Theoretical Prediction of Changes in Protein Structural Stability upon Cosolvent or Salt Addition and Amino-acid Mutation

Shota Murakami

Department of Fundamental Energy Science
Graduate School of Energy Science
Kyoto University

February 2017

Table of contents

Chapter 3
Effects of Salts or Cosolvent Addition on Solubility of a Hydrophobic Solute in Water: Relevance to Those on Thermal Stability of a Protein

Chapter 1

General Introduction

1.1. Scope and significance of this study

Proteins play a variety of imperative roles in sustaining life. The function of a protein is closely related to its three-dimensional structure which is referred to as "native structure", and it is of fundamental importance to elucidate the mechanism of structure formation, i.e., protein folding. Though a number of experimental, theoretical, and computer simulation studies have been performed, there are still lots of uncertain and controversial aspects to overcome for the elucidation. On the other hand, it is experimentally known that a protein is denatured (unfolded) by raising or lowering the temperature[1-3] and by applying high pressure.[4-6] The stability of the native structure is significantly influenced by a modification of solvent environment or by an amino-acid mutation. Here, the modification signifies the addition of cosolvents or salts to water. It is desired that all of the mechanisms of folding, denaturation (unfolding), and change in structural stability be explicated in a unified manner within the same theoretical framework: A theory which can elucidate a particular subject but fails to elucidate the others, for example, is not a reliable one.

Kinoshita and coworkers suggested that a protein is driven to fold by a large gain of the configurational entropy of water.[7-11] The protein insertion into water generates a space which the centers of water molecules cannot enter. The volume of this space is referred to as "excluded volume (EV)". A water molecule also generates an EV for *the other water molecules*, and in this sense all of the water molecules are entropically correlated. This correlation is referred to as "water crowding".[11-13] Upon protein folding, the EV generated by the protein reduces, leading to an increase in the total volume available to the translational displacement of water molecules coexisting with the protein in the system, which is followed by a mitigation of the water crowding. The large gain of water entropy mentioned above is ascribed primarily to this mitigation. The negative value of the solvation entropy of a solute is attributed to the increased water crowding which is a dominant factor of the solute hydrophobicity (see Sec. 1.3 for more details). Kinoshita and coworkers were also successful in elucidating the mechanisms of heat,[13] cold,[12,13] and pressure[13,14] denaturating of a protein using a theory based on statistical mechanics wherein the water-entropy effect described above (this is referred as "entropic EV effect") is treated as the key factor. A remaining issue to be tackled is the structural-stability change caused by a modification of solvent environment or by an amino-acid mutation. This issue is essential not only from a scientific viewpoint but also from a practical one. The cosolvent or salt addition can enhance the thermal stability of a protein.[15-24] The same result is also achievable by an amino-acid mutation. If an enzyme protein becomes thermally more stable, for example, it can be used as a catalyst at a higher temperature, leading to a significant acceleration of the chemical reaction. An example situation where such thermostabilization is required is found in a process of biomass energy generation: Thermostabilization of β-glucosidase, which catalyzes the hydrolysis of β-glycosidic bond in a

sugar molecule, is strongly desired.

It is experimentally known that the addition of sugars such as glucose and sucrose enhances the thermal stability of a protein,[17-19,22] whereas that of urea or guanidine hydrochloride lowers it.[16,20,21,24] The interesting effects of alcohols can be summarized as follows: (1) The addition of a monohydric alcohol lowers the stability and the degree of the lowering becomes stronger as the size of hydrophobic group in an alcohol molecule increases[16,19,25,26] and (2) the addition of a polyol possessing two or more hydroxyl groups per molecule enhances the stability and the enhancement becomes stronger as the number of hydroxyl groups increases.[17,19,27-29] The thermal stability can be influenced by the cosolvent addition due to the following factors: direct protein-cosolvent interactions such as hydrogen bonds and electrostatic and van der Waals attractive interactions; and indirect effects through the changes in water structure. However, detailed mechanisms of the cosolvent effects have not yet been understood. The effects of adding salts are quite complex even when alkali halides, the simplest salts, are considered: The effects are qualitatively dependent on the protein species unlike those of adding sugars, monohydric alcohols, and polyols. A salt works as a stabilizer for some proteins, whereas it works as a destabilizer for other proteins. One associates the so-called Hofmeister series[30] with the salt effects. The Hofmeister series is one of the most intricate subjects in modern chemical physics of aqueous solution that has not been solved for over one hundred years. For alkali and halide ions, the series was originally expressed in the orders $Li^+ > Na^+ > K^+$ and $Cl^- > Br^- > I^-$ in terms of their salting-out ability for hen egg white protein[30] ("salting out" signifies a decrease in solute solubility in water upon salt addition). Though the solubility data for significantly many nonpolar, polar, and charged solutes share *similar* orders, they are not quite the same: The series is often expressed in more or less different orders depending on the solute species and the physicochemical quantity considered. The series applicable to the thermal stability or structural transition of a protein is quite variable: The order for cations is not definite, and that for anions can even be reversed to $I^- > Br^- > Cl^-$.[31-33]

A molecular dynamics (MD) simulation using a realistic, all-atom model has been the most popular tool for elucidating the mechanism of the cosolvent- or salt-induced change in thermal stability.[34-46] It allows us to calculate the average number of solvent-protein hydrogen bonds, solvent number density profiles near the protein, solvent-protein interaction energy, and mobilities of solvent molecules in the bulk. However, it suffers from the following drawbacks: (1) It is not good at calculating the changes in thermodynamic quantities of solvation upon the cosolvent or salt addition; and (2) the result is often difficult to interpret because it stems from complex interplay of multiple physicochemical factors. The new MD simulation combined with the energy representation method,[47-52] which has recently been developed, is considerably more efficient than the usual one. It is capable of calculating the solvation free energy of a protein with zero net charge. However, when a protein possesses nonzero net charge, the calculation becomes problematic. Moreover, it still requires heavy computational burden.

Another approach is a theoretical method based on statistical mechanics. The most useful statistical-mechanical method is the integral equation theory[53-58] (IET). In the IET, from the system

partition function, various correlation functions are defined, and the basic equations satisfied by them are derived. The many-body correlations are also approximately taken into account. The average value of a physical quantity is calculated for an infinitely large system and an infinitely large number of system configurations. In the case of bulk solvent of a single component, for example, the temperature, number density, and interaction potential form the input data. By numerically solving the basic equations to obtain the correlation functions, we can calculate the microscopic structure and thermodynamic quantities. A molecular solvent with angle-dependent potential such as water can also be treated, in which case the basic equations must be reduced mathematically before the numerical treatment. A solvent comprising multiple components can readily be treated. Thermodynamic quantities of solvation, that is, changes in thermodynamic quantities upon solute insertion into the solvent, can also be calculated. When the IET is combined with the morphometric approach[59,60] (MA) developed by Kinoshita and coworkers, the combined method[11-14] becomes very powerful because it can treat a large solute with complex polyatomic structure immersed in simple fluid or in water and enables us to decompose a thermodynamic quantity of solvation into its physically insightful constituents. The combined method is best suited to the elucidation of the cosolvent or salt effects. It should be emphasized that the changes in thermodynamic quantities of solvation upon the cosolvent or salt addition and their physically insightful constituents give much more informative results than the items calculated by the MD simulation mentioned above. When the IET-MA combination is applied to a simplified model in which only a particular, expectedly pivotal factor is (or only some of the selected factors are) incorporated, it enables us to gain physical insights. In the first step, we explore to what extent it can reproduce the experimental data. It fails to reproduce some of the data. In the second step, we think about the reason for the failure and identify a factor or factors to be considered further.

A mutation for a protein usually results in only a slight structural change of the protein, and its denaturation temperature is correspondingly raised or lowered. Only a rather small percentage of the possible mutations bring enhanced thermal stability. The theoretical prediction of the thermal-stability change upon a mutation is a subtle task. Up to now, significantly many approaches for the prediction have been reported in the literature.[61-68] They always include parameters which are adjusted so that the prediction results can be best fitted to the experimental data for a sufficiently large set of proteins and mutations.[62,64,66-68] The inclusion is necessitated to achieve satisfactorily high prediction performance. A problem is that the resulting values of the parameters are often physically meaningless, and the physicochemical factors governing the thermal-stability change upon mutation remain rather ambiguous. It is strange that none of the previously reported prediction approaches takes account of the solvent-entropy effect emphasized by Kinoshita and coworkers to its full extent.

In this study, we employ the IET-MA combination and challenge the following subjects: the elucidation of the effects of cosolvent or salt addition on the thermal stability of a protein[15-24]; clarification of physical origins of the Hofmeister series,[30] and development of a theoretical method for predicting the thermal-stability changes upon single and multiple mutations for a protein without

using any parameters fitted to the experimental data. First, we adopt a rigid-body model where the solvent particles and a protein are treated as hard spheres and as a set of fused hard spheres accounting for the polyatomic structure, respectively. In this model, all of the accessible system configurations share the same energy and the system behavior is purely entropic in origin. The entropic effects of translational displacement of water and cosolvent molecules and ions can well be described by the rigid-body model (see Sec. 1.4 for more details). However, this model is not capable of directly accounting for the other factors such as the protein-water, -cosolvent, and -ion interaction potentials. We examine to what extent the incorporation of only the entropic EV effect can reproduce the experimental observations. When it fails, the factors to be considered further are identified by physically meaningful argument with the help of additional calculations using another, more realistic model.[53,54] We show that the effects of cosolvent or salt addition are governed by the change in the degree of solvent crowding in most cases. Just for reproducing the effects of adding urea for every protein[16,20,21,24] and anions with large sizes (i.e., Br^- and I^-) for a protein with significantly large positive net charge,[23] the protein-urea and -anion van der Waals interactions come into play and must be taken into account. When urea is added to water, some of the water molecules near the protein surface are replaced by urea molecules due to the enrichment of urea near it.[69,70] However, the total protein-water, protein-urea, water-water, water-urea, and urea-urea electrostatic interaction energy in water-urea solution is not significantly different from the total protein-water and water-water electrostatic interaction energy in pure water.[52] The same argument can be made for the large anions. As a remarkable progress, we present a new view on the Hofmeister series. This success is understandable, because the change in the entropic EV effect should be the key factor in describing the result of modifying the solvent side. By contrast, it is significantly better to take account of the enthalpic component as well as the entropic EV effect in describing the result of modifying the protein side: the thermal-stability change upon a mutation for a protein. The IET-MA combination has not yet been employed by any other research group. We hope that it will become a popular method for analyzing a variety of self-assembly processes. The outlines of Chapters 2 through 4 in this thesis are given in Sec. 1.5.

1.2. Driving force of protein folding

The following physicochemical factors have been proposed as the driving forces in protein folding: the gain of protein intramolecular hydrogen bonds (factor 1), gain of intramolecular van der Waals attractive interactions (factor 2), and burial of nonpolar groups (factor 3). However, factor 1 is unavoidably accompanied by the loss of protein-water hydrogen bonds. Factor 2 is also accompanied by the loss of protein-water van der Waals attractive interactions. The net loss is significantly larger than the net gain, which is manifested by the experimental result[8] that the enthalpy change upon apoplastocyanin (apoPC) folding at 298 K takes a large, positive value. This result indicates that the gain of water entropy is large enough to surpass a large loss of protein intramolecular entropy. The suggestion of factor 3 is based on the view that the water near a nonpolar group is entropically unstable due to its structuring (i.e., increased number of and

strengthened hydrogen bonds) and the release of such unstable water to the bulk upon protein folding gives a gain of water entropy.[71,72] Kinoshita and coworkers showed that the water-entropy gain originating from this view is certainly present but too small to elucidate the very large water-entropy gain experimentally shown[8].

Kinoshita and coworkers argued that the driving force in protein folding is the water-entropy gain arising from the translational displacement of water molecules in the system[9-11] (not limited to the water molecules near the protein surface). As illustrated in Fig. 1.1, the formation of α-helix by a portion of the backbone or that of β-sheet by portions of the backbone leads to a decrease in the total EV,[73] leading to a gain of water entropy. When the α-helix or the β-sheet is formed, not only a water-entropy gain occurs but also the energetically unfavorable break of hydrogen bonds *with water molecules* is compensated with the formation of intramolecular hydrogen bonds. Hence, these secondary structures (α-helix and β-sheet) are very advantageous units to be formed as much as possible in protein folding: This is why they frequently appear in the native structure. The close packing of side chains is also crucially important because it is followed by a large gain of water entropy.[7,73] Upon protein folding, the water crowding in the bulk is largely reduced by a decrease in the EV generated by the protein. This reduction is a principal contributor to the very large water-entropy gain driving a protein to fold. (Water possesses not only the translational entropy but also the orientational (rotational) entropy. Kinoshita and coworkers have shown that the translational contribution predominates over the orientational one.[8,58] For instance, the translational and orientational contributions to the water-entropy gain upon apoPC folding are ~95% and ~5%, respectively.[8])

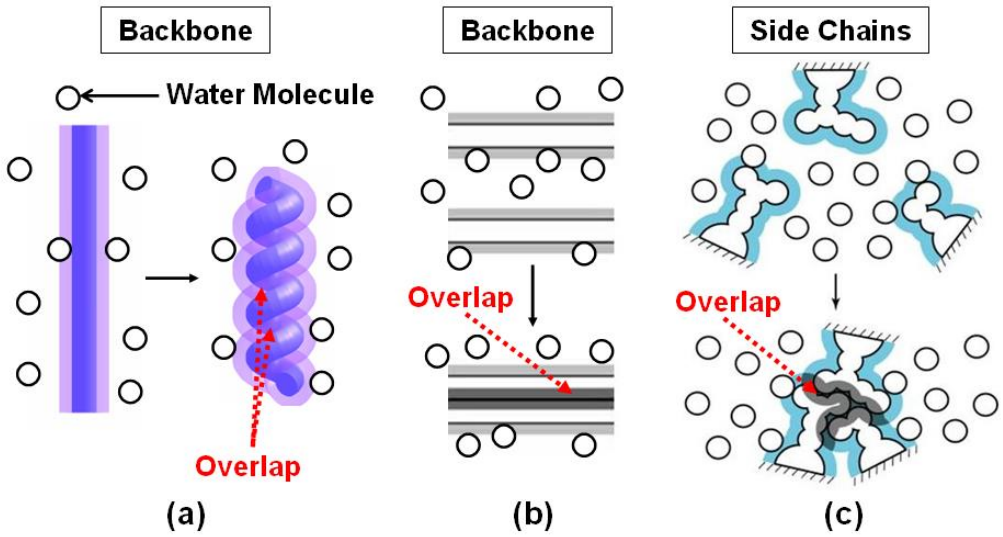

Fig. 1.1. (a) Formation of α-helix by a portion of the backbone. (b) Formation of β-sheet by portions of the backbone. (c) Close packing of side chains. Overlap of an excluded space in (a) or excluded spaces in (b) and (c) occur, and the total volume available to the translational displacement of water molecules increases by the volume of overlapped space.

1.3. Physical origin of hydrophobic effect

It is well recognized that the hydrophobic effect is essential in a variety of self-assembly processes. The structures formed by the processes are collapsed by the application of high pressures: A protein is denatured,[2,74] filamentous actin (F-actin) is dissociated into actin monomers,[75,76] amyloid fibril is destructed,[77] virus assemblies are dissociated[78] at high pressures. The power of forming the structures becomes considerably weaker when the temperature is lowered: A protein is unfolded, binding of myosin to F-actin is weakened,[79] protein aggregation is dissociated,[80] and for nonionic amphiphilic molecules the critical micelle concentration becomes higher and the average size of micelles becomes smaller[81] at low temperatures. These phenomena suggest that there are common features of the self-assembly processes and a certain physical factor universally plays a dominant role as the driving force. Kinoshita and coworkers proposed a new view claiming that this physical factor is the reduction in water crowding in the entire system upon a self-assembly process and nothing but the true origin of the hydrophobic effect.[11-13,58] In this sense, protein folding is driven by the hydrophobic effect. The pressure and temperature dependences of the hydrophobic effect described above can be elucidated only by this new view.

1.4. Simplified model for exclusive investigation of the solvent-entropy effect

As long as a theoretical analysis is focused on the configurational entropy originating from the translational displacement of water molecules, water can be modeled as a hard-sphere solvent on the condition that the solvent diameter and number density are set at the values pertinent to water. In what follows, the calculation results obtained by Kinoshita and coworkers are described. The hydration free energy μ, entropy S, and energy U under isochoric condition for a spherical solute are calculated using the angle-dependent integral equation theory[53-56,58] applied to a multipolar model for water.[53,54] For the hard-sphere solute with zero charge, the calculated values are μ=5.95k_BT, S=–9.22k_B, and U=–3.27k_BT (μ=U–TS). When a point charge of –0.5e (e is the electronic charge) is embedded at its center, the calculated values become μ=–32.32k_BT, S=–10.11k_B, and U=–42.43k_BT. Thus, μ and U are largely influenced by the solute-water interaction potential, whereas S is fairly insensitive to it. Even when the solvent is replaced by hard spheres whose number density and diameter are those of water, the result is the following: μ=9.64k_BT and S=–9.64k_B (U=0). Thus, the solvation entropy can approximately be evaluated even by modeling water as hard spheres. The values of S are calculated using the three-dimensional reference interaction site model (3D-RISM) theory for the native structures of a total of eight peptides and proteins.[82] Realistic all-atom (Coulomb plus Lennard-Jones (LJ)) potentials are employed for the peptides, proteins, and water. Even when the protein-water electrostatic potentials, which are quite strong, are shut off and only the LJ potentials are retained, $|S|$ decreases only by ~5%. Therefore, when an analysis is focused on the solvent-entropy effect, a peptide or protein can simply be modeled as a set of fused hard spheres with no partial charges.

An exception is the case where the temperature dependence of the effect plays important roles. For example, the weakening of the hydrophobic effect at low temperatures[58] and cold

denaturation[12,83,84] of a protein cannot be elucidated by the hard-sphere model for water. The temperature dependence of the effect is ascribed to the interplay of strongly attractive interaction (i.e., hydrogen bonding) and exceptionally small molecular size of water. In this study, however, the temperature dependence is not discussed with the result that the rigid-body model provides exclusive investigation of the solvent-entropy effect.

1.5. Synopsis of this thesis

In Chapter 2, we consider the effects of monohydric alcohols and polyols as the cosolvents on the thermal stability of a protein. The solvent is either pure water or water-cosolvent solution. The solvent-entropy gain upon protein folding, which is calculated at 298 K and normalized by the number of residues, is adopted as a measure of the thermal stability. A larger measure implies higher thermal stability. It is demonstrated that all of the experimental observations mentioned in Sec. 1.1 can be reproduced even in a quantitative sense by a rigid-body model focused on the entropic effect originating from the translational displacement of solvent molecules. In the rigid-body model, a protein is a fused hard spheres accounting for the polyatomic structure in the atomic detail, and the solvent is formed by hard spheres or a binary mixture of hard spheres with different diameters. The effective diameter of cosolvent molecules and the packing fractions of water and cosolvent, which are crucially important parameters, must carefully be estimated using the experimental data of properties such as the density of solid crystal of cosolvent, parameters in the pertinent cosolvent-cosolvent interaction potential, and density of water-cosolvent solution. It is argued that the degree of solvent crowding in the bulk is the key factor. When it is made more serious by the cosolvent addition, the solvent-entropy gain upon protein folding is magnified, leading to the enhanced thermal stability. When it is made less serious, the opposite is true. The mechanism of the effects of monohydric alcohols and polyols is physically the same as that of sugars. However, when the rigid-body models are employed for the effect of urea, its addition is predicted to enhance the thermal stability, which conflicts with the experimental fact. This conflict is further investigated in Chapter 3.

In Chapter 3, we revisit the effects cosolvents on the thermal stability of a protein. Those of salts are also investigated. We consider the solubility of a small nonpolar solute in water which is much simpler and more fundamental than the thermal stability. The solubility is changed upon addition of a salt or cosolvent. Hereafter, “solvent” is formed by water molecules for pure water, by water molecules, cations, and anions for water-salt solution, and by water and cosolvent molecules for water-cosolvent solution. For the small nonpolar solute, the hydrophobic effect predominates over the other physicochemical factors. Decrease and increase in the solubility, respectively, are ascribed to enhancement and reduction of the hydrophobic effect. A plenty of experimental data are available for the change in solubility of argon or methane arising from the addition. We show that the integral equation theory combined with a rigid-body model, in which the solute and solvent particles are modeled as hard spheres with different diameters, can reproduce the data for the following items: salting out by an alkali halide and salting in by tetramethylammonium bromide

("salting in" signifies an increase in solute solubility in water upon salt addition), increase in solubility by a monohydric alcohol, and decrease in solubility by sucrose or urea. The orders of cation or anion species in terms of the power of decreasing the solubility can also be reproduced for alkali halides. With the rigid-body model, the analyses are focused on the roles of entropy originating from the translational displacement of solvent particles. It is argued by decomposing the solvation entropy of a nonpolar solute into physically insightful constituents that the solvent crowding in the bulk is a pivotal factor of the hydrophobic effect: When the solvent crowding in the bulk becomes more serious, the effect is strengthened, and when it becomes less serious, the effect is weakened. It is experimentally known that the thermal stability of a protein is also influenced by the salt or cosolvent addition. The additions which decrease and increase the solubility of a nonpolar solute, respectively, usually enhance and lower the thermal stability. This suggests that the enhanced or reduced hydrophobic effect is also a principal factor governing the stability change. However, urea decreases the solubility but lowers the stability. Bromide and iodide ions decrease the solubility but lower the stability of a protein with a large, positive total charge. In these cases, the urea- or ion-protein van der Waals interaction energy as well as the hydrophobic effect needs to be taken into account in arguing the stability change. All of the experimental data can be interpreted by the two principal factors: the change in solvent crowding and the cosolvent- or ion-protein van der Waals interaction energy. The achievements are as follows: The validity of our view concerning the hydrophobic effect explained in Sec. 1.3 is corroborated; and we present a new view on the Hofmeister series. It is shown that when the hydrophobic effect is dominant as in the salting-out case for a nonpolar solute, the Hofmeister series for alkali and halide ions are expressed by $Na^+ > K^+ > Cs^+ > Li^+$ and $Cl^- > Br^- > I^-$, respectively. We also argue how it is modified when other factors are also influential.

In Chapter 4, we develop a new measure of the thermal stability of a protein, which takes account of the enthalpic component as well. We employ a realistic, multipolar model for water and the combination of the angle-dependent integral equation theory and the MA. As explained in Sec. 1.2, protein folding is accompanied by a large gain of water entropy (the entropic EV effect), loss of protein conformational entropy, and increase in enthalpy. The enthalpy increase originates primarily from the following: The energy increase due to the break of protein-water hydrogen bonds (HBs) upon folding cannot completely be cancelled out by the energy decrease brought by the formation of protein intramolecular HBs. We develop the measure on the basis of these three factors and apply it to the prediction of the thermal-stability change upon mutation. As a consequence, an approach toward the prediction is constructed. It is distinguished from the previously reported approaches in the following respects: The parameters adjusted in the manner mentioned above are not employed at all; and the entropic EV effect, which is ascribed to the translational displacement of water molecules coexisting with the protein in the system, is fully taken into account using a molecular model for water. Our approach is compared with one of the most popular approaches, FOLD-X, in terms of the prediction performance not only for single mutations but also for double, triple, and higher-fold (up to seven-fold) mutations. It is shown that on the whole our approach and FOLD-X

exhibit almost the same performance despite that the latter uses the adjusting parameters. For multiple mutations, however, our approach is far superior to FOLD-X. Five multiple mutations for staphylococcal nuclease lead to highly enhanced stabilities, but we find that this high enhancement arises from the entropic EV effect. The neglect of this effect in FOLD-X is a principal reason for its ill success. A conclusion is that the three factors mentioned above play essential roles in elucidating the thermal-stability change upon mutation. (The major conclusions drawn in this thesis are recapitulated in Chapter 5.)

References

[1]P. L. Privalov, Y. V. Griko, and S. Y. Venyaminov, J. Mol. Biol. **190**, 487 (1986).

[2]A. Pastore, S. R. Martin, A. Politou, K. C. Kondapalli, T. Stemmler, and P. A. Temussi, J. Am. Chem. Soc. **129**, 5374 (2007).

[3]P. L. Privalov, Pure Appl. Chem. **79**, 1445 (2007).

[4]C. Cléry, F. Renault, and P. Masson, FEBS Lett. **370**, 212 (1995).

[5]R. Kitahara, S. Yokoyama, and K. Akasaka, J. Mol. Biol. **347**, 277 (2005).

[6]F. Meersman, C. M. Dobson, and K. Heremans, Chem. Soc. Rev. **35**, 908 (2006).

[7]Y. Harano and M. Kinoshita, Biophys. J. **89**, 2701 (2005).

[8]T.Yoshidome, M. Kinoshita, S. Hirota, N. Baden, and M. Terazima, J. Chem. Phys. **128**, 225104 (2008).

[9]M. Kinoshita, Front. Biosci. **14**, 3419 (2009).

[10]M. Kinoshita, Int. J. Mol. Sci. **10**, 1064 (2009).

[11]M. Kinoshita, Biophys. Rev. **5**, 283 (2013).

[12]T. Yoshidome and M. Kinoshita, Phys. Chem. Chem. Phys. **14**, 14554 (2012).

[13]H. Oshima and M. Kinoshita, J. Chem. Phys. **142**, 145103 (2015).

[14]Y. Harano, T. Yoshidome, and M. Kinoshita, J. Chem. Phys. **129**, 145103 (2008).

[15]P. H. von Hippel and K.-Y. Wong, Science **145**, 577 (1964).

[16]P. H. von Hippel and K.-Y. Wong, J. Biol. Chem. **240**, 3909 (1965).

[17]J. F. Back, D. Oakenfull, and M. B. Smith, Biochemistry **18**, 5191 (1979).

[18]H. Uedaira and H. Uedaira, Bull. Chem. Soc. Jpn. **53**, 2451 (1980).

[19]K. Gekko and S. Koga, J. Biochem. **94**, 199 (1983).

[20]V. K. Dubey and M. V. Jagannadham, Biochemistry **42**, 12287 (2003).

[21]J. D. Batchelor, A. Olteanu, A. Tripathy, and G. J. Pielak, J. Am. Chem. Soc. **126**, 1958 (2004).

[22]N. K. Poddar, Z. A. Ansari, R. K. B. Singh, A. A. Moosavi-Movahedi, and F. Ahmad, Biophys. Chem. **138**, 120 (2008).

[23]E. Sedlák, L. Stagg, and P. Wittung-Stafshede, Arch. Biochem. Biophys. **479**, 69 (2008).

[24]A. Hédoux, S. Krenzlin, L. Paccou, Y. Guinet, M.-P. Flament, and J. Siepmann, Phys. Chem. Chem. Phys. **12**, 13189 (2010).

[25]A. L. Fink, Cryobiology **23**, 28 (1986).

[26]V. Bhakuni, Arch. Biochem. Biophys. **357**, 274 (1998).

[27]Y. Fujita, Y. Iwasa, and Y. Noda, Bull. Chem. Soc. Jpn. **55**, 1896 (1982).

[28]J. K. Kaushik and R. Bhat, J. Phys. Chem. B **102**, 7058 (1998).

[29]I. Haque, R. Singh, A. A. Moosavi-Movahedi, and F. Ahmad, Biophys. Chem. **117**, 1 (2005).

[30]F. Hofmeister, Arch. Exp. Pathol. Pharmakol. **24**, 247 (1888).

[31]Y. Goto, N. Takahashi, and A. L. Fink, Biochemistry **29**, 3480 (1990).

[32]Y. Goto and S. Nishikiori, J. Mol. Biol. **222**, 679 (1991).

[33]Y. Hagihara, M. Kataoka, S. Aimoto, and Y. Goto, Biochemistry **31**, 11908 (1992).

[34]F. Müller-Plathe and W. F. van Gunsteren, Polymer **38**, 2259 (1997).
[35]B. J. Bennion and V. Daggett, Proc. Natl. Acad. Sci. U. S. A. **100**, 5142 (2003).
[36]H.-L. Liu and C.-M. Hsu, J. Chin. Chem. Soc. **50**, 1235 (2003).
[37]A. Lerbret, P. Bordat, F. Affouard, A. Hédoux, Y. Guinet, and M. Descamps, J. Phys. Chem. B **111**, 9410 (2007).
[38]J. Dzubiella, J. Am. Chem. Soc. **130**, 14000 (2008).
[39]L. Hua, R. Zhou, D. Thirumalai, and B. J. Berne, Proc. Natl. Acad. Sci. U. S. A. **105**, 16928 (2008).
[40]V. Vagenende, M. G. S. Yap, and B. L. Trout, Biochemistry **48**, 11084 (2009).
[41]F.-F. Liu, L. Ji, L. Zhang, X.-Y. Dong, and Y. Sun, J. Chem. Phys. **132**, 225103 (2010).
[42]F. Mehrnejad, M. M. Ghahremanpour, M. Khadem-Maaref, and F. Doustdar, J. Chem. Phys. **134**, 035104 (2011).
[43]Q. Shao, Y. Fan, L. Yang, and Y. Q. Gao, J. Chem. Phys. **136**, 115101 (2012).
[44]Q. Shao, Y. Fan, L. Yang, and Y. Q. Gao, J. Chem. Theory Comput. **8**, 4364 (2012).
[45]W. J. Xie and Y. Q. Gao, Faraday Discuss. **160**, 191 (2013).
[46]S. Paul and S. Paul, J. Phys. Chem. B **119**, 10975 (2015).
[47]N. Matubayasi and M. Nakahara, J. Chem. Phys. **113**, 6070 (2000).
[48]N. Matubayasi and M. Nakahara, J. Chem. Phys. **117**, 3605 (2002); Erratum, **118**, 2446 (2003).
[49]N. Matubayasi and M. Nakahara, J. Chem. Phys. **119**, 9686 (2003).
[50]Y. Karino and N. Matubayasi, Phys. Chem. Chem. Phys. **15**, 4377 (2013).
[51]S. Sakuraba and N. Matubayasi, J. Comput. Chem. **35**, 1592 (2014).
[52]Y. Yamamori, R. Ishizuka, Y. Karino, S. Sakuraba, and N. Matubayasi, J. Chem. Phys. **144**, 085102 (2016).
[53]P. G. Kusalik and G. N. Patey, J. Chem. Phys. **88**, 7715 (1988).
[54]P. G. Kusalik and G. N. Patey, Mol. Phys. **65**, 1105 (1988).
[55]M. Kinoshita, D. R. Bérard, J. Compt. Phys. **124**, 230 (1996).
[56]N. M. Cann and G. N. Patey, J. Chem. Phys. **106**, 8165 (1997).
[57]J.-P. Hansen and L. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, London, 2006).
[58]M. Kinoshita, J. Chem. Phys. **128**, 024507 (2008).
[59]R. Roth, Y. Harano, and M. Kinoshita, Phys. Rev. Lett. **97**, 078101 (2006).
[60]R. Kodama, R. Roth, Y. Harano, and M. Kinoshita, J. Chem. Phys. **135**, 045103 (2011).
[61]A. Mozo-Villiarías and E. Querol, Curr. Bioinformatics **1**, 25 (2006).
[62]R. Guerois, J. E. Nielsen, and L. Serrano, J. Mol. Biol. **320**, 369 (2002).
[63]N. Pokala, and T. M. Handel, J. Mol. Biol. **347**, 203 (2005).
[64]S. Yin, F. Ding, and N. V. Dokholyan, Nat. Methods **4**, 466 (2007).
[65]V. Potapov, M. Cohen, and G. Schreiber, Protein Eng. Des. Sel. **22**, 553 (2009).
[66]A. Benedix, C. M. Becker, B. L. de Groot, A. Caflisch, and R. A. Böckmann, Nat. Methods **6**, 3 (2009).

[67]Z. Zhang, L. Wang, Y. Gao, J. Zhang, M. Zhenirovskyy, and E. Alexov, Bioinformatics **28**, 664 (2012).
[68]L. Wickstrom, E. Gallicchio, and R. M. Levy, Proteins **80**, 111 (2012).
[69]V. Prakash, C. Loucheux, S. Scheufele, M. J. Gorbunoff, and S. N. Timasheff, Arch. Biochem. Biophys. **210**, 455 (1981).
[70]S. N. Timasheff, Biochemistry **31**, 9857 (1992).
[71]W. Kauzmann, Adv. Protein Chem. **14**, 1 (1959).
[72]K. A. Dill, Biochemistry **29**, 7133 (1990).
[73]S. Yasuda, H. Oshima, and M. Kinoshita, J. Chem. Phys. **137**, 135103 (2012).
[74]P. L. Privalov, Crit. Rev. Biochem. Mol. Biol. **25**, 281 (1990).
[75]T. Ikkai and T. Ooi, Biochemistry **5**, 1551 (1966).
[76]P. S. Niranjan, P. B. Yim, J. G. Forbes, S. C. Greer, J. Dudowicz, K. F. Freed, and J. F. Douglas, J. Chem. Phys. **119**, 4070 (2003).
[77]D. Foguel, M. C. Suarez, A. D. Ferrão-Gonzales, T. C. R. Porto, L. Palmieri, C. M. Einsiedler, L. R. Andrade, H. A. Lashuel, P. T. Lansbury, J. W. Kelly, and J. L. Silva, Proc. Natl. Acad. Sci. U. S. A. **100**, 9831 (2003).
[78]C. F. S. Bonafe, C. M. R. Vital, R. C. B. Telles, M. C. Gonçalves, M. S. A. Matsuura, F. B. T. Pessine, D. R. C. Freitas, and J. Vega, Biochemistry **37**, 11097 (1998).
[79]S. Highsmith, Arch. Biochem. Biophys. **180**, 404 (1977).
[80]R. Mishra and R. Winter, Angew. Chem., Int. Ed. **47**, 6518 (2008).
[81]D. Myers, *Surfaces, Interfaces, and Colloids: Principles and Applications* (Wiley-VCH, Berlin, 1999).
[82]T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata, J. Chem. Phys. **125**, 024911 (2006).
[83]T. Yoshidome and M. Kinoshita, Phys. Rev. E **79**, 030905(R) (2009).
[84]H. Oshima, T.Yoshidome, K. Amano, and M. Kinoshita, J. Chem. Phys. **131**, 205102 (2009).

