} - \Delta\Delta G_{solv,no\ salt}$	*0.1*	0.6	0.1	-0.1	0.6	0.7	1.2	1.0	-0.3	5.6	-0.9	0.0	-0.4	1.7	0.0
	0.3	1.5	0.0	-0.8	1.8	1.6	2.5	1.3	-0.2	5.3	-0.7	-0.2	-0.1	0.2	-0.4
	0.5	1.8	1.0	-0.7	2.0	1.5	3.6	1.8	-0.2	0.3	0.5	0.1	2.0	5.7	-0.1
	1.0	-0.3	-3.7	8.4	0.4	1.1	-14.5	1.2	-0.4	4.9	-1.6	0.0	-0.6	5.5	-0.4
	2.0	9.2	-0.7	-14.4	2.1	2.4	-15.7	1.4	5.7	7.1	-1.8	-2.0	-1.0	-13.0	-0.7

Table 4-4. Computed desolvation penalties ($\Delta\Delta G_{solv,\ with\ salt} - \Delta\Delta G_{solv,no\ salt}$) of salt bridges upon association in the context of the protein at different salt concentrations relative to no salt using the implicit solvent simulations. Also the effect on the solvation energies of the salt bridges in the unbound ($\Delta G^{unbound}_{solv,\ with\ salt} - \Delta\Delta G^{unbound}_{solv,no\ salt}$) and bound ($\Delta G^{bound}_{solv,\ with\ salt} - \Delta\Delta G^{bound}_{solv,no\ salt}$) states are shown.

		Barnase-barstar			***Raf1-Rap1A***				***Growth hormone - receptor***					***Neuraminidase-Ab***	
		R59-E76	R83-D39	R87-D39	D33-K84	E37-R59	D38-R89	E54-R67	K41-E127	R64-E44	R64-D164	R167-E127	D171-R43	K432-D97	K463-E56
	Conc. (M)														
$\Delta G^{unbound}_{solv,with\ salt} - \Delta G^{unbound}_{solv,no\ salt}$	*0.1*	0.2	0.5	0.6	0.2	-0.4	0.7	-0.5	-1.6	-0.7	-1.6	-1.8	-1.0	-4.1	-1.2
	0.3	0.6	1.8	1.8	1.4	0.7	2.6	1.0	-1.7	-0.5	-1.8	-1.6	-0.3	-4.5	-1.2
	0.5	1.4	2.0	1.6	1.5	1.1	1.9	1.1	-0.5	0.5	-0.2	-0.9	-0.6	-5.6	-1.7
	1.0	1.5	2.4	2.4	2.2	2.2	2.8	1.6	-0.3	1.3	0.0	-0.4	-0.1	-3.3	-0.7
	2.0	2.1	2.8	2.4	3.0	2.0	3.4	1.3	-0.8	0.6	-0.8	-1.1	-0.3	-4.4	-0.7
$\Delta G^{unbound}_{solv,with\ salt} - \Delta G^{unbound}_{solv,no\ salt}$	*0.1*	-1.7	-0.7	-1.0	0.0	-0.1	0.3	-1.0	0.9	1.7	1.7	1.7	1.3	0.4	1.2
	0.3	-2.0	-0.2	-1.0	0.3	0.1	0.4	-0.1	1.3	2.3	2.2	1.4	0.8	-1.7	0.6
	0.5	-1.9	-0.2	-0.9	-0.1	0.1	0.5	-0.7	-0.1	2.1	1.0	0.4	0.4	0.5	2.4
	1.0	-2.1	-0.6	-0.5	-0.4	-0.4	0.7	-0.1	1.5	1.7	1.9	0.2	1.3	-0.2	0.7
	2.0	-1.7	0.1	-0.7	0.2	0.8	1.4	0.1	0.1	1.3	1.4	0.6	0.0	-0.4	0.7
$\Delta\Delta G_{solv,with\ salt} - \Delta\Delta G_{solv,no\ salt}$	*0.1*	1.9	1.2	1.6	0.2	-0.3	0.4	0.4	-2.5	-2.4	-3.3	-3.5	-2.4	-4.5	-2.4
	0.3	2.6	2.1	2.8	1.1	0.6	2.2	1.2	-3.0	-2.7	-4.0	-3.0	-1.1	-2.8	-1.8
	0.5	3.2	2.2	2.5	1.6	1.1	1.4	1.8	-0.4	-1.6	-1.2	-1.3	-1.0	-6.2	-4.2
	1.0	3.5	3.0	2.9	2.6	2.5	2.1	1.6	-1.8	-0.4	-1.9	-0.6	-1.5	-3.1	-1.3
	2.0	3.8	2.7	3.2	2.7	1.2	2.0	1.2	-0.8	-0.7	-2.2	-1.8	-0.3	-4.1	-1.4

4.6 REFERENCES

(1) Novotny, J.; Sharp, K. A.: Electrostatic fields in antibodies and antibody/antigen complexes. *Prog. Biophys. Mol. Biol.* **1992**, *58*, 203-224.

(2) Hendsch, Z. S.; Tidor, B.: Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding. *Protein science : a publication of the Protein Society* **1999**, *8*, 1381-92.

(3) Sheinerman, F. B.; Honig, B.: On the role of electrostatic interactions in the design of protein-protein interfaces. *Journal of molecular biology* **2002**, *318*, 161-77.

(4) Baldwin, R. L.: How Hofmeister ion interactions affect protein stability. *Biophys J* **1996**, *71*, 2056-2063.

(5) Hendsch, Z. S.; Tidor, B.: Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein science : a publication of the Protein Society* **1994**, *3*, 211-26.

(6) Bertonati, C.; Honig, B.; Alexov, E.: Poisson-Boltzmann calculations of nonspecific salt effects on protein-protein binding free energies. *Biophys J* **2007**, *92*, 1891-9.

(7) Yu, Z. Y.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A.: First-shell solvation of ion pairs: Correction of systematic errors in implicit solvent models. *J. Phys. Chem. B* **2004**, *108*, 6643-6654.

(8) Salari, R.; Chong, L. T.: Desolvation Costs of Salt Bridges across Protein Binding Interfaces: Similarities and Differences between Implicit and Explicit Solvent Models. *The Journal of Physical Chemistry Letters* **2010**, *1*, 2844-2848.

(9) Thomas, A. S.; Elcock, A. H.: Direct observation of salt effects on molecular interactions through explicit-solvent molecular dynamics simulations: differential effects on electrostatic and hydrophobic interactions and comparisons to Poisson-Boltzmann theory. *Journal of the American Chemical Society* **2006**, *128*, 7796-806.

(10) Huang, B.; Liu, F. F.; Dong, X. Y.; Sun, Y.: Molecular mechanism of the effects of salt and pH on the affinity between protein A and human immunoglobulin G1 revealed by molecular simulations. *The journal of physical chemistry. B* **2012**, *116*, 424-33.

(11) Wagoner, J. A.; Baker, N. A.: Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8331-8336.

(12) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M.: Solvent models for protein-ligand binding: Comparison of implicit solvent poisson and surface generalized born models with explicit solvent simulations. *J. Comput. Chem.* **2001**, *22*, 591-607.

(13) Tan, C.; Yang, L.; Luo, R.: How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. *The journal of physical chemistry. B* **2006**, *110*, 18680-7.

(14) Salari, R.; Chong, L. T.: Effects of high temperature on desolvation costs of salt bridges across protein binding interfaces: similarities and differences between implicit and explicit solvent models. *The journal of physical chemistry. B* **2012**, *116*, 2561-7.

(15) Green, D. F.; Tidor, B.: Design of improved protein inhibitors of HIV-1 cell entry: Optimization of electrostatic interactions at the binding interface. *Proteins* **2005**, *60*, 644-57.

(16) Altman, M. D.; Nalivaika, E. A.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Tidor, B.: Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. *Proteins* **2008**, *70*, 678-94.

(17) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M.: Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(18) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L.: Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474-6487.

(19) Honig, B.; Nicholls, A.: Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144-1149.

(20) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B.: Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128-137.

(21) Gilson, M. K.; Honig, B.: Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* **1988**, *4*, 7-18.

(22) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E.: GROMACS 4: Algorithms for highly efficient, load balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.

(23) Bishop, M.; Frinks, S.: Error analysis in computer simulations. *J. Chem. Phys.* **1987**, *87*, 3675-3676.

(24) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G.: A smooth particle mesh ewald method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.

(25) Aaqvist, J.: Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry* **1990**, *94*, 8021-8024.

(26) Prabhu, N. V.; Panda, M.; Yang, Q.; Sharp, K. A.: Explicit ion, implicit water solvation for molecular dynamics of nucleic acids and highly charged molecules. *J Comput Chem* **2008**, *29*, 1113-30.

(27) Gruziel, M.; Grochowski, P.; Trylska, J., The Poisson-Boltzmann model for tRNA: Assessment of the calculation set-up and ionic concentration cutoff. *J Comput Chem* **2008,** *29* (12), 1970-81.

5.0 DIRECT OBSERVATIONS OF SHIFTS IN THE BETA-SHEET REGISTER OF A PROTEIN-PEPTIDE COMPLEX USING EXPLICIT SOLVENT SIMULATIONS

This work was published as: M. T. Panteva,* R. Salari,* M. Bhattacharjee, L. T. Chong. (2011). "Direct observations of shifts in the beta-sheet register of a protein-peptide complex using explicit solvent simulations." *Biophys. J.* 100(9): L50-52. * *Equal Contribution*

5.1 INTRODUCTION

Intermolecular β-sheets are found in many protein-protein complexes.[1] The formation of the native bound structure is likely to involve encounter complexes[2] with nonnative hydrogen-bonding registers, and these misregistered states may then either dissociate or rearrange to the correct register. Intersheet rearrangements within aggregates of amyloid peptides were previously detected by isotope-edited IR spectroscopy.[3] In several studies, misregistered states were observed in atomistic molecular dynamics (MD) simulations of peptides in either implicit[4,5] or explicit solvent[6]; however, their rearrangements to the native states were not resolved in these studies. To date, investigators have only been able to capture rearrangements using implicit solvent simulations artificially accelerated by low solvent viscosity,[7] explicit solvent simulations

employing high temperature with replica exchange,[8] and Monte Carlo simulations considering only torsional degrees of freedom.[9,10]

In this work, we obtained direct observations of shifts in the register of a β-sheet using all-atom, explicit solvent MD simulations at room temperature. The β-sheet involves the formation of four backbone hydrogen bonds between the forkhead-associated domain of the cancer marker Ki67 (Ki67FHA) and a peptide fragment (residues 260–266) of the hNIFK signaling protein.[11] This is an ideal model system for simulating rearrangements of β-sheets not only because of its small size (106 residues; Figure 5-3, Supporting Information) but also because of its weak binding affinity (K_D = 42 ± 5 mM),[11] which may facilitate the rearrangements. We explored the rearrangements of two potential encounter complexes to the native state. These complexes involve either a "+2" or "−2" register shift in the β-sheet, in which the peptide is displaced by two residues in the direction of the N- and C-termini of the Ki67FHA receptor, respectively (Figure 5-1 and Table 5-1, Supporting Information).

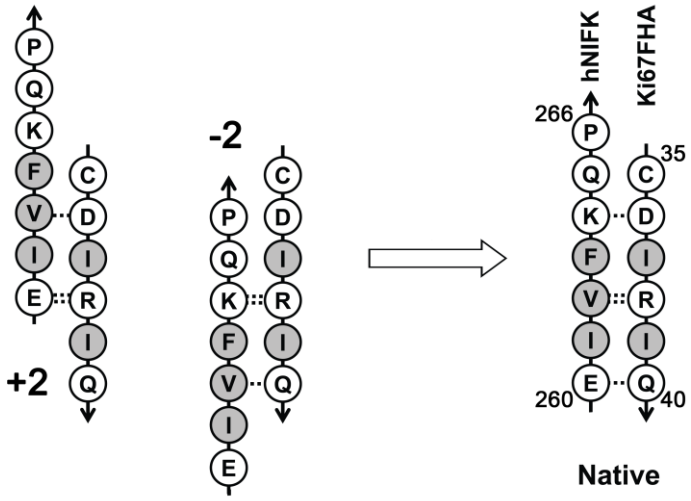

Figure 5-1. Rearrangements of misregistered states (left) to the native state of the β-sheet between the Ki67FHA receptor and hNIFK peptide were explored by simulations. Only two-

residue shifts (+2 and -2) were considered since they maintain alignment of hydrophobic residues (gray) between the strands.

5.2 METHODS

MD simulations. For simulations starting from the native complex between the FHA domain of Ki67 (Ki67FHA) and the heptapeptide fragment of hNIFK (residues 260 to 266), heavy atom coordinates were taken from the NMR solution structure of the FHA domain of Ki67 (residues 3 to 100) in complex with the 44-residue fragment of hNIFK (residues 226 to 269), which is triply phosphorylated at T238, T234, and S230 (PDB code: 2AFF).[1] For simulations starting from the unbound Ki67FHA receptor, heavy atom coordinates were taken from the NMR solution structure of the unbound receptor (PDB code: 1R21).[12] Acetyl and amino capping groups were added to the N- and C-termini, respectively, of the Ki67FHA receptor. Consistent with NMR titration experiments,[1] the hNIFK heptapeptide was modeled with charged termini. Models for the +2 and -2 misregistered states of the Ki67FHA/hNIFK peptide complex were generated by first translating the peptide in the native state towards its N- or C-terminus by two backbone hydrogen bonds in the β-sheet, respectively, using the Swiss-PdbViewer program,[13] and then repacking the side chains using the SCAP program in the Jackal 1.5 software package.[14] After energy minimization (see below), the heavy atom rms deviations of the peptide conformation in the +2 and -2 states from the native state were 2.15 and 1.98 Å, respectively. Hydrogen atoms were added to each model using ionization states present in neutral solution. Each model was solvated in truncated octahedral boxes of TIP3P water[15] with a minimum solute-wall distance of 12 Å. A total of 19 Na^+ and 19 Cl^- counterions were included to yield an ionic strength of 150 mM.

MD simulations of each model were performed in the NPT ensemble (constant number of atoms, pressure, and temperature) using the AMBER 10 software package,[16] ff99SB force field,[17] the TIP3P water model,[15] and ion parameters optimized for use with the TIP3P water model.[18] The temperature was maintained at 300 K and the pressure was maintained at 1 atm using weak Berendsen coupling[19] with a coupling time constant of 10 ps. Van der Waals and short-range electrostatic interactions were truncated at 10 Å; longer range electrostatic interactions were calculated using particle mesh Ewald (PME)[20] and periodic bound conditions. To enable a 2 fs time step, bonds to hydrogen were constrained to their equilibrium values with the SHAKE algorithm.[21]

To relieve unfavorable interactions, each model was subjected to energy minimization followed by a two-stage equilibration with harmonic position restraints applied to the peptide with a force constant of 10 kcal $mol^{-1} \bullet Å^{-2}$. During the first stage, the energy-minimized system was equilibrated for 20 ps at constant temperature (300K) and volume. During the second stage, the system was equilibrated for 2 ns at constant temperature (300K) and pressure (1 atm). After equilibration, more than 300 unrestrained, independent production simulations with different initial velocities (selected from a Maxwell distribution) were performed on the Folding@Home distributed computing resource (http://folding.stanford.edu) for each model at 300K and 1 atm for 20 to 40 ns, accumulating more than 39 μs of aggregate simulation time over a couple of months. A subset of these simulations was then extended to 300 ns on the Ranger supercomputer at the Texas Advanced Computing Center (TACC), with each simulation requiring more than a month of wall clock time, generating ~10 ns/day using 32 cores in parallel. All analysis was performed using conformations sampled every 100 ps.

Targeted MD simulations. The targeted MD approach[22] was used to generate an artificially accelerated rearrangement of the +2 state to the native state. In addition to the standard molecular mechanics potential, this approach applies a harmonic constraining force at each time step to a set of atoms in the current conformation until the target conformation is reached. This additional energy term has the following form: $1/2\ k\ (RMSD - RMSD_{target})^2$ where k is the force constant, RMSD is the mass-weighted RMSD of the current conformation from a reference structure, and $RMSD_{target}$ is the mass-weighted RMSD between a suitable target conformation and a reference structure. In our simulations, a force constant of 2 kcal/mol·Å^2 was exerted on the C^{α} atoms of the hNIFK peptide and the target structure was defined as having a C^{α} RMSD of the intermolecular β-sheet that is less than 0.5 Å from the native state. Targeted MD simulations were performed for 2 ns using a 1 fs time step. All other simulation parameters were identical to those used for the standard MD simulations described above.

Analysis. rms deviations and the formation of hydrogen bonds were determined using the *ptraj* module of the AmberTools 1.0 software package.[23] The percent burial of selected residues upon binding was computed using percent burial = $(SASA_{bound}/SASA_{unbound})$ x 100 where $SASA_{bound}$ and $SASA_{unbound}$ are the solvent accessible surface areas (SASAs) in the bound and unbound states, respectively; the unbound states consisted of the same conformations of the bound state, but without the binding partner. Solvent accessible surface areas were computed using the MSMS program[24] with the solvent probe radius set to 1.4 Å.

5.3 RESULTS

Since rearrangements of these misregistered states are likely rare, but fast events, the probability of capturing these events might be increased by performing a large number of short simulations.[25] Our simulation approach therefore involved first performing a large ensemble (>300) of short (20-40 ns), explicit solvent simulations starting from each of the misregistered states (+2 and -2 states) using the Folding@Home distributed computing network, then extending the simulations that had resulted in partial rearrangements (<2.5 Å C^{α} RMSD of the β-sheet from the native state or at least one native hydrogen bond) to a much longer timescale (300 ns) using the TACC Ranger supercomputer; the remaining simulations were not extended due to the high computational cost (~10 ns/day using 32 cores on Ranger; for simulation details, see Methods in Supporting Information). Among the short Folding@Home simulations, a small percentage resulted in partial rearrangements: 11 starting from +2 state and 10 starting from the -2 state. Only four out of the hundreds of short simulations resulted in dissociation of the peptide, suggesting that the +2 and -2 states are relatively stable and therefore likely to be relevant to the binding pathway.

