Theoretical study of catalytic mechanisms of editing reaction by aminoacyl-tRNA synthetases

「アミノアシル tRNA 合成酵素によるエディテング反応における触媒メカニズムの理論解析」

２０１９

坂　部　　翔

兵庫県立大学大学院生命理学研究科

Contents

Chapter 1

Introduction

1-1. Aminoacyl-tRNA synthetase (aaRSs)

The base sequence of messenger RNA (mRNA) is determined by the rule of Watson-Crick base pairs, which are made with each bases in the coding region of the gene. However, for the synthesis of protein (*i.e.*, the amino acid sequence), adaptor molecules are required to bridge base sequences (codons) of mRNA and amino acid sequences of protein, because no rules exist to associate them, such as the Watson-Crick base pair. Notably, a codon, composed of combinations of 3 bases (4^3 = 64 patterns), specifies each amino acid.

Thus, transfer RNA (tRNA) was discovered, acting as the adaptor molecule that connects codons and the cognate amino acids. Organisms possess ~50 distinct types of tRNAs, each of which is specific to an amino acid among 20 distinct types of amino acids. Each tRNA molecule recognizes codons that are corresponding to the specific amino acid, by the anticodon moiety of the tRNA.

The next question is what recognizes tRNAs for transferring the specific amino acids to the cognate tRNA. Also, how does it recognize the cognate tRNAs ? To clarify this problem, an enzyme family, aminoacyl-tRNA synthetases (aaRSs)[1)], plays a pivotal role, *i.e.*, decoding of codons (base sequence) toward amino acid (aa) residues (amino acid sequence) in protein in the translation. Fundamentally, each aaRS is corresponding to each of 20 amino acids, and attaches the cognate amino acid to the specific tRNAs (this is not the case with eubacteria).

This catalytic reaction is referred to as the aminoacylation, which is operated via two steps; i) the activation of an amino acid toward aa-AMP, and ii) transferring of the activated amino acid to the 3' terminal of tRNA (*i.e.*, adenosine 76; A76). Notably, aaRSs strictly recognize the specific tRNAs and amino acids, and thus are responsible for translational fidelity.

All of aaRSs can be classified into 2 classes, *i.e.*, classes I and II.[2,3)] Correspondingly, the ancestors of class I and II enzymes are completely different. Each of these classes are further classified into three subclasses a, b, and c. For example, class Ia aaRSs involve ValRS, LeuRS, and IleRS, which are the targets of this thesis. aaRSs

are considered to have existed in the primordial organisms, and thus this enzyme family is one of the key systems to understand the origin of life (discussed below).

1-2. Editing reaction of aminoacyl-tRNA synthetases

However, some aaRSs mis-attach non-cognate amino acids to the specific tRNAs, because some amino acids, such as Leu, Val, and Ile, are very similar to each other in their chemical structures. For proofreading of such mis-attachments of amino acids, aaRSs possess the editing mechanisms in each of the two aminoacylation steps that were mentioned above. These proofreading reactions are termed the "pre-transfer editing" and "post-transfer editing", corresponding to the aforementioned reaction steps i) and ii), respectively.[4-11)]

The analyses of mechanisms of the editing reactions by aaRSs have been performed by employing various methodologies such as biochemistry, molecular biology, and structural biology. However, the catalytic reaction mechanisms based on the electronic structures of the systems as well as the molecular structure levels have never been analyzed for a long time. The first work was performed with respect to the Leu system at the atomistic and electronic structure levels, as follows.

In our lab., we adopted *ab initio* quantum mechanics (QM) electronic structure calculations coupled to classical (molecular) mechanics (MM) calculations as “hybrid” molecular dynamics (MD) simulations, to analyze the catalytic mechanism of the post transfer editing occurring in the complex of LeuRS and mis-aminoacylated $tRNA^{Leu}$. The analysis revealed that O3' of A76 in the tRNA activated the nucleophilic water, which attacked to the carbonyl group located between the tRNA and amino acid moieties, which led to the cleavage of the bond, and thus the reaction was elucidated to be ribozymal. We referred to this catalytic mechanism as the hybrid ribozyme/protein catalyst.[12-17)]

Very recently, our mechanism has also been reasserted by biochemical experiments, as follows. For a “defective” mutation of the mis-aminoacylated $tRNA^{Leu}$ where O3' of A76 was replaced with H3', the post-transfer editing activity was reduced by as much as ~10^4-fold.[18)] Thus, with respect to the Leu system, the theoretical and experimental data have been shown to be well consistent.

However, it is also well known that for the Val systems, the post-transfer editing activity was reduced by only ~10-fold for the “defective” mutation of the mis-aminoacylated $tRNA^{Val}$ where O3' was replaced with H3'. Also, for the Ile system, similar data have been reported previously.[19)] These are experimental discrepancies, since the three-dimensional (3D) structures of LeuRS and ValRS/IleRS are strikingly

similar. Nevertheless, the reasons of the above-mentioned experimental discrepancies have been remained to be clarified.

To overcome these contradictions, we constructed an atomistic structural model of the complex of ValRS and mis-aminoacylated, threonyl-tRNA^{Val}. Employing this 3D structural model, we performed an over 100-ns classical MD simulation in a large water box, and thus obtained the fully-solvated relaxed structure of the molecular system. Then, we performed hybrid *ab initio* QM/MM electronic structure simulations to elucidate the crucial states of the catalytic site for the post-transfer editing reaction of the Val system.

The analysis revealed that the catalytic mechanism of the editing reaction in the Val system was also ribozymal, which showed that the reaction mechanisms were conserved in the Leu and Val systems. Moreover, in order to analyze the catalytic mechanisms of the "defective" complex of ValRS and the mis-aminoacylated tRNA^{Val} where O3' of A76 was replaced with H3', we searched for another possible activator of the nucleophilic water, instead of O3' of A76 in our fully-solvated 3D structural model of ValRS and the "defective" mis-aminoacylated tRNA^{Val}. Thus, in this thesis, we propose a possible transition from the hybrid ribozyme/protein catalyst toward a protein enzyme, occurring in the "defective" Val system.

1-3. Methodologies to analyze catalytic mechanisms of biological macromolecules

1-3-1. *ab initio* QM calculation

In order to describe the reaction mechanisms, quantum mechanics is required to be adopted toward the systems that we analyze, since the electronic structures of the system would be changed, and thus need to be monitored in detail during the reactions. For such scales, characteristic features that are different from our dairy scales appear. In fact, electrons exhibit the duality, *i.e.*, both characteristic features of particle and wave. To describe the motions and energy levels of electrons in the reaction, we need to solve the Schrödinger's equation.

Thus, in this thesis, we performed experiments of the enzymatic reactions in computers, which were carried out by solving the Schrödinger's equation. It is the experiments employing computers, and so is very precise if one could solve the equation without approximations. Actually, the density functional theory (DFT), which is currently a standard methodology to evaluate the electronic structures of molecular systems, is not an approximation, but the exact solution of the Schrödinger's equation based on the variation principle.

In any current actual calculations employing the DFT, we usually utilize the DFT for

evaluating the exchange-correlation, which involves the most characteristic feature of QM systems. In recent standard QM calculations, the DFT is coupled to the Hartree-Fock (HF) calculation method, which is referred to as the hybrid *ab initio* all-electron HF/DFT calculation. Many functionals have been proposed; in particular, the B3LYP functional is the most standard one, because it commonly and well reproduces various physical quantities. This has made the current *ab initio* QM calculation significantly reliable and employed in various systems in many scientific and engineering fields.

1-3-2. 3D structural modeling

As mentioned above, in this thesis, we analyzed the geometric and electronic structures of the catalytic site of the editing reaction in the complex of ValRS and the mis-aminoacylated, threonyl-$tRNA^{Val}$. For this aim, we built an atomistic structural model of the ValRS•threonyl-$tRNA^{Val}$ complex, based on the crystal structure of the complex. However, in the crystal structure, the editing domain (connective polypeptide 1 domain; CP1 domain) was too flexible to precisely capture the strict atomic coordinates of the CP1 domain. In fact, some secondary structures in the CP1 domain are irregularly deviated from the standard conformations. Accordingly, for the above-mentioned aim, we needed to re-construct the CP1 domain in the ValRS•threonyl-$tRNA^{Val}$ complex. Herein, we combined another crystal structure of the isolated CP1 domain.

Thus, in order to constitute the fully-solvated, atomistic structural model of the ValRS•threonyl-$tRNA^{Val}$ complex in this thesis, we combined the following two crystal structures; i) the entire structure of the *Thermus thermophilus* ValRS and $tRNA^{Val}$ (PDB ID: 1ivs),[20] and ii) the editing domain (CP1 domain) isolated from the identical ValRS with an inhibitor (*i.e.*, [N-(L-threonyl)sulfamoyl]-adenosine (Thr-AMS)) (PDB ID: 1wk9).[21] Notably, antiparallel two β-strands connect the editing domain and the main body of ValRS, and thus the CP1 domain is so flexible to be rotated in terms of the main body. Employing the structural modeling techniques, we replaced the atomic coordinates of the CP1 domain in the crystal structure of ValRS•$tRNA^{Val}$ complex, with those of the crystal structure of the isolated CP1 domain, and thereby constituted the entire structure of the ValRS•threonyl-$tRNA^{Val}$ complex. In order to relax the constructed structural model with the explicit solvent water molecules, we performed (classical) molecular dynamics (MD) simulations in a large water box.

1-3-3. Molecular Dynamics (MD) simulation

Standard classical MD simulations generate trajectories of atoms in the systems,

being subject to the Newton's equation of motion. To perform the MD calculation, an atom is considered to be a point of the mass and partial charge. Covalent bonds are approximated by springs, for which the force constants are adopted to the bond lengths and angles, being obtained by *ab initio* QM calculations. Very recently, highly advanced QM calculation methodologies have been applied to evaluate the potential and force fields of DNA, and thereby its excellent parameter set has currently become available. This has also remarkably improved the accuracy of the structural prediction of nucleic acids.

Moreover, a lot of elaborate technical and methodological advances have been achieved to enhance the computational performance and accuracy of the MD simulations with respect to biological macromolecular systems, such as the electrostatic interactions of all atoms without employing the cutoff distance under the periodical boundary condition. This has also created the significantly higher precision in the MD calculations. Very recently, another break-through has been achieved, i.e., computational acceleration of the MD calculation by employing general purpose/graphics processor unit (GP/GPU), which has enhanced the performance of the MD calculations up to ~1,000-fold. Thus, an MD simulation including several hundred thousands (~10^5) atoms has been possible up to several μs time scale.

1-3-4. Hybrid QM/MM calculation

Current *ab initio* QM calculations need a lot of computational resources. Accordingly, it is impossible to describe the whole structures of biological macromolecular systems such as protein enzymes. We adopt such calculations only to the catalytic center, and for the remainders, we employ the classical molecular mechanics (MM) representations. This is referred to as QM/MM calculation. In the QM/MM calculation scheme, distinct two types of the schemes have been proposed up to date, *i.e.*, the interactions of the QM and MM regions are to be involved, or not involved, in the calculation schemes.

In this thesis, we adopted the QM/MM calculation scheme involving the interactions between the QM and MM regions, which are referred to here as the “hybrid” QM/MM calculation. This was conducted in order to precisely evaluate the interactions in the calculations.

1-4. aaRSs and the origin of life

In organisms, protein and RNA enzymes are responsible for the catalytic reactions. For example, the reactions in ribosome, which is the “factory” for biosynthesis of

protein, are operated via ribozymes; *i.e.*, the peptidyl transferase (transfer of an amino acid to a peptide chain) and deacylation (dissociation of the peptide chain) are performed by ribozymes in ribosome. This means that protein is fundamentally synthesized by RNA enzyme (ribozyme), which shows that the evolutional understandings are closely relevant to the most essential aspect of organisms (mechanisms for decoding genetic information). Also, to provide correct amino acids in the protein sequence, aminoacyl-tRNAs are required, and thus aaRSs are essential in any organisms.

Accordingly, my goal is to understand the origin of life from a viewpoint of catalytic mechanisms of primordial biological systems. In this manner, we point out herein the importance of the combined perspective of the catalytic mechanisms and the origin of life (involving evolutional changes of the catalytic mechanisms as well as the mechanisms of the molecular recognition between aaRSs and the cognate tRNAs). In order to resolve such issues, current usual methodologies (in the molecular evolution) are not sufficient to obtain substantial understandings of the combined aspects. In this thesis, we tried to provide a step to approach the goal.

References

1) J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry* (W. H. Freeman, New York, 2002) 5th ed., p. 1208.
2) S. Cusack, C. Berthet-Colominas, M. Hartlein, N. Nassar, and R. Leberman, Nature **347**, 249 (1990).
3) G. Eriani, M. Delarue, O. Poch, J. Gangloff, and D. Moras, Nature **347**, 203 (1990).
4) O. Nureki, D. G. Vassylyev, M. Tateno, A. Shimada, T. Nakama, S. Fukai, M. Konno, T. L. Hendrickson, P. Schimmel, and S. Yokoyama, Science **280**, 578 (1998).
5) L. F. Silvian, J. M. Wang, and T. A. Steitz, Science **285**, 1074 (1999).
6) S. Fukai, O. Nureki, S. Sekine, A. Shimada, J. Tao, D. G. Vassylyev, and S. Yokoyama, Cell **103**, 793 (2000).
7) R. S. Mursinna, T. L. Lincecum, and S. A. Martinis, Biochemistry **40**, 5376 (2001).
8) T. L. Lincecum, Jr., M. Tukalo, A. Yaremchuk, R. S. Mursinna, A. M. Williams, B. S. Sproat, W. Van Den Eynde, A. Link, S. Van Calenbergh, M. Grotli, S. A. Martinis, and S. Cusack, Mol. Cell **11**, 951 (2003).
9) R. Fukunaga, S. Fukai, R. Ishitani, O. Nureki, and S. Yokoyama, J. Biol. Chem. **279**, 8396 (2004).
10) R. Fukunaga and S. Yokoyama, J. Biol. Chem. **280**, 29937 (2005).
11) Y. X. Zhai and S. A. Martinis, Biochemistry **44**, 15437 (2005).
12) Y. Hagiwara, O. Nureki, and M. Tateno, FEBS Lett. **583**, 825 (2009).
13) Y. Hagiwara, O. Nureki, and M. Tateno, FEBS Lett. **583**, 1901 (2009).
14) Y. Hagiwara, M. J. Field, O. Nureki, and M. Tateno, J. Am. Chem. Soc. 132, 2751 (2010).
15) J. Kang, Y. Hagiwara, and M. Tateno, J. Biomed. Biotechnol. 2012, **11** (2012).
16) J. Kang, H. Kino, M. J. Field, and M. Tateno, J. Phys. Soc. Jpn **86**, 044801 (2017).
17) Y. Hagiwara, T. Ohta, and M. Tateno, J. Phys.: Condens. Matter **21**, 064234 (2009).
18) M. Dulic, N. Cvetesic, I. Zivkovic, A. Palencia, S. Cusack, B. Bertosa, and I. Gruic-Sovulj, J. Mol. Biol. 430, 1 (2018).
19) B. E. Nordin and P. Schimmel, Biochemistry 42, 12989 (2003).
20) S. Fukai, O. Nureki, S. Sekine, A. Shimada, D. G. Vassylyev, and S. Yokoyama, RNA **9**, 100 (2003).
21) Fukunaga and S. Yokoyama, J. Biol. Chem. 280, 29937 (2005).

