Structural modelling of the complex of leucyl-tRNA synthetase and mis-aminoacylated tRNA^{Leu}

Yohsuke Hagiwara[1, 2], Osamu Nureki[3], and Masaru Tateno[1, 2]

[1] Center for Computational Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8571, Japan, Tel: (+81)-29-853-6496, Fax: (+81)-29-853-6496

[2] Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba Science City, Ibaraki 305-8571, Japan

[3] Institute of Medical Science, University of Tokyo, Shirokanedai 4-6-1, Minatoku, Tokyo 108-8639, Japan, Tel: (+81)-3-6409-2125, Fax: (+81)-3-6409-2127

Corresponding author: tateno@ccs.tsukuba.ac.jp

Keywords: Aminoacyl-tRNA synthetase; Error-editing reaction; Molecular dynamics simulation; Structural modelling; Ribose; Water.

Abbreviations: aaRS, aminoacyl-tRNA synthetase; leucyl-tRNA synthetase, LeuRS; valyl-tRNA synthetase, ValRS; MD, molecular dynamics; connective polypeptide 1, CP1

Abstract

To assure fidelity of translation, class Ia aminoacyl-tRNA synthetases (aaRSs) edit mis-aminoacylated tRNAs. Mis-attached amino acids and structural water molecules are not included simultaneously in the current crystal structures of the aaRS•tRNA complexes, where the 3′-ends (adenine 76; A76) are bound to the editing sites. A structural model of the completely solvated leucyl-tRNA synthetase complexed with valyl-$tRNA^{Leu}$ was constructed by exploiting molecular dynamics simulations modified for the present modelling. The results showed that the ribose conformation of A76 is distinct from those observed in the above-mentioned crystal structures, which could be derived from structural constraints in a sandwiched manner induced by the mis-attached valine and $tRNA^{Leu}$.

1. Introduction

Aminoacyl-tRNA synthetases (aaRSs) catalyse the attachment of their cognate amino acid to the 3′-end of the specific tRNA (aminoacylation). This reaction proceeds as: first, an amino acid is activated to an aminoacyl adenylate by transfer of ATP with generation of pyrophosphate; second, the amino acid moiety of the aminoacyl adenylate is transferred to the 3′-end of the specific tRNA. According to the primary and tertiary structures, aaRSs are divided into classes I and II, and further subdivided into subclasses, a, b and c, within each class [1]. The fidelity of translation is assured by the strict discrimination of cognate from non-cognate amino acids. However, for the leucine, isoleucine, valine, threonine, alanine and phenylananine systems, each of which is structurally similar to some other systems, their cognate enzymes, i.e. leucyl- (LeuRS), isoleucyl- (IleRS), valyl- (ValRS), threonyl- (ThrRS), alanyl- (AlaRS) and phenylalanyl- (PheRS) tRNA synthetases (LeuRS, IleRS and ValRS belong to class Ia, ThrRS and AlaRS to class IIa and PheRS to class IIc), have difficulties in the strict discrimination of their specific amino acids, producing mis-activated amino acids or mis-aminoacylated tRNAs. Two types of editing, pre-transfer editing and post-transfer editing, correct mis-activated amino acids or mis-aminoacylated tRNAs, respectively; a misactivated amino acid is hydrolysed to the amino acid and AMP by the pre-transfer

editing pathway and a mis-aminoacylated tRNA is hydrolysed to the amino acid and tRNA by the post-transfer editing pathway [1–10].

Several mutational analyses have been performed to elucidate the mechanisms of these editing reactions [3, 10, 11]; however, the reaction mechanisms remain unclear. To date, two crystal structures have been determined for class Ia aaRSs in complex with their cognate tRNAs, in which the 3′-termini are bound to the editing sites [12, 13]. However, in these crystal structures, no amino acid is attached to adenine 76 (A76) of tRNAs. Furthermore, even though water is assumed to participate in the nucleophilic attack, no crystallographic water molecule has been identified in the editing site, since the resolutions of the X-ray crystallographic data on LeuRS and ValRS are 3.3 and 2.9 Å, respectively. Thus, the absence of crystal structures containing all the molecular components essential for the reaction causes difficulties in the elucidation of the detailed mechanisms. To address this issue, it is necessary to obtain three-dimensional (3D) structures of complexes of aaRSs and mis-aminoacylated tRNAs, including ordered water molecules in the editing sites using computational structural modelling.

Here, using the crystal structure of the *Thermus thermophilus* LeuRS in complex with $tRNA^{Leu}$, we built a 3D structural model of LeuRS in complex with valyl-$tRNA^{Leu}$. With regard to the conformation of the ribose moiety of A76, which binds to the catalytic

pocket in the reaction, we have found differences among the above-mentioned crystal structures and the modelled structure. These differences arise from the presence of the structural constraints induced by mis-attached amino acid and $tRNA^{Leu}$, either of which is absent in the crystal structures, while in the LeuRS•valyl-$tRNA^{Leu}$ complex, the ribose moiety topologically connects those two moieties in a sandwiched manner.

2. Materials and Methods

2. 1 System setup

In this study, two systems were used: the *Thermus thermophilus* LeuRS in complex with 2′-(L-norvalyl) amino-2′-deoxyadenosine as an inhibitor (Protein Data Bank (PDB) accession code 1OBC), and the LeuRS in complex with $tRNA^{Leu}$. In the latter complex, the 3′-terminal is bound to the active site of editing (PDB accession code 2BYT) [3, 12]. The former system (1OBC) was used to establish an algorithm to identify ordered water molecules in the editing site. For this purpose, the connective polypeptide 1 (CP1) domain (amino acid residues 228–415) was truncated to reduce computational costs; although in the crystal structure of LeuRS in complex with an inhibitor, 2′-(L-norvalyl) amino-2′-deoxyadenosine (Nva2AA; see Fig. 1a), to prevent the catalytic reaction, the inhibitor is bound to the active site in the CP1 domain. We replaced it with the ‘true’

substrate as follows: N2′ was replaced with an O atom and the norvaline moiety with valine. The truncated CP-1 domain was immersed in a box of water molecules modelled by TIP3P water, and seven Na^+ ions were added to neutralize the system. Thus, the total number of atoms included in the solvated domain was 24,253. The latter system (2BYT) was used to construct a model of the LeuRS•valyl-$tRNA^{Leu}$ complex, including the solvent water. In this case, the complex of the enzyme and $tRNA^{Leu}$ was used to attach valine to the 3′-end of tRNA and to identify hydration water around the editing site. This complex was immersed in a solvent box consisting of 49,587 water molecules and the periodic boundary condition was used, where the size of the unit box was 103.0 × 138.3 × 117.1 $Å^3$. Thus, the total number of atoms in the system was 165,739.

All simulations were performed using the AMBER 9 software package [14]. The parm99 force field was applied to all atoms in the system. Electrostatic interactions were calculated by the particle-mesh Ewald (PME) method [15] 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 [16] 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 [17]. All atom types and partial charges assigned for LeuRS were taken from AMBER ff99 force filed. The

details of the parameterization for the valyl-tRNA^{Leu} moiety are described in Supplementary material (S1).

2. 2 Identification of hydration water in the active site of editing

We used a scheme consisting of the following five stages to predict ordered water molecules in the active sites. Computational details are described in Supplementary material (S1).

(i) The ligand-binding pocket of a protein is immersed *in the absence of a ligand*, by performing MD simulations with explicit solvent water molecules.

(ii) The molecular volume and atomic charges of the ligand are then reduced to ~0, following which the *ligand* is placed back onto the immersed binding pocket. Accordingly, the coordinates of the ligand and water molecules observed in the binding pocket are overlaid at the beginning of the subsequent simulation.

(iii) The volume and atomic charges of the ligand are gradually increased in the MD simulation to exclude water molecules that are overlaid with the ligand and do not contribute to the protein•ligand binding. For this purpose, we introduce the following modified energy function: $E(\lambda) = (1-\lambda)E_0 + \lambda E_1$. Here, λ is a scaling factor that varies from 0 to 1; E_0 denotes the total energy of the system, where the molecular

volume and partial charges of the ligand are set to ~0; $E(\lambda)$ represents the total energy of the system, where λ is varied in the MD simulation and E_1 denotes the total energy of a system with the ligand volume and charges set to the original values as defined in the force field. In this stage, the ligand atoms are positionally constrained using a harmonic potential. Thus, the aim of this stage is to exclude water molecules that are not required for interfacial hydrogen bond networks in the protein•ligand complex.

(iv) Conformational searches for the interfacial structures in the protein•ligand complex are further performed using the above-mentioned modified energy function without any positional constraints for both the protein and the ligand until λ reaches 1.

(v) For further equilibration of the protein•ligand complex and the solvent structure around it, standard MD simulations are performed using a standard energy function.