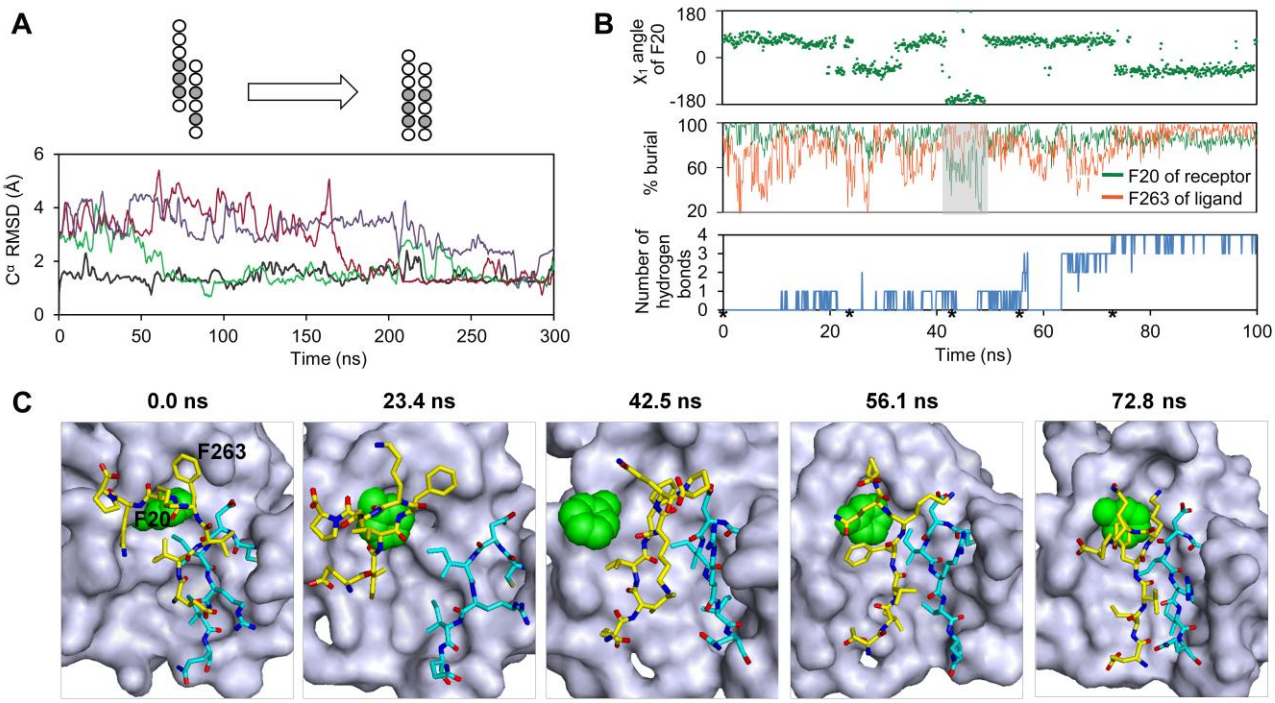

Figure 5-2. Mechanism of rearrangement from the +2 state to the native state. (A) Plot of C^{α} RMSD of the β-sheet from the native state vs. time for simulations starting from the +2 (green, red, purple) and native states (gray); for each simulation, the first 20-40 ns was performed using the Folding@Home network and then extended to 300 ns on the Ranger supercomputer. (B) Plots of the χ1 angle of receptor residue F20, % burials of F20 and peptide residue F263, and number of native hydrogen bonds in the β-sheet vs. time for the most rapid rearrangement. As highlighted by the gray box in the % burial plot, F263 anchors into a transient hydrophobic pocket of the receptor. (C) Snapshots at times indicated by asterisks in (B), tracking the positions of the F20 side chain (green), peptide (yellow), and receptor β-strand (cyan); the rest of the receptor is represented by its molecular surface (gray).

Of the eleven 300-ns simulations starting from the +2 state, three resulted in complete rearrangements to the native state (Figure 5-2A): their C^{α} RMSD's of the intermolecular β-sheet from the native state were within one standard deviation of the average value during a 300-ns simulation starting from the native state (1.5 ± 0.3 Å). Of the remaining eight simulations, seven remained bound in a non-native state and one resulted in unbinding of the peptide. None of the ten 300-ns simulations starting from the -2 state rearranged. In one of these simulations, the

peptide partially dissociates; in the other nine, receptor residue F20 sterically blocks translation of the peptide towards the native register (Movie 5-1).

All three rearrangements from the +2 state share a common mechanism (Movies 5-2 to 5-4). We illustrate this mechanism for the most rapid rearrangement (green in Figure 5-2A) by monitoring the χ_1 angle of receptor residue F20, percent burials of F20 and peptide residue F263, and number of native hydrogen bonds in the β-sheet (Figure 5-2B; plots for the other two rearrangements and the native complex are provided in Figure 5-4 in Supporting Information). As shown by the snapshots in Figure 5-2C, rearrangement begins with partial dissociation of the peptide at 23.4 ns. At 42.5 ns, F20 swings out into solution (χ_1 angle of -160^o to 60^o in upper panel of Figure 5-2B), exposing a hydrophobic pocket (pocket 1) in the receptor. The subsequent anchoring of peptide residue F263 into this transient pocket (middle panel of Figure 5-2B) seems to facilitate the formation of two more native hydrogen bonds at 56.1 ns (lower panel of Figure 5-2B). The hydrogen bonds form sequentially starting from the N- to C-terminal ends of the peptide until all four have formed at 72.8 ns. Since the mechanism in Figure 5-2C involves partial dissociation of the peptide, it is distinct from the reptation-like mechanism[26] that has been observed in implicit solvent simulations of amyloid peptides.[7,9,10] To our knowledge, our simulations provide the first direct views of β-sheet rearrangements that involve anchoring into a transient pocket. Interestingly, use of the targeted MD approach[22] to accelerate rearrangements from the +2 state did not result in the same mechanism (Figure 5-5, Supporting Information).

To investigate the importance of pocket 1 in the rearrangements, we mutated F20 to an alanine thereby leaving the pocket open. Out of ten 100-ns simulations starting from the mutant +2 state, one rearranged to the native state (Figure 5-6A, Movie 5-5, Supporting Information). In this rearrangement, however, F263 does not anchor into pocket 1; instead, it uses another pocket

(pocket 2) that is blocked when F20 swings out to expose pocket 1 (Figure 5-6B, C, Supporting Information). Pocket 2 is also used as an anchor point by F263 during the slowest rearrangement of the wild-type +2 state before F263 settles into pocket 1 (Figure 5-4B, Movie 5-4, Supporting Information). From simulations of the unbound receptor, it appears that the F20 gating of both pockets 1 and 2 is intrinsic to the receptor (Figure 5-4D, Supporting Information).

5.4 CONCLUSION

In closing, we have reported the first direct observations of shifts in the β-sheet register of a protein-peptide complex using explicit solvent MD simulations. In particular, rearrangements of the +2 misregistered state to the correct register of the Ki67FHA-hNIFK peptide complex were captured in three independent simulations. All three rearrangements share a common mechanism: the anchoring of peptide residue F263 into a transient pocket of the receptor facilitates the "crawling" of the peptide along the receptor surface to the native alignment. These rearrangements suggest that MD simulations can correct for errors in the register of nonlocal β-sheets, which may be useful in the area of structure prediction where the prediction of β-sheet registers remains a challenge.[27]

Given that protein binding interfaces are much richer in aromatic residues than the average protein surface,[28] the anchoring of aromatic residues into transient pockets may be a general mechanism of "dynamic" induced-fit binding. This general mechanism might be relevant to various proteins in which alternate registers play important roles in their biological functions.[29-33] Our results also demonstrate that MD simulations can identify transient pockets that might be used to develop new classes of pharmaceuticals, particularly for targets that appear

"undruggable".[34] Finally, although we have simulated only three β-sheet rearrangements, the generation of a large ensemble of these kinds of simulations will become more practical as computational power improves.

5.5 ACKNOWLEDGEMENT

We thank Adrian Elcock (U. Iowa) and Angela Gronenborn for insightful discussions; Seth Horne, Brandon Mills, and Matt Zwier for constructive comments on the manuscript. This research was supported by NSF (CAREER MCB-0845216 and TeraGrid TG-MCB070090). Computational resources were also provided by the Center for Simulation and Modeling and Folding@Home project (NIH R01-GM062868).

5.6 SUPPORTING INFORMATION

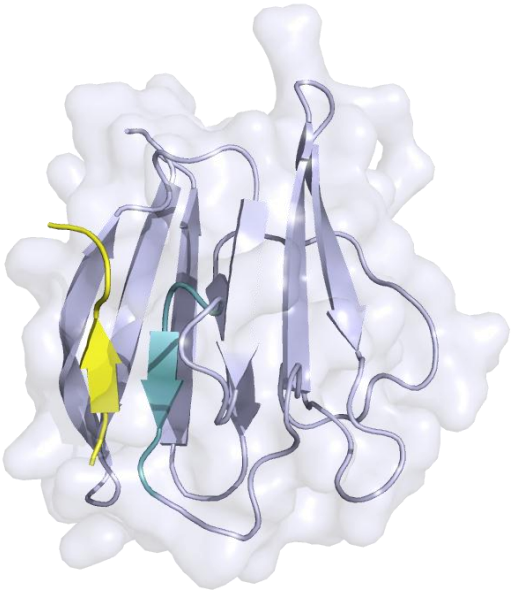

Figure 5-3. Ribbon diagram of the Ki67FHA-hNIFK peptide complex. The model is based on the NMR solution structure of the Ki67FHA-hNIFK peptide complex receptor in complex with the 44-residue fragment of hNIFK (residues 226 to 269) (PDB code: 2AFF). Upon binding the Ki67FHA receptor, the intrinsically unstructured heptapeptide fragment of hNIFK (yellow; residues 260 to 266) forms a β-sheet with the complementary receptor strand (cyan); the remaining ribbon diagram of the receptor is shown in gray. Consistent with Figure 5-2C, a molecular surface is displayed for the receptor. Molecular graphics were created using PyMol.

Table 5-1. Backbone hydrogen bonds formed in the misregistered states (+2 and -2) and native states of the β-sheet between the Ki67FHA receptor and hNIFK peptide. Models of the +2 and -2 misregistered states were created as described in Methods.

+2 complex			-2 complex			Native complex		
hNIFK	*Ki67FHA*	*NH…O distance (Å)*	*hNIFK*	*Ki67FHA*	*NH…O distance (Å)*	*hNIFK*	*Ki67FHA*	*NH…O distance (Å)*
NH V262	O D36	2.0	O K264	NH R38	2.1	NH K264	O D36	1.9
O E260	NH R38	2.0	NH K264	O R38	1.9	O V262	NH R38	1.9
NH E260	O R38	1.9	O V262	NH Q40	1.9	NH V262	O R38	1.9
						O E260	NH Q40	2.1

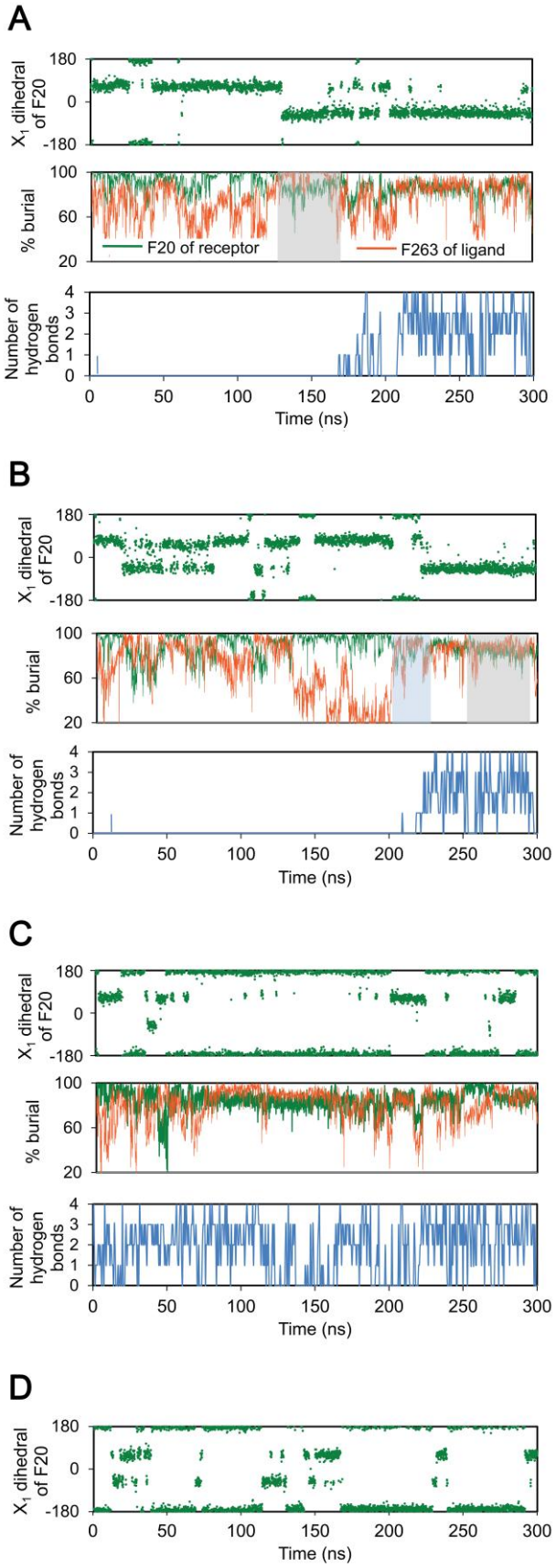

Figure 5-4. Plots of the χ1 angle of receptor residue F20, % burial of F20 and peptide residue F263, and the number of native hydrogen bonds in the β-sheet vs. time for (A) rearrangement from the +2 state to the native state occurring within 174.2 ns (red in Figure 5-2A), (B) rearrangement from the +2 state to the native state occurring within 276.7 ns (purple in Figure 5-2A), and (C) simulation starting from the native state (PDB code: 2AFF). The gray and light blue boxes in the % burial plots highlight the periods during which F263 anchors into pockets 1 and 2, respectively. A plot of the χ1 angle of F20 vs. time is also provided for a simulation starting from the unbound receptor (PDB code: 1R21) in (D). All analysis was performed using conformations sampled every 100 ps during the simulations.

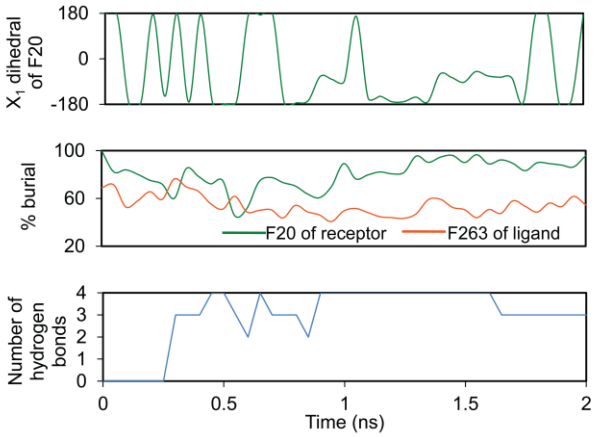

Figure 5-5. Mechanism of rearrangement from the +2 misregistered state during a 2-ns targeted MD simulation. As done in Figures 5-2B and 5-4, the mechanisms were explored by monitoring the χ1 angle of receptor residue, F20; % burial of F20 and peptide residue, F263; and the number of native hydrogen bonds in the β-sheet.

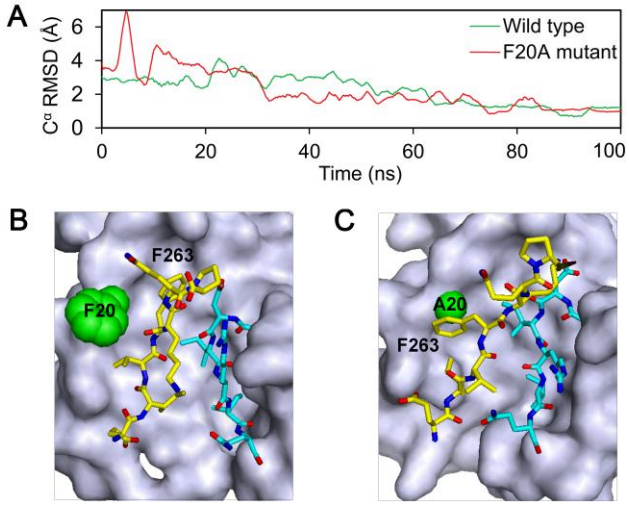

Figure 5-6. Rearrangement of the F20A mutant +2 misregistered state. (A) Plot of Cα RMSD of the β-sheet from the native complex vs. time for simulations starting from the F20A mutant +2 state (red) and native, wild-type complex (green). Snapshots reveal that peptide residue F263 anchors into different pockets of the receptor during rearrangements from the (B) wild-type +2 state and (C) F20A mutant +2 state; these pockets are referred to in the text as pockets 1 and 2, respectively. Molecular graphics were created using PyMol.

MOVIE 5-1[1] Movie of a representative simulation starting from the -2 state (first 100 ns of a 300 ns simulation) that shows receptor residue F20 sterically blocking the translation of the peptide towards the native register. Snapshots of the movie involve conformations sampled every 200 ps during the simulation. Consistent with Figure 5-2C, the movie highlights the positions of receptor residue F20 (green), peptide (yellow), and complementary β-strand of the receptor (cyan); the rest of receptor is represented by a molecular surface that is colored according to degree of burial (e.g. deep cavities in red and exposed surfaces in blue). The movie was created using VMD. Any apparent distortion of the molecules is an artifact of frame smoothing applied for purposes of visualization.

MOVIES 5-2 to 5-4 Movies of each of three independent 300-ns simulations that resulted in rearrangements of the +2 state to the native state. All other details are the same as described in the caption for Movie 5-1. As illustrated in Figure 5-2, the anchoring of peptide residue F263 into a transient hydrophobic pocket of the receptor (pocket 1) appears to facilitate rearrangement to the native state.

MOVIE S5 Movie of the 100-ns simulation characterized in Figure 5-6 that results in rearrangement of the F20A mutant +2 misregistered state. As shown in Figure 5-6C, peptide residue F263 anchors into an alternate pocket (pocket 2) even though pocket 1 is always open (as a result of the F20A mutation). The position of receptor residue A20 is highlighted in green. All other details are the same as described in the caption for Movie 5-1.

[1] Movies are available online at:

https://www.cell.com/biophysj/supplemental/S0006-3495%2811%2900383-3

5.7 REFERENCES

(1) Dou, Y; Baisnée, P.F.; Pollastri, G.; Pecout, Y.; Nowick, J.; Baldi, P. ICBS: A Database of Interactions between Protein Chains Mediated by β-sheet Formation. *Bioinformatics* **2004**, *20*, 2767.

