Prediction of Thermostabilizing Mutations for a Membrane Protein on the Basis of Statistical Thermodynamics

Yuta Kajiwara

Prediction of Thermostabilizing Mutations for a Membrane Protein on the Basis of Statistical Thermodynamics

Yuta Kajiwara

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

February 2018

Table of contents

Chapter 1

General Introduction

1.1. Scope and significance of this study

Membrane proteins play essential roles in the communication between cells and their environment.[1,2] They are indispensable to the expression of life phenomena and implicated in a number of diseases. It is no wonder that they construct more than 60% of current drug targets.[3] For the efficient drug design, it is required to obtain the structural information of a target membrane protein. Nevertheless, the studies on membrane proteins are far behind those on water-soluble proteins. This is due to their low structural stability exhibited especially when they are removed from the lipid bilayer and solubilized in detergents. It then becomes quite difficult to solve the three-dimensional (3D) structure by X-ray crystallography, which is a usual mean of the structure determination. Therefore, enhancing the stability of membrane proteins is an urgent task.

It is known that an amino-acid mutation can lead to the enhanced thermal stability of membrane proteins.[4–6] The enhancement usually leads to higher stability in detergents as well. However, thermostabilizing mutations are currently identified by experiments testing a number of trial mutants. The prevailing approach is alanine (Ala) scanning mutagenesis in which every residue is mutated only to Ala (Ala is mutated to leucine).[7–9] However, even when this approach exploring only the limited mutational space is employed, the experimental burden is extremely heavy. The development of a theoretical method is strongly desired for the identification of thermostabilizing mutations.

Three methods[10–12] which do not rely on experiments were reported for the identification of thermostabilizing mutations for membrane proteins. Chen et al.[10] targeted poorly packed nonpolar residues and nonconserved charged and polar residues which lack the stabilization by hydrogen bonding or electrostatic interaction, and mutated these residues. Sauer et al.[11] proposed an approach whose advantage is that the wild-type structure of the target protein is not required. However, the mutations other than those included in the genomic sequences data cannot be considered. These two methods are based on bioinformatics or rather empirical (i.e., they are knowledge-based methods), and the physicochemical origins of the stability changes brought by mutations are somewhat ambiguous. Bhattacharya et al.[12] developed a method using an energetic score in which only the van der Waals (vdW) and torsion energies are incorporated as two important factors. However, protein intramolecular hydrogen bonds (IHBs), which should be crucially important, are not taken into account in the score, and its application is limited to a mutation from a nonpolar residue to Ala. The problem is that the score is defined not for the folding process but for the folded state alone.

The ultimate goal of this study is to develop a theoretical method for identifying the thermostabilizing mutations for membrane proteins. There are no visual differences between the wild type and its thermostabilized mutant. The enhanced thermal stability cannot be realized by investigating microscopic characteristics revealed by an all-atom molecular dynamics simulation. Therefore, a theoretical method for identifying thermostabilizing mutations must be based on thermodynamics. For testing all of the possible mutations (not limited to mutations to Ala) with minor computational effort, we should take account of only the fewest possible number of physicochemical factors that substantially influence the thermostability of membrane proteins.

Kinoshita and coworkers demonstrated for water-soluble proteins that the entropic effect originating from the translational displacement of water molecules is critical in the elucidation of mechanisms of not only folding but also cold, pressure, and thermal denaturating.[13–20] In general, the quantification of entropy by statistical mechanics is more difficult than that of energy. This is particularly true when a large solute with complex polyatomic structure like a protein is treated. Kinoshita and coworkers overcame this difficulty using a hybrid of an integral equation theory (IET)[21] and their morphometric approach (MA)[22,23]. The former is a statistical-mechanical theory for fluids and the latter is necessitated to treat a large protein with complex polyatomic structure (see Sec 1.4 for more details). This hybrid method allows us to calculate the solvation entropy upon protein insertion with sufficient accuracy and high speed. Kinoshita and coworkers developed a free-energy function (FEF) comprising entropic and energetic components denoted by S and Λ, respectively. S and Λ are dependent on the protein structure.

Kinoshita and coworkers presumed that for a membrane protein the hydrocarbon groups (CH_2, CH_3, and CH) constituting nonpolar chains of lipid molecules act as a "solvent" just like water for a water-soluble protein. The entropic effect arising from the translational displacement of hydrocarbon groups is incorporated in S. Λ is related to the energetic effect due to protein IHBs. Let ΔF be the change in F upon protein folding. Lower ΔF implies higher stability. Taking GpA as an example, Kinoshita and coworkers succeeded in showing that the native structure can be discriminated from ~15,000 non-native structures generated by a computer simulation[13]: ΔF takes the lowest value for the native structure. The most important result was that the discrimination ended with failure when the entropic component was neglected. It was thus suggested that the entropic effect described above is crucial in discussing the structural stability of a membrane protein.

In this study, on the basis of the achievement made by Kinoshita and coworkers, we develop a theoretical method for identifying the thermostabilizing mutations for membrane proteins and demonstrate its high capability for G-protein coupled receptors (GPCRs), which form the largest family of membrane proteins. GPCRs share a common seven-transmembrane topology and mediate cellular response to a variety of extracellular signals ranging from photons and small molecules to peptides. GPCR possesses three regions: the extracellular, transmembrane (TM) and intracellular regions (see Figure 1-1). It can be presumed that the structural stability of the TM regions is

significantly higher than that of the other two regions. In other words, the free-energy change for a GPCR upon the folding is governed by that for its TM region.

The biochemical and biophysical studies suggested that for a GPCR there exists multiple conformational states.[24] The conformation of a GPCR with no ligand binding could be discussed in terms of the two-state model: There are inactive and active states which are in equilibrium with each other.[25,26] For many GPCRs, the free energy of the inactive state is lower than that of the active state; the time during which it is in the inactive state is significantly longer than that during which it is in the active state. An agonist can bind to a GPCR molecule while the GPCR is in the active state. Upon the binding, the active state is stabilized in the sense that its free energy is lowered. Likewise, the binding of an inverse agonist to a GPCR molecule leads to the stabilization of the inactive state. If the active state is stabilized by a mutation without the agonist binding, the design of an agonist with high binding affinity will significantly be facilitated. The stabilization of the inactive state by a mutation will also be highly advantageous to efficient design of an inverse agonist with high binding affinity.

In this study, we consider a GPCR in the inactive state and argue its stability by looking at its TM region. We initiate the study by discussing the essential physical factors governing the structural stability of a GPCR (see Section 1.2 for more details). GPCR folding is accompanied by a gain of solvent entropy, loss of protein conformational entropy, and lowering of energy. The solvent is formed by hydrocarbon groups constituting nonpolar chains within the lipid bilayer, and its entropy is presumed to be governed by the element originating from the translational displacement of these hydrocarbon groups. The solvent-entropy gain arising from the close packing of side chains in the side-to-side association of the seven helices is considered as the dominant contributor to the net gain. Assuming that the wild type and its mutant share the same loss of conformational entropy, we omit this loss. Protein intramolecular van der Waals (vdW) energy decreases upon the folding, but this decrease is assumed to be cancelled out by the increase in protein-solvent vdW energy. Since such cancellation does not occur for electrostatic energy, we account for the decrease in protein intramolecular electrostatic energy which is attributable primarily to the formation of intramolecular hydrogen bonds (IHBs). We then construct a free-energy function F and the change in F upon GPCR folding ΔF. A lower value of ΔF implies higher structural stability. A mutation which lowers ΔF is identified as a thermostabilizing one.

As an important first step, we deal with the case where the 3D structure of the wild type is known and can be utilized. For the adenosine A_{2a} receptor ($A_{2a}R$), an important GPCR, we examine all of the possible mutations, select some of the mutations which are predicted to be highly thermostabilizing, and perform experiments to check if they are actually thermostabilizing. The structure of a mutant is modeled using the MODELLER program.[27] The success rates are in the range from 5/7 to 7/7 depending on the criterion employed for the thermostability relative to that of the wild type in the theoretical prediction and on the thermostability measure adopted in our experiments. We conclude that the overall success rate is 6/7. From a practical point of view,

however, a theoretical prediction must be made under the condition where the 3D structure of the wild type is not available. Therefore, as the second step, we consider $A_{2a}R$ but postulate that its 3D structure is unknown. A number of candidate models are constructed for the 3D structure using the homology modeling. In the homology modeling, the amino-acid sequence of the target protein is aligned to that of another protein as a template protein, and the 3D-structure models of the target are constructed by referring to the structural information of the template. A good measure of the appropriateness of a model is the root-mean-square deviation (RMSD) between this model and the crystal structure. However, RMSD cannot be calculated in a practical situation. We therefore select the model whose ΔF is the lowest and compare the prediction results obtained using the crystal structure. The overall success rate decreases but only slightly: It is 5/7.

The most important success is that we discover "key residues" and "hot-spot residues" for GPCRs of the largest family of Class A. A key residue is the residue to be mutated in the sense that many of its mutations are highly thermostabilizing. A GPCR possesses multiple key residues. A hot-spot residue is the key residue designated by N_{BW}, which is common in different GPCRs.[28] N_{BW} is the Ballesteros-Weinstein (BW) number that is the enumeration of a GPCR residue relative to that of the most conserved residue designated as x.50 in the helix. The same BW number signifies the same position in the amino-acid sequence. We find that the residue at a position of N_{BW}=3.39 is a hot-spot residue for significantly many GPCRs of Class A in the inactive state. In particular, mutating this hot-spot residue to arginine (Arg) or lysine (Lys) is shown to remarkable enhance the thermal stability. As a product, highly thermostabilizing mutations are found out for muscarinic acetylcholine receptor 2 (M2R) and prostaglandin E receptor 4 (EP4), leading to the determination of new 3D structures. We expect that the 3D structures of many GPCRs of Class A in the inactive state can be solved for the first time in succession.

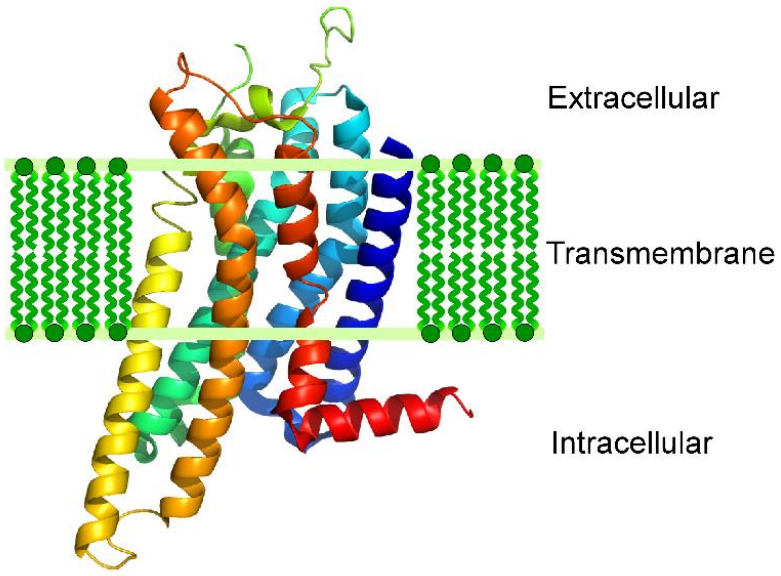


Figure 1-1. Illustration of the extracellular, transmembrane (TM), and intracellular regions of a GPCR. The adenosine A_{2a} receptor ($A_{2a}R$) is shown as an example.

1.2. Essential physical factors for driving force of a membrane protein formation

Just like water for a water-soluble protein, the hydrocarbon groups constituting nonpolar chains of lipid molecules act as "solvent" for a membrane protein. The factors affecting the thermostability of the membrane protein are the protein intramolecular vdW and electrostatic energies, solvent-solvent and protein-solvent vdW and electrostatic interaction energies, protein intramolecular entropy, and solvent entropy. A donor or an acceptor for a hydrogen bond (HB) carries a negative partial charge whereas a hydrogen atom does a positive one, and the hydrogen bonding is a primary component of the intramolecular electrostatic energy. For testing all of the possible mutations with minor computational effort, however, it is not advantageous to account for all of these factors. It is required that only the fewest number of factors be taken into consideration.
When a protein folds, the excluded volume (EV) (i.e., the volume of the space which the centers of solvent molecules cannot enter) decreases to a large extent, leading to a corresponding increase in the total volume available to the translational displacement of solvent molecules coexisting with the protein in the system.[13-20] This increase is accompanied by a large gain of solvent entropy. Close packing of side chains is the most important to the solvent-entropy gain (see Figure 1-2).[13,29] The solvent possesses the orientational (rotational) and vibrational entropies as well as the translational

entropy (TE), but Kinoshita and coworkers showed that the translational contribution predominates over the orientational entropy.[15,30] For instance, the TE and orientational (rotational) contributions to the water-entropy gain upon apoplastocyanin folding are ~95% and ~5%, respectively.[15] The dominance of the TE contribution should also be applicable to the case of nonpolar chains of lipid molecules.

In water, the formation of intramolecular hydrogen bonds (IHBs) is accompanied by the break of HBs with water molecules. This is not the case in nonpolar chains of lipid molecules, which makes the formation of IHBs even more influential. A gain of intramolecular vdW attractive interactions upon protein folding is somewhat cancelled out by the loss of protein-solvent vdW attractive interactions unavoidably accompanied. When the wild-type and mutant structures are compared in terms of the stability, the effect of protein intramolecular entropy may be relatively smaller since the structures are both quite compact.

1.3. Solvent model

Recent experimental data have shown that many membrane proteins fold and oligomerize quite efficiently in nonpolar environments which bear little similarity to a membrane (e.g., those provided by surfactant molecules and amphipols).[31] This fact can be interpreted as follows. First, the thermal motion of surrounding solvent molecules (i.e., the solvent-entropy effect) is a principal driving force in protein folding but the details of specific characteristics of the solvent molecules are not important. Further, the intramolecular hydrogen bonding is a sufficiently powerful contributor to protein folding due to the nonpolar nature of solvent (in water, the folding is unavoidably accompanied by the break of protein-water hydrogen bonds).

We note that a membrane is immersed in water. When a membrane protein takes a structure with larger EV, the membrane also generates larger EV for water molecules. Thus, water indirectly acts as the solvent. For this reason, we should rather take the view that the membrane protein is immersed in bulk solvent. According to the results of Kinoshita and coworkers' works,[16,18-20] the qualitative aspects of the solvent-entropy effect at ambient temperature and pressure can be elucidated by neglecting the solvent-solvent attractive interaction. For the membrane, we take account of only the effect of the translational displacement of hydrocarbon (CH_2, CH_3, and CH) groups which are treated as if they were not connected with one another, so that a tractable statistical-mechanical theory can be applied.

Taken together, we employ a simplified model for the solvent: an ensemble of neutral hard spheres whose diameter and packing fraction are set at those of water at 298 K and 1 atm.

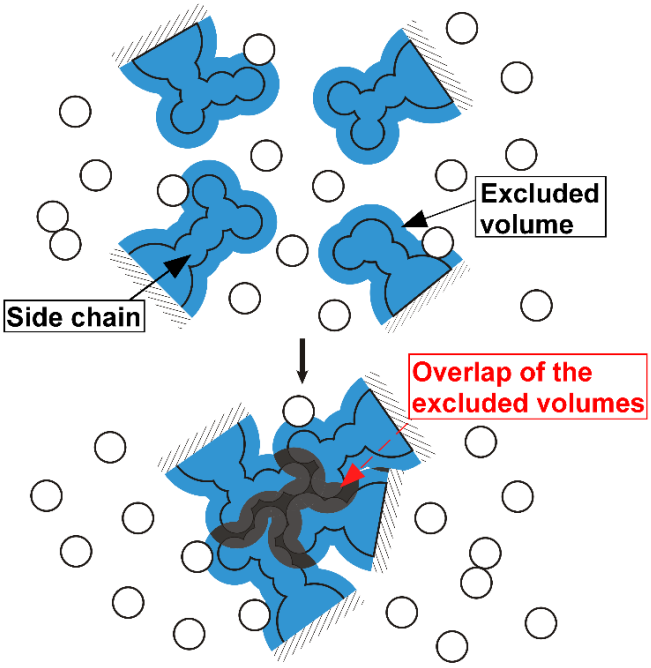


Figure 1-2. Close packing of side chains of a protein in solvent. Overlap of excluded volumes occurs, and the total volume available to the translational displacement of solvent particles (white circles) increases by the overlapped volume.

1.4. The hybrid method for calculating solvation entropy of a protein

To calculate the solvation entropy, the solvent must be treated not as a continuum but as an ensemble of particles with finite sized. The calculation entropy is usually more difficult than that of energy. In particular, the calculation of the solvation entropy of a protein is a formidable task.

Kinoshita and coworkers overcame this difficulty using a hybrid of an integral equation theory and their morphometric approach.[22,23] The former is a statistical-mechanical theory for fluids[21] and the latter is necessitate to treat a large protein with complex polyatomic structure. Using this hybrid, we can finish the calculation of *S* in ~0.54 sec per structure on our workstation with Xeon E5-2695 2.3-GHz processor. The detailed descriptions for the hybrid method are provided the following section.

1.4.1. Integral equation theory (IET)

The integral equation theory (IET)[21] is based on classical statistical mechanics. In this theory, from the system partition function, various correlation functions are defined, and the basic equations satisfied by these functions are derived. The many-body correlations are 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. Once the basic equations are numerically solved and the correlation functions are obtained, we can calculate the microscopic structure and thermodynamic quantities. The solvent structure near a solute molecule and thermodynamic quantities of solvation, that is, changes in thermodynamic quantities upon solute insertion into the solvent, can also be calculated. A thermodynamic quantity of solvation is calculated for a fixed solute structure inserted into a fixed position within the solvent. When the solvent is a simple fluid with radial-symmetric potential, a complex solute molecule with a polyatomic structure (e.g., a protein) can be handled by the three-dimensional integral equation theory (3D-IET).[32-34] However, the use of the 3D-IET is quite time consuming, which requires a special approach developed by Kinoshita and coworkers (i.e., the morphometric approach).

1.4.2. Morphometric approach (MA)

The idea of the morphometric approach (MA)[22,23] is to express the solvation entropy of a solute molecule with a prescribed structure S by the linear combination of only four geometric measures of the solute structure:

$$S/k_B=C_1V_{ex}+C_2A+C_3X+C_4Y. \tag{1-1}$$

Here, eq. 1-1 is referred to as the morphometric form, V_{ex} is the excluded volume (EV) generated by the solute molecule, A is the solvent-accessible surface area, X and Y are the integrated mean and Gaussian curvatures of the solvent-accessible surface, respectively, and k_B is the Boltzmann constant. The solvent-accessible surface is the surface that is accessible to the centers of solvent molecules. The EV is the volume that is enclosed by this surface. Since S is fairly insensitive to the solute-solvent interaction potential, the solute molecule can be modeled as a set of fused hard spheres. In the MA, the solute structure enters S only via the four geometric measures. Therefore, the four coefficients (C_1–C_4) can be determined in simple geometries: They are calculated from the values of S of isolated hard-sphere solutes immersed in the solvent. The IET is employed in the calculation. The four coefficients are determined by the least square fitting applied to the following equation (i.e., eq. 1-2 applied to the hard-sphere solutes);

$$\begin{aligned} &S/k_B=C_1(4\pi R^3/3)+C_2(4\pi R^2)+C_3(4\pi R)+C_4(4\pi), \\ &R=(d_U + d_S)/2. \end{aligned} \tag{1-2}$$

Here, d_U denotes the diameter of a hard-sphere solute, and sufficiently many different values of d_U are considered ($0.6d_S \le d_U \le 10d_S$). The absolute temperature is set at 298 K, and the number density of bulk solvent ρ_S is taken to be that of water on the saturation curve, $\rho_S d_S^3=0.7317$ (d_S=2.8 Å).

Once the four coefficients are determined, S of the solute molecule with a prescribed structure is obtained from eq. 1-2 only if the four geometric measures are calculated.

The IET-MA hybrid method described above has been shown to be quantitatively reliable in Kinoshita and coworkers' earlier publications.[20,23]

1.5. Synopsis of this thesis

In Chapter 2, we develop a rapid method which allows us to treat all of the possible mutations. It employs a free-energy function (FEF) which takes account of the translational entropy of hydrocarbon groups within the lipid bilayer as well as the protein intramolecular hydrogen bonding. The method is illustrated for $A_{2a}R$ whose wild-type structure is known and utilized. We propose a reliable strategy of finding key residues to be mutated and selecting their mutations which will lead to considerably higher stability. Representative single mutants predicted to be stabilizing or destabilizing are experimentally examined: The overall success rate is remarkably high. The melting temperature T_m for two of them is substantially higher than that of the wild type. A double mutant with even higher T_m is also obtained. Our FEF captures the essential physics of the stability changes upon mutations.

It was experimentally showed at another group that the thermal stability of a membrane protein, the adenosine A_{2a} receptor, was remarkably enhanced by an octuple mutation. In Chapter 3, we theoretically prove that the energy decrease arising from the formation of protein intramolecular hydrogen bonds and the solvent-entropy gain upon protein folding are made substantially larger by the mutation, leading to the remarkable enhancement. The mutation modifies geometric characteristics of the structure so that the solvent crowding can be reduced to a larger extent when the protein folds. Moreover, it is shown that $|\Delta\Delta F|$ (F is the FEF, ΔF denotes the change in F upon protein folding, and $\Delta\Delta F$= "ΔF for a mutant" – "ΔF for the wild type") and the melting temperature change ΔT_m for the octuple mutations are much larger than those for the four single mutations and significantly larger than those for the double mutation.

In Chapter 4, we postulate that the 3D structure of $A_{2a}R$ is unknown. We construct candidate models for the 3D structure using the homology modeling and select the model giving the lowest value to ΔF. A good measure of the appropriateness of a model is the root-mean-square deviation (RMSD) between this model and the crystal structure. However, RMSD cannot be calculated in a practical situation. We therefore select the model whose ΔF is the lowest (this model is referred to as "the best model") and follow the basic procedure explained above. We compare the prediction results obtained using the crystal structure, the best model, and six more representative models. The performance achieved in the best model is only slightly lower than that in the crystal structure.

In Chapter 5, we report a theoretical strategy by which many different GPCRs can be considered at the same time. The strategy is illustrated for three GPCRs: $A_{2a}R$, M2R, and EP4 of Class A in the

inactive state. We argue that a mutation of the residue at a position of N_{BW}=3.39 (N_{BW} is the Ballesteros-Weinstein number), a hot-spot residue, leads to substantially higher stability for significantly many GPCRs of Class A in the inactive state. The most stabilizing mutations of the residues with N_{BW}=3.39 are then identified for two of three GPCRs: M2R and EP4. These identifications are experimentally corroborated, which is followed by the determination of new structures for M2R and EP4. We expect that on the basis of the strategy, the 3D structures of many GPCRs of Class A can be solved for the first time in succession.

References

[1]Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell. Biol*. **2002**, *3*, 639−650.

[2]Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabaśi, A. L.; Vidal, M. Drug-target network. *Nat. Biotechnol.* **2007**, *25*, 1119−1126.

[3]Fredriksson, R.; Lagerström, M. C.; Lundin, L. G.; Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **2003**, *63*, 1256−1272.

[4]Lau, F. W.; Nauli, S.; Zhou, Y.; Bowie, J. U. Changing single side-chains can greatly enhance the resistance of a membrane protein to irreversible inactivation. *J. Mol. Biol.* **1999**, *290*, 559−564.

[5]Zhou, Y.; Bowie, J. U. Building a thermostable membrane protein. *J. Biol. Chem.* **2000**, *275*, 6975−6979.

[6]Scott, D. J.; Kummer, L.; Tremmel, D.; Plückthun, A. Stabilizing membrane proteins through protein engineering. *Curr. Opin. Chem. Biol.* **2013**, *17*, 427−435.

[7]Serrano-Vega, M. J.; Magnani, F.; Shibata, Y.; Tate, C. G. Conformational thermostabilization of the β1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 877−882.

[8]Magnani, F.; Shibata, Y.; Serrano-Vega, M. J.; Tate, C. G. Co-evolving stability and conformational homogeneity of the human adenosine A_{2a} receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10744−10749.

[9]Shibata, Y.; White, J. F.; Serrano-Vega, M. J.; Magnani, F.; Aloia, A. L.; Grisshammer, R.; Tate, C. G. Thermostabilization of the neurotensin receptor NTS1. *J. Mol. Biol.* **2009**, *390*, 262−277.

[10]Chen, K. M.; Zhou, F.; Fryszczyn, B. G.; Barth, P. Naturally evolved G protein-coupled receptors adopt metastable conformations. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13284−13289.

[11]Sauer, D. B.; Karpowich, N. K.; Song, J. M.; Wang, D. N. Rapid bioinformatic identification of thermostabilizing mutations. *Biophys. J.* **2015**, *109*, 1420−1428.

[12]Bhattacharya, S.; Lee, S.; Grisshammer, R.; Tate, C. G.; Vaidehi, N. Rapid Computational Prediction of Thermostabilizing Mutations for G Protein-Coupled Receptors. *J. Chem. Theory Comput.* **2014**, *10*, 5149−5160.

[13]Yasuda, S.; Oshima, H.; Kinoshita, M. Structural stability of proteins in aqueous and nonpolar environments. *J. Chem. Phys.* **2012**, *137*, 135103.

[14]Harano, Y.; Kinoshita, M. Translational-entropy gain of solvent upon protein folding. *Biophys. J.* **2005**, *89*, 2701−2710.

[15]Yoshidome, T.; Kinoshita, M.; Hirota, S.; Baden, N.; Terazima, M. Thermodynamics of apoplastocyanin folding: comparison between experimental and theoretical results. *J. Chem. Phys.* **2008**, *128*, 225104.

[16]Kinoshita, M. Roles of translational motion of water molecules in sustaining life. *Front. Biosci.* **2009**, *14*, 3419−3454.

[17]Kinoshita, M. Importance of translational entropy of water in biological self-assembly processes like protein folding. *Int. J. Mol. Sci.* **2009**, *10*, 1064−1080.

[18]Kinoshita, M. A new theoretical approach to biological self-assembly. *Biophys. Rev.* **2013**, *5*, 283−293.

[19]Oshima, H.; Kinoshita M. Effects of sugars on the thermal stability of a protein. *J. Chem. Phys.* **2013**, *138*, 245101.

[20]Oshima, H.; Kinoshita M. Essential roles of protein-solvent many-body correlation in solvent-entropy effect on protein folding and denaturation: comparison between hard-sphere solvent and water. *J. Chem. Phys.* **2015**, *142*, 145103.

[21]Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids, 3rd ed.* Academic Press; London, U.K., 2006.

[22]König, P. M.; Roth, R.; Mecke, K. R. Morphological thermodynamics of fluids: shape dependence of free energies. *Phys. Rev. Lett.* **2004**, *93*, 160601.

[23]Roth, R.; Harano, Y.; Kinoshita, M. Morphometric approach to the solvation free energy of complex molecules. *Phys. Rev. Lett.* **2006**, 97, 078101.

[24]Kobilka, B. K.; Deupi, X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci*. **2007**, *28*, 397−406.

[25]Vauquelin, G.; Van Liefde, I. G protein-coupled receptors: A count of 1001 conformations. *Fundam. Clin. Pharmacol*. **2005**, *19*, 45−56.

[26]Yao, X. J.; Ruiz, G. V.; Whorton, M. R.; Rasmussen, S. G. F.; DeVree, B. T.; Deupi, X.; Sunahara, R. K.; Kobilka, B. The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9501−9506.

[27]Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.

[28]Ballesteros, J.A.; Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G protein-coupled receptors.

Methods Neurosci. **1995**, *25*, 366−428.

[29]Yasuda, S.; Yoshidome, T.; Oshima, H.; Kodama, R.; Harano, Y.; Kinoshita, M. Effects of side-chain packing on the formation of secondary structures in protein folding. *J. Chem. Phys.* **2010**, *32*, 065105.

[30]Kinoshita, M. Molecular origin of the hydrophobic effect: Analysis using the angle-dependent integral equation theory. *J. Chem. Phys.* **2008**, *128*, 024507.

[31]Popot, J. L.; Engelman, D. M. Membranes do not tell proteins how to fold. *Biochemistry* **2016**, *55*, 5−18.

[32]Beglov, D.; Roux, B. Numerical solution of the hypernetted chain equation for a solute of arbitrary geometry in three dimensions. *J. Chem. Phys.* **1995**, *103*, 360.

[33]Ikeguchi, M.; Doi, J. Direct numerical solution of the Ornstein–Zernike integral equation and spatial distribution of water around hydrophobic molecules. *J. Chem. Phys.* **1995**, *103*, 5011.

[34]Kinoshita, M. Spatial distribution of a depletion potential between a big solute of arbitrary geometry and a big sphere immersed in small spheres. *J. Chem. Phys*. **2002**, *116*, 3493.

Chapter 2

Identification of Thermostabilizing Mutations for Membrane Proteins: Rapid Method Based on Statistical Thermodynamics

2.1. Introduction

Membrane proteins such as G-protein coupled receptors (GPCRs), transporters, and channels, which represent ~30% of the currently sequenced genomes, play imperative roles in sustaining life.[1,2] Their malfunctioning causes a diversity of serious diseases. Many of them are located at cell surfaces and they can readily be subjected to small molecular drugs circulating through the blood. It is no wonder that they construct more than 60% of current drug targets.[3] Despite the crucial importance of membrane proteins, the studies on them are far behind those on water-soluble proteins. This is due to their low structural stability exhibited especially when they are removed from the lipid bilayer and solubilized in detergents. It then becomes quite difficult to obtain the three-dimensional structure by X-ray crystallography, the usual means of structural determination. In fact, the structures of membrane proteins registered constitute only ~0.5% of all protein structures registered in Protein Data Bank. When the structure of a membrane protein is unknown, understanding its function and designing a new drug targeting it becomes extremely difficult. Even for a membrane protein whose structure has already been determined, further improvement of its stability is desirable in biochemical and pharmaceutical applications. Thus, enhancing the stability of membrane proteins is an urgent subject.

It is known that an amino-acid mutation can enhance the thermostability of a membrane protein.[4−6] The enhancement usually leads to higher stability in detergents as well. However, thermostabilizing mutations are currently identified by experiments testing a number of trial mutants. The prevailing approach is alanine (Ala) scanning mutagenesis in which every residue is mutated only to Ala (Ala is mutated to leucine).[7−9] Even when this approach limiting the mutational space is employed, the experimental burden is quite heavy. Though the development of a theoretical method is thus strongly desired for the identification of thermostabilizing mutations, it is a cumbersome task because physical chemistry of the structural stability of membrane proteins is only poorly comprehended unlike that of water-soluble proteins.

Three methods[10−12] which do not rely on experiments were reported for the identification of thermostabilizing mutations for membrane proteins. Chen et al.[10] targeted poorly packed nonpolar residues and nonconserved charged and polar residues which lack the stabilization by hydrogen bonding or electrostatic interaction, and mutated these residues. Sauer et al.[12] proposed an approach whose advantage is that the wild-type structure of the target protein is not required. However, the mutations other than those included in the genomic sequences data cannot be considered. These two methods are based on bioinformatics or rather empirical (i.e., they are knowledge-based methods),

and the physicochemical origins of the stability changes brought by mutations are somewhat ambiguous. Bhattacharya et al.[11] developed a method using an energetic score in which only the van der Waals (vdW) and torsion energies are incorporated as two important factors. However, its application is limited to a mutation from a nonpolar residue to Ala, primarily because protein intramolecular hydrogen bonds (HBs) are not taken into account in the score. Another problem is that the score is defined not for the folding process but for the folded state alone.

In the present article, we report the results of the first attempt to develop a physics-based method which allows us to rapidly explore the whole mutational space. Our free-energy function *F*, which has recently been developed by us for a protein in nonpolar environment,[13] is modified for this particular purpose. The hydrocarbon (CH_2, CH_3, and CH) groups constituting nonpolar chains of lipid molecules act as "solvent" for a membrane protein just like water for a water-soluble protein. Kinoshita and coworkers have demonstrated for water-soluble proteins that the entropic effect originating from the translational displacement of water molecules is crucial in the elucidation of mechanisms of not only folding but also cold, pressure, and thermal denaturating.[13−20] This demonstration has inspired us to account for the solvent-entropy effect for membrane proteins as well.[13] *F* takes account of the entropic effect originating from the translational displacement of hydrocarbon groups and the protein intramolecular hydrogen bonding as two essential factors: It comprises the entropic and energetic terms which are denoted by *S* and Λ, respectively. *S* and Λ are dependent on the protein structure. As the first step, it is postulated that the wild-type structure is known and utilizable in the identification. Once stabilizing mutations for a membrane protein are found, they will be applicable to the homologs which show high sequence similarity. We illustrate the method for the adenosine A_{2a} receptor ($A_{2a}R$). $A_{2a}R$ is a seven-transmembrane protein which is a member of the GPCR family. An important finding in Kinoshita and coworkers' earlier work concerning water-soluble proteins is that a mutation which largely increases the water-entropy gain upon folding often leads to very high enhancement of the thermal stability irrespective of the enthalpic factor.[21] By the analogy of this finding, we examined the following strategy: First, calculate S for all of the possible mutations and nominate key residues to be mutated in the sense that many of their mutations will lead to relatively higher enhancement of the stability; second, select stabilizing mutations of the key residues using *F*. In the second step, only the mutants which are predicted to be sufficiently more stable than the wild type should be selected. The examination of this strategy was performed by our experiments. The result for a mutation of one of the key residues to Ala, which was previously reported,[8] was also considered. The success rates were in the range from 7/9 to 9/9 depending on the criterion employed for the thermostability relative to that of the wild type in the theoretical prediction and on the thermostability measure adopted in our experiments. Despite the high success rates, the calculation of *F* can be finished in less than 1 sec per structure, which allows us to test all of the possible mutations with moderate computational effort. We then combined two of the single mutations to a double mutation. It was predicted to be highly stabilizing on the basis of *F*: It certainly lead to an increase in T_m of ~12°C (T_m is the melting

temperature). This result is remarkable because in many cases an increase in T_m exceeding 10°C can be achieved only by a triple or higher-fold mutation when Ala scanning mutagenesis is employed.

2.2. Theoretical method

For testing all of the possible mutations with minor computational effort, it is required that the calculation of a free-energy function be accomplished quite rapidly. We therefore account for only the fewest number of physicochemical factors that substantially influence the thermostability of membrane proteins. We discuss such factors by comparing membrane proteins with water-soluble proteins.

2.2.1. Entropic excluded-volume effect

Kinoshita and coworkers have shown for a water-soluble protein that the entropic effect originating from the translational displacement of solvent (i.e., water) molecules plays a pivotal role in folding and unfolding mechanisms.[13–20] For a membrane protein, the hydrocarbon (CH_2, CH_3, and CH) groups constituting nonpolar chains of lipid molecules should act as "solvent" just like water for a water-soluble protein. Upon protein folding, the excluded volume (EV) (i.e., the volume of the space which the centers of solvent molecules cannot enter) decreases to a large extent, which is followed by a corresponding increase in the total volume available to the translational displacement of solvent molecules coexisting with the protein in the system (a cartoon is given in Figure 2.1).[13–20] Namely, the folding leads to a large gain of solvent entropy. It is true that the solvent possesses not only the translational entropy (TE) but also the orientational (rotational) and vibrational entropies. In the case of water, the TE contribution dominates in the solute solvation (i.e., the solvent-entropy change upon solute insertion).[15,22] This is because upon solute insertion the reduction of orientational and vibrational freedoms occurs only for the solvent molecules in the close vicinity of the solute but that of translational freedom reaches all of the solvent molecules. This can also be applied to the case of nonpolar chains of lipid molecules.

