Structural basis for recognition of G-1-containing tRNA by histidyl-tRNA synthetase

Qingnan Tian1,2, Caiyan Wang1,2, Yuhuan Liu1 and Wei Xie1,2,*

1Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory for Biocontrol, School of Life Sciences, The Sun Yat-Sen University, Guangzhou 510275, People’s Republic of China and 2Center for Cellular & Structural biology, The Sun Yat-Sen University, 132 E. Circle Rd., University City, Guangzhou 510006, People’s Republic of China

Received December 23, 2014; Revised January 26, 2015; Accepted February 06, 2015

ABSTRACT

Aminoacyl-tRNA synthetases (aaRSs) play a crucial role in protein translation by linking tRNAs with cognate amino acids. Among all the tRNAs, only tRNAHis bears a guanine base at position -1 (G-1), and it serves as a major recognition element for histidyl-tRNA synthetase (HisRS). Despite strong interests in the histidylation mechanism, the tRNA recognition and aminoacylation details are not fully understood. We herein present the 2.55Å crystal structure of HisRS complexed with tRNAHis, which reveals that G-1 recognition is principally nonspecific interactions on this base and is made possible by an enlarged binding pocket consisting of conserved glycines. The anticodon triplet makes additional specific contacts with the enzyme but the rest of the loop is flexible. Based on the crystallographic and biochemical studies, we inferred that the uniqueness of histidylation system originates from the enlarged binding pocket (for the extra base G-1) on HisRS absent in other aaRSs, and this structural complementarity between the 5′ extremity of tRNA and enzyme is probably a result of coevolution of both.

INTRODUCTION

Aminoacyl-tRNA synthetases (aaRSs) play a central role in protein synthesis by establishing the linkage between cognate amino acid and the tRNA anticodon triplet (1). This reaction proceeds as a two-step process where in the first step, the specific amino acid is condensed with ATP to produce an aminoacyl-adenylate intermediate (aa-AMP); and in the second, the amino acid is covalently linked to the cognate tRNA acceptor stem from aa-AMP (2). aaRSs can be divided into two major classes based on their sequence homology and oligomeric structures (3,4). Class I aaRSs contain a Rossmann fold and two highly conserved sequence motifs, and aminoacylate the 2′ OH group of the 3′ terminal ribose ring of tRNAs. Class II aaRSs are characterized by three conserved signature motifs at the active site and the primary site of aminoacylation is the 3′ OH group of A76 (5). Both classes can be further divided into finer subclasses (1,5). Subclass Iia enzymes form homodimers and with the exception of seryl-tRNA synthetase (SerRS), possess the homologous C-terminal extension that acts as the anticodon binding domain (1). Subclass Iib enzymes harbor a homologous N-terminal anticodon binding domain (6–8). Subclass Iic enzymes are tautomers and their amino acids and tRNA binding elements are less conserved than other class II aaRSs (9).

Histidyl-tRNA synthetase (HisRS) belongs to class IIa category (1,4,10). It forms the typical active site of Iia enzymes and the common anticodon binding domain. On the other hand, the histidylation system displays distinct features that are only unique to this synthetase. Histidine-specific tRNAs are well known for the extra guanosine at the 5′ terminus of the acceptor stem, designated as G-1 (11), either genetically encoded or added post-transcriptionally (12–15). Successful histidylation depends on the recognition of the essential G-1:N73 base pair by HisRS. The base across from G-1 is the discriminator base, which is well conserved as A in eukaryotes and C in prokaryotes (11,16). Mutations, deletions or substitutions of this pair causes a significant drop on $V_{\text{max}}/K_m$ and this observation is in agreement with in vivo studies (19). Additionally, the 5′ phosphate group of G-1 has also been shown to influence the enzymatic activity significantly (17,20,21).