(2) Gabdoulline, R.R.; Wade, R.C. On the Protein-protein Diffusional Encounter Complex. *J. Mol. Recognit.* **1999**, *12*, 226.

(3) Petty, S.A.; Decatur, S.M. Intersheet Rearrangement of Polypeptides during Nucleation of β-sheet Aggregates. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14272.

(4) B. Strodel, C.S. Whittleston, D.J. Wales. Thermodynamics and Kinetics of Aggregation for the GNNQQNY Peptide. *J. Am. Chem. Soc*. **2007**, *129*, 16005.

(5) Wallace, J.A.; Shen, J.K. Probing the Strand Orientation and Registry Alignment in the Propagation of Amyloid Fibrils. *Biochemistry* **2010**, *49*, 5290.

(6) Thukral, L.; Smith, J.C; Daidone, I. Common Folding Mechanism of a β-hairpin Peptide via Non-native Turn Formation Revealed by Unbiased Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2009**, *131*, 18147.

(7) Li, D.W.; Han, L.; Huo, S. Structural and Pathway Complexity of β-strand Reorganization within Aggregates of Human Transthyretin(105-115) Peptide. *J. Phys. Chem. B* **2007**, *111*, 5425.

(8) Pitera, J.W.; Haque, I.; Swope, W.C. Absence of Reptation in the High-temperature Folding of the Trpzip2 β-hairpin Peptide. *J. Chem. Phys.* **2006**, *124*, 141102.

(9) Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L.M.; Dobson, C.M.; Vendruscolo, M.; Favrin, G. Structural Reorganisation and Potential Toxicity of Oligomeric Species Formed during the Assembly of Amyloid Fibrils. *PLOS Comput. Biol.* **2007**, *3*, 1727.

(10) Li, D.W.; Mohanty, S.; Irback, A.; Huo, S. Formation and Growth of Oligomers: a Monte Carlo Study of an Amyloid Tau Fragment. *PLOS Comput. Biol.* **2008**, *4*, 1000238.

(11) Byeon, I.J.; Li, H.; Song, H.; Gronenborn, A.M.; Tsai. M.D. Sequential Phosphorylation and Multisite Interactions Characterize Specific Target Recognition by the FHA Domain of Ki67. *Nat. Struct. Mol. Biol.* **2005**, *12*, 987.

(12) Li, H., I. J. Byeon, Ju, Y.; Tsai, M.D. Structure of Human Ki67 FHA Domain and its Binding to a Phosphoprotein Fragment from hNIFK Reveal Unique Recognition Sites and New Views to the Structural Basis of FHA Domain Functions. *J. Mol. Biol.* **2004**, *335*, 371.

(13) Guex, N.; Peitsch. M.C. SWISS-MODEL and the Swiss-PdbViewer: An Environment for Comparative Protein Modeling. *Electrophoresis* **1997**, *18*, 2714.

(14) Xiang, Z.; Honig, B. Extending the Accuracy Limits of Prediction for Side-chain Conformations. *J. Mol. Biol*. **2001**, *311*, 421.

(15) Jorgensen, W.L., Chandrasekhar, J.; Madura, J.D.; Imey, R.W.; Klein, M.L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.

(16) Case, D.A., Darden, T.A.; Kollman, P.A.; *et al*. AMBER Version 10. **2004**, San Francisco: University of California at San Francisco.

(17) Hornak, V., Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712.

(18) Joung, I.S.; Cheatham, T.E. 3rd. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B.* **2008**, *112*, 9020.

(19) Berendsen, H.J.C.; Postma, J.P.M; van Gunsteren, W.F.; Dinola, A.; Haak, J.R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684.

(20) Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.

(21) Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327.

(22) Schlitter, J.; Engels, M.; Kruger, P.; Jacoby, E.; Wollmer, A. Targeted Molecular Dynamics Simulation of Conformational Change - Application to the T ↔ R Transition in Insulin. *Mol. Sim.* **1993**, *10*, 291.

(23) Case, D.A.; Darden, T.A.; Kollman, P.A; *et al*. AmberTools. Version 1.0. **2008**, San Francisco: University of California.

(24) Sanner, M.F.; Olson, A.J.; Spehner, J.C. Reduced Surface: An Efficient Way to Compute Molecular Surfaces. *Biopolymers* **1996**, *38*, 305.

(25) Pande V.S.; Baker, I.; Chapman, J.; Elmer, S.P.; Khaliq, S.; Larson, S.M.; Rhee, Y.M.; Shirts, M.R.; Snow, C.D.; Sorin, E.J.; Zagrovic, B. Atomistic Protein Folding Simulations on the Submillisecond Time Scale Using Worldwide Distributed Computing. *Biopolymers* **2003**, *68*, 91.

(26) Wei, G.; Mousseau, N.; Derreumaux, P. Computational Simulations of the Early Steps of Protein Aggregation. *Prion* **2007**, *1*, 3.

(27) Hu, X.; Wang, H.; Ke, H.; Kuhlman, B. Computer-based Redesign of a β Sandwich Protein Suggests that Extensive Negative Design is Not Required for de novo β-sheet Design. *Structure* **2008**, *16*, 1799.

(28) Lo Conte, L.; Chothia, C.; Janin, J. The Atomic Structure of Protein-protein Recognition sites. *J. Mol. Biol.* **1999**, *285*, 2177.

(29) Evans, M.R.; Card, P.B.; Gardner, K.H. ARNT PAS-B Has a Fragile Native State Structure with an Alternative β-sheet Register Nearby in Sequence Space. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2617.

(30) Eneqvist, T.; Andersson, K.; Olofsson, A.; Lundgren, E.; Sauer-Eriksson, A.E. The β-slip: A Novel Concept in Transthyretin Amyloidosis. *Mol. Cell* **2000**, *6*, 1207.

(31) Goldberg J. Structural Basis for Activation of ARF GTPase: Mechanisms of Guanine Nucleotide Exchange and GTP-myristoyl Switching. *Cell* **1998**, *95*, 237.

(32) Eigenbrot, C.; Kirchhofer, D.; Dennis, M.S.; Santell, L.; Lazarus, R.A.; Stamos, J.; Ultsch M.H.. The Factor VII Zymogen Structure Reveals Reregistration of β-strands during Activation. *Structure* **2001**, *9*, 627.

(33) Tuinstra, R.L.; Peterson, F.C.; Kutlesa, S.; Elgin, E.S.; Kron, M.A.; Volkman, B.F. Interconversion between Two Unrelated Protein Folds in the Lymphotactin Native State. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5057.

(34) Wells, J.A.; McClendon, C.L. Reaching for High-hanging Fruit in Drug Discovery at Protein-protein Interfaces. *Nature* **2007**, *450*, 1001.

6.0 THE NATIVE GCN4 LEUCINE-ZIPPER DOMAIN DOES NOT UNIQUELY SPECIFY A DIMERIC OLIGOMERIZATION STATE

K. M. Oshaben, R. Salari, D. R. McCaslin, L. T. Chong, and W. S. Horne (2012). "The Native GCN4 Leucine-Zipper Domain Does Not Uniquely Specify a Dimeric Oligomerization State." *Biochemistry*, In press. [2]

6.1 INTRODUCTION

It has been estimated that ~3% of the protein-encoding regions across known genomes specify sequences that adopt α-helical coiled-coil folds.[1] Coiled-coil proteins play diverse roles in nature, including cellular scaffolding, oligomerization domains, and mediators of transmembrane signaling.[2,3] The importance of coiled coil proteins in biology, along with their emerging role in the preparation of designed biomaterials,[4] has motivated extensive efforts to elucidate the fundamental relationship between sequence and folding behavior in this common quaternary structure.

The basic sequence pattern of the coiled coil has a simplicity that belies the ubiquity and importance of the folding motif. α-Helices are defined by a repeat of 3.6 residues per turn; thus, two helical turns comprise a ~7 residue "heptad". In a canonical α-helical coiled-coil sequence,

[2] I performed the computational part of this paper.

positions *a* and *d* in several adjacent *abcdefg* heptad repeats are hydrophobic (i.e., $(XOOXOOO)_n$ sequence repeat where X is hydrophobic and O is polar). When such a sequence adopts an α-helical fold, it forms an amphiphilic structure with a hydrophobic stripe along one face of the helix. Burial of the hydrophobic stripes in two or more such helices can drive their assembly to form a rope-like superhelical quaternary structure with a slight left-handed twist.

The precise shape complementary of residues packed together at the hydrophobic interface is a critical determinant of coiled-coil folding. Moreover, inter- or intra-helix polar contacts involving core residues or flanking *e*/*g* heptad positions can tune folding and assembly behavior. As a result of the complex interplay among these forces, closely related coiled-coil sequences can give rise to folding patterns that vary in gross structural properties such as oligomerization state (dimer, trimer, etc.),[5-7] topology (parallel or antiparallel),[8] and specificity (homotypic or heterotypic association).[9] A particular challenge in elucidating the folding behavior of coiled coils is that the above properties of the quaternary structure can vary widely without significant changes in the helical structure of individual chains.

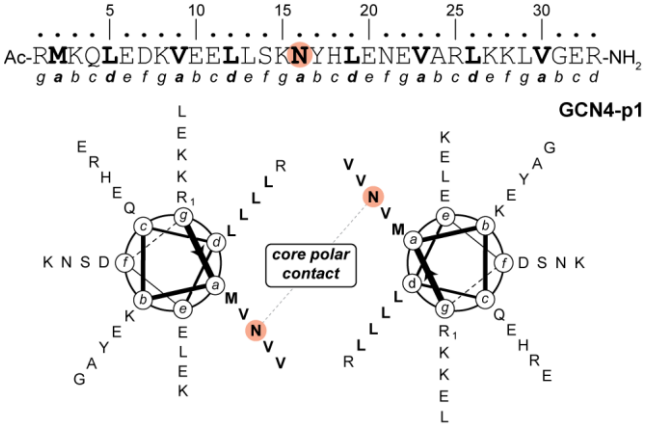

Figure 6-1. Primary sequence of GCN4-p1 and helical wheel diagram for the GCN4-p1 dimeric coiled coil. Residue Asn16, which is involved in an inter-helix polar contact in the otherwise hydrophobic core of the dimer, is highlighted.

One of the earliest and most thoroughly studied α-helical coiled-coil protein sequences is the dimerization domain of the yeast transcription factor GCN4.[10,11] The 33-residue GCN4 leucine zipper (often referred to as "GCN4-p1") is a canonical coiled coil sequence with Val and Leu residues predominating at *a* and *d* heptad positions, respectively (Figure 6-1). A single polar Asn residue found at a central *a* position plays a key role in dictating folding through formation of an inter-chain polar contact in the otherwise hydrophobic core of the dimer.[12-14] In more than two decades since it was first reported, the GCN4-p1 leucine-zipper domain has been the basis for numerous fundamental studies on α-helical coiled-coil protein folding. Although mutations are known to change its oligomerization state,[5,12-16] the wild-type sequence has generally been assumed to exclusively specify a dimer. In a landmark 1993 paper, it was shown that mutations at *a*/*d* heptad positions in GCN4-p1 could alter the preferred oligomerization state of the sequence.[5] For example, a mutant with Ile at all *a* core positions and Leu at all *d* core positions (GCN4-pIL) formed a dimer, while switching to Leu at all *a* positions and Ile at all *d* positions (GCN4-pLI) led to a tetramer. Perhaps most intriguing in that study was the observation that the Val at *a*, Leu at *d* hydrophobic core of the native sequence was poorly discriminating between dimer and trimer in the absence of the single Asn_{16}.

We report here that the native GCN4 leucine zipper domain (GCN4-p1) can adopt either a dimeric or trimeric coiled-coil fold, depending on environment. High-resolution crystal structures show how the core Asn_{16} residue is accommodated into each oligomerization state. Biophysical measurements suggest populations of both dimeric and trimeric assemblies in solution under certain experimental conditions. Microsecond molecular dynamics (MD) simulations provide insights into the relative stabilities of the two folded states in isolation. These simulations involved the application of parallel tempering, which is a special case of the

replica exchange enhanced sampling method. By making optimal use of GPU hardware, our simulations are at least ten times longer than previous replica exchange simulations of the GCN4-p1 domain, its mutants, and fragments.[17-20] Other computational studies of related systems have been limited to single-point energy calculations[21] and short, standard MD simulations (up to 100 ns).[22,23] Thus, the MD simulations we report are the most extensive to date involving the leucine-zipper fold.

6.2 METHODS

6.2.1 Peptide Synthesis and Purification

Peptides were synthesized by Fmoc solid-phase methods using manual microwave-assisted protocols[24] or in automated fashion on a Protein Technologies Tribute Automated Synthesizer. NovaPEG Rink Amide resin was used to prepare the C-terminal carboxamide, and Fmoc-Arg(Pbf)-Wang 100-200 mesh polystyrene resin was used to prepare the C-terminal carboxylic acid. Peptides were cleaved from resin by treatment with 94% trifluoroacetic acid, 2.5% water, 2.5% ethanedithiol and 1% triisopropylsilane solution for 2 to 4 hours. After the peptide was cleaved from resin, it was precipitated from the filtered cleavage solution by addition of ~40 mL cold diethyl ether. The precipitate was pelleted by centrifugation, and the ether decanted. The peptide pellet was suspended in a mixture of 0.1% TFA in water and 0.1% TFA in acetonitrile for purification. Peptides were purified by HPLC on a C_{18} preparative column using 0.1% TFA in water and 0.1% TFA in acetonitrile gradients. HPLC fractions

containing the product were combined, frozen, and lyophilized. Peptide identity was confirmed by mass spectrometry using a Voyager DE Pro MALDI-TOF instrument (monoisotopic $[M+H]^+$ *m*/*z* for GCN4-p1 C-terminal carboxamide: obsd. = 4034.7, calc. = 4036.2; GCN4-p1 C-terminal carboxylic acid: obsd. = 4038.6, calc. = 4037.2). All peptides were >95% pure by analytical HPLC on a C_{18} column.

6.2.2 Crystallization, Diffraction Data Collection, and Structure Determination

Crystallization was carried out using the hanging drop vapor diffusion method. Drops were prepared by mixing 0.7 μL of peptide stock (10 mg/mL in water) with 0.7 μL of buffer and allowed to equilibrate at room temperature over a well containing 0.7 mL of buffer solution. Crystals of the GCN4-p1 dimer were obtained from a well buffer composed of 0.1 M sodium acetate pH 4.6, 0.1 M sodium citrate tribasic dihydrate pH 5.6, and 25% w/v PEG 4000. A single crystal was flash frozen in liquid N_2 after being soaked in the above buffer supplemented with 25% v/v glycerol. Crystals of the GCN4-p1 trimer were obtained from a well buffer composed of 0.2 M ammonium sulfate, 0.1 M MES monohydrate pH 6.5 and 30% w/v PEG monomethylether 5000. A single crystal was flash frozen in liquid N_2 after being soaked in the above buffer supplemented with 10% v/v glycerol. The GCN4-p1 C-terminal carboxylic acid dimer was crystallized in 0.3 M sodium acetate pH 4.6, 0.1 M sodium citrate tribasic dihydrate pH 5.6, and 20% w/v PEG 4000. A single crystal was flash frozen in liquid N_2 after being soaked in the parent buffer supplemented with 25% v/v glycerol. The GCN4-p1 C-terminal carboxylic acid trimer was crystallized by mixing 0.7 μL of a 20 mg/ml stock solution and 0.2 μL of the buffer described above for the C-terminal carboxamide trimer. A single crystal was flash frozen in liquid N_2 after being soaked in the parent buffer supplemented with 10% v/v glycerol.

Diffraction data were collected on Rigaku Saturn 944 CCD using CuKα radiation. d*TREK was utilized to index, integrate, and scale the collected data.

Structure refinement was carried out using CCP4.[25] Phaser was used for molecular replacement, and previously published GCN4 coiled-coil derivatives were used as models; the dimer and trimer structures were solved using PDB entries 2ZTA[11] and 1IJ2,[14] respectively. A combination of refinement programs were used to complete the structure: Refmac[26] for automated refinement, Coot[27] for manual model building, ARP/wARP[28] for solvent building, and Phenix[29] for construction of composite omit maps. Phenix was also used to compare the metric symmetry between the trimer crystal forms of the C-terminal carboxamide (Table 6-1) and C-terminal carboxylic acid ($P2_1$, a = 34.6 Å, b = 58.5 Å, c = 101.3 Å, γ = 90.5°); this analysis indicated a shared primitive cell between the two lattices. Superhelix parameters and cavity volume were calculated using the TWISTER[30] and the CASTp server,[31] respectively. Buried surface area values were calculated using the PISA server.[32] Coordinates and structure factors for the refined dimer and trimer structures were deposited in the PDB under accession codes 4DMD and 4DME.

Table 6-1. Crystallographic Data Collection and Refinement Statistics.

	GCN4-p1 dimer	GCN4-p1 trimer
space group	*C*2	*C*2
unit cell		
a, *b*, *c* (Å)	83.4, 30.5, 27.8	61.2, 34.4, 78.1
α, β, γ (°)	90, 104.5, 90	90, 139.7, 90
data collection		
resolution (Å)	40.4–2.0 (2.07–2.00)	50.5–2.2 (2.28–2.20)
R_{merge} (%)	5.8 (25.3)	9.4 (22.1)
$I / \sigma(I)$	15.5 (2.2)	9.7 (3.2)
completeness (%)	99.8 (100.0)	99.5 (100.0)
redundancy	3.8 (2.4)	2.7 (2.8)
refinement		
resolution (Å)	40.4–2.00	50.54–2.20
no. reflections	4432	5198
R_{work} / R_{free} (%)	23.5 / 26.5	22.6 / 28.7
no. atoms	567	903
avg. *B* factor (Å^2)	23.9	24.7
rmsd		
bond lengths (Å)	0.016	0.010
bond angles (°)	1.8	1.4

6.2.3 Gel Permeation Chromatography

GPC was carried out on a Superdex 75 10/300 column (10 x 300 mm, 24 mL bed volume, 13 μm average particle size, GE Healthcare). The column was equilibrated with 0.15 M NaCl in 0.05 M sodium phosphate, pH 7.0. Peptides were loaded onto the column (100 μL sample at 100 μM concentration in equilibration buffer) and eluted at a flow rate of 0.8 mL/min (Figure 6-7, Supporting Information). A molecular weight calibration curve was obtained by

fitting the elution volumes of 1 mg/mL solutions of BSA, ovalbumin, aprotinin, a 17-residue synthetic peptide (Ac-YEAAAKEAAAKEAAAKA−NH_2), and vitamin B12.