Chapter 2

Atomistic Structural Modeling of Complex of Valyl-tRNA Synthetase and Mis-Aminoacylated $tRNA^{Val}$ to Identify Catalytic Activator of Editing Reaction

2-1 Introduction

Elucidating the mechanisms of the evolution from primordial biological macromolecular systems to the present systems is an important issue relevant to various fields concerning the origin of life, such as biophysics, biochemistry, molecular biology, molecular evolution, geological science, systems biology, and systems science. The enzyme family of aminoacyl-tRNA synthetases (aaRSs)[1] is one of the earliest proteins that could have been present in the primordial systems of life, and thus it is a crucial and interesting system for studying the evolutional mechanisms in ancient organisms.[2,3] Our goal is to understand these issues from biophysical viewpoints, such as the catalytic reaction mechanisms of the early biological macromolecular systems based on *ab initio* electronic structure calculations.

In protein biosynthesis, aaRSs transfer their cognate amino acids to the 3′-end of specific tRNAs. They catalyze two crucial reactions: (i) activating the amino acid residues for the formation of aminoacyl adenylates and (ii) linking the amino acid moiety with the 3′-end of the tRNA. The 20 canonical aaRSs can be classified into two highly conserved structural groups, classes I and II, each containing 10 members.[4,5] The two classes are further subdivided into three groups, Ia/IIa, Ib/IIb, and Ic/IIc. To ensure translational fidelity, aaRSs must discriminate their cognate amino acid residues from the others. However, several aaRSs have difficulty in distinguishing their cognate amino acids from non-cognate ones, owing to the co-existence of chemically and structurally similar amino acids, such as valine (Val), isoleucine (Ile), and leucine (Leu), and thus yield mis-activated amino acids and mis-aminoacylated tRNAs. To correct these errors,

the aaRSs possess an editing function to hydrolyze the mis-products.[6-13] Two types of editing have been reported so far: hydrolysis of a mis-activated amino acid, referred to as pre-transfer editing, and hydrolysis of a mis-aminoacylated tRNA, which is post-transfer editing.

In our previous study, we constructed a structural model of a class Ia aaRS, LeuRS, in complex with valyl-$tRNA^{Leu}$ [14] and identified the nucleophilic water for the editing reaction.[15] Employing *ab initio* quantum mechanical (QM)/molecular mechanical (MM) molecular dynamics (MD) simulations, we revealed that the editing mechanism in the LeuRS•valyl-$tRNA^{Leu}$ complex is ribozymal, where the $O^{3'}$ of the ribose moiety of the 3′-end nucleotide (adenine 76; A76) acts as the general base to activate the nucleophilic water.[16-18] This ribozymal reaction is also novel from the viewpoint that the protein contributes to the enhancement of the ribozyme catalytic activity through hydrogen bond formation in the catalytic core of the $tRNA^{Leu}$. While the catalytic cores of the conventional protein-dependent ribozymes (e.g., group I intron and ribosome) are solely composed of RNAs,[19] the protein moiety (LeuRS) in the LeuRS•valyl-$tRNA^{Leu}$ complex forms hydrogen bonds "directly" with the catalytic site of the substrate, i.e., the valine-attached A76 of the $tRNA^{Leu}$, and thereby enhances the ribozymal reaction.[16] Thus, we refer to the LeuRS•valyl-$tRNA^{Leu}$ complex as a "hybrid ribozyme/protein catalyst" (Fig. 1).

Very recently, Dulic et al. have experimentally shown that the defective mutation of the $O^{3'}$ atom significantly reduces the activity of the Leu system (~10^4-fold rate reduction),[20] and thus the hybrid ribozyme/protein catalyst mechanism has been reasserted by biochemical experiments. Furthermore, Kumar et al. suggested that for the complex of prolyl-tRNA synthetase (ProRS) and alanyl-$tRNA^{Pro}$, the editing reaction operated by a similar mechanism, in which the 2'-OH group of A76 of $tRNA^{Pro}$ is also important, and involved in the substrate binding and the activation of the nucleophilic water.[21] Note here that ProRS is a class IIa enzyme. Thus, the hybrid ribozyme/protein catalyst mechanism seems to be shared in the editing reactions by various aaRSs beyond the classes (Table I).

However, for the catalytic mechanisms of the editing reaction by aaRSs, an experimental conflict is yet to be resolved as follows. Valyl-tRNA synthetase (ValRS) (Fig. 1) and isoleucyl-tRNA synthetase (IleRS) also belong to the class Ia aaRSs. Owing to their high structural similarity with LeuRS, they may also share a common editing mechanism with LeuRS. However, the removal of the $O^{3'}$, *i.e.*, the chemical modification of the 3'-hydroxyl group of A76 (3'-OH) to 3'-H, was shown to have subtle effects on the editing activity in the Val and Ile systems,[22] whereas this modification

seriously reduced the activity in the Leu system as mentioned earlier (Table I).[16,20] Therefore, this discrepancy in the experimental data concerning the Val/Ile and Leu systems still remains to be settled.

In an attempt to provide an explanation for this experimental conflict, we previously constructed a preliminary model of the ValRS complex (see the Supporting Information in our previous report[16]). Based on this modeling, we suggested that the hybrid ribozyme/protein catalyst mechanism was also shared in the editing reaction of the Val system. However, for the variant $tRNA^{Val}$ in which the 3'-OH of A76 was replaced with the 3'-H, we anticipated that the editing reaction of the Val system would be operated by the protein enzyme activity of ValRS (*i.e.*, an amino acid residue acted as the general base) (Fig. 1B). However, in this analysis, the structure of the LeuRS•valyl-$tRNA^{Leu}$ complex was employed to build the Val system. Accordingly, more sophisticated structural modeling is still required to resolve the conflict in the experimental data, based on the atomistic and electronic structure analyses.

In this study, based on the crystal structures of the *Thermus thermophilus* ValRS•$tRNA^{Val}$ complex and the isolated editing domain [the editing domain is also referred to as the connective polypeptide (CP) 1 domain], we constructed a modeled structure of ValRS in complex with threonyl-$tRNA^{Val}$ (mis-aminoacylated $tRNA^{Val}$). Employing the modeled structure, we performed MD simulations to provide an explanation for the discrepancy in the experimental data. As a result of the present modeling, we identified the nucleophilic water, which was hydrogen-bonded with the 3'-OH of A76 as found in the LeuRS•valyl-$tRNA^{Leu}$ complex.[16] This suggests that the editing mechanism in the Val system is identical to that in the Leu system.

Moreover, we also performed an MD simulation of ValRS in complex with a variant of threonyl-$tRNA^{Val}$, which lacks the $O^{3'}$ of A76. The resultant configuration of the structural water molecules in the catalytic site suggested that, in the case of this tRNA variant, the reaction mechanism would change from the ribozymal mode, occurring in the wild type, to that catalyzed by the protein, ValRS, since the conserved Asp276 presumably acts as the activator of the nucleophile in the ValRS complexed with the variant $tRNA^{Val}$. These results are consistent with the experimental data,[22] thus supporting the feasibility of the structural model constructed in the present study.

In this manner, we propose that the transition from the hybrid ribozyme/protein catalyst (in the case where the wild type $tRNA^{Val}$ is bound) toward the protein enzyme (in the case where the aforementioned variant of $tRNA^{Val}$ is bound) occurred to compensate for the defective catalytic activator of the $O^{3'}$ of A76 in the variant $tRNA^{Val}$. This type of transition in the catalytic mechanisms could have played an important role

in the transition from the RNA to RNP worlds in the origin of life.

2-2 Materials and Methods

2-2-1 System setup for structural modeling

For the structural modeling, two systems were used: the *Thermus thermophilus* ValRS editing domain in complex with [N-(L-threonyl)sulfamoyl]-adenosine (Thr-AMS) as an inhibitor [Protein Data Bank (PDB) accession code 1WK9], and ValRS in complex with $tRNA^{Val}$. In the latter complex, the 3′-terminus is bound to the editing active site (PDB accession code 1IVS).[12,23] To construct a model of ValRS in complex with $tRNA^{Val}$ attached with a non-cognate amino acid, threonine, the former system (1WK9) was used for the initial coordinates of the CP1 domain, while the latter system (1IVS) was used for those of the remainder of the ValRS and $tRNA^{Val}$ (G1-A76).

However, the initial conformation of the side chain of Val215 was obtained from the crystal structure of the ValRS·$tRNA^{Val}$ complex (1IVS), i.e., its $\chi 1$ angle was set to the value in the crystal structure of the complex, since a comparison of the two crystal structures revealed differences between the complex and the isolated CP1 domain. Similarly, the $\chi 1$ and $\chi 2$ angles of Phe264 were set to the values observed in the crystal structure of the complex. The initial coordinates of the threonine moiety attached to A76 were obtained from the coordinates of the inhibitor. The crystal water molecules involved in the structure of the isolated CP1 domain (1WK9) were exploited in the present modeling, although those located close to the substrate and $tRNA^{Val}$ were removed. Here, the crystal structure of the complex (1IVS) was used to identify such water molecules (the CP1 domains of both crystal structures were superimposed). Thus, HOH57, HOH136, and HOH143 in 1WK9 were eliminated, since they were within 2 Å of the heavy atoms of the substrate and $tRNA^{Val}$. The crystal water molecules in the structure of the complex (1IVS) that are located close to the other moieties were involved in the present modeling (i.e., 281 crystal water molecules). Next, the ValRS•threonyl-$tRNA^{Val}$ complex was immersed in a solvent box consisting of 48,153 water molecules, and the periodic boundary condition was imposed, where the size of the unit box was $141.8 \times 98.8 \times 115.6$ $Å^3$. Thus, the total number of atoms in the system was 160,895.

2-2-2 MD simulations

All MD simulations were performed using the AMBER 10 software package.[24] The parm99SB force field was applied to all atoms in the system. Electrostatic interactions were calculated by the particle-mesh Ewald (PME) method[25] with a dielectric constant

of 1.0, and a cutoff distance of 12 Å was used to calculate the direct space sum for PME. The SHAKE algorithm[26] was used to restrain the bond lengths involving hydrogen atoms, allowing the time step for integrations to be set to 1 fs. Temperature and pressure were regulated using the Berendsen algorithm.[27]

The parameterization of threonyl-tRNA^{Val} was performed as follows. All atom types of threonyl-tRNA^{Val} were taken from the ff99SB force field. For the threonine moiety attached to A76, the atom types of the N-terminal threonine were assigned; for A76, those of the ribo-adenosine with a 5'-phosphate group and a 3'-OH were used. More specifically, for the $O^{2'}$ of A76, the atom type of an ester oxygen, 'OS', was used, since $O^{2'}$ bonds with the carboxyl oxygen of the valine moiety. With respect to the partial charges of tRNA^{Val} including A76, the values defined in the ff99SB force field were applied. For the threonine moiety attached to A76, the values of the N-terminal valine defined in the force field were employed, so this moiety could establish consistency with the others. These assignments of the partial charges resulted in a non-zero value (i.e., −0.29) of the total charge of the threonyl-A76 moiety. To neutralize the value, we added a positive value of 0.006 to the partial charge of each atom in the threonyl-A76 moiety.

2-2-3 Hybrid ab initio QM/MM calculations

For hybrid *ab initio* QM/MM calculations, the initial structures were obtained from the optimized structures at the MM level, with respect to the two snapshots corresponding to the closed and opened H-gate states, as mentioned later. All hybrid QM/MM calculations were performed using modified AMBER and GAMESS packages, and our in-house interface program connecting the two software packages.[28-31] Restricted Hartree-Fock (RHF)/density functional theory (DFT) all-electron calculations were performed using the B3LYP functional, along with the 6-31G* basis set. The link atom approach was used to satisfy the valence requirements where the QM/MM boundary separated the covalently-bonded atoms. While calculating the electrostatic interactions between the QM and MM atoms lying within 25 Å from the center of mass of the QM region, the partial charges of the MM atoms were incorporated into the one-electron integral term of the QM Hamiltonian, whereas the interactions between the QM and the other MM atoms were calculated at the MM level. The partial charges were obtained from the AMBER parm99 force field.

For geometry optimization, the atoms assigned as QM atoms were as follows: the substrate, i.e., threonine-attached A76, the water molecule corresponding to the nucleophile, and the moieties forming hydrogen bonds or salt bridges with the substrate

(the side chains of Thr215, Asp276, Asp279, and the water molecule forming the hydrogen bond with the nucleophilic water molecule and Asp276). Thus, 71 atoms, including capping hydrogen atoms, were set as QM atoms. The self-consistent field (SCF) calculations were performed with respect to the closed and opened H-gate states.

2-3 Results and Discussion

2-3-1 Structural modeling of the CP1 domain of ValRS in complex with threonyl-tRNA^{Val}

The crystal structure of the complex of *Thermus thermophilus* ValRS and tRNA^{Val}, in which the 3'-end nucleotide of the tRNA^{Val}, adenine 76 (A76), and a Val-AMP analog, N-[L-valyl]-N' -adeno-syldiaminosulfone (Val-AMS) (an inhibitor) are bound to the editing site of the CP1 domain, was determined at 2.90 Å resolution (PDB code: 1IVS) (Fig. 1).[23] With respect to the CP1 domain of *Thermus thermophilus* ValRS, a high resolution crystal structure of the isolated form was determined at 1.75 Å resolution (PDB code: 1WK9) after the determination of the crystal structure of the ValRS•threonyl-tRNA^{Val} complex (Fig. 1).[12] Due to the lower resolution of the latter crystal structure, it contained several energetically unfavorable conformations. In fact, the Ramachandran plot of the crystal structure and the combined validation employing the C-Alpha Based Low-resolution Annotation Method (CaBLAM)[32] showed that the catalytic domain contains six outliers (Gly208, Ala243, Asp257, Ala308, Leu312, and Asp313). In addition, there is a large cavity surrounded by hydrophobic amino acids, such as Ile212, Ile226, and Pro246, in the interior of the catalytic domain, which would destabilize the protein structure (Fig. 1).

