

Protein–Solvent Interactions

Ninad Prabhu and Kim Sharp*

Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia, Pennsylvania 19104-6059

Received April 1, 2005

Contents

1. Introduction	1616
2. Peptide–Water Interactions	1617
3. New Experimental Probes of Protein–Water Interactions	1619
4. New Solvent Models for Protein–Solvent Simulations	1619
5. Thermal Hysteresis Proteins	1620
6. Concluding Remarks	1621
7. Acknowledgments	1622
8. References	1622

1. Introduction

The central importance of solvent interactions in stabilization of specific protein structure has long been recognized. Decades ago, Tanford and Kirkwood treated in detail the interaction of charges with solvent, and they showed how desolvation/burial of charges upon protein folding was an important factor in stability.¹ The influence of their model, with further elaborations, can still be seen in much subsequent work on protein electrostatics and implicit solvent models. A little later, Kauzmann provided a seminal insight into the second major “theme” in protein–solvent interaction: The hydrophobic effect and how burial of hydrophobic amino acid side chains could stabilize proteins and play a role in determining their structure.² The view that hydrophobicity is the major contributor to protein stability is widely held,³ although current studies of solvation recognize the importance of other types of solvent–protein interaction, including van der Waals, polar, charged, ionic, and hydrogen bonding interactions. The view of solvation as a stabilizing force was further expanded to include the possibility that solvent interactions play a role in specifying structure and function: that water is in effect the “21st amino acid”. The field of experimental and theoretical studies, even for this rather specialized topic, is now too vast to be covered in any single review. We have selected four topics for discussion in this review: peptide–water interactions, new experimental probes of protein hydration, new solvent models for long protein–solvent simulations, and thermal hysteresis proteins. The selection was guided by the theme of this issue: protein folding.

The rationale for discussing peptides is that the effect of solvent on conformations/nascent folding is best understood

* Telephone: 215-573-3506. Fax: 215-898-4217. E-mail: sharpk@mail.med.upenn.edu.



Ninad Prabhu received an M.Sc. in Chemistry from the Indian Institute of Technology, Bombay, in 1991. He did his doctoral thesis with Professor Montgomery Pettitt in the area of peptide simulations, receiving his Ph.D. in Chemistry from the University of Houston in 1998. He is currently a research scientist in the Department of Biochemistry and Biophysics at the University of Pennsylvania School of Medicine.



Kim Sharp received a B.Sc. (with honors) in Biophysics from the University of Leeds in 1978. He did his doctoral thesis with Professor Donald Brooks in the area of cell surface chemistry, receiving his Ph.D. in Chemistry from the University of British Columbia in 1985. Following that he was a postdoctoral fellow and research associate in the Department of Biochemistry at Columbia University, working with Professor Barry Honig on the theory and application of electrostatics to proteins and DNA. He joined the faculty of the department of Biochemistry and Biophysics at the University of Pennsylvania School of Medicine in 1992.

in these systems. Peptides have long been used as more experimentally and computationally tractable test systems for studying protein–solvent interactions and developing simulation methods. Thus, much of what is known about the specific and quantitative effects of solvent on proteins is derived from peptide studies. This applies particularly to

protein folding. It also goes without saying that many peptides are not simply smaller versions of proteins but have their own biological importance.

New experiments that directly access the hydration of proteins, especially in the interior of folded proteins and around unfolded proteins, have major implications for how the field views the role of water removal during folding. X-ray crystallography has long been used to analyze water around proteins, and with the routine production of stunningly high resolution structures, a wealth of water structure has been revealed. Most X-ray structures, however, are now solved at cryogenic temperatures. This fact, combined with the presence of the crystal lattice, makes the relevance of this water structure to biological temperatures and the solution phase problematic at best, and beyond the scope of this review. These concerns and other theoretical issues related to crystallographically observed water have been recently reviewed elsewhere at length.^{4–6} In contrast to the rather “static” picture of protein hydration obtained from crystallography, a “revised”, more dynamic view of protein hydration has resulted from advances in several specific types of spectroscopy, discussed below.

Protein folding and unfolding occur on the microsecond to second time scale. Thus, very long time scale molecular dynamics (MD) simulations of these events are required. At this point in time, this means one must use implicit solvent models for solvent to be realistically and tractably included. We thus focus on two such implicit models which are currently practical in MD simulations: The generalized Born (GB) and Poisson–Boltzmann (PB) implicit solvent models. These two models have changed the way many simulations of proteins in water are done. They have also done much to extend the time scale of routine simulations into the multi-nanosecond to sub-microsecond scale while retaining an accurate treatment of solvent effects. Both these models have been used to study protein electrostatics in a wide variety of applications aside from MD, but this section is not intended as a general review of GB/PB implicit solvent models.

The final topic, thermal hysteresis proteins (THPs), was included because these proteins are unique in their ability to recognize and selectively bind solvent water in its solid phase (i.e. as ice). Study of these proteins is an active area of research. Although the mechanism of thermal hysteresis is not fully understood, study of the unique aspects of THP hydration is leading to qualitatively new information about protein solvation.

2. Peptide–Water Interactions

Peptides are constituted of the same amino acid building blocks as proteins but span only a fraction of the primary structure of proteins. Thus, lacking well-defined hydrophobic cores and having high solvent exposure of all their atoms relative to the case of proteins, peptides are fundamental indicators of the influence of water on amino acids, the effect of solvent on conformation, and, thus, the effects of these factors on large proteins themselves.