6.2.4 Circular Dichroism Spectroscopy

Measurements were taken on an Olis DSM17 CD Spectrometer using 0.1 cm quartz cuvettes. Peptide concentration was determined by UV absorbance at 276 nm (ε = 1450 M^{-1} cm^{-1}) from the single Tyr residue in the GCN4-p1 sequence.[33] Samples of 100 μM peptide in buffer were prepared and scanned from 200 nm to 260 nm in 1 nm increments, an integration time of 5 seconds and a bandwidth of 2 nm at 20°C. A buffer blank was subtracted each spectrum and baseline molar ellipticity at 260 nm was set to zero. Variable temperature CD was taken by monitoring molar ellipticity at 222 nm from 20-96°C in 4°C increments with a 2 minute equilibration time between data points and an integration time of 5 seconds. Thermal melt data was fit to a two-state unfolding model[34,35] to obtain the melting temperature (T_m). Although the GCN4-p1 folding equilibrium is concentration dependent, this analysis is sufficient for qualitative comparison of folded stability at a fixed concentration.

6.2.5 Concentration-Dependent Circular Dichroism

A 400 μM solution of peptide was prepared in 10 mM phosphate buffer, 6 M urea, pH 7.0. Serial two-fold dilutions were made into 10 mM phosphate buffer, 6 M urea, pH 7.0 to generate nine samples with peptide concentrations ranging from 400 μM to 1.56 μM. Samples from 400 μM to 100 μM were measured in 1 mm quartz cuvettes, 50 μM to 6.25 μM in 2 mm quartz cuvettes, and 3.125 μM to 1.56 μM in 5 mm quartz cuvettes. Buffer solution used for the

dilutions was taken from a common stock. CD measurements of each sample were performed on an Olis DSM17 Circular Dichroism Spectrometer. Molar ellipticity of the samples was monitored at 222 nm at 20.0°C using a 2 nm bandwidth and a 5 second integration time. Three independent samples of each concentration were measured to obtain the reported molar ellipticities and accompanying error bars (standard deviation of the mean).

The concentration-dependent molar ellipticities were fit to a previously published model,[36] which assumes a simple two-state transition:

$$nP_{mon} \rightarrow P_n \quad \text{(eq. 1)}$$

$$K = \frac{P_{mon}^{\ n}}{P_n} \quad \text{(eq. 2)}$$

$$P_{tot} = P_{mon} + nP_n \quad \text{(eq. 3)}$$

$$P_{tot} = P_{mon} + n\left(\frac{P_{mon}^{\ n}}{K}\right) \quad \text{(eq. 4)}$$

where P_{tot}, P_{mon} and P_n are the total concentration of peptide, the monomer, and the *n*-mer respectively, K is the dissociation constant and n is the number of the molecules in the associated state. The concentration dependent of molar ellipticity is described by the following equations:

$$[\theta_{obs}] = [\theta_{coil}]f_{coil} + [\theta_{fold}]f_{fold} \quad \text{(eq. 5)}$$

$$[\theta_{obs}] = [\theta_{coil}](1 - f_{fold}) + [\theta_{fold}]f_{fold} \quad \text{(eq. 6)}$$

$$f_{fold} = \frac{P_n}{P_{tot}} = n\frac{P_{mon}^{\ n}}{KP_{tot}} \quad \text{(eq. 7)}$$

$$[\theta_{obs}] = [\theta_{coil}](1 - n\frac{P_{mon}^{\ n}}{KP_{tot}}) + [\theta_{fold}](n\frac{P_{mon}^{\ n}}{KP_{tot}}) \quad \text{(eq. 8)}$$

in which [θ_{fold}] is the mean molar ellipiticity of the folded peptide, [θ_{coil}] is the mean molar ellipticity of the random coil, and f is the fraction of peptide in the state specified by the subscript. The value for [θ_{coil}] was determined experimentally from a thermal melt of GCN4-p1 in 10 mM phosphate buffer, pH 7.0, 6M urea. The value for the fully unfolded baseline was used as the value for [θ_{coil}]. Using non-linear least-squares regression methods in Mathematica (Wolfram Research), the best fit parameter values for the K, n, and [θ_{fold}] were determined by fitting the data to eq. 8 and using eq. 4 to determine P_{mon}. In the fit, each data point was weighted by $1/\sigma^2$, where σ is the standard deviation from the three independent measurements.

6.2.6 Sedimentation Equilibrium Measurements

Sedimentation equilibrium measurements were carried out using a Beckman Coulter Model XL-A Analytical Ultracentrifuge. GCN4-p1 was prepared by simple dissolution in one of four buffers and used without further manipulation; the initial concentrations in each buffer were determined from spectra as recorded in the centrifuge using an extinction coefficient of 1490 M^{-1} cm^{-1} at 280 nm. The buffers were (A) 0.05 M phosphate, 0.15 M NaCl, pH 7.0 with peptide concentrations of 89, 220, and 415 μM; (B) 0.1 M sodium acetate, 0.1 M sodium citrate tribasic, pH 5.3, 245 μM GCN4-p1; (C) 0.2 M ammonium sulfate, 0.1 M MES, pH 6.6, 269 μM GCN4-p1; (D) 10 mM phosphate, 6 M urea, pH 7.0, 267 μM GCN4-p1. Buffer densities at 20°C were computed using density increment functions[37] as 1.010, 1.004, 1.012, and 1.102 g/mL respectively (a contribution for MES buffer was not available but it is likely that the ammonium sulfate is the dominant contributor). The partial specific volume of GCN4-p1 (0.748 cm^3 g^{-1}) was calculated using consensus values reported for the amino acid residues with a correction applied for the acetyl and carboxamide end groups.[38,39] In the presence of denaturing concentrations of

urea, the partial specific volume can be corrected for specific interactions of urea and water with the protein;[40] this corrections leads to an effective partial specific volume of 0.764 cm^3 g^{-1} for GCN4-p1. However, the data presented here suggest that the protein is in a mostly folded state in 6 M urea, so that the true value may lie closer to 0.748 cm^3 g^{-1}. The molecular weight of GCN4-p1 is 4038 Da including the terminal blocking groups.

Approximately 100 μL of a peptide solution was placed in one sector of a 1.2 cm pathlength, charcoal filled epon centerpiece with ~110 μL of the corresponding buffer added to the reference sector. Gradients were monitored at a nominal wavelength of 276 nm. Samples were spun at various speeds at 20°C until gradients collected 3 or more hours apart were superimposable. The equilibrium data were analyzed following an approach similar to that described by Laue[5] using software written by D.R.M. for Igor Pro (Wavemetrics, Inc, Lake Oswego, OR).

Under all buffer conditions, a single macromolecular species with a small contribution from non-sedimenting absorbance (see supplemental information) was able to describe the data. The weight average molecular weights derived from global fits of data in each buffer are summarized in Table 6-2. The variations are likely a reflection of the computed nature of the partial specific volume in various salts. Figure 6-8 (Supporting Information) shows a plot of the logarithm of the measured absorbance (after subtracting the fitted non-sedimenting absorbance) as a function of squared radial distance from the center of rotation; in such plots, a single species manifests as a series of straight lines with slopes proportional to the weight average molecular weight. The solid lines are based on the fitted weight average molecular weights shown in Table 6-2 and well account for the available data.

6.2.7 Parallel Tempering Molecular Dynamics Simulations

To determine the melting temperatures (T_m) of the dimer and trimer folded states of the GCN4-p1 leucine zipper in isolation, we used parallel tempering molecular dynamics (MD) simulations.[41-43] Parallel tempering is a widely used replica exchange enhanced sampling technique that involves simultaneously performing "replica" simulations of the system at different temperatures with the aim of making configurations at higher temperatures available to simulations at lower temperatures and vice versa. During the course of these simulations, the conformations at two different temperatures T_i and T_j are swapped at regular intervals according to the Metropolis-type criterion with a probability of:

$$P_{acc}(i,\, j) = \min\{1, \exp((\beta_i - \beta_j)(U_i - U_j))\}$$

Where $\beta = (k_B T)^{-1}$ and U is the potential energy. MD simulations at each of the replica temperatures were simulated for 1 μs using the OpenMM 4.0 software[44-47] on the XSEDE Forge GPU cluster. Replicas were exponentially spaced at temperatures from 37 to 177 °C (18 replicas for the dimer and 24 replicas for trimer), resulting in an aggregate simulation time of 18 μs and 24 μs for the dimer and trimer, respectively. After the first 1 ns, exchanges were attempted every 5 ps among all replicas using the Gibbs sampling scheme.[48] Velocities were reassigned from the Maxwell-Boltzmann distribution after every exchange attempt. The overall exchange acceptance ratio was ~13% for the dimer and ~11% for the trimer. Conformations were sampled every 5 ps. To allow for sufficient equilibration of the system, only the last 900 ns of each parallel tempering simulation was subjected to analysis. We obtained ~100 and ~50 ns/day for the dimer and trimer, respectively, using one NVIDIA Fermi M2070 GPU per replica (total of 18 GPUs for the dimer and 24 GPUs for the trimer).

To estimate T_m values of the dimer and trimer, three approaches were applied, each using a different order parameter to monitor unfolding in the parallel tempering simulations. In the first approach, an unfolded conformation was defined as having a C_α rmsd from the crystal structure that is more than one standard deviation above the average value at 37 °C in the simulations. The three C-terminal residues that were missing in the crystal structure of the dimer were not included in the rmsd calculations. In the second approach, a conformation was considered unfolded if the number of helical residues was more than one standard deviation below the average value at 37 °C (helical residues were defined as having $\varphi = -60 \pm 30$ and $\psi = -47 \pm 30$[49]). Finally, in the third approach, an unfolded conformation was defined as having at least one of the chains dissociated (i.e., beyond van der Waals distance of 4.5 Å). For each approach, the average fraction unfolded was plotted vs. temperature and the temperature of 50% unfolding was considered as the melting temperature.

To compute the folding free energies of the dimer and trimer at the temperatures of interest, we applied the multistate Bennett acceptance ratio (MBAR) method[50] to our parallel tempering simulations as implemented in the PyMBAR package (https://simtk.org/home/pymbar). This method is equivalent to the weighted histogram analysis method[51] in the limit of infinitesimally thin histogram bins with the additional feature of providing a means to directly calculate statistical uncertainties in the free energy differences. In this method, the dimensionless free energy at thermodynamic state i, $\hat{f}_i$, is estimated by solving the following equation in a self-consistent fashion:

$$\hat{f}_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{\exp[-u_i(\mathbf{x}_{jn})]}{\sum_{k=1}^{K} N_k \exp[\hat{f}_k - u_k(\mathbf{x}_{jn})]}$$

where K is the total number of replicas in the system, N_j is the number of uncorrelated samples for each replica j, and u_i is the reduced potential energy for conformation x_{jn}. The difference in free energies can then be calculated as $\Delta\hat{f}_{ij} = \hat{f}_j - \hat{f}_i$ where i and j are the two thermodynamics states of interest (i.e. folded and unfolded), respectively. Folding free energies were estimated using snapshots collected every 5 ns. The extent of folding was monitored using order parameters with non-discrete values (i.e. C^{α} rmsd and the helical content). Folding free energies at temperatures that were not included among the temperate replicas of the parallel tempering simulations (i.e. below 37 °C) were estimated by extrapolation using the MBAR method.[50] Uncertainties at each temperature were estimated by the asymptotic covariance matrix of the MBAR estimating equations.[50]

MD simulations at each of the replica temperatures were performed using the AMBER ff99SB force field,[52] generalized Born implicit solvent (OBC model II; igb=5)[53] with a surface-area dependent term,[54] and physiological salt concentration of 150 mM as implemented in the AMBER 11 molecular dynamics package.[55] Constant temperature was maintained for each of the replicas using the Langevin thermostat. To accelerate unfolding and refolding events, we used a reduced solvent viscosity (collision frequency of 1 ps^{-1}). All hydrogen bonds to heavy atoms were constrained using the SHAKE algorithm[56] (relative geometrical tolerance of 10^{-6}), enabling the use of a 2-fs time step. To increase the frequency of refolding events, a square-well distance restraint with a spring constant of 20 kcal mol^{-1} $Å^{-2}$ was applied between the C_{α} atoms of the N-terminal arginine residues in each monomer (two for the dimer and three for the trimer) to keep these atoms within 45 Å of each other, resulting in the same effective concentration of monomers for each simulation.

Starting structures for the simulations of the dimer and trimer folded states were prepared using the LEAP module in the AmberTools 1.5 package.[55] Heavy atom coordinates were taken from the crystal structures of the corresponding oligomeric form. The three C-terminal residues of the dimer had not been resolved in the crystal structure and were added using the PyMol visualization software.[57] Crystallographic water molecules were removed and hydrogen atoms were added using the protonation states present in solution at pH=7. The removal of the two buried water molecules in the trimer crystal structure (Figure 6-3C) appears to not substantially affect the stability of the trimer within our implicit solvent model as the structure remained folded in the 500 ns-long standard MD simulation at 20 ^{o}C (C_{α} rmsd values of 2.0 ± 0.4 Å relative to the crystal structure, Figure 6-14, Supporting Information). No cutoff for non-bonding interactions was used. To relieve unfavorable interactions, each starting structure was subjected to energy minimization in two stages, with position restraints applied to the heavy atoms in the first stage and no position restraints in the second stage.

6.3 RESULTS

6.3.1 Crystal Structure and Analysis

Crystallization efforts yielded diffraction-quality crystals of GCN4-p1 in two different forms. From indexing, it was clear that neither corresponded to the lattice first reported for the native GCN4 leucine-zipper sequence in a dimer fold.[11] We obtained X-ray diffraction data for each crystal form and solved the structures to 2.0 Å and 2.2 Å (Table 6-1). The higher resolution data set was readily solved by molecular replacement with a model of the previously published

GCN4-p1 dimeric coiled coil (Figure 6-2A-B, the asymmetric unit is the dimer). For the other crystal system, the dimer failed to give a reasonable solution. Molecular replacement performed with a single α-helix as the search model placed three chains in the asymmetric unit in a parallel arrangement. We repeated the molecular replacement with a trimeric coiled coil formed by a known sequence variant of GCN4-p1,[14] which led to the refined structure reported here (Figure 6-2C-D, the asymmetric unit is the trimer).

Although the crystal packing in our GCN4-p1 dimer structure differed from that previously observed for the same sequence in a different lattice (PDB 2ZTA),[11] the coiled-coil quaternary folds were found to be virtually identical (C_α rmsd of 0.53 Å). The main differences between the two structures are in the conformation of a few solvent-exposed side chains and the C-terminal tail; this segment of GCN4-p1 tends to be disordered past Gly_{31} and was not consistently resolved in electron density.

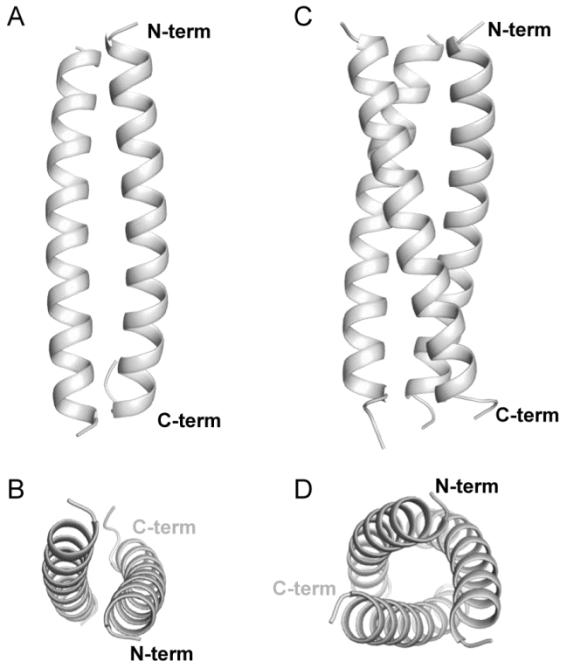

Figure 6-2. Cartoon representations of the crystal structures of the wild-type GCN4-p1 coiled coil domain in a dimeric (A,B) and trimeric (C,D) oligomerization state.

Residue Asn_{16} is known to be a key determinant of oligomerization state specificity in GCN4-p1.[11-14] This hypothesis is supported in the literature by (1) the involvement of this side chain in an intermolecular polar contact in the otherwise hydrophobic core of the dimer interface[11] (Figure 6-3B) and (2) the observation that mutation of just Asn_{16} can lead to a change in the favored oligomerization state.[12-15] We examined Asn_{16} in the trimer structure of the wild-type sequence in an effort to gain insight into how this side chain was accommodated in the core of the trimeric coiled coil (Figure 6-3). The interface between chains in the GCN4-p1 dimer is tightly packed (total buried surface area 1810 $Å^2$). The close packing is largely retained in the trimer (total buried surface area 4000 $Å^2$); however, the switch in oligomerization state is accompanied by formation of a 112 $Å^3$ cavity in the vicinity of Asn_{16} (Figure 6-3A). Although the pocket is isolated from solvent, we observed two ordered water molecules filling it in the crystal structure (Figure 6-3C). One of the two core waters appears to be stabilized by hydrogen-bonds to the three Asn_{16} carboxamide oxygen atoms lining the cavity, while the other is involved in an inter-water polar contact. A matching shift of the *e*-position Glu_{20} residues between the dimer and trimer folds leads to three new intrachain polar contacts to the Asn carboxamide (Figure 6-3C). These flanking polar interactions may help to stabilize the Asn_{16} side chains in a conformation that effectively binds the core water.

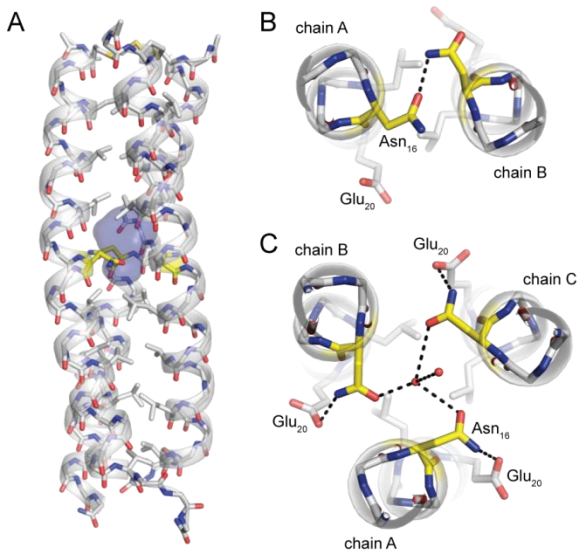

Figure 6-3. (A) A water-filled cavity is present in the hydrophobic core of the GCN4-p1 trimer. Sticks are shown for all backbone atoms as well as side-chains of hydrophobic core residues; carbons in Asn_{16} are colored yellow. (B, C) Polar interactions involving core residue Asn_{16} in the GCN4 p1 dimer (B) and trimer (C) helix bundle; two ordered water molecules resolved in the buried cavity are shown as spheres.