In contrast, with respect to the crystal structure of the isolated CP1 domain (1WK9), no outliers are included in the protein structure, as shown by the Ramachandran plot (the validation employing CaBLAM is not required for this structure). Moreover, the cavity observed in the crystal structure of the complex is absent, for the following reason: In the crystal structure of the complex, the side chain atoms of Tyr202 are assigned as exposed toward the solvent, while in the isolated CP1 domain, they are directed toward the interior of the protein, and thus form the core through tight interactions with other hydrophobic amino acid residues, such as Ile212, Ile226, and Pro246 (Fig. 1). Based on these facts, for the moiety of the CP1 domain, we employed the atomic coordinates from the high resolution crystal structure of the isolated CP1 domain to construct the modeled structure of ValRS in complex with mis-aminoacylated tRNA^{Val}, while the coordinates of the remaining parts of ValRS and tRNA^{Val} were obtained from the crystal structure of the complex.

The two crystal structures still have several conformational differences derived from the presence/absence of the $tRNA^{Val}$, rather than the difference in their resolutions. In the crystal structure of the complex, the χ1 and χ2 angles of Phe264, which interacts with Cytidine 75 (C75) of $tRNA^{Val}$, are rotated by about 100 and 60 degrees, respectively, relative to those in the crystal structure of the isolated CP1 domain (Fig. 1). In addition, the difference in the χ1 angle of Val215 is approximately 100 degrees, accompanying the conformational changes of its backbone (both φ and ψ angles of Val215 are rotated by approximately 30 degrees) (Fig. 1). With respect to these amino acid residues, we employed the conformations observed in the crystal structure of the complex (details are discussed below).

Furthermore, for A76, the initial coordinates were also obtained from the crystal structure of the complex, while the threonine moiety attached to A76 was modeled using the atomic coordinates of the threonine moiety in the crystal structure of the isolated CP1 domain complexed with a pre-transfer editing inhibitor. The recognition mode of the threonine in the crystal structure is considered reasonable, since the mutational experiments with respect to the amino acid residues that participate in the recognition revealed a serious decrease in the editing activity.[12)]

2-3-2 Stability of the modeled structure

Subsequently, employing the modeled structure of the complex, we performed a 100-ns productive MD simulation of ValRS in complex with threonyl-$tRNA^{Val}$, where threonine was attached to the $O^{2'}$ atom of the ribose of A76, and thereby obtained the relaxed structure, as discussed below. For the heavy atoms of the CP1 domain and the entire complex structure that were obtained from the snapshots of the 100-ns MD simulation, the time evolutions of root mean square deviations (RMSDs) are shown in Fig. 2. The RMSD values were less than 3.5 and 2.5 Å for the heavy atoms of the CP1 domain moiety and the entire complex structure, respectively. In fact, the structures of the active site and ligand in the MD simulation were preserved as those in the crystal structure of the ValRS•$tRNA^{Val}$ complex (Fig. 3). Moreover, the hydrogen bond networks between the threonine moiety of the substrate and the CP1 domain were quite consistent with those observed in the high resolution crystal structure of the CP1 domain in complex with an inhibitor of the pre-transfer editing substrate. In addition, the hydrogen bond network between the adenosine moiety of the substrate and the ValRS moiety in the modeled structure was quite consistent with that observed in the crystal structure of the ValRS•$tRNA^{Val}$ complex. Thus, the modeled structure of the ValRS•threonyl-$tRNA^{Val}$ complex was well converged and relaxed with the

experimental structures being preserved.

2-3-3 Identification of the nucleophilic water

The substrate of the post-transfer editing by ValRS contains a labile ester linkage between $O^{2'}$ of the ribose of A76 of $tRNA^{Val}$ and the carbonyl carbon of its amino acid moiety (threonine in this study). This carbon atom (referred to as "C" in this report; see Fig. 4A) will be attacked by a nucleophile (a water molecule) in the editing reaction. Thus, the water molecule that is closest to the C atom is considered to act as the nucleophile. The modeled structure of the ValRS•threonyl-$tRNA^{Val}$ complex, equilibrated in the MD simulation without any constraints, showed that the water molecule located at this position hydrogen-bonds with the hydrogen atom of the 3'-hydroxyl group of the ribose of A76 (referred to as 3'-HO) (Fig. 4A). This water molecule is also hydrogen-bonded with another water, which is bound to Asp276. Thus, the hydrogen bond network including the possible nucleophilic water is consistent with that in our previous model of the LeuRS•valyl-$tRNA^{Leu}$ complex.[16] The angle defined by the oxygen atom of the water (referred to as O_w), C, and the carbonyl oxygen of the threonine moiety (referred to as O) confirmed that the water molecule is located at a favorable position for the nucleophilic attack (Fig. 4A). In fact, the O_w–C vector is nearly orthogonal to the plane defined by the $O^{2'}$, C, and O atoms.

2-3-4 Rotation of the 3'-HO leads to the productive complex: the opening of the H-gate

In this manner, we have identified the nucleophilic water, in good agreement with our previous studies of the Leu system.[15] However, the average O_w–C distance, 3.4 Å, calculated using the trajectory of the 100-ns MD simulation, was made stable (see green line in Fig. 5), thus suggesting that the nucleophilic attack may be hindered. This configuration of the nucleophilic water was also observed in our previous MD simulations of *Thermus thermophilus* LeuRS in complex with the valyl-$tRNA^{Leu}$ complex.[15] In the previous study, we found that the 3'-HO, which is hydrogen-bonded with the O_w, prevents the access of the nucleophilic water (accordingly, we referred to the 3'-HO as the "H-gate", which blocked the water's approach), and that the dihedral angle that includes the hydrogen atom, $C^{2'}$–$C^{3'}$–$O^{3'}$–3'HO, must rotate for the nucleophilic water to approach the C atom.[15] Thus, for editing in the *Thermus thermophilus* ValRS•threonyl-$tRNA^{Val}$ complex, this rotation may also be required to open the H-gate.

To examine this issue, we performed another 1-ns MD simulation as a trial, where a

distance constraint between the O_w and C atoms and a dihedral angle constraint concerning $C^{2'}$–$C^{3'}$–$O^{3'}$–3'-HO were imposed using harmonic potentials. As a result, the water molecule approached the C atom, and the O_w–C distance was reduced to less than 2.6 Å (see red line in Fig. 5; also see Fig. 4B). We also performed a further 1 ns MD simulation, where only the harmonic potential regarding the O_w–C distance was imposed (without the constraint regarding the above-mentioned dihedral angle), and found that the reduction of the O_w–C distance was hindered (~2.9 Å) (see blue line in Fig. 5; also see Fig. 4A). Thus, the present analysis supported the requirement of the dihedral angle rotation (*i.e.*, the opening of the H-gate), which thus leads to the approach of the nucleophilic water to the C atom.

It should be noted here that, in the case of the *Thermus thermophilus* LeuRS, Thr248 acts to open the H-gate, i.e., its O^{γ} atom acts as the hydrogen bond acceptor for the 3'-HO. This allows the water to approach the C atom, leading to the formation of the "productive enzyme•substrate (ES) complex".[15] However, in *Thermus thermophilus* ValRS, the threonine is replaced with valine (Val215), which cannot act as the hydrogen-bond acceptor. A representative snapshot of the MD simulation with the rotated dihedral angle revealed that a water molecule functions to open the H-gate, through the formation of a hydrogen bond with the 3'-HO (Fig. 4B). This water molecule is coordinated with the phosphate group of A76, and stably exists at a favorable position to act as the hydrogen-bond acceptor for the 3'-HO (Fig. 4B). Interestingly, this functional role of the water molecule was also observed in our previous MD simulation of the T247V mutant of *Thermus thermophilus* LeuRS in complex with valyl-$tRNA^{Leu}$.[15] Thus, in the species where Thr248 is replaced with a hydrophobic amino acid residue, such as in *Thermus thermophilus* ValRS and *Deinococcus radiodurans* ValRS, a water molecule should adopt the role of the threonine for opening the H-gate.

In our previous study, we also investigated the feasibility of the dihedral angle rotation involving the 3'-HO by the potential of mean force (PMF) calculation, using the dihedral angle as the reaction coordinates. We found that the resultant free energy barrier for the rotation was ~3.7 kcal/mol, indicating that the change in this dihedral angle can be driven by thermal fluctuations.[15] This was further confirmed by means of the previous QM/MM MD simulation,[16] and may also be the case in the present Val system. Notably, we found that the above-mentioned configuration of the water molecules is also dependent on the conformation of Val215, which is different between the two crystal structures of the isolated CP1 domain and the ValRS•$tRNA^{Val}$ complex. The details are described below.

2-3-5 Possible role of $tRNA^{Val}$ as an effector influencing Val215 and the water molecules in the editing site

We found that the configuration of the water molecules in the editing site (described above) is dependent on the conformation of Val215. As mentioned above, Val215 adopts different conformations in the crystal structures of the isolated CP1 domain and the ValRS•$tRNA^{Val}$ complex. Its χ1 angle in the former structure is rotated by approximately 100 degrees relative to that in the latter structure. Thereby, the Val215 side chain was directed toward the ribose of A76 in the present modeled structure of the ValRS•threonyl-$tRNA^{Val}$ complex before the χ1 angle was changed to that observed in the crystal structure of the ValRS•$tRNA^{Val}$ complex (Fig. 1). In addition, the φ and ψ angles of Val215 are different in the two crystal structures of the isolated CP1 domain and the ValRS•$tRNA^{Val}$ complex.

In the present modeling, we employed the conformation of the Val215 side chain observed in the crystal structure of the complex as the initial structure for the MD simulations (*i.e.*, its χ1 angle was set to the value observed in the crystal structure of the ValRS•$tRNA^{Val}$ complex). In contrast, with respect to the backbone conformation of Val215, we used that in the crystal structure of the isolated CP1 domain, to examine whether the conformation changed to that observed in the crystal structure of the ValRS•$tRNA^{Val}$ complex in the MD simulations. In fact, during the MD simulation of the ValRS•threonyl-$tRNA^{Val}$ complex, the side chain conformation of Val215 was stable, whereas the backbone conformation, which was obtained from the crystal structure of the isolated CP1 domain, changed to that observed in the crystal structure of the ValRS•$tRNA^{Val}$ complex without any constraints. While the differences in the φ and ψ angles of Val215 were approximately 30 and 20 degrees, respectively, between the two crystal structures of the isolated CP1 domain (the initial conformation of the present MD simulations) and the ValRS•$tRNA^{Val}$ complex, those between the MD snapshots and the crystal structure of the complex were reduced to less than 10 degrees (Fig. 3B).

Moreover, we also performed an MD simulation of the ValRS•threonyl-$tRNA^{Val}$ complex, where the initial coordinates of Val215 (for both the side chain and backbone atoms) were obtained from the crystal structure of the isolated CP1 domain. During the MD simulation, the conformation of Val215 was stabilized through the hydrophobic interaction with the methylene group of $C^{5'}$ of A76, and did not change to that observed in the crystal structure of the ValRS•$tRNA^{Val}$ complex. The resultant conformation is shown in Fig. 6A, which depicts the interaction between Val215 and the $C^{5'}$ methylene group. In this structure, the bulky side chain of Val215 adjacent to the ribose moiety

excludes the neighboring water molecules, including the one that should act to open the H-gate, while in the productive conformation, the water molecules are accessible to the catalytic site (Fig. 6B; also see Fig. 4B). This difference is also derived from the effect of Val215 on the phosphate group of A76; the dihedral angle of $C^{5'}$-$O^{5'}$-P-O^{P} (O^{P} represents an oxygen atom attached to the phosphorus) in the former structure (Fig. 6A) is around +20 degrees (the average value calculated in the MD simulation), while that in the latter structure (Fig. 6B) is around −30 degrees (also the average value calculated in the MD simulation). Here, the value in the crystal structure of the ValRS•$tRNA^{Val}$ complex is −13.1 degrees, which shows that the experimental data is more consistent with the conformation observed in the latter structure obtained by the productive MD simulation (Fig. 6B).

In this manner, two types of structures observed in the crystal structures of the isolated CP1 domain and the ValRS•$tRNA^{Val}$ complex were also found to be stable in the MD simulations. A comparison of these two structures revealed that the positions of the water molecules are also dependent on the conformation of the phosphate group (Fig. 6). In other words, the "unproductive" conformation of Val215 also induces the "unproductive" conformation of the phosphate group, thereby leading to the exclusion of the water molecules that are necessary for the function of ValRS, such as those participating in the nucleophilic attack and the opening of the H-gate. These results further suggested the importance of the conformational change of Val215 to establish the productive complex, although both conformations of Val215 were stabilized in the MD simulations, as discussed. This conformational change of Val215 may be induced through the binding of $tRNA^{Val}$ (in fact, the unproductive conformation of Val215 was observed in the crystal structure of the isolated CP1 domain), thus implying that $tRNA^{Val}$ also contributes to the formation of the productive complex as an effector molecule.

2-3-6 Mechanism of the editing reaction of ValRS in complex with a tRNA variant

The present analyses suggest that the water molecule that is hydrogen-bonded with the 3'-OH of the A76 ribose acts as the nucleophile of the editing reaction. The high structural similarity among the class Ia aaRSs (ValRS, IleRS, and LeuRS) implies that the mechanism of the editing reaction by ValRS is shared with that of LeuRS, *i.e.*, the editing in the Val system may also be a ribozymal reaction, where the nucleophilic water is activated by the 3'-OH of the $tRNA^{Val}$, which acts as a general base for the activation.[16] However, the substitution of the hydroxyl group with a hydrogen atom, *i.e.*, the introduction of the chemical modification of the 3'-OH to the 3'-H of A76, had

relatively subtle effects on the activity (~10-fold rate reduction),[22] although the corresponding substitution in the Leu system caused a significant decrease in the editing activity (~10^4-fold rate reduction) (Table I).[16,20] This appears to be inconsistent with the ribozymal mechanism for the Val system.

In order to explain this discrepancy, we performed a further MD simulation of ValRS in complex with the aforementioned threonyl-$tRNA^{Val}$ variant (*i.e.*, the 3'-OH of A76 was replaced with the 3'-H) without constraints. The results revealed that a water molecule can be located adjacent to the C atom (thus, it could be the nucleophile), since the excluded volume of the H-gate in the wild type $tRNA^{Val}$ (3'-HO) is lost in the case of the variant $tRNA^{Val}$ (3'-H) (Fig. 4C). It should be noted here that the water molecule is hydrogen-bonded with another water molecule, which is hydrogen-bonded with the conserved Asp276, in this non-constrained MD simulation. The presence of such a hydrogen bond network including the nucleophilic water suggests an alternative mechanism of editing, even though the 3'-OH is defective; *i.e.*, Asp276 can abstract a hydrogen atom from the coordinated water molecule, and then the hydrogen atom of the nucleophilic water can be abstracted by the Asp276-coordinated water. This means that, in the ValRS complex with the variant $tRNA^{Val}$, Asp276 can compensate for the functional role of the 3'-OH, by activating the nucleophile in a water-mediated manner.