Chapter 2

Effects of Addition of Sugars, Monohydric Alcohols, Polyols on the Thermal Stability of a Protein

2.1. Introduction

It is experimentally known that the cosolvent addition enhances or lowers the thermal stability of a protein. For example, the addition of sugars such as glucose and sucrose raises the thermal denaturation temperature T_m,[1–4] whereas that of urea or guanidine hydrochloride lowers T_m.[5–8] The effects of alcohols as cosolvents are particularly interesting: (1) The addition of a monohydric alcohol lowers the stability and the degree of the lowering becomes stronger as the size of hydrophobic group in an alcohol molecule increases[3,5,9,10] and (2) that of a polyol possessing two or more hydroxyl groups per molecule enhances the stability and the enhancement becomes stronger as the number of hydroxyl groups increases.[1,3,11–13] The degree of lowering or enhancement becomes larger as the cosolvent concentration increases.[1,3,5,9–13] There can be direct and indirect effects which are responsible for the thermal-stability change. Through the direct effect, cosolvent molecules form hydrogen bonds with a protein or come in contact with it through electrostatic and van der Waals attractive interactions. Consequently, the water-protein and cosolvent-protein interaction energy in water-cosolvent solution becomes different from the water-protein interaction energy in pure water. On the other hand, the cosolvent addition changes the water structure in the bulk, having the indirect effect on the thermal stability. Hereafter, "solvent" represents "water" for pure water and "water and cosolvent" for water-cosolvent solution.

An all-atom molecular dynamics (MD) simulation has been the most popular tool employed for elucidating the mechanisms of the cosolvent-induced change in thermal stability.[14–23] Principal concerns in the simulation are the average number of solvent-protein hydrogen bonds, solvent number density profiles near the protein, solvent-protein interaction energy, and mobilities of solvent molecules in the bulk. Analyzing changes in the solvation free energy and its components upon the cosolvent addition and identifying their microscopic origins are expected to give the most informative results, but the computational burden required becomes quite heavy when the simulation is employed in such an analysis. The MD simulation combined with the energy representation method[24–29] (the MD-ER simulation) is much less time consuming and has recently been applied to the elucidation of protein denaturation caused by urea.[29] In the present study, we adopt a statistical-mechanical theory combined with a simplified model system focused on a particular physical factor. We analyze the effects of a total of nine cosolvent species at the same time, which is difficult to perform even by the MD-ER simulation.

The presence of a water molecule generates an excluded volume for the other water molecules. Thus, water molecules in the system are entropically correlated, which is referred to as "water crowding". Upon protein folding, the excluded volume generated by the protein exhibits a large

decrease, leading to reduced water crowding followed by a water-entropy gain. Protein folding is accompanied by a large gain of water entropy,[30–33] and the reduction of water crowding in the bulk is the largest contributor to the gain.[34,35] In Kinoshita and coworker's earlier work,[36] they investigated the effect of sugar addition on the thermal stability of a protein. The protein was modeled as a set of fused hard spheres accounting for its polyatomic structure and water-sugar solution was formed by a binary mixture of hard spheres with different diameters. A hybrid of the integral equation theory[37] and the morphometric approach[38,39] was employed. With the rigid-body models, all of the accessible system configurations share the same energy, and the system behavior is purely entropic in origin. The rigid-body models are best suited to the exclusive investigation of the solvent-entropy effect. They succeeded in reproducing the experimental observation that a protein becomes thermally more stable as the sugar concentration becomes higher and the stabilization effect is stronger for sucrose than for glucose.[36] The solvent-entropy gain upon protein folding, which is calculated at 298 K under the isochoric condition, was shown to be a good measure of the thermal stability.[36,40] Their conclusion was that when a sugar is added, solvent crowding becomes more serious and its reduction upon protein folding is magnified: Due to this effect, the thermal stability of a protein is enhanced.[36] Graziano[41] reached the same conclusion on the sugar addition though he employed the classical scaled particle theory applied to simplified models of a protein in folded and unfolded states neglecting its polyatomic structure.

In the present study, we examine if the experimental results for the effects of addition of monohydric alcohols and polyols on the thermal stability of a protein can be reproduced by the same rigid-body models focused on the solvent-entropy effect as a pivotal factor. Monohydric alcohols considered are methanol, ethanol, and 2-propanol, and polyols considered are ethylene glycol (2), glycerol (3), erythritol (4), xylitol (5), and mannitol (6) (the figure in the parentheses denotes the number of hydroxyl groups per molecule). We employ the hybrid of the integral equation theory and the morphometric approach. The hybrid is best suited to the physical interpretation of the calculation result. Water-cosolvent solution is modeled as a binary hard-sphere mixture. As pointed out in earlier works,[36,39,41] it is important to carefully determine the effective diameter of cosolvent molecules d_C and the packing fractions of water and cosolvent, η_S and η_C. For monohydric alcohols, d_C is estimated from the Stockmayer (12-6-3) potential pertinent to polar liquids.[42] For polyols, d_C is estimated from the volume per molecule in the solid crystal.[43–47] Once d_C is estimated (the molecular diameter of water d_S is 0.28 nm), η_S and η_C can be calculated using the experimental data of solution density for a cosolvent concentration given.[48,49] We find that all of the experimental observations can beautifully be reproduced: The stability is lowered by the addition of a monohydric alcohol and the lowering becomes stronger in the order, methanol<ethanol<2-propanol,[3,5,9,10] it is enhanced by the addition of a polyol and the enhancement becomes stronger in the order, ethylene glycol<glycerol<erythritol<xylitol<mannitol,[1,3,11–13] and the degree of lowering or enhancement becomes larger as the cosolvent concentration increases.[1,3,5,9–13] The agreement between the theoretical and experimental results are good even in a quantitative sense. We also test the effect of urea addition on the thermal stability using the same theoretical method combined with the rigid-body models. The result is that urea acts as a stabilizer, which conflicts with the experimental fact. This conflict indicates that another physical factor is also

significant unlike in the cases of sugars, monohydric alcohols, and polyols. We discuss a probable physical factor and suggest a general method which allows us to elucidate the effect of adding any cosolvent.

2.2. Physical picture of thermal denaturation of a protein in different solvents

Kinoshita and coworker's physical picture of the thermal stability of a protein[36] is illustrated in Fig. 2.1. The protein is immersed in three different solvents. For example, solvents α, β, and γ correspond to 0.1 mol/L ethanol solution, pure water, and 0.1 mol/L glycerol solution, respectively. The stability is described by the competition of $\Delta S/(k_B N_r)$ and $|\Delta S_{conf}|/(k_B N_r)$. Here, ΔS is the solvent-entropy gain and ΔS_{conf} the conformational-entropy loss upon protein folding, k_B the Boltzmann constant, and N_r the number of residues of the protein (refer to their earlier publications[36,40] for more details). Let $\Delta S_{298}/(k_B N_r)$ be $\Delta S/(k_B N_r)$ at T=298 K. The figure shows that T_m becomes higher as $\Delta S_{298}/(k_B N_r)$ increases and the order of $\Delta S_{298}/(k_B N_r)$ coincides with that of T_m. Hence, $\Sigma=\Delta S_{298}/(k_B N_r)$ is an appropriate measure of the stability. We can analyze the effect of cosolvent addition by looking at the change in the measure caused: If it is positive, for example, the stability is enhanced by the addition and its larger value implies stronger enhancement.

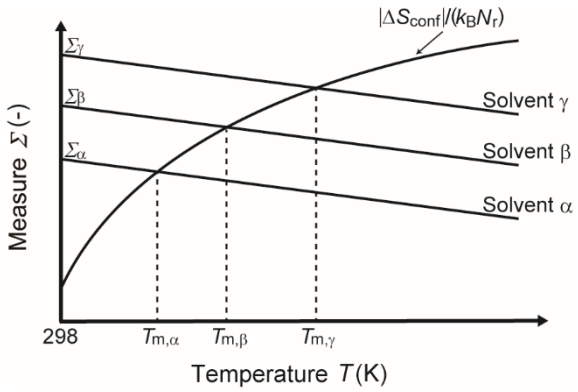

Fig. 2.1. Kinoshita and coworker's physical picture of thermal denaturation for a protein in different solvents. Three solvents, "α", "β", and "γ", are considered. Notations: ΔS=solvent-entropy gain in each solvent upon protein folding, ΔS_{conf}=conformational-entropy loss upon protein folding, k_B=Boltzmann constant, N_r=number of residues of the protein, T_m=denaturation temperature, $\Sigma=\Delta S_{298}/(k_B N_r)$, and ΔS_{298} is ΔS at 298 K. Σ is a measure of the thermal stability of the protein. The subscripts, "α", "β", and "γ", denote the values for solvents "α", "β", and "γ", respectively.

2.3. Model and theory

2.3.1. Protein and solvent models

We consider Protein G with 56 residues as a model protein. Its native structure, the folded state,

is taken from Protein Data Bank (the PDB code is 1PGB). The unfolded state is an ensemble of thermally denatured structures. A total of 500 random coils are prepared by assigning random numbers to the dihedral angles[36,50] and regarded as the unfolded state. A polyatomic structure of the protein is represented by a set of fused hard spheres. The diameter of each atom is set at the sigma value of the Lennard-Jones (LJ) potential parameters taken from CHARMM22.[51] Unrealistic overlaps of protein atoms in any structure are removed by slightly modifying it using the standard energy-minimization technique. Refer to their earlier work[36] for more details. A physical quantity of the unfolded state is calculated as the average value for the 500 random coils.

Pure water is formed by hard spheres with diameter d_S=0.28 nm. Water-cosolvent solution is modeled as a binary hard-sphere mixture: The cosolvent possesses an effective diameter denoted by d_C. The packing fractions of water and cosolvent, η_S and η_C, are dependent on the cosolvent species and concentration.

2.3.2. Estimation of effective diameter of cosolvent molecules and packing fractions of cosolvent and water

For water-cosolvent solution, d_C, η_S, and η_C must be estimated. First, we describe how to estimate d_C. For polyols, it is evaluated as $d_C=(6fm/\pi\rho N_A)^{1/3}$ where f is the close packing factor, m (g/mol) the molecular weight, ρ (g/cm^3) the density of the solid crystal,[43–47] and N_A Avogadro's constant. This evaluation method is the same as that employed in their earlier work[36] for sucrose and glucose. However, it should not be applied to monohydric alcohols. This is because the solid-crystal structure is far from the close-packed one: It possesses lots of vacancy due to the formation of intermolecular hydrogen bonds. Take water, for example: If d_C is evaluated using the equation given above, the resultant value becomes unreasonably large, much larger than the correct one 0.28 nm. For this reason, we consider the Stockmayer potential.[42] For a polar gas of a monohydric alcohol, its potential parameters were determined by adapting the Chapman-Enskog theory combined with the experimental data of viscosity.[42] We modify the parameters in the dipole-dipole interaction part so that the potential can become pertinent to a liquid state in the following manner: The dipole moment is replaced by a significantly larger, effective one proposed by Jorgensen[52] to account for the polarization effect; and the orientations of the dipole moments of two molecules are chosen so that the dipole-dipole attractive interaction can be maximized (i.e., the most probable orientations[53] are chosen). The value of d_C is then evaluated as the distance between two molecular centers at which the potential equals k_BT (T=298 K). Once d_C is estimated, η_S and η_C can be calculated using the experimental data of solution density at 293 K for a cosolvent concentration given[48,49] (the data at 298 K should be almost the same as those at 293 K). We add that η_S of pure water is also calculated from the water-density data at 293 K. For urea, we adopt the value of d_C determined by Graziano[54] and the experimental data of water-urea solution density.[48] The values of d_C thus obtained for methanol, ethanol, 2-propanol, ethylene glycol, glycerol, erythritol, xylitol, mannitol, and urea are collected in Table 2.1. The values of η_S and η_C at a cosolvent concentration of ~0.8 mol/L are also given.

In general, the packing fraction of water-cosolvent solution at a cosolvent concentration becomes higher as the cosolvent hydrophilicity increases. A monohydric alcohol tends to become

less hydrophilic as the size of hydrophobic group in its molecule increases. A polyol tends to become more hydrophilic as the number of hydroxyl groups per molecule increases. Our solvent model implicitly accounts for these effects through the use of the experimental data of solution density. Of course, this implicit treatment gives a successful result only when the solvent-entropy effect explained in "Introduction" dominates. We show in the present study that this is truly the case.

Table 2.1. Values of d_C/d_S (d_S and d_C are effective molecular diameters of water and cosolvent, respectively, and d_S=0.28 nm) and those of η_S and η_C (packing fractions of water and cosolvent, respectively) at a cosolvent concentration of ~0.8 mol/L. Those for a total of nine cosolvent species are given. We choose the cosolvent concentration for which the experimental data of water-cosolvent solution is available.[48,49]

Cosolvent (Concentration: mol/L)	d_C/d_S	η_S	η_C	$\eta=\eta_S+\eta_C$
Methanol (0.930)	1.20	0.3701	0.0111	0.3812
Ethanol (0.860)	1.42	0.3655	0.0170	0.3826
2-propanol (0.822)	1.56	0.3608	0.0216	0.3824
Ethylene glycol (0.809)	1.68	0.3666	0.0266	0.3932
Glycerol (0.771)	1.92	0.3625	0.0378	0.4002
Erythritol (0.800)	2.06	0.3570	0.0484	0.4054
Xylitol (0.800)	2.20	0.3524	0.0590	0.4113
Mannitol (0.800)	2.32	0.3464	0.0691	0.4155
Urea (0.842)	1.66	0.3691	0.0267	0.3957

2.3.3. Theoretical strategy

The principal quantity we calculate is the solvation entropy of a protein with a prescribed structure *S*. When the solvent is a simple fluid as in the present study, *S* can be calculated using the three-dimensional integral equation theory (3D-IET).[55–57] However, we wish to perform physically insightful decomposition of *S*. *S* is decomposed into the protein-solvent pair correlation component S_{Pair} and the protein-solvent-solvent triplet and higher-order (i.e., many-body) correlation component $S_{Many\text{-}body}$.[36] Each component is further decomposed into two terms. One of them is dependent on the excluded volume (EV) generated by the protein and the other is dependent on the area and integrated curvatures of the solvent-accessible surface (SAS).[36] We note that such decomposition is made possible only by applying the morphometric approach (MA).[38,39] A total of four constituents, $S_{Pair,EV}$, $S_{Pair,SAS}$, $S_{Many\text{-}body,EV}$, and $S_{Many\text{-}body,SAS}$, are thus obtained. In the MA, the four constituents are calculated from the solvation entropies and their solute-solvent pair and many-body correlation components of isolated hard-sphere solutes with various diameters. The solvation entropy and its components of each isolated hard-sphere solute is calculated using the

radial-symmetric integral equation theory (RSIET) for simple fluids combined with the hypernetted-chain closure.[37] The RSIET is for solvation of a spherical solute whereas the 3D-IET is for that of a solute possessing polyatomic structure like a protein.

2.3.4. Radial-symmetric integral equation theory

First, we treat an isolated hard-sphere solute. Hereafter, the superscript "hs" for a thermodynamic quantity of solvation represents that the quantity is for the isolated hard-sphere solute. Using the RSIET, we calculate the total and direct correlation functions (h_i and c_i, respectively) for the pair, solute-component i of the solvent: m=1 for pure water and m=2 for water-cosolvent solution (components 1 and 2 are water and cosolvent, respectively). The solvation free energy of the solute μ^{hs} is calculated through the Morita-Hiroike formula:[58,59]

$$\mu^{hs}_i/(k_BT)=4\pi\rho_i \int_0^\infty r^2\{h_i(r)^2/2-c_i(r)-h_i(r)c_i(r)/2\}dr;\ i=1,\ \ldots,\ m, \tag{2.1}$$

$$\mu^{hs}=\sum_{i=1}^{m}\mu^{hs}_i. \tag{2.2}$$

Here, ρ_i is the number density of component i ($\eta_S=\pi\rho_1 d_S^3/6$ and $\eta_C=\pi\rho_2 d_C^3/6$). The subscripts "1" and "2" correspond to those "S" and "C", respectively. S^{hs}_i/k_B and S^{hs}/k_B are then obtained as $-\mu^{hs}_i/(k_BT)$ and $-\mu^{hs}/(k_BT)$, respectively, because the solvation energy is zero in the rigid-body models employed.

The solute-solvent pair correlation component of S^{hs}, S^{hs}_{Pair}, is also calculated from the solute-solvent pair correlation function $g_i(r)=h_i(r)+1$ as[36]

$$S^{hs}_{i,Pair}/k_B=4\pi\rho_i\,[\int_0^\infty\{g_i(r)-1\}r^2dr-\int_0^\infty\{g_i(r)\ln g_i(r)\}r^2dr\,], \tag{2.3}$$

$$S^{hs}_{Pair}=\sum_{i=1}^{m}S^{hs}_{i,Pair}. \tag{2.4}$$

The solute-solvent many-body correlation component of S^{hs}, $S^{hs}_{Many\text{-}body}$, is obtained as[36]

$$S^{hs}_{i,Many\text{-}body}=S^{hs}_i-S^{hs}_{i,Pair}, \tag{2.5}$$

$$S^{hs}_{Many\text{-}body}=\sum_{i=1}^{m}S^{hs}_{i,Many\text{-}body}. \tag{2.6}$$

2.3.5. Morphometric approach

Second, we treat a protein. When the superscript "hs" is removed for a thermodynamic quantity of solvation, the quantity is for the protein.

In the MA, S_i is expressed as the linear combination of the four geometric measures of a protein with a prescribed structure:[39]

$$S_i/k_B=C_{1i}V_i+C_{2i}A_i+C_{3i}X_i+C_{4i}Y_i;\ i=1,\ \ldots,\ m, \tag{2.7}$$

where V is the EV, A is the solvent-accessible surface area, and X and Y are the integrated mean and Gaussian curvatures of the surface, respectively (m=1 for pure water and m=2 for water-cosolvent solution). V_1 and V_2, for example, denote the EVs which the centers of water molecules and those of cosolvent molecules cannot enter, respectively. V_2 is not introduced in the case of pure water. The right side of Eq. (2.7) is referred to as "morphometric form". The coefficients C_{1i}–C_{4i}, which are considered independent of the solute shape, can be determined in simple solute geometries (i.e., isolated hard-sphere solutes). The calculation of S_i is performed in the following manner:[39]

(1) Calculate $S^{hs}{}_i$ with diameter d_U using the RSIET.[37] Consider sufficiently many different values of d_U in the range, $0.02 \le d_U/d_S \le 30$.
(2) Determine C_{1i}–C_{4i} by the least square fitting applied to the following equation for hard-sphere solutes:

$$S^{hs}{}_i/k_B=C_{1i}(4\pi R_i^3/3)+C_{2i}(4\pi R_i^2)+C_{3i}(4\pi R_i)+4\pi C_{4i},\ R_i=(d_U+d_i)/2;\ i=1,\ \ldots,\ m. \tag{2.8}$$

The number of coefficients is 4 for pure water (m=1) and 8 for water-cosolvent solution (m=2).
(3) Calculate V_i, A_i, X_i, and Y_i of the protein using an extended version of Connolly's algorithm.[60,61] The x-y-z coordinates and diameters of the protein atoms are the input data. The diameter of each atom is set at the sigma value of the LJ potential parameters taken from CHARMM22.[51]
(4) Using C_{1i}–C_{4i} determined in step (2), calculate S_i from Eq. (2.7).
(5) Calculate S from

$$S=\sum_{i=1}^{m} S_i. \tag{2.9}$$

Further, $S^{hs}{}_{i,\mathrm{Pair}}$ and $S^{hs}{}_{i,\mathrm{Many\text{-}body}}$ for various solute diameters are calculated using the RSIET, and $S_{i,\mathrm{Pair}}$ and $S_{i,\mathrm{Many\text{-}body}}$ are calculated by applying the MA in the same manner. S_{Pair} and $S_{\mathrm{Many\text{-}body}}$ are obtained from the following:

$$S_{\mathrm{Pair}}=\sum_{i=1}^{m} S_{i,\mathrm{Pair}}, \tag{2.10}$$

$$S_{\mathrm{Many\text{-}body}}=\sum_{i=1}^{m} S_{i,\mathrm{Many\text{-}body}}. \tag{2.11}$$

The high accuracy of the MA was demonstrated in Kinoshita and coworkers' earlier works [35,39] (also see Appendix 2-A).

2.3.6. Four constituents of solvation entropy for a protein

Each of $S_{i,\mathrm{Pair}}$ and $S_{i,\mathrm{Many\text{-}body}}$ is further decomposed into two terms. One of them is dependent on the EV and the other is dependent on the area and integrated curvatures of the SAS. That is, the EV term is the first one in the morphometric form and the SAS term is the sum of the other three terms. The two terms are represented by subscripts "EV" and "SAS", respectively. The four constituents, $S_{i,\mathrm{Pair,EV}}$, $S_{i,\mathrm{Pair,SAS}}$, $S_{i,\mathrm{Many\text{-}body,EV}}$, and $S_{i,\mathrm{Many\text{-}body,SAS}}$, are thus obtained:

$$S_i=S_{i,\mathrm{Pair,EV}}+S_{i,\mathrm{Pair,SAS}}+S_{i,\mathrm{Many\text{-}body,EV}}+S_{i,\mathrm{Many\text{-}body,SAS}};\ i=1,\ \ldots,\ m. \tag{2.12}$$

Recall that m=1 for pure water and m=2 for water-cosolvent solution. In the case of m=2, components 1 and 2 are water and cosolvent, respectively. The solvent molecules in the bulk and those near the protein surface, respectively, contribute to the EV and SAS terms. $S_{i,\mathrm{Pair,EV}}$, $S_{i,\mathrm{Pair,SAS}}$, $S_{i,\mathrm{Many\text{-}body,EV}}$, and $S_{i,\mathrm{Many\text{-}body,SAS}}$, respectively, originate from the following factors occurring upon protein insertion.[35,36]

$S_{i,\mathrm{Pair,EV}}$: Decrease in the total volume available to each solvent molecule in the bulk,
$S_{i,\mathrm{Pair,SAS}}$: Decrease in the translational freedom of each solvent molecule near the protein surface (factor 1),
$S_{i,\mathrm{Many\text{-}body,EV}}$: Increase in the solvent crowding in the bulk,
$S_{i,\mathrm{Many\text{-}body,SAS}}$: Solvent structuring near the protein surface (factor 2) and decrease in the solvent crowding in the bulk brought by factors 1 and 2.

$S_{i,\mathrm{Pair,EV}}$ coincides with the solvent-entropy loss evaluated by the Asakura-Oosawa (AO) theory.[36,62,63] $S_{i,\mathrm{Pair,EV}}$, $S_{i,\mathrm{Pair,SAS}}$ and $S_{i,\mathrm{Many\text{-}body,EV}}$ are negative. However, $S_{i,\mathrm{Many\text{-}body,SAS}}$ is positive, which can be interpreted as follows. When several solvent molecules contact with the protein surface, the EVs generated by these solvent molecules and the protein overlap, with the result that the total volume available to *the other* solvent molecules increases by the overlapped volume. The increase acts for reducing the solvent crowding in the bulk. The entropic gain arising from this reduction is larger than the entropic loss caused by the contact of the solvent molecules mentioned above (the contact contributes to factors 1 and 2).

It is straightforward from Eqs. (2.9) and (2.12) that S can be decomposed into the four constituents (these can be denoted by (Pair, EV), (Pair, SAS), (Many-body, EV), and (Many-body, SAS)) as

$$S=S_{\mathrm{Pair,EV}}+S_{\mathrm{Pair,SAS}}+S_{\mathrm{Many\text{-}body,EV}}+S_{\mathrm{Many\text{-}body,SAS}}. \tag{2.13}$$

$S=S_1$ for pure water and $S=S_1+S_2$ for water-cosolvent solution.

2.3.7. Four constituents of solvent-entropy gain upon protein folding

The values of S of the folded and unfolded states are denoted by S_{F} and S_{UF}, respectively. The solvent-entropy gain upon protein folding ΔS is given by

$$\Delta S = S_F - S_{UF}. \quad (2.14)$$

By applying Eq. (2.13) to S_F and S_{UF}, ΔS can be decomposed into the four constituents ((Pair, EV), (Pair, SAS), (Many-body, EV), and (Many-body, SAS)). Hereafter, superscripts "pw" and "mix", respectively, represent the values for pure water and water-cosolvent solution (mixture). For example, S^{pw} is "S_1 calculated for pure water" and S^{mix} is "S_1+S_2 calculated for water-cosolvent solution". ΔS^{pw} and ΔS^{mix} can also be decomposed into the four constituents. Important quantities are

$$\Delta\Delta S = \Delta S^{mix} - \Delta S^{pw} \quad (2.15)$$

and its four constituents:

$$\Delta\Delta S = \Delta\Delta S_{Pair,EV} + \Delta\Delta S_{Pair,SAS} + \Delta\Delta S_{Many\text{-}body,EV} + \Delta\Delta S_{Many\text{-}body,SAS}. \quad (2.16)$$

$\Delta S_{Many\text{-}body,EV}$ (a positive quantity), for instance, originates from the reduction of solvent crowding in the bulk upon protein folding, and $\Delta\Delta S_{Many\text{-}body,EV}$ arises from the change in $\Delta S_{Many\text{-}body,EV}$ induced by the cosolvent addition. When the addition makes the solvent crowding in the bulk more serious, $\Delta S^{mix}_{Many\text{-}body,EV}$ is larger than $\Delta S^{pw}_{Many\text{-}body,EV}$ and $\Delta\Delta S_{Many\text{-}body,EV}$ is positive: The addition leads to a larger gain of solvent entropy upon protein folding.
The change in thermal-stability measure due to the cosolvent addition $\Delta\Sigma$ is given by

$$\Delta\Sigma = \Sigma^{mix} - \Sigma^{pw} = \Delta S^{mix}_{298}/(k_B N_r) - \Delta S^{pw}_{298}/(k_B N_r). \quad (2.17)$$

Since all of the quantities are calculated at T=298 K, the following equation holds:

$$\Delta\Sigma = \Delta\Delta S/(k_B N_r). \quad (2.18)$$

Positive and negative values of $\Delta\Sigma$, respectively, imply that the cosolvents act as a stabilizer and a destabilizer.

2.4. Results and discussion

2.4.1. Addition of monohydric alcohols and polyols

We calculate $\Delta\Sigma$ for each cosolvent species as a function of its concentration (mol/L). The result is plotted in Fig. 2.2. It is apparent that methanol, ethanol, and 2-propanol act as destabilizers. The degree of destabilization follows the order, methanol<ethanol<2-propanol: It increases as the hydrophobic group in an alcohol molecule becomes larger.[3,5,9,10] By contrast, ethylene glycol, glycerol, erythritol, xylitol, and mannitol act as stabilizers. The degree of stabilization follows the order, ethylene glycol(2)<glycerol(3)<erythritol(4)<xylitol(5)<mannitol(6) (the figure in the

parentheses denotes the number of hydroxyl groups in a molecule): It increases as the number of hydroxyl groups per molecule becomes larger.[1,3,11–13] The degree always increases as the cosolvent concentration becomes higher. Thus, all of the experimental observations are reproducible at least qualitatively by the rigid-body models focused on the solvent-entropy effect.

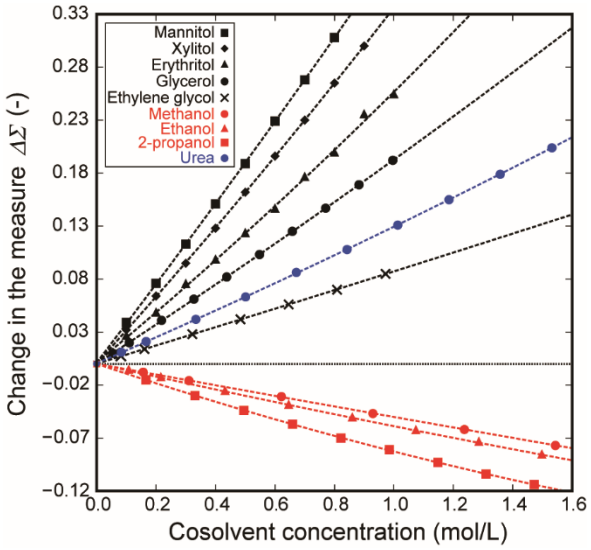

Fig. 2.2. Relation between $\Delta\Sigma$ (change in the thermal-stability measure caused by the cosolvent addition) and the cosolvent concentration calculated for a total of eight cosolvent species and urea. Positive $\Delta\Sigma$: The cosolvent acts as a stabilizer and increases the denaturation temperature of a protein T_m. Negative $\Delta\Sigma$: The cosolvent acts as a destabilizer and lowers T_m.

In general, the entropic EV effect becomes larger as the solvent packing fraction increases and/or the molecular diameter of solvent decreases.[32–34] In the present case, d_C and the total packing fraction $\eta=\eta_S+\eta_C$ are essential parameters. When d_C becomes larger with η kept constant, for example, $\Delta\Sigma$ decreases. When η becomes higher with d_C kept constant, $\Delta\Sigma$ increases. $\Delta\Sigma$ is determined by the interplay of d_C and η.

We then check the quantitative agreement between the theoretical and experimental results. Gekko and Koga[3] reported the changes in T_m caused by the addition of methanol, ethanol, and the five polyols for collagen (we deal with protein G). The relation between $\Delta\Sigma$ and ΔT_m (the change in T_m) is plotted in Fig. 2.3. The correlation coefficient R is 0.973. When only the five polyols are considered, R reaches 0.993. We can conclude that the agreement is surprisingly good even in a quantitative sense.

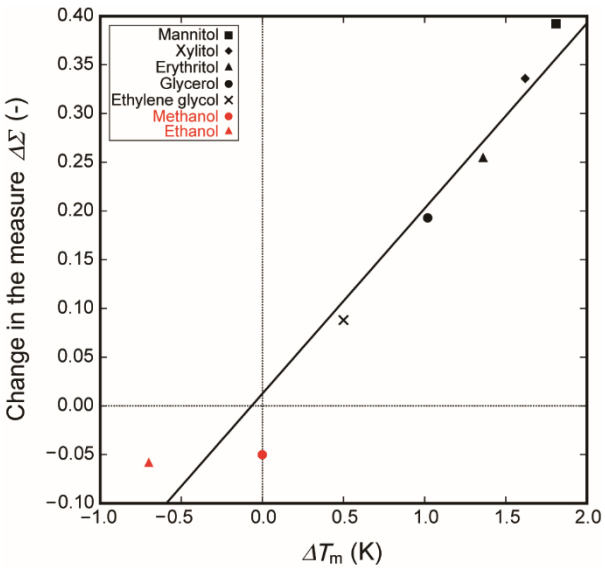

Fig. 2.3. Relation between theoretical $\Delta\Sigma$ and experimental ΔT_m for methanol, ethanol, and the five polyols. The cosolvent concentration is 1 mol/L and the protein is collagen. $\Delta\Sigma$ and ΔT_m denote changes in the thermal-stability measure and in the thermal denaturation temperature, respectively. ΔT_m was measured by Gekko and Koga.[3] ΔT_m for methanol is approximately zero in their data, though it is negative in the other groups' data.[5,9,10] The intercept of the straight line, which is drawn using the least-squares method, is close to zero. The correlation coefficient R is 0.973. When only the five polyols are considered, R reaches 0.993.

2.4.2. Physical interpretation of the results

We provide $\Delta\Delta S$, $\Delta\Delta S_{Pair,EV}$, $\Delta\Delta S_{Pair,SAS}$, $\Delta\Delta S_{Many\text{-}body,EV}$, and $\Delta\Delta S_{Many\text{-}body,SAS}$ for six representative water-cosolvent solutions (solutions 1 through 6) in Table 2.2. $\Delta\Delta S$ is negative for solutions 1 and 2 whereas it is positive for solutions 3–6. The magnitude of $|\Delta\Delta S|$ becomes larger as the solution number increases. These characteristics are also observed only in $\Delta\Delta S_{Many\text{-}body,EV}$. The stability change due to the cosolvent addition is determined primarily by $\Delta\Delta S_{Many\text{-}body,EV}$. T_m increases when the reduction of solvent crowding in the bulk upon protein folding is magnified, and T_m decreases when it is made smaller. When a cosolvent possesses high affinity with water, its addition to water increases the total packing fraction and makes the solvent crowding more serious. As a consequence, unless the cosolvent effective diameter is too large, the solvent-entropy gain upon protein folding increases. Interestingly, the application of the AO theory[36,62,63] considering $\Delta\Delta S_{Pair,EV}$ alone leads to the prediction that a cosolvent always works as a destabilizer.

Table 2.2. Values of $\Delta\Delta S$, $\Delta\Delta S_{\mathrm{Pair,EV}}$, $\Delta\Delta S_{\mathrm{Pair,SAS}}$, $\Delta\Delta S_{\mathrm{Many\text{-}body,EV}}$, and $\Delta\Delta S_{\mathrm{Many\text{-}body,SAS}}$ calculated for six representative water-cosolvent solutions, solutions 1 through 6. 1: (2-propanol, 0.495 mol/L), 2: (2-propanol, 0.822 mol/L), 3: (glycerol, 0.548 mol/L), 4: (glycerol, 0.771 mol/L), 5: (mannitol, 0.500 mol/L), and 6: (mannitol, 0.800 mol/L). "1: (2-propanol, 0.495 mol/L)", for example, represents that the cosolvent is 2-propanol and its concentration is 0.495 mol/L in solution 1. The following equations hold: $\Delta\Delta S_{\mathrm{Pair}}=\Delta\Delta S_{\mathrm{Pair,EV}}+\Delta\Delta S_{\mathrm{Pair,SAS}}$, $\Delta\Delta S_{\mathrm{Many\text{-}body}}=\Delta\Delta S_{\mathrm{Many\text{-}body,EV}}+\Delta\Delta S_{\mathrm{Many\text{-}body,SAS}}$, $\Delta\Delta S_{\mathrm{EV}}=\Delta\Delta S_{\mathrm{Pair,EV}}+\Delta\Delta S_{\mathrm{Many\text{-}body,EV}}$, and $\Delta\Delta S_{\mathrm{SAS}}=\Delta\Delta S_{\mathrm{Pair,SAS}}+\Delta\Delta S_{\mathrm{Many\text{-}body,SAS}}$. See Secs. 2.3.5 and 2.3.6 for the notations.