3. Results and Discussion

3. 1 Test calculations to establish the modelling scheme

First, we examined the scheme to identify the binding mode of a ligand and ordered water molecules in an active site using only the editing (proofreading) domain, referred to as the connective polypeptide 1 (CP1) domain, in the calculations of leucyl-tRNA synthetase (LeuRS) in complex with an inhibitor, 2′-(L-norvalyl)

amino-2′-deoxyadenosine (Fig. 1a). In this calculation, instead of the inhibitor bound to the CP1 domain in the crystal structure, we used the 'true' substrate as the ligand in the simulations. In the present modelling, the ligands mentioned are assumed to be divided into two fragments, i.e. the amino acid (fragment 1) and adenosine (fragment 2) moieties (Fig. 1c; also see Supplementary material; S1).

We applied the scheme to the crystal structure of the *Thermus thermophilus* LeuRS in complex with the inhibitor mentioned, i.e., the initial coordinate of the protein and ligand are taken from the crystal structure of the complex. As a result of the simulations described in the Materials and Methods section, we found all five of the ordered water molecules that are experimentally observed in the catalytic site (Fig. 2a). Three of them are exchangeable water molecules corresponding to crystallographic water molecules HOH484, HOH483, and HOH224 in the crystal structure of the LeuRS•inhibitor complex (1OBC). The other two waters are 'deeply buried' ones, which are recognized by the N^7 of the base and the carbonyl carbon of the valine moiety, corresponding to the crystallographic water molecules HOH222 and HOH482 in the crystal structure (1OBC), respectively. These 'deeply buried' water molecules are unlikely to exchange with bulk water molecules; in fact, when we removed these crystal water molecules, a standard MD simulation conducted for the complex of the CP1 domain and the substrate failed to

predict them (see Supplementary material; S3). Thus, we showed that the present scheme is essential for accurate identification of the ordered water molecules buried in the LeuRS•inhibitor interface.

In stage (v) of our present scheme, we continued the standard MD simulation for 1 ns to evaluate the stability of the complex structure obtained, and found that the interfacial hydrogen bond networks, which are composed of the substrate, the CP1 domain and the buried water molecules, are fundamentally stable throughout the 1-ns MD simulation (see Supplementary material; S4). Thus, the present scheme can be used efficiently to identify the ordered water molecules in the ligand-binding site, in particular, the deeply buried water molecules in the interspace between the ligand and the protein.

3. 2 Structural modelling of LeuRS in complex with valyl-tRNA^{Leu}

Structural comparison of the two crystal structures, i.e., the *Thermus thermophilus* LeuRS in complex with the tRNA^{Leu} (2BYT) and that in complex with the inhibitor mentioned (1OBC), showed almost similar hydrogen bond networks between the adenosine bases and the enzymes. On the other hand, with respect to the ribose moieties of A76 of tRNA (2BYT) and the inhibitor (1OBC), the hydrogen bond networks with the surrounding structures are different: in the LeuRS•inhibitor complex, the hydrogen

bond is present between $O^{5'}$ of the inhibitor and the hydroxyl group of Tyr332, whereas in the LeuRS•$tRNA^{Leu}$ complex, the corresponding hydrogen bond is absent; instead, the hydroxyl group of Tyr332 forms the hydrogen bond with O2′ of C75 of $tRNA^{Leu}$. Correspondingly, one can find differences in the positions and conformations of the ribose moieties between the LeuRS•inhibitor and LeuRS•$tRNA^{Leu}$ complexes, as shown in Fig. 1c. This indicates that the bound states of the ribose moiety depend on the presence/absence of either $tRNA^{Leu}$ or the amino acid moiety attached to the O2′ atom. In this study, for the LeuRS•valyl-$tRNA^{Leu}$ complex, we elucidate such differences using computational modelling techniques; it may be crucial to investigate the detailed mechanisms of the editing.

Next, using the parameter set for MD simulations, which were optimized in the test calculations, we applied the present scheme for the identification of hydration water and constructed a fully solvated structure of the LeuRS•valyl-$tRNA^{Leu}$ complex, as follows. The initial coordinates of the protein and $tRNA^{Leu}$ (G1–C75) moieties were obtained from the crystal structure of the LeuRS•$tRNA^{Leu}$ complex (2BYT). For the initial coordinates of A76 of $tRNA^{Leu}$ and the amino acid moiety attached to $O^{2'}$ of A76, the coordinates of the inhibitor in the crystal structure of the LeuRS–inhibitor complex (1OBC) were exploited.

As a result, five ordered water molecules were identified in the editing site, and the hydrogen bond networks formed by these water molecules are identical to those in the test case (Fig. 3a). It should be noted that two buried water molecules are also observed in the crystal structure of the *Thermus thermophilus* IleRS complexed with a post-transfer editing inhibitor [18], for which the configurations are well consistent with the ones identified in the modelled structure of the LeuRS•valyl-tRNA^{Leu} complex. This suggests that a conserved hydrolysis mechanism is present in these different aaRSs. The high stability of the CP1 domain and tRNA^{Leu} in a subsequent 500-ps MD simulation indicates that the calculations at stages (i) to (iv) do not induce any distortions of the structure (see Supplementary material; S5). However, since in the present modelling, the norvaline moiety was changed to valine, the binding mode of the side chain of the amino acid moiety is changed with respect to those of the crystal structure of the LeuRS•inhibitor complex. Nevertheless, it fits well in the hydrophobic pocket, as shown in Fig. 3b. Experimentally, Thr252, which consists of the catalytic pocket, acts to block the binding of leucine to the editing active site by the bulkiness of its side chain, since leucyl-tRNA^{Leu} is the 'correct' product for the LeuRS system. In fact, T252A results in mis-editing of the leucyl-tRNA^{Leu} [8, 9]. Here, in the modelled structure of the LeuRS•valyl-tRNA^{Leu} complex, the γ-methyl group of the valine moiety of the ligand is

located close to the side chain of Thr252 (the distance between C^{δ} of the leucine moiety of the ligand and O^{γ} of Thr252 is 2.1 Å), which would cause steric clash when the leucine is attached as the ligand (Fig. 3c).

With respect to the adenosine moiety, its binding mode in the LeuRS•valyl-$tRNA^{Leu}$ complex is identical to those in the crystal structures of LeuRS•$tRNA^{Leu}$ and LeuRS•inhibitor complexes (see Supplementary material; S5). On the other hand, the position of the ribose moiety is different from those in the two crystal structures; i.e. for the LeuRS•valyl-$tRNA^{Leu}$ complex, the MD simulation at stage (v) of the present modelling revealed that the ribose of A76 is located at intermediate positions between the two crystal structures (Fig. 4a and b). Next, to investigate its conformations, we measured the pseudorotation angles in the MD simulation (Table 1). This analysis revealed that the averaged value of the pseudorotation angle is almost identical to the value in the crystal structure of the *Pyrococcus horikoshii* LeuRS•$tRNA^{Leu}$ complex (PDB ID: 1WD2), in which A76 is bound to the aminoacylation site, supporting the plausibility of the observed conformation of the ribose moiety. In fact, the pseudorotation angle corresponds to an optimal conformation of the ribose in the C3′-endo pucker [19]. Furthermore, since this averaged value is within the range of values for the crystal structures of the *Thermus thermophilus* LeuRS•$tRNA^{Leu}$ and

LeuRS•inhibitor complexes, the conformation, as well as the position, of the ribose moiety obtained in the LeuRS•valyl-tRNA^{Leu} complex is also intermediate between these two crystal structures (Table 1). This is presumed to be induced by the mechanical constraints from tRNA^{Leu} and the attached amino acid moiety in a sandwiched manner, either of which is absent in these two crystal structures (Fig. 4a and b). In fact, changes of the pseudorotation angle within the C3′-endo conformation are almost barrierless [19]. This conformational flexibility of the ribose moiety enables the valyl-tRNA^{Leu} to be accommodated in the binding pocket in a manner whereby the specific recognition by the protein observed in the two crystal structures is fundamentally conserved, whereas the positions of the C1′ and O2′ atoms of the ribose are affected by the tRNA and the amino acid moiety, respectively.

It should be noted here that despite the above-mentioned changes, the recognition mode of C75 and C74 in the LeuRS•valyl-tRNA^{Leu} complex is similar to that of the crystal structure of the LeuRS•tRNA^{Leu} complex (see Supplementary material; S5). Furthermore, with respect to the amino acid moiety attached to A76, i.e. valine, its hydrogen bond networks are consistent with those of norvaline in the crystal structure of the LeuRS•inhibitor complex (see Supplementary material; S5 and S7).