In fact, this type of activation has also been proposed for other systems, such as F1-ATPase (ATP hydrolysis)[33] and the *glmS* ribozyme (transesterification).[34] Thus, based on our present analyses, we can provide an explanation concerning the compensation mechanism for the $tRNA^{Val}$ variant, and thereby fill the apparent gap (discrepancy) between the experimental data of the Val system (which suggest the minimal importance of the 3'-OH) and the Leu system (which show its importance). This is also quite consistent with our preliminary modeling, in which the modeled structure of the LeuRS•valyl-$tRNA^{Leu}$ complex was used to model the ValRS•threonyl-$tRNA^{Val}$ complex.[16] The Ile system could also be similar to the Val system as discussed in our previous report.[16] In fact, Glu or Asp residues (e.g. D341 in *Escherichia coli* IleRS) could compensate for the defect of 2'-OH of A76 in $tRNA^{Ile}$ (5-fold rate reduction) (Table I).[22,35]

In contrast, with respect to the Leu system, the corresponding Asp residue is presumably inactivated through the salt bridge with an Arg residue, as discussed in our previous report.[16] In fact, Dulic et al. showed that the mutation where the corresponding Asp residue in *Escherichia coli* LeuRS was replaced with Ala (D342A) marginally affected the editing activity (~3-fold rate reduction) (see Table I and Fig. 5 in the report of Dulic et al.).[20] The Phe system is also similar to the Leu system; *i.e.*, the

defect of 3'-OH of A76 reduced the editing activity by ~300-fold (Table I),[36] since a hydrophilic amino acid residue is not found in the proximity of the editing site as discussed in our previous report (see Table SI in the Supporting Information in our previous report).[16] Note that PheRS is a class IIc enzyme. Thus, the hybrid ribozyme/protein catalyst mechanism is shared in the editing reactions of various aaRS systems beyond the classes.

2-3-7 Evolutional transition from hybrid enzyme toward protein enzyme

Moreover, the present finding with the Val system indicates that the function of a hybrid ribozyme/protein catalyst can be transferred to a protein enzyme (Fig. 1B), simply through the defect of a ribozymal functional group (e.g. 3'-OH). This further suggests that in molecular evolution, protein enzymes evolved from hybrid ribozyme/protein catalysts, which could have evolved from more primordial systems, such as ribozymes and ribozyme-protein complexes. In the latter systems, the protein presumably acted as a chaperone for the ribozyme.[37] In fact, aaRSs are considered to be one of the earliest protein families that replaced functional ribozymes playing catalytic roles.[2,3,38-41] This picture could contribute to filling the serious gap between the RNA and RNP worlds in the current theory on the origin of life.

In fact, primordial ribozymes were considered to be pure RNA catalysts (the RNA world). In their environment, randomly generated peptides and small molecules were bound to the ribozymes, which may thus have led to the expansion of their structural repertoire, and thereby also their functional repertoire (the RNP world).[19] In the subsequent evolutionary stage, protein moieties that were more efficient in the catalysis even without the involvement of the ribozymes in the catalytic activations (*i.e.*, protein enzymes; see Fig. 1B) took over the biological functions of the original RNP complexes (note here that the definition of the protein enzyme in this study is different from the previous one[19]). Here, how did the protein enzymes accomplish the catalytic reactions without the involvement of the ribozymal catalysts? We propose that the hybrid ribozyme/protein catalysts intermediated between the original RNP complexes and the protein enzymes as the functional systems that emerged in the evolutionary process of the RNA and RNP worlds.

2-3-8 Molecular orbital analysis of activation mechanisms of editing reaction

We also investigated the electronic structures of the catalytic center for the editing reaction by the Val system, employing hybrid *ab initio* QM/MM calculations. To evaluate the effects of the distinct states of the H-gate on the electronic structure, we

performed the molecular orbital (MO) analysis of the conformations, with the H-gate opened and closed. The analysis revealed that the lowest unoccupied molecular orbital (LUMO) was located in the C atom in both conformations of the H-gate (Figs. 7A and 7B).

For the nucleophilic water, the 2*p* orbital of the O atom had the highest energy level among the occupied MOs of the water molecule. However, when the H-gate adopted the closed conformation, the energy level of this orbital was much lower than that of the highest occupied molecular orbital (HOMO) of the QM system (i.e., HOMO-11). In contrast, with respect to the open conformation of the H-gate, the identical MO of the nucleophilic water was raised up to HOMO-6 (from HOMO-11), and thus the energy level of this reactive MO seems to become elevated as the reaction proceeds (Figs. 7C, 7D, and 7E). Thus, the nucleophilic attack may be achieved when the energy level of the reactive MO in the nucleophilic water is raised up to that of the HOMO, thus resulting in the hybridization of the HOMO and the above-mentioned LUMO.

In fact, we have also observed similar dynamical changes in the electronic structure for the editing reaction by the LeuRS•valyl-tRNA^{Leu} complex.[17,18] Thus, such dynamical rearrangements of the electronic structure may be characteristic in the catalytic reactions of various biological macromolecular systems as described in our previous reports, and thus were referred to as the dynamical induction of the reactive HOMO (DIRH) and LUMO (DIRL).[17,18] This interesting issue should also be investigated for various other systems in the near future.

In this manner, the present modeling and the obtained structures may provide a solid basis for further theoretical and experimental studies in order to elucidate the catalytic mechanisms, such as those based on dynamical electronic structural changes. Hybrid QM/MM MD simulations are ongoing in our group for the analyses of the dynamical features in the editing reaction, to clarify the overall and detailed catalytic mechanisms of the Val system. Thereby, we may consolidate our understandings of the mechanisms of the catalytic reactions and evolutional transitions of the early biological macromolecular systems as well as the present biological systems, based on the dynamical electronic structure analyses.

2-4. Conclusion

In this study, we have constructed a modeled structure of the ValRS·threonyl-tRNA^{Val} complex, and identified the water molecule that presumably acts as the nucleophile in editing. This nucleophilic water is hydrogen-bonded with the 3'-OH of A76 in tRNA^{Val}, suggesting that the editing of the Val system is also ribozymal, similar

to the Leu system described in our previous study. The binding mode of the nucleophilic water is common among several aaRSs that possess the editing activity, such as the *Thermus thermophilus* IleRS (class Ia),[42] the *Pyrococcus abyssi* ThrRS (class IIa),[43-46] and the *Enterococcus faecalis* ProRS (class IIa).[21] Thus, the present work also suggests that this ribozymal mechanism is common among various aaRS systems beyond the classes, as discussed in our previous report (see Table I and the Supplemental Material, Fig. S1).[16]

However, with respect to the variant of $tRNA^{Val}$ in which the 3'-OH of A76 is replaced with 3'-H, the editing activity is still preserved. This could occur because Asp276 compensates for the functional role of the 3'-OH, by activating the nucleophile through a water-mediated hydrogen bond network. This finding indicates that the function of a hybrid ribozyme/protein catalyst can be transferred to a protein enzyme, simply through the inactivation of a ribozymal functional group (e.g. 3'-OH of A76). This phenomenon further implies the evolutional transition from RNA enzymes (the RNA world) toward protein enzymes (assisted by RNA) (the RNP world), intermediated by hybrid ribozyme/protein catalysts, in the molecular evolution of life.

For the electronic structure of the catalytic centre, the reactive orbital of the nucleophilic water (the 2*p* of the O atom) initially had a very low energy level when the H-gate was closed, although this orbital should be responsible for the nucleophilic attack. In contrast, when the H-gate was opened, the energy level of this MO was elevated, which may lead it to the HOMO during the nucleophilic attack. Thus, the hybridization of the HOMO and LUMO is expected to be induced, thereby forming a new covalent bond. Such dynamical electronic structural changes may be characteristic to various biological macromolecular systems, since the reactive orbitals are initially masked by the other MOs even in the catalytic centre. In this manner, *ab initio* electronic structure calculations may invoke novel understandings of the mechanisms of the catalytic reactions and evolutional transitions of the early biological macromolecular systems, as well as the present biological systems of life.

References

1) Berg JM, Tymoczko JL, Stryer L, Biochemistry (W. H. Freeman, New York, 2002) 5th ed., p. 1208.
2) T. J. Bullwinkle and M. Ibba, Top. Curr. Chem. **344**, 43 (2014).
3) H. Jakubowski, FEBS Lett. **590**, 469 (2016).
4) S. Cusack, C. Berthet-Colominas, M. Hartlein, N. Nassar, and R. Leberman, Nature **347**, 249 (1990).
5) G. Eriani, M. Delarue, O. Poch, J. Gangloff, and D. Moras, Nature **347**, 203 (1990).
6) O. Nureki, D. G. Vassylyev, M. Tateno, A. Shimada, T. Nakama, S. Fukai, M. Konno, T. L. Hendrickson, P. Schimmel, and S. Yokoyama, Science **280**, 578 (1998).
7) L. F. Silvian, J. M. Wang, and T. A. Steitz, Science **285**, 1074 (1999).
8) S. Fukai, O. Nureki, S. Sekine, A. Shimada, J. Tao, D. G. Vassylyev, and S. Yokoyama, Cell **103**, 793 (2000).
9) R. S. Mursinna, T. L. Lincecum, and S. A. Martinis, Biochemistry **40**, 5376 (2001).
10) T. L. Lincecum, Jr., M. Tukalo, A. Yaremchuk, R. S. Mursinna, A. M. Williams, B. S. Sproat, W. Van Den Eynde, A. Link, S. Van Calenbergh, M. Grotli, S. A. Martinis, and S. Cusack, Mol. Cell **11**, 951 (2003).
11) R. Fukunaga, S. Fukai, R. Ishitani, O. Nureki, and S. Yokoyama, J. Biol. Chem. **279**, 8396 (2004).
12) R. Fukunaga and S. Yokoyama, J. Biol. Chem. **280**, 29937 (2005).
13) Y. X. Zhai and S. A. Martinis, Biochemistry **44**, 15437 (2005).
14) Y. Hagiwara, O. Nureki, and M. Tateno, FEBS Lett. **583**, 825 (2009).
15) Y. Hagiwara, O. Nureki, and M. Tateno, FEBS Lett. **583**, 1901 (2009).
16) Y. Hagiwara, M. J. Field, O. Nureki, and M. Tateno, J. Am. Chem. Soc. **132**, 2751 (2010).
17) J. Kang, Y. Hagiwara, and M. Tateno, J. Biomed. Biotechnol. **2012**, 11 (2012).
18) J. Kang, H. Kino, M. J. Field, and M. Tateno, J. Phys. Soc. Jpn **86**, 044801 (2017).
19) T. R. Cech, Cell **136**, 599 (2009).
20) M. Dulic, N. Cvetesic, I. Zivkovic, A. Palencia, S. Cusack, B. Bertosa, and I. Gruic-Sovulj, J. Mol. Biol. **430**, 1 (2018).
21) S. Kumar, M. Das, C. M. Hadad, and K. Musier-Forsyth, J. Phys. Chem. B **116**, 6991 (2012).
22) B. E. Nordin and P. Schimmel, Biochemistry **42**, 12989 (2003).
23) S. Fukai, O. Nureki, S. Sekine, A. Shimada, D. G. Vassylyev, and S. Yokoyama, RNA **9**, 100 (2003).
24) D. A. Case, Darden, T.A., Cheatham, T.E., III, Simmerling, C.L., Wang, J., Duke,

R.E., Luo, R., Crowley, M., Walker, R.C., Zhang, W., Merz, K.M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossva´ry, I., Wong, K.F., Paesani, F., Vanicek, J., Wu, X., Brozell, S.R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D.H., Seetin, M.G., Sagui, C., Babin, V., Kollman, P.A., AMBER 10, University of California, San Francisco (2008).

25) T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).

26) J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

27) H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

28) M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

29) D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, J. Comput. Chem. **26**, 1668 (2005).

30) Y. Hagiwara, T. Ohta, and M. Tateno, J. Phys.: Condens. Matter **21**, 064234 (2009).

31) K. Jiyoung, O. Takehiro, H. Yohsuke, N. Keigo, Y. Tetsunori, N. Hidemi, and T. Masaru, J. Phys.: Condens. Matter **21**, 064235 (2009).

32) J. S. Richardson and D. C. Richardson, Ann. Rev. Biophy. **42**, 1 (2013).

33) M. Dittrich and K. Schulten, J. Bioenerg. Biomembr. **37**, 441 (2005).

34) D. J. Klein and A. R. Ferre-D'Amare, Science **313**, 1752 (2006).

35) B. E. Nordin and P. Schimmel, J. Biol. Chem. **277**, 20510 (2002).

36) J. Ling, H. Roy, and M. Ibba, Proc. Natl. Acad. Sci. U. S. A. **104**, 72 (2007).

37) L. Ribas de Pouplana and P. Schimmel, Cell **104**, 191 (2001).

38) P. O'Donoghue and Z. Luthey-Schulten, Microbiol. Mol. Biol. Rev. **67**, 550 (2003).

39) L. Ribas de Pouplana and P. Schimmel, Cell. Mol. Life Sci. **57**, 865 (2000).

40) B. Ruan, I. Ahel, A. Ambrogelly, H. D. Becker, S. Bunjun, L. Feng, D. Tumbula-Hansen, M. Ibba, D. Korencic, H. Kobayashi, C. Jacquin-Becker, N. Mejlhede, B. Min, G. Raczniak, J. Rinehart, C. Stathopoulos, T. Li, and D. Soll, Acta Biochim. Pol. **48**, 313 (2001).

41) T. J. Bullwinkle and M. Ibba, *Aminoacyl-tRNA Synthetases in Biology and Medicine*, ed. S. Kim (Springer, Dordrecht, 2013) Chap. 2, p. 43.

42) R. Fukunaga and S. Yokoyama, J. Mol. Biol. **359**, 901 (2006).

43) T. Hussain, S. P. Kruparani, B. Pal, A. C. Dock-Bregeon, S. Dwivedi, M. R. Shekar, K. Sureshbabu, and R. Sankaranarayanan, EMBO J. **25**, 4152 (2006).

44) M. M. Aboelnga, J. J. Hayward, and J. W. Gauld, J. Phys. Chem. B **122**, 1092 (2018).
45) W. F. Waas and P. Schimmel, Biochemistry **46**, 12062 (2007).
46) M. M. Aboelnga, J. J. Hayward, and J. W. Gauld, ACS Catal. **7**, 5180 (2017).

Table

Table I.