Alanine dipeptide has long been the model of choice to understand the influence of water and the resulting structure and free energy changes of the basic amino acid building block. It has two peptide bonds, the side chain is a simple methyl group, and the ends are capped to avoid the complications of having to consider the dipole moments and charged group effects originating from the zwitterion. It is

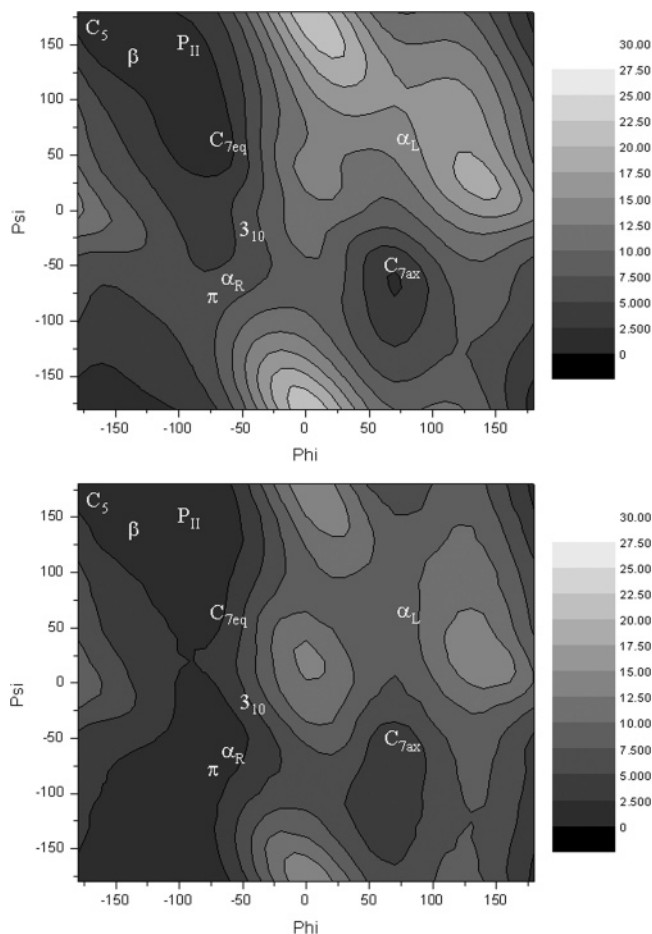


Figure 1. Ramachandran diagram for the alanine dipeptide *in vacuo* (a) computed using the CHARMM27 potential function, and including solvent water (b) computed using the finite difference Poisson–Boltzmann (PB) implicit solvent model. Conformation regions are indicated on the map.

now well-known that the transfer of alanine dipeptide from the gas or vacuum phase to the aqueous phase causes a significant change in its preferred conformation. Specifically, the C_5 and C_{7eq} regions of the Ramachandran plot, which are stabilized by intrapeptide hydrogen bonds *in vacuo* (Figure 1a), are destabilized in aqueous medium. Instead, the right-handed α helix (α_R), β sheet (β), and polyproline (P_{II}) regions share the preferred conformations (Figure 1b). Competition for hydrogen bonds from water molecules, shielding of the backbone polar groups, the orientations of the peptide dipoles (amide and carbonyl groups) that favor solvent interactions, and also favorable entropies contribute to this reordering of relative free energies of conformations.

Alanine dipeptide models have also been commonly used to test the feasibility of new computational methods such as implicit solvent models. (For older studies, see the review by Prabhu et al.⁷) More recent tests of methods addressed the folding of tripeptides with generalized Born (GB) surface area (SA) and Monte Carlo (MC) sampling using a concerted rotational algorithm⁸ and a comparison of explicit and implicit solvent models for Met-Enkephalin.⁹ Advances in computational methods such as the polarized continuum model and advances in computers have also made possible detailed studies of the solvent influenced free energy surface and short peptides by quantum mechanics calculations.^{10–12}

With increasing number of residues, peptides can adopt well-defined secondary structures in water. Observed sec-

ondary structures are the right-handed α helix, the 3–10 helix, PPII, β sheets, β turns, and β hairpins. It was believed that, in water, short peptides less than 10 residues long have random secondary conformations, with secondary structure elements rapidly interconverting between various folded and random-coil-like conformations.

Water interactions cause a softening of many of the free energy barriers between conformational states. Also, in this view there were insufficient interactions to stabilize a well-defined structure in short peptides. Advances in experimental and computational methods have revised this picture. It is possible to determine the conformational equilibrium of peptides by using NMR spectroscopy and theoretical methods.¹³ There are fewer disparate conformations than previously believed, and even very short peptides have been found to have backbone dihedral conformations that are dominated by two conformations, such as PPII and β in short alanine polymers.¹⁴ In a study¹⁵ of a nonhomologous set of protein structures, it was observed that the $\alpha(r)\beta$ turn is the most common turn type in 4-mers.

Long alanine rich chains are known to have a predominantly right α helical structure. Helix formation in water is enthalpy driven by about 1 kcal/(mol residue). In the helix, the peptide backbone is shielded from solvent and the loss of solvent–backbone interactions is partially compensated by the formation of internal hydrogen bonds. The calculation of the different parts of the thermodynamic cycle to estimate solvation energies for α helix formation is not trivial, as has been shown by Avbelj and Baldwin^{16–18} They have concluded that experimental amide transfer data which are commonly used to determine free energies for helix formation are not appropriate in this context, and free energy calculations have to take into account the peptide and solvent environment in more precise detail. Leucine pairs have been shown to stabilize alanine rich sequences in water. Luo and Baldwin¹⁹ have shown that the degree of shielding afforded to the backbone by interactions between the leucine side chains affects the stability of the helix. They have calculated that the ($i, i + 4$) Leu–Leu interaction is stronger due to lesser shielding of the peptide backbone groups from the solvent. Ionizable or polar residues inserted into alanine rich peptides have been shown to desolvate polar groups of the peptide backbone and thus increase the helical propensity of the peptide.²⁰

Peptide chains with less than 20 residues, especially those rich in alanine content, have been found to show a preference for the PPII conformation. This conformation has been a subject of much recent work, especially since it was observed that unfolded states of proteins have a PPII bias.²¹ Thus, the PPII conformation is a possible model for the unfolded state. A PPII propensity scale has been proposed for amino acids in which solvent occlusion caused by bulky side chains disfavors PPII formation.^{22,23} Proline rich sequences show the strongest PPII propensity due to steric restrictions imposed by the prolyl ring. It was found that glutamine shows the next highest propensity, which has been attributed to the PPII conformation being ideal for a favorable backbone–side chain interaction.²⁴ Bulky β branched side chains such as valine were found to have a low PPII propensity. Also, the propensities decreased with chain length.