Here, it is worthwhile to comment on hydrophobicity or hydrophobic effect that is relevant to water-soluble proteins. In the conventional view, the water adjacent to a nonpolar group is entropically unstable owing to the water structuring (i.e., increase in and enhancement of water-water hydrogen bonds): Due to this effect, a water-soluble protein is driven to fold through the burial of nonpolar groups.[23] Kinoshita and coworkers have shown that the entropic gain originating from this factor is much smaller than that from the EV effect explained above.[15,16,18] Actually, the EV effect is nothing but the true physical origin of hydrophobicity. The weakening of hydrophobic effect at low temperatures, which was experimentally observed for a variety of processes in aqueous solution such as cold denaturation of a protein, can be elucidated only by the EV effect.[18,20,24]

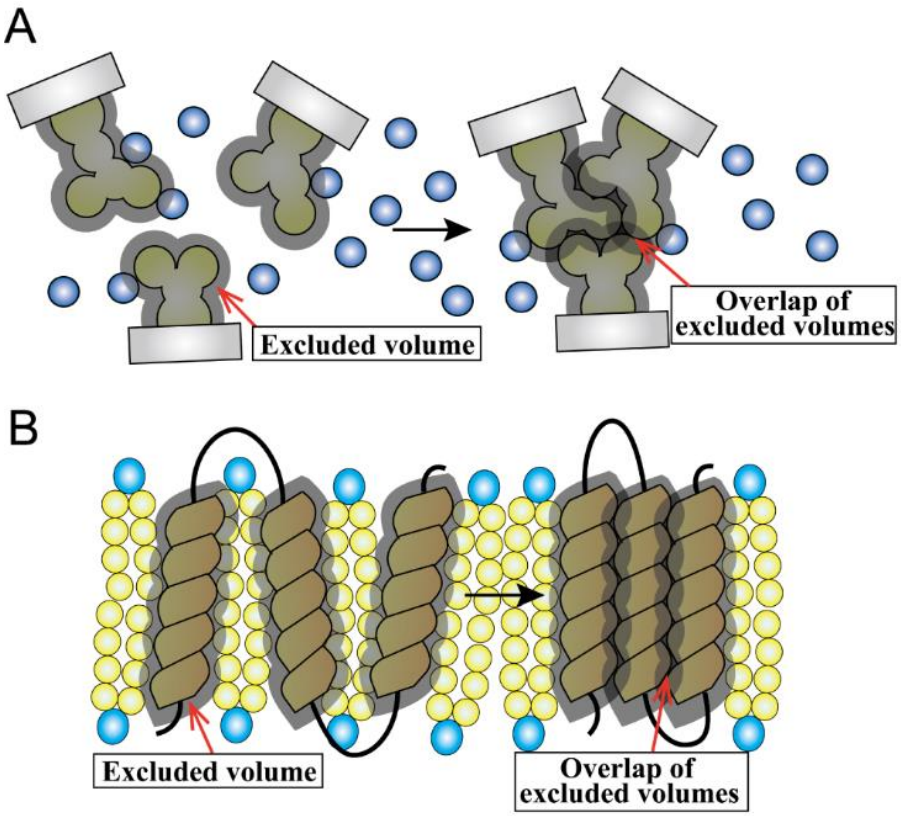


Figure 2.1. Close packing of side chains of a protein in solvent. Overlap of excluded volumes occurs, and the total volume available to the translational displacement of solvent particles increases by the overlapped volume. (A) The solvent is formed by water molecules (blue circles) for a water-soluble protein. (B) Side-to-side association of two helices of a membrane protein in solvent. The solvent is formed by hydrocarbon (CH_2, CH_3, and CH) groups (yellow circles) constituting nonpolar chains of lipid molecules.

2.2.2. Protein intramolecular hydrogen bonding

In aqueous environment, the formation of intramolecular hydrogen bonds (HBs) is accompanied by the break of HBs with water molecules. By contrast, this is not the case in nonpolar environment (i.e., nonpolar chains of lipid molecules). Therefore, the contribution to structural stability from the formation is stronger in nonpolar environment. Kinoshita and cowokers have recently shown the following:[13] Due to the difference mentioned above, α-helix and β-sheet are both favored and their relative contents are quite variable in aqueous environment, whereas there is a tendency that α-helix (and β-barrel) is exclusively chosen in nonpolar environment; and the content of secondary structures is higher in nonpolar environment than in aqueous one. It should be emphasized, however, that the formation is crucial even in aqueous environment: When a donor and an acceptor are buried with the break of HBs with water molecules, they need to form an intramolecular HB to compensate

for the break (i.e., dehydration). A donor or an acceptor carries a negative partial charge whereas a hydrogen atom a positive one, and the hydrogen bonding is thus a primary component of the intramolecular electrostatic interaction energy.

It was suggested that the intramolecular hydrogen bonding was not very important in aqueous and nonpolar environments because the energy decrease arising from it was rather small.[25] This suggestion was based on the experimental data of free-energy change upon a mutation which was postulated to bring the formation or removal of a HB or HBs. However, the mutation was unavoidably accompanied by changes in the other factors such as the entropic EV effect. The suggestion is not convincing at all.

2.2.3. On the other physicochemical factors

A gain of intramolecular vdW attractive interactions upon protein folding is somewhat cancelled out by the loss of protein-solvent vdW attractive interactions accompanied. It is empirically known for a water-soluble protein that close packing of the backbone and side-chains leads to its high structural stability. The stability is ascribed not to intramolecular vdW attractive interactions but to the solvent-entropy effect explained in Figure 2.1.[26] Since the structures of the wild type and its mutants to be considered are all quite compact, the effect of protein intramolecular entropy may be neglected. This neglect was justified in Kinoshita and coworkers' earlier works[27,28] on water-soluble proteins for discriminating the native fold from the compact but misfolded decoys (see the second paragraph in "Free-Energy Function (FEF)" for more details).

2.2.4. Free-energy function (FEF)

We incorporate only the entropic EV effect and the intramolecular hydrogen bonding in our FEF. It is expressed by

$$F/(k_{\mathrm{B}}T_0) = \Lambda/(k_{\mathrm{B}}T_0) - TS/(k_{\mathrm{B}}T_0),\ T_0 = 298\ \mathrm{K}. \qquad (2.1)$$

T is the absolute temperature and set at T_0 in the present study. Λ ($\Lambda<0$) is the energetic term representing protein intramolecular energy of the formation of intramolecular HBs. As the number of HBs increases, $|\Lambda|$ becomes larger. The entropic term $-S$ ($-S>0$) represents the magnitude of entropic loss of the hydrocarbon groups constituting nonpolar chains of the lipid bilayer (i.e., the solvent for a membrane protein) upon protein insertion. A smaller value of $-S$ implies higher efficiency of the backbone and side-chain packing.

For water-soluble proteins, Kinoshita and coworkers have developed a similar free-energy function taking account of only the effects of water entropy and HBs.[27,28] In this case, however, protein-water HBs as well as protein intramolecular HBs must be considered. The free-energy function was tested for a total of 133 proteins and shown to be capable of discriminating the native

fold from the compact but misfolded decoys with almost 100% accuracy.[27,28] It is far superior to any of the previously reported functions.

2.2.5. Solvent model

Recent experimental data have shown that many membrane proteins fold and oligomerize quite efficiently in nonpolar environments which bear little similarity to a membrane (e.g., those provided by surfactant molecules and amphipols).[29] This strongly suggests that folding of a membrane protein occurs only if the surrounding solvent molecules thermally move (i.e., the solvent-entropy effect comes into play) and the intramolecular hydrogen bonding is a sufficiently powerful contributor to the structure formation: The details of specific characteristics of nonpolar chains of lipid molecules are unimportant. Further, we note that a membrane is immersed in water. When a membrane protein takes a structure with a larger EV, the membrane also generates a larger EV for water molecules. Thus, water indirectly acts as the solvent. Consequently, we should take the view that a membrane protein is immersed in bulk solvent. We have shown that water can be modeled as "simple fluids" in any theoretical method focused on the entropic EV effect at ambient temperature and pressure.[16,18–20] These discussions have motivated us to employ a simplified model for the solvent: an ensemble of neutral hard spheres whose diameter and packing fraction are set at those of water at 298 K and 1 atm. This parameter setting for nonpolar chains of lipid molecules can be justified as follows. The solvent-entropy effect becomes larger as the solvent diameter decreases or the packing fraction increases.[16] The diameters of CH_2, CH_3, and CH groups are larger than the molecular diameter of water but their packing fraction is higher than the water value. These two properties are rather compensating, and the parameter setting becomes reasonable. With the simplified solvent model thus obtained, we can employ a statistical-mechanical approach. The model was shown to provide a good model of the membrane environment in Kinoshita and coworkers' earlier work.[13]

2.2.6. Calculation of the entropic term

The entropic term for a protein structure is calculated using a hybrid of an integral equation theory (IET)[30] and Kinoshita and coworkers' morphometric approach (MA).[31,32] The former is a statistical-mechanical theory for fluids and the latter is necessitated to treat a large protein with complex polyatomic structure. Using this hybrid, we can finish the calculation of $-S$ in less than 0.5 sec per structure despite that the solvent is not a continuum but an ensemble of particles with finite sizes. Detailed descriptions of the IET and the MA, respectively, are provided in Secs. 1.4 of Chapter 1.

2.2.7. Calculation of the energetic term

Not only $-S$ but also Λ is to be calculated quite rapidly. Hence, we employ the simplest possible method which still captures the physical essence. The essential quantity is the change in Λ upon

protein folding. As described above, we account for only intramolecular HBs. We examine all of the donors and acceptors for backbone-backbone, backbone-side chain, and side chain-side chain HBs and calculate Λ. The examination is made using the criteria proposed by McDonald and Thornton[33] with the modification that the maximum distances between H and the acceptor and between the donor and the acceptor are 3.0 Å and 4.4 Å, respectively. (H is the hydrogen atom covalently bound to the donor.) When an HB is formed, an energy decrease of D is considered. Unlike in our earlier work,13 D is made dependent on the distance between centers of H and the acceptor, d: This is why the modification of the criteria mentioned above is necessitated. For d≤1.5Å, D is set at $-14k_BT$. This value is based on the free-energy decrease arising from hydrogen-bond formation between two formamide molecules in a nonpolar liquid.[34] As d increases from 1.5Å, |D| decreases lineally and becomes zero for d≥3.0Å. We note that an H-acceptor pair is not in vacuum but in the environment where atoms with positive and negative partial charges are present. Therefore, |D| does not necessarily decrease in proportion to 1/d. The calculation of Λ can also be finished in less than 0.5 sec per structure.

In Kinoshita and coworkers' earlier work,[13] we succeeded for GpA in showing that the native structure can be discriminated from ~15,000 non-native structures generated by a computer simulation: The FEF takes the lowest value for the native structure. The discrimination is still successful even after the alteration of the calculation method for Λ mentioned above (the solvent model is the same as that explained in "Solvent Model").

2.2.8. Protocol for comparing the stabilities of the wild type and a mutant

The two-stage model[35] has been proposed for the structure formation process of a membrane protein. In this model, individual α-helices of the protein are separately stabilized as constituent domains within a lipid bilayer in the first stage, and the native structure is completed by the side-to-side association of these helices (i.e., association of the transmembrane domains accompanying the packing of side chains) in the second stage. The solvent-entropy gain upon protein folding occurring in the second stage is the most important quantity.[13] Hence, $-S$ is calculated by focusing on this side-chain packing (see Figure 2.2), which was validated in Kinoshita and coworkers' earlier work.[13] In the figure, the transmembrane domains are determined by the web server of TMDET.[36] First, we calculate the solvent-entropy gain ΔS ($\Delta S>0$) originating from the packing of separated α-helices for the wild-type structure as

$$-\Delta S_W=-S_{compact,W}-\Sigma(-S_{each})_W \qquad (2.2)$$

The subscript "W" denotes "wild type". $-S_{compact,W}$ is the magnitude of entropic loss upon insertion of the associated α-helices, and $\Sigma(-S_{each})_W$ is that upon insertion of the separated α-helices. $-S_{compact,W}$ is smaller than $\Sigma(-S_{each})_W$. We then calculate the same quantity for a mutant structure as

$$-\Delta S_{\mathrm{M}}=-S_{\mathrm{compact,M}}-\Sigma(-S_{\mathrm{each}})_{\mathrm{M}} \tag{2.3}$$

The subscript "M" denotes "mutant". $-S_{\mathrm{compact,M}}$ is smaller than $\Sigma(-S_{\mathrm{each}})_{\mathrm{M}}$. When $-\Delta\Delta S$ defined by

$$-\Delta\Delta S=-\Delta S_{\mathrm{M}}-(-\Delta S_{\mathrm{W}}) \tag{2.4}$$

is negative, the mutant is more stable than the wild type with respect to the entropic term.

The energy decrease $\Delta\Lambda$ ($\Delta\Lambda<0$) arising from the formation of intramolecular HBs is calculated for both of the first and second stages (see Figure 2.2). It is calculated by

$$\Delta\Lambda_{\mathrm{W}}=\Lambda_{\mathrm{compact,W}}-\Sigma(\Lambda_{\mathrm{extended}})_{\mathrm{W}} \tag{2.5}$$

where $\Lambda_{\mathrm{compact,W}}$ is the energy decrease for the associated α-helices, and $\Sigma(\Lambda_{\mathrm{extended}})_{\mathrm{W}}$ is that for the completely extended structures with no HBs: $\Sigma(\Lambda_{\mathrm{extended}})_{\mathrm{W}}=0$. That is, $\Delta\Lambda$ is calculated simply by counting the number of HBs in the folded structure. We calculate the same quantity for a mutant structure as

$$\Delta\Lambda_{\mathrm{M}}=\Lambda_{\mathrm{compact,M}}-\Sigma(\Lambda_{\mathrm{extended}})_{\mathrm{M}} \tag{2.6}$$

When $\Delta\Delta\Lambda$ defined by

$$\Delta\Delta\Lambda=\Delta\Lambda_{\mathrm{M}}-\Delta\Lambda_{\mathrm{W}} \tag{2.7}$$

is negative, the mutant is more stable than the wild type with respect to the energetic term. The change in the FEF ($T=T_0$) due to a mutation $\Delta\Delta$F is given by

$$\Delta\Delta F/(k_{\mathrm{B}}T_0)=\Delta\Delta\Lambda/(k_{\mathrm{B}}T_0)-\Delta\Delta S/k_{\mathrm{B}} \tag{2.8}$$

Negative $\Delta\Delta F$ implies that the mutant is more stable than the wild type.

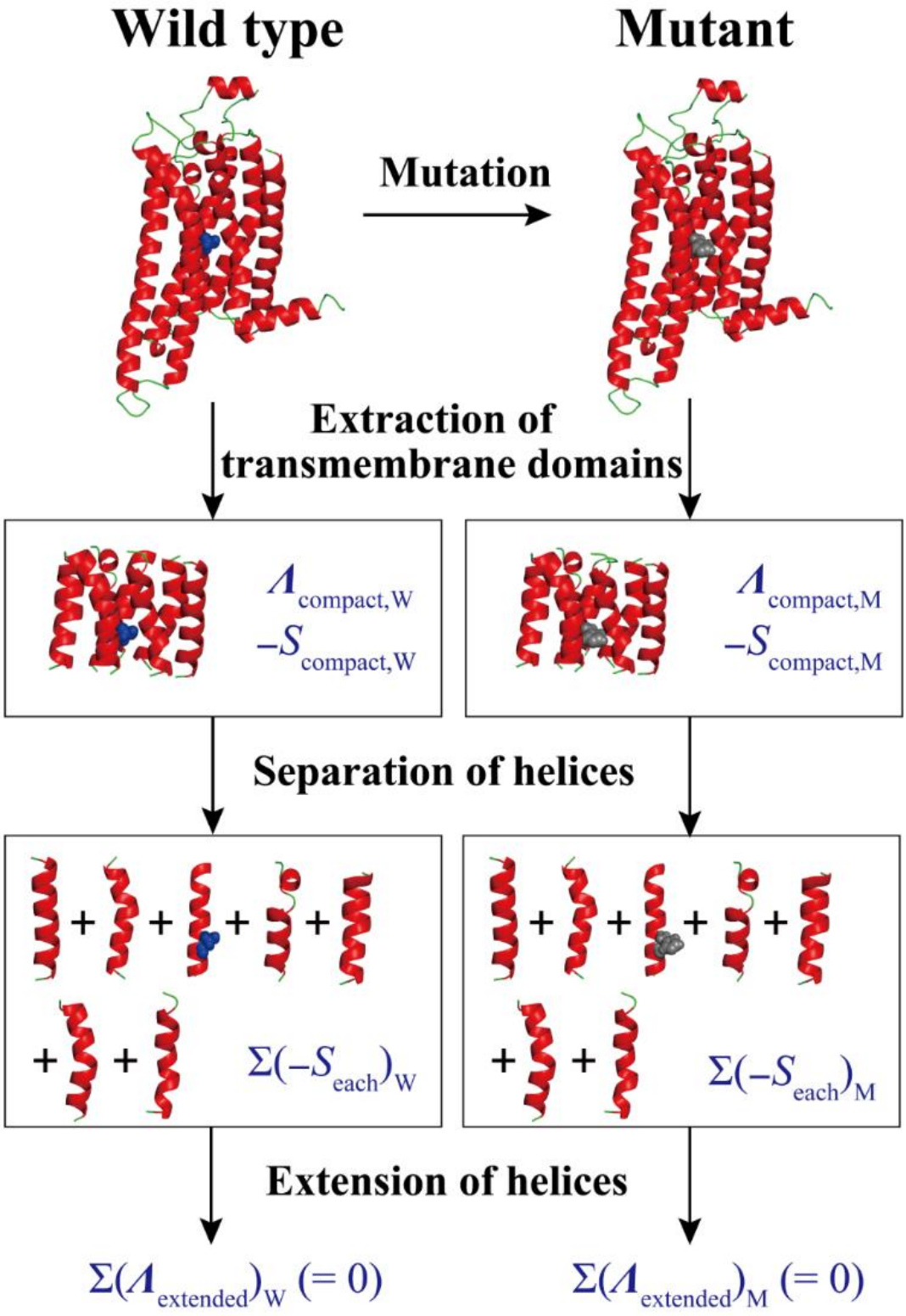


Figure 2.2. Calculation strategies for the entropic and energetic terms in our free-energy function. The case where S91 (shown in blue) is mutated to histidine (shown in gray) is given as an example. See Eqs. (2.2)–(2.4) for the entropic terms and Eqs. (2.5)–(2.7) for the energetic one.

2.2.9. Structure model for the wild type

Human adenosine A_{2a} receptor $A_{2a}R$ is chosen as a target in the present study (see Sec. 2.6.1). The crystal structure (PDB ID: 3VG9) of $A_{2a}R$ in an inactive state[37] is used as a template. Since the loop structure of P149–H155 is missing, it is merged from that of the other crystal structure of $A_{2a}R$ (PDB ID: 4EIY).[38] The Lennard-Jones potential energy for the wild-type structure thus obtained is positive and significantly large due to the unrealistic overlaps of protein atoms. Such overlaps are removed by local minimization of the energy (not our FEF) using the CHARMM biomolecular simulation program[39] through the multiscale modeling tools in structural biology (MMTSB) program.[40] The minimization is performed so that the original structures can be retained as much as possible. We employ the CHARMM22[41] as the force-field parameters. Electrostatic and nonbonded interactions are all evaluated without any cutoff. After the minimization, each structure is switched to a set of fused hard spheres for calculating the solvation entropy. We emphasize the following: All that matters is the removal of the overlaps of protein atoms with retaining the original structure, and the details of the minimization procedure are not important.

2.2.10. Structure model for a mutant

Mutant structures are generated by using the MODELLER program:[42] Starting from the wild-type structure, we move the residues which are present within a distance of 5 Å from the mutated residue. For each mutant, a total of 10 structures are generated and the unrealistic overlaps of protein atoms are removed by local minimization of the energy as mentioned above. The root-mean-square deviation (RMSD) for C_α atoms between the structures before and after the energetic local minimization is only 0.56 Å for the wild type and in the range 0.43–0.78 Å for the mutants: The original structures are almost completely retained. A thermodynamic quantity for a mutant is obtained as the average of the 10 values. We mutate each residue of $A_{2a}R$ from S6 to R309 to every amino acid residue other than the original one. Finally, we prepared 304×19×10=57760 mutants. Scanning of all of these mutations by our method takes only ~310 h using 20 central processing units on our standard workstation. Most of the computation time is devoted to the energetic local minimization process.

2.3. Experimental procedure

2.3.1. DNA construction, expression, and purification of $A_{2a}R$

The coding sequence of $A_{2a}R$ from residues 1−316 were amplified by PCR, in which Asn154 was replaced by Gln to eliminate N-linked glycosylation.[36] The DNA fragment was inserted into the plasmid pDDGFP-2[43] including TagRFP-His8 at the C-terminus by homologous recombination in Saccharomyces cerevisiae strain FGY217 as described previously.[44] The $A_{2a}R$-RFP fusion protein was expressed in FGY217,[45] and purified using 2% n-Dodecyl-β-D-maltoside (DDM) containing

0.2% cholesteril hemisuccinate (CHS) as described previously.[37] After metal affinity chromatography, A_{2a}R-RFP was cleaved using His-tagged TEV protease, and TagRFP and His-tagged TEV protease were removed by use of Ni Sepharose High Performance (GE Healthcare). The flow-through fraction as a final purified sample was concentrated to approximately 2 mg/ml by Amicon Ultra-4 (Millipore, 100K) after the solution buffer was exchanged to "20 mM HEPES (pH 7.5), 250 mM NaCl, 10% glycerol, 0.05% DDM, 0.01% CHS". A_{2a}R thus obtained is referred to as "wild type" in the present study. All amino-acid substitution mutants were introduced by site-directed polymerase chain reaction mutagenesis using PrimeSTAR Max DNA polymerase (TaKaRa), and purified by following the procedures described above. The expression levels and the amount of purified sample were estimated from the RFP fluorescence which was measured at 595 nm (excitation 535 nm) with the FilterMax F5 microplate reader (Molecular Devices).

2.3.2. Thermal stability assay

Thermal stability assay (the so-called CPM assay) was performed with N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM) as described previously.[46] Purified sample (2 mg/ml) and CPM dye (4 mg/ml in DMSO) were diluted to 0.2 mg/ml by adding the reaction buffer (20 mM Tris, 100 mM NaCl, 0.06% DDM: pH 7.5). The CPM assay was started by adding 5 µl of the diluted CPM dye into 75 µl of the protein solution, and the fluorescence of the reacted CPM dye was measured at 440 nm (excitation 350 nm) with Real-Time PCR Mx3005P (Agilent Technologies) as described previously.[45] The temperature was gradually increased from 25 to 90°C with an increment of 2°C per minute. To determine the inflection point of a melting curve, which was considered to equal the melting temperature T_m, a Boltzmann sigmoidal equation was fitted to the intensity plot using GraphPad Prism 4.0 (GraphPad Software, Inc., California, USA).

When a mutation is identified as a thermostabilizing one by our FEF, it implies that the folded structure of the corresponding mutant is more stable than that of the wild type. It is not definite if the two structures are essentially the same in the sense that the activity is completely retained in the mutant. Such an activity assay, which is to be carried out by another experiment, is beyond our scope in the present study.

2.4. Results and discussion

2.4.1. Nomination of key residues to be mutated for highly improving the thermostability

In Kinoshita and coworkers' earlier work for water-soluble proteins, Kinoshita and coworkers investigated essential physical factors governing the thermal-stability changes upon mutations.[21] An important finding was that a mutation which largely increases the water-entropy gain upon folding often leads to very high enhancement of the thermal stability irrespective of the enthalpic factor. By the analogy of this finding, we first focused on the entropic term of our FEF. Figure 2.3 shows the

values of $-\Delta\Delta S$ calculated for all of the possible mutations for $A_{2a}R$. It is observed that many of the mutations for T88 and S91 possess relatively lower values of $-\Delta\Delta S$: 12 mutants of T88 and 11 mutants of S91 are included in the top 100 mutants predicted to give the lowest values of $-\Delta\Delta S$. Therefore, we nominated T88 and S91 as key residues whose mutants were worth examining by our experiments to find highly thermostabilizing mutations.

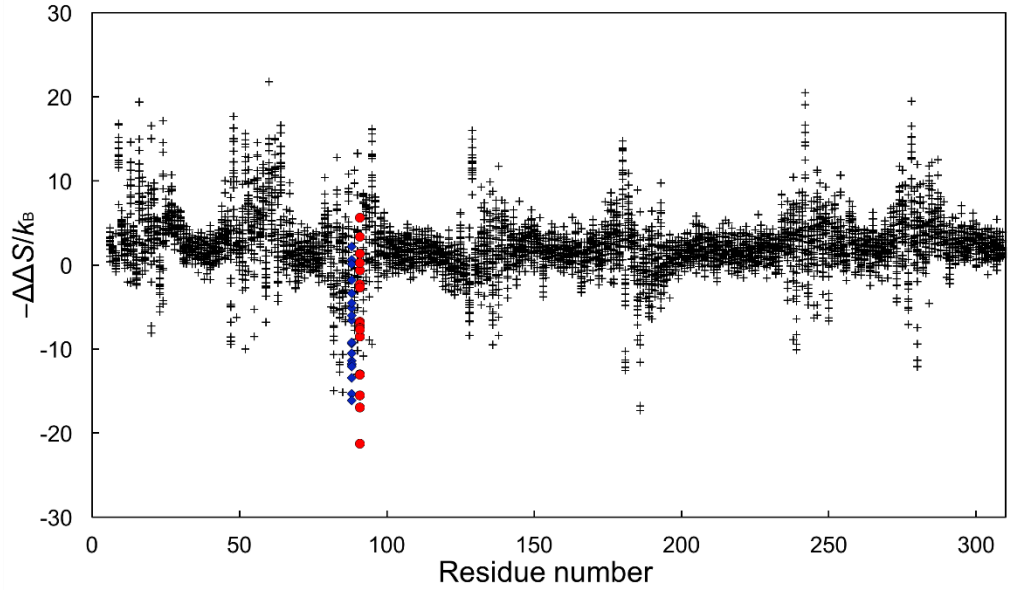


Figure 2.3. Values of $-\Delta\Delta S$ for all of the possible mutations for $A_{2a}R$. Negative and positive values of $-\Delta\Delta S$ mean that the mutation enhances and lowers the thermostability, respectively. Blue squares denote the values of $-\Delta\Delta S$ of the mutations for T88 and red circles denote those for S91.

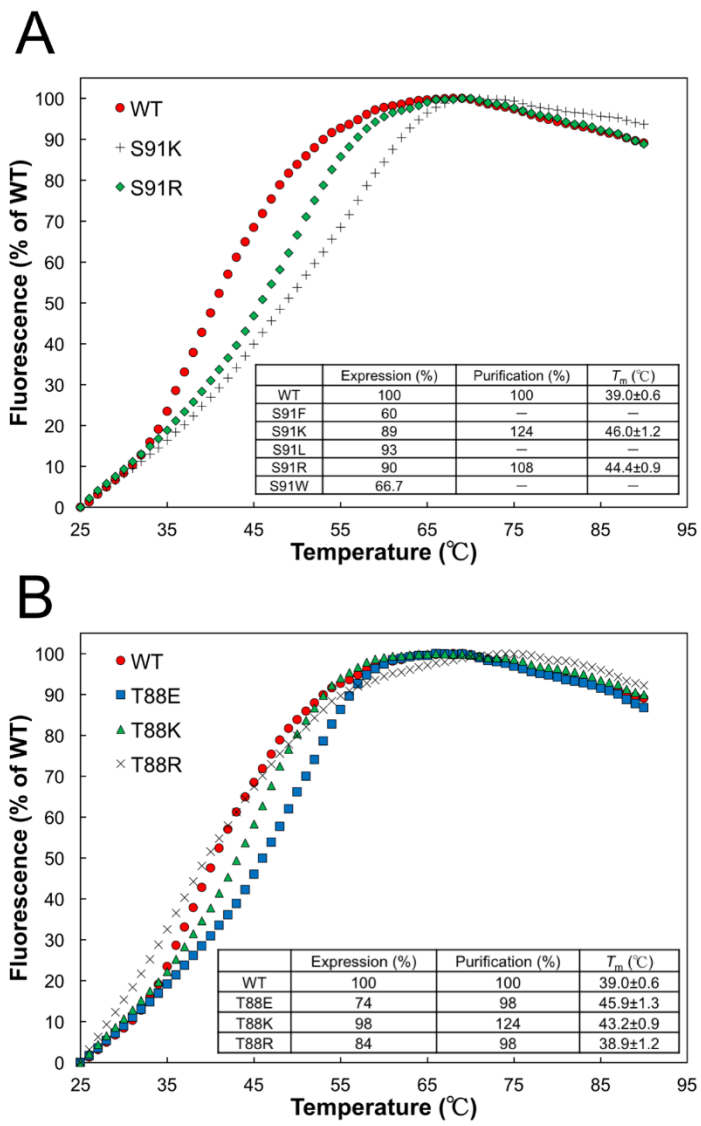
2.4.2. Stabilizing mutations for S91

The values of $-\Delta\Delta S$ calculated for the mutants of S91 are given in Table 2.1. We selected S91F, S91K, S91L, S91R, and S91W from among those giving considerably low values of $-\Delta\Delta S$ and examined their thermostabilities by our experiments. The experimental results for the five mutants are shown in the inset of Figure 2.4(A). It should be noted that when a mutant cannot be purified, it is less stable than the wild type. Purification of S91K and S91R was successful (the purification method was similar to that of the wild type). We measured T_m for S91K, S91R, and the wild type by the CPM assay and found that the values were 45.9, 43.2, and 39.0°C, respectively. On the other hand, S91L, S91F, and S91W could not be purified, suggesting that the thermal stabilities of these three mutants were significantly lowered. This result indicates that the energetic term, which was not taken into account in the above prediction, may also be important. The energetic term $\Delta\Delta\Lambda$ and the free-energy change $\Delta\Delta F$ calculated are also given in Table 2.1. It is observed for S91F and S91W that $\Delta\Delta\Lambda$ takes considerably large, positive values, leading to positive values of $\Delta\Delta F$. S91F

and S91W are predicted to be destabilizing mutations in terms of $\Delta\Delta F$, which is consistent with the experimental observation. The calculation result that $\Delta\Delta F$ for S91L is negative implies the failure of our prediction. However, $|\Delta\Delta F|$ is quite small. It can be suggested that only a mutation giving a sufficiently large, negative value of $\Delta\Delta F$ (e.g., $\Delta\Delta F/(k_BT_0)<-5$) be regarded as a stabilizing mutation. For S91K and S91R, $\Delta\Delta\Lambda$ as well as $-\Delta\Delta S$ takes considerably large, negative values, leading to the high stabilities corroborated by our experiments. Consequently, we could successfully identify stabilizing mutations, S91K and S91R, whose values of T_m increase by ~7°C and ~4°C, by our theoretical method. Although the stabilizing mutations can be identified more accurately in terms of $\Delta\Delta F$, the nomination of key residues to be mutated using $-\Delta\Delta S$ has been justified as explicated above. The entropic effect arising from the translational displacement of hydrocarbon groups within the lipid bilayer as well as the protein intramolecular hydrogen bonding is crucially important for the structural stability of $A_{2a}R$.

Table 2.1. Values of $-\Delta\Delta S$, $\Delta\Delta\Lambda$, and $\Delta\Delta F$ for S91 Mutated to Residues Other than Ser (T_0=298 K)

Mutant	$-\Delta\Delta S/k_B$	$\Delta\Delta\Lambda/(k_BT_0)$	$\Delta\Delta F/(k_BT_0)$
S91A	3.36	8.91	12.28
S91C	–0.62	–2.03	–2.66
S91D	–2.59	4.08	1.49
S91E	–6.68	0.18	–6.51
S91F	–12.95	17.06	4.11
S91G	5.57	17.25	22.81
S91H	–8.42	–10.27	–18.69
S91I	–6.89	6.71	–0.18
S91K	–13.01	–21.63	–34.65
S91L	–8.45	5.98	–2.47
S91M	–7.47	10.66	3.19
S91N	–7.61	–13.91	–21.52
S91P	1.29	26.84	28.13
S91Q	–2.26	–7.63	–9.88
S91R	–21.19	–12.53	–33.72
S91T	0.26	6.78	7.04
S91V	–2.66	4.45	1.79
S91W	–15.41	20.79	5.38
S91Y	–16.88	6.05	–10.83



	Expression (%)	Purification (%)	T_m (℃)
WT	100	100	39.0±0.6
S91F	60	–	–
S91K	89	124	46.0±1.2
S91L	93	–	–
S91R	90	108	44.4±0.9
S91W	66.7	–	–

	Expression (%)	Purification (%)	T_m (℃)
WT	100	100	39.0±0.6
T88E	74	98	45.9±1.3
T88K	98	124	43.2±0.9
T88R	84	98	38.9±1.2

Figure 2.4. (A) Comparison among the wild type of $A_{2a}R$ and its single mutations for S91 in terms of the thermostability. S91L, S91F, and S91W mutants could not be purified with the result of no curves measured. (B) Comparison among the wild type (WT) of $A_{2a}R$ and its single mutations for T88 in terms of the thermostability. These two figures show unfolding curves measured in the temperature range 25–90°C. The inset shows the results for the expression, purification, and melting temperature T_m. The expression and the purification are normalized by these of the WT. T_m was determined as the inflection point of each unfolding curve. The experiments were independently carried out three times. A representative unfolding curve is shown here and the average value is adopted for each numerical data in the inset. The standard error for T_m is also given.

2.4.3. Stabilizing mutations for T88

The values of $-\Delta\Delta S$, $\Delta\Delta\Lambda$, and $\Delta\Delta F$ calculated for the mutants of T88 are collected in Table 2.2. We selected T88E, T88K, and T88R from among those giving considerably low values of $-\Delta\Delta S$ for the experimental test. The experimental results for the three mutants are shown in Figure 2.4(B). We purified these mutants and measure the values of T_m by the CPM assay: The values for T88E, T88K, and T88R were 45.9, 43.2, and 38.9°C, respectively. T88E and T88K were actually stabilizing mutations though the order of T_m is not in accord with that of $\Delta\Delta F/(k_B T_0)$, −8.93 for T88E and −43.60 for T88K. It is interesting to note that T88E is predicted to cause destabilization if only $\Delta\Delta\Lambda$ is considered by omitting $-\Delta\Delta S$.

T88R does not bring stabilization in terms of T_m despite its low $\Delta\Delta F$: T88R and the wild type share almost the same value of T_m. We note, however, that T_m is not necessarily the best measure of the thermostability. As described in Sec. 2.6.2, we measured unfolding curves of the wild type and its mutants measured at 35°C for 30 min. The half-life time τ at which the percentage of folded protein decreased to 50% is another good measure: The values of τ for the wild type and T88R are 9.0 and 28.8 min, respectively, which indicates that T88R is significantly more stable than the wild type at 35°C. It should be emphasized that the theoretical prediction is made using F calculated at 25°C, though the temperature dependence of F for a mutant is different from that of the wild type. We can argue that the prediction result is valid up to 35°C but an inversion occurs when the temperature is further increased. Taken together, T88R is a stabilizing mutation in terms of τ. We note that T88A was identified as a stabilizing mutation in Ala scanning experiments.[8] It is observed in Table 2.2 that T88A is certainly a stabilizing mutant in terms of $-\Delta\Delta S$ and $\Delta\Delta F$, but not in terms of $\Delta\Delta\Lambda$. It is worthwhile to repeat that the translational entropy of hydrocarbon groups and the intramolecular hydrogen bonding are both crucially important.

Table 2.2. Values of $-\Delta\Delta S$, $\Delta\Delta\Lambda$, and $\Delta\Delta F$ for T88 Mutated to Residues Other than Thr (T_0=298 K)

Mutant	$-\Delta\Delta S/k_B$	$\Delta\Delta\Lambda/(k_B T_0)$	$\Delta\Delta F/(k_B T_0)$
T88A	–9.26	2.28	–6.98
T88C	–6.46	–5.15	–11.61
T88D	–1.71	9.13	7.42
T88E	–15.12	6.19	–8.93
T88F	–11.25	4.16	–7.09
T88G	–3.30	4.50	1.19
T88H	–11.71	0.66	–11.05
T88I	–5.02	1.88	–3.14
T88K	–10.39	–33.21	–43.60

T88L	–5.88	3.73	–2.15
T88M	–13.24	2.99	–10.25
T88N	–4.41	7.39	2.98
T88P	2.28	22.20	24.48
T88Q	–11.60	–5.90	–17.49
T88R	–11.99	–23.86	–35.84
T88S	0.60	7.49	8.09
T88V	0.21	6.23	6.44
T88W	–9.12	11.23	2.11
T88Y	–16.0	13.52	–2.48

2.4.4. Strategy for identifying thermostabilizing mutations

The results explained above are summarized in Figure 2.5 ((A) for S91 and (B) for T88). On the basis of the results of our tests for S91 and T88, we propose the following strategy as a reliable one: First, calculate S for all of the possible mutations and nominate key residues to be mutated in the sense that many of their mutations will lead to relatively higher enhancement of the stability; second, select stabilizing mutations of the key residues using F. In the second step, only the mutants giving $\Delta\Delta F$ sufficiently low values (e.g., $\Delta\Delta F/(k_B T_0)<-5$) should be selected.