Thus, these results are consistent with previous experimental data supporting the

feasibility of the present modelled structure of the LeuRS•valyl-tRNA^{Leu} complex. This structure enables us to computationally investigate the mechanisms of the editing by LeuRS. Furthermore, it also allows us to address the detailed processes of translocation of the 3′-end of tRNA between the editing and aminoacylation sites of LeuRS, followed by the large-scale rotations of the editing domain [12, 20]. In this manner, the present structural model of the fully solvated LeuRS•valyl-tRNA^{Leu} complex provides a structural basis for the design of new experimental and computational analyses of the mechanisms of the editing.

Acknowledgements

This work was partly supported by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) under contracts 19019003, 19340108 and 20051003. Computations were performed using computer facilities under the auspices of the Interdisciplinary Computational Science Program at the Center for Computational Sciences, University of Tsukuba and the Computer Center for Agriculture, Forestry, and Fisheries Research, MAFF, Japan.

References

[1] Sankaranarayanan, R. and Moras, D. (2001) The fidelity of the translation of the genetic code. Acta. Biochim. Pol. 48, 323-335.

[2] Nureki, L., Vassylyev, G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T.L., Schimmel, P. and Yokoyama, S. (1998) Enzyme structure with two catalytic sites for double sieve selection of substrate. Science 280, 578-582.

[3] Lincecum, T.L. Jr., Tukalo, M., Yaremchuk, A., Mursinna, R.S., Williams, A.M., Sproat, B.S., Van Den Eynde, W., Link, A., Van Calenbergh, S., Grotli, M., Martinis, S.A. and Cusack, S. (2003) Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. Mol Cell. 11, 951-963.

[4] Fukunaga, R., Fukai, S., Ishitani, R., Nureki, O. and Yokoyama, S. (2004) Crystal structures of the CP1 domain from Thermus thermophilus isoleucyl-tRNA synthetase and its complex with L-valine. J. Biol. Chem. 279, 8396-8402.

[5] Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D.G. and Yokoyama, S. (2000) Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. Cell 103, 793-803.

[6] Silvian, L.F., Wang, J. and Steitz, T.A. (1999) Insights into editing from an

ile-tRNA synthetase structure with tRNAile and mupirocin. Science, 285, 1074-1077.

[7] Fukunaga, R. and Yokoyama, S. (2005) Structural basis for non-cognate amino acid discrimination by the valyl-tRNA synthetase editing domain. J. Biol. Chem. 280, 29937-29945.

[8] Mursinna, R.S., Lincecum, T.L. Jr. and Martinis, S.A. (2001) A conserved threonine within Escherichia coli leucyl-tRNA synthetase prevents hydrolytic editing of leucyl-$tRNA^{Leu}$. Biochemistry, 40, 5376-5381.

[9] Zhai, Y., Nawaz, M.H., Lee, K.W., Kirkbride, E., Briggs, J.M. and Martinis, S.A. (2007) Modulation of substrate specificity within the amino acid editing site of leucyl-tRNA synthetase. Biochemistry, 46, 3331-3337

[10] Nordin, B.E. and Schimmel, P. (2002) Plasticity of recognition of the 3'-end of mischarged tRNA by class I aminoacyl-tRNA synthetases. J. Biol. Chem. 277, 20510-20517.

[11] Zhai, T., Martinis, S. A. (2005) Two conserved threonines collaborate in the Escherichia coli leucyl-tRNA synthetase amino acid editing mechanism. Biochemistry 44, 15437–15443

[12] Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S. and Cusack, S. (2005)

The crystal structure of leucyl-tRNA synthetase complexed with $tRNA^{Leu}$ in the post-transfer-editing conformation. Nature Struct. Mol. Biol. 12, 923-930.

[13] S. Fukai, O., Nureki, S., Sekine, A., Shimada, D., Vassylyev, G. and Yokoyama, S. (2003) Mechanism of molecular interactions for tRNA recognition by valyl-$tRNA^{Val}$ synthetase. RNA 9, 100-111.

[14] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, P. A. Kollman (2006) AMBER 9, University of California, San Francisco.

[15] Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. J. Chem. Phys. 98, 10089-10092.

[16] Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. J. Comput. Phys. 23, 327-341.

[17] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81, 3684-3690.

[18] Fukunaga, R. and Yokoyama, S. (2006) Structural basis for substrate recognition by the editing domain of isoleucyl-tRNA synthetase**.** J. Mol. Biol. 359, 901-912.

[19] Levitt, M. and Warshel, A. (1978) Extreme conformational flexibility of the Furanose ring in DNA and RNA. J. Am. Chem. Soc. 100, 2607-2613.

[20] Fukunaga, R. and Yokoyama, S. (2005) Aminoacylation complex structures of leucyl-tRNA synthetase and $tRNA^{Leu}$ reveal two modes of discriminator-base recognition. Nature Struct. Mol. Biol. 12, 915-922.

Table 1. Comparison of pseudorotation angles of the ribose moieties of A76 in the crystal structures of the *Thermus thermophilus* LeuRS•$tRNA^{Leu}$ complex (PDB ID: 2BYT), the *Pyrococcus horikoshii* LeuRS•$tRNA^{Leu}$ complex (PDB ID: 1WZ2), the *Thermus thermophilus* LeuRS•inhibitor complex (PDB ID: 1OBC) and the LeuRS•valyl-$tRNA^{Leu}$ complex obtained in the MD simulation (values in the table are averaged values using snapshots from 400 to 500 ps).

PDB ID	2BYT	1WZ2	1OBC	MD simulation
Pseudorotation (degrees)	30.7	20.3	12.2	18.0

Figure legends

Figure 1. (a) Crystal structure of *Thermus thermophilus* LeuRS in complex with the inhibitor Nva2AA (PDB code 1OBC). The left and right panels show the overall structure and the editing active site of the LeuRS, respectively. (b) Crystal structures of *Thermus thermophilus* LeuRS in complex with $tRNA^{Leu}$ (PDB code 2BYT). (c) Chemical structures of ligands of LeuRS. (d) Structural comparison between the crystal

structures of the LeuRS•inhibitor complex (PDB code 1OBC) (magenta) and the LeuRS•$tRNA^{Leu}$ complex (PDB code 2BYT) (green).

Figure 2. Results of test calculations to establish an algorithm for modelling of the enzyme•ligand complex with hydration water on the molecular interface, using the CP1 domain of LeuRS and the substrate. (a) The van der Waals surface of the catalytic pocket, and the identified binding mode of the substrate (red) and solvent water molecules (yellow). (b) Conformations of the ligand (red), protein (blue) and solvent water molecules obtained using the present calculation scheme (light blue). The crystal structures of the protein, ligand and crystallographic water molecules are represented in yellow, grey and green, respectively.

Figure 3. (a) Structural comparison of the catalytic site of the modelled structure of the LeuRS•valyl-$tRNA^{Leu}$ complex (black) with those of the two crystal structures shown in Fig. 1b. (b) Recognition mode of the valine moiety in the editing pocket of the LeuRS•valyl-$tRNA^{Leu}$ complex. (c) Mechanism of exclusion of the leucine moiety of the ligand (blue) so as not to bind to the editing pocket. This is regulated by the side chain of Thr252 (red).

Figure 4. (a) Close-up view of the conformations of the ribose moiety of A76. Colour representations are the same as in Fig. 3a. (b) Schematic drawing of the role of the conformational flexibility of the ribose moieties in the LeuRS•valyl-tRNA^{Leu} complex, which enables the A76-valine moiety to be accommodated in the catalytic pocket in a manner whereby the recognition modes observed in the crystal structures of the LeuRS•tRNA^{Leu} and LeuRS•inhibitor complexes are fundamentally conserved. Red and blue arrows represent mechanical constraints induced by tRNA^{Leu} and the misattached valine moiety, respectively.