The PPII conformation is believed to be stabilized by the solvation of the backbone.²⁵ Specifically, the interaction energy of the peptide and its first solvation shell and the

volume of the first solvent shells show correlation with the propensity to form the PPII conformation.²⁶ An implicit solvent model based on water accessible surface area was modified to include effects of optimally positioned water atoms to account for solvent peptide hydrogen bonding and was found to successfully predict the ensemble preference for the PPII conformation.²⁷ However, Pappu and co-workers²⁸ have done molecular dynamics (MD) and MC simulations of alanine dipeptide in explicit water and concluded that if the peptide–solvent interaction did indeed dominate the system free energy, a more compact, non-PPII structure would be favored. Instead, they observe that the PPII preference is dictated by the need to minimize steric conflicts between the peptide atoms. A molecular dynamics study of the initial steps of hydration of peptides demonstrated that water absorbs more readily on globular proteins than on helical ones.²⁹ Also, it was observed in simulations that β conformations have a tendency to order vicinal water and thus incur an entropic penalty,²⁶ further arguing against a solvent driven force. How water influences the tendency of peptides to adopt a PPII conformation is still not well understood.

Apart from the dominant PPII, the structural ensemble of short alanine based peptides also has significant populations of β and $\alpha(R)$ backbone conformations. The entropic benefits of a wider free energy well of the β conformation over the $\alpha(R)$ conformation (Figure 1) are possibly one reason for the large population of β structures observed.³⁰ Folded structures such as β hairpins have been observed even in alanine 8-mers.¹⁴ The β hairpins are stabilized by both intrapeptide interactions and favorable solvent interaction energy on folding.³¹ However, calculations by Brooks and co-workers³² have shown that solvation is opposed to folding and the driving force is solely due to favorable interactions between peptide atoms. They also observed that the folding process for β hairpins is highly cooperative with side chain interactions preceding intrapeptide hydrogen bond formation. The importance of solvation in the β turn formation is thus still a topic of interest. For example, while examining the reasons for the unfolding of the β hairpin YITNSNGTWT upon T3S mutation, Lei and Smith observed that the unfolded mutated structure was stabilized by intrapeptide interactions which were sterically not possible in the wild-type peptide.³³

The general feature that emerges from the many studies of the effect of solvent on peptide conformations is that solvent tends to flatten the conformational energy landscape, as exemplified for alanine dipeptide in Figure 1. However, there still remain distinct conformational preferences, and solvent effects and internal energy contributions combine to make these lie in a fairly restricted part of the φ – ψ phase space. The presence of preferred secondary structure elements in peptides has enormous consequences for the understanding of the thermodynamics and kinetics of protein folding. It used to be popular to consider unfolded proteins as having random secondary structure. Along with recent studies that show that unfolded proteins retain many elements of native structure,³⁴ it is becoming increasingly evident that even the non-native-like portions of unfolded proteins²¹ and peptides³⁵ may have fairly well-defined secondary structure (most likely PPII) which may diminish the entropic penalty associated with folding. Durani and co-workers¹⁴ have observed folds in short alanine peptides and have suggested that these may be thought of as “seeds” that initiate protein folding.

3. New Experimental Probes of Protein–Water Interactions

The structure and energetics of water around proteins in solution at biological temperatures are difficult to determine, in part because an ensemble of structures must be considered and in part because water is complex and incompletely understood even in the bulk state, let alone in the hydration shell of complex solutes such as proteins. Thus, new experimental techniques are badly needed to supplement the picture from low temperature/crystalline environment X-ray studies. There have been some recent and novel approaches to this problem, notably from advances in time-resolved spectroscopy and use of isotopes, often combined with the classic technique of site directed mutagenesis. For example, Starzyk et al. and Walsh et al. have used ^{13}C editing of Fourier transform infrared (FTIR) spectroscopy to study solvation of peptide groups in helices formation.^{36,37} By selectively labeling different residues, this type of technique provides site-specific resolution. Due to the difficulty of interpreting frequency and intensity changes in IR spectroscopy, the technique primarily gives the direction of effects, for example whether a particular group is more or less solvated. Temperature excursion IR spectroscopy³⁸ also provides information about protein–solvent interactions. This technique has been used on horseradish peroxidase to understand how protein and solvent motions are coupled and to probe the properties of water in the narrow active site channel of this enzyme.^{39,40} Time-resolved Stokes shift measurement (i.e. solvent effects on fluorescence emission frequencies) is another “high-tech” spectroscopic tool that has been used to study solvent coupled motions in calmodulin⁴¹ and monellin,⁴² and it has been applied to protein folding and surface hydration analysis.⁴³ These studies typically find a distinction between bulk water, with a characteristic relaxation time of ≈ 1 ps, and hydrating water, which typically has relaxation times of 30–60 ps. Another example of how time-resolved spectroscopy has added another dimension to experiments is in a new application of the commonly used photoacid probe Pyranine. This probe has been used by Nachliel et al. to study lac permease,⁴⁴ revealing that the binding cleft is deep (>9 Å) and hydrated by low activity water.

Magnetic relaxation dispersion (MRD), an NMR technique, can be used to quantify the amount and degree (via the relaxation time) of solvent binding and water exchange rates, when applied with suitable isotopes—notably ^{17}O for protein and ^{23}Na for nucleic acids.⁴⁵ Another NMR technique, spin diffusion, can also be used to distinguish between bulk and hydrating waters.⁴⁶ The MRD technique shows that the bulk of water at protein surfaces ($>95\%$) has, within a factor of 2, the same motional freedom as bulk water.⁴⁷ The implication is that the highly retarded water seen in Stokes shift type experiments is either buried water or the remaining 5% of surface water that is strongly hydrated. Study of the internal water in the ligand binding cavity of fatty acid binding protein using MRD shows that the amount of bound water increases upon binding of the hydrophobic ligand (!), in contrast to the picture obtained from crystallography.⁴⁸ Moreover, at least one of these waters can remain “bound” to the protein after unfolding by urea.⁴⁹ Both observations reflect the spectroscopy driven revision of the rather static picture of protein hydration inherited from crystallographic analyses.