There are six known single-residue mutations of GCN4-p1 that lead to a trimeric oligomerization state: replacement of Asn_{16} with Ala, Val, Ser, Thr, Gln, or (*S*)-2-aminobutyric acid (Abu).[12-15] Four of the former have been structurally characterized at high resolution, so we compared the wild-type trimer structure to each of these mutants (Table 6-2). Backbone C_α overlay for residues 1-30 revealed significant structural homology between the five coiled coils; rmsd values varied from 0.55-0.80 Å. The closest match was to the fold of the $Asn_{16}\rightarrow$Thr sequence variant. Idealized superhelix parameters,[58] calculated from backbone C_α positions, show that the wild-type trimer is more tightly wound than any of the point mutants, indicated by its smaller superhelical pitch.

Table 6-2. Comparison of the wild-type GCN4-p1 coiled-coil trimer structure to known trimer-forming mutants.

	GCN4-p1 mutant[a]				
	wt	**N_{16}→S**	**N_{16}→T**	**N_{16}→Q**	**N_{16}→X**
C_α rmsd to wt (Å)	n/a	0.80	0.55	0.63	0.67
Superhelix parameter					
radius (Å)	6.3	6.4	6.3	6.5	6.5
rise per residue (Å)	1.44	1.47	1.44	1.45	1.46
pitch (Å)	130	153	139	140	143

a wt is wild-type trimer structure reported here; other structures were published previously (Asn$_{16}$→Ser, PDB 1IJ3; Asn$_{16}$→Thr, PDB 1IJ2; Asn$_{16}$→Gln, PDB 1ZIM; Asn$_{16}$→Abu, PDB 1ZIJ); Abu = (*S*)-2-aminobutyric acid; *b* rmsd values calculated from overlay to wild-type using C_α atoms in residues 1-30.

The GCN4-p1 peptide used to obtain the two structures described above has a C-terminal carboxamide, whereas the previously published crystal structure of the same sequence was reported for the C-terminal carboxylic acid. GCN4-p1 and its variants tend to be disordered past the flexible Gly_{31} residue, so we deemed it unlikely that such a small change at Arg_{33} would be responsible for the trimer fold observed. To test this assumption, we synthesized a sample of GCN4-p1 with a C-terminal carboxylic acid and attempted to grow crystals using the same two buffers that gave rise to the dimer and trimer crystal forms described above. Using the acetate-based dimer buffer, we were able to obtain a crystal of the C-terminal acid that diffracted with the same unit cell and symmetry as the C-terminal amide dimer. Using the MES-based trimer buffer, we obtained a crystal form that was crystallographically related to that of the C-terminal amide trimer; the lattice had lower symmetry, however, leading to three crystallographically independent trimeric coiled coils in the asymmetric unit. Neither C-terminal acid structure was

refined. Overall, these data suggest the identity of the C-terminal functional group has no bearing on the ability of the GCN4-p1 sequence to adopt a dimer vs. trimer fold.

6.3.2 Solution Biophysical Characterization

We performed a series of experiments to further investigate the folding behavior of GCN4-p1 in solution. To our knowledge, no prior study has shown direct biophysical evidence for the trimer fold we saw in the crystal; however the assumption of a simple two-state unfolding transition has been questioned.[59] Our goals in these experiments were to test prior observations that the dimer is the favored fold in benign buffer at pH 7 and to look for measurable population of trimer under other experimental conditions.

We carried out analytical gel-permeation chromatography (GPC) to examine the association state of GCN4-p1 in pH 7 phosphate. We compared the wild-type sequence to GCN4-pII, a known mutant that adopts a trimeric fold.[5,60] Each peptide eluted as a single peak (Figure 6-7, Supporting Information). Calibration of the column with protein molecular weight standards showed that the elution volume of GCN4-p1 corresponded to an apparent molecular weight slightly larger than expected for the dimer ($MW_{app}/MW_{calc} = 2.3$), while GCN4-pII had an elution volume that exactly matched the expected trimer. Given the resolution limitations of GPC as an analytical tool for sizing small globular proteins, the 15% deviation from the MW expected for the dimer was not deemed significant within the uncertainty of the measurement.

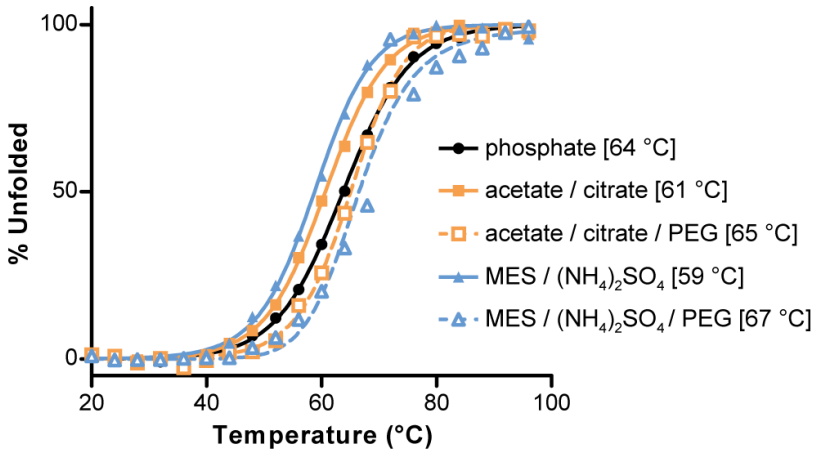

Figure 6-4. . Circular dichroism thermal melts for GCN4-p1 in different buffers. Solution conditions from top to bottom in the legend: 10 mM phosphate, pH 7; 0.1 M sodium acetate, 0.1 M sodium citrate, pH 5.3; 0.1 M sodium acetate, 0.1 M sodium citrate, 25% w/v PEG 4000, pH 5.3; 0.2 M ammonium sulfate, 0.1 M MES, pH 6.6; 0.2 M ammonium sulfate, 0.1 M MES, 30% w/v PEG 5000 monomethylether, pH 6.6. Curves are shown for fits to a two-state unfolding transition, with melting temperatures (T_m) indicated in brackets. The outlier point at 72 °C in sample (5) was observed in two independent experiments; the origin is not clear, but its presence does not impact the T_m from the fit.

A circular dichroism (CD) spectrum and thermal melt for a 100 μM solution of GCN4-p1 in pH 7 phosphate buffer were consistent with prior data.[10] This measurement served as a baseline for comparison of the folded stability of the peptide in the buffers that gave rise to the dimer and trimer crystal forms (Figure 6-4). We carried out CD thermal melts on 100 μM peptide solutions in the two crystallization conditions with and without added PEG to isolate possible effects of pH/salt from those of the precipitant. The data showed that the GCN4-p1 coiled coil was slightly destabilized by the pH and/or salt content of the two crystallization buffers relative to pH 7 phosphate. More striking was the consistent increase in coiled-coil folded stability that accompanied addition of PEG; the increase in T_m was more pronounced in the trimer buffer (ΔT_m = +8 °C) than in the dimer buffer (ΔT_m = +4 °C).

Concentration-dependent folded population can be used to determine oligomerization state in proteins where folding requires self-assembly. Folding and self-assembly are known to

be coupled in GCN4-p1;[10] however, the folded state is too stable to allow for a complete titration under concentrations accessible to CD analysis. We therefore performed a titration under partially denaturing conditions, a method developed previously for the specific purpose of determining the preferred oligomerization state of coiled-coil peptides.[36] We measured the concentration-dependent molar ellipticity of GCN4-p1 at 222 nm in 10 mM phosphate buffer, 6 M urea, pH 7 (Figure 6-5). The concentration of urea used was determined empirically to give the best range of folded population in the concentration range of peptide used in the experiment. A fit of the CD titration data to a model where the number of chains in the associated state (n) was allowed to float gave a value of $n = 2.26 \pm 0.08$, suggesting the possible presence some higher oligomerization state(s) under these conditions.

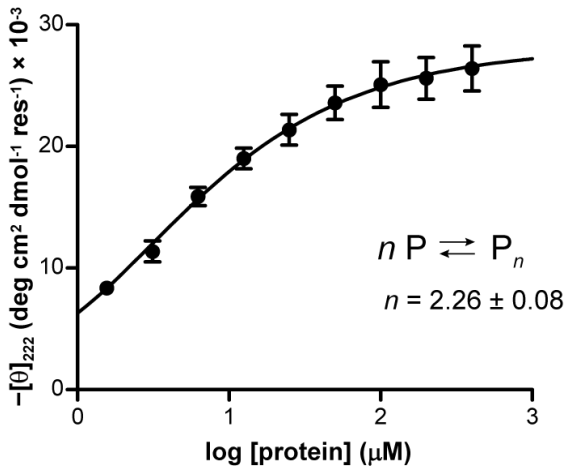

Figure 6-5. Concentration-dependent molar ellipticity of GCN4-p1 at 20 °C in 10 mM phosphate, 6 M urea, pH 7. The curve is the best fit of the data to a self-association model of monomer to *n*-mer, where *n* is allowed to float.

To further probe the self-assembly behavior of GCN4-p1 under different buffer conditions, we carried out sedimentation equilibrium measurements in four buffers: (A) pH 7 phosphate, (B) the dimer crystallization buffer without PEG, (C) the trimer crystallization buffer

without PEG, and (D) the partially denaturing urea-containing buffer used for CD titration. Over the concentration range accessible in these experiments, the analysis suggests a single macromolecular species in all buffer conditions (Figure 6-8, Supporting Information). Using a computed partial specific volume, the weight average molecular weight under each condition matches that expected for the dimer (Table 6-2). There is some uncertainty as to treatment of data obtained in 6 M urea. In the presence of high concentrations of urea corrections to the partial specific volume are needed to account for preferential interactions of the urea and water.[40] Application of a known method for this correction leads to a MW_{obs}/MW_{calc} of 2.22, in accord with the concentration-dependent CD analysis. However, the method for partial specific volume correction was originally developed to provide correct molecular weights under denaturing conditions, which is not the case here. Thus, the dimeric form of the peptide appears to best describe the oligiomeric state under all the buffer conditions tested.

6.3.3 Parallel Tempering Molecular Dynamics

In an effort to compare the relative stabilities of the GCN4-p1 dimer and trimer oligomerization states in isolation, we carried out parallel tempering MD simulations in implicit solvent on the microsecond timescale. The simulations appear converged according to three criteria. One criterion is the "end-to-end" transition time (τ_{end}), which represents the speed of diffusion of a replica in temperature space[61] and reflects the quality of mixing among replicas.[48,62-65] Our transition times are significantly shorter than the length of the simulations, resulting in a large number of round trips (184 and 135 for the dimer and trimer, respectively), with all replicas visiting most of the temperatures (Figure 6-9, Table 6-5, Supporting Information). Another criterion for convergence is the number of chain reassociations, as the

simulations should sample many dissociation and association events for accurate determination of the melting temperature. Overall, 2355 and 1791 rebinding events were observed for the dimer and trimer, respectively. The average time scale for the rebinding events (total simulation time divided by the total rebinding events) was ~8 ns for the dimer and ~13 ns for the trimer, allowing each replica to sample 125 and 75 transitions on average for the dimer and trimer simulations, respectively. Finally, obtaining converged estimates of the melting temperatures required simulation times (>100 ns, Figure 6-10, Supporting Information) more than ten times longer than previous replica exchange MD simulations on related systems.[17,18]

Table 6-3. Sedimentation Equilibrium results for GCN4-p1 under different buffer conditions.

Buffer	**MW_{obs} (Da)**[a]	**MW_{obs}/MW_{calc}**[b]
0.05 M phosphate, 0.15 M NaCl, pH 7.0	8339 ± 22	2.07 ± 0.01
0.1 M sodium acetate, 0.1 M sodium citrate, pH 5.3	7546 ± 42	1.87 ± 0.01
0.1 M MES, 0.2 M $(NH_4)_2SO_4$, pH 6.6	8276 ± 57	2.05 ± 0.01
0.05 M phosphate, 6 M urea, pH 7.0	8069 ± 66	2.00 ± 0.02

a Weight average molecular weight (MW_{obs}) from fits to a single-species model using computed densities and a partial specific volume of 0.748 mL g-1. Errors represent least squares fitting errors in the fitted parameter; *b* ratio of the weight average molecular weight to the molecular weight based on the peptide sequence (MW_{calc}).

To estimate the difference in melting temperatures between the dimer and trimer forms, we used three different order parameters to monitor the extent of unfolding: C_α rmsd from the crystal structure, helical content, and chain dissociation (see Methods). The two association states show melting temperatures that are identical within experimental error, regardless of the order parameter employed (Table 6-4, Figure 6-11, Supporting Information). It is worth noting

that the absolute values of the observed melting temperatures are artificially high due to the use of an implicit solvent model.[66-69] This systematic overestimation of T_m does not interfere with our goal to ascertain the *relative* stabilities of the dimer and trimer forms by MD simulations. The results obtained indicate that there is no significant difference in melting temperatures for the two oligomerization states under the conditions of the simulations.

Table 6-4. Melting temperatures for the dimer and trimer oligomerization states determined by parallel tempering MD simulations.

Order parameter[a]	T_m (°C)[b]	
	Dimer	**Trimer**
C_α RMSD	156 ± 4 °C	155 ± 8 °C
Fraction helicity	157 ± 4 °C	165 ± 5 °C
Chain dissociation	163 ± 4 °C	165 ± 5 °C

a Method employed to quantify folded population (see text); *b* temperature for 50% unfolding as determined by the indicated order parameter.

We also characterized the relative folding free energies of the dimer and trimer at different temperatures ranging from 0 to 180 °C. The extent of folding was monitored using two different order parameters: C_α rmsd (Figure 6-6) and the helical content (Figure 6-12, Supporting Information). Regardless of the order parameter, we observe similar qualitative trends. At temperatures below ~45 °C, the dimeric form is more stable than the trimeric form. As the temperature increases from ~45 to ~155 °C (near their melting temperatures), the dimeric and trimeric forms appear to be similar in stability, indicating that both forms can exist in solution at these temperatures. The enhanced thermal stability of the trimeric form is consistent with the greater buried surface area relative to dimer observed in the crystal structure.

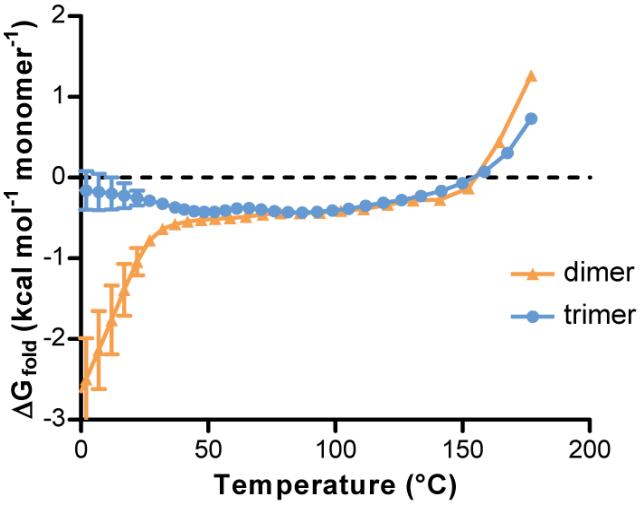

Figure 6-6. Folding free energies of the dimer vs. the trimer computed from parallel tempering simulations at various temperatures. For these computations, conformations were collected every 5 ns (total of 180 conformations at each temperature). The folded state was defined based on C_α rmsd from the crystal structure. Uncertainties are computed as described in Methods.

To identify the key conformational states that are populated by the dimer and trimer, we generated potential of mean force (PMF) surfaces of each oligomer as a function of C_α rmsd and number of inter-chain residue contacts at the temperature of each replica in the parallel tempering simulations (Figure 6-13, Supporting Information). Each simulation predominantly populates its starting state (dimer or trimer) at the lowest temperature (37 °C), with some extent of unfolding as evident by the large C_α rms deviations from the corresponding crystal structure. The apparent partial unfolding at low temperatures in the simulations is consistent with prior experimental results suggesting that the temperature-induced unfolding of the leucine zipper consists of several transitions, with the first transition starting at temperatures as low as 0 °C and the last transition involving cooperative unfolding / dissociation of the monomers.[59] Our simulation model appears to be reasonable, as found by standard molecular dynamics

simulations of each oligomer for 500 ns at 20 °C during which the oligomers remained folded the entire time (C_α rmsd values of 2.3 ± 0.9 Å for the dimer and 2.0 ± 0.4 Å for the trimer, Figure 6-14, Supporting Information). Upon increasing the temperature, our simulations of the trimeric form reveal a dimer intermediate (with ~60 inter-chain residue contacts) that becomes increasingly populated (Figure 6-13, Supporting Information). This result suggests the temperature-induced denaturation of the trimer may involve a dimeric intermediate.

6.4 DISCUSSION

Collectively, the data we report here show that the GCN4 leucine-zipper (GCN4-p1) can populate either a dimeric or trimeric coiled-coil fold depending on environment. Consistent with extensive prior work, we found the dimer fold is favored under most conditions examined, including pH 7 phosphate and pH 5.3 acetate/citrate, with added PEG or without. In a buffer composed of pH 6.6 MES/$(NH_4)_2SO_4$ with added PEG, however, we obtained a high-resolution crystal structure of wild-type GCN4-p1 in a trimeric fold. We also saw evidence for some population of trimeric coiled coil in solution under partially denaturing conditions by concentration-dependent CD analysis, although sedimentation equilibrium measurements suggest a single dimer species. Parallel tempering molecular dynamics simulations starting from the two experimentally determined crystal structures suggest that while the dimer is the dominant form at low temperatures, the folding free energy of the dimer and trimer states are similar at temperatures well below T_m. Furthermore, our simulations reveal that refolding events of the trimer involve a dimeric intermediate. It is possible that the small energy gap between the dimer

and trimer states could be closed or swapped through minor changes to environmental conditions.