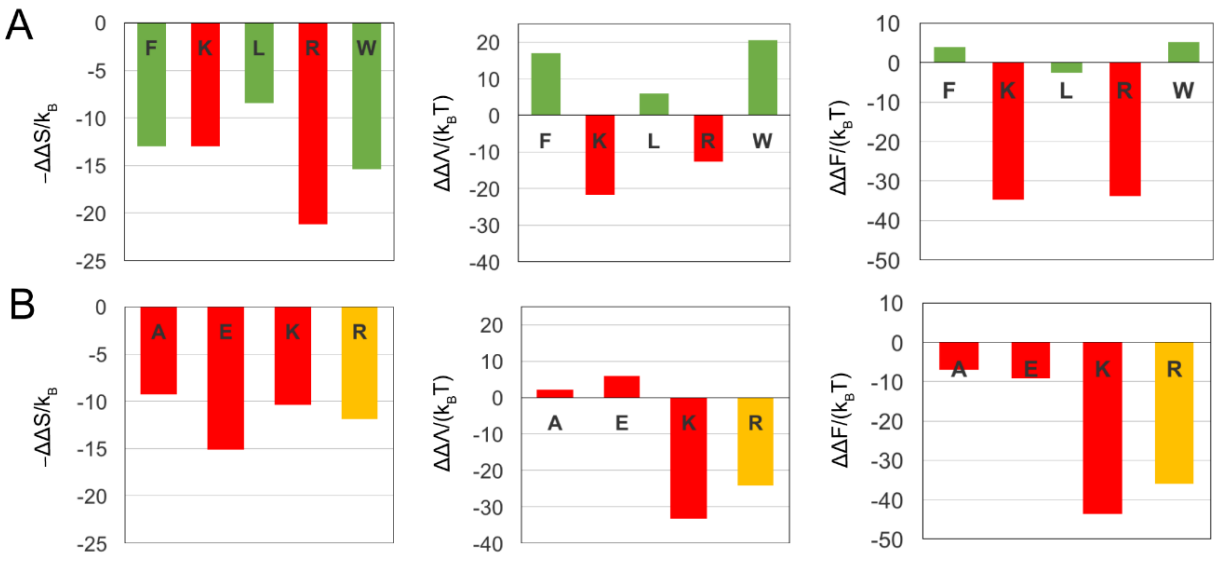
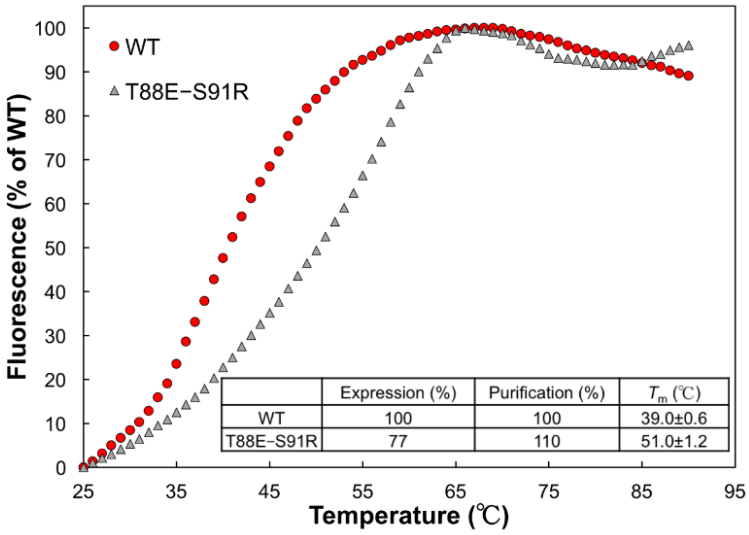


Figure 2.5. (A) Values of $-\Delta\Delta S/k_B$, $\Delta\Delta\Lambda/(k_BT_0)$, and $\Delta\Delta F/(k_BT_0)$ (T_0=298 K) for S91F, S91K, S91L, S91R, and S91W. The bars for mutants experimentally shown to be stabilizing and destabilizing in terms of the melting temperature T_m are colored in red and green, respectively. The success rates with $-\Delta\Delta S/k_B$, $\Delta\Delta\Lambda/(k_BT_0)$, and $\Delta\Delta F/(k_BT_0)$ are 2/5, 5/5, and 4/5, respectively. (B) Values of $-\Delta\Delta S/k_B$, $\Delta\Delta\Lambda/(k_BT_0)$, and $\Delta\Delta F/(k_BT_0)$ (T_0=298 K) for T88A, T88E, T88K, and T88R. The bars for mutants experimentally shown to be stabilizing and destabilizing in terms of the melting temperature T_m are colored in red and green, respectively, except that the bar for T88R is colored in orange: T88R is destabilizing in terms of T_m but stabilizing in terms of the half-life time τ. If the stability change is judged from T_m, the success rates with $-\Delta\Delta S/k_B$, $\Delta\Delta\Lambda/(k_BT_0)$, and $\Delta\Delta F/(k_BT_0)$ are 3/4, 1/4, and 3/4, respectively. When $\Delta\Delta F/(k_BT_0)<-5$ is employed as the criterion for the thermostability relative to that of the wild type in (A) and T88R is determined to be stabilizing on the basis of τ in (B), the overall success rate with $\Delta\Delta F/(k_BT_0)$ is 9/9.

2.4.5. Introducing a double mutation

It is of great interest to see how much T_m increases by a double mutation. To this end, we selected T88E−S91R (S91 and T88 are nonadjacent) because S91R and T88E give considerably low values of $-\Delta\Delta S$ as observed in Tables 2.1 and 2.2, respectively, and calculated its $-\Delta\Delta S$, $\Delta\Delta\Lambda$, and $\Delta\Delta F$. The calculation results are as follows: $-\Delta\Delta S/k_B$=−34.78, $\Delta\Delta\Lambda/(k_BT_0)$=−18.33, and $\Delta\Delta F/(k_BT_0)$=−53.11. $-\Delta\Delta S$ and $\Delta\Delta F$ are lower than those for any of the single mutations considered in Tables 2.1 and 2.2. The experimental results are shown in Figure 2.6. T_m for the double mutant was 51.0°C that is higher than that of the wild type by ~12°C. As described in Sec. 2.6.2, the values of τ for the wild type and T88E−S91R are 9.0 and 49.6 min, respectively. This result suggests that our theoretical method is applicable not only to a single mutant but also to a double one for predicting thermostabilizing mutations.



	Expression (%)	Purification (%)	T_m (℃)
WT	100	100	39.0±0.6
T88E−S91R	77	110	51.0±1.2

Figure 2.6. Comparison among the wild type of $A_{2a}R$ and its double mutant for S91R and T88E in

terms of the thermostability. This figure shows their unfolding curves measured in the temperature range 25−90°C. The data points for the wild type (WT) and S91R−T88E are indicated by red circles and gray triangles, respectively. The inset shows the results for the expression, purification, and melting temperature T_m. The expression and the purification are normalized by these of the WT. T_m was determined as the inflection point of each unfolding curve. The experiments were independently carried out three times. A representative unfolding curve is shown here and the average value is adopted for each numerical data in the inset. The standard error for T_m is also given.

2.4.6. Effect of solvent model

Last, we examine how the results are changed when the hard-sphere solvent is replaced by water. The four coefficients in the morphometric form for the solvation entropy of a protein with a prescribed structure (see Sec. 2.6.2.) are C_1=−0.26966 (Å^{-3}), C_2=0.21418 (Å^{-2}), C_3=−0.18719 (Å^{-1}), and C_4=0.05103 for the hard-sphere solvent and C_1=−0.19676 (Å^{-3}), C_2=0.04517 (Å^{-2}), C_3=0.25671 (Å^{-1}), and C_4=−0.35690 for water. The values of $-\Delta\Delta S$, $\Delta\Delta\Lambda$, and $\Delta\Delta F$ of the mutations for S91 and those of the mutations for T88 are collected in Tables 2.3 and 2.4, respectively, for the case of water. Tables 2.3 and 2.4 should be compared to Tables 2.1 and 2.2, respectively ($\Delta\Delta\Lambda$ remains unchanged). Overall, the differences among the mutants in $-\Delta\Delta S$ are magnified when the hard-sphere solvent is replaced by water: The contribution from $-\Delta\Delta S/k_B$ to $\Delta\Delta F/(k_BT_0)$ becomes larger. As a consequence, all of S91F, S91K, S91L, S91R, S91W, T88E, T88K, and T88R are predicted to be quite thermostabilizing in terms of $\Delta\Delta F$. It is important to note that the prediction ends with failure for S91F, S91L, and S91W. We can conclude that since the membrane protein is directly immersed not in water but in nonpolar environment, the hard-sphere solvent is more suitable than water for the solvent model.

Table 2.3. Values of $-\Delta\Delta S$, $\Delta\Delta\Lambda$, and $\Delta\Delta F$ for S91 Mutated to Residues Other than Ser (T_0=298 K)

Mutant	$-\Delta\Delta S/k_B$	$\Delta\Delta\Lambda/(k_BT_0)$	$\Delta\Delta F/(k_BT_0)$
S91A	–1.55	8.91	7.36
S91C	–12.10	–2.03	–14.13
S91D	–19.47	4.08	–15.39
S91E	–21.55	0.18	–21.38
S91F	–27.07	17.06	–10.01
S91G	10.24	17.25	27.49
S91H	–25.37	–10.27	–35.63
S91I	–22.83	6.71	–16.12
S91K	–33.44	–21.63	–55.07
S91L	–25.38	5.98	–19.40

S91M	–17.66	10.66	–7.00
S91N	–21.16	–13.91	–35.07
S91P	1.35	26.84	28.19
S91Q	–15.30	–7.63	–22.93
S91R	–42.67	–12.53	–55.20
S91T	–9.55	6.78	–2.78
S91V	–11.46	4.45	–7.01
S91W	–42.40	20.79	–21.62
S91Y	–31.39	6.05	–25.34

Table 2.4. Values of $-\Delta\Delta S$, $\Delta\Delta\Lambda$, and $\Delta\Delta F$ for T88 Mutated to Residues Other than Thr (T_0=298 K)

Mutant	$-\Delta\Delta S/k_B$	$\Delta\Delta\Lambda/(k_BT_0)$	$\Delta\Delta F/(k_BT_0)$
T88A	–14.56	2.28	–12.28
T88C	–22.96	–5.15	–28.11
T88D	–11.36	9.13	–2.24
T88E	–48.31	6.19	–42.12
T88F	–33.24	4.16	–29.07
T88G	–8.6	4.5	–4.1
T88H	–33.56	0.66	–32.89
T88I	–27.97	1.88	–26.09
T88K	–27.82	–33.21	–61.03
T88L	–26.72	3.73	–22.98
T88M	–37.81	2.99	–34.82
T88N	–19.9	7.39	–12.51
T88P	–9.55	22.2	12.65
T88Q	–35.8	–5.9	–41.7
T88R	–25.42	–23.86	–49.28
T88S	2.7	7.49	10.22
T88V	–21.23	6.23	–15.0
T88W	–18.86	11.23	–7.63
T88Y	–29.22	13.52	–15.7

2.5. Conclusion

We have adapted our recently developed free-energy function[13] (FEF) to the identification of thermostabilizing mutations for membrane proteins. The FEF comprises the entropic and energetic terms which are denoted by S and Λ, respectively. S is featured by the translational entropy of hydrocarbon groups of nonpolar chains within the lipid bilayer. Λ is based on the intramolecular hydrogen bonding. The calculation of the FEF can be finished in less than 1 sec per structure, and we can consider all of the possible mutations. The identification method thus obtained has been illustrated for the adenosine A_{2a} receptor ($A_{2a}R$) whose wild-type structure in an inactive state is known and utilizable. The significance of the FEF and its two terms has been examined by our experiments. Some of the important results are recapitulated as follows.

(1) S and Λ are both essential, and neither of them can be omitted.

(2) A reliable strategy is as follows: First, calculate S for all of the possible mutations and nominate key residues to be mutated in the sense that many of their mutations will lead to relatively higher enhancement of the stability; second, select some mutations of the key residues using F. In the second step, only the mutants giving $\Delta\Delta F$ sufficiently low values (e.g., $\Delta\Delta F/(k_B T_0)<-5$) should be selected.

(3) For $A_{2a}R$, T88 and S91 are the key residues.

(4) Not only the melting temperature T_m but also the half-life time τ at a moderately high temperature (35°C in the present study), which is defined as the time at which the percentage of folded protein decreased to 50%, is a good measure of the thermostability.

(5) T_m becomes higher than that of the wild type by ~7°C for S91K and T88E despite that they are single mutations. The values of τ for the wild type, S91K, and T88E are 9.0, 28.1, and 32.0 min, respectively.

(6) It is promising to combine two of the stabilizing single mutations and verify the high stability of the resultant double mutation using F.

(7) For $A_{2a}R$, the double mutation T88E−S91R leads to an increase in T_m of ~12°C (its τ reaches 49.6 min).

(8) With F, the success rates are in the range from 7/9 to 9/9 depending on the criterion employed for the thermostability relative to that of the wild type in the theoretical prediction and on the thermostability measure adopted in our experiments (see Figure 2.5). Though the number of the mutations considered is only 9, the success rate is sufficiently high for demonstrating the validity of our FEF.

As explained in Sec. 2.6.3, S91 is conserved in significantly many other G protein-coupled receptors (GPCRs). The mutations, S91K and S91R, for those GPCRs can lead to equally large stabilization. (This is not the case for T88.) It is probable that there are significantly many mutations whose stability enhancement is higher than that from Ala scanning. The next subjects to be pursued are the following: testing multiple mutations for $A_{2a}R$, refining the procedure for generating mutant structures, and applying our method to other membrane proteins including GPCRs. In the last subject, we are especially concerned with the case where the wild-type structure is unknown and must therefore be constructed using such tools as the homology modeling. Actually, we have already made significant progresses as follows. T_m of an octuple mutation (A54L, T88A, R107A, K122A, L202A, L235A, V239A, and S277A; seven residues are mutated to Ala) for the adenosine A_{2a} receptor was experimentally shown to be higher than that of the wild type by ~20°C.[47] The result of an analysis using our FEF is quite consistent with this experimental evidence, and the physical origins of the higher stability have also been clarified by the entropic and energetic terms. We have succeeded in greatly raising the stability of a GPCR whose structure is unknown, leading to the acquisition of its high-resolution structural data for the first time. These achievements will be published in future articles.

Appendix 2-A: Adenosine A_{2a} receptor

Human adenosine A_{2a} receptor ($A_{2a}R$) is an important GPCR involved in the control of various physiological activities, including regulation of glutamine and dopamine release in the brain.[48] Specific inhibitors of the receptor are in advanced clinical trials for the treatment of Parkinson's disease.[49] Purified $A_{2a}R$ in detergents is not very stable at room temperature. Half of the sample is denatured in 30 min incubation at 30°C .[50] Several crystal structures of thermostabilizing $A_{2a}R$ mutants (T4 lysozyme [or b562RIL]-fusion mutants in intercellular loop 3 [ICL3] and alanine-scanning mutants) have been obtained by breakthroughs in protein engineering.[37,46,50,51] We obtained the crystal structure (PDBID: 3VG9)[36] of $A_{2a}R$ in complex with a mouse monoclonal-antibody Fab-fragment, which is used as a model structure in this work.

Appendix 2-B: Thermal stabilities examined in terms of unfolding curves at a constant temperature

The thermal stabilities of the wild type and its mutants can be evaluated from a different viewpoint. Figure 2.7 shows unfolding curves measured at 35°C for 30 min. It is observed that T88R is more stable than the wild type.

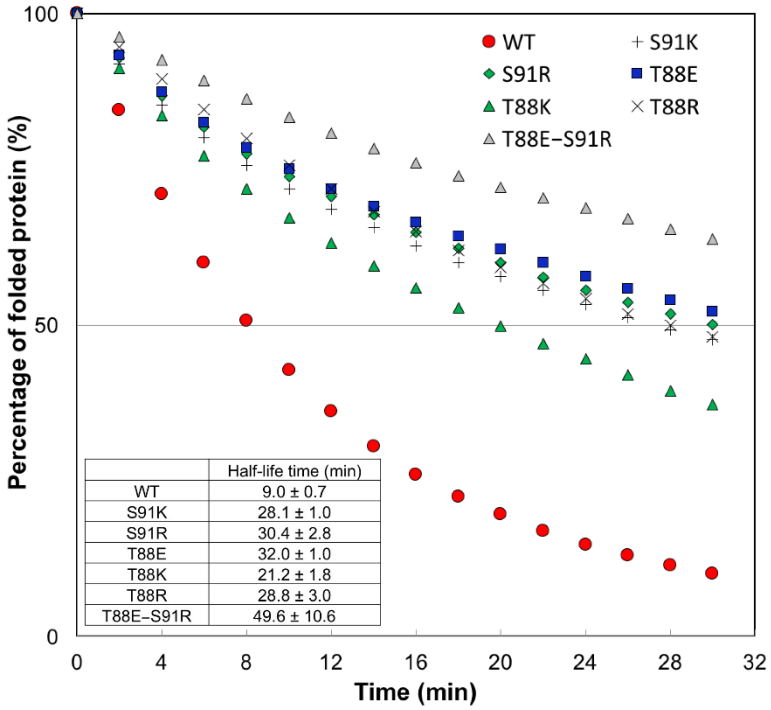


Figure 2.7. Thermal stabilities of purified $A_{2a}R$ mutants. This figure shows unfolding curves of the wild type and its mutants measured at 35°C for 30 min. The percentage of folded protein was calculated by the quotient of raw fluorescence measured at each time point divided by the maximal fluorescence as described elsewhere.[52] The inset shows the half-life time at which the percentage of folded protein decreased to 50%. The experiments were independently carried out three times. A representative unfolding curve is shown here and the average value is adopted for each numerical data in the inset. The standard error for the half-life time is also given.

Appendix 2-C: High conservation of S91 in G protein-coupled receptors

We choose ten representative GPCRs from among those belonging to Class A whose wild-type structures are experimentally available and check their sequence alignment. We then find that S91 is conserved in many of them. See Figure 2.8.

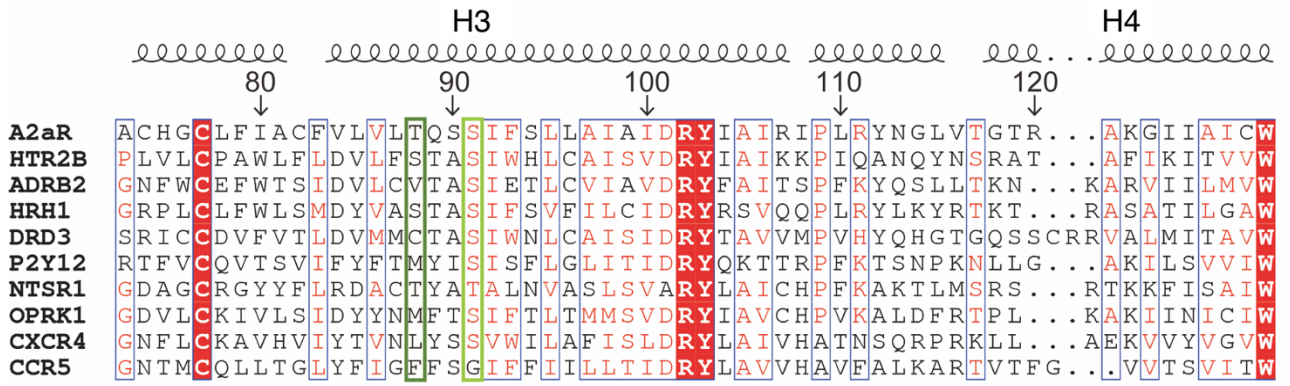


Figure 2.8. Sequence alignment of ten representative GPCRs belonging to class A; adenosine A_{2a} receptor (A2aR), serotonin 5-HT2B receptor (HTR2B), β2-adrenergic receptor (ADRB2), histamine H1 receptor (HRH1), dopamine D3 receptor (DRD3), P2Y12 receptor (P2Y12), rat neurotensin receptor 1 (NTSR1), kappa opioid receptor (OPRK1), C-X-C chemokine receptor 4 (CXCR4), and C-C chemokine receptor 5 (CCR5). The residue numbers and secondary structures of $A_{2a}R$ are indicated on its sequence. The residues corresponding to T88 and S91 of $A_{2a}R$ are surrounded by dark green and light green squares, respectively. S91 is conserved except in NTSR1 and CCR5. By contrast, the conservation of T88 is much lower.

References

[1]Boyd, D.; Schierle, C.; Beckwith, J., Protein Sci. **1998**, 7, 201−205.

[2]Wallin, E.; Heijne, G. V., Protein Sci. **1998**, 7, 1029−1038.

[3]Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabási, A. L.; Vidal, M., Nat. Biotechnol. **2007**, 25, 1119−1126.

[4]Lau, F. W.; Nauli, S.; Zhou, Y.; Bowie, J. U., J. Mol. Biol. **1999**, 290, 559−564.

[5]Zhou, Y.; Bowie, J. U., J. Biol. Chem. **2000**, 275, 6975−6979.

[6]Scott, D. J.; Kummer, L.; Tremmel, D.; Plückthun, A., Curr. Opin. Chem. Biol. **2013**, 17, 427−435.

[7]Serrano-Vega, M. J.; Magnani, F.; Shibata, Y.; Tate, C. G., Proc. Natl. Acad. Sci. U.S.A. **2008**, 105, 877−882.

[8]Magnani, F.; Shibata, Y.; Serrano-Vega, M. J.; Tate, C. G., Proc. Natl. Acad. Sci. U.S.A. **2008**, 105, 10744−10749.

[9]Shibata, Y.; White, J. F.; Serrano-Vega, M. J.; Magnani, F.; Aloia, A. L.; Grisshammer, R.; Tate, C. G., J. Mol. Biol. **2009**, 390, 262−277.

[10]Chen, K. M.; Zhou, F.; Fryszczyn, B. G.; Barth, P. Proc., Proc. Natl. Acad. Sci. U.S.A. **2012**, 109, 13284−13289.

[11]Bhattacharya, S.; Lee, S.; Grisshammer, R.; Tate, C. G.; Vaidehi, N., J. Chem. Theory Comput. **2014**, 10, 5149−5160.
[12]Sauer, D. B.; Karpowich, N. K.; Song J. M.; Wang, D. N., Biophys. J. **2015**, 109, 1420−1428.
[13]Yasuda, S.; Oshima, H.; Kinoshita, M., J. Chem. Phys. **2012**, 137, 135103.
[14]Harano, Y.; Kinoshita, M., Biophys. J. **2005**, 89, 2701−2710.
[15]Yoshidome, T.; Kinoshita, M.; Hirota, S.; Baden, N.; Terazima, M., J. Chem. Phys. **2008**, 128, 225104.
[16]Kinoshita, M. Roles of translational motion of water molecules in sustaining life. Front. Biosci. **2009**, 14, 3419−3454.
[17]Kinoshita, M., Int. J. Mol. Sci. **2009**, 10, 1064−1080.
[18]Kinoshita, M., Biophys. Rev. **2013**, 5, 283−293.
[19]Oshima, H.; Kinoshita M., J. Chem. Phys. **2013**, 138, 245101.
[20]Oshima, H.; Kinoshita M., J. Chem. Phys. **2015**, 142, 145103.
[21]Murakami, S.; Oshima, H.; Hayashi T.; Kinoshita M., J. Chem. Phys. **2015**, 143, 125102.
[22]Kinoshita, M., J. Chem. Phys. **2008**, 128, 024507.
[23]Kauzmann, W., Adv. Protein Chem. **1959**, 14, 1−63.
[24]Yoshidome, T.; Kinoshita, M., Phys. Chem. Chem. Phys. **2012**, 14, 14554−14566.
[25]Bowie, J. U., Curr. Opin. Struct. Biol. **2011**, 21, 42−49.
[26]Oda, K.; Kinoshita, M., Biophys. Physicobiol. **2015**, 12, 1−12.
[27]Yoshidome, T.; Oda, K.; Harano, Y.; Roth, R.; Sugita, Y.; Ikeguchi, M.; Kinoshita, M., Proteins **2009**, 77, 950−961.
[28]Yasuda, S.; Yoshidome, T.; Harano, Y.; Roth, R.; Oshima, H.; Oda, K.; Sugita, Y.; Ikeguchi, M.; Kinoshita, M., Proteins **2011**, 79, 2161−2171.
[29]Popot, J. L.; Engelman D. M., Biochemistry **2016**, 55, 5−18.
[30]Hansen, J.-P.; McDonald, I. R. Theory of Simple Liquids, 3rd ed. Academic Press; London, U.K., 2006.
[31]König, P. M.; Roth, R.; Mecke, K. R., Phys. Rev. Lett. **2004**, 93, 160601.
[32]Roth, R.; Harano, Y.; Kinoshita, M., Phys. Rev. Lett. **2006**, 97, 078101.
[33]McDonald, I. K.; Thornton, J. M., J. Mol. Biol. **1994**, 238, 777−793.
[34]Sneddon, S. F.; Tobias, D. J.; Brooks, C. L. III, J. Mol. Biol. **1989**, 209, 817−820.
[35]Popot, J. L.; Engelman D. M., Annu. Rev. Biochem. **2000**, 69, 881−922.
[36]Tusnády, G. E.; Dosztányi, Z.; Simon, I., Bioinformatics **2005**, 21, 1276−1277.
[37]Hino, T.; Arakawa, T.; Iwanari, H.; Yurugi-Kobayashi, T.; Ikeda-Suno, C.; Nakada-Nakura, Y.; Kusano-Arai, O.; Weyand, S.; Shimamura, T.; Nomura, N.; et al., Nature **2012**, 482, 237−240.
[38]Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; Ijzerman, A. P.; et al., Science **2012**, 337, 232−236.
[39]Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., J. Comput. Chem. **1983**, 4, 187−217.

[40]Feig, M.; Karanicolas, J.; Brooks III, C. L., J. Mol. Graph. Model. **2004**, 22, 377−395.

[41]MacKerell, Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr. R.L.; Evanseck, J.D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al., J. Phys. Chem. B **1998**, 102, 3586−3616.

[42]Šali, A.; Blundell, T. L., J. Mol. Biol. **1993**, 234, 779−815.

[43]Newstead, S.; Kim, H.; von Heijne, G.; Iwata, S.; Drew, D., Proc. Natl. Acad. Sci. U. S. A. **2007**, 104, 13936−13941.

[44]Kota, J.; Gilstring, C. F.; Ljungdahl, P. O., J. Cell Biol. **2007**, 176, 617−628.

[45]Shiroishi, M.; Tsujimoto, H.; Makyio, H.; Asada H.; Yurugi-Kobayashi, T.; Shimamura, T.; Murata, T.; Nomura, N.; Haga, T.; Iwata S.; et al., Microb. Cell Fact. **2012**, 11, 78.

[46]Alexandrov, A. I.; Mileni, M.; Chien, E. Y.; Hanson, M. A.; Stevens R. C., Structure **2008**, 16, 351−359.

[47]Doré, A. S.; Robertson, N.; Errey, J. C.; Ng, I.; Hollenstein, K.; Tehan, B.; Hurrell, E.; Bennett, K.; Congreve, M.; Magnani, F.; et al. H., Structure **2011**, 19, 1283−1293.

[48]Fredholm, B. B.; Chen, J. F.; Masino, S. A.; Vaugeois, J. M. Annu. Rev. Pharmacol. Toxicol. **2005**, 45, 385.

[49]Müller, C. E.; Jacobson, K. A. Biochim. Biophys. Acta **2011**, 1808, 1290.

[50]Xu, F.; Wu, H.; Katritch, V.; Han, G. W.; Jacobson, K .A.; Gao, Z. G.; Cherezov, V.; Stevens, R. C. Science **2011**, 332. 322.

[51]Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G. W.; Tate, C. G. Nature **2011**, 474. 521.

[52]Sonoda, Y.; Newstead, S.; Hu, N. J.; Alguel, Y.; Nji, E.; Beis, K.; Yashiro, S.; Lee, C.; Leung, J.; Cameron, A. D.; Byrne, B.; Iwata, S.; Drew, D. Structure **2011**, 19, 17.

Chapter 3

Physical Origins of Remarkable Thermostabilization by an Octuplet Mutation for the Adenosine A_{2a} Receptor

3.1. Introduction

Membrane proteins such as G-protein coupled receptors (GPCRs) are imperative in the life phenomena.[1] The problems caused in their functions lead to a diversity of diseases, and they construct more than 60% of current drug targets.[2] Though membrane proteins are thus crucially important, the studies on them have been hindered by their low structural stability in detergents followed by the difficulties in obtaining the three-dimensional structure by X-ray crystallography. In this Letter, we deal with human adenosine A_{2a} receptor ($A_{2a}R$), an important GPCR involved in the control of various physiological activities including regulation of glutamine and dopamine release in the brain.[3] Specific inhibitors of the receptor are in advanced clinical trials for the treatment of Parkinson's disease.[4] The crystallization of $A_{2a}R$ followed by the determination of its structure has been a nontrivial task due to its low structural stability.

A method for improving the stability of a membrane protein in detergents as well as its thermostability is an amino-acid mutation.[5] Alanine (Ala) scanning mutagenesis, in which every residue is mutated to Ala (Ala is mutated to leucine (Leu)), has been applied to $A_{2a}R$, leading to the finding of significantly many stabilizing single mutants.[6] Combining multiple stabilizing single mutations often brings more stability than a stabilizing single mutation. In particular, the thermostability is exceptionally enhanced by an octuple mutation (A54L, T88A, R107A, K122A, L202A, L235A, V239A, and S277A; seven residues are mutated to Ala): The denaturation temperature T_m of the octuple mutant (PDB ID of its structure is 3PWH) is higher than that of the wild type by ~20°C.[7] $A_{2a}R$ comprises the extracellular, transmembrane (TM), and intracellular regions. Five of the residues mutated are totally within the TM region and the other three are in the vicinity of the TM region (see Sec. 3.2.4). Therefore, it is probable that the structural modification occurs primarily within the TM region, leading to the remarkable stabilization. On the other hand, Murata and his coworkers[8] have succeeded in determining the structure of $A_{2a}R$ to which a mouse monoclonal-antibody Fab-fragment is bound (PDB ID: 3VG9). In their method, the success is ascribed to the achievement of crystallization by the fragment binding to $A_{2a}R$. As discussed in Sec. 3.2.4 in more detail, the binding interface is at the end of the intracellular region and far from the TM region. Presumably, the structure of the TM region is not significantly modified by the fragment binding though that of the intracellular region is more or less influenced. It follows that the physical origins of the remarkable stabilization can be investigated by comparing the TM regions of the octuple mutant of $A_{2a}R$ and $A_{2a}R$ with the fragment bound. The latter is referred to as "wild type"

hereafter, because it and the wild type should share almost the same TM-region structure. Both of the octuple mutant and the wild type are in the inactive states. Though the structures of their TM regions look almost indistinguishable in sight, there must be significant differences affecting the thermodynamic quantities governing the stability, which have not yet been understood.

In the present study, using our recently developed theory for the thermostability of a membrane protein,[9] we clarify the physical factors responsible for the remarkable stabilization by the octuple mutation. For the octuple mutant and the wild type, the changes in thermodynamic quantities upon protein folding are analyzed using the structures of their TM regions. The theoretical tool is the integral equation theory (IET)[10] combined with Kinoshita and coworkers' recently developed morphometric approach (MA)[11]: The former is a statistical-mechanical theory for fluids and the latter is introduced for treating a large, complex solute like a protein. The combined method is characterized by the concept that hydrocarbon (CH_2, CH_3, and CH) groups in nonpolar chains of lipid molecules work as "solvent" for a membrane protein and the translational displacement of solvent particles plays a pivotal role in the protein structural stability.[9,12] It is demonstrated that we are capable of explicating the remarkable stabilization: Protein folding is accompanied by a decrease in energy related to the formation of protein intramolecular hydrogen bonds and a gain of solvent entropy, but both of the energy decrease and the entropic gain are substantially larger for the octuple mutant than for the wild type.

3.2. Model and theory

3.2.1. Two essential physical factors

Following our earlier work,[9] we assume that the thermostability of a GPCR is governed by that of the TM region. For it, nonpolar chains of lipid molecules act as "solvent".[9,12] Insertion of a protein into the membrane causes an entropic loss because it reduces translational, rotational, and vibrational freedoms of the nonpolar chains. However, the reduction of translational freedom is the most serious, because it reaches all the nonpolar chains coexisting with the protein, whereas the reduction of rotational and vibrational freedoms occurs only in the close vicinity of the protein surface. Therefore, we take account of only the effect of the translational displacement of hydrocarbon (CH_2, CH_3, and CH) groups which are treated as if they were not connected with one another, so that a tractable statistical-mechanical theory can be applied (more details are described in Sec. 3.2.2). A gain of protein intramolecular van der Waals (vdW) attractive interactions upon protein folding is significantly cancelled out by the loss of protein-solvent vdW attractive interactions accompanied. This is not the case for protein intramolecular hydrogen bonds (IHBs) because there are no solvent-protein hydrogen bonds (HBs).[12] The importance of IHBs is reflected, for instance, on the high helical content of a GPCR. We assume that the wild type and the octuple mutant, which are both quite compact, share the same protein conformational entropy.

We take account of the solvent-entropy effect and the protein intramolecular hydrogen bonding as two essential factors which most influence the structural stability of a membrane protein.[9,12] The validity of accounting for only the above two factors and of employing the simplified solvent model has been confirmed by our success in discriminating the native structure from ~15000 non-native structures generated by a computer simulation for glycophorin A (GpA).[9,12]

3.2.2. Entropic excluded-volume effect by solvent

Hydrocarbon (CH_2, CH_3, and CH) groups in the nonpolar chains are referred to as "solvent particles". Upon protein folding, the excluded volume (EV) (i.e., the volume of the space which the centers of solvent particles cannot enter) decreases to a large extent, which is followed by a corresponding increase in the total volume available to the translational displacement of solvent particles in the system.[13] The effect of the close packing of side chains is particularly important: A cartoon is shown in Fig. 3.1. A notable point is that the presence of a solvent particle also generates an EV for the other solvent particles. In this sense, all of the solvent particles are entropically correlated. This correlation is referred to as "solvent crowding".[14,15] The increase in the total volume available mentioned above reduces the solvent crowding. Primarily through this effect, protein folding leads to a large gain of solvent entropy. The solvent-entropy gain originating from this EV effect is dependent on not only the EV but also the solvent-accessible surface area (SASA) and surface curvatures (see Sec. 3.2.5).

Note that a membrane is immersed in aqueous solution. When the protein structure becomes less compact and generates a larger EV, for instance, the membrane also generates a larger EV for water molecules. Thus, water also acts as the solvent. It has been shown in experiments that many membrane proteins fold and function in nonpolar environments which bear little similarity to the membrane.[16] It follows that correct folding of a membrane protein is realized only if it is immersed in nonpolar environment where the EV effect is present and the intramolecular hydrogen bonding is essential. The details of specific characteristics of the nonpolar chains are not relevant.

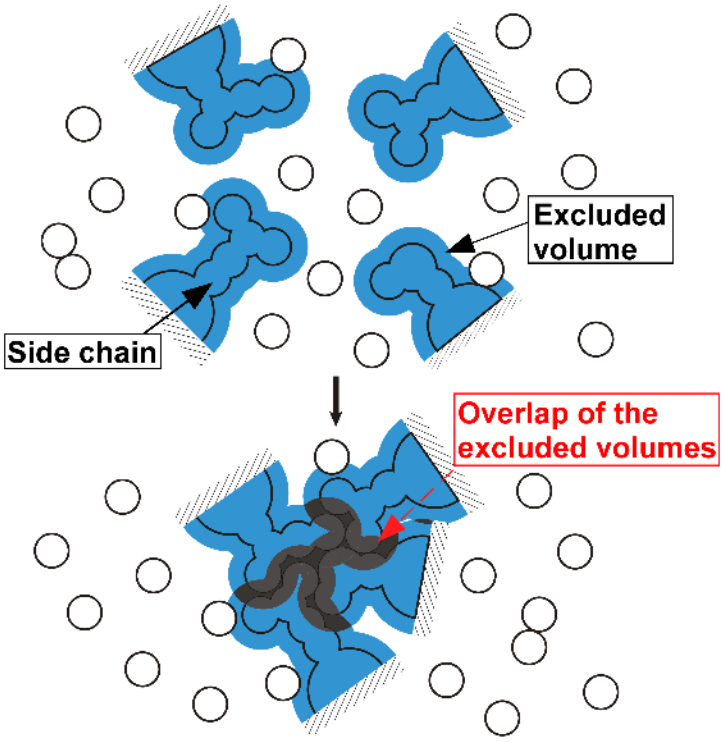


Fig. 3.1. Close packing of side chains upon protein folding in solvent. Overlap of excluded volumes occurs, and the total volume available to the translational displacement of solvent particles increases by the overlapped volume. A sphere corresponds to a solvent particle.