To summarize the results of these and other recent spectroscopy studies, some of the interesting hydration features that are emerging are as follows: that protein bound water can have 1–1.5 orders of magnitude slower rotation than bulk (crystallographic analysis only told us that observed waters are more localized in position) but that most surface water is similar to bulk in this respect; that the interior of proteins can be surprisingly wet; and that denatured proteins, like their native counterparts, often “contain” significant amounts of bound water.

4. New Solvent Models for Protein–Solvent Simulations

As far as solvent models for protein simulation are concerned, there are, loosely speaking, two computational regimes: those requiring detailed analysis of one to a few structures, and those requiring the analysis of many (perhaps $>10^4$) structures. Examples of the former are detailed binding free energy calculations, pK_a , electron-transfer calculations, and structure analysis of a specific protein. Examples of the latter are molecular dynamics simulations, docking, ligand screening, folding simulations, and combinatoric protein design. For the former class of problems, the solvent treatment of choice remains the Poisson–Boltzmann (PB)/surface area (SA) model, since it gives accurate energetics of electrostatic and hydrophobic terms in peptides and proteins,^{50–52} it can include ionic strength effects on highly charged molecules accurately,^{53,54} it is rapid and easy to apply, and there are several software packages available. Recent applications to binding that exemplify the wide applicability of the PB approach to solvation include the barnase–barstar system,^{55,56} binding of peptides to membranes,⁵⁷ antibody–antigen interactions,⁵⁸ the design of electrostatically optimal kinase ligands,⁵⁹ and prediction of DNA binding sites on proteins.⁶⁰ Other “classic” applications of the PB approach to solvation include the closely related pK_a (i.e. proton-transfer) and electron-transfer problems.^{61,62} The just referenced works, and others in the literature, refer to studies of proteins that include quantitative comparison of PB with experimental data on solvation free energies, pK_a shifts, redox midpoint shifts, binding free energies, ionic strength effects, Stark effect shifts, and prediction of binding sites. Quantitative agreement is generally very good across a broad range of applications, indicating that the PB model provides a robust and realistic treatment of electrostatics and electrostatic contributions from solvation.

For problems such as protein folding simulations, that require evaluation of the contribution from solvent to thousands of protein conformations, PB methods have until recently been considered too slow. Explicit water models, which are even more computationally intensive, are usually ruled out in these types of applications too. With the introduction of the generalized Born (GB) model of solvation,⁶³ and the discovery that the model could, with careful parametrization on ensembles of protein and nucleic acid conformations, be made to reproduce PB model energies with much less calculation,^{64–67} there has been an explosion of applications of the GB model to protein solvation. Much of this work has recently been reviewed,^{68,69} so we discuss here only very recent trends and applications of particular interest.

NMR determination of protein structures provides the conformation(s) in solution. In such determinations, the number of NMR derived constraints is typically much less than those obtained from X-ray diffraction. For both reasons

one might expect the solvent treatment in any molecular mechanics based refinement to be a factor in the quality of the final structure(s). This has been systematically examined by Xia *et al.*, who compared no solvent, an explicit water model, and a GB implicit solvent model.⁷⁰ They found that the GB model was significantly better than no solvent and quite close to the explicit water model. Given the relatively modest cost of the GB model, it suggests that inclusion of some kind of solvent treatment should be a standard part of NMR refinement.

Implicit solvent models are increasingly being used for protein design. It is argued that protein design is a particularly challenging application for solvent models, since they are being used to discriminate between quite similar proteins that may differ, for example, by a single choice of amino acid or even a single rotamer at a particular site.⁷¹ Jaramillo *et al.*⁷¹ systematically compared several different solvent models, including the GB model, the PB model, and the effective energy function.⁷² Their analysis showed that most of the models produced rather similar results and that they could detect misfold decoys quite reliably. However, their analysis produced the rather counterintuitive result that burial of polar, even charged, side chains was more favorable than burial of apolar side chains. This counters the expectation that desolvation of charged side chains would result in little favorable or mostly unfavorable net free energy for burial. This issue warrants further study. An alternative approach to putting water into protein design is to analyze specific hydration/H-bonding patterns in the protein structural database, parametrize a function based on this, and effectively “recompute” a library for solvated rotamers.⁷³ A related approach to putting H-bonding into an implicit solvent model is the “directional” atomic solvation energy.⁷⁴

The GB and PB solvent models both treat the hydrophobic term by a solvent accessible area model. While proportionality with area is a good approximation for linear alkanes, there is significant nonproportionality with area when branched and cyclic molecules are considered.⁵² A systematic approach to improving the hydrophobic area term in implicit solvent models has been made by Levy and co-workers.⁷⁵ This model has been combined with the GB model for the electrostatic term and applied to protein folding, specifically the energetics of β -hairpin and α -helix formation.⁷⁶ The model would presumably give equivalent improvements if used with PB electrostatics too.

Since the GB model does not have the true physical form for describing the effect of solvent polarization on atomic forces, a number of alternative implicit solvent models for MD have been proposed. Two hark back to the influential Langevin model for solvent polarization,⁷⁷ one which uses polarizable solvent pseudoparticles,⁷⁸ and one that uses solvent configurations from an explicit simulation to evaluate the dipolar reaction field.⁷⁹ While the PB model is routinely used as the standard of precision and for parametrization in the GB model, the PB method itself has been considered too slow for direct force calculations in MD simulations. This is no longer true. Recent developments in PB technology, such as rapid PB solvers and smoothed dielectric boundaries,^{80–82} have now made MD with in-line PB implicit solvent force calculation feasible.^{83–86} Comparison with explicit water simulations on eight proteins showed that the PB method is about 5–10 times faster and has at least as good agreement with experiment for *dynamic* properties such as NMR derived backbone and side chain order parameters.