Solution	1	2	3	4	5	6
$\Delta\Delta S/k_{\mathrm{B}}$	−2.469	−3.957	5.770	8.222	10.59	17.26
$\Delta\Delta S_{\mathrm{Pair,\,EV}}/k_{\mathrm{B}}$	−1.480	−2.419	−1.141	−1.603	−2.449	−3.944
$\Delta\Delta S_{\mathrm{Pair,\,SAS}}/k_{\mathrm{B}}$	−3.309	−5.261	3.650	5.194	5.081	8.166
$\Delta\Delta S_{\mathrm{Many\text{-}body,\,EV}}/k_{\mathrm{B}}$	−10.00	−15.74	55.45	78.96	104.5	170.0
$\Delta\Delta S_{\mathrm{Many\text{-}body,SAS}}/k_{\mathrm{B}}$	12.32	19.46	−52.19	−74.33	−96.60	−157.0
$\Delta\Delta S_{\mathrm{Pair}}/k_{\mathrm{B}}$	−4.788	−7.680	2.509	3.590	2.632	4.222
$\Delta\Delta S_{\mathrm{Many\text{-}body}}/k_{\mathrm{B}}$	2.319	3.723	3.261	4.631	7.951	13.04
$\Delta\Delta S_{\mathrm{EV}}/k_{\mathrm{B}}$	−11.48	−18.16	54.31	77.35	102.1	166.0
$\Delta\Delta S_{\mathrm{SAS}}/k_{\mathrm{B}}$	9.011	14.20	−48.54	−69.13	−91.52	−148.8

2.4.3. Addition of urea

We calculate $\Delta\Sigma$ for urea as a function of its concentration (mol/L), and the result is included in the plot of Fig. 2.2. It is observed that urea acts as a stabilizer. This is consistent with the previously reported results showing that the reversible work required for cavity formation in water-urea solution is larger than that in pure water.[54,64–66] However, urea is experimentally known as a typical destabilizer.[5–8] It follows that there is another factor which destabilizes the protein native structure and can surpass the solvent-entropy effect. Very recently, using an all-atom molecular dynamics simulation combined with the energy representation method (the MD-ER simulation), Matubayasi and his coworkers[29] have analyzed the changes in solvation free energy and protein-solvent interaction energy upon transfer of cytochrome *c* from pure-water to 8M-urea solution for folded and partially unfolded structures. The transfer free energy (i.e., the change in solvation free energy) is strongly correlated with the van der Waals component of protein-solvent interaction energy. The correlation with the electrostatic component is much weaker due to the cancellation of water and urea contributions. The protein-urea van der Waals attractive interaction is substantially stronger than the protein-water one. The protein-solvent electrostatic interaction energy remains almost unchanged even when pure water is replaced by 8M-urea solution. Following this result, we assume that the factor is the protein-urea van der Waals attractive interaction. The factor promotes a protein to take a structure with much larger solvent-accessible surface area, which works for destabilizing the protein. At the same time, however, the

solvent-entropy effect works for stabilizing it. Probably, the destabilizing effect is more dependent on the urea concentration than the stabilizing one. The concentration must be sufficiently high for the destabilizing effect to become dominant. This is why a very high urea concentration (~8 mol/L) is necessitated[67,68] for the denaturation at room temperature.

2.4.4. On the general method of elucidating cosolvent effects

We now introduce a new thermal-stability measure incorporating the energetic component as well as the entropic component, which is similar to our recently developed one.[69] It is defined as

$$\Sigma_{\mathrm{new}}=\Delta S_{298}/(k_{\mathrm{B}}N_{\mathrm{r}})-\Delta\Lambda_{298}/(k_{\mathrm{B}}TN_{\mathrm{r}}),\ T=298\ \mathrm{K} \tag{2.19}$$

where $\Delta\Lambda$ is the change in the energetic component Λ upon protein folding and the subscript "298" denotes the value calculated at T=298 K. $\Delta\Lambda$ comprises an energy decrease brought by the gain of protein intramolecular (electrostatic plus van der Waals) energy, energy increase due to the loss of protein-solvent interaction energy, and energy decrease originating from reorganization of the solvent structure near the protein surface (these are referred to as elements 1, 2, and 3, respectively). Here, we discuss "$\Delta\Lambda^{\mathrm{mix}}-\Delta\Lambda^{\mathrm{pw}}$". The folded and unfolded states of a protein in water-cosolvent solution can be considered to be almost indistinguishable from those in pure water, respectively. The values of element 1 are then cancelled out when the subtraction is taken. The contribution from element 3 is not large unless the cosolvent concentration is quite high. As explained above, it is reasonable to assume that the van der Waals term of element 2 (this is denoted by $\Delta\lambda$) makes a dominant contribution to "$\Delta\Lambda^{\mathrm{mix}}-\Delta\Lambda^{\mathrm{pw}}$". As for the electrostatic term of element 2, cases of pure water and water-cosolvent share almost the same value.[29] The key quantity is "$\Delta\lambda^{\mathrm{mix}}-\Delta\lambda^{\mathrm{pw}}$" per unit solvent-accessible surface area as a function of the cosolvent concentration. We then obtain

$$\Delta\Lambda^{\mathrm{mix}}-\Delta\Lambda^{\mathrm{pw}}=\Delta A(\Delta\lambda^{\mathrm{mix}}-\Delta\lambda^{\mathrm{pw}}) \tag{2.20}$$

where ΔA is the change in the solvent-accessible surface area upon protein folding. The change in thermal-stability measure due to the cosolvent addition is then given by

$$\Delta\Sigma_{\mathrm{new}}=\Delta\Sigma_{\mathrm{entropy}}-\Delta\Sigma_{\mathrm{energy}}, \tag{2.21a}$$
$$\Delta\Sigma_{\mathrm{entropy}}=\Delta S^{\mathrm{mix}}{}_{298}/(k_{\mathrm{B}}N_{\mathrm{r}})-\Delta S^{\mathrm{pw}}{}_{298}/(k_{\mathrm{B}}N_{\mathrm{r}}), \tag{2.21b}$$
$$\Delta\Sigma_{\mathrm{energy}}=\Delta A(\Delta\lambda^{\mathrm{mix}}{}_{298}-\Delta\lambda^{\mathrm{pw}}{}_{298})/(k_{\mathrm{B}}TN_{\mathrm{r}}),\ T=298\ \mathrm{K}. \tag{2.21c}$$

$\Delta\Sigma_{\mathrm{entropy}}$ and $\Delta\Sigma_{\mathrm{energy}}$ are, respectively, the entropic and energetic components of the measure $\Delta\Sigma_{\mathrm{new}}$. Graziano[54] proposed a physical picture wherein the reduction in the EV followed by a solvent-entropy gain and the increase in protein-solvent attractive interaction energy, which occur upon protein folding, are highlighted and the urea effects can be described by calculating these two factors in water-urea solution and in pure water. This physical picture is in line with ours though his and our theoretical approaches are substantially different (he employed the classical scaled particle theory applied to simplified models of a protein in folded and unfolded states neglecting its

polyatomic structure).

It is experimentally known that sugar and polyol molecules prefer to be hydrated in the bulk.[70–73] For sugars and polyols, $\Delta\Sigma_{\mathrm{entropy}}$ (>0) predominates over $|\Delta\Sigma_{\mathrm{energy}}|$. By contrast, urea molecules are considerably enriched near the protein surface,[67,74] probably due to the protein-urea van der Waals attractive interaction.[29] For urea, $|\Delta\Sigma_{\mathrm{energy}}|$ ($\Delta\Sigma_{\mathrm{energy}}<0$) is larger than $\Delta\Sigma_{\mathrm{entropy}}$ (>0). For monohydric alcohols, neither preferential hydration in the bulk nor considerable enrichment near the protein surface has been observed.[75] However, the result from the present study indicates that $|\Delta\Sigma_{\mathrm{entropy}}|$ ($\Delta\Sigma_{\mathrm{entropy}}<0$) is larger than $|\Delta\Sigma_{\mathrm{energy}}|$ for monohydric alcohols. According to experimental observations at normal temperature,[76–79] there is a strong trend that the addition of a monohydric alcohol induces a protein to form α-helices. This is in contrast with the urea addition promoting a protein to take considerably more extended structures. The helical structure induced by a monohydric alcohol is independent of the alcohol species, but the degree of the inducement becomes stronger as the size of hydrocarbon group in an alcohol molecule increases.[79] Kinoshita and coworkers have recently shown that in solvent environment where $\Delta\Sigma_{\mathrm{entropy}}$ takes a significantly large, negative value, the formation of α-helices (not β-sheets) is strongly promoted.[80] This result is consistent with the above discussion.

2.5. Concluding remarks

As in the case of sugars,[36] when a polyol is added, solvent crowding in the bulk is enhanced and its reduction upon protein folding is magnified with the result of more stabilized folded state. The enhancement becomes stronger as the number of hydroxyl groups per molecule increases, because a larger number leads to the stronger hydrophilicity of polyol followed by the increased total packing fraction of solvent. By contrast, the solvent crowding is weakened when a monohydric alcohol is added. The reduction of solvent crowding upon protein folding becomes smaller with the result of less stabilized folded state. This action becomes stronger as the hydrophobic group in an alcohol molecule becomes larger, because a larger size leads to the stronger hydrophobicity of alcohol with the result of the decreased total packing fraction of solvent. The solvent-entropy gain upon protein folding (factor I) can be a good measure of the protein thermal stability. For urea, the theoretical method focused on factor I fails to reproduce the experimental fact[5–8] that its addition lowers the thermal stability: It gives the opposite result. For urea, the loss of protein-solvent interaction energy upon protein folding (factor II) plays more important roles.

Factor I dominates for polyols and sugars. It is experimentally known that sugar and polyol molecules prefer to be hydrated in the bulk.[70–73] For a cosolvent which prefers to be hydrated in the bulk, factor I dominates. The result from the present study indicates that for monohydric alcohols factor I is larger than factor II (see the last paragraph in Sec. 2.4.4), though neither preferential hydration in the bulk nor considerable enrichment near the protein surface has been observed for monohydric alcohols.[75] Factor II dominates for urea. Urea is considerably enriched near the protein surface[67,74] probably because the protein-urea van der Waals attractive interaction is stronger than the protein-water one.[29] Such considerable enrichment is a sign of the dominance of factor II. The

protein-solvent electrostatic interaction energy remains almost unchanged even when pure water is replaced by water-urea solution.[29] We note that factor I still works to enhance the thermal stability in the urea case: This is why a very high urea concentration (~8 mol/L) is necessitated[67,68] for the denaturation at room temperature. The effect of adding guanidine hydrochloride should be qualitatively the same as that of the urea addition.[74] It is interesting to note that a cosolvent destabilizing the protein through factor II such as urea and guanidine hydrochloride possesses nitrogen atoms in its molecule. This could be the reason for the strong van der Waals attractive interaction with the protein surface.

Appendix 2-A: On the accuracy of morphometric approach applied to a two-component solvent

We consider the case where the solvent is water-cosolvent solution (m=2). In the morphometric (MA) approach applied to a two-component solvent comprising smaller and larger spheres,[39] the geometric measures for smaller and larger spheres are *independently* treated. Here, they consider the excluded volume (EV) generated by a solute sphere. The MA takes account of the EVs which the centers of smaller spheres and those of larger spheres cannot enter, respectively. When a pair of solute spheres is fused, for example, both of these EVs decrease. However, there is the EV which the centers of smaller spheres can enter but those of larger spheres cannot (see Fig. 2.4(a)). This interactive EV, which depends on both of the diameters of smaller and larger spheres (i.e., which is equivalent to neither V_1 nor V_2), is not treated in the MA. A decrease in the interactive EV also takes place when a pair of solute spheres is fused (see Fig. 2.4(b)), which is neglected in the MA. This neglect leads to underestimation of the EV generated by the fused solute spheres for smaller spheres, whereas it leads to overestimation of that for larger spheres. For this reason, μ_1 and μ_2 (μ is the solvation free energy of a protein with a prescribed structure, $S/k_B=-\mu/(k_BT)$, and subscripts "1" and "2", respectively, denote "water (smaller spheres)" and "cosolvent (larger spheres)") are always underestimated and overestimated by the MA, respectively. Here, they define the error of the MA as

$$D_i\ [\%]=100\{(\mu_i)^{\mathrm{MA}}-(\mu_i)^{\mathrm{3D\text{-}IET}}\}/(\mu_i)^{\mathrm{3D\text{-}IET}};\ i=1,\ 2 \tag{2.A1}$$

where the superscript "3D-IET" represents that the quantity is calculated using the three-dimensional (3D) integral equation theory[55–57] in which the complex polyatomic structure can directly be treated, and the superscript "MA" represents that the quantity is obtained via the MA. D_1 and D_2 are always negative and positive, respectively. However, the errors are fortuitously cancelled out to a significantly large extent when the summation, $\mu=\mu_S+\mu_L$, is taken with the result that the error for μ,

$$D\ [\%]=100\{(\mu)^{\mathrm{MA}}-(\mu)^{\mathrm{3D\text{-}IET}}\}/(\mu)^{\mathrm{3D\text{-}IET}}, \tag{2.A2}$$

becomes considerably small (D~1%).[39]

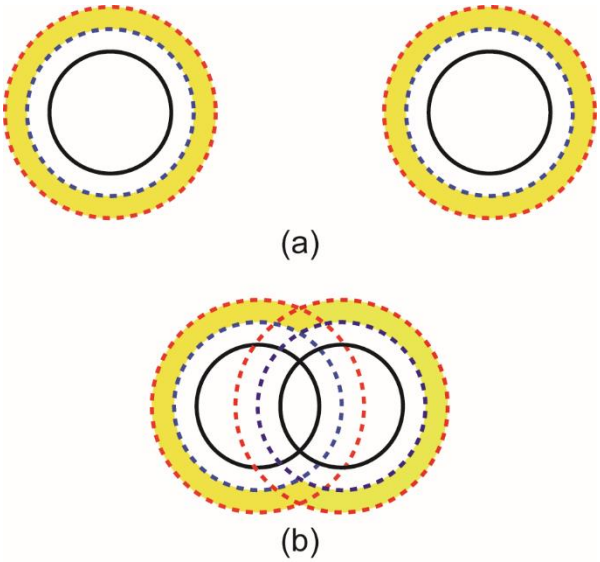

Fig. 2.4. Excluded volume (EV) which the centers of smaller and larger spheres cannot enter in the presence of a pair of isolated (a) and fused (b) solute spheres. The smaller and larger spheres form a two-component solvent. Black solid line: solute sphere. Volume encircled by blue broken line: EV for smaller spheres. Volume encircled by red broken line: EV for larger spheres. (a): The EV which the centers of smaller spheres can enter but those of larger spheres cannot (this is referred to as "interactive EV") is marked in yellow. (b): When a pair of solute spheres is fused, the interactive EV decreases by a significant amount.

References

[1]J. F. Back, D. Oakenfull, and M. B. Smith, Biochemistry **18**, 5191 (1979).

[2]H. Uedaira and H. Uedaira, Bull. Chem. Soc. Jpn. **53**, 2451 (1980).

[3]K. Gekko and S. Koga, J. Biochem. **94**, 199 (1983).

[4]N. K. Poddar, Z. A. Ansari, R. K. B. Singh, A. A. Moosavi-Movahedi, and F. Ahmad, Biophys. Chem. **138**, 120 (2008).

[5]P. H. von Hippel and K.-Y. Wong, J. Biol. Chem. **240**, 3909 (1965).

[6]V. K. Dubey and M. V. Jagannadham, Biochmistry **42**, 12287 (2003).

[7]J. D. Batchelor, A. Olteanu, A. Tripathy, and G. J. Pielak, J. Am. Chem. Soc. **126**, 1958 (2004).

[8]A. Hédoux, S. Krenzlin, L. Paccou, Y. Guinet, M.-P. Flament and J. Siepmann, Phys. Chem. Chem. Phys. **12**, 13189 (2010).

[9]A. L. Fink, Cryobiology **23**, 28 (1986).

[10]V. Bhakuni, Arch. Biochem. Biophys. **357**, 274 (1998).

[11]Y. Fujita, Y. Iwasa, and Y. Noda, Bull. Chem. Soc. Jpn. **55**, 1896 (1982).

[12]J. K. Kaushik and R. Bhat, J. Phys. Chem. B **102**, 7058 (1998).

[13]I. Haque, R. Singh, A. A. Moosavi-Movahedi, and F. Ahmad, Biophys. Chem. **117**, 1 (2005).

[14]F. Müller-Plathe and W. F. van Gunsteren, Polymer **38**, 2259 (1997).

[15]B. J. Bennion and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **100**, 5142 (2003).

[16]H.-L. Liu and C.-M. Hsu, J. Chin. Chem. Soc. **50**, 1235 (2003).

[17]A. Lerbret, P. Bordat, F. Affouard, A. Hédoux, Y. Guinet, and M. Descamps, J. Phys. Chem. B **111**, 9410 (2007).

[18]L. Hua, R. Zhou, D. Thirumalai, and B. J. Berne, Proc. Natl. Acad. Sci. U.S.A. **105**, 16928 (2008).

[19]V. Vagenende, M. G. S. Yap, and B. L. Trout, Biochemistry **48**, 11084 (2009).

[20]F.-F. Liu, L. Ji, L. Zhang, X.-Y. Dong, and Y. Sun, J. Chem. Phys. **132**, 225103 (2010).

[21]F. Mehrnejad, M. M. Ghahremanpour, M. Khadem-Maaref, and F. Doustdar, J. Chem. Phys. **134**, 035104 (2011).

[22]Q. Shao, Y. Fan, L. Yang, and Y. Q. Gao, J. Chem. Phys. **136**, 115101 (2012).

[23]S. Paul and S. Paul, J. Phys. Chem. B **119**, 10975 (2015).

[24]N. Matubayasi and M. Nakahara, J. Chem. Phys. 113, 6070 (2000).

[25]N. Matubayasi and M. Nakahara, J. Chem. Phys. **117**, 3605 (2002); Erratum. **118**, 2446 (2003).

[26]N. Matubayasi and M. Nakahara, J. Chem. Phys. **119**, 9686 (2003).

[27]S. Sakuraba and N. Matubayasi, J. Comput. Chem. **35**, 1592 (2014).

[28]Y. Karino and N. Matubayasi, Phys. Chem. Chem. Phys. **15**, 4377 (2013).

[29]Y. Yamamori, R. Ishizuka, Y. Karino, S. Sakuraba, and N. Matubayasi, J. Chem. Phys. **144**, 085102 (2016).

[30]Y. Harano and M. Kinoshita, Biophys. J. **89**, 2701 (2005).

[31]T. Yoshidome, M. Kinoshita, S. Hirota, N. Baden, and M. Terazima, J. Chem. Phys. **128**, 225104 (2008).

[32]M. Kinoshita, Front. Biosci. **14**, 3419 (2009).

[33]M. Kinoshita, Int. J. Mol. Sci. **10**, 1064 (2009).

[34]M. Kinoshita, Biophys. Rev. **5**, 283 (2013).
[35]H. Oshima and M. Kinoshita, J. Chem. Phys. **142**, 145103 (2015).
[36]H. Oshima and M. Kinoshita, J. Chem. Phys. **138**, 245101 (2013).
[37]J.-P. Hansen and L. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic, London, 2006).
[38]R. Roth, Y. Harano, and M. Kinoshita, Phys. Rev. Lett. **97**, 078101 (2006).
[39]R. Kodama, R. Roth, Y. Harano, and M. Kinoshita, J. Chem. Phys. **135**, 045103 (2011).
[40]K. Oda, R. Kodama, T. Yoshidome, M. Yamanaka, Y. Sambongi, and M. Kinoshita, J. Chem. Phys. **134**, 025101 (2011).
[41]G. Graziano, Int. J. Biol. Macromol. **50**, 230 (2012).
[42]L. Monchick and E. A. Mason, J. Chem. Phys. **35**, 1676 (1961).
[43]H. M. Berman, G. A. Jeffrey, and R. D. Rosenstein, Acta Cryst. **B24**, 442 (1968).
[44]H. van Koningsveld, Recueil des Travaux Chimiques des Pays-Bas **87**, 243 (1968).
[45]H. S. Kim and G. A. Jeffrey, Acta Cryst. **B25**, 2607 (1969).
[46]C. Ceccarelli, G. A. Jeffrey, and R. K. McMullan, Acta Cryst. **B36**, 3079 (1980).
[47]A. D. Fortes and E. Suard, J. Chem. Phys. **135**, 234501 (2011).
[48]CRC Handbook of Chemistry and Physics, 85th ed., edited by D. R. Lide (CRC Press, Boca Raton, 2004).
[49]C. Zhu, Y. Ma, and C. Zhou, J. Chem. Eng. Data **55**, 3882 (2010).
[50]T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata, J. Chem. Phys. **126**, 225102 (2007).
[51]A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[52]W. L. Jorgensen, J. Phys. Chem. **90**, 1276 (1986).
[53]M. Kinoshita and M. Harada, Mol. Phys. **74**, 443 (1991).
[54]G. Graziano, Phys. Chem. Chem. Phys. **13**, 17689 (2011).
[55]D. Beglov and B. Roux, J. Chem. Phys. **103**, 360 (1995).
[56]M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).
[57]M. Kinoshita, J. Chem. Phys. **116**, 3493 (2002).
[58]T. Morita, Prog. Theor. Phys. **23**, 829 (1960).
[59]T. Morita and K. Hiroike, Prog. Theor. Phys. **25**, 537 (1961).
[60]M. L. Connolly, J. Appl. Crystallogr. **16**, 548 (1983).
[61]M. L. Connolly, J. Am. Chem. Soc. **107**, 1118 (1985).
[62]S. Asakura and F. Oosawa, J. Chem. Phys. **22**, 1255 (1954).
[63]S. Asakura and F. Oosawa, J. Polym. Sci. **33**, 183 (1958).
[64]P. E. Smith, J. Phys. Chem. B **103**, 525 (1999).
[65]M. Ikeguchi, S. Nakamura, and K. Shimizu, J. Am. Chem. Soc. **123**, 677 (2001).
[66]S. Weerasinghe and P. E. Smith, J. Chem. Phys. **118**, 5901 (2003).

[67]V. Prakash, C. Loucheux, S. Scheufele, M. J. Gorbunoff, S. N. Timasheff, Arch. Biochem. Biophys. **210**, 455 (1981).
[68]C. A. Nelson and J. P. Hummel, J. Biol. Chem. **237**, 1567 (1962).
[69]S. Murakami, H. Oshima, T. Hayashi, and M. Kinoshita, J. Chem. Phys. **143**, 125102 (2015).
[70]K. Gekko and T. Morikawa, J. Biochem. **90**, 39 (1981).
[71]P. Lozano, J. Cano, J. L. Iborra, and A. Manjón, Enzyme Microb. Technol. **15**, 868 (1993).
[72]P. Lozano, D. Combes, and J. L. Iborra, J. Biotechnol. **35**, 9 (1994).
[73]T. Arakawa and S. N. Timasheff, Biochemistry **21**, 6536 (1982).
[74]S. N. Timasheff, Biochemistry **31**, 9857 (1992).
[75]M. G. Ortore, P. Mariani, F. Carsughi, S. Cinelli, G. Onori, J. Teixeira, and F. Spinozzi, J. Chem. Phys. **135**, 245103 (2011).
[76]N. Hirota-Nakaoka and Y. Goto, Bioorg. Med. Chem. **7**, 67 (1999).
[77]V. N. Uversky, N. V. Narizhneva, S. O. Kirschstein, S. Winter, and G. Löber, Fold. Des. **2**, 163 (1997).
[78]Y. O. Kamatari, T. Konno, M. Kataoka, and K. Akasaka, J. Mol. Biol. **259**, 512 (1996).
[79]N. Hirota, K. Mizuno, and Y. Goto, J. Mol. Biol. **275**, 365 (1998).
[80]S. Yasuda, H. Oshima, and M. Kinoshita, J. Chem. Phys. **137**, 135103 (2012).

Chapter 3

Effects of Salts or Cosolvent Addition on Solubility of a Hydrophobic Solute in Water: Relevance to Those on Thermal Stability of a Protein

3.1. Introduction

The hydrophobic effect plays imperative roles in a variety of physicochemical and biological processes in aqueous environments such as micelle formation, protein folding, molecular recognition, aggregation of biomolecules, and formation of lipid membrane. Elucidating the microscopic origin of this effect is crucial for understanding the mechanism of these processes. Though a number of theoretical and computer simulation studies have been carried out, there are still a lot of unresolved issues and controversial aspects to overcome.[1–10] Comparing the results with experimental observations is indispensible to theoretical and computer simulation studies. A problem is that physicochemical factors other than the hydrophobic effect are usually involved in the processes treated in experiments. The solubility of a nonpolar solute in water, which is extremely low, is best suited to the investigation, because it enables us to exclusively treat the hydrophobic effect. In particular, investigating the influence of adding a salt or cosolvent to water on the solubility provides a clue to explicating the hydrophobic effect.

Hereafter, "solvent" is formed by water molecules for pure water, by water molecules, cations, and anions for water-salt solution, and by water and cosolvent molecules for water-cosolvent solution. A decrease in solubility of a nonpolar solute by the salt or cosolvent addition arises from the enhancement of the hydrophobic effect. An increase in solubility, on the other hand, is ascribed to the mitigation of the effect. Up to now, plenty of experimental data have been reported for argon and methane, typical examples of small nonpolar solutes.[11–19] It is experimentally known that the solubility is decreased by the addition of an alkali halide,[19] and the degree of this decrease follows the order, $Na^+ > K^+ > Cs^+ > Li^+$ for cations with a common anion species or $Cl^- > Br^- > I^-$ for anions with a common cation species. For instance, NaCl reduces the solubility more than KCl at the same salt concentration. Further, there is experimental data for methane indicating the order, NaCl>NaBr>KCl>NaI>CsCl>LiCl.[19] The decrease in solubility is referred to as "salting out". Interestingly, the addition of tetramethylammonium bromide $[(CH_3)_4N]Br$ leads to an increase in the solubility of methane,[17] which is known as "salting in". The addition of methanol increases the solubility of argon,[14] whereas that of urea or sucrose decreases the solubility of methane.[19] (Alkali halides and $[(CH_3)_4N]Br$ are categorized as salts and methanol, sucrose, and urea as cosolvents.) The power of decreasing or increasing the solubility becomes stronger as the salt or cosolvent concentration increases.

The thermal stability of a protein, which is measured by the denaturation temperature T_m, is also influenced by the addition of a salt or colsolvent.[20–29] It is definite that the hydrophobic effect

is a principal driving force in protein folding regardless of its microscopic origin. For a protein, enhanced hydrophobic effect and reduced hydrophobic effect lead to higher T_m and to lower T_m, respectively. Unlike argon and methane, however, a protein possesses groups with positive or negative charges. Further, it possibly interacts with solvent particles through significantly strong van der Waals potential. Protein-solvent electrostatic and van der Waals interactions can come into play depending on the salt or cosolvent species added, and the hydrophobic effect is often not the only factor affecting the thermal stability. Consequently, the effects of salt or cosolvent addition are much more complex than those on the solubility of nonpolar solutes such as argon and methane. Though various salts have been tested in experiments, alkali halides provide the most fundamental information. According to the experimental results, T_m of apoflavodoxin at pH=7 (its total charge is $\sim -19e$; e is the elementary charge) is raised by the addition of an alkali halide, and the degree of this raising follows the order, $Na^+ > K^+ > Cs^+ > Li^+$ for cations with Cl^- as a common anion species or $Cl^- > Br^-$ for anions with Na^+ as a common cation species.[28] For cytochrome *c* at pH=4.5 (its total charge is $\sim 17e$), qualitatively the same behavior is observed for cations. However, upon addition of NaCl, NaBr, and NaI, respectively, T_m increases, remains almost unchanged, and decreases.[28] For ribonuclease A at pH=7 (its total charge takes a significantly large, positive value), T_m becomes slightly higher by NaCl or KCl but lower by LiBr or NaBr.[20,21] T_m of ribonuclease A is lowered by $[(CH_3)_4N]Br$, methanol, or 1-propanol.[21] The degree of this lowering is larger for 1-propanol than for methanol. If the hydrophobic effect is dominantly important, the nonpolar-solute solubility and the protein thermal stability should share qualitatively the same behavior against the salt or cosolvent addition. When the addition of a salt or cosolvent leads to lower solute solubility, that of the same salt or cosolvent makes T_m of a protein higher. However, there can be discrepancies between the solubility and the thermal stability in their qualitative behavior. For example, the presence of anions with large sizes, Br^- and I^-, can lower T_m of a protein.[28] Urea is a popular denaturant which lowers T_m to a significant extent.[21,25,26,29] These observations are inconsistent with or even opposite to those for the solute solubility. In cases where such discrepancies are observed, factors other than the hydrophobic effect need to be taken into account in describing the change in thermal stability. In Table 3.1, we summarize the experimentally available information on the effects of salt or cosolvent addition described above.

Table 3.1. Summary of experimentally available information[11,12,14,15,17–29] on the effects of salt or cosolvent addition on the solubility of a small nonpolar solute (argon or methane) and the thermal denaturation temperature T_m of a protein. "↑" and "↓" represent that the addition increases and decreases the physicochemical quantity, respectively. "$Na^+ > K^+ > Cs^+ > Li^+$ (Cl^-)", for example, represents that the degree of decreasing the solubility or raising T_m follows the order $Na^+ > K^+ > Cs^+ > Li^+$ with a common anion species of Cl^-. "$I^- > Br^-$" represents that the presence of I^- or Br^- decreases the solubility, but it often lowers T_m and the degree of lowering T_m follows the order $I^- > Br^-$.

	Solubility of a small nonpolar solute ↓	Solubility of a small nonpolar solute ↑
Denaturation temperature of a protein ↑	$Na^+ > K^+ > Cs^+ > Li^+$ (Cl^-). $Cl^- > Br^- > I^-$ (K^+). Sucrose.	No experimental data known.
Denaturation temperature of a protein ↓	$I^- > Br^-$: for a protein with a large, positive total charge. Urea.	$[(CH_3)_4N]Br$. 1-propanol > Methanol.

The influence of the presence of an ion species on a physicochemical quantity (e.g., solubility or T_m of a protein) is described by the so-called Hofmeister series.[30] For alkali and halide ions, the series was originally expressed in the orders $Li^+ > Na^+ > K^+$ and $Cl^- > Br^- > I^-$ in terms of their salting-out ability for hen egg white protein (a protein possessing a negative total charge). Though the solubility data for significantly many nonpolar, polar, and charged solutes share *similar* orders, they are not quite the same: The series is often expressed in more or less different orders depending on the solute species and the physicochemical quantity considered.[20,21,28] The Hofmeister series is one of the most intricate subjects in modern chemical physics of aqueous solution that has not been solved for over one hundred years. It is probable that the series is related, in some way, to the change in hydrophobic effect caused by the ion addition.

Kinoshita and coworkers have argued that the hydrophobic effect is attributed to an entropic one originating from the translational displacement of water molecules.[8,31–33] Upon solute insertion, an excluded space which the centers of water molecules cannot enter is generated. The volume of the excluded space is referred to as "excluded volume (EV)". The total volume available to the translational displacement of water molecules in the system (not limited to the water molecules near the solute) reduces by the EV, causing an entropic loss. They note that the presence of a water molecule also generates an EV for the other water molecules. In this sense, water molecules in the bulk are entropically correlated, and this correlation is referred to as "water crowding".[31–33] Upon solute insertion the water crowding becomes more serious, which makes a pivotal contribution to

the water-entropy loss. The enhanced water crowding is primarily responsible for the hydrophobic effect. The hydrophobic effect, which is influenced by the salt or cosolvent addition, can be analyzed by means of a rigid-body model where all of the solute and solvent particles interact through hard-body potentials with no attractive parts, as long as the temperature and pressure dependences of the factor are not considered[31–33] (i.e., the discussion is limited to the factor at normal temperature and pressure). The simplest method is to model the solute and solvent particles as hard spheres with different diameters. With this rigid-body model, all of the possible system configurations share the same energy, and the system behavior becomes purely entropic: The argument can be focused on the roles of solvent entropy. They emphasize the following: The solvation energy and free energy are largely dependent on solute-solvent interaction potential, whereas the solvation entropy is rather insensitive to it (see Appendix 3-A). The diameters and packing fractions of the solvent particles must carefully be determined. For instance, the determination is made from the interaction potentials[34] or X-ray electron density measurements[35–39] and the experimental data of solution density.[40,41] The effects of strength of ion hydration, cosolvent-water affinity, and change in the water structure in the bulk are implicitly incorporated in the packing fractions.

There are two types of approaches toward the elucidation of the effects of cosolvent or salt addition discussed above. One of them is a molecular dynamics (MD) simulation using a realistic, all-atom model. It provides us with much detailed information. However, the result is often difficult to interpret because it stems from complex interplay of multiple physicochemical factors. The other is a theoretical method based on statistical mechanics. When it is combined with a simplified model in which only a particular, expectedly pivotal factor is incorporated, it enables us to gain physical insights. In the first step, we explore to what extent it can reproduce the experimental data. Of course, it fails to reproduce some of the data. In the second step, we think about the reason for the failure and identify a factor or factors to be considered further. Both of the two approaches are required for the breakthrough.