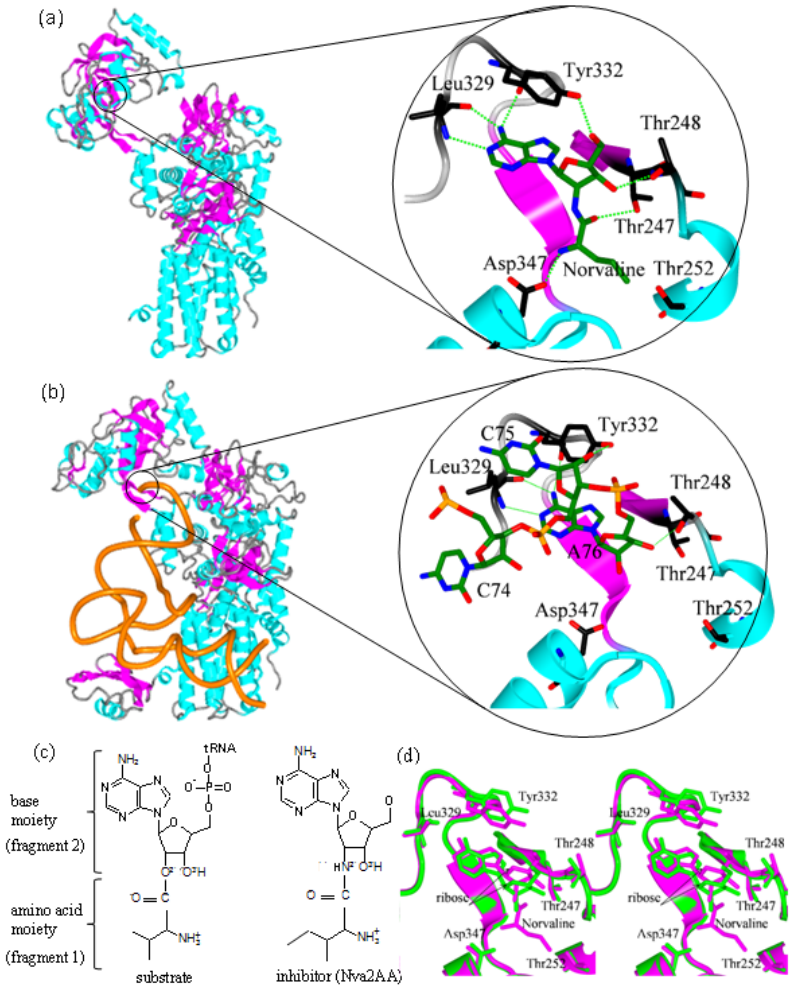

Figure 1

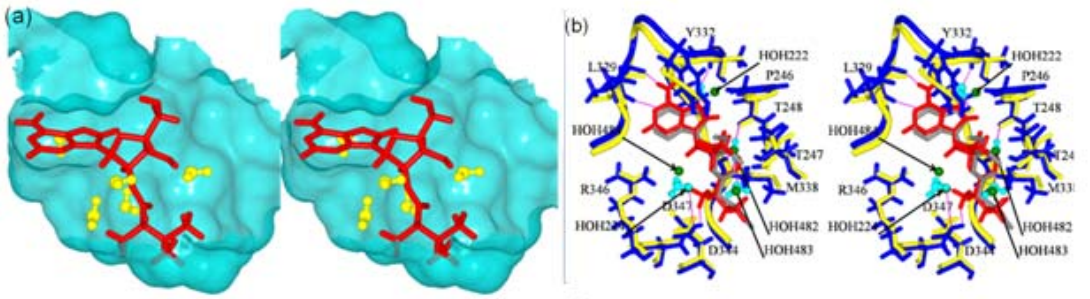

Figure 2

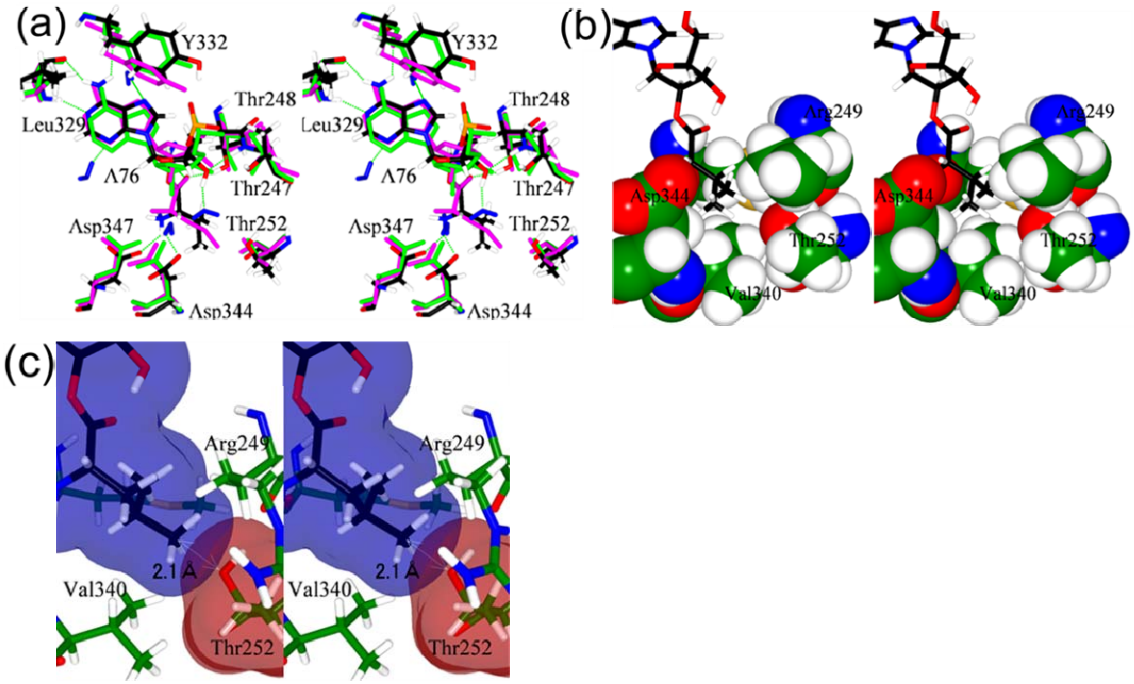

Figure 3

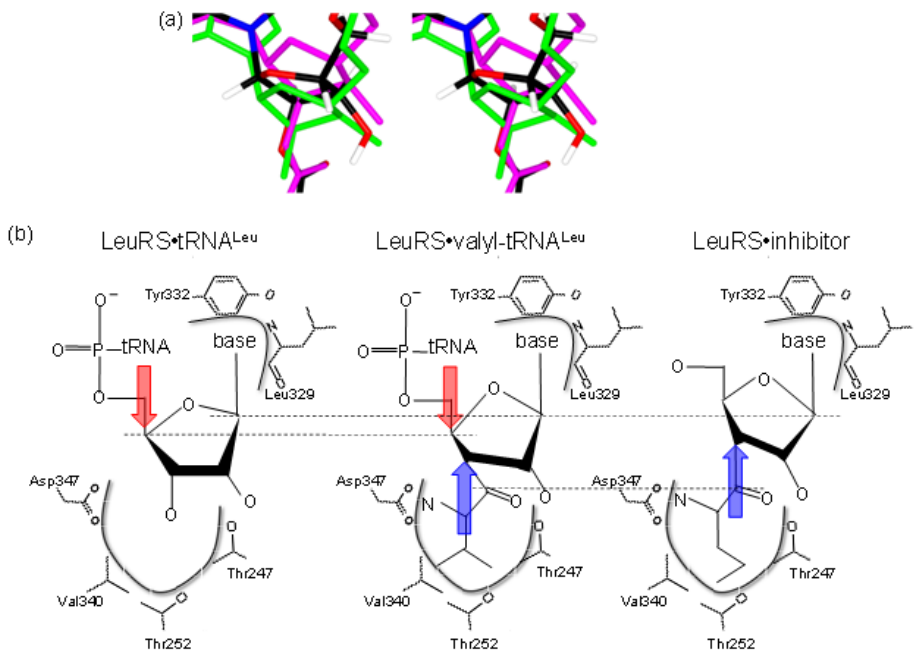

Figure 4

Supplementary materials

Structural modelling of the complex of leucyl-tRNA synthetase and mis-aminoacylated tRNA^{Leu}

Yohsuke Hagiwara[1], Osamu Nureki[2], and Masaru Tateno[1] *

[1] Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba Science City, Ibaraki 305-8571, Japan

[1] Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba Science City, Ibaraki 305-8577, Japan

[2] Institute of Medical Science, University of Tokyo, Shirokanedai 4-6-1, Minatoku, Tokyo 108-8639, Japan.

Table of contents

S1. Computational details of identification of hydration water

For test calculations to establish the modelling scheme, first, the connective polypeptide 1 (CP1) domain (amino acid residues 228–415) was truncated, and then, to completely immerse the binding pocket of the CP1 domain with solvent water molecules, MD simulation was performed for 10 ps at 300 K, where a harmonic constraint was applied to all atoms of the protein and Na^+ atoms with a force constant of 500 kcal·mol^{-1}·Å^{-2}. Furthermore, the MD simulation was performed for 5 ps at 500 K to fully solvate the ligand-binding pocket, where only main chain atoms were constrained. This procedure was applied to MD simulations conducted in stage (i). In stage (ii), the substrate was placed in the binding pocket by superimposing the crystal structure onto the simulated structure. In stage (iii) (for fragment 1), λ was increased first for the amino acid moieties of the substrate at a rate of 1.0×10^{-6}/step until λ reached 4.0×10^{-4}. Then, the rate was kept at 2.0×10^{-4}/step until λ reached 0.9. In this stage, positional constrains with a force constant of 100 kcal·mol^{-1}·Å^{-2} were applied to all atoms of the substrate. Next, in stage (iv) for fragment 1, the force constant was set to 0 for the atoms of the amino acid moieties until λ reached 1. Next, stage (iii) for fragment 2 was performed using a similar procedure; when λ reached 0.7, stage (iv) for fragment 2 was started, where all atoms of the system were free to move (Figure S1).