For MD simulations, the benchmark for solvent detail remains explicit water models, although these are significantly slower than implicit solvent models such as the PB model. Nevertheless, further development of explicit solvent models is considered desirable for improving the accuracy of protein–solvent interactions. In addition, explicit solvent models will always be required to model specific hydration structures such as individual protein–water H-bonds or small water clusters where the graininess of water is the salient feature. Water models currently used with MD simulations such as SPC/E and TIP3P are so-called fixed charged (nonpolarizable) models. It is generally agreed that including polarizability is an essential first step in improving explicit water models. A variety of polarizable water models have in fact been in development by physical chemists for many years, but no consensus model or approach has emerged that has been widely applied to protein simulations. In part, this is due to uncertainty as to how these models would work with the standard (nonpolarizable) MD force fields for proteins, AMBER, CHARMM, GROMOS, and OPLS, and the formidable task of redoing the numerous “benchmark” protein simulations with a new water model to verify that it is at least as good as, if not better than, existing water models. This picture may be changing with the implementation and initial testing of a fluctuating charge water model in a major MD package, CHARMM, by the Brooks group, one of the main developers and testers of CHARMM.⁸⁷

5. Thermal Hysteresis Proteins

In the field of protein–solvent interactions, thermal hysteresis or antifreeze proteins (THPs) play a special role. THPs are produced by a variety of polar fish, insects, and other animals that live in cold environments. These proteins lower the freezing point of water in a noncolligative fashion, i.e., through a nonequilibrium thermodynamic or kinetic mechanism. They achieve their effect by preventing the formation or growth of ice nuclei. Arguably, THPs perform the most amazing and difficult molecular recognition task in biology: They must specifically bind ice nuclei over the vast excess of 55 M liquid water. Moreover, they must do this purely by physical differences between water phases, since unlike the cases of other protein–ligand recognition events there are no chemical differences to recognize. At least four classes of THPs have been identified—types I, II, III, and IV, respectively—along with a less well-characterized family of antifreeze glycoproteins (AFGPs). For general reviews, see, e.g., refs 88 and 89. The different THP classes are unrelated evolutionarily, and they vary widely in sequence and structure. For example, type I THPs are predominantly helical, type II THPs are cysteine rich, and type III THPs are globular, predominantly β sheet proteins. The less well-characterized AFGPs are largely unstructured.⁹⁰ The key residues involved in the putative ice-binding site also vary widely between classes, forming a repeating pattern in type I THPs, a single flat site in type III THPs, and a structurally and chemically diverse and poorly characterized sugar group interface in AFGPs. Given that THPs bind ice nuclei in water, there are inherent and currently unsolved difficulties in assaying binding strength and determining high-resolution protein–ligand complexes experimentally. Thus, theory and simulation have played a more prominent role than usual in determining the ligand–protein structure and recognition mechanism. Three crucial questions are currently being studied and remain to be definitively

answered: (a) How does binding prevent ice crystal growth? (b) Where on the ice crystal does THP bind? (c) What residues and what kinds of interactions do THPs use to discriminate between liquid and solid water?

Three mechanisms of ice binding/growth poisoning have been proposed: First, recognition by some complementary array of H-bonding groups that match (or slightly mismatch) the ice lattice. Second, some kind of hydrophobic recognition. The term hydrophobic recognition is less well defined in structural terms than H-bonding interactions, but it may be taken to mean a similarity between the prevalent structures of liquid water around hydrophobic groups and the ice structure. This similarity of structure would imply a favorable interaction of such a hydrophobic group with ice vs water, since the entropic cost of organizing the ice into the preferred structure(s) has already been paid. The third proposed mechanism is accumulation of THPs at the ice surface combined with a “remodeling” of the ice planes. The last mechanism, however, still implies some specific interactions (H-bonding, hydrophobic, etc.) that would drive the accumulation. For example, see ref 89 for a review of mechanisms.

On the question of where on the ice crystal THPs bind, the major experimental tool is ice etching combined with electron microscopy. Which ice lattice planes are favored by different THPs has been reviewed by Madura et al.⁸⁹ and is beyond the scope of this review. Interestingly, THP binding is stereospecific and enantiomer selective.⁹¹ However, the overall lesson which can be drawn from many studies of this type is that no one ice face is favored. Types I, II, and III each favor different faces, and even within one class there are variations. This is not unexpected in, e.g., type III THPs, which are small globular proteins with a single ice binding patch. This argues against a strict protein/ice H-bond array matching and selectivity requirement, especially in type III THPs. A recent modeling study indeed shows that type III THPs can interact favorably with different ice surfaces.⁹²

Whether H-bonding or hydrophobic interactions dominate the ice recognition is still under debate. Strom et al. present a geometric/lattice based analysis which emphasizes the role of H-bonding and bridging interactions.⁹³ This model does not include specific protein structural information, but it would presumably apply best to type I THPs which have repeating H-bonding side chains on one side of the α helix. Other recent theoretical studies have emphasized the importance of hydrophobic and weak van der Waals interactions.^{94,95} Site directed mutation and modeling studies of Baardsnes et al.⁹⁶ on a type I THP from sculpin also support the importance of hydrophobic interactions from Ala residues and disprove the idea that Lys residues might mediate polar interactions with the ice (The latter are apparently there to keep the THP soluble). Using unnatural amino acids such as 2-aminobutyric acid, Haymet et al. have shown that hydrophobic interactions of key Thr residues in type I THPs are not themselves sufficient for activity.⁹⁷ While an increasing number of site directed mutagenesis/function and modeling studies have been published, the central, still outstanding puzzle can be quickly conveyed by one example: In a particular type I THP the mutation T \rightarrow A abolishes thermal hysteresis activity, while in a particular type III THP the reverse mutation A \rightarrow T abolishes activity. Sharp et al. have argued that the “either hydrophobic/or H-bond” model creates the following dilemma:^{98–100} To preferentially recognize ice in excess liquid water using apolar groups alone (which by