3.2.3. Solvent model

Kinoshita and coworkers have found that water can be modeled as "neutral hard spheres" in a theoretical treatment focused on the entropic EV effect at ambient temperature and pressure.[13-15] On the basis of this finding and the discussion in Sec. 3.2.2, the solvent is modeled as an ensemble of neutral hard spheres whose diameter and packing fraction are set at those of water at 298 K and 1 atm (a membrane protein is immersed in this bulk solvent). This simplified model has been shown to be valid by our success in the native-structure discrimination for GpA[9,12] mentioned in Sec. 3.2.1.

3.2.4. Wild-type and mutant structures

The wild-type[8] and mutant[7] structures are taken from the Protein Data Bank (3VG9 and 3PWH, respectively). The structural information for extracellular loop 2 (P149-H155 in 3VG9 or K150-Q157 in 3PWH) is missing. Hence, the loop structure is merged from that of the other crystal structure of $AA_{2a}R$ (PDB ID: 4EIY)[17] with the help of MODELLER.[18] The LJ potential energy for each structure thus obtained is unreasonably high owing to the overlaps of protein atoms. Such overlaps are removed by local minimization of the energy using the CHARMM biomolecular simulation program[19] through the multiscale modeling tools in structural biology (MMTSB) program.[20] We employ CHARMM22[21] as the force-field parameters. Electrostatic and nonbonded

interactions are all evaluated without any cutoff. After the minimization, each structure is switched to a set of fused hard spheres for calculating the solvation entropy. (The solvation entropy denotes the loss of solvent entropy upon insertion of a solute in the solvent.) All that matters is the removal of the unrealistic overlaps of protein atoms with retaining the original structure, and the details of the minimization procedure are unimportant. The root-mean-square deviation (RMSD) for C_α atoms between the structures before and after the energetic local minimization is only 0.54 Å for the wild type and only 0.44 Å for the octuple mutant.

$A_{2a}R$ possesses three regions: the extracellular, TM, and intracellular regions. The TM region of the wild-type structure is determined using the web server of TMDET.[22] It is then assumed that the wild-type and mutant structures share the same TM region (but their structures are different). In the wild type employed, an antibody fragment is bound to the intracellular region. The intracellular region is much larger than the extracellular region (see Fig. 3.2), and the binding interface is far from the TM region. The structure of the intracellular region is somewhat influenced by the binding, but the other two regions should undergo no significant structural changes. The wild-type and mutant structures are drawn in Fig. 3.3. The TM region comprises seven α-helices, and mutations are made for the eight residues within these α-helices. Five of the residues mutated are totally within the TM region. The other three are within the intracellular region but very close to the TM region. Hence, the structure within the TM region should be modified to a significant extent by the octuple mutation.

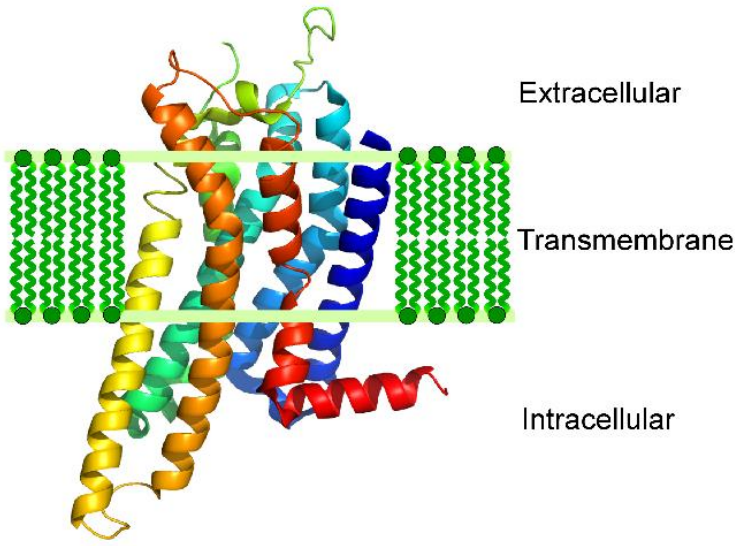


Fig. 3.2. Illustration of the extracellular, transmembrane (TM), and intracellular regions of the adenosine A_{2a} receptor ($A_{2a}R$).

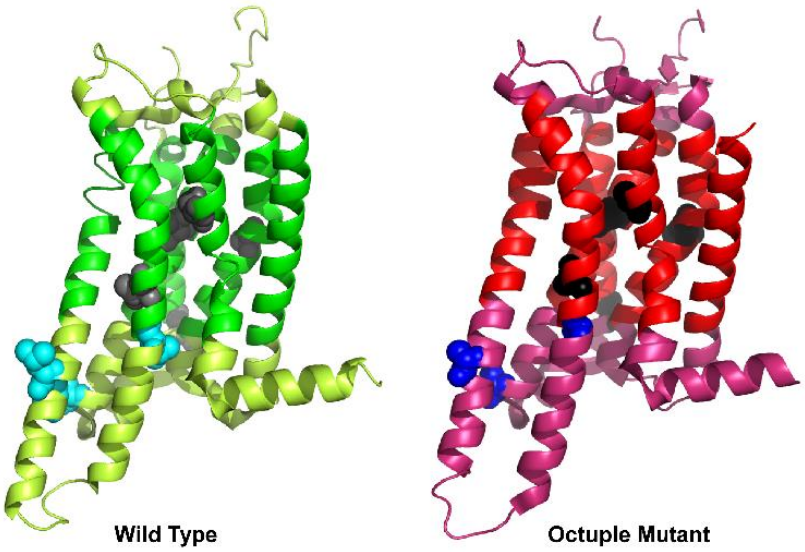


Fig. 3.3. Wild-type (left) and octuple mutant (right) structures of the adenosine A_{2a} receptor ($A_{2a}R$). For the wild type, the transmembrane (TM) region is drawn in green and the extracellular and intracellular regions are drawn in yellowish green. For the octuple mutant, the TM region is drawn in red and the extracellular and intracellular regions are drawn in pink. In the wild type, the residues before the mutation within the TM region are indicated by the dark-gray spheres and those within the intracellular region are indicated by the light-blue spheres. In the mutant, the residues after the mutation within the TM region are indicated by the black spheres and those within the intracellular region are indicated by the blue spheres.

3.2.5. Calculation of solvent-entropy gain upon protein folding

Folding of a membrane protein can be described by the two-stage model[23]: Individual α-helices of the protein are separately stabilized in the first stage; and folding is completed by the side-to-side association of these α-helices accompanying the packing of side chains in the second stage. The total gain of solvent entropy upon protein folding ΔS ($\Delta S>0$) is governed by the solvent-entropy gain in the second stage, as shown in our earlier works.[9,12] Hence, ΔS is calculated by focusing on the side-chain packing within the TM region. Using the structural data for the wild type and the octuple mutant, α-helices are simply separated, and the solvent-entropy gain upon their association is calculated by combining the IET[10] and the MA.[11]

With the MA, ΔS is decomposed into four terms as

$$\Delta S/k_B=C_1\Delta V+C_2\Delta A+C_3\Delta X+C_4\Delta Y \quad (3.1)$$

where k_B is the Boltzmann constant, V the EV, A the SASA, X and Y the integrated mean and Gaussian curvatures, respectively, and Δ denotes the change upon protein folding. Since S is rather

insensitive to the protein-solvent interaction potential,[14] a protein is modeled as a set of fused hard spheres retaining its complex polyatomic structure.

The four coefficients (C_1–C_4), which are independent of geometric characteristics of the solute, are calculated from the solvation entropies of isolated hard-sphere solutes with various diameters immersed in the model solvent. The calculation is made using the IET. Once they are determined, *S* of a protein with a prescribed structure is obtained simply by calculating its *V*, *A*, *X*, and *Y* from the x-y-z coordinates of the protein atoms and the diameter of each atom set at the sigma-value of the Lennard-Jones (LJ) potential parameters which are taken from CHARMM22.[21] Refer to our earlier publications[9,11,12,14,15] for more details.

The values of the four coefficients are as follows: C_1=−0.26966 Å^{-3}, C_2=0.21418 Å^{-2}, C_3=−0.18719 Å^{-1}, and C_4=0.05103. The contributions from $C_1\Delta V$ and $C_2\Delta A$ are much larger than those from the other two terms. It is straightforward that C_1 is negative and a decrease in the EV leads to higher solvent entropy. Positive C_2 may be counterintuitive but can be interpreted as explained in Sec. 3.2.6. [14,15]

3.2.6. Physical meaning of positive C_2

Positive C_2 implies that an increase in the SASA leads to higher solvent entropy. As described in Sec. 3.2.2, the solvent crowding is serious in the system. When solvent particles come in contact with the protein surface, the EVs generated by them and by the protein overlap. As a consequence, the total volume available to the other solvent particles increases by the overlapped volume. The increase acts for reducing the solvent crowding. The entropic gain arising from this reduction is larger than the entropic loss caused by the contact of solvent particles mentioned above. An increase in the SASA results in a larger number of solvent particles in contact with the protein surface, reducing the solvent crowding to a larger extent.[14,15]

3.2.7. Protein intramolecular hydrogen bonding

The formation of IHBs is crucial in both of the first and second stages.[9,12] What matters for a thermodynamic quantity is not its absolute value but its change upon a mutation. We assume that the unfolded states of the wild type and its mutant possess essentially no IHBs. We examine all the donors and acceptors for backbone-backbone, backbone-side chain, and side chain-side chain IHBs of the folded structure within the TM. The examination is made using the criteria proposed by McDonald and Thornton[24] with the modification that the maximum distances between H and the acceptor and between the donor and the acceptor are 3.0 Å and 4.4 Å, respectively. When an IHB is formed, an energy decrease of D is considered.[12] D is dependent on the distance between centers of H and the acceptor, d. For $d \leq 1.5$Å, D is set at $-E$. As d increases from 1.5Å, |D| decreases lineally and becomes zero for $d \geq 3.0$Å. (In the unfolded states of the wild type and its mutant, *d* is assumed to be larger than 3.0Å for all of the H-acceptor pairs.) An H-acceptor pair is not in vacuum but in the environment where atoms with positive and negative partial charges are present. Therefore, $|D|$ does not necessarily decrease in proportion to $1/d$.

The value of E is not definite. The value calculated using quantum chemistry for the formation of an IHB in gas phase is E=6.0 kcal/mol.[25] The energy lowering brought by the formation of an IHB between two formamide molecules in a nonpolar solvent is $-E$=−8.3 kcal/mol[26] which implicitly includes the contribution from entropic gain of the nonpolar solvent upon the formation. As stated above, however, an H-acceptor pair is not in vacuum but in the environment where atoms with positive and negative partial charges are present. Therefore, E can be regarded as the potential of mean force between the pair in such environment and should be much smaller than these values: It has been suggested that E be ~2.0 kcal/mol.[27]

3.3. Results and discussion

3.3.1. Solvent-entropy gain upon protein folding

The important quantities for each of the wild type and the octuple mutant are the solvation entropy S, V (EV), and A (SASA) of separate α-helices and those of associated α-helices within the TM. We calculate the changes in these quantities attributed to the side-to-side association of α-helices, ΔS, ΔV, and ΔA. The association always leads to decreases in both of V and A. The decrease in V makes the solvent entropy smaller while that in A makes it larger. However, the former effect dominates, with the result of an increase in solvent entropy and positive ΔS. Hereafter, the subscripts "W" and "M" denote values of the wild type and the mutant, respectively. $\Delta\Delta S=\Delta S_M-\Delta S_W$, which is given in Table 3.1, takes a significantly large, positive value ($\Delta\Delta S/k_B$=22.1). The stabilization upon protein folding, which is represented by the solvent-entropy gain originating from the side-to-side association of α-helices, is stronger for the mutant by ~–13.1 kcal/mol in terms of the free energy.

Table 3.1. $\Delta\Delta S=\Delta S_M-\Delta S_W$, $\Delta\Delta V=\Delta V_M-\Delta V_W$, and $\Delta\Delta A=\Delta A_M-\Delta A_W$ calculated. Here, ΔS, ΔV, and ΔA are changes in the solvent entropy, excluded volume, and solvent-accessible surface area upon protein folding, respectively, and the subscripts "W" and "M" denote values of the wild type and the octuple mutant, respectively. $\Delta S_M/k_B$=350.14, $\Delta S_W/k_B$=328.07, ΔV_M=–6025.14 Å^3, ΔV_W=–6009.67 Å^3, ΔA_M=–7074.01 Å^2, and ΔA_W=–7143.59 Å^2.

$\Delta\Delta S/k_B$	$\Delta\Delta V$ (Å^3)	$\Delta\Delta A$ (Å^2)
22.1	–15.5	69.6

3.3.2. Changes in excluded volume and solvent-accessible surface area upon protein folding

$\Delta\Delta V=\Delta V_M-\Delta V_W$ and $\Delta\Delta A=\Delta A_M-\Delta A_W$ are given in Table 1. $\Delta\Delta V$ and $\Delta\Delta A$ are negative and positive, respectively: $|\Delta V_M|>|\Delta V_W|$ and $|\Delta A_M|<|\Delta A_W|$. This result indicates that the mutant is more favorable in both of ΔV and ΔA. In general, a structure with a smaller EV may possess a smaller SASA as well: One might think that when $|\Delta V|$ is larger, $|\Delta A|$ is also larger. However, for a complex solute molecule with polyatomic structure like a protein, larger $|\Delta V|$ together with smaller $|\Delta A|$ is possible, and this is fortuitously realized by the octuple mutation. Interestingly, $\Delta\Delta A$ makes a larger contribution to $\Delta\Delta S$ than $\Delta\Delta V$. The reduction of solvent crowding is primarily responsible for the large, positive value of $\Delta\Delta S$.

We note that $V_M<V_W$ and $A_M<A_W$. This is reasonable because the larger residues are mutated to the smaller Ala residues except the mutation Ala→Leu. However, $|\Delta V_M|>|\Delta V_W|$.

3.3.3. Energy decrease upon protein folding

We find that four more IHBs are formed in the octuple mutant when we employ the criteria proposed by McDonald and Thornton[24] for the formation of an IHB. By the octuple mutation, eight of the IHBs in the wild type are broken but twelve IHBs are newly formed. It is interesting that the seven mutations to Ala and the mutation Ala→Leu lead to such an increase though the side chains of Ala and Leu cannot participate in the formation of any IHB. This result is suggestive that a mutation can modify the whole structure of the TM region of a protein and the change in the number of IHBs cannot readily be deduced. The energy change arising from the octuple mutation, $\Delta\Delta U$, which comes from the hydrogen bonding energy is $-5.48E$. This is −32.9 kcal/mol for E=6.0 kcal/mol, −45.5 kcal/mol for E=8.3 kcal/mol, and −11.0 kcal/mol for E=2.0 kcal/mol.

3.3.4. Free-energy change upon protein folding

The result of our calculation is as follows: $-T\Delta\Delta S=-13.1$ kcal/mol (T is the absolute temperature) and $\Delta\Delta U=-5.48E$ where $-E$ is the energy decrease arising from the formation of an IHB within the TM region. $\Delta\Delta U$ is estimated to be in the range from −11.0 to −45.5 kcal/mol. It follows that $40.7\le-\Delta\Delta F/(k_BT)=-(\Delta\Delta U-T\Delta\Delta S)/(k_BT)\le98.9$ (T=298 K) where ΔF is the free-energy change upon protein folding and $\Delta\Delta F$ is the change in ΔF upon a mutation.

In Chapter 2, we found four single mutations (S91K, T88E, S91R, and T88K) and a double mutation (T88E and S91R) for $A_{2a}R$ leading to significantly higher thermostability. Table 3.2 compares the values of $-\Delta\Delta F$ and ΔT_m (the change in T_m upon a mutation) for these mutations and the octuple mutation. It is observed that $-\Delta\Delta F$ and ΔT_m for the double and octuple mutations are considerably larger than those for the single mutations. In particular, those for the octuple mutation are the largest. For the single mutations, the order of $-\Delta\Delta F$ is not necessarily consistent with that of ΔT_m, but this is ascribed to the approximate treatments employed in our method. (Unfortunately, there are no experimental data for $-T\Delta\Delta S$, $\Delta\Delta U$, and $\Delta\Delta F$.)

The effect of the protein conformational entropy is neglected in our method. In general, if $\Delta\Delta F$

of a mutant is negative, the conformational-entropy loss of the mutant is larger than that of the wild type because the folded state of the mutant is more compact. Therefore, if the conformational-entropy effect was incorporated (i.e., the structural fluctuation was taken into account), $-\Delta\Delta F$ would become considerably smaller. Nonetheless, our method is quite useful in the quantification of protein thermostability with minor computational effort.

Table 3.2. Comparison among four single mutations (S91K, T88E, S91R, and T88K), a double mutation (T88E and S91R), and the octuple mutation (A54L, T88A, R107A, K122A, L202A, L235A, V239A, and S277A) for $A_{2a}R$ in terms of $-\Delta\Delta F$ and ΔT_m. $\Delta\Delta F$ is in the range, $\text{Min} \le -\Delta\Delta F/(k_B T) \le \text{Max}$ (T=298 K). T_m is the denaturation temperature and ΔT_m is the change in T_m upon a mutation.

Mutation	ΔT_m (°C)	Min	Max
Octuple	~20	40.7	98.9
T88E and S91R	~12	39.2	53.1
S91K	~7	18.2	34.7
T88E	~7	8.90	13.6
S91R	~5	24.2	33.7
T88K	~4	18.4	43.6

3.4. Conclusion

The studies on membrane proteins are far behind those on water-soluble proteins, primarily due to their low structural stability. Even the physicochemical factors governing the stability of a membrane protein are rather ambiguous and controversial. As part of research to circumvent this problematic situation, we have investigated the physical origins of the remarkable enhancement of structural stability brought by an octuple mutation for the adenosine A_{2a} receptor ($A_{2a}R$). It is assumed that the enhancement originates from the structural modification of the TM region. We look at the following quantities occurring when the protein folds: (i) the changes in the excluded volume (EV: volume of the space which the centers of solvent particles cannot enter), solvent-accessible surface area (SASA), and integrated mean and Gaussian curvatures of the protein structure; (ii) resultant solvent-entropy gain ΔS; and (iii) energy decrease ΔU arising from the

formation of protein intramolecular hydrogen bonds (IHBs). The changes in ΔS and ΔU upon the mutation, which are denoted by $\Delta\Delta S$ and $\Delta\Delta U$, respectively, are two essential parameters in our theory.

In Chapter 2, we found four single mutations and a double mutation for $A_{2a}R$ leading to significantly higher thermostability. In this Letter, it has been shown that $-\Delta\Delta F$ and ΔT_m for the octuple mutations are much larger than those for the four single mutations and significantly larger than those for the double mutation. We have thus succeeded in showing that the protein is made much more stable by the octuple mutation. Protein folding is accompanied by a decrease in energy related to the formation of protein intramolecular hydrogen bonds and a gain of solvent entropy, but both of the energy decrease and the entropic gain are substantially larger for the octuple mutant than for the wild type. By the mutation, the decreases in EV and SASA upon protein folding are made larger and smaller, respectively, leading to enhanced reduction of the solvent crowding followed by larger gain of the solvent entropy. An increase in the number of IHBs within the membrane is responsible for the energy decrease. In a future study, the value of E is to be optimized.

It is an urgent task to develop a reliable theoretical or computational method for indentifying thermostabilizing mutations of membrane proteins. A nice review[28] has recently been written concerning this development. In the previously reported methods, the solvent-entropy effect we emphasize is not taken into consideration, and the molecular dynamics (MD) simulation is regarded as the most useful tool. In our opinion, however, the MD simulation is not suited to the calculation of thermodynamic quantities such as the solvation entropy and the quantification of protein thermostability, owing to its heavy burden. Our method based on statistical thermodynamics is expected to become a powerful alternative to the MD simulation.

References

[1]Boyd, D.; Schierle, C.; Beckwith, J. How many membrane proteins are there? *Protein Sci.* **1998**, *7*, 201−205.

[2]Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabási, A. L.; Vidal, M. Drug-target network. *Nat. Biotechnol.* **2007**, *25*, 1119−1126.

[3]Fredholm, B. B.; Chen, J. F.; Masino, S. A.; Vaugeois, J. M. *Annu. Rev. Pharmacol. Toxicol*. 2005, *45*, 385−412.

[4]Müller, C. E.; Jacobson, K. A. *Biochim. Biophys. Acta.* **2011**, *1808*, 1290−1308.

[5]Lau, F. W.; Nauli, S.; Zhou, Y.; Bowie, J. U. Changing single side-chains can greatly enhance the resistance of a membrane protein to irreversible inactivation. *J. Mol. Biol.* **1999**, *290*, 559−564.

[6]Magnani, F.; Shibata, Y.; Serrano-Vega, M. J.; Tate, C. G. Co-evolving stability and conformational homogeneity of the human adenosine A_{2a} receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10744−10749.

[7]Doré, A. S.; Robertson, N.; Errey, J. C.; Ng, I.; Hollenstein, K.; Tehan, B.; Hurrell, E.; Bennett, K.; Congreve, M.; Magnani, F.; Tate, C. G.; Weir, M.; Marshall, F. H. Structure of the adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **2011**, *19*, 1283−1293.

[8]Hino, T.; Arakawa, T.; Iwanari, H.; Yurugi-Kobayashi, T.; Ikeda-Suno, C.; Nakada-Nakura, Y.; Kusano-Arai, O.; Weyand, S.; Shimamura, T.; Nomura, N.; Cameron, A. D.; Kobayashi, T.; Hamakubo, T.; Iwata, S.; Murata, T. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* **2012**, *482*, 237−240.

[9]Yasuda, S.; Kajiwara, Y.; Takamuku, Y.; Suzuki, N.; Murata, T.; Kinoshita, M. Identification of thermostabilizing mutations for membrane proteins: Rapid method based on statistical thermodynamics. *J. Phys. Chem. B* **2016**, *120*, 3833−3843.

[10]Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids, 3rd ed.* Academic Press; London, U.K., 2006.

[11]Roth, R.; Harano, Y.; Kinoshita, M. Morphometric approach to the solvation free energy of complex molecules. *Phys. Rev. Lett.* **2006**, 97, 078101.

[12]Yasuda, S.; Oshima, H.; Kinoshita, M. Structural stability of proteins in aqueous and nonpolar environments. *J. Chem. Phys.* **2012**, *137*, 135103.

[13]Harano, Y.; Kinoshita, M. Translational-entropy gain of solvent upon protein folding. *Biophys. J.* **2005**, *89*, 2701−2710.

[14]Kinoshita, M. A new theoretical approach to biological self-assembly. *Biophys. Rev.* **2013**, *5*, 283−293.

[15]Oshima, H.; Kinoshita M. Essential roles of protein-solvent many-body correlation in solvent-entropy effect on protein folding and denaturation: comparison between hard-sphere solvent and water. *J. Chem. Phys.* **2015**, *142*, 145103.

[16]Popot, J. L.; Engelman D. M. Membranes do not tell proteins how to fold. *Biochemistry* **2016**, 55, 5−18.

[17]Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; Ijzerman, A. P.; et al. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **2012**, *337*, 232−236.

[18]Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.

[19]Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

[20]Feig, M.; Karanicolas, J.; Brooks III, C. L. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **2004**, *22*, 377−395.

[21]MacKerell, Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr. R.L.; Evanseck, J.D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

[22]Tusnády, G. E.; Dosztányi, Z.; Simon, I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics* **2005**, *21*, 1276−1277.

[23]Popot, J. L.; Engelman D. M. Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* **2000**, *69*, 881−922.

[24]McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*, 777−793.

[25]Mitchell, J. B. O.; Price, S. L. *Chem. Phys. Lett.* On the relative strengths of amide…amide and amide…water hydrogen bonds. **1991**, *180*, 517−523.

[26]Sneddon, S. F.; Tobias, D. J.; Brooks C. L. III. Thermodynamics of amide hydrogen bond formation in polar and apolar solvents. *J. Mol. Biol.* **1989**, *209*, 817−820.

[27]Bowie, J. U. Membrane protein folding: how important are hydrogen bonds? *Curr. Opin. Struct. Biol.* **2011**, *21*, 42−49.

[28]Vaidehi, N.; Grisshammer, R.; Tate, C. G. *Trends Pharmacol. Sci*. **2016**, *37*, 37−46.

Chapter 4

Identification of Thermostabilizing Mutations for Membrane Proteins Whose Three-Dimensional Structure Is Unknown

4.1. Introduction

Membrane proteins are indispensable to the sustenance of life. For instance, G-protein coupled receptors (GPCRs) play imperative roles in the cell signaling, and their malfunctioning causes a diversity of diseases.[1,2] They form a considerably large portion of current drug targets.[3] For determining the three-dimensional (3D) structure by X-ray crystallography, membrane proteins are removed from the lipid bilayer and solubilized in detergents. However, this treatment gives rise to the collapse of their structure, making the structure determination infeasible. Their low structural stability has thus been a major stumbling block in the structure-guided drug design. An amino-acid mutation is known to possibly enhance the thermostability of a membrane protein.[4–6] The enhancement usually leads to higher structural stability in detergents as well. It is unfortunate that thermostabilizing mutations are currently identified by experiments testing a number of trial mutants. Moreover, alanine (Ala) scanning mutagenesis, in which every residue is mutated only to Ala (Ala is mutated to leucine), has been prevailing in the experiments.[7–9] It is strongly desired that a theoretical method be developed, but the prediction of thermostabilizing mutations is one of the most subtle subjects.

An important point is that all of the possible mutations (not limited to Ala scanning mutagenesis) are to be examined. We recently developed a theoretical method for identifying thermostabilizing mutations of a membrane protein.[10] The method possesses the following three features: The importance of translational entropy of hydrocarbon groups within the lipid bilayer is emphasized; all of the possible mutations can rapidly be examined; and when a particular mutation has turned out be thermostabilizing, physical origins of the stabilization can be made clear because the method is a physics-based one. None of the previously reported methods[11–13] without relying on experiments possessed all of the three features. In our method,[10] the free-energy function (hereafter, this is denoted by F) is employed, and it comprises entropic and energetic terms. The translational-entropy effect is incorporated in the entropic term S. Hereafter, ΔX denotes the change in X upon protein folding and $\Delta\Delta X$ does the change in ΔX upon mutation. ΔS, for example, represents the gain of translational entropy of hydrocarbon groups upon protein folding. Our basic procedure is as follows.[10] First, key residues are nominated using $\Delta\Delta S$. Here, the key residues are those to be mutated in the sense that many of their mutations will lead to relatively higher enhancement of the stability. (We showed for a water-soluble protein that a mutation which largely increases the water-entropy gain upon protein folding often leads to very high enhancement of the

thermal stability irrespective of the enthalpic factor.[14,15] The first step is based on this finding.) Second, select stabilizing mutations of the key residues using $\Delta\Delta F$. In the second step, only the mutants which are predicted to be sufficiently more stable than the wild type are selected.

The ratio of the number of stabilizing mutations to that of all of the possible mutations for a membrane protein is quite low. In general, therefore, an arbitrarily chosen mutation is thermostabilizing only with unacceptably low probability. In Chapter 2 mentioned above, the high performance of the theoretical method was demonstrated for the adenosine A_{2a} receptor ($A_{2a}R$), a member of the GPCR family. A total of 10 mutations including a double mutation predicted to be stabilizing or destabilizing were checked by referring to the experimental results. The success rate, which depends on the criterion employed for the thermostability relative to that of the wild type in the theoretical prediction and on the thermostability measure adopted in our experiments, was in the range from 8/10 to 10/10: It is remarkably high. However, there was one shortcoming in the demonstration: We utilized the experimentally determined 3D structure (hereafter, this structure is referred to as "crystal structure") of $A_{2a}R$ in an inactive state. As described above, from a practical point of view, a theoretical prediction must be made under the condition where the 3D structure is not available.

In this work, we again consider $A_{2a}R$ but postulate that its 3D structure is unknown. A number of candidate models are constructed for the 3D structure using the homology modeling. A good measure of the appropriateness of a model is the root-mean-square deviation (RMSD) between this model and the crystal structure. However, RMSD cannot be calculated in a practical situation. We therefore select the model whose ΔF is the lowest (this model is referred to as "the best model") and follow the basic procedure explained above. We compare the prediction results obtained using the crystal structure, the best model, and six more representative models. The findings are as follows: ΔF is a very good measure of the appropriateness of a 3D-structure model; ΔF is better than RMSD as the measure; even when the crystal structure is replaced by the best model, the key residues, T88 and S91, are successfully nominated and the prediction performance for them is only slightly lowered.

4.2. Model and theory

It is worthwhile to recall that we examine all of the possible mutations. Nevertheless, both of the following two contradicting requirements must be met: The identification of thermostabilizing mutations is to be made with moderate computational effort; and the performance needs to be sufficiently high. Hence, we must be very careful in constructing the theoretical method: It is desired that the method be as simple as possible but still captures the essential physics. The details of our method for identifying thermostabilizing mutations are provided in Chapter 2 and need not be repeated in this article. In what follows, however, some of its important points are recapitulated.

4.2.1. Essential physical factors governing the thermostability of a membrane protein

Just like water for a water-soluble protein, the hydrocarbon (CH_2, CH_3, and CH) groups constituting nonpolar chains of lipid molecules act as "solvent" for a membrane protein. The factors affecting the thermostability of the membrane protein are the protein intramolecular Lennard-Jones (LJ) and electrostatic energies, solvent-solvent and protein-solvent LJ and electrostatic interaction energies, protein intramolecular entropy, and solvent entropy. A donor or an acceptor for a hydrogen bond (HB) carries a negative partial charge whereas a hydrogen atom does a positive one, and the hydrogen bonding is a primary component of the intramolecular electrostatic energy. For testing all of the possible mutations with minor computational effort, however, it is not advantageous to account for all of these factors. It is required that only the fewest number of factors be taken into consideration.

We showed for a water-soluble protein that the entropic effect originating from the translational displacement of solvent (i.e., water) molecules plays a pivotal role in the structural stability.[16–21] This should also be true for a membrane protein. When a protein folds, the excluded volume (EV) (i.e., the volume of the space which the centers of solvent molecules cannot enter) decreases to a large extent, leading to a corresponding increase in the total volume available to the translational displacement of solvent molecules coexisting with the protein in the system. This increase is accompanied by a large gain of solvent entropy.[16–21] Close packing of side chains makes a principal contribution to the solvent-entropy gain:[18] A simple illustration is given in Figure 4.1(a). The solvent possesses the orientational (rotational) and vibrational entropies as well as the translational entropy (TE), but we showed for the case of water that the TE contribution dominates.[17,22] The reason for this is as follows: Upon protein folding, the increase in orientational and vibrational freedoms occurs only for the solvent molecules in the close vicinity of the protein, whereas that of translational freedom reaches all of the solvent molecules in the system. The dominance of the TE contribution should also be applicable to the case of nonpolar chains of lipid molecules.

In water, the formation of intramolecular hydrogen bonds (IHBs) is accompanied by the break of HBs with water molecules. This is not the case in nonpolar chains of lipid molecules, which makes the formation of IHBs even more influential.[19] A gain of intramolecular van der Waals (vdW) attractive interactions upon protein folding is somewhat cancelled out by the loss of protein-solvent vdW attractive interactions unavoidably accompanied. When the wild-type and mutant structures are compared in terms of the stability, the effect of protein intramolecular entropy may be relatively smaller since the structures are both quite compact.

On the basis of the above argument, we incorporate only the two factors, the energy decrease arising from the formation of protein IHBs and the solvent-entropy gain upon protein folding, in our free-energy function F.[10] F combined with the simplified models for calculating the two factors described below was justified in our earlier work:[10,19] We succeeded for a membrane protein GpA in

showing that the native structure can be discriminated from ~15000 non-native structures generated by a computer simulation because *F* takes the lowest value for the native structure.

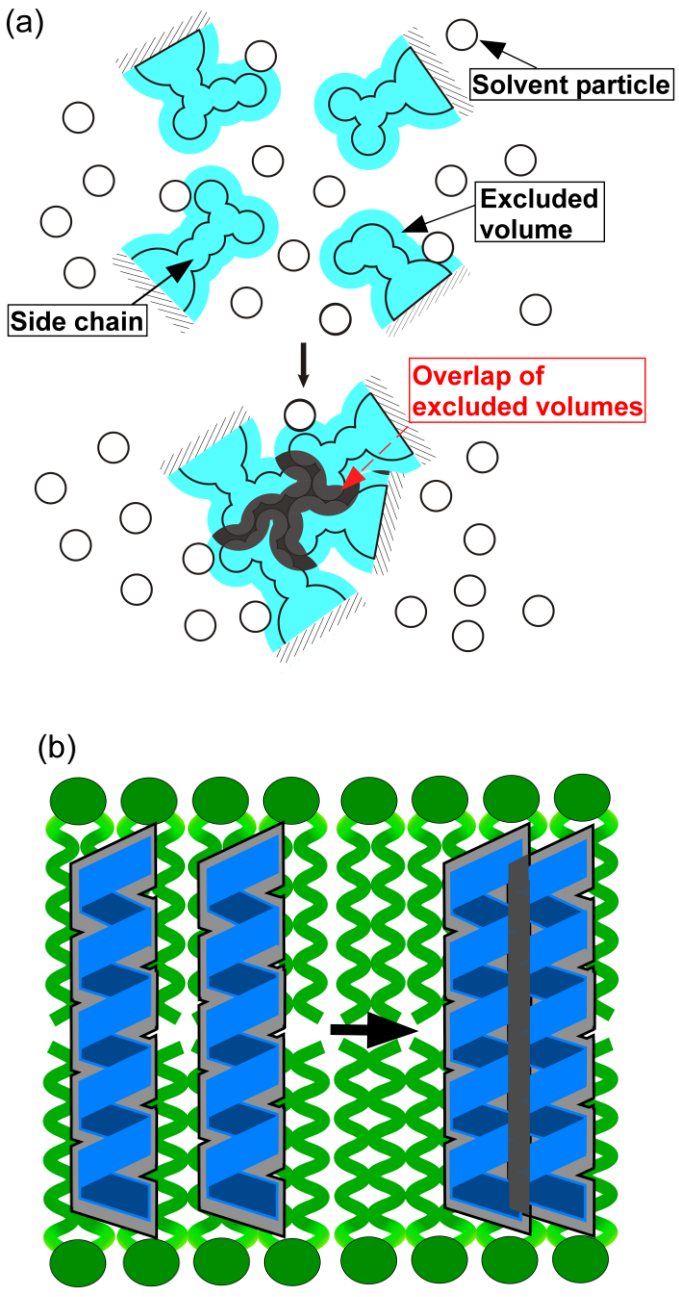


Figure 4.1. (a) Close packing of side chains. It is accompanied by the overlap of excluded volumes. The total volume available to the translational displacement of solvent particles increases by the overlapped volume, leading to an increase in solvent entropy. A sphere corresponds to a solvent particle. (b) Association of α-helices of a membrane protein. Close packing of side chains occurs within the lipid bilayer.

4.2.2. Free-energy function

Our free-energy function is expressed by

$$F/(k_{\mathrm{B}}T_0) = \Lambda/(k_{\mathrm{B}}T_0) - TS/(k_{\mathrm{B}}T_0),\ T_0 = 298\ \mathrm{K} \tag{4.1}$$

where k_{B} is the Boltzmann constant, T is the absolute temperature, and T is set at T_0.[10] Λ ($\Lambda<0$) represents the protein intramolecular energy of the formation of IHBs. $|\Lambda|$ becomes larger as the number of IHBs increases. $-S$ ($-S>0$) signifies the magnitude of entropic loss of the hydrocarbon groups constituting nonpolar chains of the lipid bilayer (i.e., the solvent for a membrane protein) upon protein insertion. Higher efficiency of the backbone and side-chain packing leads to a smaller value of $-S$.

For a water-soluble protein, Kinoshita and coworkers developed a similar free-energy function accounting for only the effects of water entropy and HBs.[23,24] In this case, protein-water HBs as well as protein IHBs must be considered. The function was tested for a total of 133 proteins. It is capable of discriminating the native fold from a number of misfolded decoys with almost 100% accuracy, which indicates that it is far superior to any of the previously reported functions.

To calculate the entropic term, the solvent must be treated not as a continuum but as an ensemble of particles with finite sizes. The calculation of entropy is usually more difficult than that of energy. In particular, the calculation of the solvation entropy of a protein is a formidable task. Kinoshita and coworkers overcame this difficulty using a hybrid of an integral equation theory (IET) and Kinoshita and coworkers' morphometric approach (MA).[25,26] The former is a statistical-mechanical theory for fluids[27] and the latter is necessitated to treat a large protein with complex polyatomic structure. Using this hybrid, we can finish the calculation of $-S$ in ~0.5 sec per structure on our workstation with Xeon E5-2695 2.3-GHz processor. Detailed descriptions of the hybrid are provided in Kinoshita and coworkers' earlier publications.[17,21,26]

The change in F upon protein folding, ΔF (= "F of the folded state" − "F of the unfolded state"), is written as

$$\Delta F/(k_{\mathrm{B}}T_0) = \Delta\Lambda/(k_{\mathrm{B}}T_0) - \Delta S/k_{\mathrm{B}} \tag{4.2}$$

Here, $\Delta\Lambda$ and ΔS are the changes in Λ and S upon protein folding, respectively.