The total simulated time required for stages (iii) and (iv) for both fragments was 10.8 ps. The final snapshot was subjected to a standard MD simulation at 300 K for 1-ns. The same procedure was employed in the structural modelling of LeuRS in complex with valyl-tRNA^{Leu}, although the simulation time of the standard MD simulation in the stage (v) was 500 ps due to their huge size of the system, i.e. total number of atoms involved in the system is 165,739.

The parameterization of the valyl-tRNA^{Leu} was done as follows. All atom types of the valyl-tRNA^{Leu} were taken from the AMBER ff99 force field. For the valine moiety attached to A76, atom types of the N-terminus valine were assigned; for A76, those of the ribo-adenosine with 5'-phosphate group and 3'-OH were used. More specifically, for $O^{2'}$ of A76, the atom type of ester oxygen, 'OS', was used, since $O^{2'}$ is bonded to the carboxyl oxygen of the valine moiety. With respect to the partial charges of tRNA^{Leu} including A76, the values defined in the ff99 force field were applied; for the valine moiety attached to A76, the values of the N-terminus valine defined in the force field were employed, leading this moiety to establish the consistency with the others. These assignments of the partial charges result in a non-zero (but small negative) value of the total charge of the valyl-A76 moiety, i.e., −0.11. To neutralize the value, we added the positive value of 0.002 to the partial charge of each atom in the valyl-A76 moiety. Atom

types and partial charges of the valyl-A76 moiety are summarized as shown in Fig. S1.

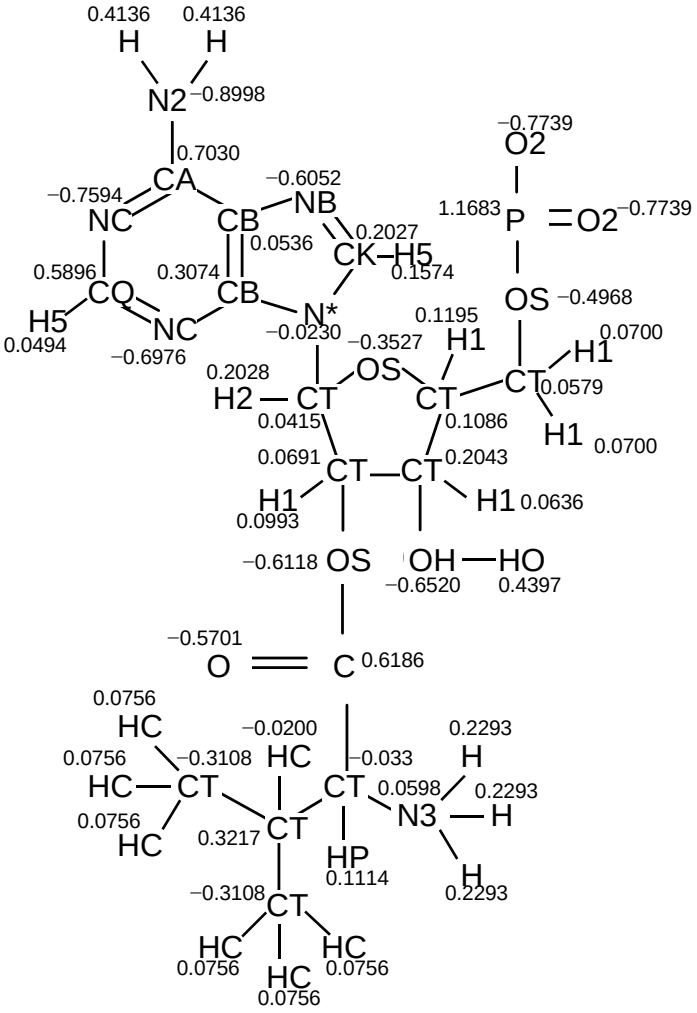

Figure S1. Assignments of atom types and partial charges for the valyl-A76 moiety.

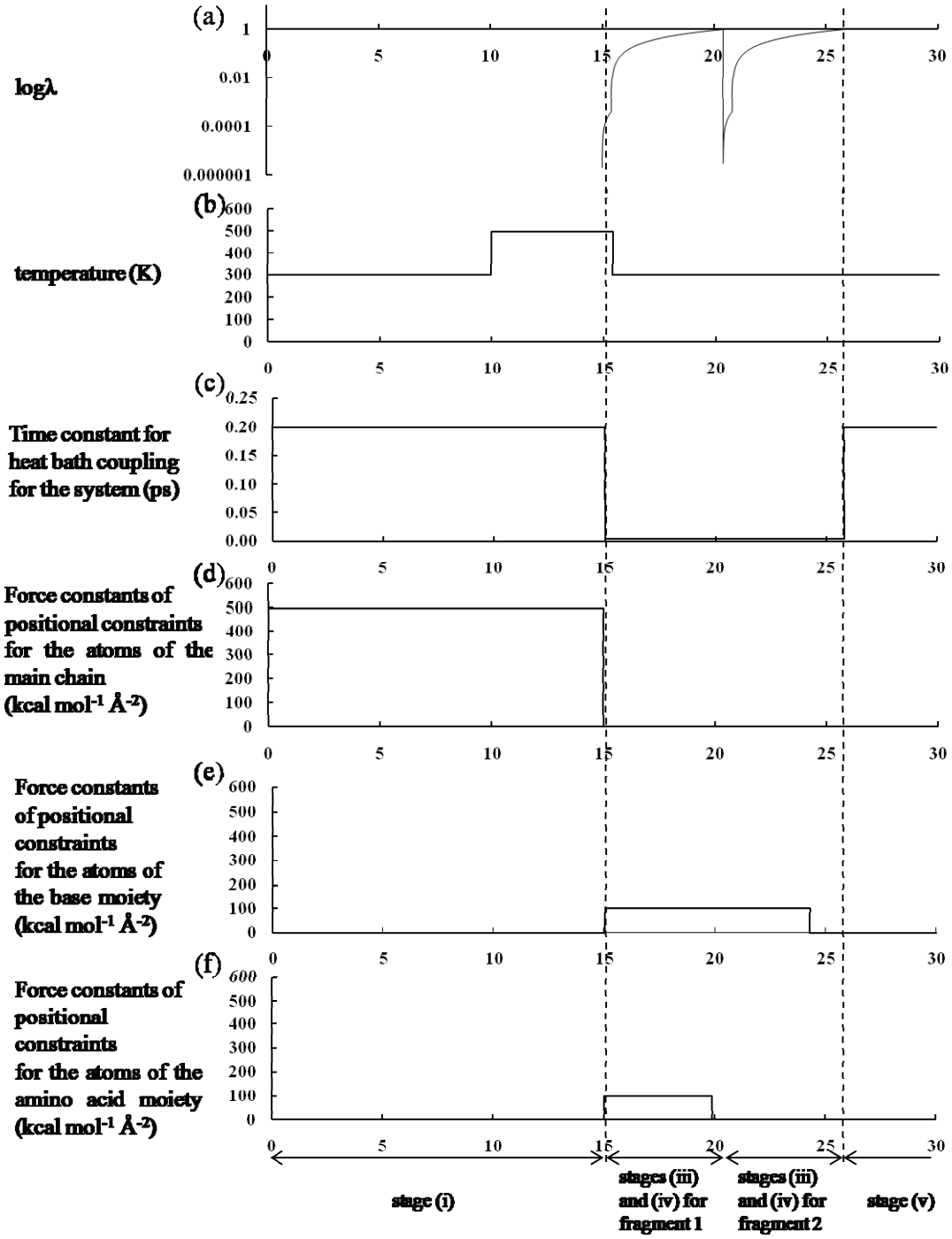

Figure S2. Regulation of the parameters used in the present calculations. (a) log □, (b) temperature, (c) time constants for heat bath coupling in ps, (d) force constants of positional constraints for the backbone atoms of the protein moiety in kcal·mol^{-1}·Å^{-2},

(e) force constants of positional constraints for the atoms of the base moiety of the substrate in kcal·mol^{-1}·Å^{-2}, (f) force constants of positional constraints for the atoms of the amino acid moiety of the substrate in kcal·mol^{-1}·Å^{-2}.