Summary of nucleophile activators and reduction rate of hybrid ribozyme/protein catalysts (*i.e.*, LeuRS, ValRS, IleRS, PheRS, ProRS, and *Pyrococcus abyssi* ThrRS) and a variant aaRS (D342 of *Escherichia coli* LeuRS is replaced with Ala).

	Class	Species	Attached site	Activator	Reduction Rate	Reference
LeuRS	Class Ia	*T. thermophilus*	$O^{2'}$	3'-OH	-	
LeuRS	Class Ia	*E.coli*	$O^{2'}$	Δ(3'-OH)[a)]	10^4-fold	20)
LeuRS (D342A)	Class Ia	*E.coli*	$O^{2'}$	3'-OH	3-fold	20)
ValRS	Class Ia	*E.coli*	$O^{2'}$	Δ(3'-OH)[a)] and D283	10-fold	16), 22), Present study
IleRS	Class Ia	*E.coli*	$O^{3'}$	Δ(2'-OH)[b)] and D341	5-fold	16), 22), 35)
ProRS	Class IIa	*E.coli*	$O^{3'}$	Δ(2'-OH)[b)]	1000-fold	21)
ThrRS	Class IIa	*E.coli*	$O^{3'}$	Δ(H73)[c),d),e)]	10^4-fold	45),46)
ThrRS	Class IIa	*P. abyssi*	$O^{3'}$	2'-OH[d)]	-	43),44)
PheRS	Class IIc	*E.coli*	$O^{2'}$	Δ(3'-OH)[a)]	300-fold	36)

a) Δ(3'-OH) represents the replacement of the 3'-OH group of A76 with the 3'-H atom.
b) Δ(2'-OH) represents the replacement of the 2'-OH group of A76 with the 2'-H atom.
c) Δ(H73) represents the replacement of His73 with Ala.
d) Escherichia coli ThrRS acts as a protein enzyme (see Fig. 1(b)).
e) See the Supplemental Material, Fig. S1.

Figures

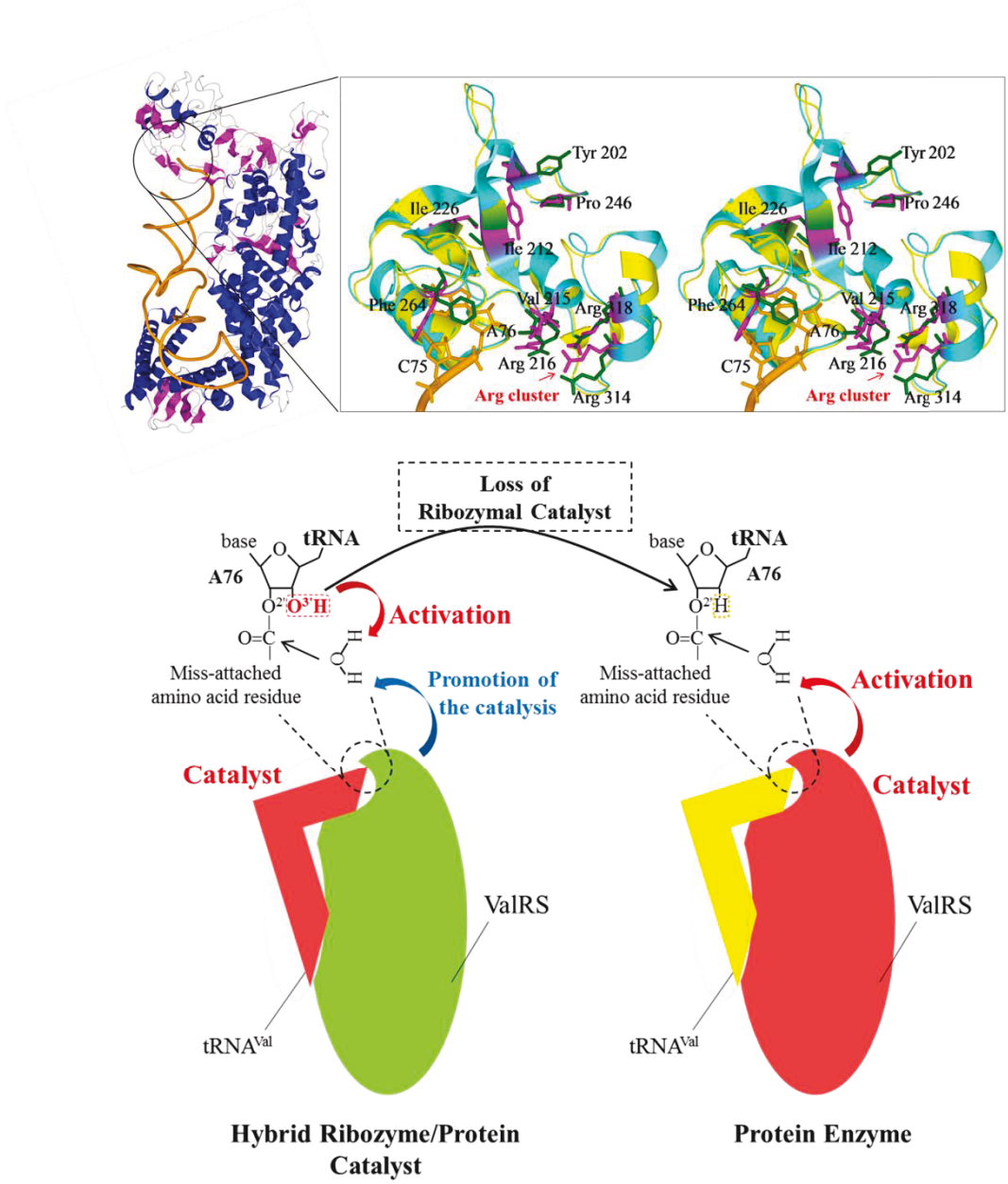


Fig. 1. (A) (left) Crystal structure of *Thermus thermophilus* ValRS in complex with $tRNA^{Val}$ (PDB code: 1IVS). (right) Catalytic site of the editing reaction (stereoview). The crystal structure of the complex (1IVS) is colored yellow (for amino acid and RNA backbones), green (for amino acid side chains), and orange (for nucleic acids). The crystal structure of the isolated CP1 domain (1WK9) is colored light blue (for amino

acid backbone) and magenta (for amino acid side chains). (B) Schematic representation of fundamental reaction schemes of hybrid ribozyme/protein catalyst (left) and protein enzyme (right). The black circles (broken line) show the catalytic site, and the macromolecules involving the catalysts are colored red. In this hybrid ribozyme/protein catalytic system (left), the ribozymal factor (*i.e.*, 3'-OH of A76 of $tRNA^{Val}$) activates the nucleophilic water molecule (as represented by the red arrow), and the protein (ValRS) moiety promotes the catalysis (the blue arrow).[15] In our previous preliminary study based on the atomistic modeled structure of the Leu system, which is closely related to the Val system, we suggested that the defective mutation of the ribozymal factor (*i.e.*, replacement of the aforementioned 3'-OH with 3'-H) could be compensated by an amino acid residue of ValRS.[15] In the present study, we theoretically constructed the atomistic structural model of the Val system [*i.e.*, the ValRS•threonyl-$tRNA^{Val}$ (mis-aminoacylated) complex], and analyzed the validity of our previous proposal that, for the Val system with a "defective" ribozymal activator of $tRNA^{Val}$, the protein (ValRS) moiety can activate the nucleophilic water molecule (red arrow) [this is referred to here as the protein enzyme (note that the definition of protein enzyme described here is different from that employed by Cech[19])]. As discussed in the text, this transition from the hybrid ribozyme/protein catalyst to the protein enzyme may fill a gap found in the evolutionary transition from the RNA world to the current RNP world, which could possibly occur in primordial biological macromolecular systems (see Sect. 2-3-7).

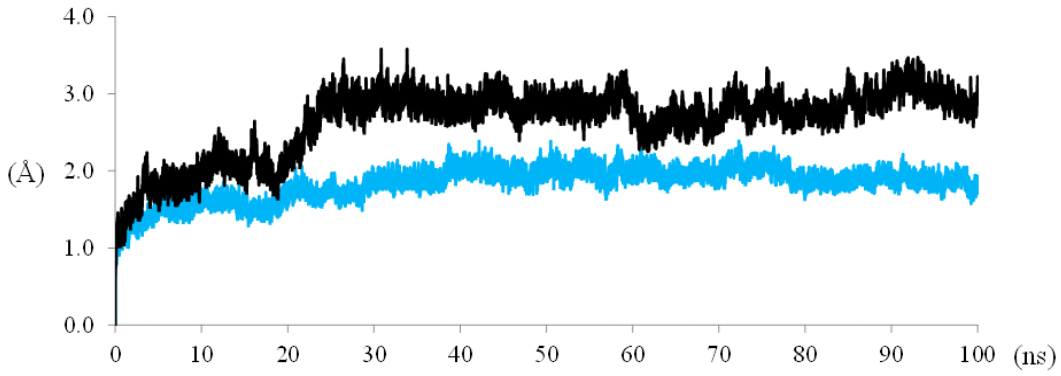


Fig. 2. Time evolution of the RMSDs calculated from the MD simulation of the ValRS•threonyl-tRNA^{Val} (mis-aminoacylated) complex. The horizontal and vertical axes show the simulation time (ns) and RMSD values (Å), respectively. For the heavy atoms of the entire structure of the complex (black line) and the CP1 domain moiety (blue line), the RMSDs are plotted.

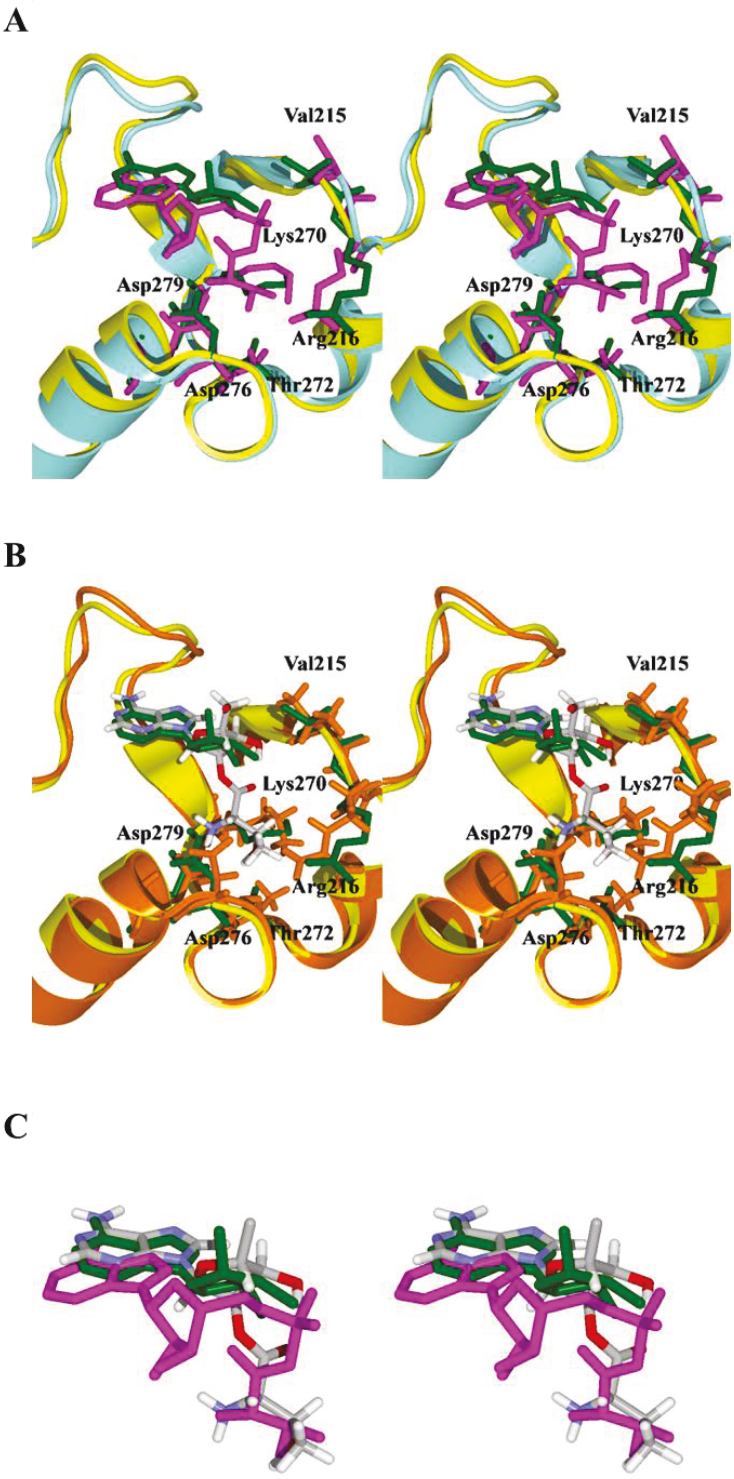


Fig. 3. (A) Stereoview of the superposition of the active sites of two crystal structures, *i.e.*, the structure of the isolated CP1 domain (its backbone is colored light blue, and its

side chains and threonyl-A76 are colored magenta) and the structure of the ValRS•tRNA^{Val} complex (its backbone is colored yellow, and its side chains and threonyl-A76 are colored green). (B) Stereoview comparison of an MD snapshot of the ValRS•valyl-tRNA^{Val} complex (its backbone is colored orange, and its side chain and threonyl-A76 are colored gray) with the crystal structure of the ValRS•tRNA^{Val} complex. (C) Stereoview of the superposition of the substrates in the MD snapshot, the ValRS•tRNA^{Val} complex, and the isolated CP1 domain.