consensus prefer water that is more ordered) would result in low affinity, since these groups are *hydrophobic*. Conversely, use of polar groups alone (which have a high affinity for the water molecule) would result in low specificity since they do not favor ice-like water structure, just the opposite. The resolution to the specificity/affinity dilemma is that THPs create, by a combination of neighboring polar and apolar interactions, a unique water structure (seen only in THPs, and then only in their binding sites) around polar groups that is more ice-like than that around apolar groups. This more ice-like water structure is characterized by an increase in low angle water–water H-bonds and a concomitant decrease in high angle water–water H-bonds. This model rationalizes the effect of the A/T and other mutants on type I and III THP activity, and it provides a unified mechanism for ice nucleus binding across THP classes.^{99,100} The implication of both polar and apolar interactions was in fact proposed quite early in MD studies of just type I THPs.¹⁰¹ The problem of the THP mechanism(s) has stimulated new theoretical approaches to studying hydration and water structure, principally involving the angular structure of water, which should be useful for understanding protein hydration in general. Key features of THP function still remain unclear, however: Although the unique hydration around THP ice binding sites can be recognized when it occurs, the precise structural/side chain type cannot be specified in advance; that is, we are unable to predict or design *de novo* ice binding sites at this time. It is still unclear where on the ice crystal each THP binds and whether binding on particular crystal planes is necessary or sufficient to halt growth. It is not known exactly how growth is halted: by physical obstruction, ice-lattice distortion, or some combination of these or other effects.

6. Concluding Remarks

Numerous peptide studies have by now shown that solvent radically changes conformational preferences from those expected from purely intramolecular, or *in vacuo*, interactions. While the exact set of conformations adopted by a peptide varies with sequence and is still difficult to elucidate due to the conformational flexibility of peptides, the consensus is that most peptides primarily populate a rather small set of conformations relative to the number expected from all the allowed regions of the Ramachandran plot. These conformational preferences are key for understanding activity of peptides in solution. They also have implications for protein folding. One important inference is that the unfolded state of proteins is much less random than supposed; that is, the conformational entropy penalty for protein folding is considerably less than theoretical estimates from random coil models. Thus, peptide conformational preferences could act to stabilize proteins, even though the conformational preferences do not resemble the native state conformation. Moreover, nascent structures could act as nuclei for protein folding, and reduce the time for protein folding from the very slow time predicted from an extensive random search of conformational space to that observed in experiments.

Recent spectroscopic studies of the dynamics of water around proteins have considerably complicated our rather one-dimensional picture of protein hydration as a dry interior in the folded state, with a monolayer of “bound” water at the surface or in cavities. The interior of a protein can be surprisingly “wet”, while the unfolded state can have significant amounts of bound water. Combined, these

observations have unexpected implications for the stability and kinetics of protein folding. First, in these cases the amount of folding-related dehydration would be less than that expected from changes in accessible area calculated from the structure, with a consequent reduction in the magnitude of energy changes (whether positive or negative) associated with dehydration of polar or hydrophobic groups. Second, if the number of waters that have to be “squeezed” out of the protein for it to attain its final structure is less than previously thought, this would alter the kinetics of folding, probably increasing the folding rate since removal of the last hydrating waters is considered rate-limiting in some models for protein folding. In any event, it suggests a detailed treatment of explicit waters and the kinetics of their removal is required in protein folding simulations in order to correctly model folding kinetics.

For the long MD simulations required to model protein folding, implicit water models provide the best tradeoff between realism and computational feasibility. The most widely used model is the generalized Born (GB) model. This has been extensively tested against the PB model for ensembles of static structures, and it has been shown, with suitable parametrization, to reproduce the accuracy of the PB model with considerably less computation. The recent implementation of the PB model in MD now provides an alternative option for long simulations. It also allows for the direct comparison and testing of the GB model against the PB model using actual MD simulations, rather than static snapshots. These developments should improve the accuracy of implicit solvent treatments in folding simulations. It is clear, however, that an implicit treatment of the entire solvent is inherently limited and that indeed for folding simulations it may omit an important aspect of the kinetics of removal, or indeed nonremoval, of individual waters as the protein core assumes its final form. Explicit solvent simulations, at least for part of the solvent near the protein, are required at some point to address these issues.

Thermal hysteresis proteins are special in their ability to recognize and selectively bind solvent water in its solid phase (i.e. as ice). Thus, study of these proteins provides a unique perspective on protein–solvent interactions. Two aspects of protein hydration that have emerged from these studies are as follows: First, the angular structure of water, described for example by the water–water H-bonding angle, is important. Perturbation of this angular structure by hydrophobic and polar groups is particularly sensitive, and it is one measure of the “hydrophobicity” of a protein surface region. Second, studies of THP active sites reveal the importance of neighboring polar and hydrophobic groups in modulating the hydration and in how the polarity/hydrophobicity of a group depends on its neighbor context.

In summary, our understanding of protein–solvent interactions continues to evolve away from the early, dichotomized viewpoint, whereby the unfolded protein is a random coil, fully unfolded and fully hydrated, which then folds to a particular structure with a dry predominantly hydrophobic interior, and where there is clean division of the solvent exposed protein surface into either hydrophobic or polar regions based on the identity of protein side chains. Peptide studies tell us that the unfolded state is a lot less random and that it has nascent structure, while new spectroscopic probes tell us that there are sometimes buried residues and trapped waters. Such probes also tell us that the folded protein may have a surprisingly wet core. Studies of thermal

hysteresis proteins show that the hydrating water structure around a protein side chain, and hence its hydrophobicity or polarity, depends on its structural and sequence context. These and future studies will lead to a more nuanced, and hopefully more realistic, view of protein hydration.

7. Acknowledgments

We thank Manoranjan Panda and Joy Yang for many discussions on protein–solvent interactions and the NIH (PO1 GM48130) and NSF (MCB02-35440) for financial support.