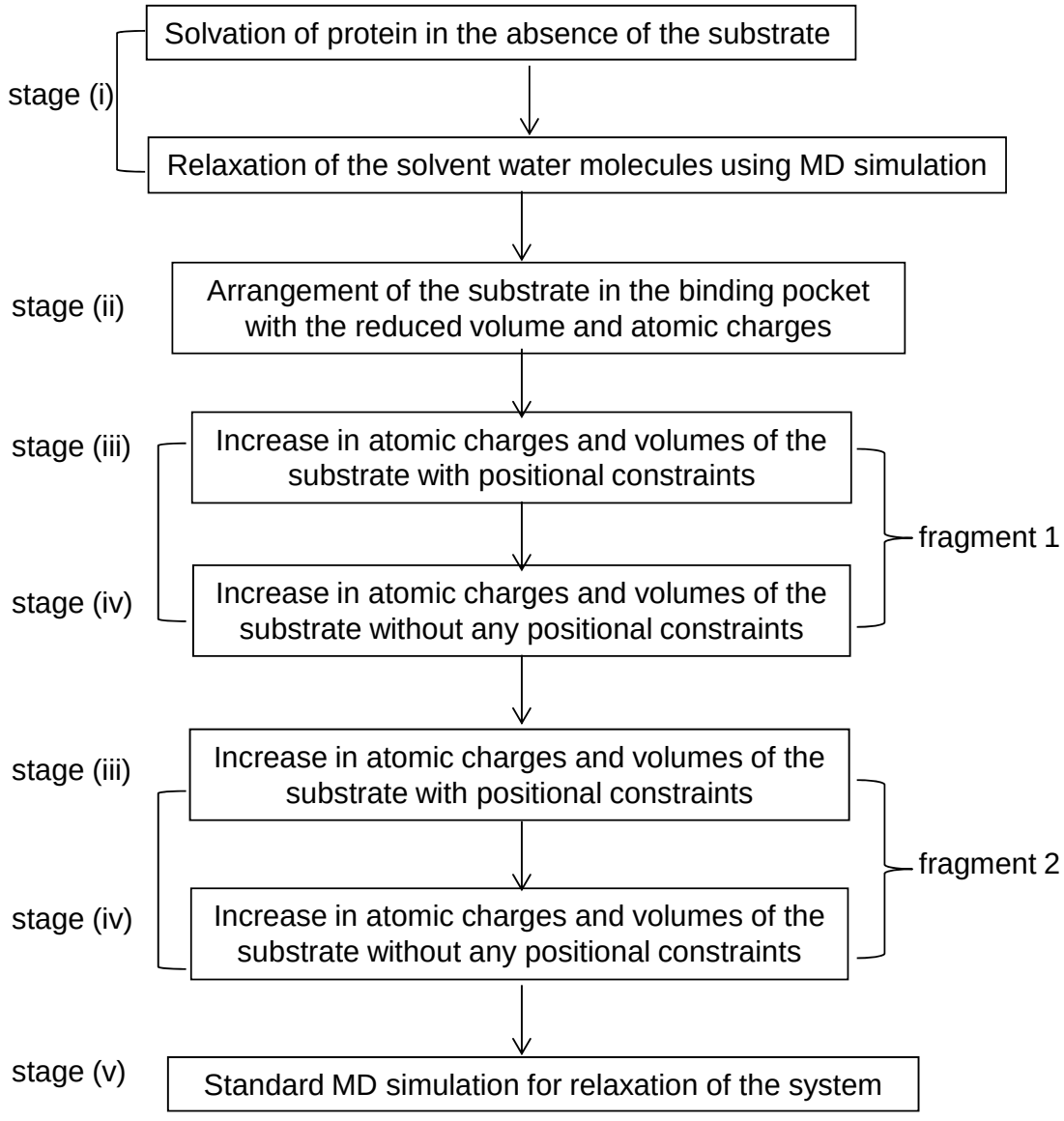

Figure S3. Flowchart of the scheme for the present modelling.

S2. The calculation for the complex of the CP1 domain of leucyl-tRNA synthetase and its substrate

In stage (i), the binding pocket of the CP1 domain of LeuRS was filled with 23 water molecules using a standard MD simulation (Figure S3a); six ordered water molecules are found in the active site, all of which correspond to structural water molecules observed in the crystal structure of LeuRS in the free state (Figure S3a). Note that in stage (ii), the coordinates of the "ligand" and water molecules are actually overlaid (Figure S3b). In stages (iii) and (iv), the MD simulations using the modified energy function are performed for fragment 1, with the positional constraints for the substrate and then without any constraints, respectively. Subsequently, stages (iii) and (iv) are applied to fragment 2. Thus, at the end of stage (iv), we have identified three deeply "buried" water molecules in the interspace between the CP1 domain and the substrate (Table S1). The first forms a hydrogen bond with the N7 atom of the base moiety; this corresponds to the crystallographic water HOH222. The second forms a hydrogen bond with the carbonyl carbon of the amino acid moiety, corresponding to the crystallographic water HOH482, and the third with the N6 atom of the base moiety, which is referred to as WAT1 (Figure S4a). The third water molecule does not correspond to any crystallographic ones and thus further relaxation phases, such as the

subsequent stage (v) in our scheme, are likely to be required. In fact, a slight distortion of the adenosine base of the substrate is also found in this stage; i.e., the N6 atom is not on the plane of the adenosine base.

Nevertheless, we identified another structural water molecule coordinated to the N3 atom of the substrate, which is not “buried” in the molecular interspaces but can be exchanged with other unordered solvent water molecules, corresponding to the crystallographic water HOH484. Moreover, the exchangeable water molecule corresponding to the crystallographic one (HOH224) was identified, which was suggested as the nucleophile in the enzymatic reaction. In this manner, four water molecules among five crystallographic ones observed in the crystal structure of the complex are found in this stage. In stage (v), a standard MD simulation is performed to further relax all atoms in the system without any constraints. In the early phase of this simulation (at ~10 ps), a water corresponding to the crystallographic HOH483, which was not observed in previous stages, was further identified. It should be noted that the WAT1 mentioned above is moved toward the outside of the ligand-binding pocket (Fig. S4b).

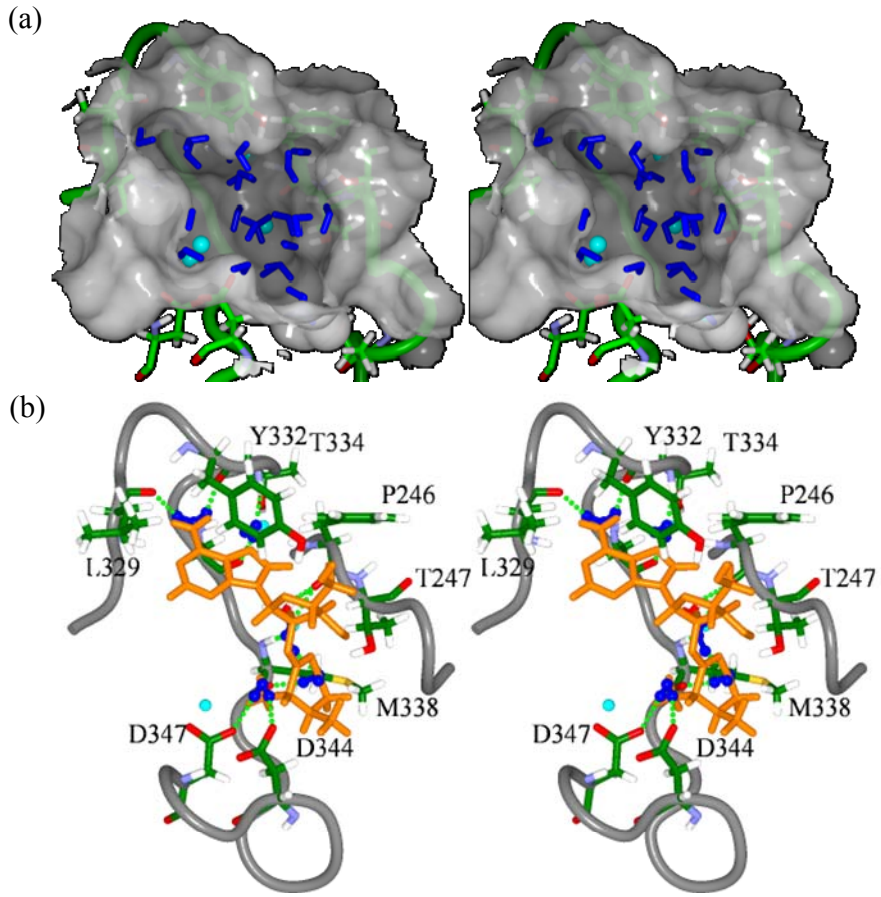

Figure S4.

(a) Snapshot of the binding pocket at 15 ps of the MD simulation. Water molecules depicted as light blue spheres are crystallographic water molecules observed in the crystal structure of the LeuRS in the free state. The water molecules observed in the MD simulation are represented as blue sticks. (b) Initial structure of stage (iii) with respect to the amino acid moiety of the substrate. The manner of the representation of the water molecules is the same as in (a).