In this study, we adopt a statistical-mechanical method. We consider the solubilities of argon and methane and demonstrate that all of the experimental data described in the first paragraph are well reproducible by the rigid-body model combined with the integral equation theory[42] based on statistical mechanics. By decomposing the solvation entropy of a nonpolar solute into physically insightful constituents using Kinoshita and coworkers' morphometric approach,[43,44] we show that the effects of salt or cosolvent addition are determined primarily by the resultant change in solvent crowding. The validity of their view concerning the hydrophobic effect is thus corroborated. Further, we discuss the cases where the results for the solubility of a nonpolar solute and for the thermal stability of a protein are inconsistent or qualitatively opposite: for example, the cosolvent addition reduces the solute solubility whereas it lowers T_m of a protein. In such cases, the rigid-body model fails to correctly describe the change in T_m. As a product, we present a physical picture by which the effects of salt or cosolvent addition on the thermal stability of a protein can be elucidated in a unified manner. A new view on the physical origin of the Hofmeister series is also proposed. We

discuss how it is expressed when the change in hydrophobic effect dominates and how it is modified when other factors also come into play.

3.2. Model and theory

3.2.1. Models of solute, water molecules, anions, cations, and cosolvent molecules

Hereafter, the subscripts "U", "S", "+", "–", and "C" denote "solute", "water", "cation", "anion", and "cosolvent", respectively. The solute, water molecules, anions, cations, and cosolvent molecules are modeled as hard spheres with diameters d_U, d_S, d_+, d_-, and d_C, respectively (d_S=0.28 nm).

3.2.2. Estimation of solute diameter

As the solutes, we choose argon and methane for which the solubility data are the most abundant. The solute diameter d_U is determined as follows. We calculate the distance between the centers of solute molecules at which the solute-solute Lennard-Jones (LJ) potential[45] equals k_BT (k_B is the Boltzmann constant and T=298 K): d_U is set at this distance (d_U=1.14d_S for argon and 1.28d_S for methane). We have verified that the qualitative aspects of the results are independent of d_U within the framework of our rigid-body model (a detailed discussion is given in Sec. 3.3.5).

3.2.3. Estimation of diameters of cations and anions

The diameters of alkali and halide ions are taken from the values adopted by Kusalik and Patey[46] in their theoretical studies on aqueous electrolyte solution. They are determined from X-ray electron density measurements. For $(CH_3)_4N^+$, the diameter proposed by Graziano[47] is employed. The values of d_+ and d_- are given in Table 3.2.

Table 3.2. Values of d_i/d_S (i=+, –, or C; d_S=0.28 nm): d_S, d_+, $d_–$, and d_C denote the diameters of water molecules, anions, cations, and cosolvent molecules, respectively.

Ion species or cosolvent	d_i/d_S
Li^+	0.68
Na^+	0.84
K^+	1.08
Cs^+	1.28
$Cl^–$	1.16
$Br^–$	1.28
$I^–$	1.44
$(CH_3)_4N^+$	1.80
Methanol	1.20
Ethanol	1.42
1-propanol	1.58
Urea	1.66

3.2.4. Estimation of effective diameters of cosolvent molecules

We consider methanol, ethanol, 1-propanol, and urea as the cosolvents. For a polar gas of an alcohol, the parameters in the Stockmayer potential were determined by adapting the Chapman-Enskog theory combined with the experimental data of viscosity.[34] The parameters in the dipole-dipole interaction part are modified so that the potential can be applied to a liquid state in the following manner. First, the dipole moment is set at a significantly larger, effective value proposed by Jorgensen[48] for taking account of the polarization effect. Second, the most probable orientations, which maximize the pair dipole-dipole attractive interaction, are chosen. The value of d_C is then evaluated as the distance between two molecular centers at which the potential equals k_BT (T=298 K). For urea, we adopt the value of d_C determined by Graziano.[49] The values of d_C thus obtained are given in Table 3.2.

3.2.5. Estimation of packing fractions of anions, cations, and cosolvent molecules

Once d_+, $d_–$, and d_C are estimated, η_+, $\eta_–$, and η_C (η denotes the packing fraction) can be calculated using the experimental data of solution density for salt and cosolvent concentrations given. The data for salt and cosolvent solutions were measured at 298 K and 293 K,[40,41] respectively, but the data at 298 K should be almost the same as those at 293 K. η_S of pure water is calculated from the water-density data at 298 K. The values of η_S, η_+, and $\eta_–$ for water-salt solution and those of η_S and η_C for water-cosolvent solution at a concentration of 1 mol/L are given in Table 3.3.

Table 3.3. Values of η_S and η_i (i=+, –, or C): η_S, η_+, η_-, and η_C denote the packing fractions of water molecules, anions, cations, and cosolvent molecules, respectively ($\eta_T=\eta_S+\eta_+ +\eta_-$ or $\eta_T=\eta_S+\eta_C$). Water-salt or water-cosolvent solution at 1 mol/L is considered. For pure water, η_S=0.3831.

Salt or cosolvent	η_S	η_+	η_-	η_C	η_T
LiCl	0.3760	0.0022	0.0108	–	0.3890
NaCl	0.3760	0.0041	0.0108	–	0.3909
KCl	0.3721	0.0087	0.0108	–	0.3916
KBr	0.3694	0.0087	0.0145	–	0.3927
KI	0.3651	0.0087	0.0207	–	0.3945
NaBr	0.3734	0.0041	0.0145	–	0.3920
NaI	0.3692	0.0041	0.0207	–	0.3939
CsCl	0.3673	0.0145	0.0108	–	0.3926
$[(CH_3)_4N]Br$	0.3404	0.0404	0.0145	–	0.3953
Methanol	0.3691	–	–	0.0120	0.3810
Ethanol	0.3626	–	–	0.0198	0.3825
1-propanol	0.3566	–	–	0.0273	0.3839
Urea	0.3664	–	–	0.0317	0.3980

3.2.6. Solute solvation entropy and solubility

The Ostwald coefficient $\exp\{-\mu/(k_BT)\}$ (μ is the solvation free energy of a solute) is a good measure of the solute solubility.[8,50] Since the solvation energy is zero in the rigid-body model employed, $\mu=-TS$ and $\exp\{-\mu/(k_BT)\}=\exp(S/k_B)$ (S is the solvation entropy, solvent-entropy loss upon solute insertion).

3.2.7. Basic strategy

The key quantity is S. S is decomposed into the solute-solvent pair correlation component S_{Pair} and the solute-solvent-solvent triplet and higher-order (i.e., many-body) correlation component $S_{Many\text{-}body}$.[31–33,51,52] The decomposition is performed using the integral equation theory (IET) for simple fluids.[42] Each component is further decomposed into two terms by applying the morphometric approach (MA).[43,44] One of them is dependent on the excluded volume (EV) generated by the solute and the other is dependent on the area and integrated curvatures of the solvent-accessible surface (SAS). These decompositions yield a total of four constituents, $S_{Pair,EV}$, $S_{Pair,SAS}$, $S_{Many\text{-}body,EV}$, and $S_{Many\text{-}body,SAS}$:

$$S=S_{Pair,EV}+S_{Pair,SAS}+S_{Many\text{-}body,EV}+S_{Many\text{-}body,SAS}. \quad (3.1)$$

The method for calculating the four constituents is explained in Sec. 3.2.9.

3.2.8. Integral equation theory for multicomponent system

We consider a hard-sphere solute with diameter d_U immersed in a hard-sphere mixture forming the solvent. The number of components in the mixture is m: m=1 for pure water, m=3 for water-salt solution (water molecules, cations, and anions), and m=2 for water-cosolvent solution (water and cosolvent molecules). In the IET,[42] the input data is formed by the temperature, number density of each solvent component, interaction potentials for $m(m-1)/2$ solvent-solvent pairs, and those for m solute-solvent pairs. By numerically solve the Ornstein-Zernike (OZ) and closure equations, we calculate the total and direct correlation functions for $m(m-1)/2$ solvent-solvent pairs and those for m solute-solvent pairs. Thermodynamic quantities of solvation are then calculated using the solute-solvent total and direct correlation functions. The OZ equation is exact, but the closure equation is approximate: the hypernetted-chain approximation is employed in this study. Unlike in the MD simulation, in essence, the system size and the number of system configurations considered for taking an ensemble average of a physical quantity are both infinitely large.

The solvation free energy of the solute μ is calculated from the Morita-Hiroike formula:[53,54]

$$\mu_i/(k_BT)=4\pi\rho_i \int_0^\infty r^2\{h_i(r)^2/2-c_i(r)-h_i(r)c_i(r)/2\}dr;\ i=1,\ \ldots,\ m, \tag{3.2}$$

$$\mu=\sum_{i=1}^{m}\mu_i. \tag{3.3}$$

Here, h_i and c_i, respectively, are the total and direct correlation functions for the pair, solute-component i of the solvent, and ρ_i is the number density of component i ($\eta_i=\pi\rho_i d_i^3/6$). For water-salt solution, the subscripts "1", "2", and "3" correspond to those "S", "+", and "–", respectively. For water-cosolvent solution, the subscripts "1" and "2" correspond to those "S" and "C", respectively. In the rigid-body model, $S_i/k_B=-\mu_i/(k_BT)$ and $S/k_B=-\mu/(k_BT)$. It is obvious that

$$S=\sum_{i=1}^{m}S_i. \tag{3.4}$$

The solute-solvent pair correlation component of S, S_{Pair}, is calculated from the solute-solvent pair correlation function $g_i(r)=h_i(r)+1$ as[51]

$$S_{i,Pair}/k_B=4\pi\rho_i\ [\int_0^\infty \{g_i(r)-1\}r^2dr-\int_0^\infty \{g_i(r)\ln g_i(r)\}r^2dr], \tag{3.5}$$

$$S_{\mathrm{Pair}}=\sum_{i=1}^{m} S_{i,\mathrm{Pair}}. \tag{3.6}$$

The solute-solvent many-body correlation component of S, $S_{\mathrm{Many\text{-}body}}$, is obtained as

$$S_{i,\mathrm{Many\text{-}body}}=S_i-S_{i,\mathrm{Pair}}, \tag{3.7}$$

$$S_{\mathrm{Many\text{-}body}}=\sum_{i=1}^{m} S_{i,\mathrm{Many\text{-}body}}. \tag{3.8}$$

In the calculations, we use the computer program developed by Kinoshita and Lado,[55] where an arbitrary number of components can be treated and a robust, highly efficient algorithm is employed for numerically solving the basic equations.

3.2.9. Morphometric approach

In the MA, imaginary hard-sphere solutes with sufficiently many different diameters are treated. It comprises the following steps:[44]

(1) Calculate Z^{I}_i ($Z^{\mathrm{I}}_i=S^{\mathrm{I}}_i$, $S^{\mathrm{I}}_{i,\mathrm{Pair}}$, or $S^{\mathrm{I}}_{i,\mathrm{Many\text{-}body}}$) of a hard-sphere solute with diameter d^{I} using the IET. The superscript "I" denotes a value for an imaginary solute. Consider sufficiently many (12 in this study) different values of d^{I} in the range, $0.02\le d^{\mathrm{I}}/d_{\mathrm{S}}\le 30$.
(2) Determine C_{1i}–C_{4i} by the least square fitting applied to the following equation for a hard-sphere solute:

$$Z^{\mathrm{I}}_i/k_{\mathrm{B}}=C_{1i}V^{\mathrm{I}}_i+C_{2i}A^{\mathrm{I}}_i+C_{3i}X^{\mathrm{I}}_i+C_{4i}Y^{\mathrm{I}}_i, \tag{3.9}$$

$$V^{\mathrm{I}}_i=4\pi R^{\mathrm{I},3}_i/3,\ A^{\mathrm{I}}_i=4\pi R^{\mathrm{I},2}_i,\ X^{\mathrm{I}}_i=4\pi R^{\mathrm{I}}_i,\ Y^{\mathrm{I}}_i=4\pi,\ R^{\mathrm{I}}_i=(d^{\mathrm{I}}+d_i)/2;\ i=1,\ \ldots,\ m. \tag{3.10}$$

V^{I}_i, A^{I}_i, X^{I}_i, and Y^{I}_i, which are defined for the imaginary hard-sphere solutes, represent the EV, SAS area, and integrated mean and Gaussian curvatures of the SAS, respectively. There are 12 sets of (Z^{I}_i, V^{I}_i, A^{I}_i, X^{I}_i, Y^{I}_i) for the fitting. The numbers of coefficients are 4, 12, and 8 for pure water (m=1), water-salt solution (m=3), and water-cosolvent solution (m=2), respectively.
(3) The dependences of Z_i for a hard-sphere solute with diameter d_{U} on the four geometric measures become clear once C_{1i}–C_{4i} are determined:

$$Z_i/k_{\mathrm{B}}=C_{1i}V_i+C_{2i}A_i+C_{3i}X_i+C_{4i}Y_i, \tag{3.11}$$

$$V_i=4\pi R_i^3/3,\ A_i=4\pi R_i^2,\ X_i=4\pi R_i,\ Y_i=4\pi,\ R_i=(d_{\mathrm{U}}+d_i)/2;\ i=1,\ \ldots,\ m. \tag{3.12}$$

(4) The EV and SAS terms of Z (Z=S, S_{Pair}, or $S_{\mathrm{Many\text{-}body}}$), Z_{EV} and Z_{SAS}, for the hard-sphere solute with diameter d_{U} are given by

$$Z_{\mathrm{EV}}/k_{\mathrm{B}}=\sum_{i=1}^{m} C_{1i}V_i, \tag{3.13}$$

$$Z_{\mathrm{SAS}}/k_{\mathrm{B}}=\sum_{i=1}^{m} (C_{2i}A_i+C_{3i}X_i+C_{4i}Y_i). \tag{3.14}$$

The EV term of $S_{i,\mathrm{Pair}}$, $S_{i,\mathrm{Pair,EV}}$, can formally be extracted from Eq. (3.5) as

$$S_{i,\mathrm{Pair,EV}}/k_{\mathrm{B}}=-\rho_i V_i. \tag{3.15}$$

We have verified that $S_{i,\mathrm{Pair,EV}}$ obtained by decomposing $S_{i,\mathrm{Pair}}$ on the basis of the MA is identical to $S_{i,\mathrm{Pair,EV}}$ calculated from Eq. (3.15).

3.2.10. Four constituents of solvation entropy

Recall that "solvent particles" are water molecules for pure water, water molecules, cations, and anions for water-salt solution, and water and cosolvent molecules for water-cosolvent solution. The four constituents possess the following physical meanings (see Fig. 3.1).[31–33,51,52] In the equilibrium state, a layer within which the solvent density is higher than that in the bulk is formed near the solute. The solvent particles in the bulk contribute to $S_{\mathrm{Pair,EV}}$ and $S_{\mathrm{Many\text{-}body,EV}}$, whereas only those near the solute contribute to $S_{\mathrm{Pair,SAS}}$ and $S_{\mathrm{Many\text{-}body,SAS}}$. $S_{\mathrm{Pair,EV}}$, which coincides with the solvent-entropy loss evaluated by the Asakura-Oosawa (AO) theory,[56,57] arises from a decrease in the total volume available to each solvent particle in the bulk. $S_{\mathrm{Pair,SAS}}$ is ascribed to a reduction of the translational freedom of each solvent particle near the solute (factor 1). $S_{\mathrm{Many\text{-}body,EV}}$ originates from an increase in the solvent crowding in the bulk. $S_{\mathrm{Many\text{-}body,SAS}}$ is attributable to the solvent structuring near the solute (factor 2) and the decrease in solvent crowding in the bulk brought by factors 1 and 2. $S_{\mathrm{Many\text{-}body,SAS}}$ can be discussed by decomposing it as

$$S_{\mathrm{Many\text{-}body,SAS}}=S_{\mathrm{Many\text{-}body,SAS,Structure}}+S_{\mathrm{Many\text{-}body,SAS,Crowding}} \tag{3.16}$$

where $S_{\mathrm{Many\text{-}body,SAS,Structuring}}$ denotes the entropic loss due to factor 2 and $S_{\mathrm{Many\text{-}body,SAS,Crowding}}$ denotes the entropic gain brought by the decrease in solvent crowding. Unfortunately, $S_{\mathrm{Many\text{-}body,SAS,Structuring}}$ and $S_{\mathrm{Many\text{-}body,SAS,Crowding}}$ cannot separately be evaluated.

$S_{\mathrm{Pair,EV}}$, $S_{\mathrm{Pair,SAS}}$, $S_{\mathrm{Many\text{-}body,EV}}$, and $S_{\mathrm{Many\text{-}body,SAS,Structure}}$ are negative. However, $S_{\mathrm{Many\text{-}body,SAS,Crowding}}$ is positive. The counterintuitive sign of $S_{\mathrm{Many\text{-}body,SAS,Crowding}}$ can be interpreted as follows. When some solvent particles come very close to the solute, the translational displacement of these solvent particles is more restricted (effect 1): They undergo an entropic loss.

This loss is approximately in proportion to the number of solvent particles in the vicinity of the solute and therefore to the SAS area. $S_{\text{Many-body,SAS,Structure}}$ is thus negative. However, the EVs generated by the solute and by these solvent particles overlap, leading to an increase in the total volume available to the translational displacement of *the other solvent particles* (i.e., solvent particles which are not in the vicinity of the solute). This increase is followed by a reduction of their crowding (effect 2): An entropic gain is conferred on the other solvent particles. This gain is approximately proportional to the net overlapped volume and therefore to the number of solvent particles in the vicinity of the solute and to the SAS area. The density structure of solvent within the solute-induced layer is determined by the competition of effects 1 and 2. Effect 2 is larger than effect 1 with the result of a gain of solvent entropy upon the formation of the density structure. Hence, a larger SAS area is entropically more favorable in this sense, leading to positive $S_{\text{Many-body,SAS,Crowding}}$. Kinoshita and coworkers have shown that $S_{\text{Many-body,SAS,Crowding}}$ is considerably larger than $|S_{\text{Many-body,SAS,Structure}}|$ and $S_{\text{Many-body,SAS}}$ is positive.[31–33,51,52] It is important to note that the information on solvent crowding is implicated in "$S_{\text{Many-body,EV}}$+$S_{\text{Many-body,SAS,Crowding}}$".

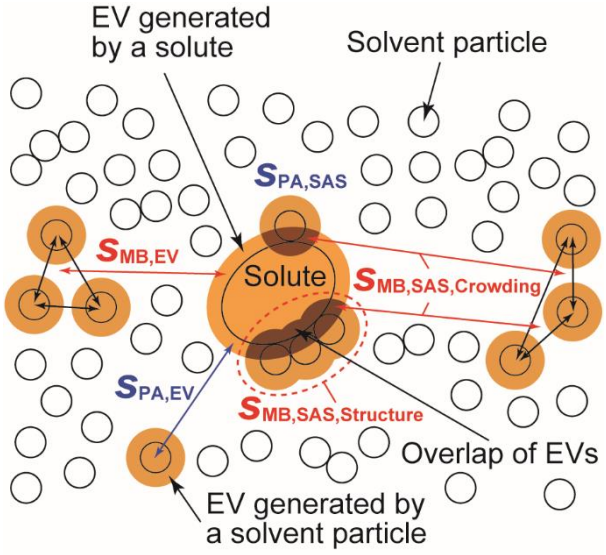

Fig. 3.1. Illustration of solute-solvent pair (PA) and many-body (MB) correlation components and their excluded-volume (EV) and solvent-accessible surface (SAS) terms of the solvation entropy, solvent-entropy loss upon solute insertion.

We define ΔX as "X for salt or cosolvent solution minus X for pure water":

$$\Delta X = X \text{ (salt or cosolvent solution)} - X \text{ (pure water)}, \tag{3.17a}$$

$$X = S_{\text{Pair,EV}},\ S_{\text{Pair,SAS}},\ S_{\text{Many-body,EV}},\ S_{\text{Many-body,SAS}},\ \text{or}\ S. \tag{3.17b}$$

3.3. Results and discussion for solubility of nonpolar solutes

3.3.1. Addition of salts (alkali-halide ions): Salting out

Choosing argon as the solute, we calculate ΔS for LiCl, NaCl, KCl, CsCl, KBr, and KI solutions. The salt concentration is 1 mol/L. Ω defined by[58]

$$\Omega=(\rho^*_S+\rho^*_+ +\rho^*_-)\exp(-\Delta S/k_B)/\rho^*_{S0} \tag{3.18}$$

is a measure of the ratio of the solubility in pure water to that in water-salt solution. Ω can be derived on the basis of the Ostwald coefficient mentioned in Sec. 3.2.6. Here, $\rho^*_i=\rho_i d_S^3$ (i=S, +, –) for water-salt solution and the subscript "0" denotes a value for pure water. The values of Ω are given in Table 3.4. $\Omega>1$ implies that the solubility is decreased by the salt addition. The degree of this decrease follows the orders, $Na^+>K^+>Cs^+>Li^+$ and $Cl^->Br^->I^-$. In experimental studies, the effect of salt addition is customarily discussed by looking at the salting coefficient k_{sol} defined by[11,58–60]

$$k_{sol}=\log(\Lambda_0/\Lambda)/C_{sol} \tag{3.19}$$

where Λ_0 and Λ, respectively, denote the solubilities of the solute in pure water and in water-salt solution of molarity C_{sol}. Posive k_{sol} implies a decrease in the solubility. We plot k_{sol} against Ω in Fig. 3.2 (CsCl is omitted because there is no experimental data for it). The two quantities are highly correlated.

Table 3.4. Values of Ω for LiCl, NaCl, KCl, CsCl, KBr, KI, and $[(CH_3)_4N]Br$ solutions. Those for water-methanol, water-ethanol, water-1-propanol, water-urea solutions are also given. Ω is defined by Eq. (3.18) for water-salt solution or by Eq. (3.20) for water-cosolvent solution. $\Omega>1$ implies a decrease in the solute solubility and $\Omega<1$ implies an increase in it. The salt or cosolvent concentration is 1 mol/L. For methanol, an additional concentration, 1.85 mol/L, is also considered. The solute is argon for LiCl, NaCl, KCl, CsCl, KBr, and KI solutions and methane for the other solutions.

Salt or cosolvent	Ω
LiCl	1.711
NaCl	1.995
KCl	1.897
CsCl	1.743
KBr	1.868
KI	1.798
$[(CH_3)_4N]Br$	0.691
Methanol	0.649
Methanol (1.85 mol/L)	0.437
Ethanol	0.549
1-propanol	0.458
Urea	2.076

There are experimental data for the Ostwald absorption coefficient[12,14] γ of methane indicating that the solubility is decreased by the salt addition and the degree of this decrease follows the order, 1M-NaCl>1M-NaBr>1M-KCl>1M-NaI>1M-CsCl>1M-LiCl>0.5M-NaCl.[19] The Ostwald absorption coefficient is the ratio of the volume of gas absorbed to the volume of absorbing liquid (argon or methane is in gas sate at ambient temperature and pressure). We plot γ_0/γ (γ_0 is the Ostwald absorption coefficient for pure water) against Ω in Fig. 3.3. The two quantities are highly correlated.

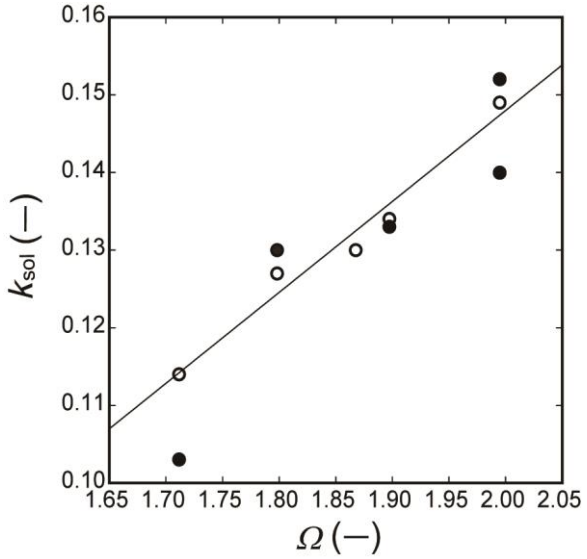

Fig. 3.2. Correlation between experimental values of k_{sol} (k_{sol} is the salting coefficient) and theoretical values of Ω defined by Eq. (3.18) for 1M-LiCl, 1M-NaCl, 1M-KCl, 1M-KBr, and 1M-KI solutions. The solute is argon. The numbers of experimental data for these salts are 3, 3, 2, 1, and 2, respectively, and the data are rather scattering (the two data for 1M-LiCl share the same value). The correlation coefficient is 0.943. When the data shown in the closed circles are removed (the straight line is drawn using the least-squares method) , the correlation coefficient reaches 0.989.

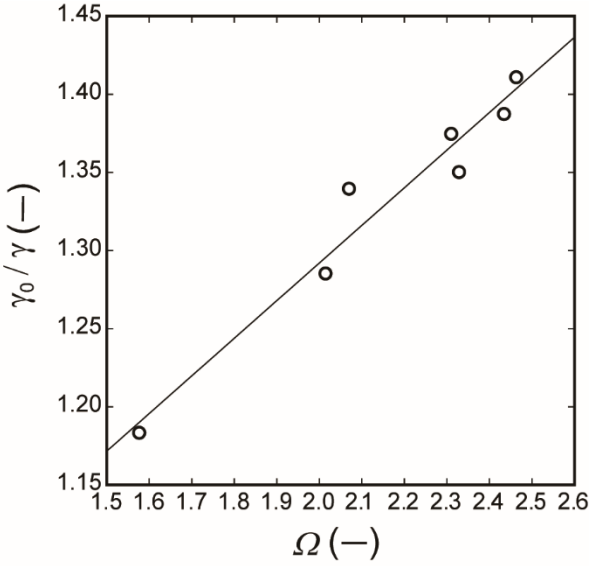

Fig. 3.3. Correlation between experimental values of γ_0/γ and theoretical values of Ω defined by Eq. (3.18) for 1M-NaCl, 1M-NaBr, 1M-KCl, 1M-NaI, 1M-CsCl, 1M-LiCl, and 0.5M-NaCl solutions. Here, γ_0 and γ are the Ostwald absorption coefficients for pure water and for salt solution, respectively. The solute is methane. The correlation coefficient is 0.975.

3.3.2. Addition of tetramethylammonium brimide: Salting in

It has experimentally been shown for methane that the addition of $[(CH_3)_4N]Br$ increases the solubility,[17] namely, the salting coefficient is negative. We calculate Ω for 1M-$[(CH_3)_4N]Br$. The result is given in Table 3.4. $\Omega<1$ implies that the solubility is increased by the cosolvent addition. The solubility of methane certainly becomes higher, which is consistent with the experimental observation.

3.3.3. Addition of methanol, ethanol, and 1-propanol

It has been shown in experiments for argon that the addition of methanol increases the solubility.[14] The degree of this increase follows the order, methanol (1.85 mol/L)>methanol (1 mol/L). For water-cosolvent solution, Ω is defined by

$$\Omega=(\rho^*_S+\rho^*_C)\exp(-\Delta S/k_B)/\rho^*_{S0} \tag{3.20}$$

where $\rho^*_i=\rho_i d_S^3$ (i=S, C) for cosolvent solution and $\rho^*_{S0}=\rho_S d_S^3$ for pure water. The values of Ω calculated for the two concentrations of methanol are smaller than 1 and consistent with the experimental data as observed in Table 3.4.

For comparison, we calculate Ω for methanol, ethanol, and 1-propanol at a concentration of 1 mol/L. The solubility increases by the alcohol addition and the degree of this increase follows the order, 1-propanol>ethanol>methanol: The degree becomes higher as the size of hydrophobic group in an alcohol molecule increases.

3.3.4. Addition of urea

It is experimentally known that the addition of urea decreases the solubility of methane.[19] As observed in Table 3.4, the value of Ω calculated is consistent with the experimental data. Another finding from our calculation is that the addition of sucrose also decreases the solubility. We have verified that the qualitative aspects of the results described in Secs. 3.3.1–4 are not altered at all by a change in the solute diameter d_U (also see Sec. 3.3.5).

3.3.5. Physical interpretation of the results

We compare ΔS, $\Delta S_{Pair,EV}$, $\Delta S_{Pair,SAS}$, $\Delta S_{Many\text{-}body,EV}$, and $\Delta S_{Many\text{-}body,SAS}$ for six representative solutions (solutions 1 through 6) in Table 3.5. To make a quantitative comparison, argon is considered as the solute in all of the solutions. The comparison is also made for a hard-sphere solute with $d_U=10d_S$ in Table 3.6 to show that the qualitative aspects of our conclusions are independent of the solute size. ΔS is negative for solutions 1–4 whereas it is positive for solutions 5 and 6. Negative ΔS implies that the magnitude of solvent-entropy loss upon solute insertion becomes larger due to the salt or cosolvent addition, whereas positive ΔS implies that it becomes smaller. $|\Delta S|$ becomes smaller as the solution number increases for solutions 1–4, and ΔS for solution 6 is larger than that of solution 5. $|\Delta S_{Many\text{-}body,EV}|$ and $\Delta S_{Many\text{-}body,SAS}$ are much larger than $|\Delta S_{Pair,EV}|$ and $|\Delta S_{Pair,SAS}|$.

Table 3.5. Values of ΔS, $\Delta S_{\text{Pair,EV}}$, $\Delta S_{\text{Pair,SAS}}$, $\Delta S_{\text{Many-body,EV}}$, and $\Delta S_{\text{Many-body,SAS}}$ scaled by k_B (k_B is the Boltzmann constant) for six representative solutions, solutions 1 through 6. 1: 1M-NaCl, 2: 1M-KCl, 3: 1M-KI, 4: 1M-LiCl, 5: 1M-$[(CH_3)_4N]$Br, 6: 1M-1-propanol. The solute is argon, a hard-sphere solute with diameter $1.14d_S$ (d_S is the molecular diameter of water). The following equations hold: $\Delta S_{\text{Pair}}=\Delta S_{\text{Pair,EV}}+\Delta S_{\text{Pair,SAS}}$, $\Delta S_{\text{Many-body}}=\Delta S_{\text{Many-body,EV}}+\Delta S_{\text{Many-body,SAS}}$, $\Delta S_{\text{EV}}=\Delta S_{\text{Pair,EV}}+\Delta S_{\text{Many-body,EV}}$, $\Delta S_{\text{SAS}}=\Delta S_{\text{Pair,SAS}}+\Delta S_{\text{Many-body,SAS}}$, and $\Delta S=\Delta S_{\text{Pair,EV}}+\Delta S_{\text{Pair,SAS}}+\Delta S_{\text{Many-body,EV}}+\Delta S_{\text{Many-body,SAS}}$. See Sec. 3.2.10 for the notations.

Solution	1	2	3	4	5	6
$\Delta S/k_B$	–0.673	–0.633	–0.597	–0.520	0.210	0.549
$\Delta S_{\text{Pair,EV}}/k_B$	–0.068	–0.052	–0.018	–0.056	0.145	0.121
$\Delta S_{\text{Pair,SAS}}/k_B$	–0.178	–0.182	–0.176	–0.130	0.000	0.127
$\Delta S_{\text{Many-body,EV}}/k_B$	–2.130	–2.145	–2.781	–1.654	–2.825	0.182
$\Delta S_{\text{Many-body,SAS}}/k_B$	1.703	1.746	2.378	1.321	2.891	0.119
$\Delta S_{\text{Pair}}/k_B$	–0.246	–0.234	–0.194	–0.187	0.144	0.248
$\Delta S_{\text{Many-body}}/k_B$	–0.427	–0.399	–0.403	–0.333	0.066	0.301
$\Delta S_{\text{EV}}/k_B$	–2.198	–2.197	–2.799	–1.711	–2.680	0.303
$\Delta S_{\text{SAS}}/k_B$	1.525	1.563	2.202	1.191	2.890	0.246

First, we discuss Table 3.6 ($d_U=10d_S$). ΔS is governed by $\Delta S_{\text{Many-body}}$ in all of the solutions. Since $\Delta S_{\text{Many-body,SAS,Crowding}}$ is considerably larger than $|\Delta S_{\text{Many-body,SAS,Structure}}|$ as mentioned in Sec. 3.2.10, $\Delta S_{\text{Many-body}}$ can be approximated by "$\Delta S_{\text{Many-body,EV}}+\Delta S_{\text{Many-body,SAS,Crowding}}$". It follows that the solvent crowding in the bulk dominates. The order of $\Delta S_{\text{Many-body}}$, 1<2<3<4<5<6, is the same as that of ΔS. Though $\Delta S_{\text{Many-body,EV}}$ makes the largest contribution to ΔS except in solution 5, the order of $\Delta S_{\text{Many-body,EV}}$, 1<3<2<4<5<6, is not completely the same as that of ΔS, implying that the contribution from $S_{\text{Many-body,SAS,Crowding}}$ cannot be neglected. $|\Delta S_{\text{EV}}|>>\Delta S_{\text{SAS}}$: The contribution from the solvent structure formed near the solute is not significantly large. We then discuss Table 3.5 ($d_U=1.14d_S$). ΔS can be described by $\Delta S_{\text{Many-body}}$ though its contribution is relatively smaller than in the case of Table 3.6. The solvent crowding is still a pivotal contributor. The orders of ΔS and $\Delta S_{\text{Many-body}}$ are 1<2<3<4<5<6 and 1<3<2<4<5<6, respectively. The order of $\Delta S_{\text{Many-body,EV}}$ is significantly different from either of them. In particular, the signs of ΔS and $\Delta S_{\text{Many-body,EV}}$ in solution 5 are opposite. These results suggest that the contribution from $S_{\text{Many-body,SAS,Crowding}}$ is significantly large. The solvent structure formed near the solute has a larger effect on ΔS than in the case of Table 3.6, which manifests, for example, relatively larger $|\Delta S_{\text{Pair,SAS}}|/|\Delta S|$: This result is reasonable because the contribution from the solvent near the solute increases in proportion to d_U^2 while that from the bulk solvent does in proportion to d_U^3.

Table 3.6. Values of ΔS, $\Delta S_{\text{Pair,EV}}$, $\Delta S_{\text{Pair,SAS}}$, $\Delta S_{\text{Many-body,EV}}$, and $\Delta S_{\text{Many-body,SAS}}$ scaled by k_B (k_B is the Boltzmann constant) for six representative solutions, solutions 1 through 6. 1: 1M-NaCl, 2: 1M-KCl, 3: 1M-KI, 4: 1M-LiCl, 5: 1M-$[(CH_3)_4N]$Br, 6: 1M-1-propanol. The solute is a hard-sphere solute with diameter $10d_S$ (d_S is the molecular diameter of water). Also see Table 3.5.