Substitution of the 5′ monophosphate of G-1 by a triphosphate group reduces histidylation efficiency remarkably while deletion of the phosphate group causes a 510-fold decrease (20). Although both the accepter stem and the anticodon of tRNAHis participate in and contribute to the recognition by HisRS, the anticodon is a much weaker identity determinant or element, a specific motif in tRNAs that triggers aminoacylation, especially in Escherichia coli (18,22). For this reason, HisRS is capable of aminoacy-
lating a wide variety of RNA substrates (23–25). In addition to the full-length tRNAHis, fragments of tRNAHis (minihelical or microhelical RNAs) also make good substrates for both the Escherichia coli and yeast enzymes. The 1-73 pair in these shorter RNAs still serves as the major determinant for recognition while the sequence context of the 5'-terminal G-1 residue serves as the major identity element on the tRNAHis substrate at a molecular level. The addition of G-1 does not improve aminoacylation efficiency (28).

Crystal structures of several prokaryotic as well as eukaryotic HisRSs have been reported previously. The first structure of HisRS solved is from Escherichia coli (EcHisRS), which is in complex with histidyl-adenylate (HAMP, PDB code 1HTT) (29). Each monomer of EcHisRS contains three conserved domains: an N-terminal catalytic domain (residues 1-130), a C-terminal domain (residues 210-280), and an insertion domain (28). Soon, crystal structures of Thermus thermophilus HisRS (TtHisRS) in complex with histidine (PDB code 1ADJ) or histidyl-adenylate (HAMP, PDB code 1ADJ) and in the apo form (PDB code 1H4V) also became available (30,31). Different from the EcHisRS structure, his6-tag at the N-terminus, C-terminal domain, and an insertion domain. The TtHisRS becomes fully ordered and forms four a-helices. It is proposed to clamp onto the acceptor stem once tRNA binds. Besides bacteria, crystal structures of eukaryotic HisRSs have also been determined from trypanosomes and humans (PDB codes 3HR3, 3LC0, 3HRK, 4G84 and 4G85) (32,33). Eukaryotic HisRSs feature an extra N-terminal whep-TRS domain and nuclear magnetic resonance studies have shown that this domain forms an antiparallel two-helix bundle (PDB code 1X59) (33). Structural comparison with prokaryotic HisRSs indicates that eukaryotic HisRSs adopted a different fold for the insertion domain as well as a different HAM-binding mode. Along with biochemical and in vivo studies, past crystallographic research reveals the activation mechanism of histidine by the enzyme, but the aminoacylation details are still missing due to the lack of the key structure of the HisRS-tRNA complex. Here we report the crystal structure of the complex between TtHisRS and tRNAHis in the presence of histidine and a nonhydrolyzable inhibitor adenylyl-imidodiphosphate (AMPPNP). The cocystal structure provides the first model of HisRS-tRNA complexes and addresses the recognition problem of the identity elements on the tRNAHis substrate at a molecular level.

**MATERIALS AND METHODS**

**Cloning, expression and purification of TtHisRS**

The gene encoding TtHisRS (hisS, Accession NO. AAS80708) was amplified by PCR from the genomic DNA of Thermus thermophilus HB27 strain (ATCC Number BAA-163D-5). After double digestion by the NdeI and NotI restriction enzymes, the PCR product was ligated into a modified pET-28a (+) vector (Novagen) in which the thrombin site in the pET-28a (+) vector (recognition peptide LVPRGS) was replaced by a PreScission protease cleavage site (recognition peptide LEVLFQGP). The resulting construct contains a His6-tag at the N-terminus, and was transformed into Escherichia coli BL21(DE3) for overexpression. The cells were grown in LB medium to A600 = 0.8 at 37°C, and subsequently induced for overnight with 0.1 mM IPTG. The cells were harvested by centrifugation at 4000 rpm for 20 min and resuspended in the lysis buffer containing 40 mM Tris-HCl (pH 8.0), 250 mM NaCl, 10 mM imidazole, 1 mM PMSF and 1 mM β-mercaptoethanol. After sonication, the cell lysate was centrifuged by centrifugation at 4000 rpm for 1 h and loaded onto an empty column. The bound protein was washed with ~10 column volumes of lysis buffer and was eluted in the elution buffer containing 40 mM Tris-HCl (pH 8.0), 250 mM NaCl, 250 mM imidazole and 1 mM PMSF. The protein of interest was then treated with a PreScission protease overnight in the presence of 1 mM DTT to cleave off the his6-tag and subsequently applied onto a HiTrap column (GE Healthcare) to remove uncut TtHisRS. The unbound portion was pooled, and further purified by gel-filtration chromatography with a Superdex 200 column (GE healthcare) in a buffer containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl and 1 mM DTT. The purity of the protein was assessed by SDS-PAGE and the pure protein was flash-frozen in liquid nitrogen before being stored at −80°C.