4.2.3. Calculation of the entropic term ΔS

We note that a membrane is immersed in water. When a membrane protein takes a structure with larger EV, the membrane also generates larger EV for water molecules. Thus, water indirectly acts as the solvent. For this reason, we should rather take the view that the membrane protein is immersed in bulk solvent. According to the results of Kinoshita and coworkers' earlier works,[20,21]

the qualitative aspects of the solvent-entropy effect at ambient temperature and pressure can be elucidated by neglecting the solvent-solvent attractive interaction. For the membrane, we take account of only the effect of the translational displacement of hydrocarbon (CH_2, CH_3, and CH) groups which are treated as if they were not connected with one another, so that a tractable statistical-mechanical theory can be applied. Taken together, we employ a simplified model for the solvent: an ensemble of neutral hard spheres whose diameter and packing fraction are set at those of water at 298 K and 1 atm. The validity of this model was proved in discriminating the native structure from ~15000 non-native structures for GpA.[10,19]

It is convenient to consider protein folding in terms of the two-stage model:[28] Individual α-helices of a membrane protein are separately stabilized as constituent domains within the lipid bilayer in the first stage, and the native structure is completed by the side-to-side association of these helices (i.e., association of the transmembrane (TM) portions accompanying the packing of side chains) in the second stage. The solvent-entropy gain occurring in the second stage is considerably larger than that in the first stage.[10,18] Hence, ΔS is calculated by focusing on this side-chain packing (see Figure 4.1(b)): ΔS originates from the packing of separated α-helices and can be written as

$$-\Delta S=-S_{\mathrm{compact}}-\Sigma(-S_{\mathrm{each}}) \tag{4.3}$$

$-S_{\mathrm{compact}}$ is the magnitude of entropic loss upon insertion of the associated α-helices into the solvent, and $\Sigma(-S_{\mathrm{each}})$ is that upon insertion of the separated α-helices into the solvent. $-S_{\mathrm{compact}}$ is smaller than $\Sigma(-S_{\mathrm{each}})$ with the result of $\Delta S>0$. This method for calculating ΔS was validated in our earlier works.[10,18] ΔS is calculated in ~1 sec on our workstation with Xeon E5-2695 2.3-GHz processor. The TM portions are determined by the web server of TMDET.[29]

4.2.4. Calculation of the energetic term $\Delta\Lambda$

We employ the simplest possible method which still captures the physical essence, because $\Delta\Lambda$ as well as ΔS must be calculated with sufficiently high speed. The formation of IHBs is crucial in both of the first and second stages.[10] $\Delta\Lambda$ is calculated simply by counting the number of IHBs in the folded structure. We examine all of the donors and acceptors for backbone-backbone, backbone-side chain, and side chain-side chain IHBs of the folded structure within the TM region. The examination is made using the criteria proposed by McDonald and Thornton[30] with the modification that the maximum distances between H and the acceptor and between the donor and the acceptor are 3.0 Å and 4.4 Å, respectively. An energy decrease of D occurs when an IHB is formed. D is dependent on the distance between centers of the H and the acceptor d. D is set at $-E$ for $d\leq 1.5$Å. As d increases from 1.5Å, $|D|$ decreases lineally and becomes zero for $d\geq 3.0$Å. (We note that $|D|$ does not necessarily decrease in proportion to $1/d$ since the IHB is not in vacuum.)

Apparently, it is assumed for the folded state that d is larger than 3.0Å for all of the H-acceptor pairs and there are essentially no IHBs.[10] $\Delta\Lambda$ ($\Delta\Lambda<0$) arising from the formation of IHBs is written as

$$\Delta\Lambda=\Lambda_{compact}-\Sigma(\Lambda_{extended}) \quad (4.4)$$

where $\Lambda_{compact}$ is the energy decrease for the associated α-helices, and $\Sigma(\Lambda_{extended})$ is that for extended structures with no IHBs: $\Sigma(\Lambda_{extended})=0$.

Following our earlier works,[10,19] E is set at $14k_BT_0$ that is the energy lowering brought by the formation of a HB between two formamide molecules in a nonpolar solvent.[31] However, this parameter setting for E is somewhat controversial.[32,33] The value of E calculated using quantum chemistry for the formation of a HB in gas phase is ~$10k_BT_0$.[34] Setting E at a larger value (e.g., $14k_BT_0$) could account for the effect of the intramolecular electrostatic energy which does not pertain to the hydrogen bonding. This is because protein folding results in a lower value of the intramolecular electrostatic energy mentioned above as well as an increase in the number of IHBs. On the other hand, an H-acceptor pair is not in vacuum but in the environment where atoms with positive and negative partial charges are present. Therefore, E could be regarded as the potential of mean force between the pair in such environment, which leads to another view that E is significantly smaller than $10k_BT_0$. We leave the optimization of E to a future work. The problem of uncertainty of E is revisited in "Comparison between experimental result and theoretical prediction".

4.2.5. Prediction of stability change on mutation

We calculate the values of ΔS for the wild-type structure and a mutant structure, yielding $\Delta\Delta S$ defined as

$$-\Delta\Delta S=-\Delta S_M-(-\Delta S_W) \quad (4.5)$$

Here, the subscripts "W" and "M" denote "wild type" and "mutant", respectively. $\Delta\Delta\Lambda$ defined by

$$\Delta\Delta\Lambda=\Delta\Lambda_M-\Delta\Lambda_W \quad (4.6)$$

is also obtained by calculating the values of $\Delta\Lambda$ for the wild-type structure and a mutant structure denoted by $\Delta\Lambda_W$ and $\Delta\Lambda_M$, respectively.

The change in ΔF caused by a mutation, $\Delta\Delta F$, is given by

$$\Delta\Delta F/(k_BT_0)=\Delta\Delta\Lambda/(k_BT_0)-\Delta\Delta S/k_B \quad (4.7)$$

Negative $\Delta\Delta F$ implies that the mutant is more stable than the wild type. Following the suggestion made in our recent work,[10] only those with $\Delta\Delta F/(k_B T_0)<-5$ are identified as thermostabilized mutants in this work.

4.2.6. Experimentally determined three-dimensional (crystal) structure of $A_{2a}R$

In the recent work,[10] we employed the 3D structure of $A_{2a}R$ in an inactive state determined by Murata and coworkers (PDB ID: 3VG9).[35] They succeeded in the crystallization followed by the structure determination using X-ray crystallography by means of the binding of a mouse monoclonal-antibody Fab-fragment to $A_{2a}R$. In this structure, the binding interface is at the end of the intracellular region and far from the TM region. Hence, the TM portions of this structure and those of the wild-type structure should share almost the same characteristics.[35]

4.2.7. Construction of candidate models for three-dimensional structure of $A_{2a}R$

In the homology modeling, the amino-acid sequence (AAS) of the target protein is aligned to that of another protein as a template protein, and the 3D-structure models of the target are constructed by referring to the structural information of the template. The procedure of constructing candidate models for the 3D structure of $A_{2a}R$ in an inactive state comprises the following steps:

Step (1). Multiple alignment for the AAS of $A_{2a}R$.
Using PSI-BLAST,[36] we search GPCRs in inactive states whose 3D structures are experimentally known: 13 class A GPCRs are found. We then align the AAS of $A_{2a}R$ to the AASs of the 13 GPCRs using the multiple sequence alignment. A template should possess sufficiently high TM region identity (TMI) with the target.[37] Those with TMIs exceeding 25% are chosen from among the 13 GPCRs. The measurement of TMI is performed by employing the web server of TMHMM2.0.[38] The following 5 GPCRs are chosen as the templates: the human dopamine D3 receptor (DD3R, PDBID: 3PBL), human β2 adrenergic receptor (β2AR, PDBID: 2RH1), human histamine H1 receptor (HH1R, PDBID: 3RZE), human sphingosine-1-phosphate receptor (S1PR1, PDBID: 3V2W), and turkey β1 adrenergic receptor (β1AR, PDBID: 2VT4). The TMIs of DD3R, β2AR, HH1R, S1PR1, and β1AR are, respectively, 35, 34, 33, 32, and 25%.

Step (2). Generation of candidate models for the 3D structure of $A_{2a}R$.
We generate 100 candidate models using MODELLER[39] for each template. A total of 500 candidate models for the wild-type structure are thus obtained. The LJ potential energy for a model can be unreasonably high owing to the overlaps of protein atoms. Such overlaps are removed by local minimization of the energy using the CHARMM biomolecular simulation program[40] through the multiscale modeling tool structural biology (MMTSB) program.[41]

4.2.8. Modeling of mutant structures

For a 3D-structure model given, we generate its mutant structures using MODELLER.[39] Starting from the 3D-structure model, we move the residues which are present within a distance of 5 Å from the mutated residue. For each mutant, a total of 10 model structures are generated and the unrealistic overlaps of protein atoms are removed by local minimization of the energy as mentioned above. A thermodynamic quantity for a mutant is obtained as the average of the 10 values.

The computation time required for the local minimization is much longer than that for calculating ΔF. Therefore, we wish to keep the number of model structures for each mutant as small as possible. Further, when the number is largely increased, more unsuitable model structures are unavoidably generated and the result can possibly become worse. We have verified that increasing the number to 20 changes the average value in a quantitative sense but the qualitative aspects of our conclusions are not altered at all: 10 is a good compromise. A more detailed discussion is provided in "Effects of the number of model structures for each mutant".

4.3. Results and discussion

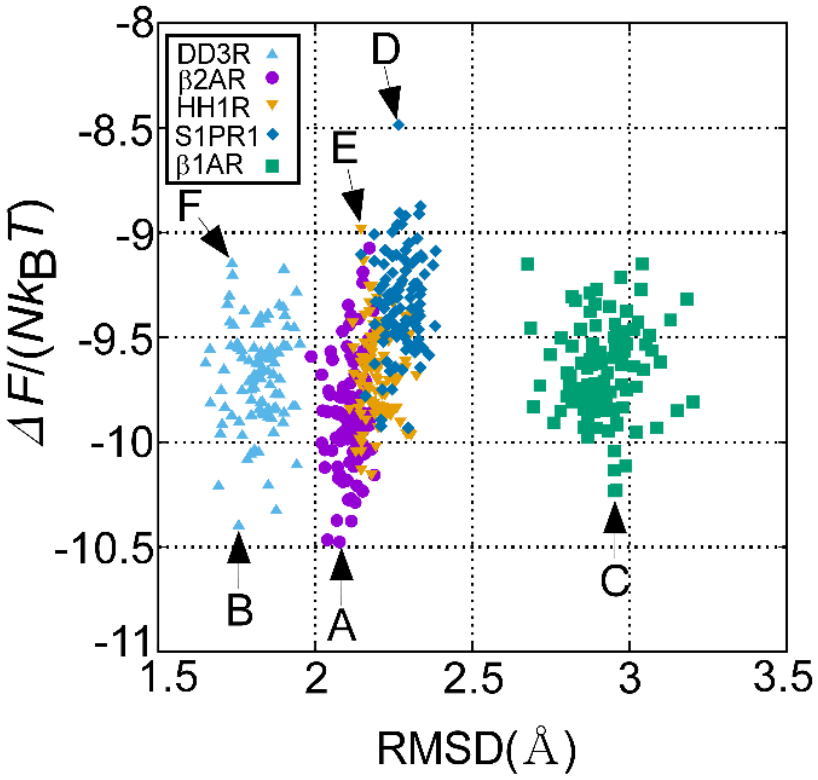
4.3.1. Values of RMSD and ΔF of the 500 candidate models for three-dimensional structure of $A_{2a}R$

For the 500 candidate models of the 3D structure, we plot ΔF against RMSD between a model and the crystal structure calculated for the TM portions in Figure 4.2. The number of residues N forming the TM portions for a model is dependent on the template from which the model is generated. Hence, ΔF is normalized by N (the values of N for DD3R, β2AR, HH1R, S1PR1, and β1AR are 143, 134, 152, 149, and 144, respectively). RMSD is calculated using the C_α atoms common in the two structures considered. For a template giving higher TMI, the average value of RMSDs of the resulting candidate models tends to be smaller. On the other hand, the correlation between RMSD and ΔF is rather low. RMSD can be a good measure in the selection of the best model, but it is unknown in a practical situation. Therefore, we wish to use ΔF as the measure instead.

We choose 6 candidate models which are referred to as structures A, B, C, D, E, and F, respectively (see Figure 4.2). The values of $\Delta\Lambda$, $-T_0\Delta S$, and ΔF normalized by Nk_BT_0 and RMSD for the crystal structure and structures A–F are compared in Table 4.1. Among the 6 structures, $\Delta F/(Nk_BT_0)$ of structure A is the lowest and RMSD of structure F is the smallest. The values of ΔF for the crystal structure and structures A–C are relatively lower than those of structures D–F. In terms of RMSD, the 6 structures can be divided into three groups, (F, B), (A, E, D), and C, for which RMSD takes smaller, intermediate, and larger values, respectively. We employ the crystal structure and structures A–F as the 3D-structure model for the wild type and compare the prediction results by following the basic procedure described in the Introduction.

Ideally, $\Delta F/(Nk_BT_0)$ of the crystal structure is considerably lower than that of any structure model generated. (It is interesting to note that $-\Delta S/k_B$ takes the lowest vale for the crystal structure.) As observed in Table 4.1, however, the crystal structure and structures A and B share almost the same value of $\Delta F/(Nk_BT_0)$. This is because the crystal structure as well as our free-energy function is not perfect. For a crystal structure with higher resolution, $\Delta F/(Nk_BT_0)$ may be lower.

Structures A, B, and F are compared with the crystal structure in Figure 4.3. There are visual differences among the three structures. For instance, unlike in structure A, one of the 7 helices is partially unfolded in structures B and F. However, it is difficult to guess how the visual differences



affect the performance in identifying thermostabilizing mutations. Detailed analyses based on statistical thermodynamics are absolutely necessitated.

Figure 4.2. Relation between RMSD and ΔF for the 500 candidate models of 3D structure of $A_{2a}R$. The five proteins, DD3R, β2AR, HH1R, S1PR1, and β1AR, are employed as the templates in the homology modeling, and 100 models are generated for each template. RMSD is the root-mean-square deviation for C_α atoms between a model and the 3D structure experimentally determined (i.e., crystal structure). ΔF is the free-energy change upon protein folding (see Equation (4.2)). F=free-energy function, T=T_0=298 K, k_B=Boltzmann constant, and N=number of residues forming the transmembrane portions. ΔF denotes the change in F upon protein folding.

Table 4.1. Changes in free-energy function and its energetic and entropic terms upon protein folding and RMSD.				
Structure	$\Delta\Lambda/(Nk_BT)$	$-\Delta S/(Nk_B)$	$\Delta F/(Nk_BT)$	RMSD (Å)
Crystal	–7.94	–2.40	–10.34	0
A	–8.53	–1.95	–10.48	2.08
B	–8.13	–2.27	–10.40	1.76
C	–8.13	–2.10	–10.23	2.96
D	–6.73	–1.76	–8.49	2.26
E	–7.39	–1.58	–8.98	2.15
F	–7.32	–1.83	–9.15	1.73
F=free-energy function=$\Lambda - TS$, T=T_0=298 K, k_B=Boltzmann constant, N=number of residues forming the transmembrane portions. ΔX denotes the change in X upon protein folding. RMSD=root-mean-square deviation between a structure and the crystal structure.				

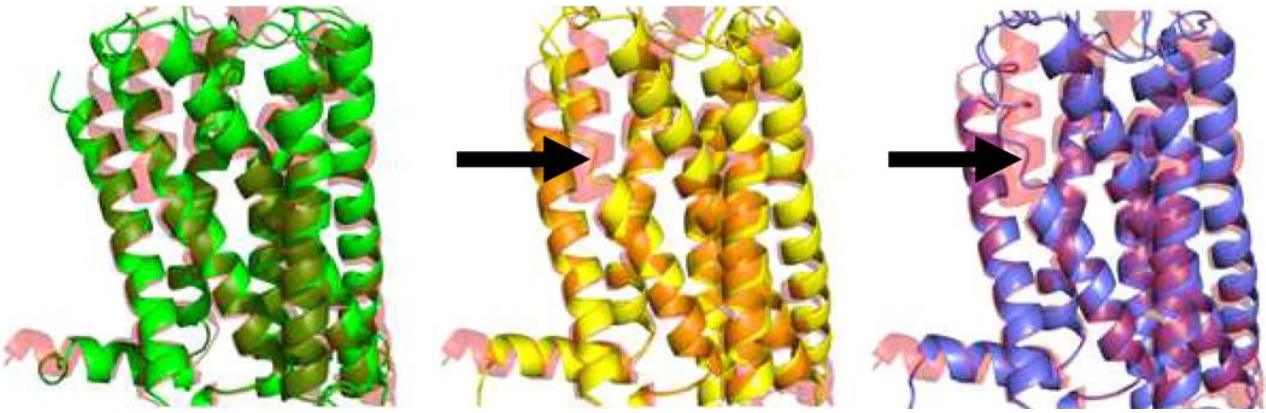


Figure 4.3. Visual comparison between structure A (green), B (yellow), or F (blue) and the crystal structure (red). In each illustration, the two structures are superposed so that the root-mean-square deviation for C_α atoms between them can be minimized. There are visual differences among the three structures. For example, the portion of α-helix indicated by the solid arrow is partially unfolded in structures B and F.

4.3.2. Nomination of key residues

We mutate each residue of $A_{2a}R$ from S6 to R309 to every amino acid residue other than the original one. Hence, a total of 304×19=5776 mutations are examined. For these mutations, in Figure 4.4 we plot $-\Delta\Delta S$ in case J against $-\Delta\Delta S$ in the reference case. Structure J (J=A, C, D) is employed as the 3D-structure model in case J, and the crystal structure is employed as it in the reference case. $-\Delta\Delta S$ in case J and $-\Delta\Delta S$ in the reference case are only poorly correlated though the correlation coefficient is in the order, A>C>D. (The correlation between $\Delta\Delta F$ in case J and $\Delta\Delta F$ in the reference case is also very low.)

We are more concerned with the key residues nominated. The criterion for the nomination is the following: Count the number of the mutations of a residue which leads to $-\Delta\Delta S/k_B<-10$; when it is larger than or equal to M, the residue is nominated as one of the key residues. Setting M at too large a value will result in only a very small number of key residues. Setting M at too small a value will give unacceptably many key residues and the examination experiments should be rather laborious. In this study, we set M at 4. The key residues nominated are considerably dependent on the 3D-structure model employed. Those in case J and the reference case are as follows:

Crystal structure (reference case): T88, S91, and N181.
Structure A (case A): G136, P189, V186, A20, V55, S91, C185, and T88.
Structure C (case C): S94, S47, and S91.
Structure D (case D): V186 and S91.

In the recent work employing the crystal structure,[10] we nominated T88 and S91 as the key residues. They are both nominated only in case A (this can be the most important).

It is remarkable that S91 is nominated in all of the four cases. We showed that S91 is conserved in significantly many other class A GPCRs.[10] (This is not the case for T88.) S91 is a hot-spot residue a mutation for which potentially leads to remarkable stabilization. In the next section, we consider the 19 mutations of S91 and compare the prediction results obtained from the crystal structure and structures A through F.

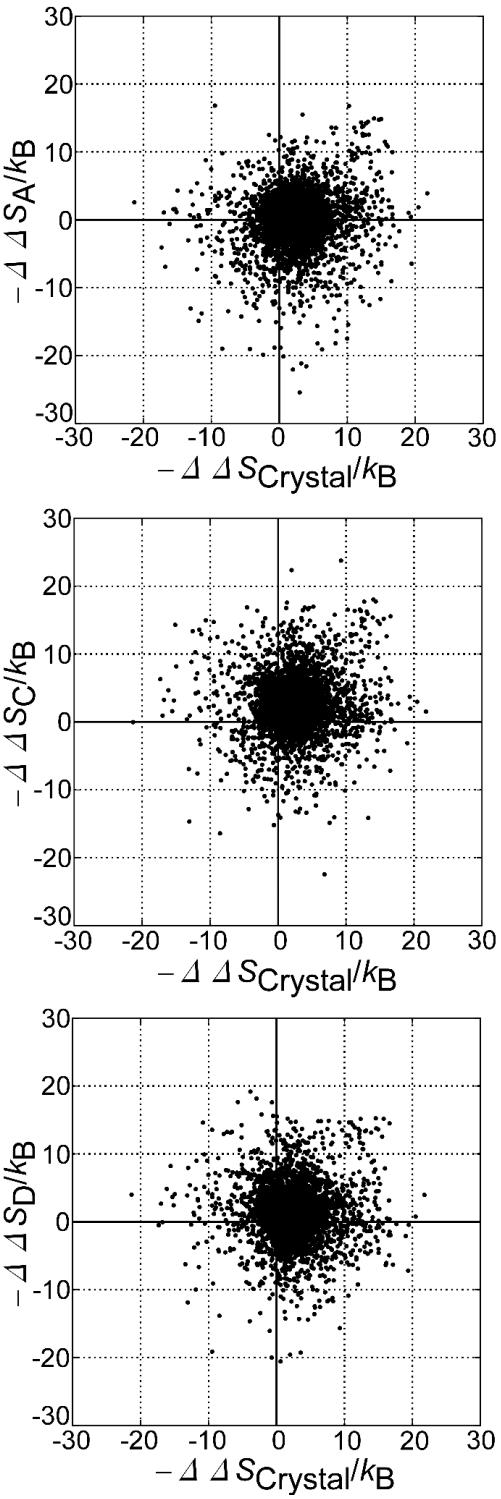


Figure 4.4. Correlation between $-\Delta\Delta S$ (see Equation (5)) in case J and $-\Delta\Delta S$ in the reference case. Structure J (J=A, C, D) is employed as the 3D-structure model in case J, and the crystal structure is employed as it in the reference case. The subscripts "J" and "Crystal", respectively, denote values in case J and in the reference case. All of the 304×19=5776 mutations are considered. The correlation coefficients between $-\Delta\Delta S_J$ and $-\Delta\Delta S_{crystal}$ for J=A, C, and D are 0.106, 0.081, and 0.005, respectively.

4.3.3. Mutations for S91

Figure 4.5 shows the relation between $\Delta\Delta F$ in case J (J=A, B, C, D, E, F) and $\Delta\Delta F$ in the reference case. The correlation between $\Delta\Delta F$ in case J and $\Delta\Delta F$ in the reference case is sufficiently high. The correlation coefficients are in the order, C~A>B>E>F~D. Though the correlation calculated for all of the mutations is quite low as observed in Figure 4.4, the mutations for S91 are exceptional. It is presumable that $\Delta\Delta F$ for a hot-spot residue is not greatly dependent on the 3D-structure model employed.

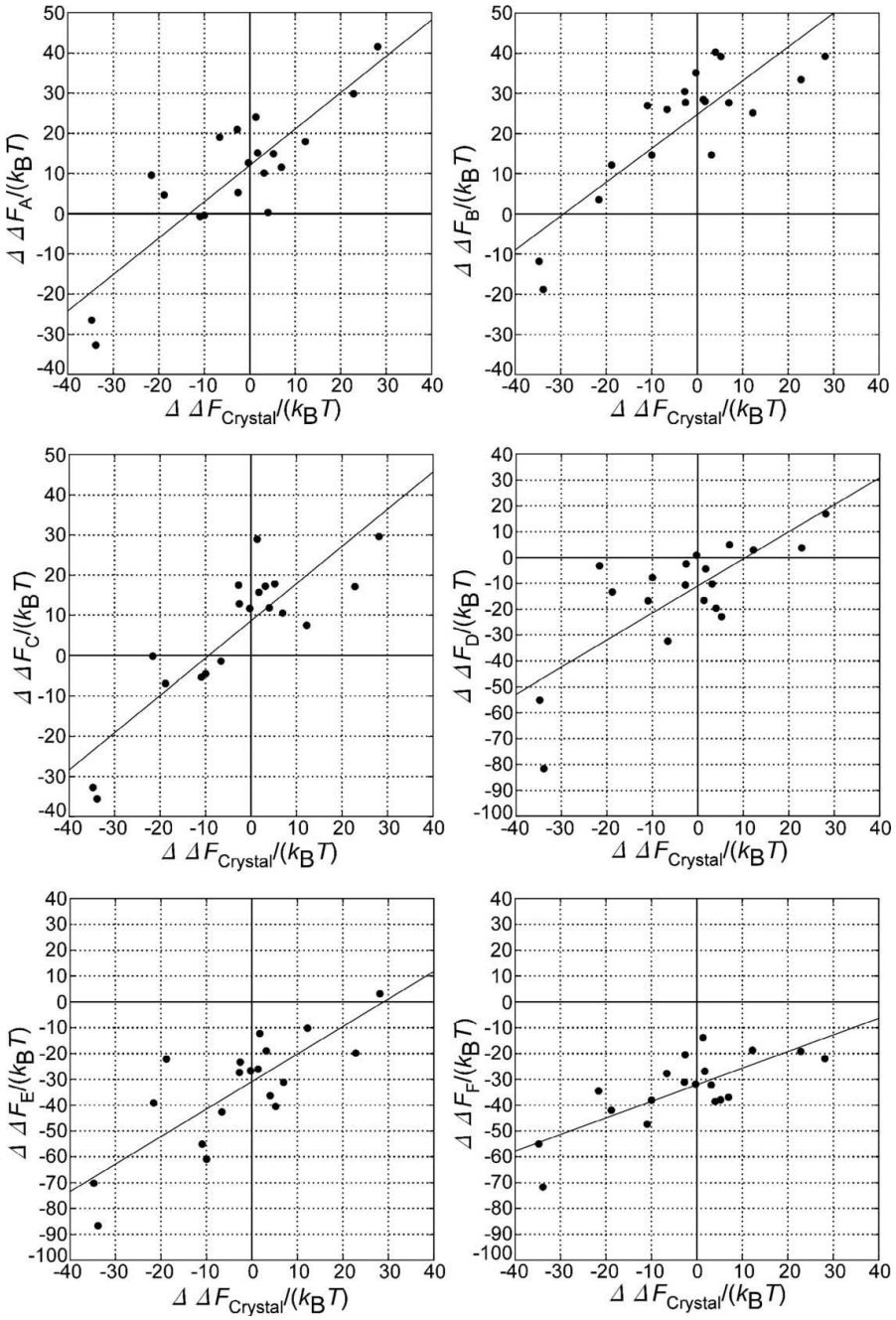


Figure 4.5. Correlation between $\Delta\Delta F$ (see Equation (7)) in case J and $\Delta\Delta F$ in the reference case. Structure J (J=A, B, C, D, E, F) is employed as the 3D-structure model in case J, and the crystal structure is employed as it in the reference case. The subscripts "J" and "Crystal", respectively, denote values in case J and in the reference case. 19 mutations for S91 are considered. The correlation coefficients between $\Delta\Delta F_J$ and $\Delta\Delta F_{\text{crystal}}$ for J=A, B, and C are 0.856, 0.842, and 0.864, respectively, and those for J=D, E, and F are 0.758, 0.797, and 0.761, respectively.

4.3.4. Comparison between experimental result and theoretical prediction

In the recent work using the crystal structure,[10] we examined 9 single mutations for T88 and S91 and a double mutation T88E−S91R on the basis of the experimental results. The same examination is performed for case J (J=A, B, C, D, E, F). We employ the stability criterion in the theoretical prediction and the thermostability measure in our recent experiments which lead to the success rate, 10/10. Therefore, a mutation with $\Delta\Delta F/(k_B T_0)<-5$ is regarded as a stabilizing one[10] (that with $\Delta\Delta F/(k_B T_0)\geq -5$ is regarded as a destabilizing one). We note that 7 stabilizing and 3 destabilizing mutations are chosen though only a very small percentage of mutations are stabilizing. The results of the examination are illustrated in Figures 4.6 and 4.7. The results from the 7 cases are significantly different in a quantitative sense, reflecting the high sensitivity of $\Delta\Delta F$ to the wild-type structure model. The success rates in the 7 cases are compared in Table 4.1. The most important matter is to predict whether a mutation is stabilizing or destabilizing. We are not very concerned with the degree of stabilization or destabilization, because it is difficult to predict the degree for all of the possible mutations with sufficient accuracy and rapidness.

The success rates in cases A, B, and C are 8/10, whereas the success rate in the reference case is 10/10. The success rates in cases D, E, and F are in the rage from 6/10 to 7/10. The increase in the denaturation temperature brought by the double mutation T88E−S91R was experimentally shown to be considerably larger than that brought by any of the single stabilizing mutations. This experimental fact is reproduced only in the reference case and case E. A problem shared by cases D, E, and F is that all of the 3 destabilizing mutations are incorrectly predicted to be stabilizing.

Overall, the prediction performance in cases A, B, and C is higher than that in cases D, E, and F. It is lower than that in the reference case but only slightly. The result from case A is the most similar to that from the crystal structure. As the measure of the appropriateness of a 3D-structure model, ΔF is as good as RMSD or even better. It can reasonably be suggested that the model giving the lowest value of ΔF (structure A in this case) be selected as the best one.

In cases A, B, and C, T88A and T88E are incorrectly predicted to be destabilizing. For these mutations, $-\Delta\Delta S<0$ but $\Delta\Delta\Lambda>0$ and the latter dominates with the result of rather large, positive $\Delta\Delta F$. It is possible that this result stems partly from an inappropriate value of E.[32,33] The optimization of E in calculating the energetic term $\Delta\Lambda$ is expected to further improve the prediction performance.

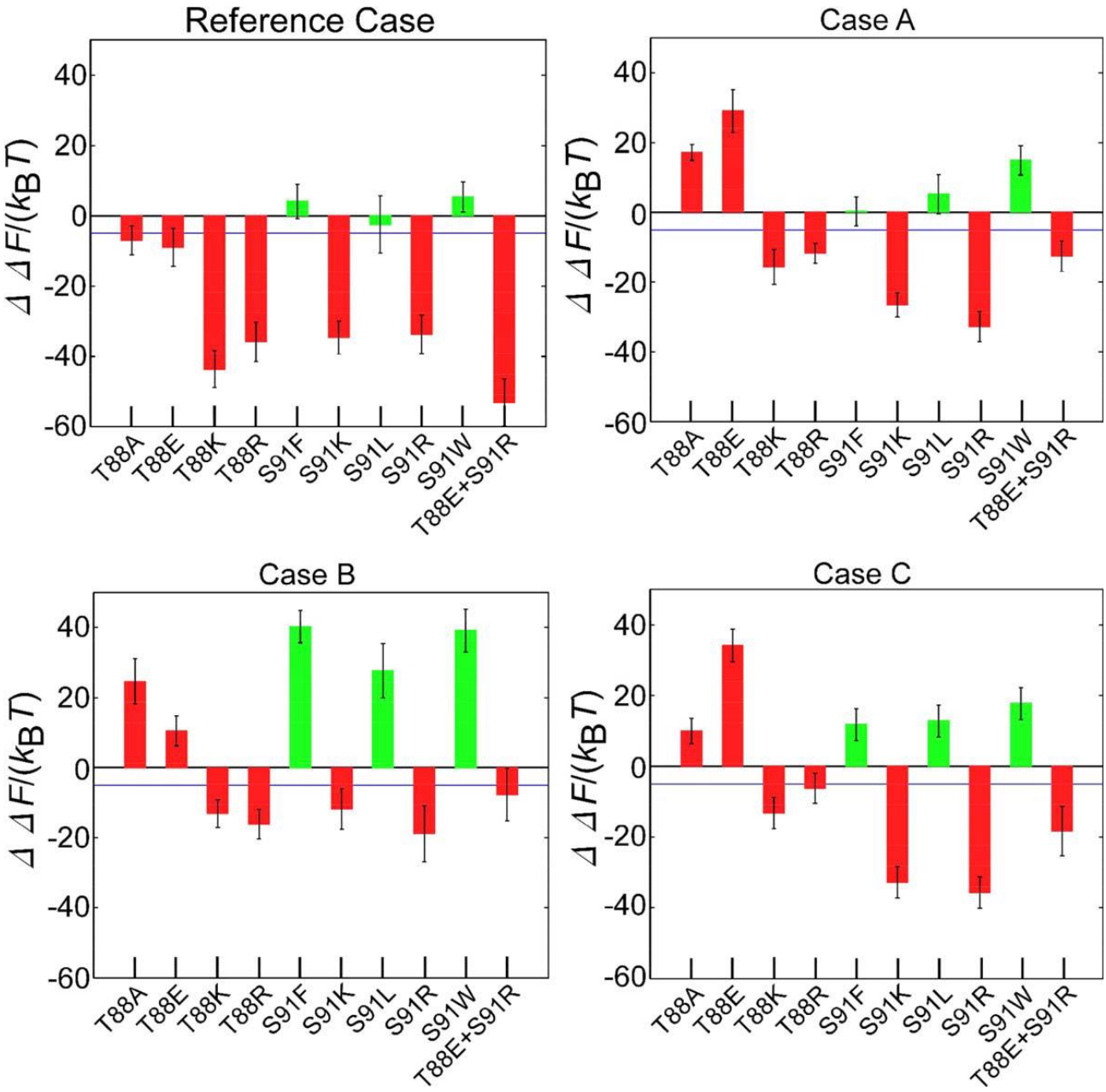


Figure 4.6. Comparison between theoretical prediction and experimental result. Four mutations for T88, five mutations for S91, and a double mutation are considered. The comparison is made for cases *J* (*J*=A, B, C) and the reference case. Structure *J* is employed as the 3D-structure model in case *J*, and the crystal structure is employed as it in the reference case. The values of $\Delta\Delta F/(k_BT)$ (T=T_0=298 K) are given as the rectangles. The mutations experimentally shown to be stabilizing and destabilizing are colored in red and in green, respectively. Standard errors are also shown as the error bars. In the theoretical prediction, a mutation giving $\Delta\Delta F/(k_BT_0)<-5$ is regarded as a stabilizing one. The number of model structures for each mutant is 10.

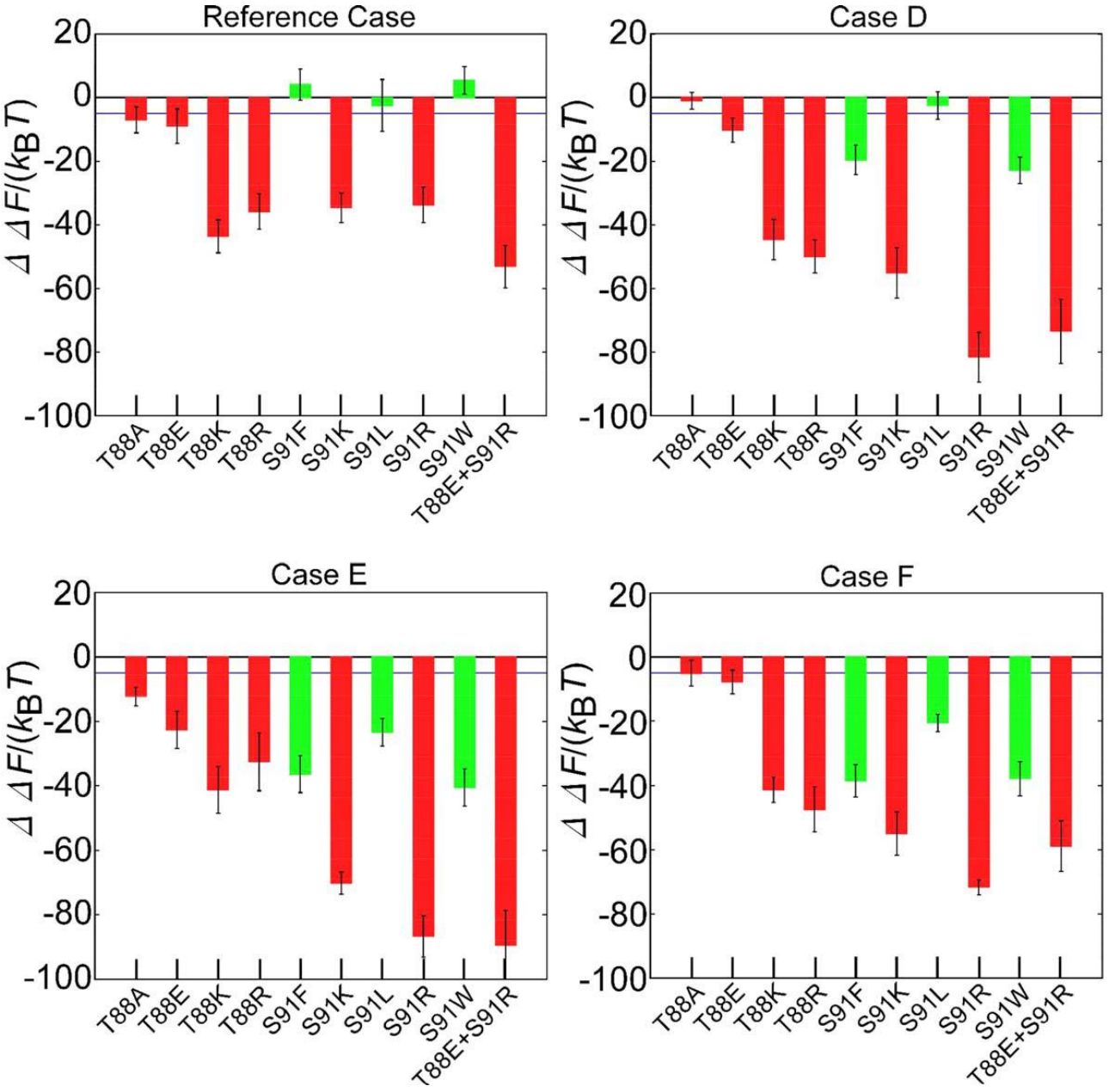


Figure 4.7. Comparison between theoretical prediction and experimental result. Four mutations for T88, five mutations for S91, and a double mutation are considered. The comparison is made for cases *J* (*J*=D, E, F) and the reference case. Structure *J* is employed as the 3D-structure model in case *J*, and the crystal structure is employed as it in the reference case. The values of $\Delta\Delta F/(k_BT)$ (T=T_0=298 K) are given as the rectangles. The mutations experimentally shown to be stabilizing and destabilizing are colored in red and in green, respectively. Standard errors are also shown as the error bars. In the theoretical prediction, a mutation giving $\Delta\Delta F/(k_BT_0)<-5$ is regarded as a stabilizing one. The number of model structures for each mutant is 10.