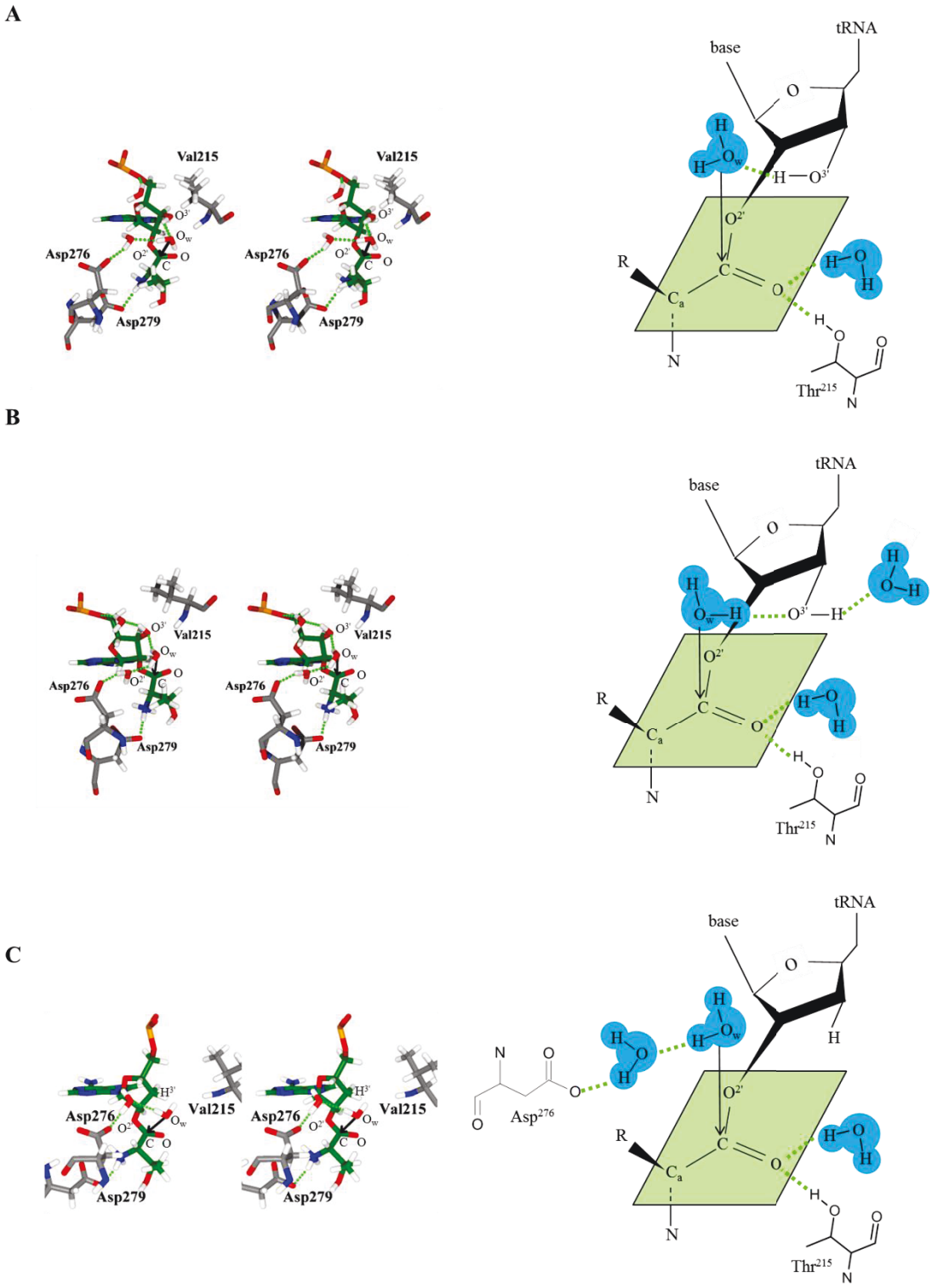


Fig. 4. (right) Schematic representation of configurations of the catalytic site in the

modeled structure of the ValRS•threonyl-tRNA^{Val} complex. (left) MD snapshots corresponding to the right panels. (A) Configuration where a nucleophilic water accesses the C atom of the carbonyl group of the substrate in the modeled structure of the ValRS•threonyl-tRNA^{Val} complex. When the H-gate is closed, the access of the nucleophilic water to the C atom of the carbonyl group is hindered (B) Catalytic configuration where the H-gate is opened in the modeled structure of the complex, allowing the nucleophilic water activated by 3'-OH to approach the C atom. (C) Configuration where a nucleophilic water accesses the C atom of the carbonyl group of the substrate in the modeled structure of ValRS in complex with the threonyl-tRNA^{Val} variant, which possesses 3'-H instead of 3'-OH. Here, the Asp276-coordinated water can activate the nucleophilic water even in the complex of ValRS and tRNA^{Val} with the defective ribozymal activator.

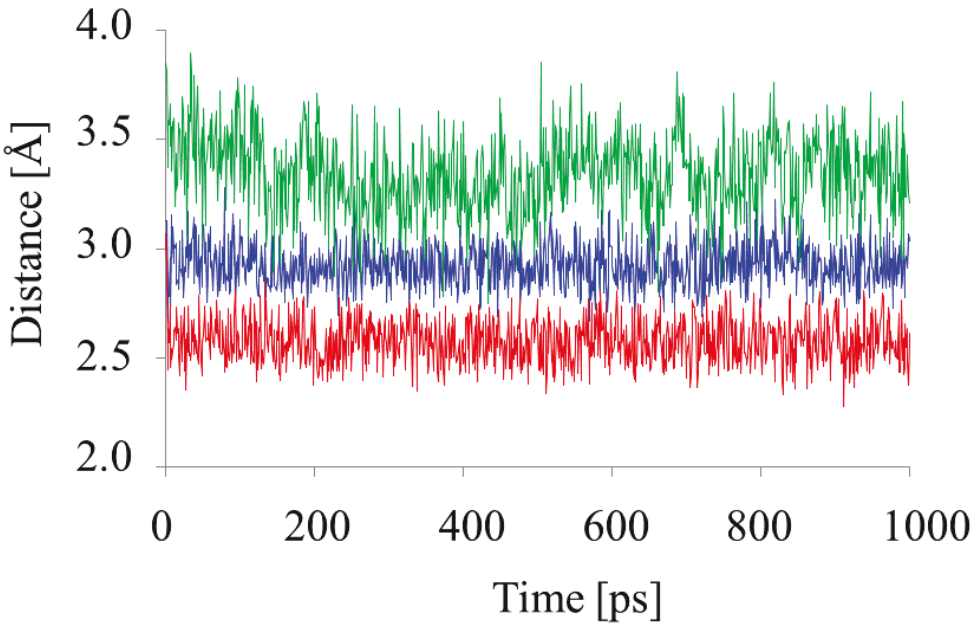


Fig. 5. Trajectories of the distance between the O_w and C atoms calculated in the three MD simulations of the ValRS•threonyl-tRNA^{Val} complex: the MD simulation without any constraints (green), the MD simulation with only the distance constraint between the O_w and C atoms (blue), and the MD simulation with the identical distance constraint and the dihedral angle constraint for H-gate opening (red). The horizontal and vertical axes show the simulation time (ps) and the distance (Å), respectively. Only the distance constraints did not lead the water molecule to the nucleophilic attack, thus suggesting that the opening of the H-gate is required to the initiation for the editing reaction.

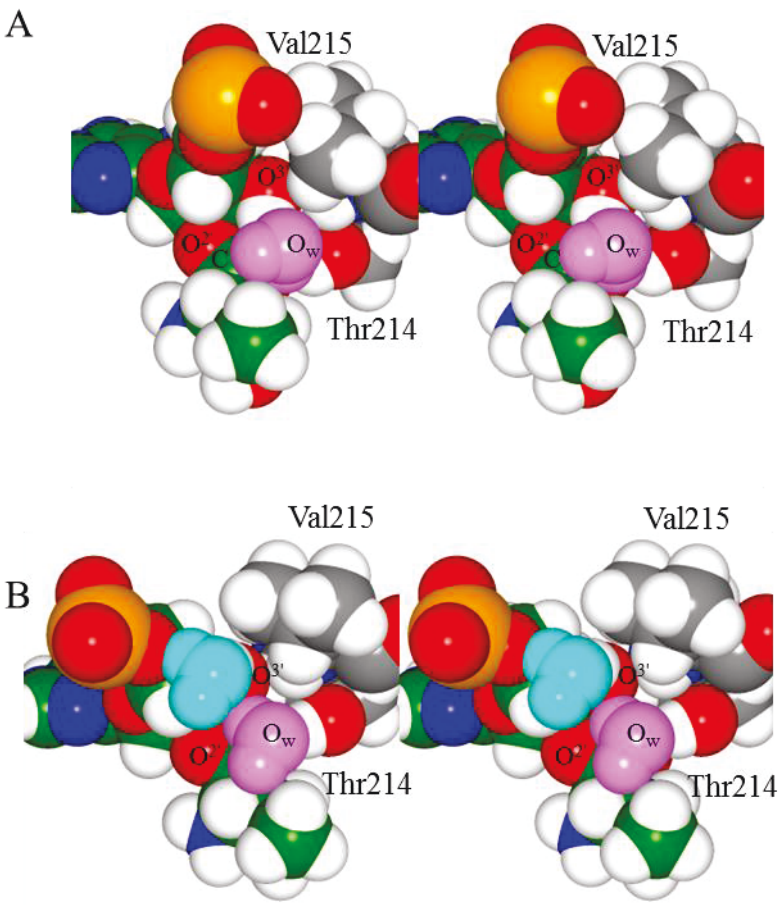


Fig. 6. Stereoview of a snapshot of the MD simulation (A) initiated from the structure where the side chain conformation of Val215 was set to that observed in the crystal structure of the isolated CP1 domain (1WK9) and (B) initiated from the structure where the side chain conformation of Val215 was set to that observed in the crystal structure of the complex (1IVS). In (B), the H-gate is open, corresponding to the configuration depicted in Fig. 4B. The molecules are represented by CPK models. A phosphorus atom is colored orange. The nucleophilic water is colored magenta. For (B), the water that opens the H-gate is colored cyan.

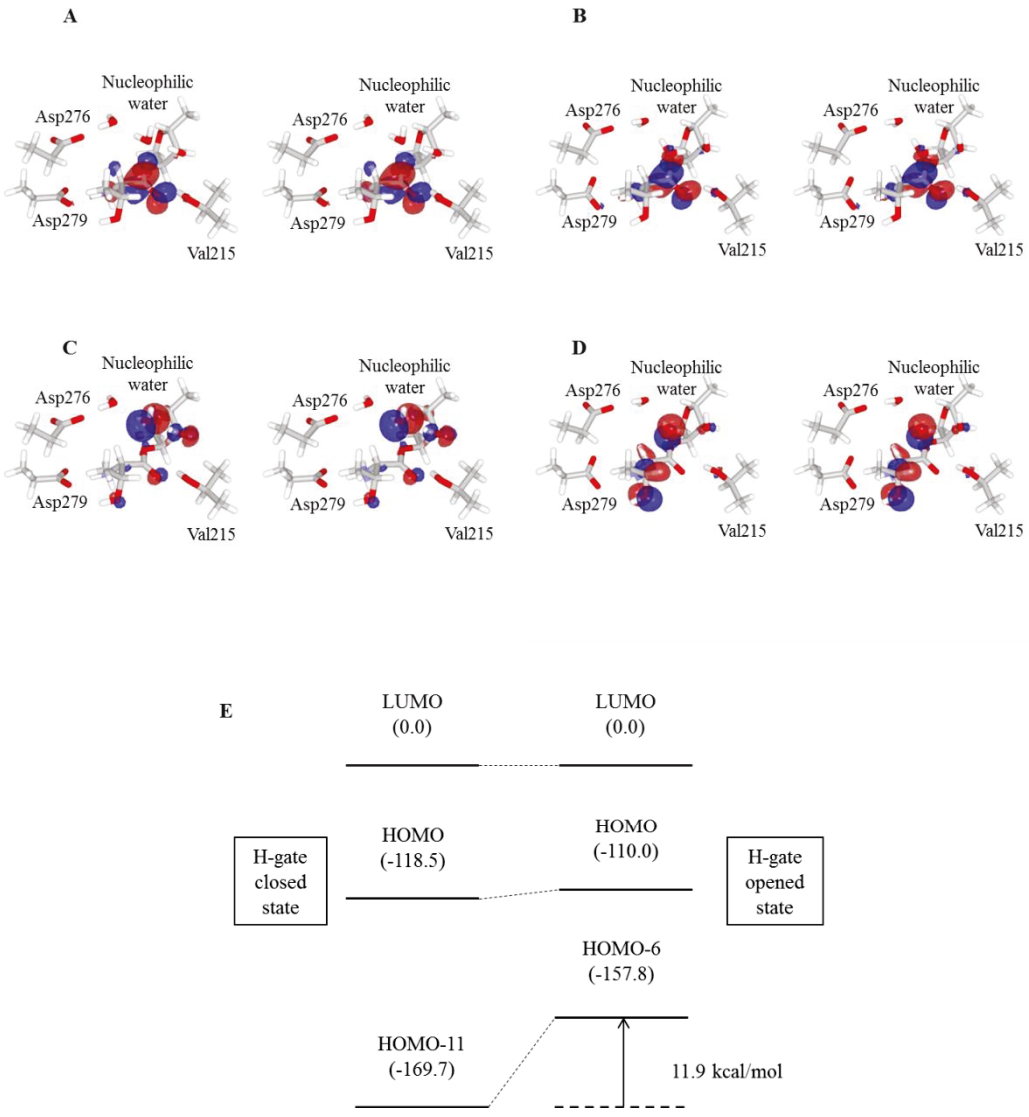


Fig. 7. Stereoview of molecular orbital (MO) of the catalytic site including the nucleophilic water molecule. LUMO for (A) H-gate closed and (B) H-gate opened conformations. HOMO-11 for (C) H-gate closed and HOMO-6 for (D) H-gate opened conformations. The ribose and threonine moieties of the substrate, Val215, Asp276, Asp279, and the three water molecules are included as the QM region. (E) Energy diagrams concerning the LUMOs and HOMO-11 (HOMO-6) in H-gate closed state (H-gate opened state).