The strongest direct evidence for the trimeric state of GCN4-p1 we have obtained is the high-resolution crystal structure. The possibility exists that crystal contacts present in the trimer lattice may influence the equilibrium between dimer and trimer in solution, effectively selecting a single species over the course of crystallization. The previous assumption presupposes some non-trivial amount of the trimer already present. Thus, it is informative to examine what about the trimer experimental conditions may shift the equilirium toward a trimeric fold. High protein concentration would be expected to promote trimer formation by Le Chatelier's principle. The lack of deviation from the ideal single-species model in the sedimentation equilibrium data analysis (Figure 6-8, Supporting Information) suggests that concentration alone is not sufficient to generate measurable amounts of trimer in solution up to the concentration accessible in the experiment (~0.6 mM). Although this is not as concentrated as the crystallization drop (~2 mM), a solution with the same concentration of peptide under different buffer conditions gave rise to a dimeric fold. The above analysis suggests that something about the trimer crystallization buffer must be important.

Sedimentation equilibrium measurements suggest that the buffer/salt components from the trimer crystallization conditions do not lead to a measurable amount of trimer in solution; however, the PEG from the crystallization buffer was omitted to prevent complications from a sedimenting buffer component. PEG is commonly used as a non-ionic precipitating agent in protein crystallization but also as a "crowding agent," a chemical additive that simulates the highly crowded environment proteins experience in a cell. The role of crowding in protein folding thermodynamics has been the subject of intensive study.[70,71] Theoretical calculations

suggest that crowding increases the association affinity of a macromolecular oligomerization event and that the extent of stabilization increases dramatically with the number of monomers in the assembly (i.e., a crowding-based increase in K_a will be more pronounced in a trimer vs. dimer).[71] The addition of PEG stabilizes the GCN4-p1 folded state, as evidenced by CD thermal melts. The extent of stabilization is more pronounced in the trimer crystallization buffer than the conditions that favor dimer, suggesting the possibility that the addition of PEG has shifted the equilibrium toward trimer under the appropriate buffer conditions. Crowding alone cannot be responsible for the observed trimer fold, however, since a similar concentration of non-ionic precipitant was present in conditions that gave rise to the dimer crystal structure. We hypothesize that the effect of PEG as a crowding agent coupled with the buffer conditions work in concert to shift the GCN4-p1 equilibrium toward trimer under the conditions where we obtained the crystal structure in this oligomerization state.

The above data suggest that the native GCN4 leucine zipper is on the verge of two folded coiled-coil oligomerization states, the dimeric fold observed in prior studies and the trimeric structure that we report here. The favored state is dependent on context. In partially denaturing buffers, the trimer may form to some extent, although the dimer is still favored. Crowding by addition of PEG, similar to the crowded environment of a cell interior, can favor trimer formation under certain conditions. The potential biological relevance of this observation is strengthened by a prior study that showed the GCN4 leucine zipper can act as a functional substitute for the oligomerization domain of heat shock transcription factor (HSF), a trimeric DNA-binding protein.[72] In that study, GCN4-HSF chimeras with the leucine zipper from GCN4 and DNA binding region from HSF were able to bind three-box DNA segments with affinity similar to wild-type HSF. These observations led the authors to propose that DNA binding might

enforce stoichiometry and alter the preferred oligomerization state by bringing a defined number of GCN4 leucine-zipper domains in close proximity. Our results suggest that the crowded cell environment, as well as the templating effect of the DNA, may be responsible for the functional equivalence of the GCN4/HSF chimera to wild-type HSF.

The hydrophobic core composition of GCN4-p1 (Val at *a* heptad positions and Leu at *d* heptad positions) has long been known to be poorly discriminating between dimeric and trimeric oligomerization states.[5,13,73] A variant with this core (GCN4-pVL) forms a mixture of the two assemblies in solution.[5] The addition of an *a*-position Asn residue to the Val at *a*, Leu at *d* context shifts the equilibrium toward dimer; however, our results show that the trimer state is still accessible to the wild-type GCN4-p1 protein under appropriate conditions. Recent bioinformatics analyses have quantified the ability of different amino acid residues to discriminate between dimer and trimer oligomerization states when placed in the hydrophobic core of coiled-coil proteins.[74] Consistent with prior work, *a*-position Asn residues were found to favor dimeric states in that study; however, the control exerted is not absolute. A search of an online repository of coiled coil structures[75] reveals nine *a*-position Asn residues in parallel, homotrimeric coiled-coil proteins (Table 6-6, Supporting Information). In most of these cases, a pocket exists in the hydrophobic core in the vicinity of the Asn that is occupied by an ordered water or ion – similar to the water-filled pocket we observed in the GCN4-p1 trimer. Presumably, this pocket contributes to the destabilization of the trimer state relative to the dimer (where the core is well-packed).

In summary, we have reported here that the folding behavior of the GCN4 dimerization domain, a canonical coiled coil, is much more complex than was previously appreciated. High-resolution crystal structures show the wild-type sequence can adopt a dimeric or trimeric fold,

depending on environment. Biophysical measurements suggest populations of both oligomerization states in the presence of crowding agents and possibly under partially denaturing conditions. Parallel tempering molecular dynamics simulations suggest that the dimer and trimer folds are quite close in energy. The microsecond time scales involved make these the most extensive set of simulations on the GCN4 leucine zipper published to date. Our findings have implications in ongoing efforts to establish predictive algorithms for coiled-coil folds and the selection of model systems for design and mutational studies where oligomerization state specificity is an important consideration. Moreover, our results provide further evidence for the complexity of folding behavior in even the simplest coiled-coil proteins. One take home message is the potential importance of environmental conditions in determining the preferred oligomerization state. By necessity, most biophysical studies of coiled-coil protein folding are carried out in dilute aqueous solution at or below room temperature; however, it is important to consider the potential consequences of using such biophysical observations to infer association behavior in the complex and crowded cellular environment.

6.5 ACKNOWLEDGEMENT

We thank the University of Pittsburgh and the National Science Foundation (CAREER award DMR-1149067 to W.S.H. and MCB-0845216 to L.T.C.) for funding. This work used an NSF XSEDE allocation to W.S.H. (TG-MCB110174) and L.T.C. (TG-MCB100109) on the Forge GPU cluster and the University of Pittsburgh's Center for Simulation and Modeling Linux Cluster. We are grateful for helpful discussions with Dr. John Chodera and his assistance with using the OpenMM package and MBAR method. Sedimentation equilibrium studies were

performed were obtained at the University of Wisconsin - Madison Biophysics Instrumentation Facility, which was established with support from the University of Wisconsin - Madison and grants BIR-9512577 (NSF) and S10 RR13790 (NIH).

6.6 SUPPORTING INFORMATION

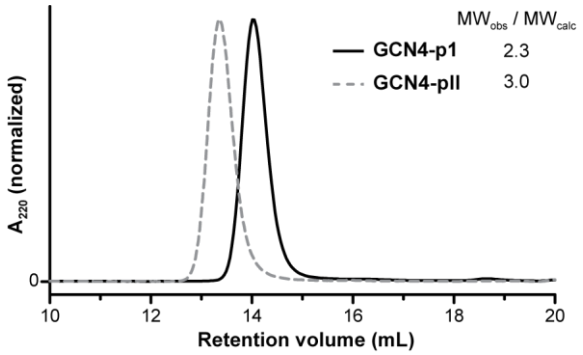

Figure 6-7. Gel-permeation chromatograms for GCN4-p1 and GCN4-pII. Injections consisted of 100 μM peptide and were eluted with 0.05 M phosphate, 0.15 M NaCl, pH 7. The ratio of molecular weight as determined by calibration of the column (MW_{obs}) to the molecular weight of monomeric peptide (MW_{calc}) is shown.

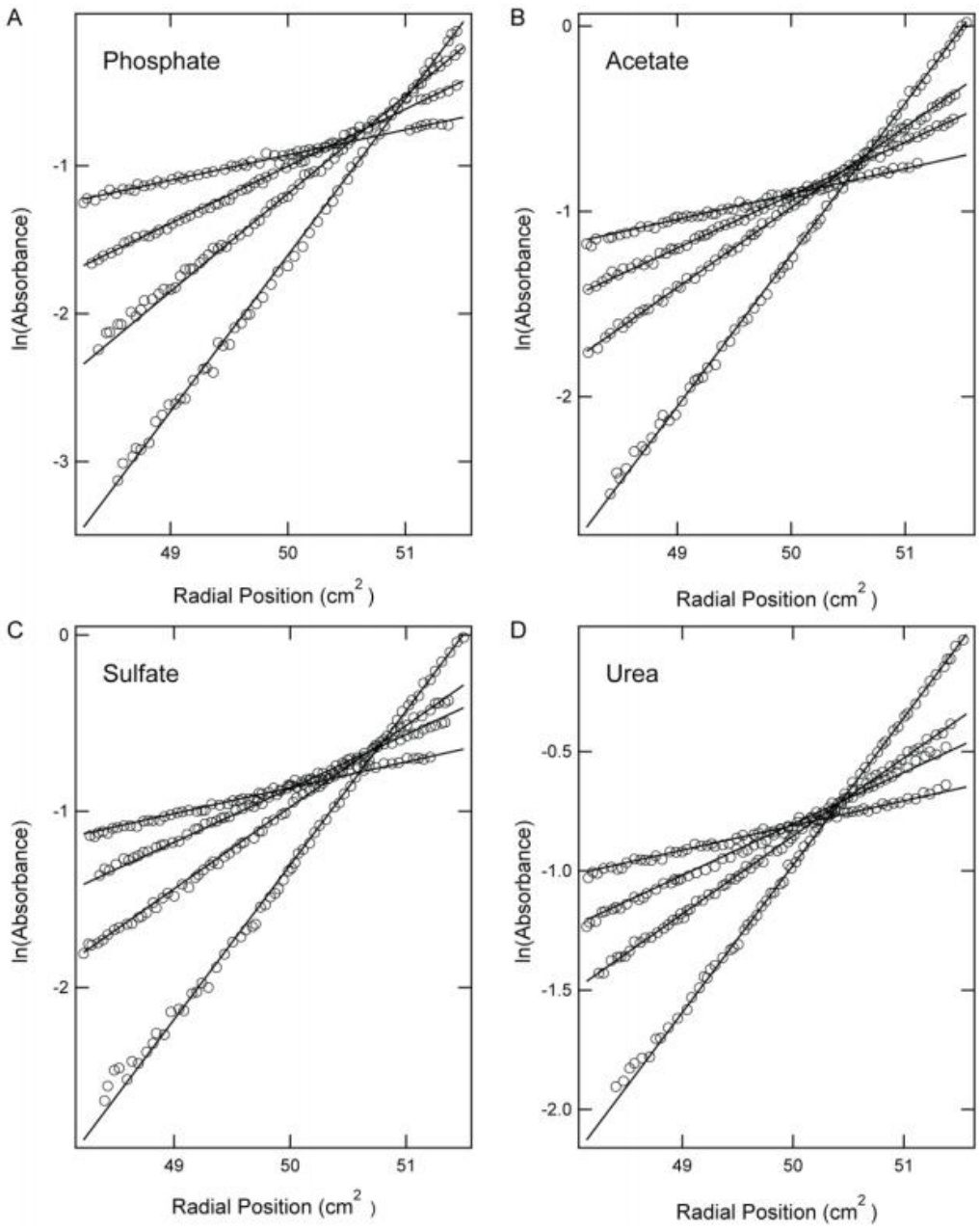

Figure 6-8. Analytical ultracentrifugation data for GCN4-p1 in four different buffers: (A) 0.05 M phosphate, 0.15 M NaCl, pH 7.0; (B) 0.1 M sodium acetate, 0.1 M sodium citrate tribasic, pH 5.3; (C) 0.2 M ammonium sulfate, 0.1 M MES, pH 6.6; (D) 10 mM phosphate, 6 M urea, pH 7.0. The points have been corrected for the non-sedimenting absorbance from the fitted model. The lines are the resulting best fit as a single macromolecular species from results shown in Table 6-3.

Sedimentation Equilibrium Data Analysis. Gradients were globally fit to a model which included a single macromolecular species and a small contribution from a non-sedimenting component (Figure S2). For panel A the analysis included speeds of 11000, 14000, 19300, 22000, 29000, 380000, 48000 at three initial loading concentrations (The data shown are for the loading concentration of 220 μM and speeds 19300, 29000, 38000 and 48000 rpm; data at other loading concentrations and speeds in this buffer are also linear). This experiment was designed to look for evidence of multiple macromolecular species, but none was found. Conditions in panels B, C and D were carried out to examine the effects of other buffer conditions similar to those employed in crystallization and spectroscopic studies on the association state. These were run at the same time with each buffer using a single loading concentration of peptide (between 245 and 267 μM). Data were collected at speeds of 18000, 26000, 32000 and 44000 rpm. At speeds greater than 50000 rpm, as yet unexplained spectral anomalies were observed at high radial positions; therefore we restricted the speeds for analysis to those 48000 and below; we note that even at higher speeds data at smaller radial positions were consistent with the finding from the global analysis at lower speeds. In our global analysis the reduced molecular weight served as the fitting variable. The reduced molecular weight is $M(1 - \nu\rho)$ where M is the weight average molecular weight, ν is the partial specific volume and ρ is the solvent density; this approach permits the impact of errors in the partial specific volume and density on M to be evaluated after the fitting process. Inclusion of a reduced molecular weight for a second species did not improve fits, and generally assuming the same partial specific volume yielded a second molecular weight not significantly different from the other component. In Table 2 the weight average molecular weights for each data set were computed from the reduced molecular weight using partial specific volumes and densities as indicated in

the text. The small molecular weight of the peptide makes it impossible even at maximal attainable speeds to fully deplete all of the peptide mass from the solution column; so the presence of non-sedimenting absorbances is not directly addressable by experiment. Such an absorbance can arise from imperfections in the cell windows, small molecule contaminants (TFA) or even slight mismatches between the sample prepared simply by dilution and the fresh buffer used as the reference. Data obtained at 60000 rpm near the solution meniscus were higher absorbance than expected even for a system composed entirely of monomer; moreover spectra recorded near the meniscus were distinct from those of the peptide recorded initially. Therefore a non-sedimenting component was included in the fitting process. In all cases the fitted value was less than the value observed at 60000 rpm near the meniscus but greater than zero and varied among the samples. Figure S2, shows a plot of the logarithm of measured absorbance (minus the fitted non-sedimenting contribution) versus the squared distance for the center of rotation. In this presentation gradients composed of a single species are transformed into a series of straight lines whose slopes depend on speed and are also proportional to the weight average molecular weight. The lines shown are based on the results of the global fitting of data to the single species model. Clearly, with available data the single species model is adequate under all conditions and over the range of concentrations accessible in the centrifugation experiment. It should be noted that the data shown in panel A is only a subset of that collected, all of which are described by the single species in Table 2.

Figure 6-9. Convergence of parallel tempering MD simulations based on all replicas visiting most of the temperatures for the (A) dimer and (B) trimer. Each subplot represents one replica's visits of the different temperatures between 37 and 177 °C.

Table 6-5. Convergence of parallel tempering MD simulations based on the average number of round trips and the end-to-end transition times (τ_{end}) for the dimer and trimer.

	Average number of round trips per replica[a]	τ_{end} (ns)
Dimer	10.2 ± 4.0	98.0 ± 38.4
Trimer	5.6 ± 2.1	178.6 ± 70.0

[a] Uncertainties in the number of round trips represent one standard deviation.

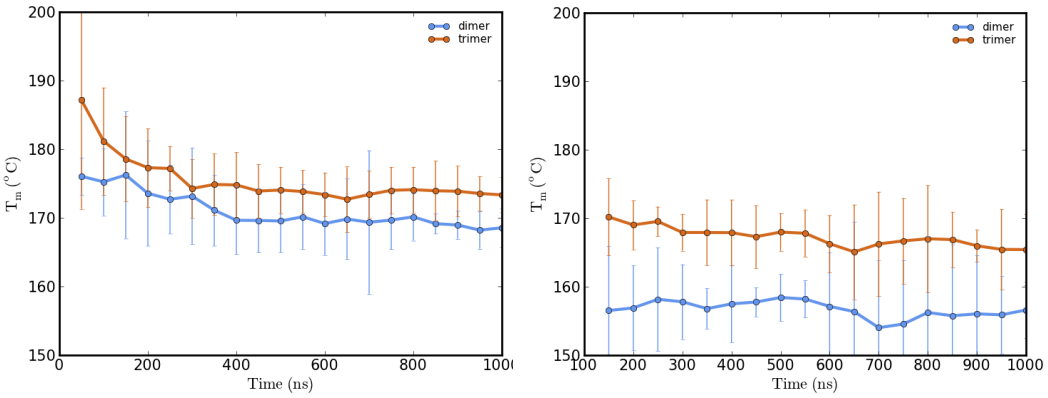

Figure 6-10. Convergence of melting temperatures of the dimer and trimer as estimated from parallel tempering MD simulations. A) Estimated T_m values as a function of simulation length. The melting temperatures change significantly during the first 100 ns. B) Estimated T_m values as a function of simulation length starting from 100 ns. The T_m values appear converged over various simulation lengths of 100-1000 ns, indicating at least 100 ns was required for convergence. Melting temperatures were estimated based on helical content (conformations were considered folded if the number of helical residues was more than 20 for the dimer or 30 for the trimer). Error bars represent one standard deviation in the estimated melting temperatures of five consecutive blocks in the corresponding time segment.

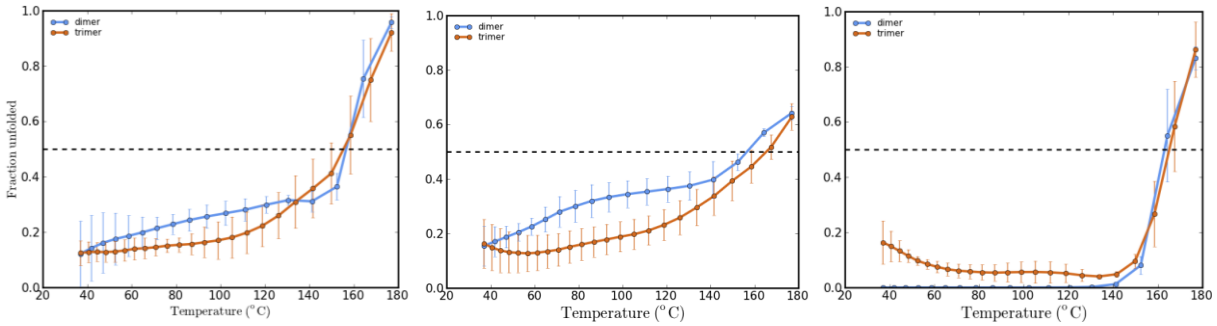

Figure 6-11. Fraction unfolded as a function of temperature based on parallel tempering MD simulations of the dimer and trimer forms, as monitored using three different order parameters. (A) C_α rmsd from the crystal structure. Conformations were considered unfolded if the C_α rmsd was more than one standard deviation higher than the average value at 37 °C. The average C_α rmsd at 37 °C relative to the crystal structure was 12.0 ± 1.0 Å for the dimer and 13.4 ± 2.4 Å for the trimer. (B) Fraction helicity. Conformations were considered unfolded if the number of helical residues was more than one standard deviation below the average at 37 °C. The average number of helical residues at 37 °C was 25.4 ± 4.9 and 38.6 ± 8.2 for the dimer and trimer, respectively. (C) Chain dissociation. Conformations were considered unfolded if at least one monomer was dissociated beyond 4.5 Å. Uncertainties represent one standard deviation in the average of the corresponding quantity calculated for five consecutive blocks of 180 ns in the simulations.