Solution	1	2	3	4	5	6
$\Delta S/k_B$	−235.0	−217.0	−200.4	−183.2	94.35	199.3
$\Delta S_{\text{Pair,EV}}/k_B$	−8.951	−4.327	4.161	−8.610	35.54	24.58
$\Delta S_{\text{Pair,SAS}}/k_B$	−9.419	−9.483	−8.908	−6.947	1.500	7.350
$\Delta S_{\text{Many-body,EV}}/k_B$	−278.1	−264.7	−269.9	−215.1	6.974	184.7
$\Delta S_{\text{Many-body,SAS}}/k_B$	61.49	61.54	74.27	47.45	50.33	−17.29
$\Delta S_{\text{Pair}}/k_B$	−18.37	−13.81	−4.747	−15.56	37.04	31.93
$\Delta S_{\text{Many-body}}/k_B$	−216.6	−203.2	−195.6	−167.7	57.31	167.4
$\Delta S_{\text{EV}}/k_B$	−287.0	−269.0	−265.7	−223.7	42.52	209.2
$\Delta S_{\text{SAS}}/k_B$	52.07	52.05	65.36	40.50	51.83	−9.939

In summary, the solubility change arising from the salt or cosolvent addition is determined primarily by $\Delta S_{\text{Many-body}}$ representing the change in solvent crowding in the bulk. This is particularly true for a large solute. Upon solute insertion, the solvent crowding is enhanced. The degree of this enhancement becomes higher or lower by the salt or cosolvent addition. When it becomes higher the solubility decreases, and when it becomes lower the solubility increases.

In general, the degree of the enhancement becomes higher as the solvent packing fraction increases or the molecular diameter of solvent decreases.[8,61] Hence, in salt or cosolvent solution, d_+ and d_- or d_C and the total packing fraction η_T are essential parameters. In cosolvent solution, for example, when d_C becomes larger with η_T kept constant, the solubility increases. When η_T becomes higher with d_C kept constant, the solubility decreases. The solubility is determined by the interplay of d_C and η_T. When a cosolvent possesses high affinity with water, its addition to water increases η_T and makes the solvent crowding more serious. As a consequence, unless the cosolvent effective diameter is too large, the solubility decreases. In salt solution, as the hydration of anions or cations becomes stronger, their addition to water increases the total packing fraction η_T and makes the solvent crowding more serious, giving rise to lower solubility.

Among the cation species tested, Li^+ is the most strongly hydrated: The absolute value of the hydration free energy is the largest (even larger than that of F^-)[62,63] and the water structure is perturbed to the largest extent. Interestingly, the total packing fraction of a solution including Li^+ is unexpectedly low. It is presumable that the strong hydration leads not only to local increases in the packing fraction but also to small vacant spaces formed.[58] As a consequence, it becomes less difficult for a solution containing Li^+ to accommodate a solute, and the order of the salting-out ability for cations is not $Li^+ > Na^+ > K^+ > Cs^+$ but $Na^+ > K^+ > Cs^+ > Li^+$.

3.4. Relevance to structural stability of a protein

3.4.1. Two principal factors (factors I and II) governing protein structural stability

We change the subject to the structural stability of a protein. The solvent-entropy gain upon protein folding ΔS is given by[51,52]

$$\Delta S = S_F - S_U \tag{3.21}$$

where S_F and S_U are the solvation entropies of folded and unfolded states of a protein, respectively. We note that the meaning of "Δ" in Eq. (3.21) is different from that in Eq. (3.17). ΔS is a measure of the thermal stability of the protein. We define $\Delta\Delta S$ as[51,52]

$$\Delta\Delta S = \Delta S \text{ (salt or cosolvent solution)} - \Delta S \text{ (pure water)}. \tag{3.22}$$

When the solvent crowding, a pivotal component of the hydrophobic effect, becomes more serious upon salt or cosolvent addition, $|S_F|$, $|S_U|$, and ΔS increase with the result of positive $\Delta\Delta S$. Positive $\Delta\Delta S$ implies enhanced thermal stability and higher T_m.

We have recently suggested that the addition of urea intensify the solvent crowing and enhances the hydrophobic effect.[52] Nevertheless, T_m becomes lower upon urea addition. This can be interpreted as follows. It is known that urea is significantly enriched near the protein surface,[64,65] with the result that significantly many water molecules near the surface are replaced by urea molecules. Interestingly, the replacement leads to essentially no change in protein-solvent (solvent is water in pure water and water and urea in solution) electrostatic interaction energy but to significant lowering of protein-solvent van der Waals interaction energy, as Matubayasi and coworkers[66] showed in their MD simulation based on the energy representation method. A significant role of protein-urea van der Waals interaction in denaturation of a peptide was also reported by Paul and coworkers[67] who performed an MD simulation. We note the following: "ε of O in a urea molecule ($\varepsilon_{O,urea}$)">"ε of N in a urea molecule ($\varepsilon_{N,urea}$)">"ε of O in a water molecule" where ε is one of the LJ parameters.[68] (Larger ε of an ion or atom in a cosolvent molecule leads to stronger protein-ion or protein-cosolvent van der Waals interaction.) When a urea molecule with flattened shape contacts the protein surface, two nitrogen atoms and an oxygen atom contact it, in which case the effective value of ε of a urea molecule should be as large as "$\varepsilon_{O,urea}+2\varepsilon_{N,urea}$". Due to the lowering of protein-solvent van der Waals interaction energy, a protein structure with larger SAS area is more favored, giving rise to deteriorated thermal stability and lower T_m. This effect is larger than the effect of the increased solvent crowding.

On the basis of the argument described above, we have pointed out two factors which are essential in describing the change in thermal stability of a protein, the hydrophobic effect (factor I) and protein-solvent van der Waals interaction energy (factor II).[52] When factor I dominates, the effects of salt or cosolvent addition on T_m of a protein and on the solubility of a nonpolar solute can

be comprehended in the same way: A salt or cosolvent which decreases the solubility always raises T_m; and when the degree of the solubility decrease is higher, T_m is raised to a larger extent. For urea, however, this is not the case and factor II is more important than factor I. Though the urea addition decreases the solubility of a nonpolar solute, it lowers T_m.

3.4.2. Cases where factor I dominates

When ions or cosolvent molecules are not significantly enriched near the protein surface, factor II cannot be influential and factor I dominates. The complex interplay of water-water, water-ion or water-cosolvent, protein-water, and protein-ion or protein-cosolvent interactions determines whether ions or cosolvent molecules are enriched or depleted and the degree of this enrichment or depletion. It is known that the addition of sugars and polyols raised T_m of a protein.[22–24,26,27] They share the feature that there are significantly many hydroxyl groups per molecule, making them highly hydrophilic. Consequently, they favor to be hydrated in bulk water. By the addition, solvent crowding becomes more serious, leading to strengthened hydrophobic effect. The addition also decreases the solubility of a nonpolar solute. In fact, there are experimental data indicating that the solubility of methane becomes lower by the addition of sucrose.[19] Monohydric alcohol molecules are neither enriched nor depleted near the protein surface.[69] Adding monohydric alcohol not only heightens the solubility of a nonpolar solute[14] but also lowers T_m of a protein,[21,24,70,71] and the effect becomes larger as the size of hydrophobic group in an alcohol molecule increases. The effects of the addition of monohydric alcohol on T_m of a protein and on the solubility of a nonpolar solute can be comprehended in the same way, indicating that factor I dominates. (A more detailed discussion on the effect of monohydric alcohol added was given in our recent publication.[52])

In general, cations and anions are more or less enriched near oppositely charged groups on the protein surface through electrostatic attractive interaction. Nevertheless, factor I is often dominant. For instance, adding NaCl to water makes the solvent crowding more serious. It not only lowers the solubility of a nonpolar solute[19] but also raises T_m of a protein.[21,28] By contrast, the solvent crowding becomes less serious by the addition of $[(CH_3)_4N]Br$. It increases the solubility of a nonpolar solute[17] and lowers T_m of a protein.[21] As argued in Sec. 3.4.4, factor II also becomes significant only when the enrichment is remarkable and at the same time the ion species features very large ε.

3.4.3. Contact of counterions with solute atoms possessing partial charges

Before going on to Sec. 3.4.4, we discuss the contact of anions with solute atoms possessing positive partial charges and that of cations with solute atoms possessing negative partial charges. Since the sign of the ion charges is opposite to that of the partial charges of solute atoms in our discussion, the ions are referred to as “counterions”. The degree of the contact is strongly dependent on the size of counterions and determined by the competition of two factors. They are the work required for destructing the hydration shell of a counterion (i.e., for dehydrating a counterion)

(factor A) and the stabilization by the direct electrostatic attractive interaction occurring when a counterion contacts a solute atom (factor B). A larger counterion is less strongly hydrated, leading to less work required: The contact is more favored in this sense. However, the distance between the centers of a larger counterion and a solute atom is longer when the former contacts the latter: The contact is less favored in this sense. (In a strict sense, the work required for destructing the hydration shell of a solute atom also comes into play, but the emphasis is placed on the counterions in the present discussion.)

In Kinoshita and coworker's earlier work,[72] it was shown that factor A dominates for anions whereas factor B dominates for cations, which could arise from the asymmetrical distributions of positive and negative charges within a water molecule.[58] Namely, larger anions and smaller cations are more enriched near groups with positive charges and near those with negative charges, respectively. What we wish to state is that Br^- and I^- tend to be enriched near oppositely charged groups much more than Cs^+.

3.4.4. Cases where factor II becomes essential

When urea is chosen as the cosolvent, the experimental and theoretical results for the two properties, solubility of a nonpolar solute and thermal stability of a protein, are qualitatively opposite. Inconsistency is found between the two properties for a protein possessing a positive total charge, when anions with large sizes, Br^- and I^-, are added to water. These anions readily come in contact with positively charged groups in the protein surface, as discussed in Sec. 3.4.3. Significantly many water molecules near the surface are thus replaced by the anions. By the analogy of the urea case, the replacement could lead to essentially no change in protein-solvent (solvent is water in pure water and water including the anions in solution) electrostatic interaction energy but to significant lowering of protein-solvent van der Waals interaction energy. We note that "ε of I^-">"ε of Br^-">>"ε of O in a water molecule". This argument does not hold in the case where a protein possessing a negative total charge is considered and cations with large sizes such as Cs^+ are added to water (see Sec. 3.4.3), due to its lower degree of enrichment and smaller ε (ε of Cs^+ is even smaller than that of Cl^-; see Appendix 3-B).

For a solute with a sufficiently large partial charge, the solute-solvent interaction energy is the dominant component of its solvation free energy. According to Kinoshita and coworkers' experience in analyses on solvation of such a solute, the solvation free energy in salt solution is not significantly different from that in pure water (see Appendix 3-C). Instead, it is generally known that the incorporation of van der Waals interaction in the solute-solvent potential leads to significant lowering of the solvation free energy.[73] Thus, our discussion based on the analogy of the urea case seems to be quite reasonable, though it is to be investigated in further studies.

Factor II is unimportant for the solubility of nonpolar solutes like argon and methane, because the solute-ion or solute-cosolvent electrostatic potential is negligibly small and the van der Waals potential has only minor effects on the solubility. Therefore, factor I always dominates, and even the addition of Br^-, I^-, and urea always decreases the solubility.

3.5. Further discussions on cosolvent effects and hydrophobicity

3.5.1. Cosolvent effects observed in molecular dynamics simulation studies

The effects of cosolvent addition on the potential of mean force (PMF) between nonpolar solutes in water have been investigated using MD simulations. In these studies, when the PMF shifts in more attractive direction upon cosolvent addition, the solute hydrophobicity is considered to be strengthened. When the PMF shifts in more repulsive direction, it is considered to be weakened. The contact and separated solute pairs could correspond to folded and unfolded states of a protein (the EV and the SAS area of the former are smaller than those of the latter), respectively. A cosolvent which makes the PMF shift in more attractive direction would enhance the structural stability of a protein (i.e., raise its T_m). On the other hand, a cosolvent inducing a PMF shift in more repulsive direction would lower it.

As argued above, our attention should be paid to the solute-solvent van der Waals interaction as well. Upon urea addition, the methane-methane PMF exhibits a shift in more attractive direction,[74] which is in qualitative accord with the experimental evidence that urea lowers the solubility of methane in water. However, the neopentane-neopentane PMF is differently influenced by urea: The PMF shifts in more repulsive direction.[75] It is probable that the solubility of neopentane increases upon urea addition. The experimental data[76] show that urea decreases the solubilities of methane and ethane but increases those of propane and alkanes possessing more than three carbon atoms per molecule. This is probably because neopentane-urea van der Waals interaction also becomes significant. Thus, it can be suggested that the hydrophobic effect can exclusively be investigated only for small nonpolar solutes like methane and argon.

The methane-methane PMF exhibits a shift in more attractive direction upon TMAO (trimethylamine-N-oxide) addition,[74] but the neopentane-neopentane PMF does not display an appreciable change.[75] The effect of neopentane-TMAO van der Waals interaction should be larger than that of methane-TMAO one. However, the difference between neopentane and methane cases for TMAO is less significant than that for urea. TMAO is experimentally known as a stabilizer of the native structure of a protein.[77] This TMAO action is attributable to the increased solvent crowding, which is consistent with the calculation result[49] showing that the reversible work required for cavity creation becomes larger upon TMAO addition.

The microstructure of water-cosolvent mixture near an amide or peptide molecule and its structural change caused by the cosolvent addition have been explored using MD simulations. Since the effects of adding monohydric alcohol, polyol, and sugar were discussed in our earlier work,[51,52] the description is focused on other cosolvents. Urea makes the peptide conformation more extended[78] whereas TMAO makes it more compact.[79] It is observed that urea molecules come in contact with the more extended peptide molecule. The interplay of two cosolvents has also been investigated. For example, the counteracting effect of TMAO against urea-induced denaturation of a protein[80] is explained as follows. The TMAO molecules cannot form hydrogen bonds by themselves but do form firm hydrogen bonds with both of water and urea molecules, leading to considerable

energy lowering.[81] Consequently, urea is stabilized with water and TMAO in the bulk, reducing the urea enrichment near the protein surface.[80,81] Not only this reduced enrichment but also the increased solvent crowding caused by the TMAO addition to water-urea solution should be responsible for the counteracting effect.

Here, we give a very important remark. In an MD simulation, cares must be taken in modeling a cosolvent molecule and setting the box size, numbers of water and cosolvent molecules, and force parameters (in particular, those affecting the EV generated by a cosolvent molecule). The manner of modeling and setting must coincide with the total packing fraction η_T of real water-cosolvent solution. We note that η_T, which has substantially large effects on the degree of solvent crowding, is considerably influenced by the manner. Inappropriate modeling and setting may result in a misleading result even in a qualitative sense. In fact, there is a report claiming that the neopentane-neopentane PMF exhibits a shift in much more repulsive direction upon TMAO addition and suggesting that the hydrophobicity of neopentane is remarkably weakened,[82] which is substantially different from the result[75] mentioned above (in our view, the neopentane-neopentane PMF should exhibit in more attractive direction upon TMAO addition). There is an MD simulation result claiming that the trehalose addition induces a shift of the neopentane-neopentane PMF in more repulsive direction.[83] In our view, however, trehalose acts just like sugars such as glucose and sucrose: Its addition makes the solvent crowding more serious and strengthens the hydrophobic effect; and trehalose is stabilized with water in the bulk and not enriched near the protein surface,[84] with the result that trehalose-protein van der Waals interaction does not play a significant role. This view can be supported by the experimental observations that trehalose is a popular stabilizer of the protein structure[85] and it counteracts urea-induced denaturation.[67] This discrepancy needs to be examined in further studies.

3.5.2. Dependency of hydrophobic effect on solute-water attractive interaction and solute size

The dependency of hydrophobic effect on the solute-water attractive interaction and the solute size is a nontrivial issue.[86–88] The attractive interaction lowers the solvation energy but has no essential effects on the water crowding in the bulk emphasized in this study (also see Appendix 3-A). As for the solute-size effect, it has been pointed out that the behavior of a sufficiently large solute is qualitatively different from that of a small one. The hydration free energy μ of the former at ambient pressure is scaled by the water-accessible surface area A and expressed by $\mu \sim \gamma A$ where $\gamma > 0$ is the surface tension of water.[7,87] Since γ increases as T becomes lower, and the scaling indicates that the hydrophobic effect becomes stronger at low temperatures. A protein is not large enough to obey the scaling for the following reason.[31] There is much experimental evidence showing that the hydrophobic effect is mitigated at low temperatures for small nonpolar solutes, amphiphilic molecules, and biomolecules like proteins. For example, upon lowering of T, the solubility of methane increases,[8,89] the critical micelle concentration becomes higher,[90] the average size of micelles for nonionic amphiphilic molecules becomes smaller,[90] most of the proteins are denatured,[91,92] and protein aggregation is dissociated.[93] Thus, the basic behavior of the hydrophobic

effect is qualitatively independent of the solute size up to the protein-size scale. Kinoshita and coworkers have argued that the folding and unfolding mechanisms of a protein (including cold and pressure denaturating) can unifiedly be explicated within the same theoretical framework emphasizing the effect of water crowding.[32,33]

3.6. Concluding remarks

By "solvent", we refer to water molecules for pure water, to water molecules, anions, and cations for water-salt solution, and to water and cosolvent molecules for water-cosolvent solution. The effects of salt or cosolvent addition on the solubility of a nonpolar solute have been analyzed using the integral equation theory[42] combined with a rigid-body model where the solute and solvent particles are modeled as hard spheres with different diameters. The model enables us to exclusively investigate the roles of solvent entropy. Argon and methane considered as the solutes are highly hydrophobic, and the solubility is governed by the hydrophobic effect and not significantly influenced by the solute-solvent interaction potential. The theoretical results for the salt or cosolvent addition are qualitatively compared with experimental data not only for the solute solubility but also for the thermal stability of a protein expressed in terms of its denaturation temperature T_m. The physical origins of the results have also been clarified by decomposing the solvation entropy into a total of four constituents using the integral equation theory combined with Kinoshita and coworkers' morphometric approach.[43,44] It is shown that the effects of salt addition and those of cosolvent addition can be understood in a unified manner.

The addition of an alkali halide reduces the solubility of a nonpolar solute, and the degree of this reduction follows the orders, $Na^+>K^+>Cs^+>Li^+$, $Cl^->Br^->I^-$, and NaCl>NaBr>KCl>NaI>CsCl>LiCl.[19] The addition of urea also reduces the solubility. By contrast, the addition of $[(CH_3)_4N]Br$ or methanol increases the solubility.[14,17] All of these experimental observations can be reproduced by our theoretical calculations. (To the best of our knowledge, it is the first time that the salting in has theoretically been reproduced.) The correlation coefficients between theoretical and experimental quantities expressing the solute solubility well exceed 0.9. It is shown that the effects of salt or cosolvent addition can be argued in terms of the mitigation or enhancement of the solvent crowding (i.e., entropic correlation among solvent particles in the bulk), a principal contributor to the hydrophobic effect. Among the cation species tested, Li^+ is the most strongly hydrated due to its smallest size.[62,63] Presumably, the strong hydration leads not only to local increases in the packing fraction but also to small vacant spaces formed.[58] As a consequence, the accommodation of a solute in a solution containing Li^+ becomes less difficult, making Li^+ come as the last in the order. To study the salting out of a nonpolar solute, Graziano[47,94] calculated the reversible work of cavity creation in water or in alkali-chloride solution using the classical scaled particle theory. He suggested that an increase in the work caused by the addition of an alkali chloride is responsible for the phenomenon. Though the result reported in this article is closely

related to his suggestion, the physical insight into the phenomenon is much deepened using our theoretical approach.

As for the thermal stability of a protein, if the enhanced hydrophobic effect dominates, the addition of a salt or cosolvent raises T_m when it lowers the solubility of a nonpolar solute. This is the case for NaCl, KCl, CsCl, and LiCl. The addition of a salt or cosolvent lowers T_m when it heightens the solute solubility, which is the case for $[(CH_3)_4N]Br$ and methanol: The reduced hydrophobic effect dominates. For urea, however, its addition reduces the solute solubility[19] but lowers T_m.[21,25,26,29] Qualitatively the same discrepancy is observed for anions with large sizes, Br^- and I^-, when a protein with a considerably large, positive total charge.[19,20] In these cases, a physical factor other than the hydrophobic effect needs to be taken into account for the change in thermal stability of a protein. The factor is probably protein-solvent van der Waals interaction energy. Urea is known to be significantly enriched near the protein surface.[64,65] Significant enrichment of Br^- and I^- near positively charged groups on the protein surface should also occur. Many water molecules near the surface are replaced by urea molecules or the anions. Since a urea molecule, Br^-, or I^- features very large ε (one of the Lennard-Jones parameters; larger ε of an ion or atom in a cosolvent molecule leads to stronger protein-ion or protein-cosolvent van der Waals interaction), the replacement leads to significant lowering of protein-solvent van der Waals interaction energy. A protein structure with a larger solvent-accessible surface area is then more favored, giving rise to deteriorated thermal stability.

Our results suggest that when the hydrophobic effect is dominant as in the salting-out case for a nonpolar solute, the Hofmeister series for alkali and halide ions are expressed by $Na^+>K^+>Cs^+>Li^+$ and $Cl^->Br^->I^-$, respectively. In these orders, the solvent crowding in the bulk is made more serious by the ion addition. However, the series for cations is often quoted in the order $Li^+>Na^+>K^+$, considering the salting-out ability for hen egg white protein.[30] For this protein possessing a negative total charge (i.e., significantly many negatively charged groups), its solubility is reduced when the negative charges are more screened by cations. As argued in Sec. 3.4.3, for the cations the screening effect becomes larger as the cation size decreases,[72] leading to the order $Li^+>Na^+>K^+$ (the decrease in protein solubility becomes larger as its negative charges are more screened). This order is also observed in the salting out of benzoic acid with oxygen atoms with considerably large negative partial charges.[20] We remark that the mechanism of salting out of a charged solute is different from that of a nonpolar one: Enhanced hydrophobic effect and screening of the charge by counterions are responsible for the latter and for the former, respectively. Further, something very interesting happens for proteins with many positively charged groups. They are extremely unfolded under conditions of acidic pH and low ionic strength but refolded to molten-globule-like conformations by the salt addition.[95–97] The power of causing the conformational transition, which arises from the screening of positive charges by the anions, follows the order $I^->Br^->Cl^-$, the reverse of $Cl^->Br^->I^-$. This is because for the anions the screening effect becomes larger as the anion size increases.[72] Taken together, the series describing the ion effect on a physicochemical quantity is exhibited in rather diverse ways because multiple physical

factors can possibly come into play, but this exhibition is reasonably interpretable.

Appendix 3-A: Insensitivity of solvation entropy to solute-solvent and solvent-solvent interaction potentials

The solvation entropy is determined primarily by the excluded-volume effect originating from the translational displacement of solvent molecules coexisting with the solute (especially by the increased solvent crowding in the bulk) and not significantly dependent on solute-solvent interaction potential. Kinoshita and coworkers calculate the solvation free energy μ, entropy S, and energy E under the isochoric condition for a hard-sphere solute with diameter 0.28 nm (equal to the molecular diameter of water). The solvent is the multipolar-model water developed by Patey and coworkers.[46,98,99] The calculation is performed using the angle-dependent integral equation theory for molecular fluids[46,98,99] (T=298 K). For the solute with zero charge, the calculated values are μ=5.95k_BT, S=–9.22k_B, and E=–3.27k_BT. When the point charge 0.5e (e is the elementary charge) is embedded at its center and the solute-solvent electrostatic potential is incorporated, the calculated values are μ=–32.32k_BT, S=–10.11k_B, and E=–42.43k_BT. Thus, S is fairly insensitive to the solute-solvent interaction potential while μ and E are largely influenced by it. (The reason why μ is largely dependent on the solute-solvent interaction potential is just that E is quite sensitive to it.) For the hard-sphere solute described above, they then calculate μ, S, and E by replacing the model water by a hard-sphere solvent for which the molecular diameter and the packing fraction are set at those of water. The calculation is performed using the integral equation theory for simple fluids.[42] The results are μ=9.64k_BT, S=–9.64k_B, and E=0. S is rather insensitive to the solvent-solvent interaction potential as well.

We then consider a large solute possessing polyatomic structure. Imai *et al.*[100] considered the native structures of a total of eight peptides and proteins and calculated S using the three-dimensional reference interaction site model (3D-RISM) theory[101–103] combined with all-atom (Coulomb plus Lennard-Jones (LJ)) potentials and the SPC/E water model.[104] Even when the protein-water electrostatic potentials, which are quite strong, are shut off and only the LJ potentials are retained, $|S|$ decreases only by ~5%. Therefore, when an analysis is focused on the solvent-entropy effect, a peptide or protein can simply be modeled as a set of fused hard spheres with no partial charges.[51,52,61]

Upon solute insertion into water, translational and rotational freedoms of water molecules are more restricted. Both of the enhanced, translational and rotational restrictions contribute to the water-entropy loss. However, the translational contribution is much larger than the rotational one (this was shown not only for a spherical solute[8] but also for a protein[105]). This is because the enhancement of rotational restriction is limited to the water molecules near the solute whereas that of translational one reaches water molecules in the whole system. Even when the model water is replaced by the hard-sphere solvent for which the molecular diameter and the packing fraction are

set at those of water, *S* does not exhibit a large change. A reason for this is the predominance of the translational contribution.

For a system comprising a nonpolar, polar, or charged solute and solvent mixture of water molecules and ions or cosolvent molecules with charges, modeling it as a mixture of hard spheres is not capable of reproducing many of important properties of the system. However, the entropic effect originating from the translational displacement of molecules and ions (this is referred to as "solvent-entropy effect" in this article) is an exception. The entropic effect in the system can well be reproduced by the hard-sphere mixture, as long as the hard-sphere diameters and the total packing fraction are carefully determined on the basis of experimental data as carried out in this study (see the following sections: 3.2.3, 3.2.4, and 3.2.5).

Appendix 3-B: Values of the Lennard-Jones parameter ε of alkali and halide ions

There are two parameters, σ and ε, in the LJ potential. When σ is given, ε can be calculated from the ion polarizability and the total number of electrons of the ion using the Mavroyannis-Stephen theory.[106] Setting σ at the diameter given in Table 3.2, we calculate ε for each of the alkali and halide ions: The result is presented in Table 3.7. The values of ε of halide ions (anions) are larger than those of alkali ions (cations). In particular, those of Br^- and I^- are quite large.

Table 3.7. Values of the Lennard-Jones parameter ε calculated for alkali and halide ions.

Ion	ε (kcal/mol)
Li^+	0.028
Na^+	0.081
K^+	0.204
Cs^+	0.732
Cl^-	0.859
Br^-	1.092
I^-	1.219

Appendix 3-C: Solvation free energies of a solute with a sufficiently large partial charge in pure water and salt solution

Using the angle-dependent integral equation theory combined with the multipolar water model,[46,98,99] Kinoshita and coworkers calculate the solvation free energies of a solute in pure water

and salt solution. The diameter and partial charge of the solute are set at 0.28 nm (equal to the molecular diameter of water) and 0.5*e* (*e* is the elementary charge), respectively. The salt is EqEq implying that the diameters of cations (Eq^+) and anions (Eq^-) are the same, and d_+ and d_- are set at d_S.[99] The solvation free energy μ is −19.0 kcal/mol in pure water. In 0.5M and 1.0M salt solutions, μ=−19.5 kcal/mol and μ=−19.6 kcal/mol, respectively. Adding the salt to water at 0.5M decreases μ only by −0.5 kcal/mol, and increasing the concentration from 0.5M to 1.0M decreases μ only by −0.1 kcal/mol.

References

[1]R. A. Pierotti, Chem. Rev. **76**, 717 (1976).

[2]B. Lee, Biopolymers **24**, 813 (1985).

[3]J. L. Finney and A. K. Soper, Chem. Soc. Rev. **23**, 1 (1994).

[4]R. M. Levy and E. Gallicchio, Annu. Rev. Phys. Chem. **49**, 531 (1998).

[5]L. R. Pratt, Annu. Rev. Phys. Chem. **53**, 409 (2002).

[6]N. T. Southall, K. A. Dill, and A. D. J. Haymet, J. Phys. Chem. B **106**, 521 (2002).

[7]D. Chandler, Nature **437**, 640 (2005).

[8]M. Kinoshita, J. Chem. Phys. **128**, 024507 (2008).

[9]P. Ball, Nature **452**, 291 (2008).

[10]J. H. Jordan and B. C. Gibb, Chem. Soc. Rev. **44**, 547 (2015).

[11]T. J. Morrison and N. B. B. Johnstone, J. Chem. Soc. 3655 (1955).

[12]A. Ben-Naim and M. Egel-Thal, J. Phys. Chem. **69**, 3250 (1965).

[13]A. Ben-Naim, J. Phys. Chem. **71**, 1137 (1967).

[14]A. Ben-Naim, J. Phys. Chem. **71**, 4002 (1967).

[15]H. L. Clever and C. J. Holland, J. Chem. Eng. Data, **13**, 411 (1968).

[16]S. K. Shoor, R. D. Walker, Jr., and K. E. Gubbins, J. Phys. Chem. **73**, 312 (1969).

[17]W.-Y. Wen and J. H. Hung, J. Phys. Chem. **74**, 170 (1970).

[18]W. L. Masterton, D. Bolocofsky and T. P. Lee, J. Phys. Chem. **75**, 2809 (1971).

[19]A. Ben-Naim and M. Yaacobi, J. Phys. Chem. **78**, 170 (1974).

[20]P. H. von Hippel and K.-Y. Wong, Science **145**, 577 (1964).

[21]P. H. von Hippel and K.-Y. Wong, J. Biol. Chem. **240**, 3909 (1965).

[22]J. F. Back, D. Oakenfull, and M. B. Smith, Biochemistry **18**, 5191 (1979).

[23]H. Uedaira and H. Uedaira, Bull. Chem. Soc. Jpn. **53**, 2451 (1980).

[24]K. Gekko and S. Koga, J. Biochem. **94**, 199 (1983).

[25]V. K. Dubey and M. V. Jagannadham, Biochemistry **42**, 12287 (2003).

[26]J. D. Batchelor, A. Olteanu, A. Tripathy, and G. J. Pielak, J. Am. Chem. Soc. **126**, 1958 (2004).

[27]N. K. Poddar, Z. A. Ansari, R. K. B. Singh, A. A. Moosavi-Movahedi, and F. Ahmad, Biophys. Chem. **138**, 120 (2008).

[28]E. Sedlák, L. Stagg, and P. Wittung-Stafshede, Arch. Biochem. Biophys. **479**, 69 (2008).

[29]A. Hédoux, S. Krenzlin, L. Paccou, Y. Guinet, M.-P. Flament, and J. Siepmann, Phys. Chem. Chem. Phys. **12**, 13189 (2010).

[30]F. Hofmeister, Arch. Exp. Pathol. Pharmakol. **24**, 247 (1888).

[31]T. Yoshidome and M. Kinoshita, Phys. Chem. Chem. Phys. **14**, 14554 (2012).

[32]M. Kinoshita, Biophys. Rev. **5**, 283 (2013).

[33]H. Oshima and M. Kinoshita, J. Chem. Phys. **142**, 145103 (2015).

[34]L. Monchick and E. A. Mason, J. Chem. Phys. **35**, 1676 (1961).

[35]H. M. Berman, G. A. Jeffrey, and R. D. Rosenstein, Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **24**, 442 (1968).
[36]H. van Koningsveld, Recl. Trav. Chim. Pays-Bas **87**, 243 (1968).
[37]H. S. Kim and G. A. Jeffrey, Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **25**, 2607 (1969).
[38]C. Ceccarelli, G. A. Jeffrey, and R. K. McMullan, Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **36**, 3079 (1980).
[39]A. D. Fortes and E. Suard, J. Chem. Phys. **135**, 234501 (2011).
[40]E. W. Washburn, *International Critical Tables of Numerical Data, Physics, Chemistry and Technology* (McGraw-Hill, New York, 1928), Vol. 3.
[41]*CRC Handbook of Chemistry and Physics*, 85th ed., edited by D. R. Lide (CRC Press, Boca Raton, 2004).
[42]J.-P. Hansen and L. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, London, 2006).
[43]R. Roth, Y. Harano, and M. Kinoshita, Phys. Rev. Lett. **97**, 078101 (2006).
[44]R. Kodama, R. Roth, Y. Harano, and M. Kinoshita, J. Chem. Phys. **135**, 045103 (2011).
[45]J. O. Hirschfelder, C. F. Curtiss and R. B. Bird, *Molecular Theory of Gases and Liquids* (New York: John Wiley and Sons, Inc., 1964).
[46]P. G. Kusalik and G. N. Patey, J. Chem. Phys. **88**, 7715 (1988).
[47]G. Graziano, J. Chem. Eng. Data. **54**, 464 (2009).
[48]W. L. Jorgensen, J. Phys. Chem. **90**, 1276 (1986).
[49]G. Graziano, Phys. Chem. Chem. Phys. **13**, 17689 (2011).
[50]B. Guillot and Y. Guissani, J. Chem. Phys. **99**, 8075 (1993).
[51]H. Oshima and M. Kinoshita, J. Chem. Phys. **138**, 245101 (2013).
[52]S. Murakami and M. Kinoshita, J. Chem. Phys. **144**, 125105 (2016).
[53]T. Morita, Prog. Theor. Phys. **23**, 829 (1960).
[54]T. Morita and K. Hiroike, Prog. Theor. Phys. **25**, 537 (1961).
[55]M. Kinoshita and F. Lado, Mol. Phys. **83**, 351 (1994).
[56]S. Asakura and F. Oosawa, J. Chem. Phys. **22**, 1255 (1954).
[57]S. Asakura and F. Oosawa, J. Polym. Sci. **33**, 183 (1958).
[58]M. Kinoshita and F. Hirata, J. Chem. Phys. **106**, 5202 (1997).
[59]W. F. McDevit and F. A. Long, J. Am. Chem. Soc. **74**, 1773 (1952).
[60]W. L. Masterton and T. P. Lee, J. Phys. Chem. **74**, 1776 (1970).
[61]Y. Harano and M. Kinoshita, Biophys. J. **89**, 2701 (2005).
[62]Y. Marcus, J. Chem. Soc., Faraday Trans. **87**, 2995 (1991).
[63]H. L. Friedman and C. V. Krishnan, in *Water: A Comprehensive Treatise*, edited by F. Franks (Plenum, New York, 1973), Vol. 3.
[64]V. Prakash, C. Loucheux, S. Scheufele, M. J. Gorbunoff, and S. N. Timasheff, Arch. Biochem. Biophys. **210**, 455 (1981).