Appendix of Chapter 2

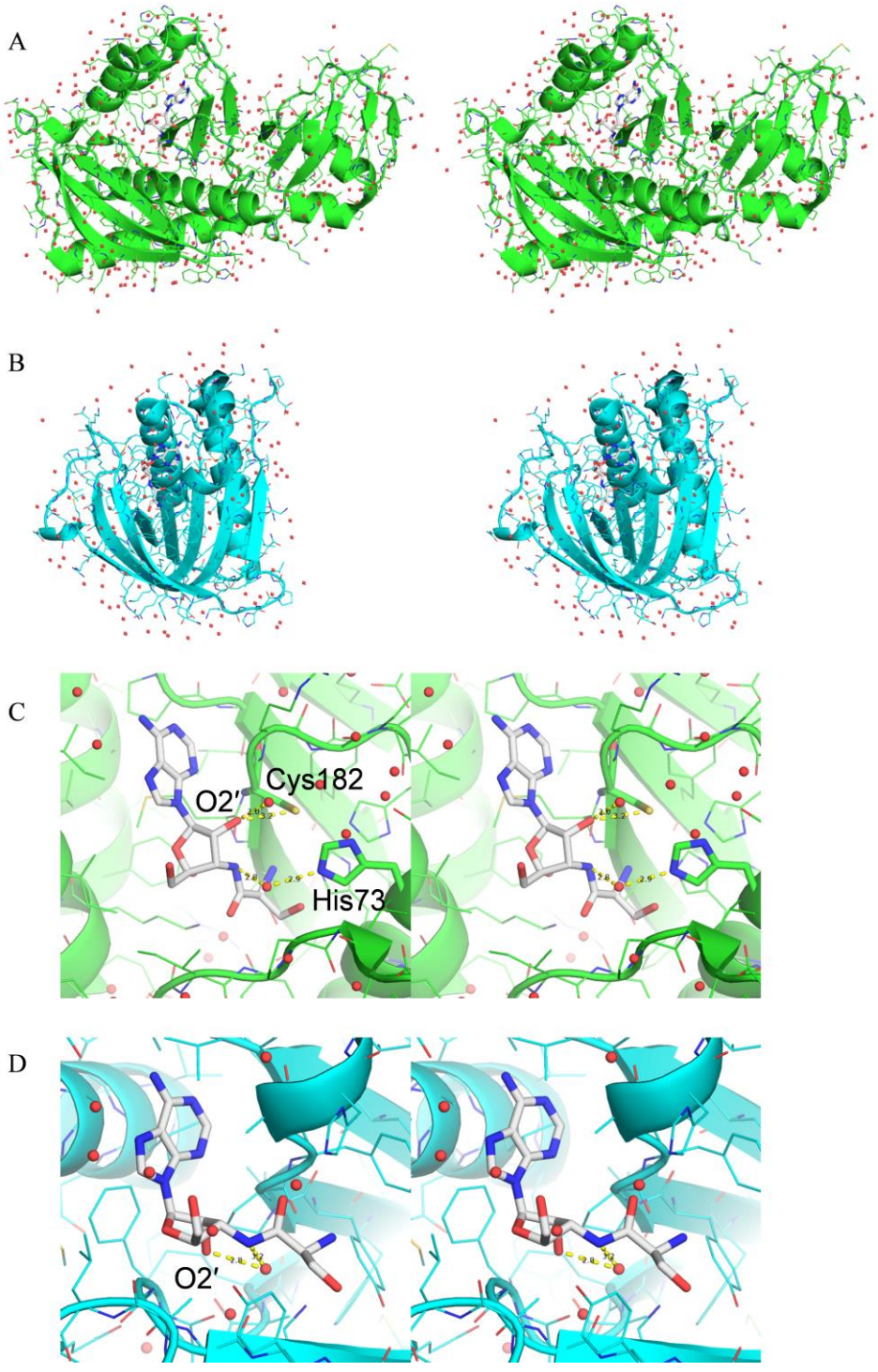


Fig. S1. (A) Stereoview of the crystal structure of the editing domain of *Escherichia coli* ThrRS (in which Cα atoms are colored green) in complex with seryl-3′-adenosine (in which the carbon atoms are colored grey) (PDB code: 1TKY)[1]. (B) Stereoview of the crystal structure of the editing domain of *Pyrococcus abyssi* ThrRS (in which Cα atoms are colored cyan) in complex with seryl-3′-adenosine (in which the carbon atoms are colored grey) (PDB code: 2HL1)[2]. (C) Stereoview of the active site in the editing domain of *Escherichia coli* ThrRS (in which Cα atoms are colored green) in complex with seryl-3′-adenosine (in which the carbon atoms are colored grey) (PDB code: 1TKY)[1]. (D) Stereoview of the active site in the editing domain of *Pyrococcus abyssi* ThrRS (in which Cα atoms are colored cyan) in complex with seryl-3′-adenosine (in which the carbon atoms are colored grey) (PDB code: 2HL1)[2].

References

1) W. F. Waas and P. Schimmel, Biochemistry **46**, 12062 (2007).
2) T. Hussain, S. P. Kruparani, B. Pal, A. C. Dock‐Bregeon, S. Dwivedi, M. R. Shekar, K. Sureshbabu, and R. Sankaranarayanan, EMBO J. **25**, 4152 (2006).

Chapter 3

Comprehensive Discussions

3-1. Dynamical properties in domain rotations

As mentioned in the previous chapters, the CP1 domain of ValRS could be flexible and fluctuated in terms of the main body of ValRS, which may induce the rotations of the CP1 domain with respect to the linker segments (two β-strands) to connect to the main body. Thus, the precise atomic positions of the CP1 domain moiety of ValRS have not been determined up to date; in fact, outliers of the ϕ and ψ angles are found in the CP1 domain of ValRS. Accordingly, for the CP1 domain that was isolated from ValRS, the crystallographic analysis was performed, and thereby the definite atomic coordinates of the CP1 domain was determined. However, we need to swap the CP1 domains in the complex with $tRNA^{Val}$, with the isolated CP1 domain structure.

In this thesis, we have actually conducted such a structural modeling, and thereby have analyzed the 3D structure of the CP1 domain and the catalytic mechanisms of the post-transfer editing reaction. As we expected, the time evolution of RMSD values of the CP1 domain was drifted with respect to the main body (see Chapter 2). This could be due to the rotational motions of the CP1 domain. Accordingly, we preliminarily present the dynamical aspects of the domain rotations of the CP1 domain in terms of the main body of ValRS.

In order to examine the properties in the domain motions, we extracted some snapshots of the complex from the MD trajectory prior and posterior to the drifting of RMSD values of the entire complex. As a consequence, we identified the rotation of the CP1 domain as shown in Fig. 8. Such domain motions could be relevant to the functional mechanisms of ValRS, although it can be a future issue.

3-2. Hybrid *ab initio* QM/MM MD simulation

In the last chapter, we discussed the mechanisms of the initial stages of the editing reaction by the complex of ValRS and the mis-aminoacylated $tRNA^{Val}$. Our goals are as follows; i) elucidation of the post transfer editing mechanism of the complex of ValRS and mis-aminoacylated $tRNA^{Val}$, and ii) that of mechanisms to preserve the catalytic activity of the complex of ValRS and the “defective” mutant of the mis-aminoacylated $tRNA^{Val}$ where O2' of A76 was replaced with H2'.

In order to achieve these purposes, we adopted hybrid *ab initio* QM/MM MD

simulation, which is a state-of-the-art methodology in theoretical investigation techniques of biological macromolecular systems. To analyze the mechanisms of the editing reaction of ValRS in the complex with the wild type or "defective" mutants of the mis-aminoacylated $tRNA^{Val}$, we have currently performed five hybrid *ab initio* QM/MM MD simulations, where distinct five conditions, classified into two types, are imposed, as follows: *ab initio* QM/MM MD simulations of the post transfer editing reactions of ValRS and i) the wild type threonyl-$tRNA^{Val}$, and ii) the defective threonyl-$tRNA^{Val}$. These conditions are further divided into the following subclasses: The initial stage of the reaction of the complex with the wild type $tRNA^{Val}$ is activated by i-a) O2' of A76, i-b) the Asp276 side chain, i-c) the phosphate group of A76, ii-a) the Asp276 side chain, and ii-b) the phosphate group of A76.

As a preliminary analysis employing our QM/MM MD simulation of the post-transfer editing reaction by the complex of ValRS and threonyl-$tRNA^{Val}$, we have obtained the activation barrier of the three pathways, *i.e.*, i-a) 26.1 kcal/mol, i-b) 32.1 kcal/mol, and i-c) 31.2 kcal/mol, which shows that O2' of A76 is more efficient, by ~10^4-fold, than the other two pathways (*i.e.*, Asp276 and the phosphate group of A76). Thus, we have indicated that the pathway exhibiting the lowest energy barrier is the pathway activated by O2' of A76, which means that the post-transfer editing reaction by the complex of ValRS and threonyl-$tRNA^{Val}$ was ribozymal. Currently, we are further performing *ab initio* QM/MM MD simulations of the other two pathways of the complex with the defective mutant of the threonyl-$tRNA^{Val}$ where O2' of A76 is replaced with H2'.

Notably, in this thesis, to improve the codes of the computer programs that were employed in the afore-mentioned hybrid *ab initio* QM/MM MD calculations, we newly incorporated additional MM force field parameters for the water molecules located in the QM/MM boundaries. More detailed descriptions are provided in the subsequent subsection.

3-3. Implementation of a modified force field of the water molecule

In this thesis, we employed TIP3P as the force field of the water molecule for the molecular dynamics simulation. It should be noted here that TIP3P does not involve the van der Waals radius of the hydrogen atoms in the water molecule. To appropriately evaluate the interactions between the MM water and QM atoms in the vicinity of the QM/MM boundaries, the van deer Waals interactions of the hydrogen atoms in the water molecule are important. Accordingly, we added the Lennard-Jones potential of the hydrogen atoms in those water molecules that are located in the vicinity of the

boundaries of the QM and MM regions, by modifying the code for the MD calculation in the Amber package (Fig. 9).[1-3)]

3-4. Development of an algorithmic scheme for automatic attachment of hydrogen atoms to crystal structures

For the construction of the structural models, we mostly employ the atomic coordinates that were provided by the X-ray crystal structure analyses. As is well known, the hydrogen atoms are not observable in most of X-ray crystal structure analyses. In this thesis, we developed a new program that automatically generates the coordinates of the hydrogen atoms. The characteristic feature of our program is to explore the stable positions of the hydrogen atoms by hindering the steric clush with the surrounding heavy atoms (Figs. 10-14). Also, we employed this program in this thesis, to build the structural models of topoisomerase IIβ (topoIIβ) and eukaryote tetrameric cytidine deaminase (TCDA) for conducting hybrid *ab initio* QM/MM MD simulations of the catalytic reactions (in progress). In the subsequent section, these on-going works are briefly described.

3-5. Applications of hybrid *ab initio* QM/MM MD simulations

In order to analyze the catalytic mechanisms of topoIIβ and TCDA, we built atomistic structural models of these molecular systems, and performed MD simulations to relax the modeled structures in the fully-solvated environments.

topoIIβ binds to supercoiled DNA, and relaxes the topology of the DNA structure by cleaving a double stranded (ds) DNA (i.e., the two phosphate backbones are cleaved), passing another dsDNA through the cleaved dsDNA (Fig. 15), and finally ligating the two phosphate backbones that are cleaved in the first stage of the reaction cycle (Fig. 16). This reaction and function of topoIIβ are essential, for example, for proceeding the transcriptional process, and thereby the enzyme contributes to the transcriptional regulation.

However, the mechanisms of the catalytic reaction cycle have been unknown up to date, because dsDNA bound to topoIIβ was cleaved in the crystal structure, which means that the initial phase of the catalytic reaction cycle occurred in the crystal. Accordingly, in this thesis, we built a structural model of the complex of topoIIβ and dsDNA prior to the cleavage of the phosphate backbones that occurred in the crystal structures.

TCDA catalyzes the reaction of the deamination of a base, and thereby changes cytosine to uridine in mRNA, which thus repairs the mutational defect in the base

sequence of a gene (Fig. 17). Accordingly, this reaction is termed the editing, and is important to ensure the fidelity of the transcription. The reaction has also been studied for the purpose to artificially change the codons in mRNA and thereby has been examined for employing it as a medical gene therapeutic technique. For understandings and applications of the mechanisms of the catalytic reaction cycle by TCDA, we conducted a structural modeling to build a fully-solvated structure of TCDA.

In this manner, we obtained atomistic models of the aforementioned molecular systems, topoIIβ and TCDA, and then performed hybrid *ab initio* QM/MM MD simulations in this thesis (still in progress). Our strategy is further applicable to various biological macromolecular systems to elucidate the functional mechanisms based on the dynamical properties of the electronic structures. This will lead us to essential understandings of our crucial question, *i.e.*, what is life ?

References

1) Daniel J. Price and Charles L. Brooks III, J. Chem. Phys. **121**, 10096 (2004).
2) Harry A. Stern, F. Rittner, B. J. Berne, and Richard A. Friesner, J. Phys. Chem. **115**, 2237 (2001).
3) Asger Halkier, Wim Klopper, Trygve Helgaker, Poul Jo/rgensen, and Peter R. Taylor, J. Chem. Phys. **111**, 9157 (1999).

Figures

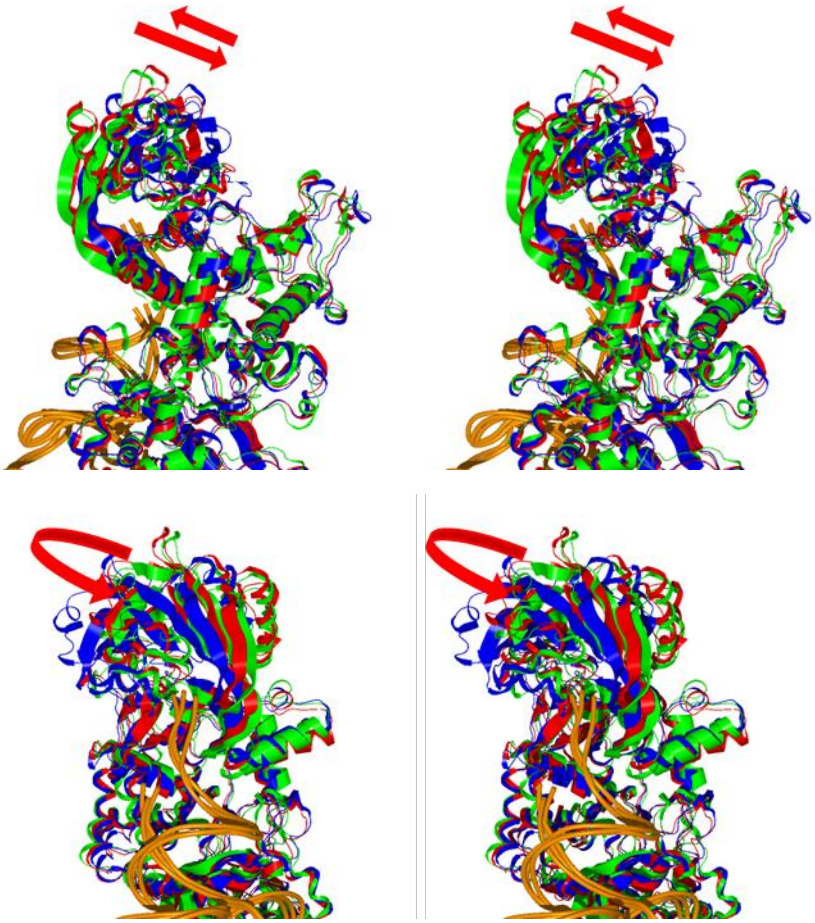


Fig. 8. Stereoview representing geometrical features in rotations of the CP1 domain with respect to the main body of ValRS (the upper and lower panels are rendered from different views by rotating around the vertical axis involved in a paper surface, by 90 degrees). In this figure, the following three snapshots extracted from the MD trajectories were employed to be rendered; i) the initial structure of the complex (green), ii) and iii) snapshots of the MD simulation prior and posterior to the drift of RMSD values in terms of the entire structure of the complex (blue and red, respectively). To characterize the rotation of the CP1 domain, those three structures were superimposed in terms of the heavy atoms of the ValRS main body and $tRNA^{Val}$, and thus the domain rotation was identified as shown in red arrows.

```
if( i== HydrogenAtomBoundaryWat .and. j==InteractiveAtom
  .or.
  j== HydrogenAtomBoundaryWat .and. i== InteractiveAtom) then

          dfee=df
          f12 = 2*sqrt(0.1521*0.046)*(3.15061/2+0.2)**12*r6*r6
          f6 = 2*sqrt(0.1521*0.046)*(3.15061/2+0.2)**6*r6
          df = dfee + (12.d0*f12 - 6.d0*f6)*delr2inv
end if

dfx = delx*df
dfy = dely*df
dfz = delz*df
```