8. References

- (1) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333–5339.
- (2) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1–63.
- (3) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
- (4) Makarov, V.; Pettitt, B. M.; Feig, M. *Acc. Chem. Res.* **2002**, *35*, 376–84.
- (5) Feig, M.; Pettitt, B. M. *Structure* **1351**, *6*, 1351–4.
- (6) Pettitt, B. M.; Makarov, V. A.; Andrews, B. K. *Curr. Opin. Struct. Biol.* **1998**, *8*, 218–21.
- (7) Prabhu, N. V.; Perkyns, J. S.; Blatt, H. D.; Smith, P. E.; Pettitt, B. M. *Biophys. Chem.* **1999**, *78*, 113–26.
- (8) Ulmschneider, J. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2004**, *126*, 1849–57.
- (9) Shen My, M. Y.; Freed, K. F. *Biophys. J.* **2002**, *82*, 1791–808.
- (10) Wang, Z. X.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699–716.
- (11) Hudaky, I.; Hudaky, P.; Perczel, A. *J. Comput. Chem.* **2004**, *25*, 1522–31.
- (12) Nagy, P. I.; Alagona, G.; Ghio, C.; Takacs-Novak, K. *J. Am. Chem. Soc.* **2003**, *125*, 2770–85.
- (13) Groth, M.; Malicka, J.; Rodziewicz-Motowidlo, S.; Czaplowski, C.; Klauzel, L.; Wiczak, W.; Liwo, A. *Biopolymers* **2001**, *60*, 79–95.
- (14) Ramakrishnan, V.; Ranbhor, R.; Durani, S. *J. Am. Chem. Soc.* **2004**, *126*, 16332–3.
- (15) Santa, H.; Ylisirio, M.; Hassinen, T.; Laatikainen, R.; Perakyla, M. *Protein Eng.* **2002**, *15*, 651–7.
- (16) Avbelj, F.; Luo, P.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10786–91.
- (17) Avbelj, F. *J. Mol. Biol.* **2000**, *300*, 1335–59.
- (18) Baldwin, R. L. *Biophys. Chem.* **2002**, *102*, 203–10.
- (19) Luo, P.; Baldwin, R. L. *Biophys. Chem.* **2002**, *96*, 103–8.
- (20) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13075–9.
- (21) Ferreone, J. C.; Hilser, V. J. *Protein Sci.* **2003**, 447–57.
- (22) Creamer, T. P.; Campbell, M. N. *Adv. Protein Chem.* **2002**, *62*, 263–82.
- (23) Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P. *Proteins* **2003**, *53*, 68–75.
- (24) Chellgren, B. W.; Creamer, T. P. *Biochemistry* **2004**, *43*, 5864–9.
- (25) Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2003**, *125*, 8178–85.
- (26) Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. *Proteins* **2004**, *55*, 502–7.
- (27) Fleming, P. J.; Fitzkee, N. C.; Mezei, M.; Srinivasan, R.; Rose, G. D. *Protein Sci.* **2005**, *14*, 111–8.
- (28) Drozdov, A. N.; Grossfield, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2004**, *126*, 2574–81.
- (29) Kohtani, M.; Jarrold, M. F. *J. Am. Chem. Soc.* **2002**, *124*, 11148–58.
- (30) Levy, Y.; Jortner, J.; Becker, O. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2188–93.
- (31) Wang, H.; Sung, S. S. *Biopolymers* **1999**, *50*, 763–76.
- (32) Wu, X.; Wang, S.; Brooks, B. R. *J. Am. Chem. Soc.* **2002**, *124*, 5282–3.
- (33) Lei, H.; Smith, P. E. *Biophys. J.* **2003**, *85*, 3513–20.
- (34) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719–22.
- (35) Kentsis, A.; Mezei, M.; Gindin, T.; Osman, R. *Proteins* **2004**, *55*, 493–501.
- (36) Walsh, S. T. R.; Cheng, R. P.; Wright, W. W.; Alonso, D. O. V.; Daggett, V.; Vanderkooi, J. M.; DeGrado, W. F. *Protein Sci.* **2003**, *12*, 520–531.
- (37) Starzyk, A.; Barber-Armstrong, W.; Sridharan, M.; Decatur, S. M. *Biochemistry* **2005**, *44*, 369–76.