[65]S. N. Timasheff, Biochemistry **31**, 9857 (1992).
[66]Y. Yamamori, R. Ishizuka, Y. Karino, S. Sakuraba, and N. Matubayasi, J. Chem. Phys. **144**, 085102 (2016).
[67]S. Paul and S. Paul, J. Phys. Chem. B **119**, 10975 (2015).
[68]E. M. Duffy, D. L. Severance, and W. L. Jorgensen, Isr. J. Chem. **33**, 323 (1993).
[69]M. G. Ortore, P. Mariani, F. Carsughi, S. Cinelli, G. Onori, J. Teixeira, and F. Spinozzi, J. Chem. Phys. **135**, 245103 (2011).
[70]A. L. Fink, Cryobiology **23**, 28 (1986).
[71]V. Bhakuni, Arch. Biochem. Biophys. **357**, 274 (1998).
[72]M. Kinoshita and Y. Harano, Bull. Chem. Soc. Jpn. **78**, 1431 (2005).
[73]R. Underwood, J. Tomlinson-Phillips, and D. Ben-Amotz, J. Phys. Chem. B **114**, 8646 (2010).
[74]R. Sarma and S. Paul, J. Phys. Chem. B **116**, 2831 (2012).
[75]R. Sarma and S. Paul, J. Chem. Phys. **135**, 174501 (2011).
[76]D. B. Wetlaufer, S. K. Malik, L. Stoller, and R. L. Coffin, J. Am. Chem. Soc. **86**, 508 (1964).
[77]I. Baskakov and D. W. Bolen, J. Biol. Chem. **273**, 4831 (1998).
[78]D. K. Klimov, J. E. Straub, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **101**, 14760 (2004).
[79]S. S. Cho, G. Reddy, J. E. Straub, and D. Thirumalai, J. Phys. Chem. B **115**, 13401 (2011).
[80]R. Sarma and S. Paul, J. Phys. Chem. B **117**, 5691 (2013).
[81]S. Paul and G. N. Patey, J. Am. Chem. Soc. **129**, 4476 (2007).
[82]S. Paul and G. N. Patey, J. Phys. Chem. B **112**, 11106 (2008).
[83]S. Paul and S. Paul, J. Chem. Phys. **139**, 044508 (2013).
[84]S. Shimizu, J. Chem. Phys. **120**, 4989 (2004).
[85]J. K. Kaushik and R. Bhat, J. Biol. Chem. **278**, 26458 (2003).
[86]M. V. Athawale, S. N. Jamadagni, and S. Garde, J. Chem. Phys. **131**, 115102 (2009).
[87]S. Garde and A. J. Patel, Proc. Natl. Acad. Sci. U.S.A. **108**, 16491 (2011).
[88]A. J. Patel and S. Garde, J. Phys. Chem. B **118**, 1564 (2014).
[89]H. S. Ashbaugh and L. R. Pratt, Rev. Mod. Phys. **78**, 159 (2006).
[90]D. Myers, *Surfaces, Interfaces, and Colloids: Principles and Applications* (Wiley-VCH, Berlin, 1999).
[91]A. Pastore, S. R. Martin, A. Politou, K. C. Kondapalli, T. Stemmler, and P. A. Temussi, J. Am. Chem. Soc. **129**, 5374 (2007).
[92]P. L. Privalov, Crit. Rev. Biochem. Mol. Biol. **25**, 281 (1990).
[93]R. Mishra and R. Winter, Angew. Chem., Int. Ed. **47**, 6518 (2008).
[94]G. Graziano, J. Chem. Phys. **129**, 084506 (2008).
[95]Y. Goto, N. Takahashi, and A. L. Fink, Biochemistry **29**, 3480 (1990).
[96]Y. Goto and S. Nishikiori, J. Mol. Biol. **222**, 679 (1991).
[97]Y. Hagihara, M. Kataoka, S. Aimoto, and Y. Goto, Biochemistry **31**, 11908 (1992).
[98]P. G. Kusalik and G. N. Patey, Mol. Phys. **65**, 1105 (1988).
[99]M. Kinoshita, S. Iba, and M. Harada, J. Chem. Phys. **105**, 2487 (1996).

[100]T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata, J. Chem. Phys. **125**, 024911 (2006).
[101]D. Beglov and B. Roux, J. Chem. Phys. **103**, 360 (1995).
[102]A. Kovalenko and F. Hirata, J. Chem. Phys. **110**, 10095 (1999).
[103]E. L. Ratkova, D. S. Palmer, and M. V. Fedorov, Chem. Rev. **115**, 6312 (2015).
[104]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[105]T. Yoshidome, M. Kinoshita, S. Hirota, N. Baden, and M. Terazima, J. Chem. Phys. **128**, 225104 (2008).
[106]C. Mavroyannis and M. J. Stephen, Mol. Phys. **5**, 629 (1962).

Chapter 4

Physics of Thermal-Stability Changes upon Mutation of a Protein

4.1. Introduction

Maintaining the function of a protein at a temperature above the thermal denaturation temperature T_m of its native state is a principal objective of researchers in a variety of fields related to biophysics and biochemistry.[1] This enhancement of the thermal stability, in general, can also lead to a higher stability against cooling, addition of chemical compounds, and change in pH. One of the methods of enhancing the thermal stability is the mutation. Though there are a number of possible mutations, only a small percentage of them actually lead to the enhancement with the protein function retained. It is strongly desired that the change in the thermal stability resulting from a mutation be predicted using a theoretical approach. The theoretical prediction is to be made on the condition that only the folded structure of the wild type is known (i.e., the folded structure of a mutant is unknown and there is no definite information on unfolded states of the wild type and the mutant). In the present article, we are concerned with the physicochemical factors governing the thermal-stability changes upon mutations as well as the theoretical approaches toward the prediction based on chemical physics.

Up to now, significantly many approaches[1-8] for the theoretical prediction have been reported. They are concerned primarily with the free-energy difference between the folded and unfolded states ΔG at a given temperature (e.g., 25 °C) and the change in ΔG upon mutation $\Delta\Delta G$. When the performance of an approach is examined, $\Delta\Delta G_{cal}$ is compared with $\Delta\Delta G_{exp}$ (the subscripts "cal" and "exp" denote the calculated and experimental values, respectively). Though the most straightforward way of evaluating the thermal stability is to look at T_m, ΔG has preferentially been employed.[1] There are two principal reasons for this: The construction of reliable ΔG is an important target in protein research; and ΔG is certainly related to T_m. In most of the previously reported approaches, ΔG comprises the components relevant to the protein intramolecular energy, hydration free energy, and protein intramolecular conformational entropy, and each component is further decomposed into multiple terms. Weighting coefficients are multiplied to the constituent terms of ΔG and determined in advance so that $\Delta\Delta G_{cal}$ can be best fitted to $\Delta\Delta G_{exp}$ for a sufficiently large set of proteins and mutations.[2,4,6-8] It has been pointed out that the performance of an approach is remarkably dependent on the set of proteins and mutations employed for the fitting.[8] Though the weighting coefficients determined should be close to unity in principle, some of the values obtained from the fitting are far from unity and even negative,[2-4,6-8] which is physically unreasonable. There is an approach[3] in which the weighting coefficients are not employed. However, it introduces two parameters in the modified Lennard-Jones (LJ) potential function defined between protein atoms. Moreover, $\Delta\Delta G_{cal}$ is multiplied by an adjusting parameter η. A total of three parameters are thus introduced, and they are determined so that the root-mean-squared error between $\Delta\Delta G_{cal}$ and $\Delta\Delta G_{exp}$ can be minimized. The parameter η should be close to unity but the determined value is

0.2.[3] Recently, a different type of approach[9] has been reported. In this approach, electrostatics based on the Debye-Hückel (DH) theory and a simplified treatment of the hydration free energy are introduced into a Go-like model in which only the interactions present in the native structure are taken into account. The dielectric constant of water in the DH theory is set at 29: This value is far smaller than the rightful one 78. The interaction energy per native heavy-atom contact ξ is determined so that T_m of the wild type can be matched with the experimental value. (There are two more parameters fitted to the experimental data.) T_m of a mutant is then predicted and the change in T_m by the mutation, ΔT_m, thus obtained is compared with the experimental value. The meanings of the Go-like model and the parameter ξ employed, which largely varies from protein to protein,[9] are not physically sound.

Kinoshita and coworkers have shown that the driving force of protein folding is a large gain of water entropy[10-12] (their physical picture of the folding is explained in Sec. 4.2.1). The water in the bulk makes a larger contribution to the gain than the water near the protein surface. Further, the protein-water-water triplet and higher-order correlations play critical roles in the gain. This entropic excluded-volume (EV) effect cannot be taken into account by a continuum model for water.[10-12] Using their theoretical method emphasizing the effect, which is combined with a molecular model for water, Kinoshita and coworkers have succeeded in reproducing the large water-entropy gain upon apoplastocyanin (apoPC) folding[13] experimentally estimated and in explicating the mechanisms of cold[14] and pressure[15] denaturating and of sugar-induced enhancement of the thermal stability.[16] (The importance of the entropic EV effect has been pointed out by other groups in different ways.[17,18]) The most serious drawback shared by all of the previously reported approaches for predicting the thermal-stability changes upon mutations[1-9] is that the entropic EV effect is not taken into account. In the present article, we report the results of the first attempt to examine a theoretical approach for predicting the thermal-stability changes upon mutations, which possesses the following features: It is completely free from the parameters fitted to the experimental data; and it accounts for the entropic EV effect to its full extent. It is not an empirical approach but a physics-based one.

In earlier works, Kinoshita and coworkers proposed a measure of the thermal stability of a protein. The measure was defined as the water-entropy gain upon folding at 25 °C divided by the number of residues.[19,20] A larger measure implied higher T_m. The validity of the measure was corroborated for homologous proteins: the yeast, bacterial, and human orthologues of frataxins[19] and four cytochromes *c* treated as models of mesophilic, moderately thermophilic, thermophilic, and hyperthermophilic proteins,[20] respectively. On the other hand, the structural differences among the wild type and mutants of a protein are much smaller than those among homologous proteins. Nevertheless, these slight differences give rise to denaturation-temperature changes in various magnitudes. We first applied the measure to the present subject but found that the result was not very successful. Therefore, we decided to construct a new measure on the basis of Kinoshita and coworkers' previously developed free-energy function[21,22] including the enthalpic component as well. The function has been tested for the discrimination of the native fold from misfolded decoys: The success rate of the discrimination is almost 100 percent,[21,22] demonstrating its superiority over any of the previously reported functions in which the incorporation of the entropic EV effect is

inadequate. However, the decoy structures are considerably different from the native structure. Further, it has only been shown that the function takes the lowest value for the native structure. Taken together, the subject tackled in the present study is significantly different from and much more delicate than the subjects treated in Kinoshita and coworkers' earlier publications.[19-22]

In the present article, the performance of our approach is compared with that of FOLD-X[2,23,24] for 10 proteins and a total of 207 mutations including 18 double and 13 triple and higher-fold (up to seven-fold) mutations. The number of mutations changing the total charge of the protein is 81. FOLD-X is one of the most popular, successful approaches. In the version of FOLD-X employed by us, the number of the weighting coefficients is reduced to 1 but its recommended value is only 0.33.[23,24] It is still empirical in the sense that the terms in its free-energy function are adjusted using the data from protein-engineering experiments and the experimental mutational free-energy changes.[23,24] We find that our approach and FOLD-X exhibit almost the same performance. For multiple mutations, however, our approach is far superior to FOLD-X. This superiority is crucially important, because most of the mutations realizing remarkable enhancement of the stability are multiple mutations. Five multiple mutations for staphylococcal nuclease lead to highly enhanced stabilities. An important finding is that this high enhancement originates from the entropic EV effect. The neglect of this effect in FOLD-X is a principal reason for its ill success. The number of the factors taken into account in the new measure for our approach is still relatively small: They are the water-entropy gain, loss of the protein conformational entropy, break of protein-water hydrogen bonds (HBs), and formation of protein intramolecular HBs upon folding. The success mentioned above indicates that these physicochemical factors govern the thermal-stability changes upon mutations.

In our opinion, a variety of subjects regarding proteins (protein folding,[13] cold denaturation,[14] pressure denaturation,[15] thermal denaturation,[19,20] effects of cosolvent addition on the structural stability,[16] effects of mutations on the structural stability, etc.) should be elucidated in a unified manner within the same theoretical framework: A theory which can elucidate a particular subject but fails to elucidate the others, for example, is not a good one. Our goal is to develop a theoretical method which is capable of explicating a number of protein-related issues and demonstrate its versatility. A pivotal factor in the method is the entropic effect originating from the translational displacement of water molecules coexisting with a protein. The present manuscript reports part of this development and demonstration.

4.2. New measure of thermal stability of a protein

4.2.1. Picture of protein folding

As illustrated in Fig. 4.1, a backbone and side chains of a protein generate excluded spaces which the centers of water molecules cannot enter. The volume of an excluded space is "excluded volume" (EV). Upon formation of α-helix and β-sheet by the backbone (see Figs. 4.1(a) and (b)), the overlap of the EVs occurs and the total EV decreases by the overlapped volume.[25] Likewise, large reduction of the total EV occurs when the side chains are closely packed (see Fig. 4.1(c)).[25] A

decrease in the total EV leads to an increase in the volume of the configurational phase space for water molecules, which is accompanied by a gain of water entropy. It is often claimed that intramolecular hydrogen bonds (HBs) and van der Waals attractive interactions within a protein play essential roles in driving a protein to fold. However, the folding undergoes serious dehydration comprising the loss of protein-water electrostatic and van der Waals attractive interactions and structural reorganization of water molecules near the protein surface. The importance of the loss of protein-water electrostatic attractive interactions was first pointed out by Honig and Nicholls.[26] Protein-water HBs form a pivotal component of protein-water electrostatic attractive interactions. Terazima *et al.*[27] showed that apoPC folding at 25 °C exhibits a significantly large enthalpic increase, proving that the dehydration effect dominates.[13] (The experimental technique of Terazima *et al.* enables us to directly measure the enthalpic change upon protein folding at a given temperature.) Taken together, a large water-entropy gain surpasses the enthalpic increase and loss of the protein conformational entropy.[10-13]

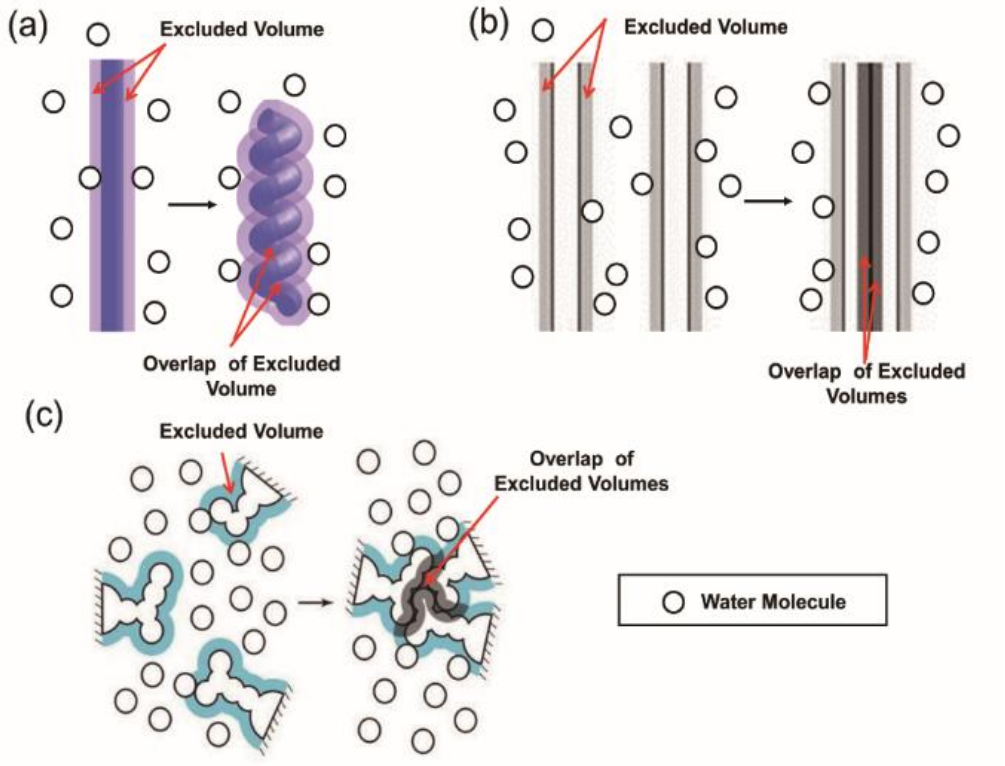

Fig. 4.1. (a) Formation of α-helix by a portion of the backbone. (b) Formation of β-sheet by portions of the backbone. (c) Close packing of side chains. The total excluded volume decreases by the overlapped volume, which leads to a corresponding increase in the total volume available to the translational displacement of water molecules coexisting with the protein.

4.2.2. Free-energy function for a protein and free-energy change upon protein folding

The free-energy function F defined by[21,22]

$$F = (\Lambda - TS_{VH})/(k_B T_0),\ T_0 = 298\ \text{K} \tag{4.1}$$

has recently been developed by Kinoshita and coworkers on the basis of the physical picture described in Sec. 4.2.1. Here, Λ is the sum of protein intramolecular energy and hydration energy, S_{VH} the hydration entropy, T the absolute temperature, and k_B Boltzmann's constant. Λ, S_{VH}, and F are functions of the protein structure. F was originally developed for the discrimination of the native fold from many misfolded decoys. Since only very compact structures were treated in the discrimination, the protein conformational entropy was not incorporated in F. In the present study, we introduce the free-energy change upon protein folding, $\Delta\Phi = k_B T_0(F_F - F_U) - T\Delta S_C$ (the subscripts "F" and "U" denote the values of the folded (native) and unfolded (denatured) states, respectively, and ΔS_C is the change in the conformational entropy) expressed by

$$\Delta\Phi/(k_B T_0) = (\Delta\Lambda - T\Delta S_{VH} - T\Delta S_C)/(k_B T_0) \tag{4.2}$$

where ΔX denotes the change in X upon folding: $\Delta\Lambda > 0$, $\Delta S_{VH} > 0$, and $\Delta S_C < 0$.

The protein insertion can be considered under either isochoric (constant-volume) or isobaric (constant-pressure) condition. The hydration free energy μ takes the same value irrespective of the insertion condition. However, this is not the case for the hydration energy E_{VH} and S_{VH}.[28] Fortunately, protein folding occurs with the system pressure and volume almost unchanged (the EV of a more compact structure is smaller but the partial molar volume is almost independent of the compactness).[10,12] It follows that the hydration energy and entropy under isochoric condition, respectively, are almost equal to the hydration enthalpy and entropy under isobaric condition. We adopt isochoric condition. Isochoric condition is free from compression or expansion of the bulk water by which the physical interpretation of a change in a thermodynamic quantity of hydration is made rather difficult.[28]

4.2.3. Water-entropy gain upon protein folding

ΔS_{VH} is a function of the number density of water ρ and T. If ρ is kept constant, ΔS_{VH} becomes larger as T increases.[29] If T is kept constant, ΔS_{VH} becomes smaller as ρ decreases.[29] It is experimentally known that above 298 K ρ decreases progressively with increasing T: In this temperature region the effect of ρ dominates and ΔS_{VH} is a decreasing function of T.[29] We then consider the wild type and a mutant of a protein whose number of residues is N_r. $S_{VH,U}/(k_B N_r)$ is not significantly dependent on the mutation if the total number of S-S bonds remain unchanged upon mutation. However, this is not true for $S_{VH,F}/(k_B N_r)$ that is quite sensitive to the packing efficiency of the backbone and side chains in the folded state.[19,20] As a result, $\Delta S_{VH}/(k_B N_r)$ ($\Delta S_{VH} = S_{VH,F} - S_{VH,U}$; ΔS_{VH} is the water-entropy gain upon folding) varies significantly from mutant to mutant. ($\Delta S_{VH}/(k_B N_r)$ of the wild type is also significantly different from that of a mutant.) See Fig. 4.2 illustrating our physical picture of the thermal stability of the wild type and a mutant.

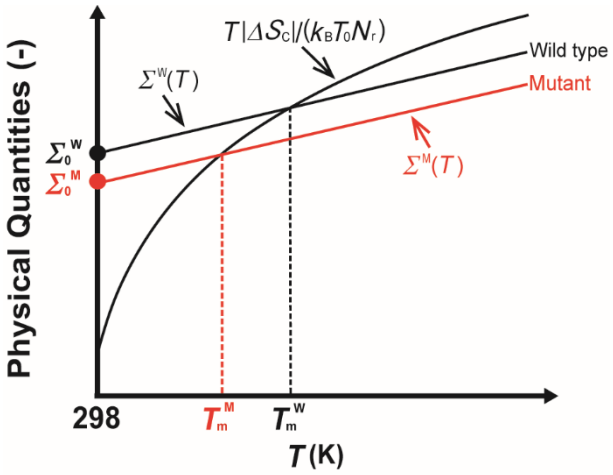

Fig. 4.2. Illustration of our physical picture for thermal stabilities of the wild type and a mutant of a protein. $\Sigma(T)$ is defined by Eq. (4.3), ΔS_C is the conformational-entropy loss upon protein folding, N_r the number of residues, k_B Boltzmann's constant, T_m the thermal denaturation temperature, $\Sigma_0 = \Sigma(T_0)$ (T_0 = 298 K), and the superscripts, "W" and "M", denote values or quantities of the wild type and the mutant, respectively. $\Sigma^W(T)$ and $\Sigma^M(T)$ do not necessarily change linearly with T. The two solid lines of $\Sigma^W(T)$ and $\Sigma^M(T)$ are not necessarily parallel.

4.2.4. Loss of protein conformational entropy upon protein folding

The protein conformational entropy is denoted by S_C. Since the folded state is under structural constraint on account of its closely packed properties, $S_{C,F}$ remains roughly constant against an increase in T. $S_{C,U}$ is much more influenced by T. $S_{C,U}$ is related to the ranges of dihedral angles allowed, which depend on the torsion energy and T.[19,20] At low T, angles giving only low torsion energy are accessible. As T increases, the allowed range of each angle is increasingly widened and $S_{C,U}$ becomes larger. As T increases further, the enlargement of $S_{C,U}$ is diminished due to the steric repulsion among atoms in a residue and in neighboring residues. Therefore, $dS_{C,U}/dT > 0$ and $d^2S_{C,U}/dT^2 < 0$ with the result of $d|\Delta S_C|/dT > 0$ and $d^2|\Delta S_C|/dT^2 < 0$ ($\Delta S_C = S_{C,U} - S_{C,F}$).[20] This temperature dependence of ΔS_C has been verified by Fitter's experimental study.[30] If the total number of S-S bonds remains unchanged upon mutation, $S_{C,U}/(k_B N_r)$ is almost constant while $S_{C,F}/(k_B N_r)$ is essentially zero, and $|\Delta S_C|/(k_B N_r)$ can be considered independent of the mutation.

4.2.5. Enthalpy change upon protein folding

Protein folding is accompanied by a decrease in the protein intramolecular energy (factor 1) and an increase in the hydration energy (factor 2).[13] Factor 2 originates from a loss of protein-water attractive (electrostatic and van der Waals) interactions and a gain of water-water attractive interactions ascribed to the structural reorganization of water near the protein surface. Factor 1 remains unchanged against an increase in T. Experimental results[27] have shown that factor 2 dominates at 298 K and protein folding gives rise to an enthalpy increase and that factor 2 becomes weaker as T increases. The gain of water-water attractive interactions, which is largely influenced

by T, is also significant in factor 2. $\Delta\Lambda$ is a strongly decreasing function of T: It takes large, positive and negative values at 298 K and in the vicinity of T_m, respectively.

4.2.6. New measure of thermal stability of a protein

We define $\Sigma(T)$ by

$$\Sigma(T) = -\Delta F/N_r = (T\Delta S_{VH} - \Delta\Lambda)/(k_B T_0 N_r) \quad (4.3)$$

where ΔX denotes the change in X upon protein folding. As illustrated in Fig. 4.2 considering the wild type and a mutant of a protein, the thermal stability can be argued by the competition of $\Sigma(T)$ and $T|\Delta S_C|/(k_B T_0 N_r)$. The argument is based on the free-energy change upon folding defined by Eq. (4.2). $\Delta\Lambda$ is a strongly decreasing function of T whereas $T\Delta S_{VH}$ remains roughly constant against an increase in T. Hence, $\Sigma(T)$ is an increasing function of T. Below T_m, the folded state is more stable than the unfolded state because $\Sigma(T)$ is larger than $T|\Delta S_C|/(k_B T_0 N_r)$. The inversion occurs above T_m, causing thermal unfolding. $\Sigma_0 = \Sigma(T_0)$ (T_0 = 298 K) can be a measure of the thermal stability. The larger Σ_0 is, the higher T_m is. The slopes of the two solid lines for the wild type and the mutant can be different, but the difference is assumed to be not large enough to invert the order of the thermal stability, $T_m^W > T_m^M$ (the superscripts, "M" and "W", denote values for the mutant and the wild type, respectively) in Fig. 4.2. We examine the correlation between $\Delta\Sigma_0$ and ΔT_m defined as

$$\Delta\Sigma_0 = \Sigma_0^M - \Sigma_0^W, \quad (4.4a)$$
$$\Delta T_m = T_m^M - T_m^W. \quad (4.4b)$$

4.3. Model and theoretical method

4.3.1. Models of water and proteins

A multipolar model[31,32] is employed for water. A point dipole and a point quadrupole of tetrahedral symmetry are embedded at the center of a hard sphere with diameter d_S = 2.8 Å. We adopt the self-consistent mean field (SCMF) theory developed by Kusalik and Patey[31,32] to account for the effect of the molecular polarizability. S_{VH} is determined primarily by the geometric characteristics of a protein structure and rather insensitive to the protein-water interaction potentials. The insensitivity has been substantiated in Kinoshita and coworkers' earlier studies.[20,33] For example, even when the protein-water electrostatic potentials are completely shut off, $|S_{VH}|$ of a protein decreases only by ~5%. Therefore, we model a protein as a set of fused hard spheres just for calculating S_{VH}.[10-16]

4.3.2. Angle-dependent integral equation theory for molecular liquids

The angle-dependent integral equation theory (ADIET)[29,31,32] is employed for calculating S_{VH} of a spherical solute (see step (1) described in Sec. 4.3.3). The water-water and solute-water potentials and correlation functions are dependent on the Euler angles representing the orientations

of water molecules. The details of basic equations and numerical solution procedure were described in Kinoshita and coworker's earlier publications.[29,34] S_{VH} is evaluated via the temperature derivative of μ calculated using the hypernetted-chain closure and the Morita-Hiroike formula adapted to molecular liquids.[29] By the ADIET combined with the multipolar water model, the dielectric constant of water is calculated to be 84 (this value is in close proximity with the experimental one 78).[29] Further, μ of a nonpolar solute calculated is in perfect agreement with that obtained from a Monte Carlo simulation.[29]

4.3.3. Calculation of hydration entropy of a protein with a prescribed structure

When a structure of a protein is given, its S_{VH} is calculated by a hybrid of the ADIET and the morphometric approach (MA).[35] In the MA, the geometric characteristics of a solute molecule are represented by only the four measures, V_{ex}, A, X, and Y.[35] V_{ex} is the EV, A is the water-accessible surface area (ASA), and X and Y are the integrated mean and Gaussian curvatures of the accessible surface, respectively. S_{VH} is expressed by the linear combination of the four geometric measures:[35]

$$S_{VH}/k_B = C_1V_{ex} + C_2A + C_3X + C_4Y. \quad (4.5)$$

The four coefficients (C_1–C_4) are dependent only on the thermodynamic state of the solvent and independent of the solute shape. Hence, they can be evaluated in the simplest geometries: for hard-sphere solutes (isolated atoms) with various diameters. We refer to the effect expressed by Eq. (4.5) as the EV effect. At normal temperature and pressure, C_1V_{ex} usually makes the largest contribution to S_{VH}/k_B.

The calculation consists of the four steps summarized below.[13-16]

(1) Calculate S_{VH} of a hard-sphere solute with diameter d_U immersed in the multipolar-model water using the ADIET. Consider different values of d_U in the range, $0.6 \le d_U/d_S \le 10$, to obtain a sufficiently large set of data for S_{VH} and R ($R = (d_U + d_S)/2$).
(2) Determine C_1–C_4 by applying the least-squares method to the following equation:

$$S_{VH}/k_B = C_1(4\pi R^3/3) + C_2(4\pi R^2) + C_3(4\pi R) + C_4(4\pi). \quad (4.6)$$

Equation (4.6) is the linear combination of the four geometric measures for spherical solutes. The determined values are as follows: $C_1 = -0.1968$ Å^{-3}, $C_2 = 0.0452$ Å^{-2}, $C_3 = 0.2567$ Å^{-1}, and $C_4 = -0.3569$.

(3) Calculate V_{ex}, A, X, and Y of a protein with a prescribed structure using an extended version[35] of Connolly's algorithm.[36,37] The x-y-z coordinates and diameters of the protein atoms are the input data. The diameter of each atom is set at the sigma value of the LJ potential parameters taken from CHARMM22.[38]
(4) Obtain S_{VH} from Eq. (4.5) to which C_1–C_4 determined in step (2) are substituted.

The maximum value of d_U considered in step (1) must be sufficiently large so that the effects

of the four geometric measures can fully be taken into account. Kinoshita and coworkers have verified that $d_U = 10d_S$ is large enough: Altering $10d_S$ to $30d_S$ results in essentially no changes in C_1–C_4 determined. They emphasize that protein molecules are not assumed to be ideally spherical. Step (4) is applicable to a variety of protein structures including random coils or fully extended structures. More detailed descriptions are provided in their earlier publications[10-12,35,39,40] two of which demonstrated the high accuracy of the MA.[35,40]

4.3.4. Calculation of energetic component for a protein with a prescribed structure

We briefly summarize the procedure for calculating the energetic component Λ (see the thermodynamic cycle illustrated in Fig. 4.3).[21,22] A fully extended structure is chosen as the reference one because it possesses the maximum number of HBs with water molecules and no intramolecular HBs: $\Lambda = 0$. Let us consider a transition from the reference structure to a more compact one. A gain of protein intramolecular interactions, loss of protein-water interactions, and energy change caused by structural reorganization of water molecules near the protein surface are assumed to be cancelled out except for the following: When the break of HBs with water molecules is not compensated by the formation of intramolecular HBs, a serious energetic increase is caused and to be taken into account. The torsion energy, which is sufficiently low for any structure considered, can be neglected.

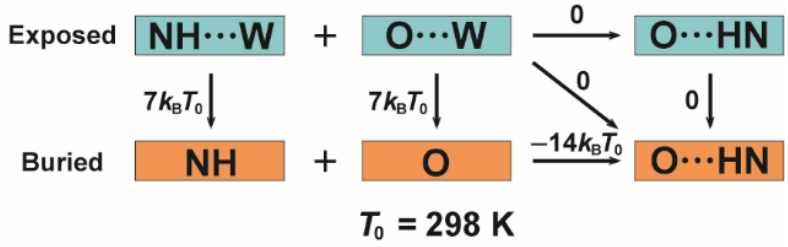

Fig. 4.3. Thermodynamic cycle for calculating the energetic component Λ. "W" and "···" represent a water molecule and a hydrogen bond, respectively, and k_B is Boltzmann's constant. This figure is illustrated for the case where N is the donor and O is the acceptor though there are four different donor-acceptor combinations: (N, O), (O, N), (O, O), and (N, N).