All point mutations for TtHisRS activity assays were created through QuickChange (Stratagene, CA, USA). The primers used for cloning and mutagenesis were listed in Supplementary Table S1. The TtHisRS mutants used for aminoacylation assays were purified by a single step of Ni-NTA affinity chromatography. The human SerRS (hSerRS) and human GlyRS-SF (hGlyRS) proteins (an active form without the whep-TRS domain) were purified through a three-step purification process consisting of affinity (Ni-NTA, Qiagen), anion-exchange (Histrap, GE Healthcare) and gel-filtration chromatography (Superdex 200, GE Healthcare). The protein was finally eluted in a buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl and 1 mM DTT. Most proteins achieved 90% purity as assessed by SDS-PAGE and their structural integrity was checked with the Superdex 200 column. Most mutants are well folded except for the deletion mutant with the Glu325-Gly328 internal truncation. 5% Glycerol was added to mutants intended for activity assays before storage.

**In vitro Transcription of the tRNA Substrates**

The preparation of the tRNA transcripts was described in a previous protocol (34). Thermus thermophilus tDNAHis(GUG) (Supplementary Table S2) plus the T7 promoter was ligated into the pUC18 vector using complementary primers with the restriction sites HindIII and XbaI. The transcription template was obtained by PCR amplification of the ligated DNA fragments followed by phenol extraction. The in vitro transcription reaction of tRNAHis (with the 5' terminal GTP) for crystallization was performed at 37°C for 3 h with 0.3 μM T7 RNA polymerase in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM MgCl₂, 5 mM DTT, 1 mM spermidine and 2 mM of each NTP. tRNAHis transcript for aminoacylation assays,
which starts with the 5′-GMP, was produced in the reaction mix with identical recipe supplemented with 20 mM GMP. The transcript was purified by a 12% denaturing urea-polyacrylamide gel (UREA-PAGE). After ethanol precipitation, the RNA pellet was redissolved in TE buffer (20 mM Tris-HCl (pH 8.0) and 1 mM EDTA). tRNA was annealed by heating to 65°C and allowed to cool to room temperature after addition of 10 mM MgCl2. The annealed RNA was aliquoted and stored at −80°C for further use. Wild type (WT) tRNA Gly/ Ser (Supplementary Table S2) and their G-1-added mutants were prepared in the same fashion and also start with 5′ GMP. The G-1 base was inserted into the pUC18/tDNA Ser(TGA) or pUC18/tDNA Gly(CCC) plasmids through QuikChange, and G73 in tRNA Ser was further changed to C73 to pair with G-1.

Crystallization, data collection and structure determination

The cocrystals of HisRS-tRNA His-histidine-AMPPNP complex were grown using the hanging drop vapor diffusion method at room temperature. To form the complex, protein was mixed with tRNA His at a molar ratio 1:1:1 in a buffer containing 20 mM HEPES (pH 7.0), 150 mM NaCl, 5 mM MgCl2, 1 mM DTT, 2 mM histidine and 2 mM AMPPNP. The final protein concentration present in the complex was 2.0 mg/ml. The sample was filtered with a 0.22 μm centrifugal device (Millipore) prior to crystallization. The drops were equilibrated against a reservoir solution containing 5% PEG 4000 and 0.1 M NaOAc (pH 4.5). The cocrystals of the quaternary complex appeared after two or three days. Fully-grown crystals were flash frozen in liquid nitrogen after being soaked in a cryoprotectant containing all the reservoir solution components supplemented with 20% glycerol (v/v).