4.3.5. Effects of the number of model structures for each mutant

We increase the number of model structures for each mutant from 10 to 20 and examine how the results change: They are illustrated in Figures 4.8 and 4.9. Of course, the average values exhibit quantitative changes and the error bars become shorter. However, the qualitative aspects of our conclusions are not altered at all. For example, the prediction performance in cases A, B, and C is higher than that in cases D, E, and F. The result from case A is the most similar to that from the crystal structure. It is interesting to note that the SRs in the reference case and case D become slightly lower as indicated within the parentheses in Table 4.1. This is because when the increase from 10 to 20 is used, more unsuitable model structures are unavoidably generated. A larger number of model structures not only make the computational load heavier but also possibly lead to a worse result. (The error bars may not be meaningful.)

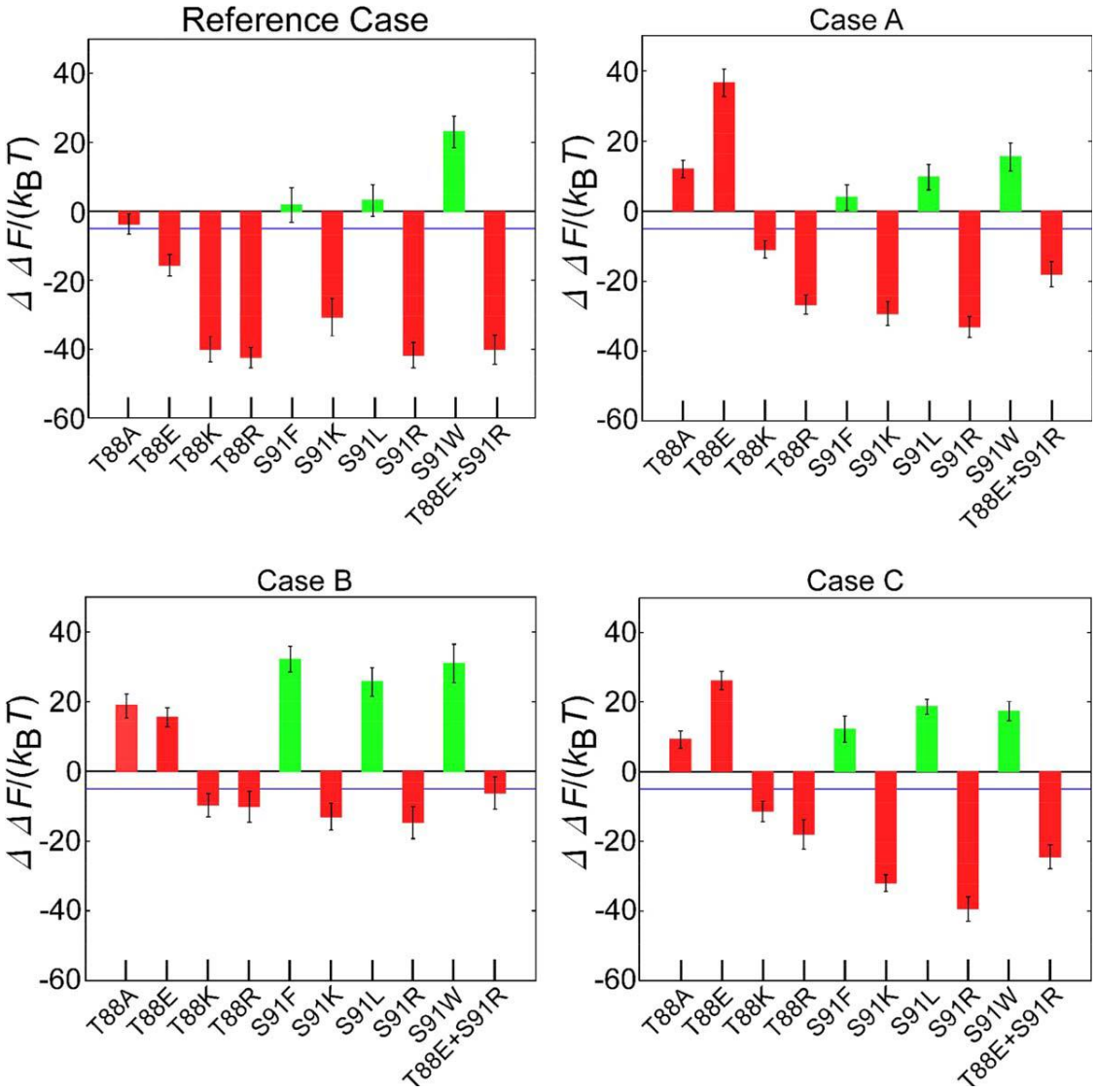


Figure 4.8. The same quantities as those in Figure 4.6 are shown by increasing the number of model structures for each mutant from 10 to 20.

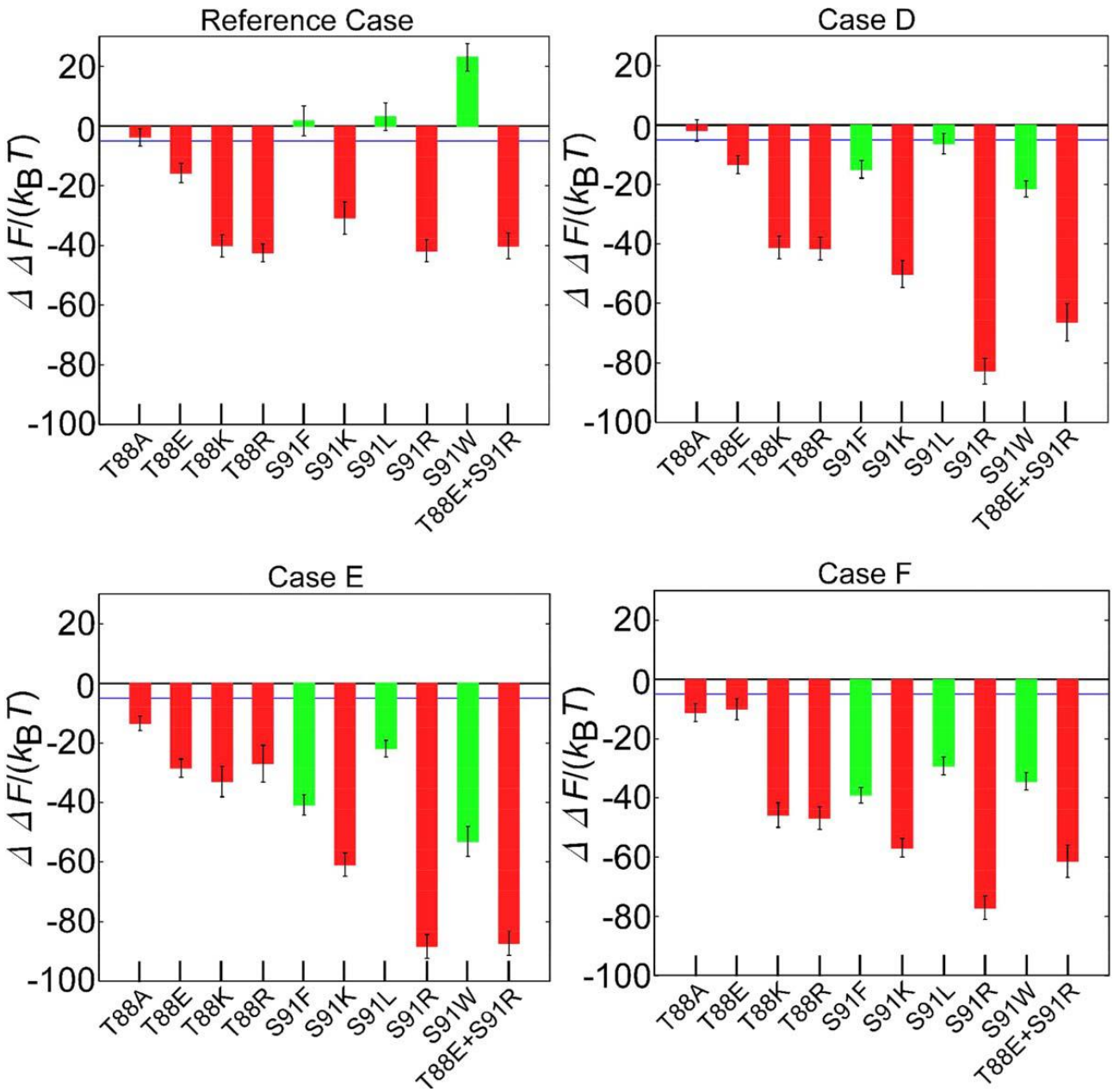


Figure 4.9. The same quantities as those in Figure 4.7 are shown by increasing the number of model structures for each mutant from 10 to 20.

4.3.6. Comparison between Δ*F* and RMSD as the accuracy measure for 3D-structure model of the wild type

To further show that Δ*F* is better than RMSD as the accuracy measure for the 3D-structure model of the wild type, we consider an additional structure model, the one with the smallest value of RMSD in Figure 4.2 (structure G). For structure G, $\Delta\Lambda/(Nk_BT_0)$=–7.56, $-\Delta S/k_B$=–2.07, $\Delta F/(Nk_BT_0)$=–9.63, and RMSD=1.65 (also see Table 4.1). $\Delta F/(Nk_BT_0)$ for structure G is higher than that of structure A, B, or C but lower than that of structure D, E, or F. We draw a picture presenting the result of performance test in the case where structure G is employed (case G): It is shown in

Figure 10 which should be compared to Figures 4.6 and 4.7. The success rates for “Stabilizing”, “Destabilizing”, and “Overall” (see Table 4.2) are 5/7, 2/3, and 7/10, respectively.

As observed in Figures 4.6 and 4.7, the results in cases A, B, and C are qualitatively similar. For example, the prediction fails for T88A and T88E. The results in cases D, E, and F are also qualitatively similar. For example, the prediction fails for S91F, S91L, and S91W. Figure 4.10 indicates that the result in case G is qualitatively an intermediate between the results in cases A–C and those in cases D–F: The prediction fails for T88A, T88E, and S91W. ΔF is much more correlated with the result of performance test than with RMSD. As illustrated in Figure 4.3, one of the 7 helices is partially unfolded in structure B but not in structure A. Nevertheless, the results from structures A and B sharing almost the same ΔF are qualitatively similar. These are indicative that the thermodynamic properties of a model are reflected not in RMSD or visual structural differences but in ΔF.

The performance achieved in the reference case should be retained when RMSD of the model is very small, but it is difficult to construct such a model using the homology modeling. In general, the values of RMSD of the models are significantly large. Moreover, RMSD is unknown in a practical situation. Under these conditions, ΔF works better than RMSD as the accuracy measure despite its simplicity.

Table 4.2. Comparison among the 7 cases in terms of the success rate.

Case	Stabilizing	Destabilizing	Overall
Reference	7/7 (6/7)	3/3	10/10 (9/10)
A	5/7	3/3	8/10
B	5/7	3/3	8/10
C	5/7	3/3	8/10
D	7/7 (6/7)	0/3	7/10 (6/10)
E	7/7	0/3	7/10
F	7/7	0/3	7/10

In the 7 cases, respectively, the crystal structure and structures A, B, C, D, E, and F are employed as the wild-type structure model. “Stabilizing” and “Destabilizing” represent the success rates for the set of mutations which is experimentally shown to be stabilizing and destabilizing, respectively. “Overall” denotes the overall success rate. The number of model structures for each mutant is 10: When it is increased to 20, the success rate is changed in the reference case and case D as indicated within the parentheses.

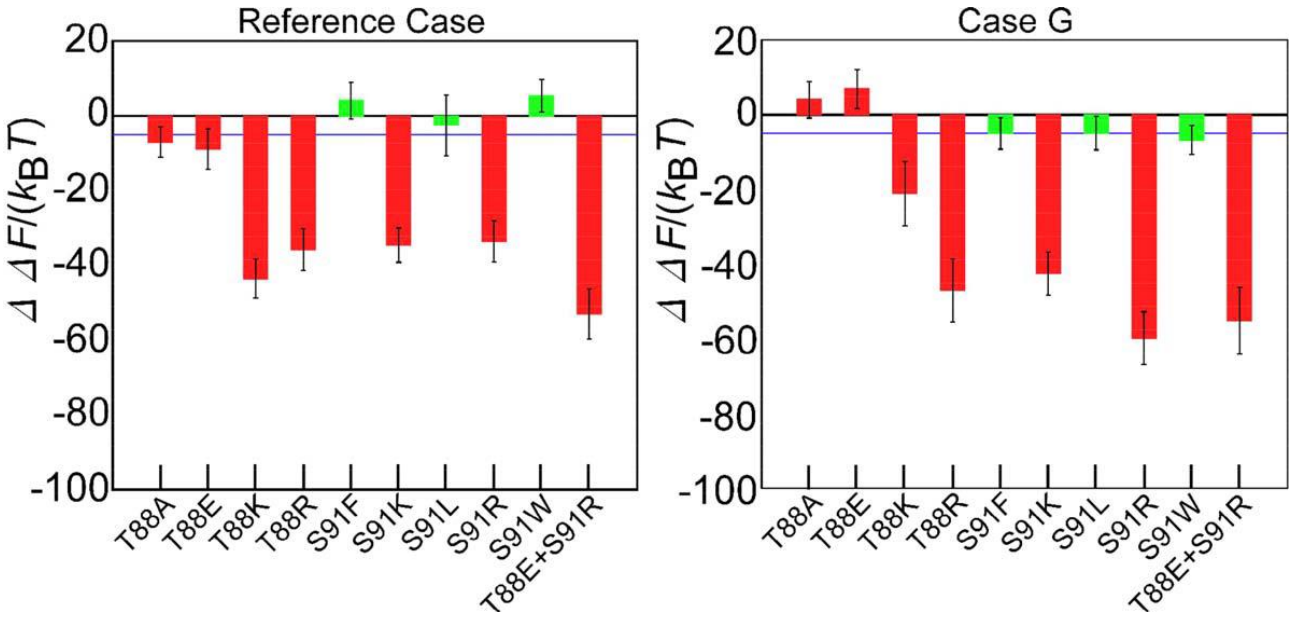


Figure 4.10. Comparison between theoretical prediction and experimental result. Four mutations for T88, five mutations for S91, and a double mutation are considered. The comparison is made for the reference case and case G. In case G, the structure model with the smallest value of RMSD in Figure 2 is employed. Also see the captions for Figures 4.6 and 4.7. For S91F and S91L, $\Delta\Delta F/(k_B T_0)>-5$.

4.4. Conclusions

Our recently developed theoretical method for identifying thermostabilizing mutations of a membrane protein[10] has been applied to a practical situation where the 3D structure of the protein is not experimentally available. The theoretical method is combined with the homology modeling, and the combined approach has been illustrated for the adenosine A_{2a} receptor ($A_{2a}R$). Though the 3D structure of $A_{2a}R$ has experimentally determined, it is not utilized in this work. First, template proteins possessing sufficiently high transmembrane region identity (TMI) with the target protein (TMI exceeding 25%) are searched using the multiple sequence alignment. Second, the 3D-structure models of the target are constructed by referring to the structural information of each template. Third, from among sufficiently many models thus generated, the best model is selected. The theoretical method then becomes applicable to the identification. How should we select the best model?

The best model can be the one whose root-mean-square deviation (RMSD) for C_α atoms, which is defined between a model and the 3D structure experimentally determined (i.e., crystal structure), is very small. However, it is difficult to construct such a model. Moreover, RMSD cannot be

calculated in a practical situation. We therefore make the following proposition: The model with the lowest value of ΔF (the change in our free-energy function F upon protein folding) is the best one. We have examined the prediction performance in cases where the crystal structure and structures A through G (see Figure 4.2 and "Comparison between ΔF and RMSD as the accuracy measure for 3D-structure model of the wild type"). The performance from structure A is the closest to that from the crystal structure: Both of T88 and S91 are nominated as the key residues; the correlation coefficient between $\Delta\Delta F$ in case A and $\Delta\Delta F$ in the reference case ($\Delta\Delta F$ is the change in ΔF upon mutation) for the 19 mutations for S91 is considerably high (~0.86); and the overall success rates from the two structures are 8/10 and 10/10, respectively. The proposition mentioned above has been proved to be valid. ΔF is better than RMSD as the measure for selecting the best 3D-structure model. We have also argued that the thermodynamic properties of a model are reflected not in RMSD or visual structural differences but in ΔF.

S91 is a hot-spot residue in the sense that a mutation for it is expected to give remarkable stabilization. It appears that $\Delta\Delta F$ for a hot-spot residue is not greatly dependent on the 3D-structure model employed.

The results can be dependent on the computer programs employed for constructing the structure of the target protein and generating the model structures of each mutant. If other computer programs were employed, the performance could become even higher, but we render such a test to a future study. It is important to apply our theoretical method combined with the homology modeling to a number of GPCRs whose 3D structures are unknown, identify thermostabilizing mutations, enable us to accomplish the crystallization followed by the structure determination by X-ray crystallography. In fact, we have succeeded in accomplishing this for a prostaglandin receptor: The details will be published in a future article. One of the remaining issues is to cope with a case where no template proteins with sufficiently high TMI can be found. At the same time, E in calculating the energetic term of ΔF (i.e., $\Delta\Lambda$) should be optimized.[32,33] Works in these directions are in Chapter 5.

References

[1]Boyd, D.; Schierle, C.; Beckwith, J. How many membrane proteins are there? *Protein Sci.* **1998**, *7*, 201−205.

[2]Wallin, E.; Heijne, G. V. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **1998**, *7*, 1029−1038.

[3]Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabási, A. L.; Vidal, M. Drug-target network. *Nat. Biotechnol.* **2007**, *25*, 1119−1126.

[4]Lau, F. W.; Nauli, S.; Zhou, Y.; Bowie, J. U. Changing single side-chains can greatly enhance the resistance of a membrane protein to irreversible inactivation. *J. Mol. Biol.* **1999**, *290*, 559−564.

[5]Zhou, Y.; Bowie, J. U. Building a thermostable membrane protein. *J. Biol. Chem.* **2000**, *275*, 6975−6979.

[6]Scott, D. J.; Kummer, L.; Tremmel, D.; Plückthun, A. Stabilizing membrane proteins through protein engineering. *Curr. Opin. Chem. Biol.* **2013**, *17*, 427−435.

[7]Serrano-Vega, M. J.; Magnani, F.; Shibata, Y.; Tate, C. G. Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 877−882.

[8]Magnani, F.; Shibata, Y.; Serrano-Vega, M. J.; Tate, C. G. Co-evolving stability and conformational homogeneity of the human adenosine A_{2a} receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10744−10749.

[9]Shibata, Y.; White, J. F.; Serrano-Vega, M. J.; Magnani, F.; Aloia, A. L.; Grisshammer, R.; Tate, C. G. Thermostabilization of the neurotensin receptor NTS1. *J. Mol. Biol.* **2009**, *390*, 262−277.

[10]Yasuda, S.; Kajiwara, Y.; Takamuku, Y.; Suzuki, N.; Murata, T.; Kinoshita, M. Identification of thermostabilizing mutations for membrane proteins: rapid method based on statistical thermodynamics. *J. Phys. Chem. B*, **2016**, *120*, 3833−3843.

[11]Chen, K. M.; Zhou, F.; Fryszczyn, B. G.; Barth, P. Naturally evolved G protein-coupled receptors adopt metastable conformations. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13284−13289.

[12]Bhattacharya, S.; Lee, S.; Grisshammer, R.; Tate, C. G.; Vaidehi, N. Rapid Computational Prediction of Thermostabilizing Mutations for G Protein-Coupled Receptors. *J. Chem. Theory Comput.* **2014**, *10*, 5149−5160.

[13]Sauer, D. B.; Karpowich, N. K.; Song, J. M.; Wang, D. N. Rapid bioinformatic identification of thermostabilizing mutations. *Biophys. J.* **2015**, *109*, 1420−1428.

[14]Oda, K.; Kinoshita, M. Physicochemical origin of high correlation between thermal stability of a protein and its packing efficiency: A theoretical study for staphylococcal nuclease mutants. *Biophys. Physicobiol.* **2015**, *12*, 1−12.

[15]Murakami, S.; Oshima, H.; Hayashi T.; Kinoshita M. On the physics of thermal-stability changes upon mutations of a protein. *J. Chem. Phys.* **2015**, *143*, 125102.

[16]Harano, Y.; Kinoshita, M. Translational-entropy gain of solvent upon protein folding. *Biophys. J.* **2005**, *89*, 2701−2710.

[17]Yoshidome, T.; Kinoshita, M.; Hirota, S.; Baden, N.; Terazima, M. Thermodynamics of apoplastocyanin folding: comparison between experimental and theoretical results. *J. Chem. Phys.* **2008**, *128*, 225104.

[18]Yasuda, S.; Yoshidome, T.; Oshima, H.; Kodama, R.; Harano, Y.; Kinoshita, M. Effects of side-chain packing on the formation of secondary structures in protein folding. *J. Chem. Phys.* **2010**, *132*, 065105.

[19]Yasuda, S.; Oshima, H.; Kinoshita, M. Structural stability of proteins in aqueous and nonpolar environments. *J. Chem. Phys.* **2012**, *137*, 135103.

[20]Kinoshita, M. A new theoretical approach to biological self-assembly. *Biophys. Rev.* **2013**, *5*, 283−293.

[21]Oshima, H.; Kinoshita, M. Essential roles of protein-solvent many-body correlation in solvent-entropy effect on protein folding and denaturation: Comparison between hard-sphere solvent and water. *J. Chem. Phys.* **2015**, *142*, 145103.

[22]Kinoshita, M. Molecular origin of the hydrophobic effect: Analysis using the angle-dependent integral equation theory. *J. Chem. Phys.* **2008**, *128*, 024507.

[23]Yoshidome, T.; Oda, K.; Harano, Y.; Roth, R.; Sugita, Y.; Ikeguchi, M.; Kinoshita, M. Free-energy function based on an all-atom model for proteins. *Proteins* **2009**, *77*, 950−961.

[24]Yasuda, S.; Yoshidome, T.; Harano, Y.; Roth, R.; Oshima, H.; Oda, K.; Sugita, Y.; Ikeguchi, M.; Kinoshita, M. Free-energy function based on an all-atom model for proteins. *Proteins* **2011**, *79*, 2161−2171.

[25]König, P. M.; Roth, R.; Mecke, K. R. Morphological thermodynamics of fluids: shape dependence of free energies. *Phys. Rev. Lett.* **2004**, *93*, 160601.

[26]Roth, R.; Harano, Y.; Kinoshita, M. Morphometric approach to the solvation free energy of complex molecules. *Phys. Rev. Lett.* **2006**, 97, 078101.

[27]Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids, 3rd ed.* Academic Press; London, U.K., 2006.

[28]Popot, J. L.; Engelman D. M. Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* **2000**, *69*, 881−922.

[29]Tusnády, G. E.; Dosztányi, Z.; Simon, I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics* **2005**, *21*, 1276−1277.

[30]McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*, 777−793.

[31]Sneddon, S. F.; Tobias, D. J.; Brooks, C. L. III Thermodynamics of amide hydrogen bond formation in polar and apolar solvents. *J. Mol. Biol.* **1989**, *209*, 817−820.

[32]Bowie, J. U. Membrane protein folding: how important are hydrogen bonds? *Curr. Opin. Struct. Biol.* **2011**, *21*, 42−49.

[33]Kajiwara, Y.; Ogino, T.; Yasuda, S.; Takamuku, Y.; Murata, T.; Kinoshita, M. Physical origins of remarkable thermostabilization by an octuple mutation for the adenosine A_{2a} receptor. *Chem. Phys. Lett.* **2016**, *657*, 119−123.

[34]Mitchell, J. B. O.; Price, S. L. *Chem. Phys. Lett.* On the relative strengths of amide…amide and amide…water hydrogen bonds. **1991**, *180*, 517−523.

[35]Hino, T.; Arakawa, T.; Iwanari, H.; Yurugi-Kobayashi, T.; Ikeda-Suno, C.; Nakada-Nakura, Y.; Kusano-Arai, O.; Weyand, S.; Shimamura, T.; Nomura, N.; et al. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* **2012**, *482*, 237−240.

[36]Morgulis, A.; Coulouris, G.; Raytselis, Y.; Madden, T. L.; Agarwala, R.; Schäffer, A. A. Database indexing for production MegaBLAST searches. *Bioinformatics* **2008**, *15*, 1757−1764.

[37]Mobarec, J. C.; Sanchez, R.; Filizola, M. Modern Homology Modeling of G-Protein Coupled Receptors: Which Structural Template to Use? *J. Med. Chem*. **2009**, *52*, 5207−5216.

[38]Krogh, A.; Larsson, B.; von Hejine, G.; Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol*. **2001**, *305*, 567−580.

[39]Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.

[40]Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, 4, 187−217.

[41]Feig, M.; Karanicolas, J.; Brooks, C. L., III. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.

Chapter 5

Hot-Spot Residues to be Mutated Common in G Protein-Coupled Receptors of Class A: Identification of Thermostabilizing Mutations Followed by Determination of Three-Dimensional Structures for Two Example Receptors

5.1. Introduction

G protein-coupled receptors (GPCRs), seven helical transmembrane (TM) proteins, are responsible for signal transduction pathways and indispensable to the sustenance of life.[1] Since their malfunctioning causes a diversity of diseases, they are important therapeutic targets.[2] It is known that approximately 800 GPCRs are encoded in the human genome.[3] The crystallization followed by the determination of three-dimensional (3D) structure of a GPCR is a crucially important task in biochemistry, structural biology, and drug design. However, it has been hindered by the inherent instability of a GPCR in detergent, and the 3D structures of most of GPCRs remain unresolved. An amino-acid mutation is known to possibly enhance its thermostability,[4–6] and this enhancement usually leads to higher stability in detergent as well. The prevailing approach toward identifying a thermostabilizing mutation is alanine (Ala) scanning mutagenesis[7–9] in which every residue is mutated to Ala (Ala is mutated to leucine), the resulting change in stability is experimentally examined, and mutations leading to relatively higher stabilities are chosen. This approach suffers the drawback that the mutational space is quite limited and still the experimental burden is often unacceptably heavy. Methods which do not rely on experiments were reported for the identification of stabilizing mutations.[10–14] In particular, our method[13,14] possesses the following features: It is based on statistical thermodynamics and the physical origins of the enhanced stability are always unambiguous; a free-energy function (FEF) comprising energetic and entropic components is developed and the entropic effect originating from the thermal motion of hydrocarbon groups constituting nonpolar chains of the lipid bilayer is incorporated in the FEF as an essential physical factor; and all of the possible mutations can be examined (i.e., the whole mutational space can be explored) with minor computational effort. There is no other method sharing all of these features. The prediction performance of our method was shown to be considerably high for the adenosine A_{2a} receptor ($A_{2a}R$) in the inactive state.[13,14] In the application of any method to another GPCR, however, the entire examination process may be required all over again.

In this study, we develop a theoretical strategy by which thermostabilizing mutations of many different GPCRs can be identified at the same time: On the basis of the result of examination of the mutational space for a reference GPCR, the examination for many other GPCRs can be accomplished quite efficiently. As an important first step, we illustrate the strategy for three GPCRs

of Class A in the inactive state. The principal conclusion drawn is that a mutation of the residue at a position of N_{BW}=3.39 leads to substantially higher stability for significantly many GPCRs of Class A in the inactive state. Moreover, mutating it to arginine (Arg; R) or lysine (Lys; K) is most recommended. N_{BW} is the Ballesteros-Weinstein (BW) number that is the enumeration of a GPCR transmembrane residue in the format, *x.yz*, where *x* is the helix number, and *yz* is the position of the residue relative to that of the most conserved residue designated as *x*.50 in the helix. The same BW number signifies the same position in the amino-acid sequence. GPCRs are usually grouped into 6 classes in terms of sequence homology and functional similarity: Classes A–F. Among them, Class A forms the largest GPCR family, comprising nearly 85% of GPCRs. Unfortunately, the 3D structures have been determined for only ~4% of GPCRs of Class A. The three GPCRs referred to above are A_{2a}R, muscarinic acetylcholine receptor 2 (M2R), and prostaglandin E receptor 4 (EP4). M2R is one of the five muscarinic acetylcholine receptor subtypes (M1-M5) and essential for the physiologic control of cardiovascular function.[15] EP4 is one of the four receptor subtypes for prostaglandin E (EP1-EP4). EP4 signaling is closely related to various physiological and pathophysiological actions such as carcinogenesis, bone remodeling, and cardiac hypertrophy.[16] Resolving the 3D structures of M2R and EP4 provides deep insights into the development of their subtype-selective ligands. The crystal structures of A_{2a}R and of M2R with a non-selective inverse agonist are available, but the crystal structure of EP4 has not been reported yet. As a product from this study, highly thermostabilizing mutations are found out for M2R and EP4, leading to the determination of new 3D structures for M2R and EP4 (structures of M2R with a new antagonist and of EP4 in complex with an antagonist and a functional antibody; the details including the structural data are reported in separate articles[17,18]).

We pointed out the possibility of existence of “key residues” and “hot-spot residues” in our earlier works[13,14] dealing with only A_{2a}R. In this study, their existence is demonstrated and they are more clearly defined. A key residue is the residue to be mutated in the sense that many of its mutations are highly thermostabilizing. A GPCR possesses multiple key residues. A hot-spot residue is the key residue designated by N_{BW}, which is common in different GPCRs. We show that the residue at a position of N_{BW}=3.39 is a hot-spot residue for significantly many GPCRs of Class A in the inactive state in the following manner: First, the key residues in A_{2a}R determined using the entropic component of our FEF and their corresponding BW numbers are listed up; second, those in M2R and EP4 are determined in the same manner and their corresponding BW numbers are checked; it is then found that the BW number 3.39 appears for all of the key residues in A_{2a}R, M2R, and EP4. The models of wild-type and mutant structures for A_{2a}R, M2R, and EP4 are constructed on the basis of the crystal structures for the inactive state (for EP4 whose crystal structure is unavailable, we construct the models as explained in a later section). Since the energetic component of our FEF was rather uncertain,[19] it is improved in this study. We then select S110R and G106R for M2R and EP4 (S110 in M2R and G106 in EP4 share the same BW number, 3.39), respectively, because these mutations are predicted to provide the highest stabilizations by the improved FEF. We

experimentally verify the considerably large stabilizations by these mutations, leading to successful determination of the 3D structures mentioned above. When the stabilization followed by the structure determination is undertaken for another GPCR of Class A in the inactive state, all we have to do is to examine the stability changes arising from the 19 mutations of the hot-spot residue using the improved FEF and select the mutation which is predicted to provide the highest stabilization.

Our strategy should be applicable to GPCRs of Class A in the active state as well. A hot-spot residue is found out by following the same procedure. As the only change to be made, it is started by constructing their structural models on the basis of the crystal structures for the active state. This study may lead to the breakthrough that the 3D structures of many GPCRs of Class A in the inactive and active states can be solved for the first time in succession.

5.2. Theoretical method

5.2.1. Basic stance in developing theoretical method

Our method was developed so that the following two contradicting requirements could be met: The identification of thermostabilizing mutations is to be made by examining all of the possible mutations (the number of mutations is quite large) with minor computational effort; and the performance needs to be sufficiently high. Hence, simplifying treatments were logically employed and a special technique was introduced. Still, the essential physics of the structural stability of a membrane protein was retained. More details were described in our earlier publications,[13,14,19,20] and some of the important aspects of our method are capitulated in Chapter 1.

5.2.2. Free-energy function

We outline the free-energy function (FEF) employed in our method.[13,14,19,20] The FEF is expressed by

$$F/(k_B T_0)=\Lambda/(k_B T_0)-TS/(k_B T_0) \quad (5.1)$$

where T is the absolute temperature, T_0=298 K, k_B is the Boltzmann constant, and Λ and $-TS$ are the energetic and entropic components of F, respectively. F, Λ, and $-TS$ are normalized by $k_B T_0$, and T is set at T_0 in this study. It is assumed that the thermostability of a GPCR is governed by that of its transmembrane (TM) region. Hence, only the TM region is considered hereafter.

$|\Lambda|$ ($\Lambda<0$) and $-S$ ($-S>0$) are strongly dependent on the protein structure. Λ is the protein intramolecular energy. $-S$ represents the magnitude of entropic loss of the solvent upon protein insertion. Here, for a membrane protein, the hydrocarbon groups constituting nonpolar chains of the lipid bilayer act as the solvent. As the protein structure becomes more compact in the sense that the backbone and side chains are more efficiently (closely) packed, the excluded volume generated by

the protein decreases, the total volume available to the translational displacement of "solvent particles" increases, and the translational, configurational entropy of the solvent becomes higher (see Figure 1.2 in Chapter 1). In other words, $-S$ becomes smaller for a more compact structure. Larger $|\Lambda|$ and smaller $-S$ are more favored: The protein is driven to take the structure minimizing F. A great advantage is that the calculation of F per protein structure can be finished in ~1 sec on a standard workstation.

For a physical quantity Z, we define ΔZ as "Z of the folded structure" − "Z of the unfolded structure": ΔZ denotes the change in Z upon protein folding. $\Delta F/(k_B T_0)$ is written as

$$\Delta F/(k_B T_0)=\Delta\Lambda/(k_B T_0)-\Delta S/k_B. \quad (5.2)$$

5.2.3. Change in energetic component upon protein folding

A gain of protein intramolecular van der Waals attractive interactions occurs upon protein folding. However, it undergoes a loss of protein-solvent van der Waals attractive interactions. The gain and the loss are rather compensating. On the other hand, a gain of intramolecular electrostatic interaction energy is much larger than the loss of protein-solvent electrostatic interaction energy undergone, because the solvent is nonpolar. As a consequence, protein folding is accompanied by considerable lowering of electrostatic interaction energy. To this lowering, the energy decrease arising from the formation of protein intramolecular hydrogen bonds (IHBs) should make a pivotal contribution. Therefore, we focus on the formation of IHBs in the calculation of $\Delta\Lambda$. Any structure can be chosen as the reference one for calculating $\Delta\Lambda$. In our method, the reference structure is a fully extended structure possessing no IHBs.