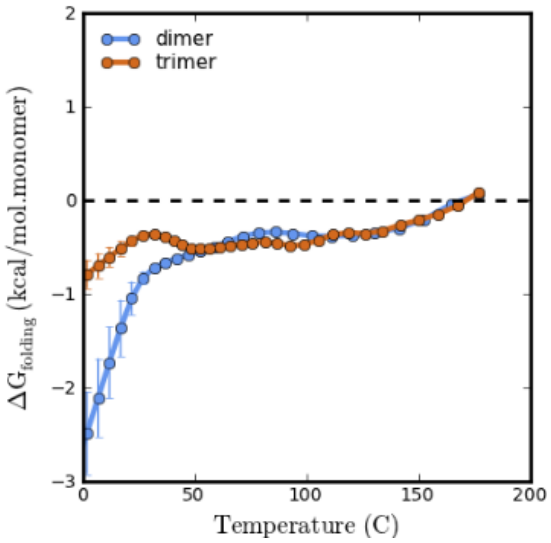

Figure 6-12. Folding free energies of the dimer and trimer computed from parallel tempering MD simulations at various temperatures, based on snapshots collected every 5 ns (total of 180 snapshots at each temperature). The folded state was defined based on helical content (see Methods). Uncertainties are provided for all data points and computed as described in Methods.

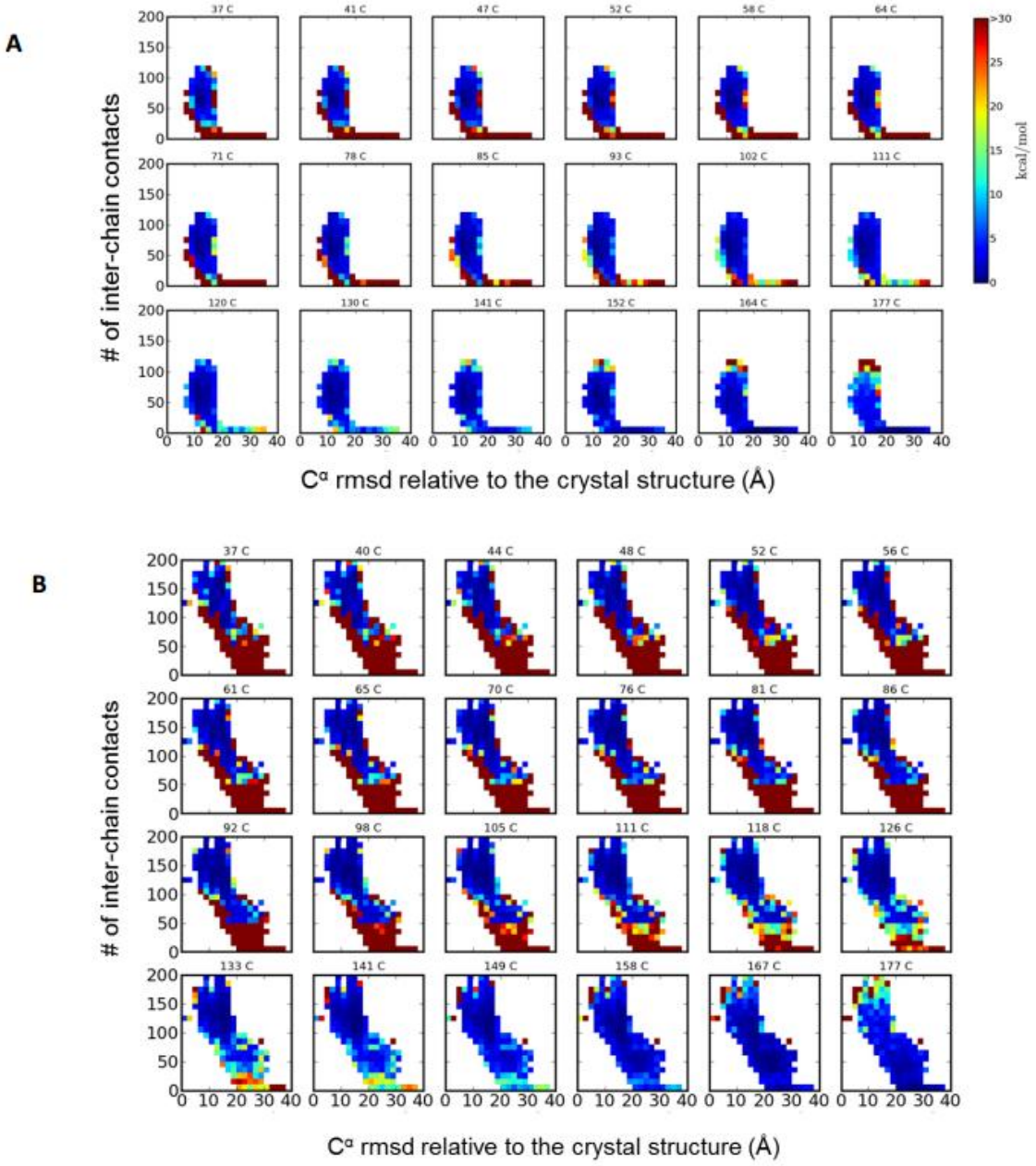

Figure 6-13. Comparison of the 2-D free energy profiles of the dimer (A) and trimer (B) as a function of C_α rmsd and the total number of inter-chain residue contacts at each replica temperature of the parallel tempering MD simulations. The color of each bin represents the corresponding free energy from blue (most favorable) to red (least favorable). The PMF plots were constructed using snapshots collected every 5 ns with bins of 2 Å width and 10 contacts height. Two residues were considered in contact if the distance between any two heavy atoms from the two residues was less than 5.5 Å.

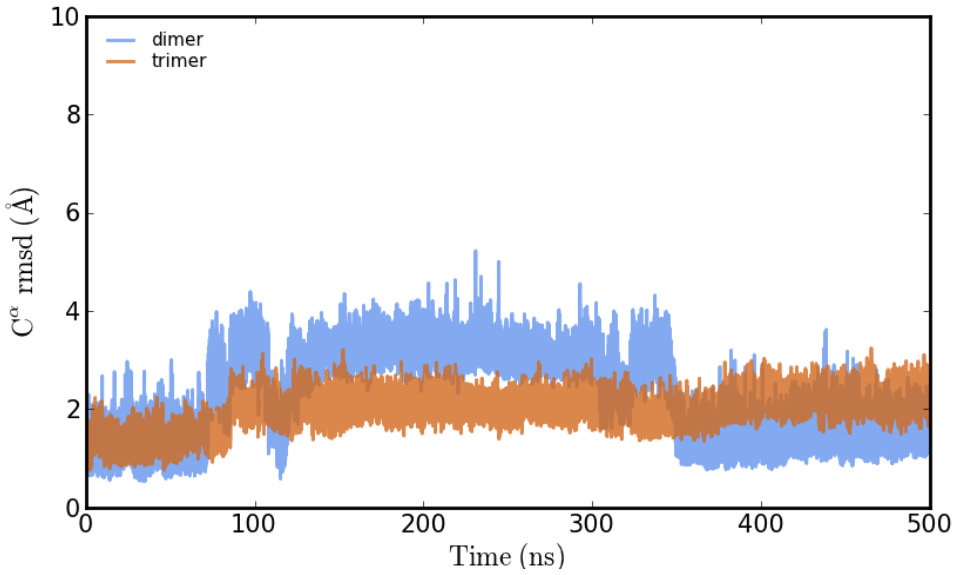

Figure 6-14. Reliability of the simulation model as gauged by low C_{α} rms deviations from the crystal structure of the dimer and trimer in standard MD simulations at 20 ^{o}C. The average C_{α} rmsd values are 2.3 ± 0.9 Å for the dimer and 2.0 ± 0.4 Å for the trimer. Uncertainties represent one standard deviation.

Table 6-6. Summary of *a*-position Asn residues in parallel, homotrimeric coiled coils in the PDB.[a]

PDB[b]	Resolution (Å)[c]	Core Asn Residue #	Core Pocket Near Asn[d]	
			Volume (Å^3)	Content
1EBO	3.00	85	49	chloride
1WP7	2.20	155	50	water
1WT6	1.60	37	100	water
		51	34	water
2FYZ	2.20	133	50	water
		148	120	nothing
2IEQ	1.75	14	206	water
2POH	2.10	63	67	water
2W6B	2.80	608	35	nothing

[a] Identified by a search of the CC+ database (ref. 76 from main text). [b] PDB accession code. [c] Refinement resolution of the crystal structure. [d] Volume of the core pocket in the vicinity of the *a*-position Asn and the contents of the pocket, if any, resolved in the crystal structure

6.7 REFERENCES

(1) Rackham, O. J. L., Madera, M., Armstrong, C. T., Vincent, T. L., Woolfson, D. N., and Gough, J. The evolution and structure prediction of coiled coils across all genomes. *J. Mol. Biol*. **2010**, 403, 480-493.

(2) Lupas, A. N., and Gruber, M. The structure of a-helical coiled coils. *Adv. Protein Chem*. **2005**, 70, 37-78.

(3) Woolfson, D. N., Bartlett, G. J., Bruning, M., and Thomson, A. R. New currency for old rope: From coiled-coil assemblies to α-helical barrels. *Curr. Opin. Struct. Biol*. **2012**, 22, 432-441.

(4) Marsden, H. R., and Kros, A. Self-assembly of coiled coils in synthetic biology: Inspiration and progress. *Angew. Chem. Int. Ed*. **2012**, 49, 2988-3005.

(5) Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine-zipper mutants. *Science* **1993**, 262, 1401-1407.

(6) Liu, J., Zheng, Q., Deng, Y., Cheng, C. S., Kallenbach, N. R., and Lu, M. A seven-helix coiled coil. *Proc. Natl. Acad. Sci*. **2006** 103, 15457-15462.

(7) Zaccai, N. R., Chi, B., Thomson, A. R., Boyle, A. L., Bartlett, G. J., Bruning, M., Linden, N., Sessions, R. B., Booth, P. J., Brady, R. L., and Woolfson, D. N. A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol*. **2011**, 7, 935-941.

(8) Oakley, M. G., and Hollenbeck, J. J. The design of antiparallel coiled coils. *Curr. Opin. Struct. Biol*. **2001**, 11, 450-457.

(9) Schnarr, N. A., and Kennan, A. J. Peptide tic-tac-toe: Heterotrimeric coiled-coil specificity from steric matching of multiple hydrophobic side chains. *J. Am. Chem. Soc*. **2002**, 124, 9779-9783.

(10) O'Shea, E. K., Rutkowski, R., and Kim, P. S. Evidence that the leucine zipper is a coiled coil. *Science* **1989**, 243, 538-542.

(11) O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. X-Ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **1991**, 254, 539-544.

(12) Gonzalez, L., Brown, R. A., Richardson, D., and Alber, T. Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nat. Struct. Biol.* **1996**, 3, 1002-1010.

(13) Gonzalez, L., Woolfson, D. N., and Alber, T. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol.* **1996**, 3, 1011-1018.

(14) Akey, D. L., Malashkevich, V. N., and Kim, P. S. Buried polar residues in coiled-coil interfaces. *Biochemistry* **2001**, 40, 6352-6360.

(15) Knappenberger, J. A., Smith, J. E., Thorpe, S. H., Zitzewitz, J. A., and Matthews, C. R. A buried polar residue in the hydrophobic interface of the coiled-coil peptide, GCN4-p1, plays a thermodynamic, not a kinetic role in folding. *J. Mol. Biol.* **2002**, 321, 1-6.

(16) Ciani, B., Bjelic, S., Honnappa, S., Jawhari, H., Jaussi, R., Payapilly, A., Jowitt, T., Steinmetz, M. O., and Kammerer, R. A. Molecular basis of coiled-coil oligomerization-state specificity. *Proc. Natl. Acad. Sci. USA* **2010**, 107, 19850-19855.

(17) Yadav, M. K., Leman, L. J., Price, D. J., Brooks, C. L., Stout, C. D., and Ghadiri, M. R. Coiled coils at the edge of configurational heterogeneity: Structural analyses of parallel and antiparallel homotetrameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution. *Biochemistry* **2006**, 45, 4463-4473.

(18) Su, L., and Cukier, R. I. Hamiltonian replica exchange method studies of a leucine zipper dimer. *J. Phys. Chem. B* **2009**, 113, 9595-9605.

(19) Cukier, R. I. A hamiltonian replica exchange method for building protein-protein interfaces applied to a leucine zipper. *J. Chem. Phys.* **2011**, 134, 045104.

(20) Gee, J., and Shell, M. S. Two-dimensional replica exchange approach for peptide-peptide interactions. *J. Chem. Phys.* **2011**, 134, 064112.

(21) Ramos, J., and Lazaridis, T. Energetic determinants of oligomeric state specificity in coiled coils. *J. Am. Chem. Soc.* **2006**, 128, 15499-15510.

(22) Missimer, J. H., Steinmetz, M. O., Jahnke, W., Winkler, F. K., van Gunsteren, W. F., and Daura, X. Molecular-dynamics simulations of C- and N-terminal peptide derivatives of GCN4-p1 in aqueous solution. *Chem. Biodivers.* **2005**, 2, 1086-1104.

(23) Lee, H., and Larson, R. G. Prediction of the stability of coiled coils using molecular dynamics simulations. *Mol. Simulat.* 2007, 33, 463-473.

(24) Price, J. L., Horne, W. S., and Gellman, S. H. Structural consequences of β-amino acid preorganization in a self-assembling α/β-peptide: Fundamental studies of foldameric helix bundles. *J. Am. Chem. Soc.* **2012**, 132, 12378-12387.

(25) Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **1994**, 50, 760-763.

(26) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **1997**, 53, 240-255.

(27) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **2004**, 60, 2126-2132.

(28) Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc*. **2008**, 3, 1171-1179.

(29) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect D: Biol. Crystallogr*. **2010**, 66, 213-221.

(30) Strelkov, S. V., and Burkhard, P. (2002) Analysis of a-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol.* **2020**, 137, 54-64.

(31) Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **2006**, 34, W116-W118.

(32) Krissinel, E., and Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, 372, 774-797.

(33) Gill, S. C., and Vonhippel, P. H. Calculation of protein extinction coefficients from amino-acid sequence data. *Anal. Biochem.* **1989**, 182, 319-326.

(34) Becktel, W. J., and Schellman, J. A. Protein stability curves. *Biopolymers* 1987, 26, 1859-1877.

(35) Shortle, D., Meeker, A. K., and Freire, E. Stability mutants of staphylococcal nuclease: Large compensating enthalpy entropy changes for the reversible denaturation reaction. *Biochemistry* **1988**, 27, 4761-4768.

(36) Betz, S., Fairman, R., Oneil, K., Lear, J., and Degrado, W. Design of 2-stranded and 3-stranded coiled-coil peptides. *Phil. Trans. R. Soc. Lond. B* **1995**, 348, 81-88.

(37) Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. Computer-aided interpretation of analytical sedimentation data for proteins, *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, **1992**, pp 90-146, The Royal Society of Chemistry, Cambridge, UK.

(38) Perkins, S. J. Protein volumes and hydration effects. *Eur. J. Biochem*. **1986**, 157, 169-180.

(39) Durchschlag, H., and Zipper, P. Calculation of the partial volume of organic compounds and polymers. *Prog. Colloid Polym. Sci.* **1994**, 94, 20-39.

(40) Prakash, V., and Timasheff, S. N. The calculation of partial specific volumes of proteins in 8 M urea solution. *Anal. Biochem*. **1981**, 117, 330-335.

(41) Hukushima, K., and Nemoto, K. Exchange monte carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn*. **1996**, 65, 1604-1608.

(42) Hansmann, U. Parallel tempering algorithm for conformational studies of biological macromolecules. *Chem. Phys. Lett*. **1997**, 281, 140-150.

(43) Sugita, Y., and Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett*. **1999**, 314, 141-151.

(44) Friedrichs, M. S., Eastman, P., Vaidyanathan, V., Houston, M., Legrand, S., Beberg, A. L., Ensign, D. L., Bruns, C. M., and Pande, V. S. Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* **2009**, 30, 864-872.

(45) Eastman, P., and Pande, V. OpenMM: A hardware-independent framework for molecular simulations. *Comput. Sci. Eng*. **2010**, 12, 34-39.

(46) Eastman, P., and Pande, V. S. Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *J. Comput. Chem*. **2010**, 31, 1268-1272.

(47) Eastman, P., and Pande, V. S. CCMA: A robust, parallelizable constraint method for molecular simulations. *J. Chem. Theory. Comput*. **2010**, 6, 434-437.

(48) Chodera, J. D., and Shirts, M. R. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *J. Chem. Phys*. **2011**, 135, 194110.

(49) Garcia, A. E., and Sanbonmatsu, K. Y. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. USA* **2002**, 99, 2782-2787.

(50) Shirts, M. R., and Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys*. **2008**, 129, 124105.

(51) Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules I: The method. *J. Comput. Chem*. **1992**, 13, 1011-1021.

(52) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, 65, 712-725.

(53) Onufriev, A., Bashford, D., and Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, 55, 383-394.

(54) Schaefer, M., Bartels, C., and Karplus, M. Solution conformations and thermodynamics of structured peptides: Molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **1998**, 284, 835-848.

(55) Case, D. A., Darden, T. A., Cheatham, I., T.E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvary, I., Wong, K. F., Paesani, F., Vanicek, X., Liu, J., Wu, X., Brozell, S., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. AMBER, 11 ed., University of California, San Francisco, **2010**.

(56) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, 23, 327-341.

(57) The PyMOL Molecular Graphics System Version 1.5.0.4, Schrödinger, LLC.

(58) Crick, F. H. C. The fourier transform of a coiled-coil. *Acta Crystallogr*. **1953**, 6, 685-689.