A 2.55 Å diffraction data set was collected using beamline 17U (BL17U) at the Shanghai Synchrotron Radiation Facility (SSRF, Shanghai, P.R. China) and was processed with the program HKL2000 (35). Cell content analysis suggested that the asymmetric unit contains one HisRS monomer in complex with one tRNA molecule. Molecular replacement (MR) was first performed with Phenix using the apo TtHisRS structure (PDB code 1H4V) and a partial structure of Candida albicans tRNA His (CatRNA His, PDB code 3WC1) (36) consisting of the acceptor stem and T/C region as the search models. The two components were then searched simultaneously to produce a partial solution. To obtain the complete complex structure, a second round of MR was carried out with the partial solution and residues 8–49 of human tRNA Gly/ Ser (PDB code 4KR2) (34) as the search models. The initial phases for both the protein and a complete tRNA molecule were evident once the solution was found. However, the insertion domain (Arg172-Glu319) was found to undergo large conformational changes and this region was manually rebuilt with the program Coot (37). The rebuilt model was fed to phenix.refine and multiple cycles of refinement were alternated with model rebuilding (38). Titration-Liberation-Screw (TLS) refinement was carried out at the late stage of the refinement using a total of 22 TLS groups. The final R-factor was 20.73% (R free = 26.17%) (Table 1). The Ramachandran plot of the final models had 96.05%, 3.46%, 0.49% of the residues in the most favorable, generously allowed and disallowed region as indicated by the program Molprobity (39). All the figures about structures were created with Pymol (www.pymol.org) and the domain architecture was prepared by DOG (40).

Aminoacylation assay

The histidylation assay mixture contained 50 mM Tris-HCl (pH 7.7), 100 mM KCl, 10 mM MgCl2, 8 mM DTT, 2.5 mM ATP, 20 μM L-[14C]-histidine (PerkinElmer, 320 mCi/mmol) and 5 μM annealed tRNA His(GUG). 50 nM HisRS or mutants were added to initiate the reactions. All the activity assays were performed with 5′-GMP tRNAs due to their higher activities as the substrates than that of the 5′-GTP tRNAs due to their higher activities as the substrates than that of the 5′-GMP tRNAs. The reaction was carried out at 60°C, and aliquots were removed at the designated time points, spotted onto 5% trichloroacetic acid (TCA)-soaked filter pads, and washed twice with 5% cold TCA as well as 95% ethanol. The filter pads were dried and the radioactivity was measured by scintillation counting.
The aminoacylation assays for hSerRS/hGlyRS were carried out under similar conditions to those of HisRS assays. hSerRS assays contained 50 mM Tris-HCl (pH 7.7), 100 mM KCl, 10 mM MgCl₂, 8 mM DTT, 2.5 mM ATP, 20 μM cold serine, 1 μM L-[³H]-serine (PerkinElmer, 22 Ci/mmol), 10 μM annealed tRNA¹ser(GUA), whereas that of hGlyRS contained 150 mM HEPES (pH 7.5), 20 mM KCl, 4 mM MgCl₂, 2 mM DTT, 3 mM ATP, 20 μM cold glycine, 1 μM L-[³H]-glycine (PerkinElmer, 48.7 Ci/mmol) and 5 μM annealed tRNA¹ser(Gly(CCC)). 200 nM hSerRS and 50 nM hGlyRS were added last, and all the reactions were performed at room temperature.

**RESULTS**

**Overview of the complex**

The cocrystal structure is determined at 2.55 Å resolution. The space group of the crystal belongs to C222₁, with a completeness of 97.86% and a Wilson B-factor of 48.7 Å². Gel filtration chromatography indicates that TtHisRS exists as a single species with an approximate mass of 80 kDa, suggesting a dimer in solution. Continuous electron density is observed for the polypeptide from Thr2 to Gly421, except for the disorder from Gly53 to Lys63. This region is located on a loop connecting ɣ1 and ɣ3. Ala57-Arg62 of this region is also unstructured in the apo TtHisRS structure, but it becomes fully ordered in the histidine- or HAM-bound structures (30). The tRNA model is almost intact except for nucleotide A38 and the map shows clear electron density for the triphosphate group of G-1. The final model contains one AMP (coming from AMPPNP), one histidine and 76 water molecules (Table 1).

The protein structure in the complex is similar to that of the apoprotein. The N-terminal catalytic domain (Met1-Asp171 and Leu229-Gly320) is formed by six-stranded antiparallel β-sheet surrounded by three long α-helices. This domain contains the necessary functional elements for HAM synthesis as well as tRNA¹his aminoacylation (41). In addition, the extra domain (Arg172-Phe228) splits the catalytic domain. The C-terminal domain (Leu331-Gly421) displays the typical α/β fold shared by the Class Ia enzymes (Figure 1A and B).