An energy decrease of $E\leq 0$ is considered when an IHB is formed. E is dependent on the distance between centers of H and the acceptor, d. For $d\leq 1.5$ Å, E is set at E^*. As d increases from 1.5 Å, $|E|$ is assumed to decrease lineally and become zero for $d\geq 3.0$ Å. That is, an H-acceptor pair with $d<3.0$ Å forms an IHB. We examine all the donors and acceptors for backbone-backbone, backbone-side chain, and side chain-side chain IHBs of the folded structure within the TM region. (The examination is made using the criteria proposed by McDonald and Thornton[21] with the modification that the maximum distances between H and the acceptor and between the donor and the acceptor are 3.0 Å and 4.4 Å, respectively.) Λ calculated is expressed as $xN_{IHB}E^*$ where N_{IHB} denotes the number of IHBs and $0<x\leq 1$. In Chapters 2–4, E^* was set at $-14k_B T_0$ and the effect of changing this value was discussed. Though the performance of our method is not largely influenced by the value of E^* set,[19] we optimize E^* for this study as explained in "Improvement of Energetic Component".

5.2.4. Change in entropic component upon protein folding

Protein folding can be considered by the two-stage model.[22] In the first stage, individual α-helices of the protein are separately stabilized within the lipid bilayer. In the second stage, the side-to-side association of these helices, which accompanies the close packing of side chains, takes

place, leading to the formation of the folded structure. The solvent-entropy gain in the second stage is the dominant contributor to ΔS. Therefore, ΔS can be approximated by

$$-\Delta S=-S_{\mathrm{compact}}-\Sigma(-S_{\mathrm{separated}}) \tag{5.3}$$

$-S_{\mathrm{compact}}$ is the magnitude of entropic loss upon insertion of the associated α-helices, and $-S_{\mathrm{separated}}$ is that upon insertion of a separated α-helix and the summation is taken over all of the separated α-helices.

$-S_{\mathrm{compact}}$ and $-S_{\mathrm{separated}}$ are calculated using a hybrid of an integral equation theory (IET)[23] and the morphometric approach (MA),[24,25] which was developed by Kinoshita and coworkers. The former is a statistical-mechanical theory for fluids and the latter is required for the treatment of a large protein possessing complex polyatomic structure with minor computational effort. In the MA, the polyatomic structure is represented by the four geometric measures, the excluded volume, solvent-accessible surface (SAS) area, and integrated mean and Gaussian curvatures of the SAS. S/k_{B} is expressed by the linear combination of the four geometric measures, which is referred to as "morphometric form". The four coefficients in the morphometric form are dependent only on the solvent species and its thermodynamic state, and they are calculated by the IET applied to isolated hard-sphere solutes with various diameters in advance. (The model of the solvent formed by nonpolar chains of lipid molecules and the IET-MA hybrid were described in our earlier publications[13,14,19,20] and summarized in Chapter 1.)

5.2.5. Protocol for comparing the thermostabilities of the wild type and a mutant

We calculate $-\Delta\Delta S$ defined by

$$-\Delta\Delta S=-\Delta S_{\mathrm{M}}-(-\Delta S_{\mathrm{W}}) \tag{5.4}$$

where the subscripts "M" and "W" denote "mutant" and "wild type", respectively. Negative $-\Delta\Delta S$ implies that the mutant is more stable than the wild type in terms of the entropic component. We also calculate $\Delta\Delta\mathit{\Lambda}$ defined by

$$\Delta\Delta\mathit{\Lambda}=\Delta\mathit{\Lambda}_{\mathrm{M}}-\Delta\mathit{\Lambda}_{\mathrm{W}}. \tag{5.5}$$

Negative $-\Delta\Delta\mathit{\Lambda}$ implies that the mutant is more stable than the wild type in terms of the energetic component. When $\Delta\Delta F/(k_{\mathrm{B}}T_0)$ given by

$$\Delta\Delta F/(k_{\mathrm{B}}T_0)=\Delta\Delta\mathit{\Lambda}/(k_{\mathrm{B}}T_0)-\Delta\Delta S/k_{\mathrm{B}} \tag{5.6}$$

is negative, the mutant is considered more stable than the wild type.

5.2.6. Construction of structure model for the wild type

The crystal structure of M2R in the inactive state with the non-selective inverse agonist, QNB((3R)-1-azabicyclo[2.2.2]oct-3-yl hydroxy(diphenyl)acetate), has already been determined but that of EP4 is still unknown. For M2R, we employ the experimentally determined crystal structure (PDB code: 3UON).[15] T4 lysozyme (T4L), which is incorporated in the third intracellular loop (ICL3), and the non-selective inverse agonist are removed. The missing loop structure is then added using the MODELLER program.[26] In this process, residues from P233 to P372 in the ICL3 are removed for further accelerating the calculation, and S232 is connected with A373. Since we consider only the TM region, the result obtained is not affected by the removal.

For EP4, the wild-type structure model is constructed in accordance with the homology modeling. Adopting $A_{2a}R$ in the inactive state in which T4L is incorporated (PDB code: 3EML)[27] as the template and applying the amino-acid sequence alignment previously suggested,[28] we generate 200 candidate models by means of the MODELLER program.[26] $A_{2a}R$ possesses a TM region identity of ~17% with EP4. (As in the case of M2R, residues from M218 to S259 in the ICL3 are removed, and F217 is connected with F260.) It has been shown that our FEF is a very good measure of the appropriateness of a candidate model.[14] More specifically, the model giving ΔF the lowest value is selected as the best one. In Figure 5.1, we plot $\Delta F/(k_B T_0)$ against the model number: $\Delta F/(k_B T_0)$ takes the lowest value for model 67.

The Lennard-Jones potential energies of the wild-type structure models thus obtained for M2R and EP4 take significantly large, positive values due to the slight overlaps of protein atoms. Such overlaps can readily be removed by the local minimization of protein intramolecular energy using the CHARMM biomolecular simulation program[29] through the multiscale modeling tools in structural biology (MMTSB) program.[30] The CHARMM22[31] is employed for the force-field parameters. After the minimization, each structure is switched to a set of fused hard spheres for calculating the solvation entropy. Since all that matters is the removal of the overlaps of protein atoms with retaining the original structure, the details of the minimization procedure are not very important. The determination of the TM region by the web server of TMDET[32] is then performed for each resultant structure.

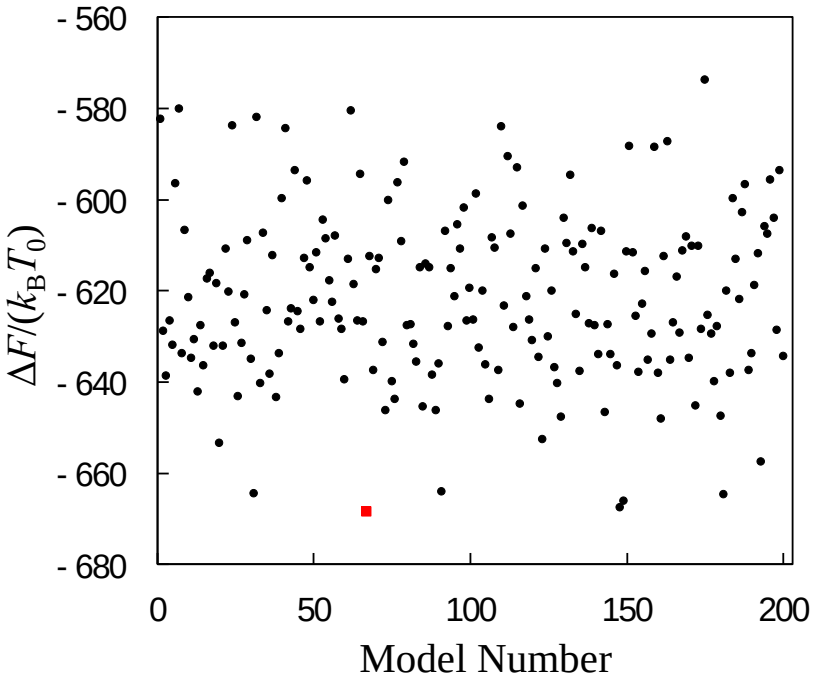


Figure 5.1. Values of $\Delta F/(k_B T_0)$ for 200 candidate models of the wild-type structure of EP4. ΔF is the free-energy change upon protein folding, T_0=298 K, and k_B is the Boltzmann constant. The red square denotes the model which gives $\Delta F/(k_B T_0)$ the lowest value.

5.2.7. Construction of structure model for a mutant

Starting from the wild-type structure model for M2R or EP4, we generate a total of 10 structures by employing the MODELLER program[26] on the condition that it can move only the residues which are present within a distance of 5 Å from the mutated residue. The slight overlaps of protein atoms are removed by the local minimization of protein intramolecular energy mentioned above. By this minimization, the structural change induced by the mutation extends to the whole protein. The structure model for a mutant is formed by the 10 structures. A thermodynamic quantity for a mutant is obtained as the average of its values calculated for the 10 structures. Since a great number of mutants have to be examined, the number of structures generated should be kept sufficiently small (10 is a good choice).

5.2.8. Improvement of energetic component

The value of E^* calculated using quantum chemistry for the formation of a hydrogen bond (HB) in gas phase is $-10k_B T_0$ (T_0=298 K).[33] If an H-acceptor pair was in vacuum, E^* could be set at $-10k_B T_0$. However, it is in the environment where atoms with positive and negative partial charges are present. E^* can be regarded as the potential of mean force between the pair in such environment.

In this respect, $|E^*|$ should be significantly smaller than $10k_BT_0$ (factor 1). For example, it has been suggested that E^* be $\sim -3.4k_BT_0$.[34] On the other hand, protein folding is accompanied by the lowering of interaction energy, but this lowering arises from not only the intramolecular hydrogen bonding but also the other electrostatic interaction. When this lowering is incorporated in E^*, the resulting $|E^*|$ should be larger than $10k_BT_0$ (factor 2). Further, the energy decrease brought by the formation of a HB between two formamide molecules in a nonpolar solvent, which is considered to mimic the environment within protein interior, is $-14k_BT_0$.[35] This value implicitly includes the contribution from entropic gain of the nonpolar solvent (i.e., that of protein atoms) arising from the H-acceptor contact forming a HB. Thus, the evaluation of E^* is rather uncertain.

For this study, we optimize E^* on the basis of the competition of factors 1 and 2 which should be most important. First, the electrostatic interaction energies in the folded and reference structures of M2R (they are denoted by $E_{\text{elec,folded}}$ and $E_{\text{elec,reference}}$, respectively) are calculated by means of the CHARMM program. The TM region determined using the crystal structure is applied to the two structures. The reference structure (i.e., a fully extended structure) within the TM region is generated using the TINKER program package[36] with the dihedral angles (φ, ψ, and ω) set at (180°, 180°, and 180°). $E_{\text{elec,folded}}-E_{\text{elec,reference}}$, which is calculated to be −341.2 kcal/mol, represents the lowering of electrostatic interaction energy *including* the energy decrease arising from formation of IHBs. On the other hand, $\Delta\Lambda=\Lambda_{\text{folded}}-\Lambda_{\text{reference}}=\Lambda_{\text{folded}}$ ($\Lambda_{\text{reference}}=0$ because the reference structure possesses no IHBs) is calculated to be $121.1E^*$. Equating −341.2 kcal/mol with $121.1E^*$ yields $E^*=-2.818$ kcal/mol$\sim -5k_BT_0$. We employ the setting, $E^*=-5k_BT_0$, in all of the calculations in this study. Thus, for the protein intramolecular interaction energy, the effect of electrostatic interaction other than hydrogen bonding is implicitly incorporated in E^*. (We have verified that the qualitative aspects of our conclusions drawn in Chapters 2–4 are not altered by this improvement.)

5.2.9. Nomination of Key Residues and Identification of Hot-Spot Residues

Many of the 19 mutations of a key residue lead to substantially higher thermostabilities. A hot-spot residue is the key residue designated by the BW number and common in significantly many different GPCRs. An important finding in an earlier study by Kinoshita and coworkers for water-soluble proteins is that a mutation leading to substantially higher stability is always characterized by a large increase in the water-entropy gain upon protein folding.[37] By the analogy of this finding, we use the entropic component rather than the FEF in nominating the key residues. The criterion for the nomination is the following (X and M are positive values set by us): Count the number of the mutations of a residue which leads to $-\Delta\Delta S/k_B<-X$; when this number is larger than or equal to M, the residue is nominated as one of the key residues. Setting X or M at too large a value will result in only very few key residues. The nomination of the hot-spot residues consists of the following five steps:

(1) Determine the key residues for $A_{2a}R$. This was already performed in Chapters 2 and 4 by setting

X and *M* at 10 and 4, respectively. The key residues are T88(3.36), S91(3.39), and N181(5.42). Here, the value in the parentheses is the BW number.
(2) Determine the key residues for M2R. Adjust the values of *X* and *M* so that the number of the key residues takes a reasonable value (e.g., in the range from 3 to 10). Clarify the BW number of each key residue.
(3) Determine the key residues for EP4. Adjust the values of *X* and *M* so that the number of the key residues takes a reasonable value. Clarify the BW number of each key residue.
(4) Check the key residues and the corresponding BW numbers obtained in steps (1)−(3). Find the key residues which share the same BW number. The residue with this BW number is the hot-spot residue.
(5) For the hot-spot residue in M2R or EP4, evaluate the thermostabilities of the possible 19 mutations using $\Delta\Delta F/(k_B T_0)$, and select the mutation for which $\Delta\Delta F/(k_B T_0)$ takes the lowest value.

5.3. Results and discussion

5.3.1. Nomination of key residues and identification of hot-spot residues

In Chapters 2 and 4, we nominated the key residues for $A_{2a}R$ in the inactive state as follows: T88(3.36), S91(3.39), and N181(5.42). *X* and *M* were set at 10 and 4, respectively. The value in the parentheses is the BW number. It was experimentally corroborated that the denaturation temperature T_m becomes higher than that of the wild type by ~7 °C for S91K, ~4 °C for S91R, ~7 °C for T88E, and ~4 °C for T88K despite that they are single mutations.[13] Further, the double mutation T88E−S91R leads to an increase in T_m of ~12 °C.[13]

Figure 5.2 shows the values of $-\Delta\Delta S/k_B$ calculated for all of the possible mutations for M2R or EP4. It is observed that the number of stabilizing mutations and the degree of stabilization vary largely from residue to residue. For EP4, setting *X* and *M* at 10 and 4, respectively, results in the following nomination of the key residues: P24(1.39), L57(2.42), G60(2.45), S73(2.58), G106(3.39), A112(3.45), S114(3.47), G190(5.43), and C202(5.55). For finding sufficiently many key residues for M2R, on the other hand, *X* is to be set at a smaller value: With the setting, *X*=5 and *M*=4, the nominated key residues are S64(2.45), S110(3.39), V111(3.40), and P198(5.50). The residues underlined share the same BW number 3.39. Therefore, the residue with N_{BW}=3.39 is a hot-spot residue common in the three GPCRs and probably in significantly many other GPCRs of Class A.

G60 in EP4 and S64 in M2R share the same BW number 2.45. In particular, for EP4 all of the 19 mutations of G60 lead to negative values of $-\Delta\Delta S/k_B$, and one of these mutations gives $-\Delta\Delta S/k_B$ the lowest value in Figure 5.2. However, the residue with N_{BW}=2.45 is not a key residue for $A_{2a}R$. Therefore, we identify the residue with N_{BW}=3.39 as a hot-spot residue.
The sequence aliment of $A_{2a}R$, M2R, and EP4 is shown in Figure 5.3. The most conserved residues at the *x*-th helix designated as *x*.50 and the key and hot-spot residues are highlighted in the figure.

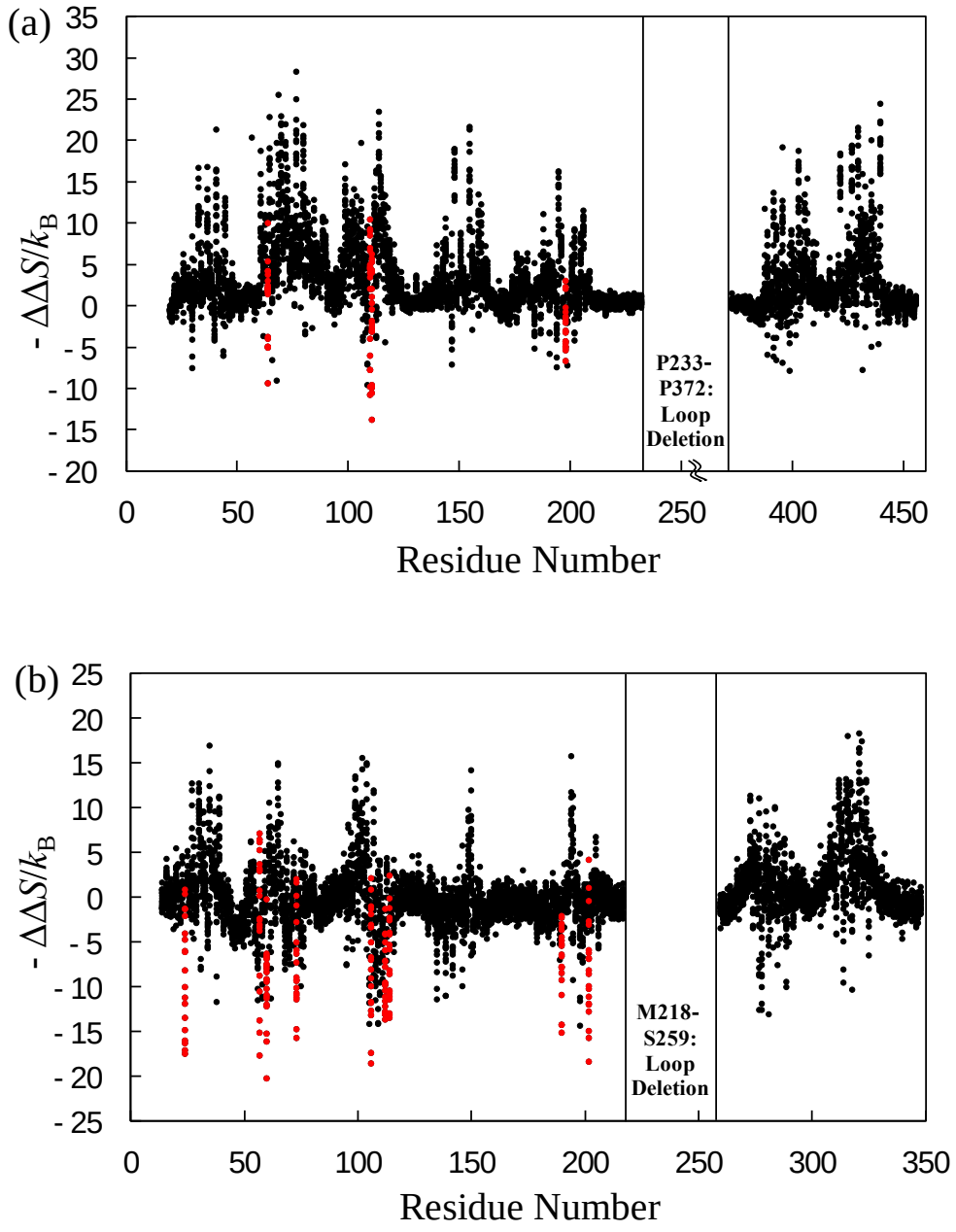


Figure 5.2. Values of $-\Delta\Delta S/k_B$ for all of the possible mutations for M2R (a) and for EP4 (b). In (a), residues from P233 to P372 are removed and S232 is connected with A373. In (b), residues from M218 to S259 are removed and F217 is connected with F260. ΔS is the solvent-entropy gain upon protein folding, $\Delta\Delta S$= "ΔS of a mutant" – "ΔS of the wild type", and k_B is the Boltzmann constant. The solvent is formed by the hydrocarbon groups constituting nonpolar chains of the lipid bilayer. Negative and positive values of $-\Delta\Delta S/k_B$ signify that the mutation enhances and lowers the thermostability in terms of the entropic component, respectively. Red circles denote those of $-\Delta\Delta S/k_B$ arising from the 19 mutations for a key residue nominated.

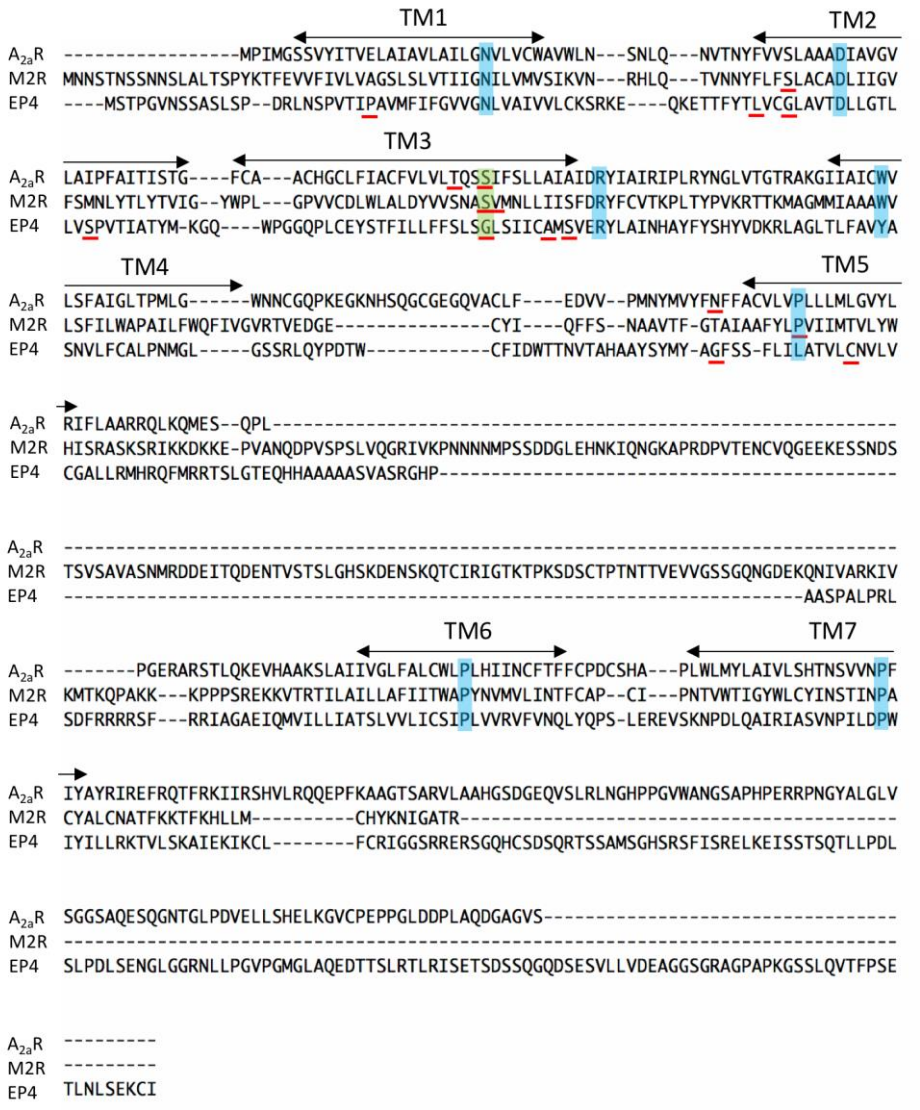


Figure 5.3. Sequence alignment of $A_{2a}R$, M2R, and EP4. The helices (TM1–TM7) within the TM region determined for M2R using the web server of TMDET[32] are indicated. The key residues for each GPCR are underlined in red. The hot-spot residues with N_{BW}=3.39 are marked in light green. The most conserved residues at the *x*-th helix designated as *x*.50 are marked in light blue.

5.3.2. Selection of the best mutation for hot-spot residue

We calculate the values of $\Delta\Delta F/(k_B T_0)$ for the 19 mutations of the hot-spot residue in M2R or EP4. The results are collected in Table 5.1. For M2R, the mutations to R, K, Y, F, and H lead to significantly enhanced stabilities in this order. For EP4, those to R, Y, K, Q, F, N, and H results in such stabilities in this order, and M2R and EP4 share R, K, Y, F, and H as the residues to which the hot-spot residue are to be mutated. $\Delta\Delta F/(k_B T_0)$ takes the lowest value when the residue is mutated to arginine (Arg) for both of M2R and EP4. For S110R, the energetic and entropic components contribute to the stability almost equally. For G106R, on the other hand, the contribution to the

stability from the entropic component is much larger than that from the energetic one. The positive charge in the side chain of Arg and its geometric feature play essential roles in the energetic and entropic stabilizations, respectively. As explained in "Change in Entropic Component upon Protein Folding", the side-to-side association of the helices within the TM region, which accompanies the close packing of side chains, increases the total volume available to the translational displacement of "solvent particles". Denoting this increase by ΔV_{TD}, we find that ΔV_{TD} becomes larger by ~$8d_S^3$ (d_S is the diameter of the solvent particles and set at 2.8 Å; see Chapter 1) upon the mutation S110R or G106R. Also, the number of IHBs increases by ~2 upon the mutation S110R or G106R.

In this study, the structural models of M2R and EP4 are constructed on the basis of the crystal structures for the inactive state. It is intriguing that the mutation of the hot-spot residue to Arg (R) is theoretically the most stabilizing for both M2R and EP4 (it is also the best one for $A_{2a}R$ when the improved energetic component is applied). The mutation to Lys (K) is also highly stabilizing. We infer that this result is relevant to the allosteric modulation by sodium ion.[38] More detailed discussions are given in a separate article[17] and in 5.5.1.

We construct the mutants, S110R for M2R and G106R for EP4, in experiments and examine their thermostabilities. As a consequence, it is corroborated that the stabilities of both of the two mutants are considerably enhanced. This enhancement leads to the successful determination of a 3D structure of EP4 in complex with the antagonist ONO-AE3-208 and a functional antibody[18] and a 3D structure of M2R with the M2/M4-selective antagonist,[17] AF-DX384((±)-5,11-dihydro-11-{[(2-{2-[(dipropylamino)methyl]-1-piperidinyl}ethyl)amino]carbonyl}-6H-pyrido(2,3-b)(1,4)-benzodiazepine-6-one). The details of the experimental results are described in separate articles.[17,18] (Brief description on the thermostabilizing mutations is given in 5.5.2.)

Table 1. Values of $-\Delta\Delta S/k_B$, $\Delta\Delta\Lambda/(k_BT_0)$, and $\Delta\Delta F/(k_BT_0)$ for S110 in M2R (a) and for G106 in EP4 (b) Mutated to Each of 19 Residues[a]

(a)

	$-\Delta\Delta S/k_B$	$\Delta\Delta\Lambda/(k_BT_0)$	$\Delta\Delta F/(k_BT_0)$
A	6.85	–0.14	6.71
R	–10.95	–11.05	–22.00
N	4.48	5.66	10.14
D	9.11	0.02	9.13
C	8.39	–1.13	7.264
Q	3.30	–3.04	0.26
E	8.88	–0.66	8.21
G	10.28	4.14	14.41

H	–4.14	–2.76	–6.90
I	4.83	1.58	6.41
L	3.87	1.02	4.89
K	–7.89	–9.75	–17.65
M	1.85	2.43	4.27
F	–6.14	–1.18	–7.32
P	4.64	4.59	9.22
T	4.38	3.27	7.65
W	–7.94	3.62	–4.32
Y	–10.00	–4.65	–14.65
V	6.57	0.51	7.08

(b)

	$-\Delta\Delta S/k_B$	$\Delta\Delta \mathcal{A}/(k_B T_0)$	$\Delta\Delta F/(k_B T_0)$
A	–3.40	–0.83	–4.23
R	–17.46	–9.33	–26.78
N	–7.06	–4.36	–11.42
D	1.99	2.49	4.48
C	–5.14	–1.72	–6.86
Q	–11.81	–2.98	–14.79
E	–3.21	5.96	2.75
H	–10.04	–0.72	–10.76
I	–6.78	3.14	–3.65
L	–8.12	1.00	–7.12
K	–11.90	–5.34	–17.24
M	–9.10	0.79	–8.31
F	–13.29	0.15	–13.14
P	0.78	5.32	6.10
S	–2.01	1.01	–1.00
T	–3.47	–0.32	–3.78
W	–12.79	6.19	–6.60
Y	–18.62	–3.02	–21.64
V	–1.44	0.98	–0.46

[a] T_0=298 K and k_B is the Boltzmann constant. In the first column, "A", for example, represents the mutation to Ala (A): S110A in (a) and G106A in (b). The mutation to Arg (R) underlined gives the

most enhanced thermostability.

5.3.3. Comparison between wild-type structure model constructed and crystal structure experimentally determined for EP4

This section is prepared to make our conclusion even more convincing. In Figure 5.4, the wild-type model structure for EP4 constructed using the homology modeling is compared with the crystal structure of the G106R mutant experimentally determined as an achievement of this study. Though the two structures are somewhat different in the first and second helices, they are quite similar to each other on the whole. Especially, the backbone conformation near G106 in the wild-type structure and that near R106 (this residue arises from the mutation G106R for the wild type) in the crystal one are almost indistinguishable. This result is suggestive that the wild-type structure model is reasonably constructed by the homology modeling and our FEF is very useful in selecting the best model from among a number of candidate models generated. Further, it should be noted that the calculation result is successful despite the difference between the two structures mentioned above.

Figure 5.5 shows the values of $-\Delta\Delta S/k_B$ calculated for all of the possible mutations for EP4 with the experimentally determined crystal structure. It is found that mutating R106 to any other residue causes a larger entropic loss upon protein insertion into the solvent, i.e., the hydrocarbon groups constituting nonpolar chains of the lipid bilayer. It is also found that $\Delta\Delta\Lambda/(k_BT_0)$ and $\Delta\Delta F/(k_BT_0)$ for all of the possible mutations show qualitatively the same tendency. Namely, mutating R106 to any other residue gives rise to higher values of $\Delta\Delta\Lambda/(k_BT_0)$ and $\Delta\Delta F/(k_BT_0)$. In the experimentally determined crystal structure, the presence of R106 plays critical roles in stabilizing the protein structure in terms of the FEF and its energetic and entropic components, which is consistent with the calculation result that the mutation G106R leads to considerably higher thermostability. We have performed qualitatively the same analysis for M2R: Mutating R110 (this residue arises from the mutation S110R for the wild type) to any other residue causes a higher value of $\Delta\Delta F/(k_BT_0)$, which provides another manifestation of considerably higher stability conferred upon M2R by the mutation S110R.

To further check the robustness of our conclusion, we start from the crystal structure of the G106R mutant experimentally determined for EP4, replace R106 by glycine (Gly), and employ the resultant structure as the wild-type model structure. We then calculate $-\Delta\Delta S/k_B$, $\Delta\Delta\Lambda/(k_BT_0)$, and $\Delta\Delta F/(k_BT_0)$ for the 19 mutations of G106. The mutant structure is constructed as described in "Construction of Structure Model for a Mutant". In comparison to the values of the three thermodynamic quantities in Table 5.1(b), those in Table 5.2 tend to exhibit upward shifts (i.e., shifts in positive directions) on the whole. However, the two results in these two tables share qualitatively the same characteristics: The mutations to R, K, H, Y, and N lead to significantly enhanced stabilities in this order; G106R is the most stabilized mutant; and G106K is also

significantly stabilizing. The important aspects are captured in both of Tables 5.1(b) and 5.2. Taken together, the combination of the homology modeling and our FEF is very useful in constructing the wild-type model structure.

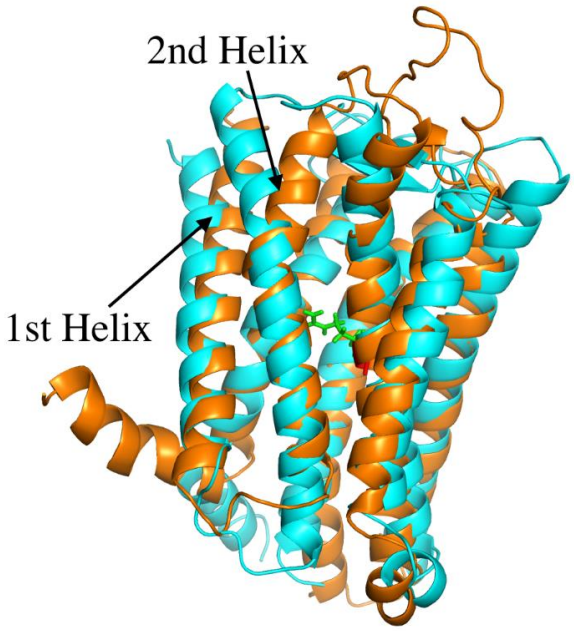


Figure 5.4. Visual comparison between the model structure for EP4 constructed using the homology modeling (orange) and the crystal structure experimentally determined as an achievement of this study (light blue). G106 in the model structure and R106 in the crystal one are shown by red and green stick representations, respectively.

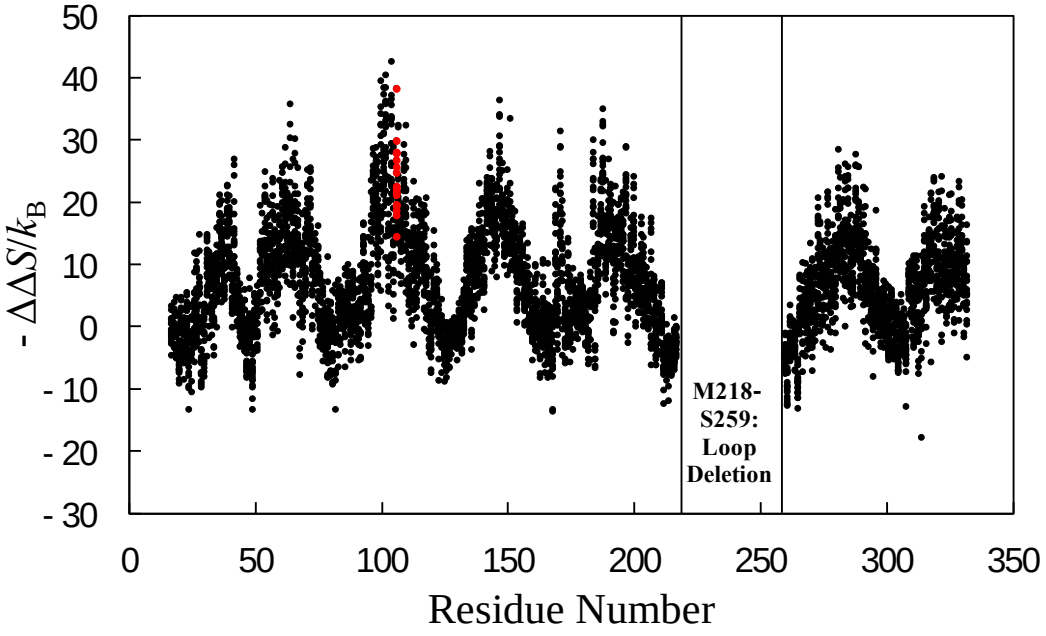


Figure 5.5. Values of $-\Delta\Delta S/k_B$ for all of the possible mutations for the crystal structure experimentally determined for EP4 as an achievement of this study. Residues from M218 to S259 are removed and F217 is connected with F260. ΔS is the solvent-entropy gain upon protein folding, $\Delta\Delta S$= "ΔS of a mutant" – "ΔS of the wild type", and k_B is the Boltzmann constant. Red circles denote those of $-\Delta\Delta S/k_B$ arising from the 19 mutations for R106.

Table 5.2. Values of $-\Delta\Delta S/k_B$, $\Delta\Delta\Lambda/(k_BT_0)$, and $\Delta\Delta F/(k_BT_0)$ for G106 in EP4 Mutated to Each of 19 Residues[a]

	$-\Delta\Delta S/k_B$	$\Delta\Delta\Lambda/(k_BT_0)$	$\Delta\Delta F/(k_BT_0)$
A	17.51	1.98	19.50
R	–8.50	–14.48	–22.99
N	–1.54	–3.84	–5.38
D	3.98	–3.68	0.30
C	4.91	–0.47	4.44
Q	5.37	–4.77	0.60
E	6.85	–7.52	–0.66
H	–2.63	–5.44	–8.07
I	–0.84	–1.42	–2.26
L	–3.12	–2.09	–5.21
K	–1.67	–13.55	–15.22
M	0.74	–2.10	–1.37
F	–2.83	–1.89	–4.72
P	4.29	–1.38	2.91
S	7.34	–4.21	3.14
T	6.71	–4.22	2.50
W	–7.60	2.32	–5.28
Y	–0.44	–5.78	–6.22
V	0.89	–0.42	0.47

[a]See Table 5.1 for the notation. This table should be compared to Table 5.1(b). The mutation to Arg (R) underlined gives the most enhanced thermostability.