Fig. 9. A part of the source code (written by FORTRAN) of sander.MPI in the Amber 9 package, that was modified in this thesis. This is a part for the calculations of the interactions between a hydrogen atom and other MM atoms employing the force field parameters that was proposed by the Brooks group. The indices *i* and *j* specify the two atoms for which the force should be evaluated. For the atom pairs specified by *i* and *j*, which are provided by the user, the van deer Waals interactions involving hydrogen atoms of the water molecules in the vicinity of the QM/MM boundary, that are specified by the user, should be added to the forces of the other atom pairs.

```
int  _nearAtom(int Num_A_all,int fcs_A,struct Atom *A,int *cntH,int InAtNum[],int *testcnt)
{
                    [skipped]

          tent=sqrt(pow(A0->crd[count][0]-A->crd[fcs_A][0],2)
                    + pow(A0->crd[count][1]-A->crd[fcs_A][1],2)
                    + pow(A0->crd[count][2]-A->crd[fcs_A][2],2) );
                     if(
                              (tent < 2.9)
                              && ( A->Residue_Num[fcs_A] != A0->Residue_Num[count] )
                              && ((A0->Atm[count][0] == 'O' || A0->ChainID[count][0] =='O' )
                              && ( strncmp(A0->Residue_Name[count],"WAT",3) != 0 ))
                    ){
                              count2++;
                              if(count2==1){
                                        InAtNum[0]=A->Atom_Num[fcs_A]-1;
                                        *testcnt=*testcnt+1;
                              }
                              x_p++;
                              InAtNum[count2]=A0->Atom_Num[count]-1;
                              if(count2==1){
                                        *cntH=*cntH+1;
                              }
                    }

                    [skipped]
}
```

Fig. 10. A part of the source code (written by C) of our automatic hydrogen attachment program that was developed in this thesis. This is a part to restore the information for the calculations of the coordinates of the movable hydrogen atoms, such as the atom types, residue types, and atomic coordinates of the atoms surrounding the movable hydrogen atoms.

```
void revolve(struct Atom A1, int memC, int memO, int mem1, double crd[][3])
{
                    [skipped]

          for (angle = 0; angle < 360; angle = angle + 1){
                    crd[angle][0] = OH.x + (HHy.x*cos(angle*pi / 180.00000))
                    + (HP2.x*sin(angle*pi / 180.00000));
                    crd[angle][1] = OH.y + (HHy.y*cos(angle*pi / 180.00000))
                    + (HP2.y*sin(angle*pi / 180.00000));
                    crd[angle][2] = OH.z + (HHy.z*cos(angle*pi / 180.00000))
                    + (HP2.z*sin(angle*pi / 180.00000));
          }

                    [skipped]

}
```

Fig. 11. A function in the source code (written by C) of our automatic hydrogen attachment program that was developed in this thesis. This function generates all the possible coordinates of a hydrogen atom for which the position should be determined, by rotating the hydrogen atom around the covalent bond of the heavy atoms, and then restores the ensemble of the created atomic coordinates.

```
struct sStore SER(int atom_num, struct Atom A, int Num_A, int InAtNum[], struct Atom *B){

                                [skipped]

        for (count = 0; count < 10; count++){
                InAtNum2[count] = InAtNum[count];
        }

                                [skipped]

        int res;
        get_atom_memory(atom_num - 1, &A1, &atom_num1);
        res = A1->Residue_Num[atom_num1];

                                [skipped]

        revolve(A, memC, memO, mem1, rcrd1);

                                [skipped]
}
```

Fig. 12. A function in the source code (written by C) of our automatic hydrogen attachment program that was developed in this thesis. This function generates all the possible coordinates of H^{γ} atom of a serine (Ser) residue, for which the position should be determined, by rotating the hydrogen atom around the O^{γ}-C^{β} covalent bond (*i.e.*, the "revolve" function), and then restores the ensemble of the created atomic coordinates.

```
struct sStore MET(int atom_num, struct Atom A, int Num_A, int InAtNum[], struct Atom *B){

        [skipped]

        for (count = 0; count < 10; count++){
        InAtNum2[count] = InAtNum[count];
        }

        [skipped]

        revolve(A, memC, memS, mem1, rcrd1);
        revolve (A, memC, memS, mem2, rcrd2);
        revolve (A, memC, memS, mem3, rcrd3);

        [skipped]

}
```

Fig. 13. A function in the source code (written by C) of our automatic hydrogen attachment program that was developed in this thesis. This function generates all the possible coordinates of three H^{ε} atoms (forming a methyl group) of a methionine (Met) residue, for which the positions should be determined, by simultaneously rotating the three hydrogen atoms around the C^{ε}-S^{δ} covalent bond (*i.e.*, the "revolve" function), and then restores the ensemble of the created atomic coordinates.

```
struct Vector WATER(struct Atom A, int mem1, int an){
        double rotate_crd2[360][3];
        int number_max;

        if (strncmp(A.Atom_Name[mem1], "H1", 2) == 0){
                revolve(A, member1 + 1, member1 - 1, member1, rotate_crd2);
        }
        else if (strncmp(A.Atom_Name[mem1], "H2", 2) == 0){
                revolve(A, member1 - 1, member1 - 2, member1, rotate_crd2);
        }
        Vector new_crd;
        number_max = maxdistance(rcrd2, A, an);

        new_crd.x = rotated_crd2[number_max][0];
        new_crd.y = rotated_crd2[number_max][1];
        new_crd.z = rotated_crd2[number_max][2];

   return new_crd;
}
```

Fig. 14. A function in the source code (written by C) of our automatic hydrogen attachment program that was developed in this thesis. This function generates all the possible coordinates of two hydrogen atoms in each water molecule, for which the positions should be determined, by simultaneously rotating the two hydrogen atoms with the H-O-H angle preserved (109.5°) (*i.e.*, the "revolve" function), and then restores the ensemble of the created atomic coordinates.

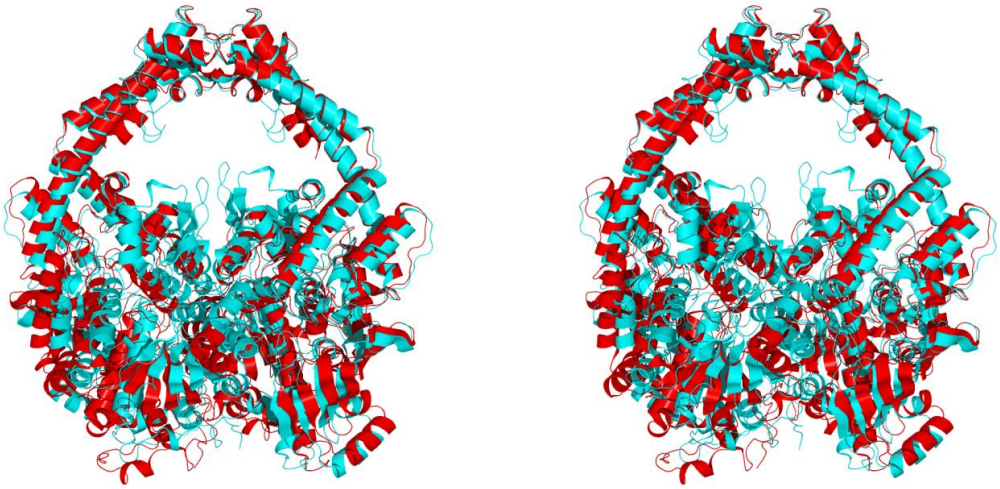
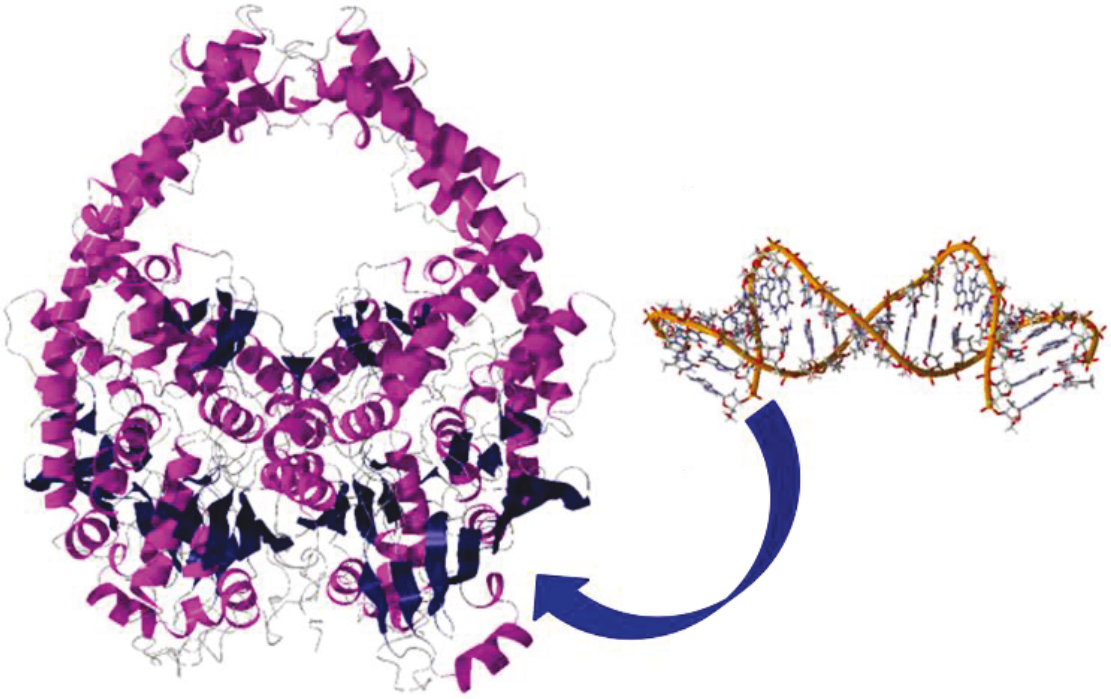


Fig. 15. Superposition of the crystal structures of human topoIIβ (red) and yeast topoII (cyan). The distinct types of states in the reaction cycle were determined in the crystal structures, *i.e.*, the open and closed conformations of human topoIIβ and yeast topoII, respectively.

a)



b)

Fig. 16. Docking simulation of human topoIIβ and the target DNA employing the restrained (targeted) energy minimization (optimization), where the B-type DNA (a) was utilized as the initial conformation of the DNA and docked with topoIIβ. The helical axis of the bound DNA molecule (b) was bent through the formation of the complex with topoIIβ.

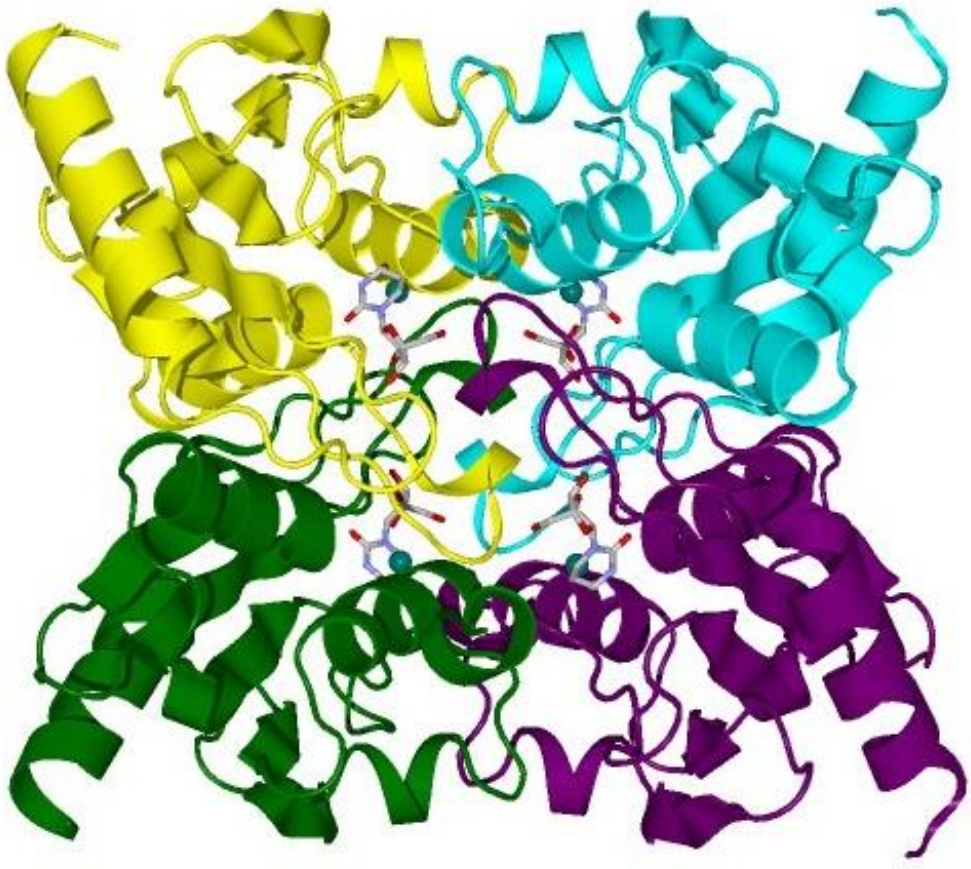


Fig. 17. Crystal structures of mouse tetrameric cytidine deaminase (TCDA), which catalyzes the hydrolytic deamination of aromatic bases in mRNA. (i.e., the transition from cytidine to uridine occurs). Tetrameric cytidine deaminase (TCDA) consists of homo tetramer (the subunits are depicted by distinct colors).

Acknowledgements

I am grateful to Professor Masaru Tateno (University of Hyogo) for his continuing guidance, encouragement, valuable scientific discussion, and advice on my life. I would thank Professor Jiyoung Kang (University of Hyogo; Present address: Medical School, Yonsei University, Korea) for her intimate and kind supervise of my work. I also appreciate fruitful discussions and insightful advice of Professor Kazuyuki Takai (Ehime University), Professor Thomas Simonson (Ecole polytechnique palaiseau, Biology, France), Professor Tohru Yoshihisa, Professor Tsunehiro Mizushima (University of Hyogo), Mr. Hiroshi Nishigami, and Mr. Ryu-ichiro Terada.