HisRS is an α₂ homodimeric enzyme and the molecular dimer axis coincides with the crystallographic 2-fold axis. tRNA¹his bind to the protein dimer in a symmetrical fashion, interacting with both subunits (Figure 1C). However, the majority of the contacts come from subunit one, and with the active site and the anticodon binding domain, burying a surface area of 3706.4 Å². The contacts with subunit two are mainly between G-1 and the 70′DRG72′ tripptide (the primes indicate the residues from another subunit), and between the D-stem end (G10, A26 and C27) and the Gly94′-Pro99′ fragment, resulting in an additional buried surface area of 779.9 Å².

**Recognition of acceptor stem**

The most interesting feature of tRNA¹his is the presence of the G-1 nucleotide. Identification of the elements responsible for the specific recognition of the critical G-1:C73 base pair holds the key to the histidylation mechanism. The interactions at the 5′ end of tRNA are mainly on G-1 and G1, whereas at the 3′ end, contacts are on nucleotides C68-A76. Surprisingly, most of these interactions are nonspecific and electrostatic interactions between positively charged residues and the backbone phosphates of tRNA (Supplementary Table S3). tRNA¹his used for crystallization begins with a guanosine triphosphate and three salt bridges are formed between the β- and γ-phosphate oxygens and Arg7 as well as Arg122. Arg122 is absolutely conserved and Arg7 is usually replaced by a similar lysine residue in other organisms (Supplementary Figure S1). These interactions are not biologically relevant because the natural tRNA¹his bears a 5′ GMP. Besides recognizing the γ-phosphate oxygen, Arg7 also forms a salt bridge with C68 backbone phosphate. The most critical interactions may come from the two salt bridges with Arg115 and an additional one from Arg122, between their guanidino groups and the oxygen atoms of G-1 α-phosphate. Furthermore, Arg115 donates a hydrogen bond to the O6 and N7 atoms of G-1, respectively, through its main chain nitrogen, and the former allows the enzyme to specifically recognize G-1 instead of A-1. Aside from the interactions with subunit one, the subunit forming the majority of protein-RNA contacts, G-1 also accepts four more hydrogen bonds and two salt bridges from the 7′RGGR74′ tetrapeptide across the subunit. However, two hydrogen bonds on the base ring involve Arg7′, whose side chain is not well defined in the electron density map. In addition, Arg7′ NH1 atom is relatively far from its hydrogen bonding partner N9 of G-1 (3.48 Å), therefore the contributions from these interactions may be rather weak (Figure 2A).

Other than the contacts on G-1, G1 also makes three hydrogen bonds with Arg115 side chain through its purine ring, as well as a salt bridge through its backbone phosphate (Figure 2B). We mutated the arginines to alamines individually and discovered that the both the R7A and R115A mutations reduce enzymatic activities by ~80%, while the R122A mutant is only 15% as active as the WT, attesting to the significance of G-1 interactions (Figure 2C and Supplementary Figure S2).

Contacts on the 3′ end are also mainly nonspecific, and the key residues responsible for the recognition are concentrated in three regions. Gln117, Lys118, Arg120 and Glu114 from the conserved motif 2, interact with C70-C72, C74 and A76 (Figure 2B). Gln117 donates two hydrogen bonds to A71 and C72 backbone oxygens using its NE2 atom, while Lys118 hydrogen bonds to C70 and A71 using its main chain nitrogen. The invariant Arg120 is responsible for the specific recognition of C74 while the semiconserved Glu114 makes specific contacts with A76 purine ring. The second region responsible for C72-C75 recognition (Arg197, Arg204 and Lys209) is located in the α5 and ɣ4 helices within the insertion domain, and these contacts are mostly salt bridges between the side chains of positively charged residues and tRNA backbone phosphates. Among these, the NZ atom of the highly conserved Lys209 electrostatically interacts with C72 phosphate oxygen, whereas Arg197 and Arg204 make a total of nine salt bridges with C73-A76. However, neither arginine is conserved. The R197A mutant retains more than half of the WT activity while the activity...
of K209A is comparable to that of WT. The third region involves residues Gly260, Tyr264 and Ala278, and is between the insertion domain and motif 3. Gly260 and Ala278 form nonspecific hydrogen bonds using their main chain nitrogens. The 3′ terminal A76 alone makes seven hydrogen bonds, three of which are donated from the 6-exocyclic amino group to the carboxylate group of Glu114, and to the main chain carbonyl oxygen of Glu64.