5.4. Conclusions

Using the free-energy function (FEF) whose energetic component is improved, we have developed a theoretical strategy by which thermostabilizing mutations of many different G protein-coupled receptors (GPCRs) can be considered at the same time. "Key residues" and "hot-spot residues" are clearly defined. Many of the 19 mutations of a key residue lead to substantially higher stabilities. A hot-spot residue is the key residue designated by the Ballesteros-Weinstein (BW) number N_{BW} and common in significantly many different GPCRs. On the basis of the idea that a mutation leading to substantially higher stability is characterized by a large increase in the solvent-entropy gain upon protein folding (the solvent is formed by the hydrocarbon groups constituting nonpolar chains of lipid molecules), the number of the mutations of a residue which leads to $-\Delta\Delta S/k_{\mathrm{B}}<-X$ is counted; when it is larger than or equal to M, the residue is nominated as one of the key residues. (S is the entropic component of the FEF, ΔS denotes the change in S upon protein folding, $\Delta\Delta S$= "ΔS for a mutant" – "ΔS for the wild type", and k_{B} is the Boltzmann constant.) The strategy has been illustrated for three GPCRs of Class A, the adenosine A_{2a} receptor (A_{2a}R), muscarinic acetylcholine receptor 2 (M2R), and prostaglandin E receptor 4 (EP4), in the inactive state. The models of wild-type and mutant structures are constructed on the basis of the crystal structures for the inactive state. More specifically, a hot-spot residue for the GPCRs has been found out in the following manner:

The key residues determined for A_{2a}R are T88(3.36), S91(3.39), and N181(5.42) (X=10 and M=4). Here, the value in the parentheses is the BW number N_{BW}. Those for M2R are S64(2.45), S110(3.39), V111(3.40), and P198(5.50) (X=5 and M=4). Those for EP4 are P24(1.39), L57(2.42), G60(2.45), S73(2.58), G106(3.39), A112(3.45), S114(3.47), G190(5.43), and C202(5.55) (X=10 and M=4). The residues underlined share the same BW number and the residue with N_{BW}=3.39 is a hot-spot residue.

We have then evaluated the thermostabilities of the possible 19 mutations of the hot-spot residue in M2R or EP4 using the FEF and selected the mutation for which $\Delta\Delta F/(k_{\mathrm{B}}T_0)$ (F is the FEF, ΔF denotes the change in F upon protein folding, $\Delta\Delta F$= "ΔF for a mutant" – "ΔF for the wild type", and T_0=298 K) takes the lowest value. The selected mutations for M2R and EP4 are S110R and G106R, respectively. (S110K and G106K are also highly stabilizing for M2R and EP4, respectively.) It has then been corroborated in experiments that S110R and G106R actually lead to considerably enhanced stabilities. The crystallization based on the lipidic cubic phase (LCP) method and the determination of new three-dimensional (3D) structures for EP4[18] and M2R[17] have thus been made possible.

It is intriguing that the mutation of the hot-spot residue to arginine (Arg) is theoretically the most stabilizing for all of A_{2a}R, M2R, and EP4 in the inactive state. That to lysine (Lys) is also highly stabilizing. This can be relevant to the allosteric modulation by sodium ion which is believed

to be common in multiple GPCRs of Class A (see "High Stabilization Brought by Mutating the Hot-spot Residue to Arginine: Possible Relevance to Allosteric Modulation by Sodium Ion"). The residue with N_{BW}=3.39 is conserved in many GPCRs of Class A, and the concept that it is a hot-spot residue should be applicable to them in the inactive state. As a matter of fact, we have already succeeded in thermostabilizing several additional GPCRs of Class A by mutating their residues with N_{BW}=3.39 (the details will be reported after their new 3D structures are solved). If the mutations of the hot-spot residue (including those to Arg and Lys) are not successful for a particular GPCR, stabilizing mutations can be identified by examining those of the key residues: The identification is performed *individually*.

We intend to identify hot-spot residues other than the residue with N_{BW}=3.39 for GPCRs of Class A in the inactive state. This identification can be performed by changing the values of X and M and nominating another set of key residues for each of A_{2a}R, M2R, and EP4. In fact, we have already found additional hot-spot residues, residues with different values of N_{BW}. Interestingly, their mutations are not relevant to the allosteric modulation by sodium ion. (Using our FEF, we have recently identified highly stabilizing mutations unrelated to the allosteric modulation for GPCRs of Class A other than A_{2a}R, M2R, and EP4: The details will be reported after their new 3D structures are solved.) We are now moving on the experimental confirmation of this finding. If multiple hot-spot residues are identified, remarkable enhancement of the thermostability can be implemented by multiple mutations.

There is no reason why our strategy is inapplicable to GPCRs of Class A in the active state. All we have to do is to construct the models of wild-type and mutant structures on the basis of the crystal structures for the active state. Suppose that we wish to thermostabilize multiple target GPCRs of Class A in the active state to solve their new 3D structures. The basic procedure can be summarized as follows. As the reference GPCR, choose a GPCR whose transmembrane (TM) region identities with the target GPCRs are sufficiently high and whose 3D structure is known; nominate the key residues in the reference and target GPCRs; identify a hot-spot residue; evaluate the thermostabilities of the possible 19 mutations of the hot-spot residue using our FEF for each target GPCR; and select the best mutation for which $\Delta\Delta F/(k_B T_0)$ takes the lowest value. We believe that the stability of the mutation selected for each target GPCR is considerably enhanced with very high probability. The hot-spot residue thus identified may be applicable to another target GPCR. The wild-type structure model is constructed using the homology modeling and our FEF is a very good measure of appropriateness of the model: The validity of this approach has been verified for EP4 in this study. In the illustration for EP4, the TM region identity is only ~17%. Nevertheless the stabilizing mutation is successfully identified. Further, it is worthwhile to consider the possibility of applying our strategy to GPCRs of the other classes (B–F). For example, the 3D structures have already been solved for two GPCRs of Class B.[39,40] Therefore, one of them can be treated as the reference GPCR. GPCRs of Class B seem to be good next examples for which our strategy is tested.

In the present stage of our study, only the mutations of residues within the TM region are

considered. It is required that a modification be made to include the mutations of residues outside the TM region, i.e., those immersed in water. We note that a hot-spot residue can exist in water. In the modification, the protein is divided in to two portions: One of them is within the TM region and the other is immersed in water. The FEF and its calculation procedure are applicable to the portion within the TM region in their present forms. For the portion immersed in water, however, both of the energetic and entropic components must be altered. However, we have much experience in the development of the FEF and its energetic and entropic components for water-soluble proteins.[20,37,41,42] When a donor or an acceptor is buried upon protein folding, the break of a donor-water or acceptor-water hydrogen bond occurs. The effect of this energetic penalty must be taken into account in the energetic component. Further, the absolute value of energy decrease due to the formation of an intramolecular hydrogen bond *within a water-soluble protein* should be larger than that in vacuum $-10k_BT_0$ (T_0=298 K).[37,41,42] The four coefficients in the morphometric form must be changed to those pertinent to water at ambient temperature and pressure in calculating the entropic component. For the portion immersed in water, the relative importance of the entropic component over the energetic one increases to a large extent.

Appendix 5-A: High stabilization brought by mutating the hot-spot residue to arginine: Possible relevance to allosteric modulation by sodium ion

In this study, the structural models of M2R and EP4 are constructed on the basis of the crystal structures for the inactive state. It is intriguing that the mutation of the hot-spot residue to Arg (R) is theoretically the most stabilizing for both M2R and EP4 (it is also the best one for $A_{2a}R$ when the improved energetic component is applied). The mutation to Lys (K) is also highly stabilizing. We infer that this result is relevant to the allosteric modulation by sodium ion.

It is conjectured that a partially hydrated sodium ion Na^+, which is bound to the surface of a pocket in the middle of the seven-transmembrane helical domain, plays important roles in stabilizing the inactive state for multiple GPCRs of Class A including $A_{2a}R$.[38] The net surface charge of the pocket should be negative due to the presence of side chains possessing oxygen atoms with negative partial charges (e.g., those of glutamic acid and glutamine),[22] and there is a vacant space within the pocket. When Na^+ is bound to the pocket surface, the space is narrowed by surface-Na^+ electrostatic attractive interaction. This brings not only the energetic stabilization but also the entropic one. The narrower space results in a more compact protein structure and a smaller excluded volume generated by the protein, leading to the entropic stabilization. S91 of $A_{2a}R$, S110 of M2R, or G106 of EP4 is on the pocket surface. Even in the absence of Na^+, when the residue is mutated to the positively charged residue Arg or Lys, qualitatively the same effect comes into play, which stabilizes the inactive state. That is, S110R, S110K, G106R, or G106K mimicks the allosteric modulation by Na^+. Our FEF is capable of reproducing these features.

As a matter of fact, in the crystal structure experimentally determined for the S110R mutant of M2R, it was observed that Arg is accommodated within the pocket as if the sodium ion is replaced by it (a more detailed discussion is given in a separate article[17]).

Appendix 5-B: Brief summary of experimental results for thermostabilities of the M2R and EP4 mutants

For M2R, we compared the wild type and the S110R mutant in terms of the thermostability using a tryptophan fluorescence-detection size-exclusion chromatography–based thermostability assay (FSEC-TS). The non-selective inverse agonist, N-methyl scopolamine (NMS), was bound to each of them. The wild type and the mutant exhibited peak shifts starting at 55°C and at 65°C, respectively, which indicate that the mutant is actually more thermostable. The apparent melting temperatures of the wild type and the mutant determined by analyzing the temperature dependence of the fluorescence intensity were 60.5°C and 65.5°C, respectively. More detailed description is given in a separate article.[17]

Qualitatively the same comparison was made for the wild type and the G106R mutant of EP4 using a FSEC-TS. The ligand, ONO-AE3-208, was bound to each of them. The wild type exhibited a peak shift starting at 40°C, whereas the S110R mutant did not. The apparent melting temperatures of the wild type and the mutant, which were determined by the microscale fluorescence stability assay with the thiol-specific dye N-[4-(7-dietyhlamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM), were 49.0°C and 55.5°C, respectively. More detailed description is given in a separate article.[18]

References

[1]Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J. Seven-transmembrane receptors. Nat. Rev. Mol. Cell. Biol. **2002**, 3, 639−650.

[2]Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabaśi, A. L.; Vidal, M. Drug-target network. Nat. Biotechnol. **2007**, 25, 1119−1126.

[3]Fredriksson, R.; Lagerström, M. C.; Lundin, L. G.; Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol. Pharmacol. **2003**, 63, 1256−1272.

[4]Lau, F. W.; Nauli, S.; Zhou, Y.; Bowie, J. U. Changing single side-chains can greatly enhance the resistance of a membrane protein to irreversible inactivation. J. Mol. Biol. **1999**, 290, 559−564.

[5]Zhou, Y.; Bowie, J. U. Building a thermostable membrane protein. J. Biol. Chem. **2000**, 275, 6975−6979.

[6]Scott, D. J.; Kummer, L.; Tremmel, D.; Plückthun, A. Stabilizing membrane proteins through protein engineering. Curr. Opin. Chem. Biol. **2013**, 17, 427−435.

[7]Serrano-Vega, M. J.; Magnani, F.; Shibata, Y.; Tate, C. G. Conformational thermostabilization of the β1-adrenergic receptor in a detergent-resistant form. Proc. Natl. Acad. Sci. U. S. A. **2008**, 105, 877−882.

[8]Magnani, F.; Shibata, Y.; Serrano-Vega, M. J.; Tate, C. G. Co-evolving stability and conformational homogeneity of the human adenosine A_{2a} receptor. Proc. Natl. Acad. Sci. U. S. A. **2008**, 105, 10744−10749.

[9]Shibata, Y.; White, J. F.; Serrano-Vega, M. J.; Magnani, F.; Aloia, A. L.; Grisshammer, R.; Tate, C. G. Thermostabilization of the neurotensin receptor NTS1. J. Mol. Biol. **2009**, 390, 262−277.

[10]Chen, K. M.; Zhou, F.; Fryszczyn, B. G.; Barth, P. Naturally evolved G protein-coupled receptors adopt metastable conformations. Proc. Natl. Acad. Sci. U. S. A. **2012**, 109, 13284−13289.

[11]Bhattacharya, S.; Lee, S.; Grisshammer, R.; Tate, C. G.; Vaidehi, N. Rapid Computational Prediction of Thermostabilizing Mutations for G Protein-Coupled Receptors. J. Chem. Theory Comput. **2014**, 10, 5149−5160.

[12]Sauer, D. B.; Karpowich, N. K.; Song, J. M.; Wang, D. N. Rapid bioinformatic identification of thermostabilizing mutations. Biophys. J. **2015**, 109, 1420−1428.

[13]Yasuda, S.; Kajiwara, Y.; Takamuku, Y.; Suzuki, N.; Murata, T.; Kinoshita, M. Identification of thermostabilizing mutations for membrane proteins: rapid method based on statistical thermodynamics. J. Phys. Chem. B, **2016**, 120, 3833−3843.

[14]Kajiwara, Y.; Yasuda, S.; Takamuku, Y.; Murata, T.; Kinoshita, M. Identification of thermostabilizing mutations for a membrane protein whose three-dimensional structure is unknown. J. Comp. Chem. **2017**, 38, 211−223.

[15]Haga, K.; Kruse, A. C.; Asada, H.; Kobayashi, T. Y.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T., et al. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. Nature **2012**, 482, 547−551.

[16]Yokoyama, U.; Iwatsubo, K.; Umemura, M.; Fujita, T.; Ishikawa Y. The prostanoid EP4 receptor and its signaling pathway. Pharmacol Rev. **2013**, 65, 1010−1052.

[17]Suno, R.; Maeda, S.; Yasuda, S.; Yamashita, K.; Hirata, K.; Horita, S.; Tawaramoto, M. S.; Tsujimoto, H.; Murata, T.; Kinoshita, M., et al. Submitted.

[18]Toyoda, Y.; Morimoto, K.; Suno, R.; Horita, S.; Yamashita, K.; Hirata, K.; Sekiguchi, Y.; Yasuda, S.; Shiroishi, M.; Shimizu, T., et al. Submitted.

[19]Kajiwara, Y.; Ogino, T.; Yasuda, S.; Takamuku, Y.; Murata, T.; Kinoshita, M. Physical origins of remarkable thermostabilization by an octuple mutation for the adenosine A_{2a} receptor. Chem. Phys. Lett. **2016**, 657, 119−123.

[20]Yasuda, S.; Oshima, H.; Kinoshita, M. Structural stability of proteins in aqueous and nonpolar environments. J. Chem. Phys. **2012**, 137, 135103.

[21]McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. 1994, 238, 777−793.

[22]Popot, J. L.; Engelman, D. M. Helical membrane protein folding, stability, and evolution. Annu.

Rev. Biochem. **2000**, 69, 881−922.

[23]Hansen, J.-P.; McDonald, I. R. Theory of Simple Liquids, 3rd ed.; Academic Press: London, U.K., **2006**.

[24]König, P. M.; Roth, R.; Mecke, K. R. Morphological thermodynamics of fluids: shape dependence of free energies. Phys. Rev. Lett. **2004**, 93, 160601.

[25]Roth, R.; Harano, Y.; Kinoshita, M. Morphometric approach to the solvation free energy of complex molecules. Phys. Rev. Lett. **2006**, 97, 078101.

[26]Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **1993**, 234, 779−815.

[27]Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. Science **2008**, 322, 1211−1217.

[28]Margan, D.; Borota, A.; Mracec M.; Mracec M. 3D homology model of the human prostaglandin E2 receptor EP4 subtype. Rev. Roum. Chim. **2012**, 57, 39−44.

[29]Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. **1983**, 4, 187−217.

[30]Feig, M.; Karanicolas, J.; Brooks, C. L., III. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J. Mol. Graphics Modell. **2004**, 22, 377−395.

[31]MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B **1998**, 102, 3586−3616.

[32]Tusnády, G. E.; Dosztańyi, Z.; Simon, I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics **2005**, 21, 1276−1277.

[33]Mitchell, J. B. O.; Price, S. L. Chem. Phys. Lett. On the relative strengths of amide…amide and amide…water hydrogen bonds. **1991**, 180, 517−523.

[34]Bowie, J. U. Membrane protein folding: how important are hydrogen bonds? Curr. Opin. Struct. Biol. **2011**, 21, 42−49.

[35]Sneddon, S. F.; Tobias, D. J.; Brooks C. L. III. Thermodynamics of amide hydrogen bond formation in polar and apolar solvents. J. Mol. Biol. **1989**, 209, 817−820.

[36]Ponder, J. W.; Richards, F. M. An efficient newton-like method for molecular mechanics energy minimization of large molecules. J. Comput. Chem. **1987**, 8, 1016−1024.

[37]Murakami, S.; Oshima, H.; Hayashi, T.; Kinoshita, M. On the physics of thermal-stability changes upon mutations of a protein. J. Chem. Phys. **2015**, 143, 125102.

[38]Katritch, V.; Fenalti, G.; Abola, E. E.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Allosteric sodium in class A GPCR signaling. Trends Biochem. Sci. **2014**, 39, 233−244.

[39]Hollenstein, K.; Kean, J; Bortolato, A.; Cheng, R. K.; Dore, A. S.; Jazayeri, A.; Cooke, R. M.;

Weir, M.; Marshall, F. H. Structure of class B GPCR corticotropin-releasing factor receptor 1. Nature **2013**, 499, 438−443.

[40]Jazayeri, A.; Dore, A. S.; Lamb, D.; Krishnamurthy, H.; Southall, S. M.; Baig, A. H.; Bortolato, A.; Koglin, M.; Robertson, N. J.; Errey, J. C., et al. Extra-helical binding site of a glucagon receptor antagonist. Nature **2016**, 533, 274−277.

[41]Yoshidome, T.; Oda, K.; Harano, Y.; Roth, R.; Sugita, Y.; Ikeguchi, M.; Kinoshita, M. Free-energy function based on an all-atom model for proteins. Proteins: Struct. Funct. Genet. **2009**, 77, 950−961.

[42]Yasuda, S.; Yoshidome, T.; Harano, Y.; Roth, R.; Oshima, H.; Oda, K.; Sugita, Y.; Ikeguchi, M.; Kinoshita, M. Free-energy function for discriminating the native fold of a protein from misfolded decoys. Proteins: Struct. Funct. Genet. **2011**, 79, 2161−2171.

Chapter 6

General Conclusion

In this study, we have developed a theoretical method for identifying the thermostabilizing mutations for membrane proteins and demonstrate its high capability for G-protein coupled receptors (GPCRs). The free-energy function (FEF; F) incorporated in the method comprises entropic and energetic components (S and Λ, respectively; $F=\Lambda-TS$ where T is the absolute temperature). The most important originality of the FEF is that the entropic effect originating from the translational displacement of hydrocarbon groups (CH_2, CH_3, and CH) constituting nonpolar chains of lipid molecules is taken into account. Upon GPCR folding, a large gain of entropy of hydrocarbon groups occurs, for which the close packing of side chains in the side-to-side association of the seven helices within the transmembrane region is primarily responsible. The entropic gain is calculated using a hybrid of an integral equation theory (IET) and the morphometric approach (MA) developed by Kinoshita and coworkers. The former is a statistical-mechanical theory of fluids and the latter is necessitated to quantify the entropic gain for a large protein with complex polyatomic structure with sufficient accuracy and high speed. Upon GPCR folding, the protein intramolecular electrostatic energy is significantly lowered mainly due to the formation of intramolecular hydrogen bonds (IHBs). The energetic component is related to this lowering. The features of our method can also be described as follows: It is based on statistical thermodynamics and the physical origins of the enhanced thermostability are always unambiguous; and all of the possible mutations can be examined (i.e., the whole mutational space can be explored) with minor computational effort. In what follows, we recapitulate the main conclusions drawn.

In Chapter 2, we have applied our FEF to the identification of thermostabilizing mutations in the case where the three-dimensional (3D) structure of the wild type is experimentally known and can be utilized. The structure of a mutant is modeled using the MODELLER program. The adenosine A_{2a} receptor ($A_{2a}R$) in the inactive state is considered as an example. We propose the following strategy for the identification: First, calculate ΔS (ΔX denotes the change in X upon GPCR folding) for all of the possible mutations and nominate key residues to be mutated in the sense that many of their mutations will lead to relatively higher enhancement of the stability; second, select some of the mutations of the key residues using ΔF. A mutant giving a sufficiently low value to $\Delta\Delta F$ ("ΔF of the mutant"−"ΔF of the wild type") leads to significantly higher stability. A total of nine mutations (seven of them are predicted to be thermostabilizing) are experimentally examined. The success rates are in the range from 7/9 to 9/9 depending on the criterion employed for the thermostability relative to that of the wild type in the theoretical prediction and on the thermostability measure adopted in our experiments. (For the seven mutations mentioned above, the success rates are in the range from 5/7 to 7/7: We conclude that the overall success rate is 6/7.) The melting temperature T_m

becomes higher than that of the wild type by ~7°C for mutations of T88E and S91K despite that they are single mutations. We further examine a double mutation of T88E−S91R: It leads to an increase in T_m of ~12°C.

In Chapter 3, we have investigated the physical origins of the remarkable enhancement of structural stability brought by an octuple mutation (A54L, T88A, R107A, K122A, L202A, L235A, V239A, and S277A; seven residues are mutated to Ala) for $A_{2a}R$ in an inactive state. T_m of the octuple mutant was experimentally shown to be higher than that of the wild type by ~20°C. Both of the changes in solvent-entropy gain ΔS and in energy decrease $\Delta \Lambda$ upon the mutation, which are denoted by $\Delta\Delta S$ and $\Delta\Delta \Lambda$, respectively, are significantly large and low. We have theoretically proved that the solvent-entropy gain and the energy decrease arising from the formation of protein intramolecular hydrogen bonds upon GPCR folding are made substantially larger by the mutation, leading to the remarkable enhancement.

In Chapter 4, the theoretical method for identifying thermostabilizing mutations developed in Chapter 2 has been applied to a practical situation where the 3D structure of the wild type is not experimentally available. Our theoretical method is combined with the homology modeling, and the combination is illustrated for $A_{2a}R$ in the inactive state. Though the 3D structure of $A_{2a}R$ has experimentally been determined, it is assumed to be unavailable. First, template proteins possessing sufficiently high transmembrane region identity (TMI) with the target protein (TMI exceeding 25%) are searched using the multiple sequence alignment. Second, the 3D-structure models of the target protein are constructed by referring to the structural information of each template. Third, from among sufficiently many models thus generated, the best model is selected.

The best model can be the one whose root-mean-square deviation (RMSD) for C_α atoms, which is defined between a model and the 3D structure experimentally determined (i.e., crystal structure), is very small. However, RMSD cannot be calculated in a practical situation. We therefore make the following proposition: The model with the lowest value of ΔF (the change in our free-energy function F upon protein folding) is the best one. We have examined the prediction performance in cases where several models including the best model and the crystal structure are utilized. The performance from the best structure is the closest to that from the crystal structure: Both of T88 and S91 are correctly nominated as the key residues; the correlation coefficient between $\Delta\Delta F$ in the best model and $\Delta\Delta F$ in the crystal structure for the 19 mutations for S91 is considerably high (~0.86); and when the seven mutations predicted to be thermostabilizing in Chapter 2 are considered and the overall success rate from the crystal structure is taken to be 6/7, that from the best model is 5/7. We also show that ΔF is better than RMSD as the measure for selecting the best 3D-structure model.

In Chapter 5, using our FEF, we have developed a theoretical strategy by which thermostabilizing mutations of many different GPCRs can be considered at the same time. “Key residues” and “hot-spot residues” are clearly defined. Many of the 19 mutations of a key residue lead to substantially higher stabilities. A hot-spot residue is the key residue designated by the Ballesteros-Weinstein (BW) number N_{BW} and common in significantly many different GPCRs. The

number of the mutations of a residue which leads to $-\Delta\Delta S/k_B<-X$ is counted; when it is larger than or equal to M, the residue is nominated as one of the key residues. The strategy has been illustrated for three GPCRs of Class A, the adenosine A_{2a} receptor (A_{2a}R), muscarinic acetylcholine receptor 2 (M2R), and prostaglandin E receptor 4 (EP4), in the inactive state. The models of wild-type and mutant structures are constructed on the basis of the crystal structures for the inactive state. More specifically, a hot-spot residue for the GPCRs has been found out in the following manner: The key residues determined for A_{2a}R are T88(3.36), S91(3.39), and N181(5.42) (X=10 and M=4). Here, the value in the parentheses is the BW number N_{BW}. Those for M2R are S64(2.45), S110(3.39), V111(3.40), and P198(5.50) (X=5 and M=4). Those for EP4 are P24(1.39), L57(2.42), G60(2.45), S73(2.58), G106(3.39), A112(3.45), S114(3.47), G190(5.43), and C202(5.55) (X=10 and M=4). The residues underlined share the same BW number and the residue with N_{BW}=3.39 is a hot-spot residue.

We have then evaluated the thermostabilities of the possible 19 mutations of the hot-spot residue in M2R or EP4 using the FEF and selected the mutation for which $\Delta\Delta F/(k_BT_0)$ takes the lowest value. The selected mutations for M2R and EP4 are S110R and G106R, respectively. (S110K and G106K are also highly stabilizing for M2R and EP4, respectively.) It has then been corroborated in experiments that S110R and G106R actually lead to considerably enhanced stabilities. The crystallization based on the lipidic cubic phase (LCP) method and the determination of new three-dimensional (3D) structures for EP4 and M2R have thus been made possible.

We believe that when both of the wild-type and mutant structures are experimentally available, our theoretical method does not fail to correctly predict which of them is thermally more stable. That is, our FEF captures the essential physics of the dependence of thermostability on the GPCR structure. However, it is suggestive that the capability of our theoretical method may become progressively lower as the accuracy of the wild-type and mutant structures (especially the wild-type structure) decreases. In Chapter 4, the success rate does not become significantly worse because we could find out a template protein with sufficiently high TMI (the highest value is 34%). For EP4 considered in Chapter 5, on the other hand, TMI of the template protein is only 17%, with the result that the success rate is deteriorated to 2/10. It is thus required that a more reliable method of modeling the wild-type structure (and its mutant structure) be developed. There is no reason why our theoretical method is inapplicable to GPCRs in the active state. We intend to apply it to GPCRs in the active state and achieve equally successful results.

In the present stage of our study, only the mutations of residues within the TM region are considered. It is required that a modification be made to include the mutations of residues outside the TM region, i.e., those immersed in water. It is experimentally known that the thermophilic rhodopsin (TR; it possesses the seven-transmembrane topology but is not a GPCR) is much more thermostable than GPCRs. According to our preliminary theoretical analysis, the higher thermostability is realized by stabilizing the water region. We are inclined to think that the mutations of residues within the water region may endow a GPCR with enhanced stability.

Moreover, even more stabilized TR could be utilized as advanced performance materials such as a sensor chip. The extension to the transporters is also interesting and important.

The further development of this study is expected to enable us to purify and produce a large amount of a mutated membrane protein, leading to efficient structural and functional analyses toward structure-guided drug design. Thus, it potentially has a great impact on the comprehension of life phenomena and causes of various diseases, development of new drugs, and progress of life science.

List of Publications

Chapter 2

"Identification of Thermostabilizing Mutations for Membrane Proteins: Rapid Method Based on Statistical Thermodynamics",
Satoshi Yasuda, Yuta Kajiwara, Yuuki Takamuku, Nanao Suzuki, Takeshi Murata, and Masahiro Kinoshita,
Journal of Physical Chemistry B, **120**(16), 3833-3843 (2016).

Chapter 3

"Physical Origins of Remarkable Thermostabilization by an Octuple Mutation for the Adenosine A_{2a} Receptor",
Yuta Kajiwara, Takahiro Ogino, Satoshi Yasuda, Yuuki Takamuku, Takeshi Murata, and Masahiro Kinoshita,
Chemical Physics Letters, **657**, 119-123 (2016).

Chapter 4

"Identification of Thermostabilizing Mutations for a Membrane Protein Whose Three-Dimensional Structure is Unknown",
Yuta Kajiwara, Satoshi Yasuda, Yuuki Takamuku, Takeshi Murata, and Masahiro Kinoshita,
Journal of Computational Chemistry, **38**(4), 211-223 (2017).

Chapter 5

"Hot-Spot Residues to be Mutated Common in G Protein-Coupled Receptors of Class A: Identification of Thermostabilizing Mutations Followed by Determination of Three-Dimensional Structures for Two Example Receptors",
Satoshi Yasuda, Yuta Kajiwara, Yosuke Toyoda, Kazushi Morimoto, Ryoji Suno, So Iwata, Takuya Kobayashi, Takeshi Murata, and Masahiro Kinoshita,
Journal of Physical Chemistry B, **121**(26), 6341-6350 (2017).

"Crystal structure of the human prostaglandin E receptor EP4",
Yosuke Toyoda, Kazushi Morimoto, Ryoji Suno, Shoiciro Horita, Keitaro Yamashita, Kuio Hirata, Yusuke Sekiguchi, Satoshi Yasuda, Mitsunori Shiroishi, Tomoko Shimizu, Yuji Urushibata, Yuta Kajiwara, Tomoaki Inazumi, Yunhon Hotta, Hidetsugu Asada, Takanori Nakane, Yuki Shiimura, Tomoya Nakagita, Kyoshiro Tsuge, Suguru Yoshida, Tomoko Kuribara, Takamitsu Hosoya, Yukihiko Sugimoto, Norimichi Nomura, Miwa Sato, Takatsugu Hirokawa, Masahiro Kinoshita, Takeshi Murata, Kiyoshi Takayama, Masaki Yamamoto, Shuh Narumiya, So Iwata, Takuya Kobayashi,

Submitted.

Copyright and publisher's Link

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Other Publication

"Physical Origin of Thermostabilization by a Quadruple Mutation for the Adenosine A_{2a} Receptor in the Active State",

Yuta Kajiwara, Satoshi Yasuda, Simon Hikiri, Tomohiko Hayashi, Mitsunori Ikeguchi, Takeshi Murata, and Masahiro Kinoshita,

Submitted.

List of Patents

International Patent

“膜タンパク質の熱安定化変異体予測装置、熱安定化変異体予測方法、および、プログラム”
村田 武士, 木下 正弘, 安田 賢司, 高椋 勇樹, 水谷 健二, 鈴木 七緒, 梶原 佑太
H27.06.24, 国際出願番号：PCT/JP2015/068277, 出願番号：特願 2016-529643.

List of Presentations

International Conference
Poster Presentation

“Theoretical prediction of mutations leading to enhanced structural stability of GPCRs”,
Yuta Kajiwara, Satoshi Yasuda, Yuki Takamuku, Nanao Suzuki, Takeshi MUurata, Masahiro Kinoshita,
Pacifichem2015, Honolulu, Hawaii, USA, 2015.12.

“Theoretical Prediction of Mutations Improving Thermal Stability of Adenosine A2a Receptor”,
Yuta Kajiwara, Satoshi Yasuda, Yuki Takamuku, Takeshi Murata, Masahiro Kinoshita, 59th Annual Meeting of the Biophysical Society, Baltimore, USA, 2015.2.

“Expression, purification and crystallization of 5-HT2A receptor”,
Nanao Suzuki, Akane Saito, Kenta Hitomi, Yuuki Takamuku, Kenji Mizutani, Yuta Kajiwara, Satoshi Yasuda, Masahiro Kinoshita, Takeshi Murata, GPCR workshop 2015, Big island, Hawaii, USA, 2015.12.

“Development of Thermostabilization Method by Mutations for GPCR:I. Theoretical Prediction on the Basis of Statistical Thermodynamics”,
Satoshi Yasuda, Yuta Kajiwara, Yuuki Takamuku, Nanao Suzuki, Takeshi Murata, and Masahiro Kinoshita, GPCR workshop 2015, Big island, Hawaii, USA, 2015.12.

“Development of Thermostabilization Method by Mutations for GPCR:II. Experimental Verification of the Theoretical Prediction”,
Yuuki Takamuku, Satoshi Yasuda, Yuta Kajiwara, Nanao Suzuki, Masahiro Kinoshita, Takeshi Murata, GPCR workshop 2015, Big island, Hawaii, USA, 2015.12.

“Towards structure determination of the human prostanoid receptor bound to the antibody”,
Yosuke Toyoda, Kazushi Morimoto, Ryoji Suno, Yusuke Sekiguchi, Keitaro Yamashita, Kunio Hirata, Satoshi Yasuda, Hidetsugu Asada, Takanori Nakane, Yuki Shiimura, Tomoya Nakagita, Tomoaki Inazumi, Kyoshiro Tsuge, Yuta Kajiwara, Tomoko Shimizu, Yuji Urushibata, Suguru Yoshida, Tomoko Kuribara, Takamitsu Hosoya, Masahiro Kinoshita, Yukihiko Sugimoto, Norimichi Nomura, Takeshi Murata, Kiyoshi Takayama, Masaki Yamamoto, Shuh Narumiya, So Iwata, Takuya Kobayashi, GPCR workshop 2015, Big island, Hawaii, USA, 2015.12.

Domestic Conference

Poster Presentation

“Theoretical prediction of mutations improving thermal stability of adenosine A2a Receptor”,
Yuta Kajiwara, Satoshi Yasuda, Yuki Takamuku, Takeshi Murata, Masahiro Kinoshita,
第 52 回日本生物物理学会，札幌, 2014.9.

“Prediction of Thermostabilizing Mutations for G Protein-Coupled Receptors:Construction of an Efficient Method”,
Yuta Kajiwara, Satoshi Yasuda, Yuuki Takamuku, Nanao Suzuki, Takeshi Murata, Masahiro Kinoshita,
第 53 回 日本生物物理学会年会，金沢, 2015.9

“Identification of Thermostabilizing Mutations for a Membrane Protein Whose Three-Dimensional Structure is Unknown”,
Yuta Kajiwara, Satoshi Yasuda, Takeshi Murata, Masahiro Kinoshita,
第 54 回 日本生物物理学会年会，つくば, 2016.11

“Physical origin of stabilization by a quadruple mutations for the adenosine A2a receptor in the ative state”,
Yuta Kajiwara, Satoshi Yasuda, Mitsunori Ikeguchi, Takeshi Murata, Masahiro Kinoshita,
第 55 回 日本生物物理学会年会，熊本, 2017.9.

“Theoretical Identification of Hot-Spot Residues to be Mutated Common in G Protein-Coupled Receptors of Class A”,
Satoshi Yasuda, Yuta Kajiwara, Yosuke Toyoda, Kazushi Morimoto, Ryoji Suno, So Iwata, Takuya Kobayashi, Takeshi Murata, Masahiro Kinoshita,
第 55 回 日本生物物理学会年会，熊本, 2017.9.

“Identification of Thermostabilizing Mutations for a G-protein coupled receptor in the active state”,
Simon Hikiri, Ryosuke Nakano, Nanao Suzuki, Satoshi Yasuda, Yuta Kajiwara, Masahiro Kinoshita,
第 55 回 日本生物物理学会年会，熊本, 2017.9.

Acknowledgement

This thesis has been accomplished with support of many people. I would like to express my gratitude to them.

I would like to express my sincere gratitude to Prof. Masahiro Kinoshita at the Institute of Advanced Energy at Kyoto University for his diligent guidance, valuable suggestions, fruitful discussions and encouragement throughout the work. During my master and doctor courses for 5 years at Kyoto University, he has taught me a lot of things not only about scientific knowledge, but also philosophy and attitude as a scientist, wisdom for life, and precious advices for my future.

I am deeply grateful to Prof. Takeshi Murata and Assistant Prof. Satoshi Yasuda at Graduate School of Science at Chiba University for their guidance, kind advices and helpful discussions. Without their support, this work would not have been possible.

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 and my understanding for science.

I am deeply grateful to Dr. Tomohiko Hayashi at Kinoshita's laboratory for their encouragements and constructive discussions.

I would like to thank Dr. Simon Hikiri at Graduate School of Science at Chiba University. He always supports not only my research but also my life in a doctor course.

I would like to thank Dr. Yosuke Toyoda, Dr. Kazushi Morimoto, and Dr. Ryoji Suno at Graduate School of Medicine at Kyoto University for kind advices and helpful discussions.

Sincere appreciation should be expressed to Dr. Hiraku Oshima (Present Address: RIKEN) and Dr. Shota Murakami (Present Address: Tome Institute CO., Ltd.) for their useful suggestions, computational skills, and warmful encouragement.

The computer program for the morphometric approach was developed with Prof. Roland Roth at Universität Tübingen, and Associate Prof. Yuichi Harano of Graduate School of Pharmaceutical Science at Himeji Dokkyo University. This work was supported by Grant-in-Aid for Scientific Research (No. 17H03663) and (B) (No. 25291035) from Japan Society for the Promotion of Science (JSPS), by a research grant from Koyoanagi Fundation, by ImPACT program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan), and by the Platform Project for Supporting in Drug Discovery and Life Science Research (Platform for Drug Discovery, Informatics, and Structural Life Science) from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) and Japan Agency for Medical Research and Development (AMED).

Finally, I greatly thank my family for supporting my mind.

February, 2018

Yuta Kajiwara