The structural transition defined above is followed by two cases: (i) A donor and an acceptor are buried in the protein interior after the break of HBs with water molecules, but they form an intramolecular HB; and (ii) a donor or an acceptor is buried but it finds no partner for an HB. No penalty is imposed in case (i) but a penalty of $7k_BT_0$ is imposed in case (ii).[21,22] $7k_BT_0$ is based on the estimation that the free-energy lowering brought by hydrogen-bond formation between two formamide molecules in a nonpolar solvent is $-14k_BT_0$.[41] The nonpolar solvent mimics the environment of protein interior, and $-14k_BT_0$ includes the entropic gain of the nonpolar solvent upon the hydrogen-bond formation. The ASA of each donor or acceptor is calculated using Connolly's algorithm.[36,37] The donor or acceptor is considered buried if its ASA is smaller than

0.001 Å^2. On the basis of the criteria proposed by McDonald and Thornton,[42] we determine whether an intramolecular HB is formed or not. All of the donors and acceptors are examined for backbone-backbone, backbone-side chain, and side chain-side chain intramolecular HBs to calculate A.[21,22]

4.3.5. Proteins and mutations considered

We choose the proteins treated in experiments under the condition which satisfies the following: The values of T_m and/or $\Delta\Delta G_{exp}$ were measured in aqueous solution whose pH is in the range from 6 to 8; no denaturants were utilized in the determination of T_m and/or $\Delta\Delta G_{exp}$; and the folded structures of the wild types were determined using the X-ray crystallography. The structure models obtained from NMR are excluded for the following reasons. The NMR models are constructed by a structure calculation upon which the structural information experimentally obtained as a set of constraints is imposed.[43] Typical constraints are the nuclear Overhauser effect (NOE), residual dipolar coupling (RDC), hydrogen bonding, and dihedral angle restraints. Unless the amount of constraints is sufficiently large, the models constructed are substantially influenced by the structure calculation employed. About 20–40 candidate models are usually prepared, but they are often significantly different from one another.[43] The prediction of the thermal-stability changes upon mutations is made under the condition that the wild-type structure is given whereas a mutant structure is unknown. Since this is a subtle task, it is required that the wild-type structure be sufficiently certain. We exclude the NMR models to concentrate on the investigation of our thermal-stability measure. It is true, however, that some NMR models give successful results while others do not, but this is beyond the scope of our work. (Of course, when the amount of constraints is large enough, the models from NMR can be better than those from the X-ray crystallography because the former accounts for the structure fluctuation in aqueous solution.[43])

We test 10 proteins[44-54] listed in Table 4.1. Four of them possess S-S bonds. The crystallization for the X-ray crystallography was made in aqueous solution whose pH was in the range from 6 to 8 with the exception of ribosomal protein L30e,[55] ribonuclease (RNase) HI,[56] and ribose binding protein:[57] Their pH-values were 5.6, 9.0, and 5.0, respectively. We could not find the pH-values in the crystallization for chicken lysozyme[58] and RNase A.[59] In order to confirm that the proteins considered are sufficiently diverse, we analyze the amino-acid sequence homology for all of the protein pairs using "CLUSTAL W".[60] For any two of RNase A, RNase HI, RNase Sa, and RNase T1, the homology is in the range 3–9%. For chicken lysozyme and T4 lysozyme, it is 11%. Among the 45 pairs, only 7 of them exhibit the homology exceeding 10%. Thus, the amino-acid sequences of the 10 proteins are quite different. The values of T_m and ΔG_{exp} for the wild types of the 10 proteins are collected in Table 4.2. ΔG_{exp}, the free-energy change upon *unfolding* at 25 °C, was calculated using the experimental data (see Secs. 4.4.2 and 4.4.3 for $\Delta\Delta G_{exp}$).

Table 4.1. Proteins and mutations considered. The multiple mutations are underlined.

Protein name (Number of S-S bonds)	PDB Code	Mutations (Total number)
Barnase[44]	1BNI	L14A, I88A, I96A, I88V, I96V (5)
Chicken lysozyme[54] (4)	4LYZ	T40S, T40I, I55L, I55V, I55M, I55F, I55A, I55T, S91T, S91V, S91A, S91D, S91Y, (T40S/I55V), (T40S/S91T), (T40S/S91V), (T40S/S91A), (I55L/S91T), (I55V/S91T), (I55V/S91A), (I55V/S91V), (I55A/S91T), (T40S/I55V/S91T), (T40S/I55V/S91A) (24)
Ribosomal protein L30e[49]	1H7M	D2A, E6A,R8A, K9A, D12A,K15A, R21A, K22A, K28A, K33A, R39A, R42A, D44A, K46A, E47A, D48A, E50A, R54A, E62A, E64A, E69A, R76A, H78A, D87A, E90A, R92A (26)
Ribonuclease A[52] (4)	1RTB	F46L, F46V, F46A (3)
Ribonuclease HI[47]	2RN2	H62A, H83A, H114A, H124A, (H62A/H83A), (H124A/H127A), (H62A/H83A/H124A/H127A) (7)
Ribonuclease Sa[46] (1)	1RGG	D79F, D79Y, D79A, D79I, D79R, D79L, D79K, D79W, D79H, D79N, D79E, Q94K, D33A (13)
Ribonuclease T1[53] (2)	1RN1	W59Y, Y24W, Y42W , Y45W, H40T, H92A , (Y24W/W59Y), (Y42W/W59Y), (Y45W/W59Y), (H40T/W59Y), (W59Y/H92A) (11)
Ribose-binding protein[48]	1URP	S9A, N13A, F15A, F16A, N64A, D89A, S103A, I132A, F164A, N190A, F214A, D215A, Q235A, (S9A/I132A), (S9A/I132A/S103A), (S9A/I132A/S103A/N13A), (S9A/I132A/S103A/N13A/Q235A) (17)
Staphylococcal nuclease[50,51]	1EY0	V23T, V39T, V51T, V66T, V74T, V99T, V104T, V111T, V114T, V39S, V51S, V66S, V114S, Y27F, Y54F, Y85F, Y91F, Y93F, Y113F, Y115F, Y54L, Y85L, Y91L, Y113L, Y115L, S128A , T13S, T22S, T33S, T41S, T44S, T62S, T82S, T120S, T13V, T22V, T33V, T41V, T44V, T62V, T82V, T120V, T13C, T22C, T33C, T41C, T44C, T62C, T82C, T120C, T22I, T33I, T41I, T44I, T62I, T82I, T120I, D19F, K28F, K48F, K49F, E52F, E57F, M65F, E67F, E73F, Q80F, K84F, E101F, A112F, K116F, E122F, Q123F, K127F, S128F, E135F, K136F, T62A, T62F, T62G, T62H, T62K, T62L, T62M, T62N, T62Q, (P117G/H124L/S128A), (T41I/P117G/H124L/S128A), (T33V/T41I/P117G/H124L/S128A), (T41I/S59A/P117G/H124L/S128A), (T33V/T41I/S59A/P117G/H124L/S128A) (91)
T4 Lysozyme[45]	1L63	N40A, K43A, S44A, E45A, L46A, D47A, K48A, (N40A/K43A/S44A/E45A/L46A/D47A/K48A), (N40A/S44A/E45A/D47A/K48A), (E45A/K48A) (10)

Table 4.2. Values of thermal denaturation temperature, T_m, and free-energy change upon *unfolding* at 25 °C, ΔG_{exp}, for wild types of the 10 proteins listed in Table 4.1. ΔG_{exp} was calculated using the experimental data, and ΔG_{exp} for staphylococcal nuclease was calculated at 20 °C. ΔG_{exp} is not presented in the literature for chicken lysozyme, RNase (ribonuclease) HI, RNase Sa, ribose-binding protein, and T4 Lysozyme.

Protein name	T_m (°C)	ΔG_{exp} (kcal/mol)
Barnase[44]	53.9	9.5
Chicken lysozyme[54]	74.0	–
Ribosomal protein L30e[49]	93.8	11.4
Ribonuclease A[52]	59.7	9.30
Ribonuclease HI[47]	50.2	–
Ribonuclease Sa[46]	47.8	–
Ribonuclease T1[53]	57.2	7.82
Ribose-binding protein[48]	57.5	–
Staphylococcal nuclease[50,51]	52.7	5.4
T4 Lysozyme[45]	62.2	–

The mutants considered for each protein are also listed in Table 4.1. A total of 207 mutations (176 are single, 18 are double, and 13 are triple and higher-fold) are chosen. It is observed in the table that we consider a variety of mutations including those changing the total charge (e.g., mutating from a nonpolar residue to a charged one): The number of such mutations is 81. The 5 multiple mutations in the case of staphylococcal nuclease[51] include the mutation from proline (Pro) to glycine (Gly). Chicken lysozyme,[54] RNase A,[52] RNase Sa,[46] and RNase T1[53] possess 4, 4, 1, and 2 S-S bonds, respectively, but the number of S-S bonds remains unchanged upon any mutation considered. By the mutation Pro→Gly, the unfolded state becomes more extended. The presence of an S-S bond makes the unfolded state less extended. There is a trend that proteins with S-S bonds and mutations including Pro or Gly are avoided due to the resulting lower prediction performance,[3,9] but we challenge such proteins and mutations. (Our result is discussed in Sec. 4.4.1.)

4.3.6. Preparation of folded state for wild type and mutant

Models of the folded structures of the wild types are taken from Protein Data Bank (PDB). After giving hydrogen atoms to the models using the CHARMM[61] and MMTSB[62] programs, we slightly modify the models using the energy minimization described in Kinoshita and coworkers' earlier publication[20] to remove unrealistic overlaps of protein atoms. The modification is carried out using the CHARMM22 parameters[38] with the CMAP correction[63] and the GBMV implicit solvent model.[64,65] We put positional restraints in the harmonic form on all of the heavy atoms during the energy minimization. The force constant for the restraints is $2M$ kcal/(mol·Å^2) where M is the mass

of each atom.

It is required that the thermal-stability change upon mutation be predicted without any experimentally determined structural data for the resultant mutant. Starting from the folded structure of the wild type taken from the PDB code as the template, we construct the folded-structure model of a mutant using Modeller[66] (Ver. 9.11). We generate 10 candidate models on the condition that the modification of the coordinates is limited to the protein atoms within the distance of 2 Å from the center of the mutated residue. The model with the lowest value of Kinoshita and coworkers' free-energy function F at $T = T_0$ (see Eq. (4.1)) is chosen as the best one. Since the 10 models are all compact, the protein conformational entropy need not be taken into consideration. When unrealistic overlaps of protein atoms are observed, they are removed in the manner mentioned above with the alteration that the minimization is terminated once the LJ potential energy becomes negative. This alteration is for preserving the original structure as much as possible.

4.3.7. Preparation of unfolded state for wild type and mutant

Structural properties of the unfolded state of a protein are not exactly known. A prevailing method for modeling the unfolded state is to generate a sufficiently large set of random coils. In Kinoshita and coworkers' earlier works,[19,20] it was shown that the previous measure of the thermal stability[19,20] retains its efficacy even when a small number of extended structures are employed as the model of the unfolded state.[20] Therefore, we adopt five extended structures. The new measure differs from the previous one in the respect that Λ is taken into account. Since Λ is calculated by regarding a fully extended structure as the reference one, the employment of a small number of extended structures is justifiable. All of the S-S bonds are broken in the preparation of the unfolded state. The effect of this treatment is discussed in Sec. 4.4.1. The main-chain dihedral angles (ϕ, ψ) of the 5 extended structures are in the range from (–130°, 130°) to (–170°, 170°) with a step of (–10°, 10°) (ω is set at 180°).[20] For Pro, however, (ϕ, ψ) is set at (–60°, 40°). The most probable conformer for each side chain is taken from Dunbrack's Backbone-Dependent Rotamer Library.[67,68] The extended structures are modified using the energy minimization explained above with the alteration that we put the position restraints only on C_α atoms to allow for more structural flexibility.

4.3.8. Performance measures defined for theoretical prediction method

In addition to the correlation coefficient between two quantities of interest (e.g., $\Delta\Sigma_0$ and ΔT_m in our theoretical approach), a variety of measures are examined in the performance test. Here, mutations which lead to higher and lower thermal stabilities are referred to as "good mutation" and "bad mutation", respectively. There are two cases: Case (a) where a mutation is actually (i.e., experimentally) a good one; and case (b) where a mutation is actually a bad one. In case (a), when it is predicted to be a good one, it is counted as a true positive (TP); and when it is predicted to be a bad one, it is counted as a false negative (FN).[69] In case (b), when it is predicted to be a bad one, it is counted as a true negative (TN); and when it is predicted to be a good one, it is counted as a false positive (FP).[69]

The performance measures for a prediction method are then defined as follows:[69]

$$\text{Accuracy} = M_A = (TP + TN)/(TP + FN + TN + FP), \quad (4.7a)$$
$$\text{Recall} = M_R = TP/(TP + FN), \quad (4.7b)$$
$$\text{Precision} = M_P = TP/(TP + FP), \quad (4.7c)$$
$$\text{Specificity} = M_S = TN/(TN + FP), \quad (4.7d)$$
$$\text{Negative predictive value} = M_N = TN/(TN + FN). \quad (4.7e)$$

‘‘Recall’’ is defined for the actual good mutations. It represents the proportion of those which are successfully predicted to be good mutations. ‘‘Precision’’ is defined for the predicted good mutations. It represents the proportion of those which are actually good mutations. ‘‘Specificity’’, which is defined for the actual bad mutations, represents the proportion of those which are successfully predicted to be bad mutations. ‘‘Negative predictive value’’ is defined for the predicted bad mutations and it represents the proportion of those which are actually bad mutations. “Accuracy” is defined for all of the mutations and the proportion of the successful predictions. The measures are concerned with whether the thermal stability becomes higher or lower and irrelevant to the degree of the stability change. We remark that M_A is most frequently considered.[8]

4.4. Results and discussion

4.4.1. Relation between $\Delta\Sigma_0$ and ΔT_m

ΔT_m is plotted against $\Delta\Sigma_0$ in Fig. 4.4(a). In a strict sense, $\Delta\Sigma_0$ is not linearly correlated to ΔT_m. However, negative $\Delta\Sigma_0$ implies negative ΔT_m and positive $\Delta\Sigma_0$ implies positive ΔT_m, and larger $|\Delta\Sigma_0|$ leads to larger $|\Delta T_m|$. Hence, the performance can be evaluated through the plot of ΔT_m against $\Delta\Sigma_0$ and the measures, in particular, M_A. The correlation coefficient in Fig. 4.4(a) is 0.409 and M_A is 73.4%. Even the 5 multiple mutations including the mutation Pro→Gly in the case of staphylococcal nuclease does not appreciably deviate from the plot. Further, the data points for the proteins possessing S-S bonds do not exhibit apparent deviation, either. When a protein possesses more S-S bonds, the unfolded state becomes less extended with the result that $|\Delta S_C|$ decreases for all T. This effect (effect I) makes the folded state more stable. By contrast, the EV of the unfolded state and the decrease in the EV upon folding become smaller, which is followed by decreased ΔS_{VH} for all T. This effect (effect II) makes the folded state less stable. It is not definite which of the two effects dominates. However, the number of S-S bonds remains unchanged upon any mutation considered, which should be responsible for the exhibition of no apparent deviation.

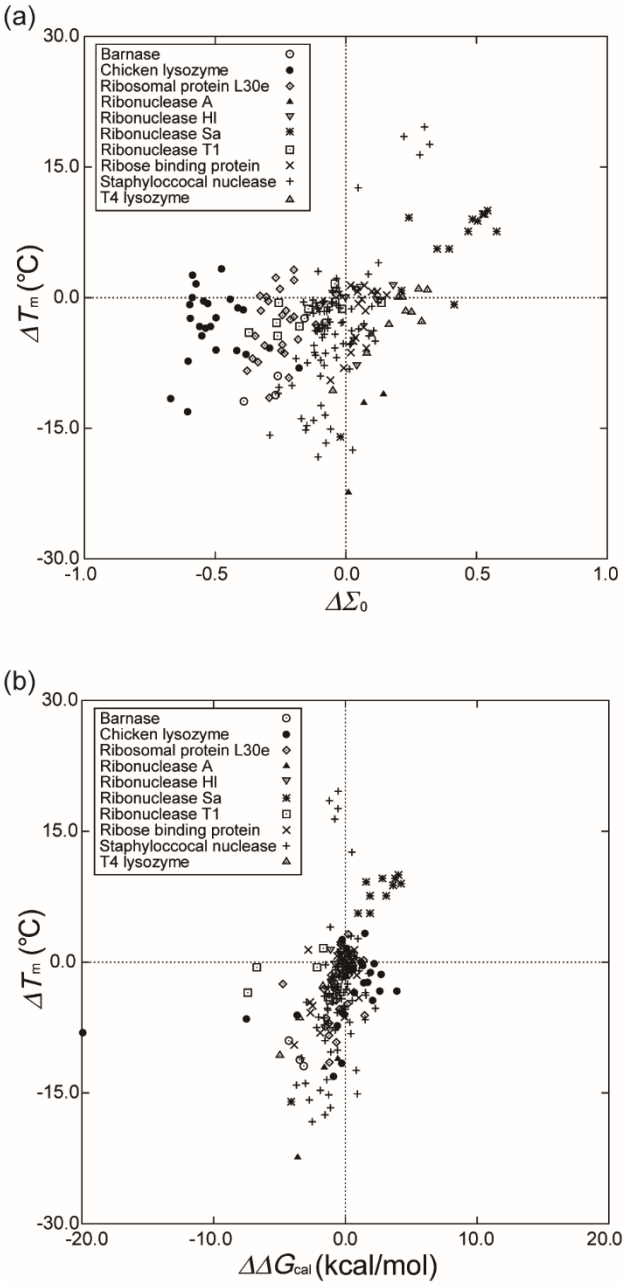

Fig. 4.4. (a) Relation between ΔT_m and $\Delta\Sigma_0$ calculated by our approach for the 207 mutants listed in Table 4.1. T_m is the thermal denaturation temperature, $\Sigma(T)$ is defined by Eq. (4.3), and $\Sigma_0 = \Sigma(T_0)$ (T_0 = 298 K). The unit of T_m is °C here. Δ denotes "the value for a mutant minus that for the wild type". (b) Relation between ΔT_m and $\Delta\Delta G_{cal}$. $\Delta\Delta G$ is the change in ΔG upon mutation ($\Delta\Delta G = \Delta G^M - \Delta G^W$; the superscripts, "W" and "M", denote values of the wild type and the mutant, respectively), ΔG ($\Delta G > 0$) is the free-energy change upon *unfolding*, and the subscript "cal" denotes the value calculated by FOLD-X.

4.4.2. Comparison with FOLD-X in terms of performance by looking at ΔT_m

FOLD-X, which is one of the most popular, successful approaches, can freely be used.[2,23,24] We use Version 3 that is essentially the same as the newest one, Version 4 (Version 4 is just easier to use than Version 3). We then compare our approach and FOLD-X in terms of the performance. It should be noted that all of the previously reported approaches including FOLD-X share roughly the same prediction performance.[5,9] In FOLD-X, the force field employed is different from ours and the folded states of the wild type and a mutant are also differently prepared. We use FOLD-X in its original way without any modification. We plot ΔT_m against $\Delta\Delta G_{cal}$ calculated by FOLD-X in Fig. 4.4(b). Here, $\Delta\Delta G$ is the change in ΔG upon mutation ($\Delta\Delta G = \Delta G^M - \Delta G^W$), ΔG ($\Delta G > 0$) is the free-energy change upon *unfolding* at 25 °C, and the subscript "cal" denotes the calculated value. Positive $\Delta\Delta G$ implies that the stability is enhanced by the mutation. Since $\Delta\Delta G$ in FOLD-X is defined as "$\Delta G^W - \Delta G^M$", the calculated value is multiplied by −1. The correlation coefficient in Fig. 4.4(b) is 0.395. In the previous tests, the correlation coefficient for FOLD-X was about 0.5 or better,[5] so the protein data set considered in the present study should be more difficult to theoretically treat. In Table 4.3, we compare FOLD-X with our approach in terms of the five performance measures defined by Eq. (4.7). All of the measures are considerably higher in ours than in FOLD-X. M_A, for instance, is 73.4% in ours and 66.7% in FOLD-X.

Table 4.3. Comparison between our approach and FOLD-X in terms of the performance measures defined by Eq. (4.7). The data points in Fig. 4.4 are considered.

Approach	M_A	M_R	M_P	M_S	M_N
Ours	0.734	0.623	0.485	0.773	0.856
FOLD-X	0.667	0.604	0.400	0.688	0.835

4.4.3. Comparison with FOLD-X in terms of performance by looking at $\Delta\Delta G_{exp}$

It is worthwhile to examine the relation between ΔT_m and $\Delta\Delta G_{exp}$. The subscript "exp" denotes the experimentally determined value. There is no data of $\Delta\Delta G_{exp}$ for chicken lysozyme, RNase HI, ribose binding protein, and single mutations of staphylococcal nuclease. There are 73 mutants with the data of $\Delta\Delta G_{exp}$ available. Cares must be taken in handling the data of $\Delta\Delta G_{exp}$ because ΔG was defined for folding in some proteins and for unfolding in others. Since the values of $\Delta\Delta G_{exp}$ were calculated at 52.6 °C for RNase Sa[46] and at 62.2 °C for T4 lysozyme[45] in the references, we recalculated them at 25 °C using the thermodynamic data given.[52] For the multiple mutations of staphylococcal nuclease,[51] the values of $\Delta\Delta G_{exp}$ were calculated at 20 °C in the reference. However, the thermodynamic data were not given. We decided to adopt the values at 20 °C because they should be very close to those at 25 °C. The relation between ΔT_m and $\Delta\Delta G_{exp}$ is plotted in Fig. 4.5. As expected, they are highly correlated: The correlation coefficient reaches 0.881.

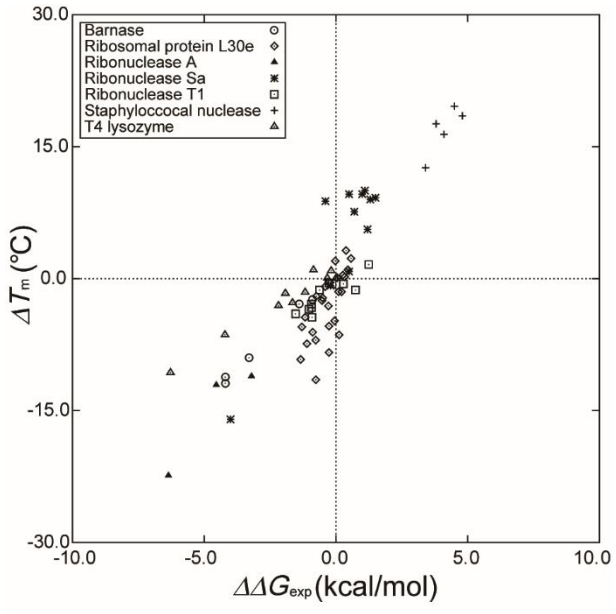

Fig. 4.5. Relation between ΔT_m and $\Delta\Delta G_{exp}$ for the 73 mutants with the data of $\Delta\Delta G_{exp}$ available. See the caption of Fig. 4.4 for the notation. The subscript "exp" denotes an experimental value.

We then employ $\Delta\Delta G_{exp}$ instead of ΔT_m in the performance evaluation. Figure 4.6(a) shows the plot of $\Delta\Delta G_{exp}$ against $\Delta\Sigma_0$ for our approach and Fig. 4.6(b) shows the plot of $\Delta\Delta G_{exp}$ against $\Delta\Delta G_{cal}$[2,23,24] for FOLD-X. The 73 mutants are considered (i.e., the other 134 mutants are excluded). The correlation coefficients in Figs. 4.6(a) and 4.6(b) are 0.314 and 0.473, respectively. The five performance measures for our approach and FOLD-X are compared in Table 4.4. Except for M_R, the measures are considerably higher in FOLD-X than in ours. M_A is 64.4% in ours and 71.2% in FOLD-X. Overall, the performance of FOLD-X is higher than ours. This result is not surprising because the weighting coefficients employed in FOLD-X are determined in advance so that $\Delta\Delta G_{cal}$ can be best fitted to $\Delta\Delta G_{exp}$ for a sufficiently large set of proteins and mutations. By contrast, our approach includes no such fitting procedure.

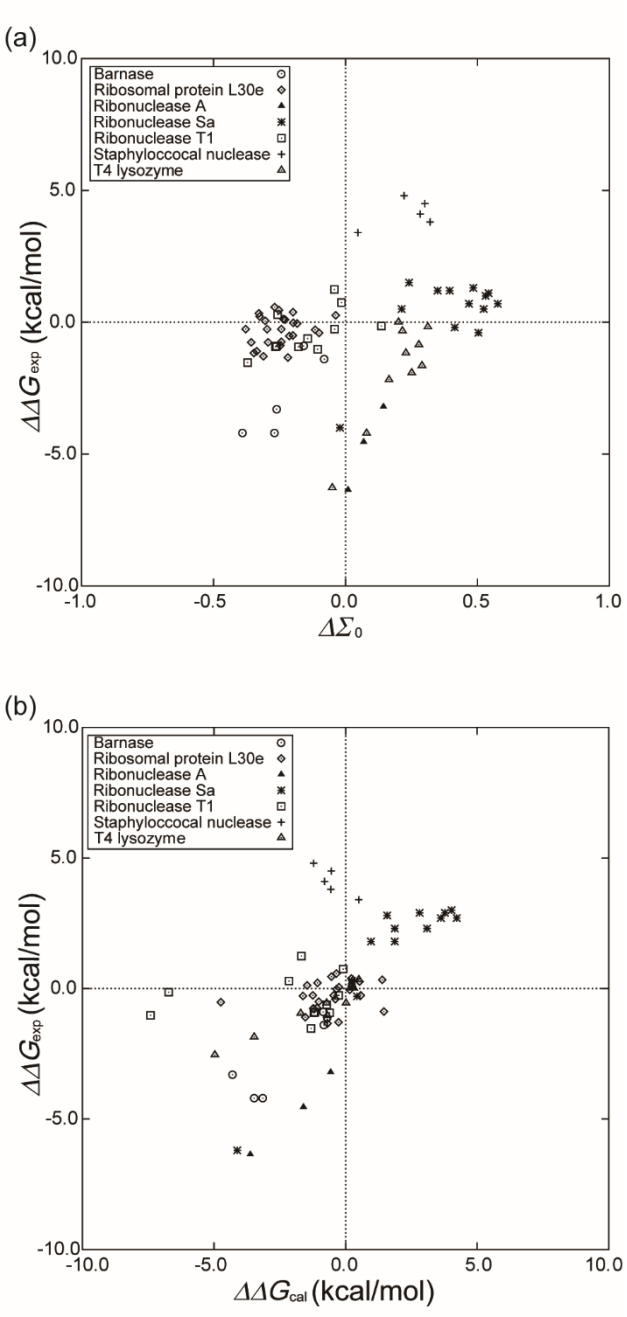

Fig. 4.6. (a) Relation between $\Delta\Delta G_{\mathrm{exp}}$ and $\Delta\Sigma_0$ calculated by our approach for the 73 mutants with the data of $\Delta\Delta G_{\mathrm{exp}}$ available. (b) Relation between $\Delta\Delta G_{\mathrm{exp}}$ and $\Delta\Delta G_{\mathrm{cal}}$. The subscript "cal" denotes the value calculated by FOLD-X. See the captions of Figs. 4.4 and 4.5 for the notation.

Table 4.4. Comparison between our approach and FOLD-X in terms of the performance measures defined by Eq. (4.7). The data points in Fig. 4.6 are considered.

Approach	M_A	M_R	M_P	M_S	M_N
Ours	0.644	0.571	0.533	0.689	0.721
FOLD-X	0.712	0.571	0.640	0.800	0.750

To confirm that the higher performance of FOLD-X mentioned above is attributed not to the exclusion of the 134 mutants but to the replacement of ΔT_m by $\Delta\Delta G_{exp}$, we plot ΔT_m against $\Delta\Sigma_0$ for ours and against $\Delta\Delta G_{cal}$ for FOLD-X only for the 73 mutants. The results are shown in Figs. 4.7(a) and 4.7(b) where the correlation coefficients are 0.583 and 0.540, respectively. FOLD-X exhibits lower performance with respect to the correlation coefficient. The five performance measures for ours and FOLD-X are compared in Table 4.5: M_R and M_N are higher in ours than in FOLD-X, but the opposite is true for the other three measures (M_A is 75.3% in ours and 79.5% in FOLD-X). However, the differences are not large.

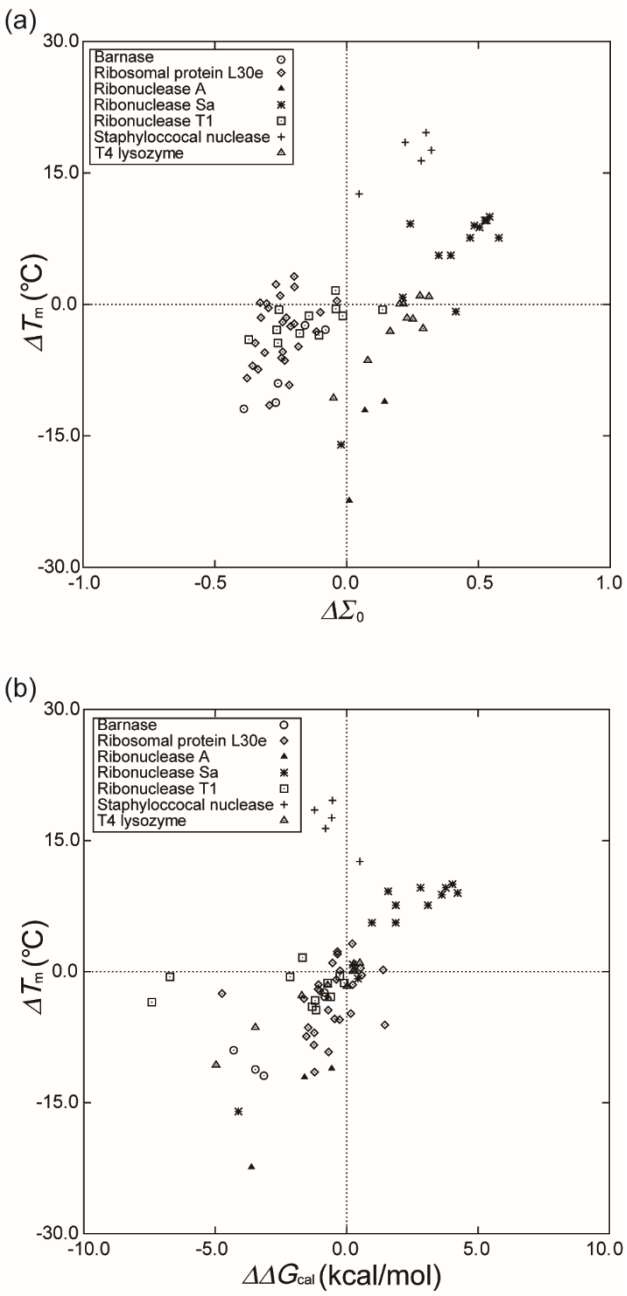

Fig. 4.7. (a) Relation between ΔT_m and $\Delta\Sigma_0$ calculated by our approach for the 73 mutants considered in Fig. 4.6. (b) Relation between ΔT_m and $\Delta\Delta G_{cal}$. The subscript "cal" denotes the value calculated by FOLD-X. See the caption of Fig. 4.4 for the notation.

Table 4.5. Comparison between our approach and FOLD-X in terms of the performance measures defined by Eq. (4.7). The data points in Fig. 4.7 are considered.

Approach	M_A	M_R	M_P	M_S	M_N
Ours	0.753	0.714	0.667	0.778	0.814
FOLD-X	0.795	0.679	0.760	0.867	0.813

In summary, our approach and FOLD-X are better suited to ΔT_m and $\Delta\Delta G_{exp}$, respectively, in the exhibition of higher performance. The reason why ours is better suited to ΔT_m may be the following. ΔT_m can directly be measured, whereas $\Delta\Delta G_{exp}$ at 25 °C is evaluated using the enthalpy and heat-capacity changes upon unfolding *at the denaturation temperature* and the assumption that the heat-capacity change is independent of T. Actually, the assumption is not quite correct because the heat-capacity change increases progressively as T becomes lower. (See Kinoshita's earlier publication[11] for a detailed discussion.) ΔT_m may be more reliable than $\Delta\Delta G_{exp}$.

4.4.4. Comparison with FOLD-X in terms of performance for multiple mutations

Taking only the 31 multiple mutations, we plot ΔT_m against $\Delta\Sigma_0$ for our approach and against $\Delta\Delta G_{cal}$ for FOLD-X in Figs. 4.8(a) and 4.8(b), respectively. The correlation coefficient in Fig. 4.8(a) is 0.549 whereas that in Fig. 4.8(b) is only 0.024. The correlation coefficient in FOLD-X for the 18 double mutations is 0.177 whereas that for the triple and higher-fold mutations is *negative*. There is a strong trend in FOLD-X that the prediction performance becomes progressively worse as the time of the mutations increases. In Table 4.6, we compare FOLD-X with our approach in terms of the five performance measures. All of the measures in ours are much higher than those in FOLD-X. M_A is 74.2% in ours and 48.4% in FOLD-X.

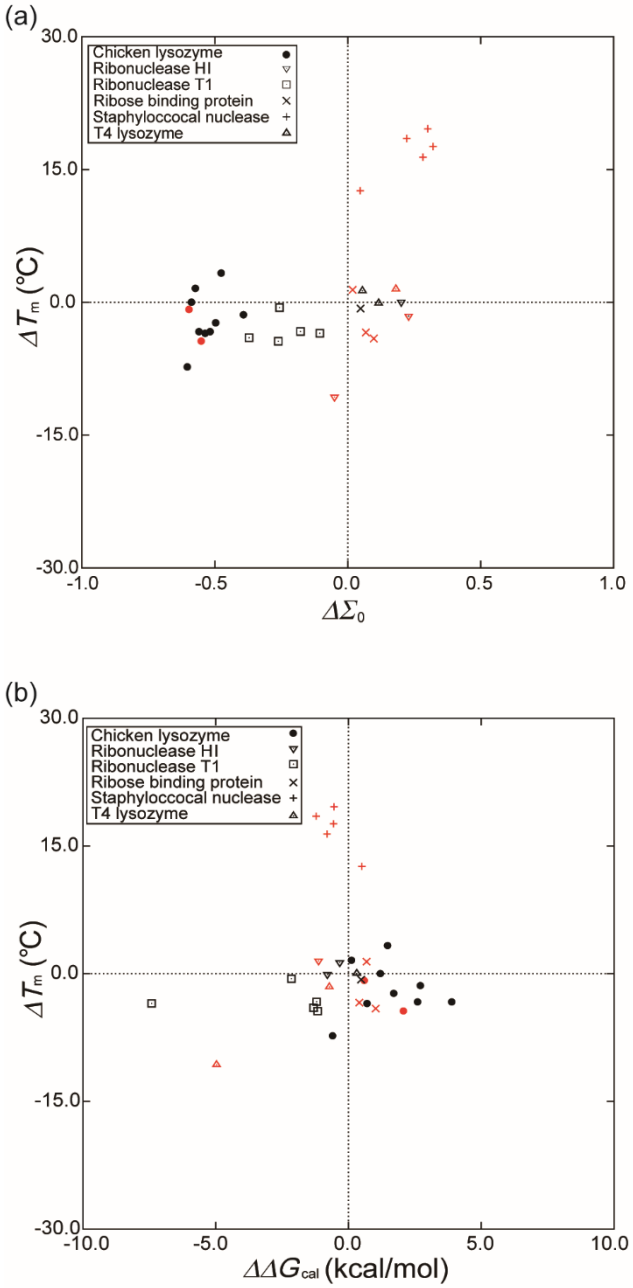

Fig. 4.8. (a) Relation between ΔT_{m} and $\Delta\Sigma_0$ calculated by our approach for the 31 multiple mutants in Table 4.1. (b) Relation between ΔT_{m} and $\Delta\Delta G_{\mathrm{cal}}$. The subscript "cal" denotes the value calculated by FOLD-X. The keys for the triple and higher-fold mutations are drawn in red. See the caption of Fig. 4.4 for the notation.