Consequently, the acceptor stem is clamped between the insertion domain, and the Arg115/Arg122 arginine pair. Surprisingly, most interactions on the G-1:C73 base pair are neither specific nor from conserved residues. In contrast, G1, C74 and A76 each form multiple base-specific hydrogen bonds. We introduced mutations to these residues to test their importance (except for residues with main chain contacts). We found that while Glu64 and Glu114 are absolutely essential for aminoacylation, the rest residues are dispensable or play minor roles in substrate recognition (Figure 2C). Therefore the contribution to aminoacylation from these nonspecific interactions is negligible. Nonetheless, these nonspecific interactions together with the contacts on the essential identity determinants, help to adjust and place the acceptor end of tRNA.
Recognition of anticodon loop

The anticodon binding domain in aaRSs is responsible for recognizing the anticodon loop. In our cocystal structure, it forms the most specific hydrogen bonds with the anticodon trinucleotide. This domain approaches the anticodon loop from the major groove (Figure 1B and C). Bases of U33 to G37 are splayed out, and the GUG triplet forms extensive interaction network with the enzyme (Figure 2D). The residues for the anticodon recognition mainly cluster in the α12-β15 region, and all the interactions are base-specific. Particularly, the wobble base G34 forms six hydrogen bonds with residues Glu386-Glu388 through the base ring, two of which are with the main chain nitrogens of Glu386 and Asp387 respectively. U35 forms three hydrogen bond contacts with the side chain of Gln404. G36 forms two hydrogen bonds with the main chain carbonyl oxygen of Lys399 and one with the side chain of Lys397. Activity tests show that the two most severe mutations are K397Q and E388A, which dropped the activities by 9- and 5-fold, respectively, whereas the Q404A activity is on a similar level to that of WT (Figure 2C). The weak contribution to aminoacylation from the C-terminal domain is similar to its *Escherichia coli* counterpart, in which the catalytic domain of the enzyme alone retains ~0.6% aminoacylation efficiency of that of the full-length enzyme (41).

In contrast to the anticodon loop, the region outside of the anticodon is largely flexible. Despite the relative high resolution of the crystal structure, the tRNA molecule shows poor electron density in this region except for the 34GUGG37 fragment, and A38 is completely disordered. We also obtained another form of cocystal structure at a lower resolution, which contains a TtHisRS dimer bound by two tRNAs in the asymmetric unit (Tian et al., unpublished). The dimeric complex displays even larger flexibility in the anticodon loop and only nucleotides G34-G37 are visible.

Active site

Although AMPPNP was added during the crystallization process, only the AMP moiety is found present in the final refined structure for reasons that are not clear to us. Additionally, the histidine molecule is not fully covered by the map and its main chain density is not evident either in the 2Fo-Fc or Fo-Fc map (Figure 2E and Supplementary Figure S3). On the other hand, the imidazole ring occupies a similar position to the tRNA-free structures (PDB code 1ADJ and 1ADY), forming two hydrogen bonds with the terminal carboxylate group of the motif 2 residue Glu130. In addition, Thr83 also forms two more hydrogen bonds with the ring. These contacts hold the side chain of histidine from the side and top, and Gly304 and Phe305 form the bottom
of the imidazole-binding pocket, adding two more hydrogen bonds (Figure 2E). However, the main chain of histidine is in a different orientation from previous histidine-bound structure. Particularly, the amino group points to the A76 ribose as well as the AMP phosphate, and makes two hydrogen bonds with the latter; the carbonyl oxygen is directed towards histidine-2 peptide (30) and are involved in two hydrogen bonds with the main chain of the Gly284Gly285 dipeptide. The orientation of histidine may not represent its genuine intermediate state. Then the nonproductive stalemate dipeptide located on a loop as well as the small size for glycine residues Gly72 Gly73 across the subunit. These two glycines are very conserved across species and the accommodation of G-1 may originate from the flexibility of the dipeptide located on a loop as well as the small size for glycine (Figure 4A, B and Supplementary Figure S7). To explore the tolerance capacity of the pocket, we reduced its