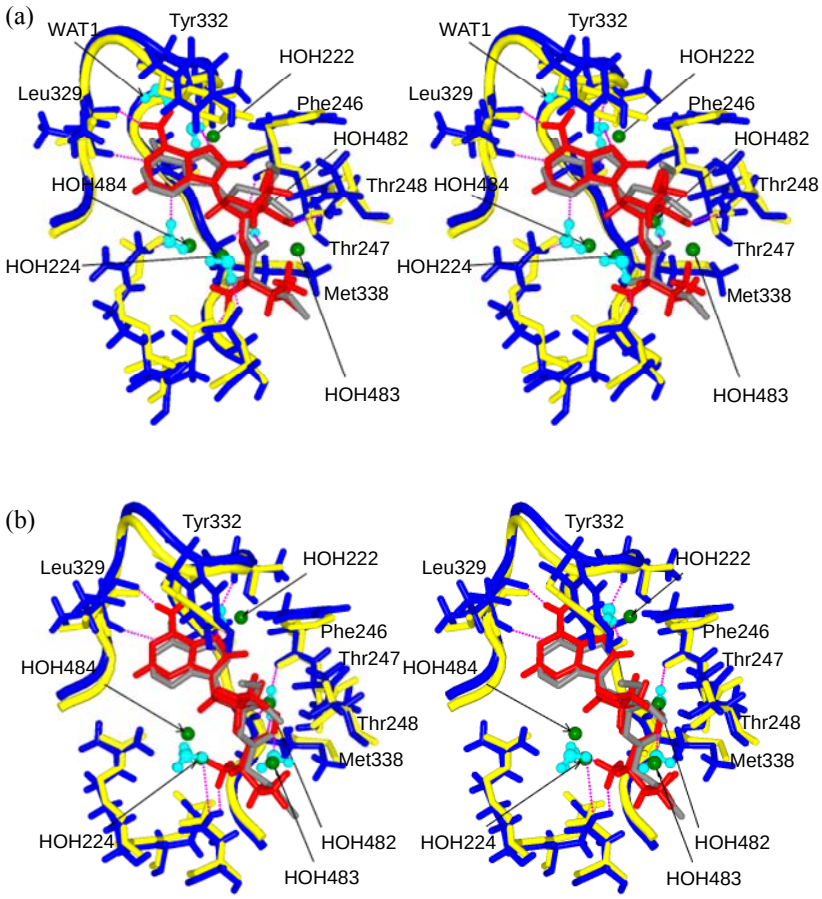

Figure. S5

(a) Obtained conformations of the ligand (red), protein (blue), and solvent water molecules (light blue). Those coordinates are taken from a snapshot at the end of stage (iv). The crystal structures of the protein, ligand, and crystallographic water molecules are represented in yellow, gray, and green, respectively. (b) Their coordinates are taken from a snapshot of the standard MD simulation conducted in stage (v) at 50 ps.

Table S1.

Distances between oxygen atoms of crystallographic water molecules and corresponding water molecules identified in the MD simulations conducted in stage (iv) and (v) (MD snapshots are used to calculate the atomic distances).

crystallographic water	HOH222	HOH224	HOH482	HOH483	HOH484
distance in stage (iv)[1]	1.37	0.95	0.37	-	1.77
distance in stage (v)[1]	1.41	0.86	0.33	0.50	1.16

1) Distances are defined as those between oxygen atoms of water molecules observed in the crystal structure and in the calculated structure.

S3. Comparison with a standard MD simulation of the complex of the CP1 domain and the substrate

The "deeply buried" water molecules identified are unlikely to be exchanged with unordered water molecules in the solvent. In fact, this is the case with the MD simulation conducted in stage (v). To further confirm this issue, we performed another standard MD simulation for 1 ns, initiated from the crystal structure of the complex, in which all the crystallographic water molecules bound to the active site in the CP1 domain were removed. As a result, the water molecules located near the surface of the ligand, such as HOH223, HOH224, and HOH483, were identified in the MD simulation. However, those buried in the molecular interspace, i.e., HOH222 and HOH482, were

not found in the active site. Thus, the standard MD simulations failed to predict them, because unordered water molecules in bulk cannot be placed in the interspace because of the excluded volume of the substrate and the protein. This shows that our procedure accurately predicts the ordered water molecules and the protein•substrate interfacial structures.

S4.Stability of intermolecular hydrogen bonds identified in the calculation of the CP1 domain.

Most of all hydrogen bonds identified in the stage (iii) and (iv) are stable in the MD simulation conducted in stage (v) execpt that the Tyr332 side chain fluctuated (Fig. S5). This is consistent with experimental data (e.g., B-factor) observed in the crystal structure of the complex (Table S2). Thereby, the hydrogen bond between $O^{5\prime}$ of the substrate and HO^{ζ} of the Tyr332 side chain is not stable; the variance of the distance between the two atoms is 0.62, whereas those of the others related to the hydrogen bonds are less than 0.1.

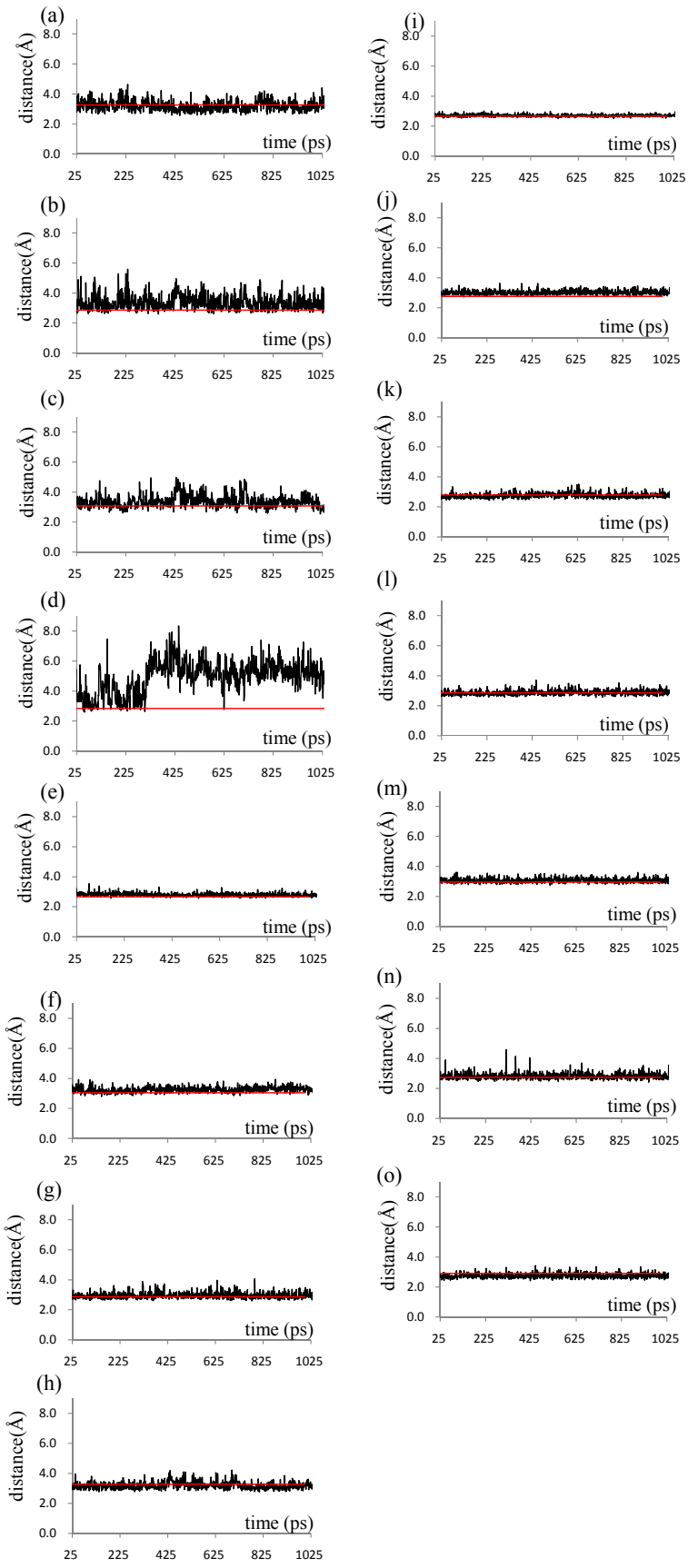
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)
distance(Å)
time (ps)
0.0
2.0
4.0
6.0
8.0
25
225
425
625
825
1025

Figure S6.

(a)-(i) Trajectories of distances of atoms forming the hydrogen bonds between the substrate and the protein in the stage (v);

(a) Distances between O3' of the substrate and the hydroxyl oxygen of Thr247.

(b) Distances between the carboxyl oxygen of the substrate and the hydroxyl oxygen of Thr247.

(c) Distances between O3' of the substrate and the hydroxyl oxygen of Thr248.

(d) Distances between O5' of the substrate and the hydroxyl oxygen of Tyr332.