(59) Dragan, A. I., and Privalov, P. L. Unfolding of a leucine zipper is not a simple two-state transition. *J. Mol. Biol*. **2002**, 321, 891-908.

(60) Harbury, P. B., Kim, P. S., and Alber, T. Crystal structure of an isoleucine-zipper trimer. *Nature* **1994**, 371, 80-83.

(61) Sindhikara, D. J., Emerson, D. J., and Roitberg, A. E. Exchange often and properly in replica exchange molecular dynamics. *J. Chem. Theory. Comput*. **2010**, 6, 2804-2808.

(62) Earl, D. J., and Deem, M. W. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys*. **2005**, 7, 3910-3916.

(63) Trebst, S., Troyer, M., and Hansmann, U. H. Optimized parallel tempering simulations of proteins. *J. Chem. Phys*. **2006**, 124, 174903.

(64) Nadler, W., and Hansmann, U. Optimizing replica exchange moves for molecular dynamics. *Phys. Rev. E* **2007**, 76.

(65) Zheng, W., Andrec, M., Gallicchio, E., and Levy, R. M. Simulating replica exchange simulations of protein folding with a kinetic network model. *Proc. Natl. Acad. Sci. USA* **2007**, 104, 15340-15345.

(66) Pitera, J. W., and Swope, W. Understanding folding and design: Replica-exchange simulations of "Trp-cage" miniproteins. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 7587-7592.

(67) Nymeyer, H., and Garcia, A. E. Simulation of the folding equilibrium of alpha-helical peptides: A comparison of the generalized born approximation with explicit solvent. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 13934-13939.

(68) Jas, G. S., and Kuczera, K. Equilibrium structure and folding of a helix-forming peptide: Circular dichroism measurements and replica-exchange molecular dynamics simulations. *Biophys. J*. **2004**, 87, 3786-3798.

(69) Yeh, I. C., Lee, M. S., and Olson, M. A. Calculation of protein heat capacity from replica-exchange molecular dynamics simulations with different implicit solvent models. *J. Phys. Chem. B* **2008**, 112, 15064-15073.

(70) Zhou, H. X., and Dill, K. A. Stabilization of proteins in confined spaces. *Biochemistry* **2001**, 40, 11289-11293.

(71) Zhou, H. X., Rivas, G. N., and Minton, A. P. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys*. **2008**, 37, 375-397.

(72) Drees, B. L., Grotkopp, E. K., and Nelson, H. C. M. The GCN4 leucine zipper can functionally substitute for the heat shock transcription factor's trimerization domain. *J. Mol. Biol*. **1997**, 273, 61-74.

(73) Woolfson, D. N., and Alber, T. Predicting oligomerization states of coiled coils. *Protein Sci.* **1995**, 4, 1596-1607.

(74) Fletcher, J. M., Boyle, A. L., Bruning, M., Bartlett, G. J., Vincent, T. L., Zaccai, N. R., Armstrong, C. T., Bromley, E. H. C., Booth, P. J., Brady, R. L., Thomson, A. R., and Woolfson, D. N. A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synthetic Biology* **2012**, 1, 240-250.

(75) Testa, O. D., Moutevelis, E., and Woolfson, D. N. CC+: A relational database of coiled-coil structures. *Nucleic Acids Res*. **2009**, 37, D315-D322.

7.0 CONCLUSION

The focus of **Chapters 2** to **4** of this dissertation was on investigating how PB-based implicit solvent models compare to the explicit solvent models in estimating the desolvation penalties of salt bridges at the protein binding interfaces. The results of **Chapter 2** indicated an overall agreement between the two solvent models and **Chapter 3** extended this conclusion to capturing the temperature-dependence of the desolvation penalties using the two solvent models, provided certain temperature-dependent parameters in the implicit model are adjusted. Since implicit solvent models are generally parameterized for small solutes, and given the dramatic difference of solvent representation between the two models, this significant overall agreement for the large protein complexes is surprising and interesting. Nevertheless, the results also showed that certain discrepancies do exist which should be kept in mind especially when quantitative results are desired.

In these chapters, in order to allow direct comparison of the two water models in calculating desolvation costs of charged residues upon complex formation, the desolvation free energies of salt bridges were calculated relative to hydrophobic isosteres. Experimentally however, such a direct comparison is not possible and various components of the free energy generally cannot be determined. Instead, the contribution of salt bridges to the stability of protein folding or binding is typically estimated relative to other amino acids or different protonation states of the salt bridge residues. In the first approach, the salt bridge is mutated to another

residue (such as alanine) and then the change in stability of folding or binding is evaluated.[1-5] For the second approach, the pK_a shifts of each of the salt bridging residues (and therefore change in free energy) is measured in titration experiments.[6-8] Such studies have identified salt bridges that, relative to their experimental reference states, are stabilizing,[6, 5, 7] destabilizing[1, 4, 7] or contributing minimally to the stability.[3] For example, in the salt bridge network of R83-D39-R87 at the binding interface of barnase-barstar, R87A and D39A mutations decrease the stability of the complex by 7.7 and 5.5 kcal/mol respectively.[5] Another example is the Arc repressor, where replacing the entire buried salt bridge network R31-E36-R40 by hydrophobic residues M31-W36-L40 increases the stability of the protein by 4.5 kcal/mol.[4] It should be noted that such mutations are not conservative and it is difficult to distinguish between various components of free energy such as those due to removing the charges and those resulting from changes in the micro-environment of the protein at the salt bridge location upon mutation.

In **Chapters 2** to **4**, direct comparison of the desolvation penalties between the two solvent models required keeping the protein structures completely rigid. The next step in studying the role of salt bridges to the binding stability is to include protein flexibility in the calculations and explore the desolvation penalties and the total protein-protein binding free energies using unconstrained explicit solvent simulations. Such simulations in principle will be able to test the PB-based prediction that salt bridges generally do not contribute to the stability of protein-protein complexes. Free energy calculations can also be used to perform *in silico* mutagenesis of salt bridge residues to other naturally occurring residues and compare the resulting free energies to the experimental values. Since such simulations require "growing" or "disappearing" of atoms, special care must be taken to obtain converged simulations.[9] Depending on the size of the system it might be affordable to run the simulations long enough to assure

convergence, but typically an enhanced sampling technique is required to obtain sufficient sampling of the phase-space, such as Hamiltonian replica exchange simulations[10] and non-equilibrium methods based on the Jarzynski relationship.[11]

Can new salt bridges be engineered to improve the stability of protein-protein binding at room temperature? As previously mentioned, the uncompensated desolvation penalty of the charged residues upon complex formation is the major determining factor for the salt bridge contribution to the stability of binding. Therefore, minimizing the desolvation penalty while maintaining (or improving) favorable interactions of salt bridge residues with the rest of proteins is the primary goal in reaching stability when engineering a new salt bridge. *In silico* methods can help in screening potential locations for the new salt bridge and comparing their corresponding desolvation penalties. Results of **Chapter 2** showed that implicit solvent calculations can give good estimates of the desolvation penalties compared to explicit solvent and since they are fast, they can be used in quickly screening the desolvation free energies of various locations of the new salt bridge. Then the candidates with low desolvation penalties can be further evaluated in explicit solvent simulations or experimentally.

In **Chapter 3**, it was shown that both implicit and explicit solvent models predict reduced desolvation penalties for salt bridges at high temperatures and therefore it provides one explanation for how salt bridges contribute to the stability of proteins at higher temperatures. In the past, engineered salt bridges[2] and optimization of charge-charge interactions[12] have been used to enhance the thermostability of monomeric proteins and therefore in principle, it should be possible to introduce certain salt bridges at the binding interface and design protein-protein complexes that only form at higher temperatures. Such complexes can have interesting applications, for example acting as temperature-sensitive switches.

The results of **Chapter 4** indicated that while there is no significant linear correlation between implicit and explicit solvent models in capturing the nonspecific effects of salt on the desolvation penalties, the results are generally of similar magnitude at lower salt concentrations. However, the specific ion effects that are missed in the implicit solvent calculations can be substantial.

The focus of **Chapters 2** to **4** of this dissertation was mainly on certain thermodynamic aspects of salt bridge formation. Looking at the big picture, it's worth mentioning that the fact that most of the salt bridges in proteins don't contribute significantly to the stability at room temperature implies their potential additional roles. For example, salt bridges have been proposed to increase the specificity of binding compared to the hydrophobic residues[13] and it has been shown that they are involved in various functions such as active transportation through cell membrane, catalysis and folding.[14-20]

Chapter 5 provided direct observation of rearrangements in an intermolecular hydrogen bond register to its native state. The common mechanism for the rearrangements involved a phenylalanine residue on the peptide anchoring to a transient pocket on the protein and facilitating the rearrangement. The findings in this chapter highlight that such transient pockets, which cannot be identified using static crystal structures, can have important roles in binding and could be potentially used in designing new ligands for targets that lack well-defined hydrophobic cavities at the binding interface in the crystal structure.[21] Anchoring points on the protein surface have been proposed to be crucial in early stages of protein-protein recognition.[22] In addition, since it was shown that the targeted molecular dynamics simulations cannot capture the unbiased rearrangement mechanism, enhanced sampling methods that intend to overcome hidden energy

barriers (such as Orthogonal Space Random Walk)[23] can be tested to see if they are able to reproduce the rearrangement mechanism.

In **Chapter 6**, microsecond-long parallel tempering molecular dynamics simulations were employed to compare the relative stabilities of two GCN4 leucine zipper oligomers in isolation. The simulations showed that while the dimer fold is more stable at room temperature, both dimer and trimer folds have similar stabilities at the temperatures well below the melting temperature. In conjunction with the experimental findings, it was shown that the newly-discovered trimer fold has similar stability to the well-known dimer fold under certain conditions and can be populated in solution. The results also highlighted that even with applying an enhanced sampling technique, obtaining converged simulations still required substantially long times. Finally, while parallel tempering molecular dynamics simulations provide valuable information about the relative stabilities of the two oligomers, a thorough picture of the oligomer formation also requires kinetics information. Therefore, techniques that allow gathering kinetic data, such as the Weighted Ensemble method,[24-26] can complement the thermodynamics results obtained from parallel tempering simulations.

7.1 REFERENCES

(1) Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. Strength and co-operativity of contributions of surface salt bridges to protein stability, *Journal of Molecular Biology* **1990**, 216, 1031-1044.

(2) Sun, D. P., Sauer, U., Nicholson, H., and Matthews, B. W. Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis, *Biochemistry* **1991**, 30, 7142-7153.

(3) Marqusee, S., and Sauer, R. T. Contributions of a Hydrogen-bond Salt Bridge Network to the Stability of Secondary and Tertiary Structure in Lambda-repressor, *Protein Science* **1994**, 3, 2217-2225.

(4) Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. Are Buried Salt Bridges Important For Protein Stability And Conformational Specificity, *Nature Structural Biology* **1995**, 2, 122-128.

(5) Frisch, C., Schreiber, G., Johnson, C. M., and Fersht, A. R. Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation, *J Mol Biol* **1997**, 267, 696-706.

(6) Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. PH-Induced Denaturation Of Proteins - A Single Salt Bridge Contributes 3-5 Kcal Mol To The Free-Energy Of Folding Of T4-Lysozyme, *Biochemistry* **1990**, 29, 2403-2408.

(7) Marti, D. N., and Bosshard, H. R. Electrostatic interactions in leucine zippers: Thermodynamic analysis of the contributions of Glu and his residues and the effect of mutating salt bridges, *Journal of Molecular Biology* **2003**, 330, 621-637.

(8) Bosshard, H. R., Marti, D. N., and Jelesarov, I. Protein Stabilization By Salt Bridges: Concepts, Experimental Approaches And Clarification Of Some Misunderstandings, *Journal of Molecular Recognition* **2004**, 17, 1-16.

(9) Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K., and Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges, *Curr Opin Struct Biol* **2011**, 21, 150-160.

(10) Meng, Y., Dashti, D. S., and Roitberg, A. E. Computing Alchemical Free Energy Differences with Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) Simulations, *J Chem Theory Comput* **2011**, 7, 2721-2727.

(11) Xiong, H., Crespo, A., Marti, M., Estrin, D., and Roitberg, A. E. Free Energy Calculations with Non-Equilibrium Methods: Applications of the Jarzynski Relationship, *Theoretical Chemistry Accounts* **2006**, 116, 338-346.

(12) Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M., and Makhatadze, G. I. Engineering a Thermostable Protein via Optimization of Charge−Charge Interactions on the Protein Surface†, *Biochemistry* **1999**, 38, 16419-16423.

(13) Hendsch, Z. S., and Tidor, B. Do Salt Bridges Stabilize Proteins - A Continuum Electrostatic Analysis, *Protein Science* **1994**, 3, 211-226.

(14) Hoersch, D., Otto, H., Joshi, C. P., Borucki, B., Cusanovich, M. A., and Heyn, M. P. Role of a conserved salt bridge between the PAS core and the N-terminal domain in the activation of the photoreceptor photoactive yellow protein, *Biophys J* **2007**, 93, 1687-1699.

(15) Reyes-Lopez, C. A., Gonzalez-Mondragon, E., Benitez-Cardoza, C. G., Chanez-Cardenas, M. E., Cabrera, N., Perez-Montfort, R., and Hernandez-Arana, A. The conserved salt bridge

linking two C-terminal beta/alpha units in homodimeric triosephosphate isomerase determines the folding rate of the monomer, *Proteins* **2008**, 72, 972-979.

(16) Barouch-Bentov, R., Che, J., Lee, C. C., Yang, Y., Herman, A., Jia, Y., Velentza, A., Watson, J., Sternberg, L., Kim, S., Ziaee, N., Miller, A., Jackson, C., Fujimoto, M., Young, M., Batalov, S., Liu, Y., Warmuth, M., Wiltshire, T., Cooke, M. P., and Sauer, K. A conserved salt bridge in the G loop of multiple protein kinases is important for catalysis and for in vivo Lyn function, *Molecular cell* **2009**, 33, 43-52.

(17) Wells, G. A., Muller, I. B., Wrenger, C., and Louw, A. I. The activity of Plasmodium falciparum arginase is mediated by a novel inter-monomer salt-bridge between Glu295-Arg404, The FEBS journal 2009, 276, 3517-3530.

(18) Badieyan, S., Bevan, D. R., and Zhang, C. A salt-bridge controlled by ligand binding modulates the hydrolysis reaction in a GH5 endoglucanase, *Protein engineering, design & selection* **2012**, 25, 223-233.

(19) Wang, W., Rasmussen, T., Harding, A. J., Booth, N. A., Booth, I. R., and Naismith, J. H. Salt bridges regulate both dimer formation and monomeric flexibility in HdeB and may have a role in periplasmic chaperone function, *J Mol Biol* **2012**, 415, 538-546.

(20) Yang, J., Wu, J., Steichen, J. M., Kornev, A. P., Deal, M. S., Li, S., Sankaran, B., Woods, V. L., Jr., and Taylor, S. S. A conserved Glu-Arg salt bridge connects coevolved motifs that define the eukaryotic protein kinase fold, *J Mol Biol* **2012**, 415, 666-679.

(21) Wells, J. A., and McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces, *Nature* **2007**, 450, 1001-1009.

(22) Rajamani, D., Thiel, S., Vajda, S., and Camacho, C. J. Anchor residues in protein-protein interactions, *Proc Natl Acad Sci U S A* **2004**, 101, 11287-11292.

(23) Zheng, L., Chen, M., and Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems, *Proc Natl Acad Sci U S A* **2008**, 105, 20227-20232.

(24) Huber, G. A., and Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions, *Biophysical Journal* **1996**, 70, 97-110.

(25) Zhang, B. W., Jasnow, D., and Zuckerman, D. M. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin, *Proc Natl Acad Sci U S A* **2007**, 104, 18043-18048.

(26) Zwier, M. C., Kaus, J. W., and Chong, L. T. Efficient Explicit-Solvent Molecular Dynamics Simulations of Molecular Association Kinetics: Methane/Methane, Na+/Cl-, Methane/Benzene, and K+/18-Crown-6 Ether, *J. Chem. Theory Comput.* **2011**, 7, 1189-1197.

APPENDIX A

MULTISTATE BENNETT ACCEPTANCE RATIO

In the molecular dynamics (MD) simulations of proteins, it can be challenging to collect a sufficient amount of data to estimate physical quantities of interest, such as free energies and potential of mean force curves. Enhanced sampling techniques such as different variants of Replica Exchange Molecular Dynamics and Umbrella sampling, attempt to improve the phase-space sampling by performing multiple equilibrium simulations at different thermodynamic states or under different biasing potentials. Therefore it is essential to apply a statistical reweighting method in order to obtain free energy differences or equilibrium expectations of the properties of the system. Two such methods that are most notable are the Multistate Bennett Acceptance Ratio (MBAR) and the Weighted Histogram Analysis Method (WHAM). Here a brief introduction of the MBAR method is presented.

Suppose that we have K different equilibrium states, each with N_i uncorrelated samples. The difference in free energy of two states i and j can be calculated using:

$$\Delta F_{ij} = kT(\hat{f}_j - \hat{f}_i)$$

where $\hat{f}_i$ and $\hat{f}_j$, are the dimensionless free energies of states i and j and defined as:

$$\hat{f}_i = -\ln q_i$$

here q_i is the partition function of state i. The dimensionless free energies are then estimated by solving the following equation for each state self-consistently:

$$\hat{f}_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{e^{-u_i(x_{jn})}}{\sum_{k=1}^{K} N_k e^{\hat{f}_k - u_k(x_{jn})}}$$

in which $u_i(x)$ is the reduced potential function of configuration x (e.g. $u_i(x)=U_i(x)/k_BT_i$ in the canonical ensemble) and N_i is the number of uncorrelated equilibrium samples from each of the K thermodynamic states.

The MBAR method is equivalent to WHAM in the limit of zero-width histogram bins, but unlike WHAM, it does not depend on the energy histograms. In WHAM, discretizing the continuous energy distribution into bins introduces biases. In addition, the MBAR method also provides a direct assessment of the uncertainties for the estimated free energy values. Finally, it also can be applied to data collected using non-Boltzmann sampling schemes.

7.2 REFERENCES

(1) Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics* **2008,** *129* (12), 124105.

(2) Fajer, M.; Swift, R. V.; McCammon, J. A., Using multistate free energy techniques to improve the efficiency of replica exchange accelerated molecular dynamics. *J Comput Chem* **2009,** *30* (11), 1719-25.

(3) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comp. Chem.* **1992,** *13*, 1011-1021.