Role of the interdomain linker

HisRSs contain a 9–14-residue interdomain hinge, which links the two otherwise stand-alone domains. The N- and C-terminal domains within the same subunit barely make any direct contacts with or without tRNA. Judging from several bacterial HisRS structures, this hinge appears to be relatively tense and little changes occur to this region after the binding of tRNA substrate. One would speculate that this loop bears much tension in order to communicate between the two domains. Bacterial HisRSs usually possess shorter linker sequences whereas eukaryotic HisRSs have an insertion (a TTET motif in vertebrates) at the end of the linkers (Supplementary Figure S1). To assess the importance of the linker length, we intentionally shortened (by removing the 325EEKG328 tetrapeptide) or extended (by inserting the TTET sequence after Gly328) this loop and tested activities. We discovered that the insertion mutant retains ≈3/4 activity of that of WT and the deletion mutant only 1/5 (Figure 3C). We also noticed that the deletion of the EEKG motif is highly detrimental to protein stability because the deletion mutant failed to form a peak on the size exclusion column, indicating some folding problems. Thus, the hinge joining the two domains is import for the crosstalk between the two domains. Structural alterations resulting from the anticodon binding domain are coupled to those of active site, caused by substrate binding via this unique linker. Through evolution, the length of the linker has attained an optimal value. While too little distance causes serious clashes between the N- and C-terminal domains, too much distance reduces the communication efficiencies between them. The two domains act as rigid units, and the linker length may affect directly their relative positions and orientations, which in turn affect the interactions with the tRNA substrate and hence aminoacylation.

The binding pocket of G-1

The binding pocket of G-1 is composed of two important glycine residues Gly72 Gly73 across the subunit. These two glycines are very conserved across species and the accommodation of G-1 may originate from the flexibility of the dipeptide located on a loop as well as the small size for glycine (Figure 4A, B and Supplementary Figure S7). To explore the tolerance capacity of the pocket, we reduced its
size by creating the G72V single- as well as the G72VG73V double mutants, and the later has proved to reduce the enzymatic activities by more than 4/5 (Figure 3C). Structural comparison with other class IIa synthetases in complex with their cognate tRNAs reveals that such a pocket is probably only unique to HisRS. We therefore created models of tRNA^{His} in complex with other class IIa enzymes by superimposing the catalytic domains of hGlyRS (34), hSerRS (42) or EcThrRS (43), respectively, onto that of tRNA-bound TtHisRS. In our models, G-1 of tRNA^{His} clashes into Glu323′Asn324′ as well as the motif 2 residues around Glu365 in EcThrRS (PDB code 1QF6), Gly233′Pro234′ as well as the motif 2 residues around Ile280 in hGlyRS (PDB code 4KR2), and Lys253′Ser262′ as well as the moti-
tRNA\textsubscript{His} was used in the activity assays due to its report- and mutants were tested for activities. G(-1)MP-containing Structure-based mutagenesis was subsequently conducted with much more information on the structural details. edly higher activity than the GTP-form of tRNA\textsubscript{His}. Our by showing that SerRS or hGlyRS only retained cylation assays also lent support to these theoretical models, activity for their G-1-containing tRNA substrates, compared to their WT tRNA substrates, respectively (Figure 4F).

**DISCUSSION**

The substrate recognition of tRNA\textsubscript{His} is an outstanding problem that has remained unclear for many years. In this work, we determined the crystal structure from TtHisRS-tRNA complex and studied the structural basis of histidyla- tion. Our crystal structure is in good agreement with the modelling studies carried out by Åberg et al. (30) but with much more information on the structural details. Structure-based mutagenesis was subsequently conducted and mutants were tested for activities. G(-1)MP-containing tRNA\textsubscript{His} was used in the activity assays due to its report- edly higher activity than the GTP-form of tRNA\textsubscript{His}. Our pilot experiments show that appropriate enzyme concentra- tions for initial velocity measurements range from 35 to 70 nM (Supplementary Figure S8). Hence we fixed the concentra- tion of the enzyme at 50 nM throughout the assays. A few conclusions can be drawn from our crystallographic and biochemical studies.