(e) Distances between N6 of the substrate and the carboxyl oxygen of Leu329.

(f) Distances between N6 of the substrate and the carboxyl oxygen of Tyr332.

(g) Distances between N1 of the substrate and the amide nitrogen of Leu329.

(h) Distances between O3' of the substrate and the amide nitrogen of Thr247.

(i) Distances between amino nitrogen of the substrate and the carboxyl oxygen of Asp347.

(j)-(o) Trajectories of distances of atoms forming the hydrogen bonds between the identified ordered waters and the protein in the stage (v);

(j) Distances between the oxygen of a water and the amide nitrogen of Thr334.

(k) Distances between the oxygen of a water and the amide nitrogen of Ala338.

(l) Distances between the oxygen of a water and the N7 of the substrate.

(m) Distances between the oxygen of a water and the carboxyl oxygen of the substrate.

(n) Distances between the oxygen of a water and the amide nitrogen of M338.

(o) Distances between the oxygen of a water and the carboxyl oxygen of P246.

Table S2.

The comparison of the B factor of the amino acids forming the hydrogen bond networks with the ligand in the crystal structure. The average value of B factor of the backbone atoms, the side chain atoms and all atoms of each amino acid residue are shown.

	averaged values of backbone atoms	averaged values of side chain atoms	averaged values of all atoms of amino acids
T247	22.1	22.1	22.1
T248	24.6	26.3	25.3
L329	31.1	33.5	32.3
Y332	30.2	38.9	36.0
M338	22.2	20.9	24.6
D347	21.4	21.5	21.4

S5. Stability of modeled structure of LeuRS•valyl-tRNA complex

The MD simulation of the LeuRS•valyl-tRNA complex in stage (v) shows that the CP1 domain which possesses the editing active site is stable and well equilibrated for 500 ps

(Fig. S6). Furthermore, the intermolecular hydrogen bonds between the substrate, i.e. valine-attached A76 and the protein are also stable during the MD simulation, all of which are consistent with the hydrogen bonds observed in the crystal structures of *Thermus thermophilus* LeuRS•inhibitor (1OBC) and *Thermus thermophilus* LeuRS•tRNA^{Leu} complexes (2BYT).

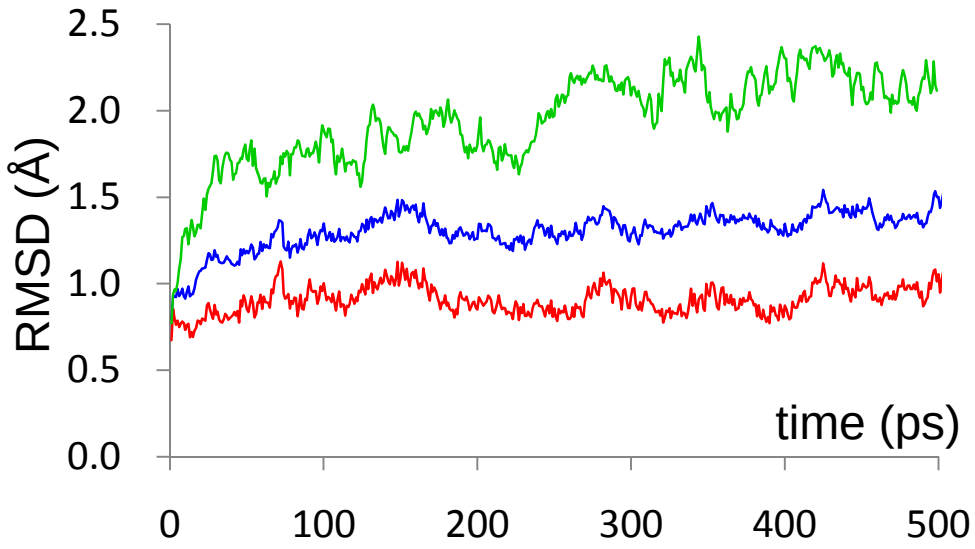

Figure S7. Time evolution of RMSDs of the tRNA^{Leu} backbone and the CP1 domain of the LeuRS•valyl-tRNA^{Leu} complex with respect to the crystal structure of the LeuRS•tRNA^{Leu} complex (PDB ID: 2BYT) obtained in the MD simulation of stage (v) of the present modelling scheme. The green line shows the RMSD of the tRNA^{Leu} backbone, and the blue and red ones show RMSDs with respect to the backbone atoms and all heavy atoms of the entire structure of the complex, respectively

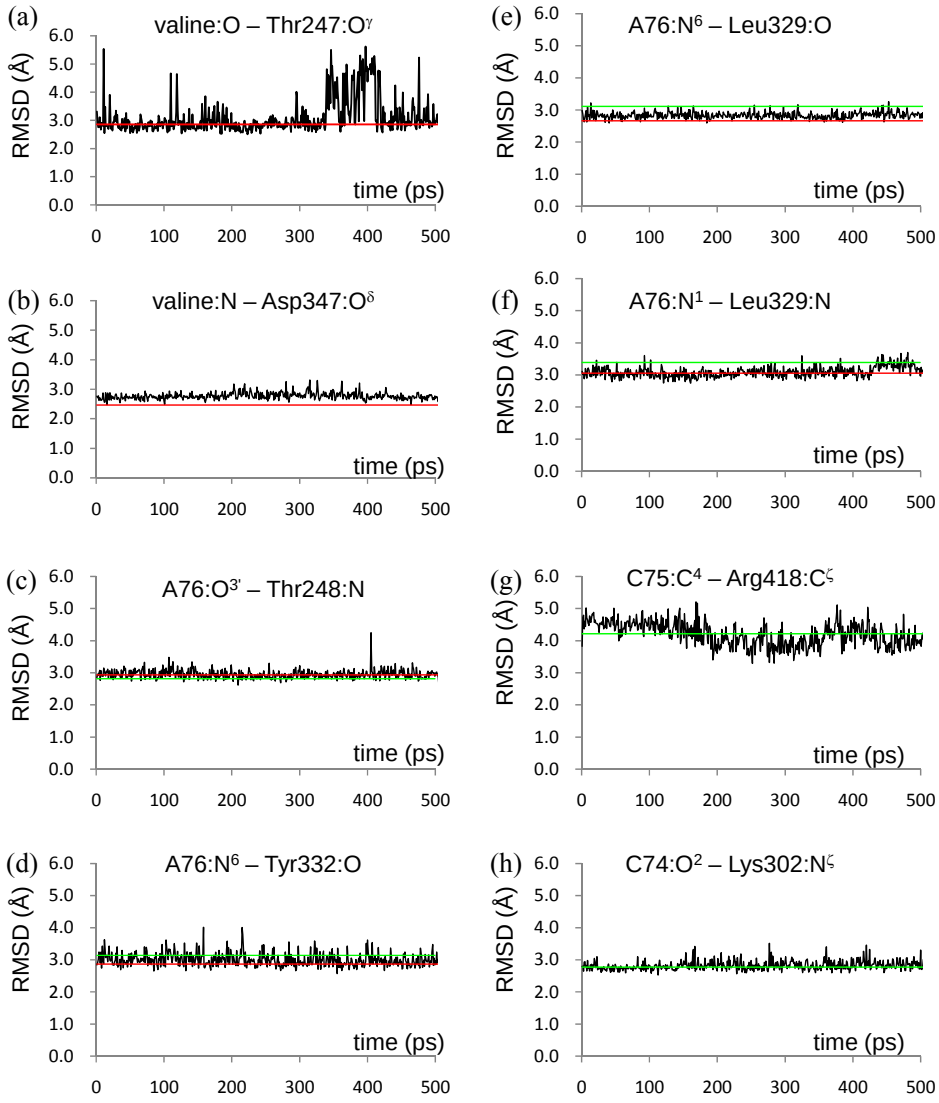

Figure S8. Trajectory of the distance between the atoms which form the hydrogen bonds in the LeuRS•valyl-tRNA^{Leu} (black) The values in the crystal structures of the LeuRS•inhibitor and LeuRS•tRNA^{Leu} are colored in green and red, respectively.

(a) Distances between the carboxyl oxygen of the valine moiety and the hydroxyl oxygen of Thr247.

(b) Distances between amino nitrogen of the valine moiety and the carboxyl

oxygen of Asp347.

(c) Distances between $O^{3'}$ of the A76 and the amide nitrogen of Thr247.

(d) Distances between N^6 of the A76 and the carboxyl oxygen of Tyr332.

(e) Distances between N^6 of the A76 and the carboxyl oxygen of Leu329.

(f) Distances between N^1 of the A76 and the amide nitrogen of Leu329.

(g) Distances between C^4 of the C75 and the C^{ζ} of Arg418.

(f) Distances between O^2 of the C74 and the N^{ζ} of Arg302.