- (38) Zelent, B.; Nucci, N. V.; Vanderkooi, J. M. *J. Phys. Chem. A* **2004**, *108*, 11141–11150.
- (39) Zelent, B.; Kaposi, A. D.; Nucci, N.; Sharp, K.; Dalosto, S. D.; Wright, W. W.; Vanderkooi, J. *J. Phys. Chem. B* **2004**, *108*, 10317–24.
- (40) Kaposi, A. D.; Prabhu, N. V.; Dalosto, S. D.; Sharp, K. A.; Wright, W. W.; Stavrov, S. S.; Vanderkooi, J. M. *Biophys. Chem.* **2003**, *106*, 1–14.
- (41) Chagnenet-Barret, P.; Choma, C.; Gooding, E.; DeGrado, W. F.; Hochstrasser, R. M. *J. Phys. Chem. B* **2000**, *104*, 9322–29.
- (42) Peon, J.; Pal, S. K.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10964–9.
- (43) Kamal, J. K.; Zhao, L.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13411–6.
- (44) Nachliel, E.; Pollak, N.; Huppert, D.; Gutman, M. *Biophys. J.* **2001**, *80*, 1498–506.
- (45) Halle, B.; Denisov, V. P. *Methods Enzymol.* **2001**, *338*, 178–201.
- (46) Kutysenko, V. P.; Cortijo, M. *Protein Sci.* **2000**, *9*, 1540–7.
- (47) Modig, K.; Liepinsh, E.; Otting, G.; Halle, B. *J. Am. Chem. Soc.* **2004**, *126*, 102–14.
- (48) Modig, K.; Rademacher, M.; Lucke, C.; Halle, B. *J. Mol. Biol.* **2003**, *332*, 965–77.
- (49) Modig, K.; Kurian, E.; Prendergast, F. G.; Halle, B. *Protein Sci.* **2003**, *12*, 2768–81.
- (50) Sitkoff, D.; Lockhart, D. J.; Sharp, K. A.; Honig, B. *Biophys. J.* **1994**, *64*, 1–10.
- (51) Sitkoff, D.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.
- (52) Sitkoff, D.; Sharp, K. A.; Honig, H. *Biophys. Chem.* **1994**, *51*, 397–409.
- (53) Misra, V.; Hecht, J.; Sharp, K.; Friedman, R.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 264–280.
- (54) Misra, V.; Sharp, K.; Friedman, R.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 245–263.
- (55) Wang, T.; Tomic, S.; Gabdouliline, R. R.; Wade, R. C. *Biophys. J.* **1618**, *87*, 1618–30.
- (56) Dong, F.; Vijayakumar, M.; Zhou, H. X. *Biophys. J.* **2003**, *85*, 49–60.
- (57) Arbuzova, A.; Wang, L.; Wang, J.; Hangyas-Mihalyne, G.; Murray, D.; Honig, B.; McLaughlin, S. *Biochemistry* **2000**, *39*, 10330–9.
- (58) Luo, H.; Sharp, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10399–10404.
- (59) Sims, P. A.; Wong, C. F.; McCammon, J. A. *J. Comput. Chem.* **1416**, *25*, 1416–29.
- (60) Tsuchiya, Y.; Kinoshita, K.; Nakamura, H. *Proteins* **2004**, *55*, 885–94.
- (61) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2004**, *55*, 689–704.
- (62) Liang, Z. X.; Kurnikov, I. V.; Nocek, J. M.; Mauk, A. G.; Beratan, D. N.; Hoffman, B. M. *J. Am. Chem. Soc.* **2004**, *126*, 2785–98.
- (63) Still, C.; Tempczyk, A.; Hawley, R.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (64) Dominy, B.; Brooks, C. J. *J. Phys. Chem. B* **1999**, *103*, 3765–73.
- (65) Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 9571–9576.
- (66) Tsui, V.; Case, D. *J. Am. Chem. Soc.* **2000**, *122*, 2489–98.
- (67) Qui, D.; Shenkin, P.; Hollinger, F.; Still, W. *J. Phys. Chem.* **1997**, *101*, 3005–14.
- (68) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243–252.
- (69) Feig, M.; Brooks, C. L., 3rd. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–24.
- (70) Xia, B.; Tsui, V.; Case, D. A.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **2002**, *22*, 317–31.
- (71) Jaramillo, A.; Wodak, S. J. *Biophys. J.* **2005**, *88*, 156–71.
- (72) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133–152.
- (73) Jiang, L.; Kuhlman, B.; Kortemme, T.; Baker, D. *Proteins* **2005**, *58*, 893–904.
- (74) Mallick, P.; Weiss, R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16041–6.
- (75) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479–99.
- (76) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004**, *56*, 310–21.
- (77) Warshel, A.; Creighton, S. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W., Weiner, P., Eds.; ESCOM: Leiden, 1989.
- (78) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Comput. Chem.* **2004**, *25*, 1015–29.
- (79) Morreale, A.; de la Cruz, X.; Meyer, T.; Gelpi, J. L.; Luque, F. J.; Orozco, M. *Proteins* **2004**, *57*, 458–67.
- (80) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244–1253.
- (81) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59–75.
- (82) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608–640.
- (83) Lu, B. Z.; Chen, W. Z.; Wang, C. X.; Xu, X. J. *Proteins* **2002**, *48*, 497–504.
- (84) Lu, Q.; Luo, R. *J. Chem. Phys.* **2003**, *119*, 11035–47.
- (85) Fogolari, F.; Brigo, A.; Molinari, H. *Biophys. J.* **2003**, *85*, 159–66.
- (86) Prabhu, N. V.; Zhu, P.-J.; Sharp, K. A. *J. Comput. Chem.* **2004**, *25*, 2049–64.
- (87) Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1504–1514.
- (88) Yeh, Y.; Feeney, R. E. *Chem. Rev.* **1996**, *96*, 601–617.
- (89) Madura, J. D.; Baran, K.; Wierzbicki, A. *J. Mol. Recognit.* **2000**, *13*, 101–13.
- (90) Harding, M. M.; Anderberg, P. I.; Haymet, A. D. *Eur. J. Biochem.* **2003**, *270*, 1381–92.
- (91) Wierzbicki, A.; Madura, J. D.; Salmon, C.; Sonnichsen, F. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1006–10.
- (92) Antson, A. A.; Smith, D. J.; Roper, D. I.; Lewis, S.; Caves, L. S.; Verma, C. S.; Buckley, S. L.; Lillford J. *Mol. Biol.* **2001**, *305*, 875–89.
- (93) Strom, C. S.; Liu, X. Y.; Jia, Z. *J. Biol. Chem.* **2004**, *279*, 32407–17.
- (94) Jorov, A.; Zhorov, B. S.; Yang, D. S. *Protein Sci.* **2004**, *13*, 1524–37.
- (95) Cheng, Y.; Yang, Z.; Tan, H.; Liu, R.; Chen, G.; Jia, Z. *Biophys. J.* **2002**, *83*, 2202–10.
- (96) Baardsnes, J.; Jelokhani-Niaraki, M.; Kondejewski, L. H.; Kuiper, M. J.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **2001**, *10*, 2566–76.
- (97) Haymet, A. D.; Ward, L. G.; Harding, M. M. *FEBS Lett.* **2001**, *491*, 285–8.
- (98) Gallagher, K.; Sharp, K. A. *Biophys. Chem.* **2003**, *105*, 195–209.
- (99) Yang, C.; Sharp, K. *Proteins* **2005**, *59*, 266–274.
- (100) Yang, C.; Sharp, K. A. *Biophys. Chem.* **2004**, *109*, 137–148.
- (101) Cheng, A.; Merz, K. M., Jr. *Biophys. J.* **1997**, *73*, 2851–73.

CR040437F