First, the recognition of tRNA\textsubscript{His} is principally a type of backbone recognition, in which the -1:73 base-pairing geometry plays an important role. The enzyme makes very few specific contacts with tRNA except for the anticodon triplet. The interactions on the acceptor stem are mainly on the 3’ end while those on the 5’ end are concentrated on G-1 and G1. Most charged electrostatic interactions be- tween conserved Arg/Lys residues and the backbone phosphates bear little consequences on aminoacylation. Additionally, many polar interactions are formed by the main chains of the protein. The only specific recognition on G-1 arises from the hydrogen bond through O6, allowing it to be possibly distinguished from A-1 by the enzyme. Interest- ingly, we identified two semi-conserved residues Glu64 and Glu114 that cause the most activity loss. Glu64 is located at the tip of a disordered loop that may contact with the minor groove of the acceptor stem and its only direct contact with the tRNA molecule is a possible hydrogen bond with the N6 atom of A76. The density of neither hydrogen-bonding partner is clearly resolved, similar to the case of Glu114. Glu114 is located to an adjacent loop that contacts with the major groove. The two glutamate residues are spatially close and may form some interactions (Supplementary Figure S9). The reasons for the significant effects on histidyla- tion from these two residues are not clear, and whether this phenomenon is TtHisRS-specific awaits further investiga- tion.

Second, the conformational changes for protein upon the formation of the complex are not dramatic, with detectable rearrangements occurring in the insertion domain. There are also several local structural changes induced by tRNA binding that lead to the ordering or disordering of several flexible loops. However, these conformational changes are relatively small, compared to those incurred by the binding of histidine or synthesis of HAM. Consequently, we propose that the synthesis histidyl-adenylate is likely to be the rate-limiting step in histidyla- tion, to solicit the large conformational changes needed for the first step. Only minor rearrangements ensue in the insertion domain as well as in other regions in order to complete the reaction.

In contrast, the anticodon of tRNA becomes highly mobile upon the formation of complex. In the structure of CatRNA\textsubscript{His} in complex with its processing enzyme Thg1, the anticodon loop is intact and its B-factor is on a similar level to that of entire tRNA (36). Therefore, the flexibility of the anticodon loop is a consequence of the induced fit brought about by the binding of tRNA\textsubscript{His}. The high flexi- bility of anticodon loop from both cocrystal structures sug- gests that this region is dispensable for aminoacylation, and offers an explanation for the observation of mini- or micro-helices or even transplanted yeast tRNA\textsubscript{Asp} being fair sub- strates for HisRS (22). It is reasonable to deduce that the heavy reliance of HisRS on G-1 as the critical identity deter- minant is a safety mechanism, compensating for the weak contribution of the anticodon to aminoacylation.

Third, the accommodation of the extra base G at the -1 position appears to be related to the enlarged substrate-binding pocket of HisRS for tRNA\textsubscript{His} at the very 5’ extremity. The conserved Gly72 Gly73’ dipeptide from across the subunit allows G-1 to fit, while the steric hindrance from hGlyRS, hSerRS and EcThrRS excludes the entry of their cognate tRNAs with an implanted G-1 base. This hypothe- sis is confirmed by the aminoacylation activities of hSerRS and hGlyRS for G-1 tRNAs, which reduced 2/3 activities of the enzymes.

*Caulobacter crescentus* HisRS (CcHisRS) is reported to have a G-1-less tRNA\textsubscript{His}. The sequence alignment indicates that a pair of aspartates substitute the conserved G72’G73’ dipeptide and thus blocks the binding site of G-1 (Supplementary Table S3). The covariation of both the enzyme/tRNA allows for the unusual histidyl-adenylate system lacking G-1 recognition, but the activity of CcHisRS is considerably lower than HisRSs from other organisms, due to the loss of interactions with G-1. To compensate for this loss of substrate binding, CcHisRS has to rely more on the anticodon recognition (28).

**ACCESSION CODES**

The atomic coordinates and structure factors have been de- posited in the Protein Data Bank with the accession code 4RDX.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENT**

We thank Shanghai Synchrotron Radiation Facility (SSRF) for access to beamline 17U (BL17U).

**FUNDING**

National Sciences Foundation of China [31100579]; Guangdong Innovative Research Team Program...
REFERENCES