Table 4.6. Comparison between our approach and FOLD-X in terms of the performance measures defined by Eq. (4.7). The data points in Fig. 4.8 are considered.

Approach	M_A	M_R	M_P	M_S	M_N
Ours	0.742	0.750	0.643	0.737	0.824
FOLD-X	0.484	0.500	0.375	0.474	0.600

There are two possible reasons for the ill success of FOLD-X for multiple mutations. The first reason is that adjustment of the terms in its free-energy function and determination of the weighting coefficient in the van der Waals term are performed only for single mutations.[2,23,24] The second one is much more important. As observed in Fig. 4.8, T_m of staphylococcal nuclease is made higher by 12.6 to 19.6 °C due to the 5 multiple mutations.[51] FOLD-X fails to reproduce these highly enhanced stabilities. By contrast, our approach is successful in reproducing them. We find the following: As shown in Fig. 4.9, this success is thanks to the entropic component of $\Delta\Sigma_0$, $\Delta S_{VH}/(k_B N_r)$; and the energetic component, $-\Delta\Lambda/(k_B T_0 N_r)$, works to *lower* the stabilities. Thus, the highly enhanced stabilities are brought by the entropic EV effect.

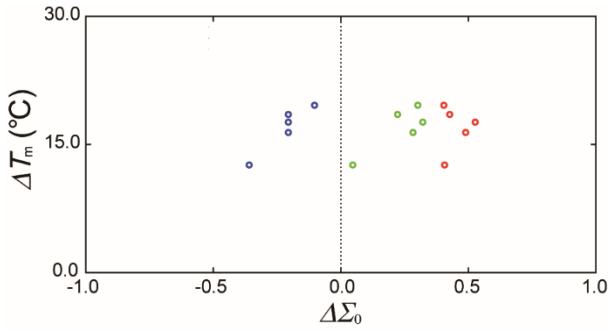

Fig. 4.9. Relation between ΔT_m and $\Delta\Sigma_0$ calculated by our approach for the 5 multiple mutations of staphylococcal nuclease. Red: Only the entropic component of $\Delta\Sigma_0$ is considered; $\Delta\Sigma_0 = \Delta S_{VH}/(k_B N_r)$. Blue: Only the energetic component of $\Delta\Sigma_0$ is considered; $\Delta\Sigma_0 = -\Delta\Lambda/(k_B T_0 N_r)$. Green: Both of the entropic and energetic components of $\Delta\Sigma_0$ are considered; $\Delta\Sigma_0 = \Delta S_{VH}/(k_B N_r) - \Delta\Lambda/(k_B T_0 N_r)$. This figure should be compared with the data points for the five multiple mutations of staphylococcal nuclease in Fig. 4.8(b).

The free-energy function of FOLD-X is parameterized using only the experimental data for single mutations. If it were re-parameterized using the experimental data including those for multiple mutations, the result from FOLD-X would certainly be improved. It should be emphasized, however, that the entropic EV effect is not incorporated in the free-energy function of FOLD-X. Without incorporating this crucial effect, even if the result is improved by the re-parameterization, the improvement is not physically sound.

4.4.5. Cases where structural data for folded states of mutants are experimentally available

The performance of our approach will become higher if the preparation method for the folded state of a mutant is improved. As an illustration, we limit the prediction to the mutants whose structural data for folded states are experimentally available (a total of 22 mutants). The experimental data are used for these mutants. The method of preparing the folded structure for each mutant is the same as that for the wild type described in Sec. 4.3.6. We find that the correlation coefficient between $\Delta T_{\rm m}$ and $\Delta\Sigma_0$ reaches 0.678.

There are two different data for chicken lysozyme: The PDB codes are 4LYZ[54] and 1HEL.[70] The two data give the results which are different in a quantitative sense. In all of the sections described above, we adopt 4LYZ leading to better performance-test results for both of our approach and FOLD-X. When the structural data experimentally determined are used for the mutants, however, 1HEL is adopted because the result for chicken lysozyme is quantitatively better for 1HEL.

4.5. Concluding remarks

We have examined a new measure of the thermal stability of a protein by applying it to the prediction of the thermal-stability changes upon mutations. In the measure, the entropic excluded-volume (EV) effect[10-16] is incorporated to its full extent using a molecular model for water. A unique approach has thus been obtained for the prediction using the measure. The performance of our approach would become much higher by the introduction of fitting parameters (e.g., by optimizing two weighting coefficients multiplied to $T\Delta S_{\rm VH}$ and to $\Delta\Lambda$ in Eq. (4.3), respectively, by adjusting $7k_{\rm B}T_0$ or $-14k_{\rm B}T_0$ in Fig. 4.3, etc.), in which case, however, the thermal-stability measure or the free-energy function becomes somewhat physically ambiguous. Hence, the use of those parameters is avoided.

We have compared our approach with FOLD-X[2,23,24] in terms of the prediction performance for 10 proteins and a total of 207 mutations including 31 multiple mutations. Further, 4 of the 10 proteins possess S-S bonds and the 5 multiple mutations for staphylococcal nuclease[51] include Pro→Gly, which presents much challenge. FOLD-X is one of the most popular, successful approaches. Our approach and FOLD-X exhibit almost the same performance on the whole, despite that ours employ no fitting parameters unlike FOLD-X. The factors taken into account in our thermal-stability measure are the water-entropy gain, loss of the protein conformational entropy, break of protein-water hydrogen bonds (HBs), and formation of protein intramolecular HBs upon folding (i.e., the protein thermal stability is described by the competition of $\Sigma(T)$ and $T|\Delta S_{\rm C}|/(k_{\rm B}T_0N_{\rm r})$ as illustrated in Fig. 4.2). In our approach, when the structures of folded states of the mutants are experimentally available and they are utilized, the correlation coefficient between $\Delta T_{\rm m}$ and $\Delta\Sigma_0$ reaches 0.678. These successful results indicate that the three physicochemical factors govern the thermal-stability changes upon mutations.

For multiple mutants, ours is far superior to FOLD-X. This superiority is important, because most of the mutations realizing remarkable enhancement of the stability are multiple mutations. For

example, T_m of staphylococcal nuclease could be made higher by 12.6 to 19.6 °C only by multiple mutations.[51] Moreover, even when two single mutations, (a) and (b), lead to higher thermal stability, the double mutation of (a) plus (b) do not necessarily result in further higher stability. It is thus crucial to successfully predict the stability change brought by a multiple mutation. We have argued that the highly enhanced stabilities for five multiple mutations of staphylococcal nuclease are realized by the entropic EV effect and that a principal reason for the ill success of FOLD-X is the neglect of this effect. We believe that the entropic EV effect, which is quite sensitive to the protein structural changes, generally plays a pivotal role in achieving high enhancement of the stability by means of mutation. It is worthwhile to incorporate this effect in the free-energy function of FOLD-X and re-parameterize it using the experimental data including those for multiple mutations.

Khechinashvili *et al.*[71-73] made thermodynamic analyses on thermal denaturation of mesophilic and thermophilic proteins and draw the following conclusions: The thermal stability is not correlated with the protein intramolecular energy and it has entropic nature; and the conformational-entropy gain upon denaturation becomes more substantial as the temperature increases, which plays a pivotal role in the inducement of thermal denaturation. The first conclusion is consistent with our claim that the water-entropy effect is a crucially important factor in the thermal stability, and the second conclusion is in line with our physical picture of thermal denaturation illustrated in Fig. 4.2.

In the present article, we are interested in not only the physicochemical aspects of the thermal-stability changes upon mutations but also the prediction performance. A future study pursuing higher performance is of value: It can be obtained by improving the energetic and entropic components of the free-energy function. In particular, the protonation states of some residues might be changeable during the folding or unfolding process, and this effect should be examined. Our method of preparing the folded structure of a mutant seems to be cruder than that employed in FOLD-X. In fact, when the experimentally determined mutant structures are used (when they are available), the performance of our method is considerably improved as explained in Sec. 4.4.5. The simplified model of the unfolded state (i.e., a set of five extended structures described in Sec. 4.3.8) is also to be reconsidered.

References

[1]A. Mozo-Villiarías and E. Querol, Curr. Bioinformatics **1**, 25 (2006).

[2]R. Guerois, J. E. Nielsen, and L. Serrano, J. Mol. Biol. **320**, 369 (2002).

[3]N. Pokala, and T. M. Handel, J. Mol. Biol. **347**, 203 (2005).

[4]S. Yin, F. Ding, and N. V. Dokholyan, Nat. Methods **4**, 466 (2007).

[5]V. Potapov, M. Cohen, and G. Schreiber, Protein Eng. Des. Sel. **22**, 553 (2009).

[6]A. Benedix, C. M. Becker, B. L. de Groot, A. Caflisch, and R. A. Böckmann, Nat. Methods **6**, 3 (2009).

[7]Z. Zhang, L. Wang, Y. Gao, J. Zhang, M. Zhenirovskyy, and E. Alexov, Bioinformatics **28**, 664 (2012).

[8]L. Wickstrom, E. Gallicchio, and R. M. Levy, Proteins **80**, 111 (2012).

[9]A. N. Naganathan, J. Phys. Chem. B **117**, 4956 (2013).

[10]M. Kinoshita, Front. Biosci. **14**, 3419 (2009).

[11]M. Kinoshita, Int. J. Mol. Sci. **10**, 1064 (2009).

[12]M. Kinoshita, Biophys. Rev. **5**, 283 (2013).

[13]T. Yoshidome, M. Kinoshita, S. Hirota, N. Baden, and M. Terazima, J. Chem. Phys. **128**, 225104 (2008).

[14]T. Yoshidome and M. Kinoshita, Phys. Chem. Chem. Phys. **14**, 14554 (2012).

[15]Y. Harano, T. Yoshidome, and M. Kinoshita, J. Chem. Phys. **129**, 145103 (2008).

[16]H. Oshima and M. Kinoshita, J. Chem. Phys. **138**, 245101 (2013).

[17]S.-H. Chong and S. Ham, Chem. Phys. Lett. **535**, 152 (2012).

[18]G. Graziano, Phys. Chem. Chem. Phys. **16**, 21755 (2014).

[19]K. Amano, T. Yoshidome, Y. Harano, K. Oda, and M. Kinoshita, Chem. Phys. Lett. **474**, 190 (2009).

[20]K. Oda, R. Kodama, T. Yoshidome, M. Yamanaka, Y. Sambongi, and M. Kinoshita, J. Chem. Phys. **134**, 025101 (2011).

[21]T. Yoshidome, K. Oda, Y. Harano, R. Roth, Y. Sugita, M. Ikeguchi, and M. Kinoshita, Proteins **77**, 950 (2009).

[22]S. Yasuda, T. Yoshidome, Y. Harano, R. Roth, H. Oshima, K. Oda, Y. Sugita, M. Ikeguchi, and M. Kinoshita, Proteins **79**, 2161 (2011).

[23]J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, Nucleic Acids Res. **33**, 382 (2005).

[24]J. W. H. Schymkowitz, F. Rousseau, I. C. Martins, J. Ferkinghoff-Borg, F. Stricher, and L. Serrano, Proc. Natl. Acad. Sci. USA **102**, 10147 (2005).

[25]S. Yasuda, H. Oshima, and M. Kinoshita, J. Chem. Phys. **137**, 135103 (2012).

[26]B. Honig and A. Nicholls, Science **268**, 1144 (1995).

[27]N. Baden, S. Hirota, T. Takabe, N. Funasaki, and M. Terazima, J. Chem. Phys. **127**, 175103 (2007).

[28]N. M. Cann and G. N. Patey, J. Chem. Phys. **106**, 8165 (1997).

[29]M. Kinoshita, J. Chem. Phys. **128**, 024507 (2008).

[30]J. Fitter, Biophys. J. **84**, 3924 (2003).
[31]P. G. Kusalik, and G. N. Patey, J. Chem. Phys. **88**, 7715 (1988).
[32]P. G. Kusalik, and G. N. Patey, Mol. Phys. **65**, 1105 (1988).
[33]T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata, J. Chem. Phys. **125**, 024911 (2006).
[34]M. Kinoshita and D. R. Bérard, J. Comput. Phys. **124**, 230 (1996).
[35]R. Roth, Y. Harano, and M. Kinoshita, Phys. Rev. Lett. **97**, 078101 (2006).
[36]M. L. Connolly, J. Appl. Crystallogr. **16**, 548 (1983).
[37]M. L. Connolly, J. Am. Chem. Soc. **107**, 1118 (1985).
[38]A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[39]T. Hayashi, H. Oshima, T. Mashima, T. Nagata, M. Katahira, M. Kinoshita, Nucleic Acids Res. **42**, 6861 (2014).
[40]H. Oshima and M. Kinoshita, J. Chem. Phys. **142**, 145103 (2015).
[41]S. F. Sneddon, D. J. Tobias, and C. L. Brooks III, J. Mol. Biol. **209**, 817 (1989).
[42]I. K. McDonald and J. M. Thornton, J. Mol. Biol. **238**, 777 (1994).
[43]H. Mishima, S. Yasuda, T. Yoshidome, H. Oshima, Y. Harano, M. Ikeguchi, and M. Kinoshita, J. Phys. Chem. B **116**, 7776 (2012).
[44]J. T. Kellis, Jr., K. Nyberg, and A. R. Fersht, Biochemistry **28**, 4914 (1989).
[45]D. W. Heinz, W. A. Baase, and B. W. Matthews, Proc. Natl. Acad. Sci. USA **89**, 3751 (1992).
[46]S. R. Trevino, K. Gokulan, S. Newsom, R. L. Thurlkill, K. L. Shaw, V. A. Mitkevich , A. A. Makarov, J. C. Sacchettini J. M. Scholtz and C. N. Pace, J. Mol. Biol. **354**, 967 (2005).
[47]S. Kanaya, M. Oobatake, H. Nakamura, and M. Ikehara, J. Biotechnol. **28**, 117 (1993).
[48]N. C. Vercillo, K. J. Herald, J. M. Fox, B. S. Der, and J. D. Dattelbaum, Protein Sci. **16**, 362 (2007).
[49]C.-F. Lee, G. I. Makhatadze, and K.-B. Wong, Biochemistry **44**, 16817 (2005).
[50]M. P. Byrne and W. E. Stites, Biophys. Chem. **125**, 490 (2007).
[51]J. Chen, Z. Lu, J. Sakon, and W. E. Stites, J. Mol. Biol. **303**, 125 (2000).
[52]T. Kadonosono, E. Chatani, R. Hayashi, H. Moriyama, and T. Ueki, Biochemistry **42**, 10651 (2003).
[53]W.-D. Schubert, G. Schluckebier, J. Backmann, J. Granzin, C. Kisker, H. -W. Choe, U.Hahn, W. Pfeil, and W. Saenger, Eur. J. Biochem. **220**, 527 (1994).
[54]P. Shih, D. R. Holland, and J. F. Kirsch, Protein Sci. **4**, 2050 (1995).
[55]Y. W. Chen, M. Bycroft, and K.-B.Wong, Biochemistry **42**, 2857 (2003).
[56]S. Kanaya, A. Kohara, M. Miyagawa, T. Matsuzaki, K. Morikawa, and M. Ikehara, J. Biol. Chem. **264**, 11546 (1989).
[57]A. J. Björkman and S. L. Mowbray, J. Mol. Biol. **279**, 651 (1998).
[58]R. Diamond, J. Mol. Biol. **82**, 371 (1974).

[59]D. L. Birdsall and A. McPherson, J. Biol. Chem. **267**, 22230 (1992).
[60]J. D. Thompson, D. G. Higgins, and T. J. Gibson, Nucleic Acids Res. **22**, 4673 (1994).
[61]B. R. Brooks, R. E. Bruccoleri , B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[62]M. Feig, J. Karanicolas, and C. L. Brooks III, J. Mol. Graphics Modell. **22**, 377 (2004).
[63]A. D. Mackerell, Jr., M. Feig, and C. L. Brooks III, J. Comput. Chem. **25**, 1400 (2004).
[64]M. S. Lee, M. Feig, F. R. Salsbury, Jr., and C. L. Brooks III, J. Comput. Chem. **24**, 1348 (2003).
[65]J. Chocholoušová and M. Feig, J. Comput. Chem. **27**, 719 (2006).
[66]A. Šali and T. L. Blundell, J. Mol. Biol. **234**, 779 (1993).
[67]R. L. Dunbrack, Jr. and F. E. Cohen, Protein Sci. **6**, 1661 (1997).
[68]R. L. Dunbrack, Jr., Curr. Opin. Struct. Biol. **12**, 431 (2002).
[69]H. Oshima, S. Yasuda, T. Yoshidome, M. Ikeguchi, and M. Kinoshita, Phys. Chem. Chem. Phys. **13**, 16236 (2011).
[70]K. P. Wilson, B. A. Malcolm, and B. W. Matthews, J. Biol. Chem. **267**, 10842 (1992).
[71]N. N. Khechinashvili, M. V. Fedorov, A. V. Kabanov, S. Monti, C. Ghio, and K. Soda, J. Biomolec. Struct. & Dynamics **24**, 255 (2006).
[72]N. N. Khechinashvili, S. A. Volchkov, A. V. Kabanov, and G. Barone, Biochim. Biophys. Acta **1784**, 1830 (2008).
[73]N. N. Khechinashvili, A. V. Kabanov, M. S. Kondratyev, and R. V. Polozov, J. Biomolec. Struct. & Dynamics **32**, 1396 (2014).

Chapter 5

General Conclusion

In this study, we have investigated the effects of cosolvent or salt addition to water and of an amino-acid mutation on the thermal stability of a protein. A rigid-body model, in which water molecules, cosolvent molecules, and ions are treated as hard spheres with different diameters and a protein comprises a set of fused hard spheres representing its polyatomic structure. In the rigid-body model, all of the accessible system configurations share the same energy, and the system behavior is purely entropic in origin. Hereafter, "solvent" is formed by water molecules for pure water, by water and cosolvent molecules for water-cosolvent solution, and by water molecules, cations, and anions for water-salt solution. The diameters and packing fractions of the solvent particles are carefully determined on the basis of the interaction potentials or X-ray electron density measurements and the experimental data of solution density. The effects of strength of ion hydration, cosolvent-water affinity, and change in the water structure in the bulk are implicitly incorporated in the packing fractions: We exclusively investigate the solvent-entropy effect for *a real solution*. An advantage of the rigid-body model is that the theoretical analysis can be focused on the entropic effect originating from the translational displacement of solvent molecules coexisting with the protein in the system, i.e., the entropic excluded-volume (EV) effect (in fact, this is the true physical meaning of the hydrophobic effect). In the first step, we explore to what extent the rigid-body model can reproduce the experimental data. Of course, it fails to reproduce some of the data. In the second step, we think about the reason for the failure and identify a factor or factors to be considered further. We employ the integral equation theory (IET) combined with the morphometric approach (MA). This combination enables us to decompose the solvation entropy into physically insightful constituents and clarify the physical origins of the results obtained. In Chapter 4 where the mutation for a protein is considered, the angle-dependent integral equation theory (ADIET) applied to a realistic multipolar model for water is adopted, because of the failure of the hard-sphere model for water.

In Chapter 2, we have investigated the effects of adding urea, monohydric alcohols, and polyols to water on the thermal stability of a protein. Methanol, ethanol, and 2-popanol are considered for monohydric alcohols, and ethylene glycol (2), glycerol (3), erythritol (4), xylitol (5), and mannitol (6) (the figure in the parentheses denotes the number of hydroxyl groups in a molecule) are considered for polyols. It is argued that the solvent-entropy gain upon protein folding, which is calculated at 298 K and normalized by the number of residues, can be a measure of the thermal stability. It is shown that the addition of monohydric alcohol lowers the thermal stability and the degree of the lowering follows the order, methanol<ethanol<2-propanol: The degree becomes larger as the size of hydrophobic group in an alcohol molecule increases. The addition of polyol enhances the thermal stability and the degree of the enhancement follows the order, ethylene glycol<glycerol<erythritol<xylitol<mannitol: The enhancement becomes stronger as the number of

hydroxyl groups increases. These results are in qualitatively perfect agreement with the experimental observations. We also find that the change in the measure calculated is highly correlated with ΔT_m (the change in the denaturation temperature T_m) experimentally measured for collagen (the correlation coefficient exceeds 0.97). For the addition of urea, however, our method predicts that it enhances the thermal stability, which is in conflict with the experimentally known fact. The lowered stability caused by the urea addition is ascribed to a factor other than the entropic EV effect.

Kinoshita and coworkers showed that a protein is driven to fold by the reduction of water crowding. "Water crowding" is entropic correlation among water molecules which arises from the factor that the presence of a water molecule generates an EV for the other water molecules. In the change of thermal stability upon cosolvent or salt addition, the mitigated or strengthened water crowding (i.e., the change in the hydrophobic effect) should also be an essential physical factor. However, other physical factors potentially affect the thermal-stability change, as manifested in the urea case considered in Chapter 2.

In Chapter 3, we have returned to a much simpler issue, cosolvent or salt effect on the solubility of a small nonpolar solute in water. Decrease and increase in the solubility, respectively, are ascribed solely to enhancement and reduction of the hydrophobic effect. Investigating the influence of adding a salt or cosolvent to water on the solubility allows us to concentrate on the hydrophobic effect. Again, the rigid-body model is adopted: The nonpolar solute is also modeled as a hard sphere. We calculate the solvation entropies of argon or methane in pure water, water-salt solution, and water-cosolvent solution. The solute solubility is governed by the solvation entropy. Alkali halides and $[(CH_3)_4N]Br$ are considered as the salts and methanol, ethanol, 1-propanol, and urea as the cosolvents. The major results obtained are as follows: (i) The addition of alkali halide decreases the solubility and this salting-out ability follows the orders, $Na^+ > K^+ > Cs^+ > Li^+$, $Cl^- > Br^- > I^-$, and NaCl>NaBr>KCl>NaI>CsCl>LiCl; (ii) $[(CH_3)_4N]Br$ increases the solubility; (iii) methanol, ethanol, 1-propanol increase the solubility in the order, 1-propanol>ethanol>methanol; and (iv) urea decreases the solubility. There results are in qualitatively perfect accord with the experimental observations. By comparing the experimental data of the solubility changes for argon and methane upon cosolvent or salt addition with those of the thermal-stability changes for several proteins, we notice that the cosolvent or salt addition which decreases the solubility enhances the thermal stability (i.e., raises T_m) and the cosolvent or salt addition which increases the solubility lowers the thermal stability, *when the hydrophobic effect dominates*. For urea, however, its addition decreases the solubility but lowers T_m. Qualitatively the same discrepancy is observed for anions with large sizes, Br^- and I^-, when a protein with a considerably large, positive total charge. In these cases, a physical factor other than the hydrophobic effect needs to be taken into account for the change in thermal stability of a protein. By examining molecular dynamics simulation results in the literature and performing some fundamental calculations by ourselves, we conclude that the factor is protein-solvent van der Waals interaction energy. Urea is known to be significantly enriched near the protein surface. Significant enrichment of Br^- and I^- near positively charged groups on the

protein surface should also occur. Many water molecules near the surface are replaced by urea molecules or the large anions. Since a urea molecule, Br^-, or I^- features very large ε (one of the Lennard-Jones parameters; larger ε of an ion or atom in a cosolvent molecule leads to stronger protein-ion or protein-cosolvent van der Waals interaction), the replacement leads to significant lowering of protein-solvent van der Waals interaction energy. A protein structure with a larger solvent-accessible surface area is then more favored, giving rise to deteriorated thermal stability. Despite the replacement of some of the water molecules near the protein surface by urea molecules, the total protein-water, protein-urea, water-water, water-urea, and urea-urea electrostatic interaction energy in water-urea solution is not significantly different from the total protein-water and water-water electrostatic interaction energy in pure water. The same argument can be made for the large anions.

A significant progress has been achieved in understanding the Hofmeister series in Chapter 3. When the hydrophobic effect is dominant, the solvent crowding in the bulk is made more serious by the addition of alkali halides. The series in terms of the power are expressed by $Na^+>K^+>Cs^+>Li^+$ and $Cl^->Br^->I^-$. These orders have been observed in experiments for the salting out of nonpolar solutes and the enhanced thermal stability for several proteins with zero and negative net charges. However, the series for cations was originally quoted in the order $Li^+>Na^+>K^+$ in the salting-out ability for hen egg white protein. Since this protein possesses significantly many groups with negative charges, its solubility is reduced when the negative charges are more screened by cations. As shown by Kinoshita and coworkers for the cations, the screening effect becomes larger as the cation size decreases. The decrease in protein solubility becomes larger as its negative charges are more screened, leading to the order $Li^+>Na^+>K^+$. This order is also observed in the salting out of benzoic acid with oxygen atoms with considerably large negative partial charges. Thus, *the mechanism of salting out of a charged solute is different from that of a nonpolar one*: Enhanced hydrophobic effect and screening of the charge by counterions are responsible for the latter and for the former, respectively. Further, proteins with many positively charged groups are extremely unfolded under conditions of acidic pH and low ionic strength but refolded to molten-globule-like conformations by the salt addition. The power of inducing the conformational transition, which arises from the screening of positive charges by the anions, follows the order $I^->Br^->Cl^-$, the reverse of $Cl^->Br^->I^-$. This is because for the anions the screening effect becomes larger as the anion size increases. Thus, the series describing the ion effect on a physicochemical quantity is exhibited in rather diverse ways because multiple physical factors can possibly come into play, but this exhibition is reasonably interpretable.

In Chapters 2 and 3, the solvation entropy is decomposed into the solute-solvent pair correlation component and the solute-solvent-solvent triplet and higher-order (i.e., many-body) correlation component. Each component is further decomposed into the terms scaled by the EV and the solvent-accessible surface, respectively. By analyzing a total of four constituents of the solvation entropy thus obtained, the following conclusion has been drawn: The many-body correlation component is substantially larger than the pair correlation one, in particular, the effect of

solvent crowding is essential; and when the hydrophobic effect dominates, enhanced thermal stability of a protein as well as decreased solubility of a small nonpolar solute is ascribed to more serious solvent crowding caused by the cosolvent or salt addition. Kinoshita and coworkers showed that the mechanisms of folding and thermal, cold, and pressure denaturating of a protein can be elucidated in a unified manner by their theoretical method wherein the effect of water crowding is treated as the key factor. In this study, it has newly been shown that the cosolvent and salt effects on the thermal stability can also be reproduced by the same method with the exceptions mentioned above. We propose that the effects of other popular cosolvents can be argued within the same theoretical framework. For example, urea and guanidine hydrochloride share the same mechanism, and polyols, sugars, and trimethylamine-N-oxide also share the same one. This proposition should be corroborated in future studies.

In Chapter 4, we have tackled a subtle subject, theoretical prediction of the thermal-stability change upon mutation for a protein. The prediction is to be made on the condition that only the folded structure of the wild type is known (i.e., the folded structure of a mutant is unknown). A total of 207 mutations for significantly many proteins are considered and the theoretical results are compared with the experimentally measured data. More specifically, we examined how the change in the thermal stability measure upon mutation is correlated with the experimental value of ΔT_{m}. The structure of a mutant is constructed using the Modeller program (Ver. 9.11). We have developed a new measure of the thermal stability of a protein in which the enthalpic component is also taken into account. The combination of the angle-dependent integral theory applied to a realistic multipolar model for water and the morphometric approach is employed in calculating the entropic component. In the enthalpic component, we incorporated the change in energy which stems from the break of protein-water hydrogen bonds and the formation of protein intramolecular hydrogen bonds upon protein folding. The performance of the resultant prediction method is as high as that of FOLD-X for single mutations. FOLD-X is one of the most popular, successful approaches using rather unphysical parameters fitted to the experimental data. By contrast, our method uses no such parameters. Nevertheless, we find that our method is far superior to FOLD-X for multiple mutations. Five multiple mutations for staphylococcal nuclease lead to highly enhanced stabilities, but we show that this high enhancement arises from the entropic EV effect. The neglect of this effect in FOLD-X is a principal reason for its ill success. Another finding is that the performance of our method becomes much higher when the prediction is limited to the mutants whose structural data for folded states are experimentally available and they are utilized (a total of 22 mutants). Therefore, the manner of constructing the mutant structure mentioned above needs to be reconsidered in future studies. The incorporation of the effects of protein-water and protein intramolecular interactions other than hydrogen bonds should also be explored.

Kinoshita and coworkers showed that in the mechanisms of folding and thermal, cold, and pressure denaturating of a protein, the entropic EV effect governed by water crowding (i.e., the hydrophobic effect in the true sense) is the key factor. In this study, it is shown that the thermal-stability change upon the cosolvent or salt addition is no exception, and the entropic EV

effect plays essential roles. At the same time, however, we find the cases where a theory accounting for this effect alone is not successful and identify the additional physical factors to be taken into consideration. For the thermal-stability change upon mutation, the enthalpic component is shown to be as important as the entropic EV effect. The results from this study should be very useful in determining a strategy of enhancing the thermal stability of a protein by modifying the solvent environment and by single or multiple mutations. Such determination is crucially important in energy science and nanotechnology. For example, thermostabilization of β-glucosidase, which catalyzes the hydrolysis of β-glycosidic bond in a sugar molecule, is strongly desired in biomass energy generation. It is considered that thermostabilization of a motor protein such as kinesin leads to the development of a new functional molecule.

List of Publications

Chapter 2

"Effects of monohydric alcohols and polyols on the thermal stability of a protein",
Shota Murakami and Masahiro Kinoshita,
The Journal of Chemical Physics, **144**, 125105(1-10) (2016).

Chapter 3

"Effects of salt or cosolvent addition on solubility of a hydrophobic solute in water: Relevance to those on thermal stability of a protein",
Shota Murakami, Tomohiko Hayashi, and Masahiro Kinoshita,
The Journal of Chemical Physics, **146**, 055102(1-15) (2017).

Chapter 4

"On the physics of thermal-stability changes upon mutations of a protein",
Shota Murakami, Hiraku Oshima, Tomohiko Hayashi, and Masahiro Kinoshita,
The Journal of Chemical Physics, **143**, 125102(1-13) (2015).

Copyright and publisher's Link

Chapter 2

Chapter 3

Chapter 4

List of Presentations

International Conference

Poster Presentation

“Theoretical Prediction of Thermal-Stability Changes upon Mutations of a Protein”,
○Shota Murakami, Hiraku Oshima, Tomohiko Hayashi, Masahiro Kinoshita,
Biophysical Society 59th Annual Meeting, Baltimore Convention Center, 2015.2.11

“Effects of cosolvent addition on the thermal stability of a protein”,
○Shota Murakami and Masahiro Kinoshita,
PACIFICHEM 2015, Hawaii Convention Center, 2015.12.17

"Effects of salt or cosolvent addition on thermal stability of a protein: Relevance to those on solubility of a hydrophobic solute in water”,
Shota Murakami, ○Tomohiko Hayashi, and Masahiro Kinoshita,
Biophysical Society 61th Annual Meeting, New Orleans Ernest N. Morial Convention Center, 2017.2.12

Domestic Conference

Poster Presentation

“Theoretical Prediction of Thermal-Stability Changes upon Mutations of a Protein”,
○Shota Murakami, Hiraku Oshima, Tomohiko Hayashi, and Masahiro Kinoshita,
第 51 回日本生物物理学会年会, 国立京都国際会館, 2013 年 10 月 30 日.

“Theoretical Prediction of Thermal-Stability Changes upon Mutations of a Protein”,
○Shota Murakami, Hiraku Oshima, Tomohiko Hayashi, and Masahiro Kinoshita,
第 52 回日本生物物理学会年会, 札幌コンベンションセンター, 2014 年 9 月 27 日.

“Effects of cosolvent addition on the thermal stability of a protein”,
○Shota Murakami and Masahiro Kinoshita,
第 53 回日本生物物理学会年会, 金沢大学角間キャンパス自然科学本館, 2015 年 9 月 14 日.

Acknowledgement

This thesis has been accomplished with support of many people.

I wish to express my sincere thanks and appreciation to Prof. Masahiro Kinoshita at the Institute of Advanced Energy of Kyoto University for his ardent guidance, appropriate suggestions, fruitful discussions, and encouragement throughout my academic work. During my enrollment as a member of his laboratory, he has given me much scientific knowledge and significantly many advices for my future.

I greatly thank Prof. Takashi Morii and Prof. Masato Katahira at the Institute of Advanced Energy of Kyoto University for their critical reading of the thesis and many useful comments. Their experimental knowledge is of benefit to the improvement of the thesis.

Sincere appreciation should be expressed to Dr. Hiraku Oshima (Present Address: RIKEN) at Kinoshita's laboratory for his useful suggestions and warmful encouragement. He has made a great contribution to the study described in Chapter 4.

I am deeply grateful to Dr. Tomohiko Hayashi at Kinoshita's laboratory for his encouragement and constructive discussions. He has made a great contribution to the studies described in Chaptes 3 and 4.

The support, encouragement, and helpful advices by Mr. Yuta Kajiwara at Kinoshita's laboratory are gratefully acknowledged.

I have received education from Dr. Hiraku Oshima, Dr. Tomohiko Hayashi, and Dr. Satoshi Yasuda (Present Address: Chiba University) at Kinoshita's laboratory. They have enhanced my computational skill, physics knowledge, and information-gathering capacity.

This research was supported by Grant-in-Aid for Scientific Research on Innovative Areas (B) (No. 25291035) from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by Grant-in-Aid for JSPS (Japan Society for the Promotion of Science) fellows.

Finally, I thank my family for supporting my mind.

February, 2017
Shota Murakami